Demand Estimation with Unobserved Choice Set Heterogeneity*

Gregory S. Crawford† Rachel Griffith‡ Alessandro Iaria§

April 13, 2017

Abstract

We present a method to estimate preferences in the presence of unobserved choice set heterogeneity. We build on the insights of Chamberlain's Fixed-Effect logit and exploit information in observed purchase decisions in either panel or cross-section environments to construct "sufficient sets" of products that enable us to "difference out" the true but unobserved choice sets from the likelihood function. We can then recover preference parameters without having to specify the process of choice set formation. We illustrate our ideas by estimating demand for chocolate bars on-the-go using individual-level data from the UK. Our results show that failing to account for unobserved choice set heterogeneity can lead to statistically and economically significant biases in the estimation of preference parameters.

*Griffith gratefully acknowledges financial support from the European Research Council (ERC) under ERC-2009-AdG-249529 and ERC-20150AdG-694822, and the Economic and Social Research Council (ESRC) under RES-544-28-0001 and ES/N011562/1. Iaria gratefully acknowledges financial support from the research grant Labex Ecodec: Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

†Dept. of Economics, University of Zürich and CEPR, gregory.crawford@econ.uzh.ch

‡University of Manchester and IFS, rgriffith@ifs.org.uk

§CREST, ENSAE, Université Paris-Saclay, and IFS, alessandro.iaria@ensae.fr

1 Introduction

Traditional approaches to estimating demand in a discrete choice framework pay a lot of attention to preference heterogeneity across consumers, but tend to rely on complete knowledge of the choice set from which decision-makers are selecting (e.g., Berry et al. (1995), Nevo (2001)). The theoretical literature is rich with reasons why choice sets may be restricted and heterogenous across consumers, including limited consumer attention, search, or endogenous product choice. It is often the case that these choice sets will be unobserved to the econometrician. Recent advances in the applied literature have accommodated choice set heterogeneity by bringing in additional information on their distribution across consumers (Goeree (2008), Van Nierop et al. (2010)) so that they can intergrate out over the distribution of unobserved choice sets following Manski (1977). However, there are cases when we may not have knowledge of the process of choice set formation, and so the distributional assumptions become ad hoc, and empirical methods to estimate preferences in this situation would be useful.

Our contribution in this paper is to propose an approach that builds on the insights of Chamberlain (1980)'s Fixed-Effect logit and allows us to difference out the unobserved choice set heterogeneity and obtain consistent estimators of preference parameters. We propose a broad class of conditional logit models that exploits information in observed purchase decisions to construct "sufficient sets" of products that enable us to "difference out" the true but unobserved choice sets from the likelihood function. These sufficient sets can be tightly linked to theory, including limited consumer attention (Eliaz and Spiegler (2011), Matejka and McKay (2015), Goeree (2008)), consumer search (De los Santos et al. (2012), Dinerstein et al. (2014)), screening rules (Gilbride and Allenby (2004)), regulatory restrictions on choice sets (Gaynor et al. (2016), Walters (2014)), and firms' endogenous product choices (Eizenberg (2014), Draganska et al. (2009)).

Our estimation procedure is broadly applicable and can accommodate a wide variety of economic situations that may lead to unobserved choice set heterogeneity in the data, including limited consumer attention (Eliaz and Spiegler (2011), Matejka and McKay (2015), Goeree (2008)), consumer search (De los Santos et al. (2012), Dinerstein et al. (2014)), screening rules (Gilbride and Allenby (2004)), regulatory restrictions on choice sets (Gaynor et al. (2016), Walters (2014)), and firms' endogenous product choices (Eizenberg (2014), Draganska et al. (2009)). In empirical applications, papers in these literatures generally specify a particular choice set formation process and show the importance of accounting for choice set formation on estimated preferences. The method we propose in this paper is complementary in that we provide methods to test alternative assumptions in the data. While our approach requires the properties of the logit model, it can accommodate flexible forms of unobserved preference heterogeneity such as nested logit and mixed logit (random coefficients).

To build intuition, we provide examples of sufficient sets and how these resolve the issue of unobserved choice set heterogeneity. We show the usefulness of our approach in both monte carlo exercises and an empirical illustration. The illustration estimates price and advertising sensitivity for on-the-go demand for chocolate among adult women in the UK. We estimate the model using a multinomial logit of choice from the "universal choice set", consisting of all products in the market, which is common practice in the demand literature. We compare that to estimates from a conditional logit model which relies on one of the sufficient sets we propose in the paper: a household's past purchase history. We show that incorrectly including alternatives in households' choice sets biases estimated price sensitivities towards zero and estimated advertising sensitivities away from zero, results consistent with theories of choice set formation that highlight imperfect consumer attention in choice environments. It also biases upwards estimated willingness-to-pay for advertising exposure and estimated willingness-to-pay for individual products.

While our illustration uses panel data, we reiterate that the methods developed in the paper apply in any situation where observed choices allow us to infer a *subset* of products that are in consumers' true choice sets. In a cross-section, this could be similar individuals making a decision in a common economic environment. For example, consider the question of whether greater availability of fast-food outlets causes obesity. We would like to be able to identify whether it is the availability of fast-foods that leads to increased demand, or whether preferences are the driving factor. We could collect precise geographic data on the location of fast food outlets and where children live and attend school (as in Currie et al. (2010)), but we still might not be sure about which outlets lie within individual children's choice sets. Using our approach, we could use information on choices made by other children living on the same street, and/or attending the same school, to identify which outlets are in the choice sets of different children.

Our work relates to several literatures in economics and marketing. There is a fast-growing theoretical literature in which limited attention is used to rationalize apparently incongruent consumer and firm behaviors. This literature motivates our interest in developing empirical approaches that accommodate these theories. Eliaz and Spiegler (2011) propose a model of boundedly rational consumers that make purchase decisions from subsets of all the available products in the market, and explore how firms can use marketing to manipulate these subsets. Gabaix (2014) presents a general model of rational inattention. In a first stage, an agent optimally reduces the complexity of a choice situation, while in a second stage she makes a final decision in this simplified environment. The simplification of the decision process may take various forms: from reducing the set of alternatives to be considered, as in models of search (see De los Santos et al. (2012)) and screening rules (see Gilbride and Allenby (2004)), to ignoring payoff-relevant product characteristics as in models of salience (see Bordalo et al. (2014)) and focus (see Kőszegi and Szeidl (2013)). In a related paper, Matejka and McKay (2015) also model rational consumer inattention. They assume consumers have imperfect information about payoffs but, before choosing, may gain information at a cost. Using information theory to determine consumers' optimal information-processing strategies, they show that they result in probabilistic choices that follow a generalized logit form: the probability of a consumer selecting a product depends on her prior beliefs about that product, the true payoff it would provide, and a parameter that scales with the cost of information.

Our contribution also relates to the literature on the identification of preferences and consideration sets uniquely from choice data. In a general model of consumer inattention, Masatlioglu et al. (2012) explore whether and how preferences and attention (or consideration) can be separately identified when both are non-stochastic, while choices and choice sets are observed. If, for example, a consumer's choice x changes when a product y is removed from the choice set, then it must be that both y was considered ("Revealed Attention") and that she prefers x over y ("Revealed Preference"). Manzini and Mariotti (2014) observe that the deterministic nature of the model proposed by Masatlioglu et al. (2012) may be at odds with stochastic choice data, and extend the framework to allow for stochastic consideration. They prove an important "lack of identification" result, namely that a choice model with unconstrained choice set formation process does not lead to any observable restriction on choice data alone (neither consideration nor preferences). In order to identify choice models with unobserved choice sets uniquely from choice data, one needs to make functional form assumptions about the choice set formation process. Along these lines, Abaluck and Adams (2016) propose a set of mild assumptions on the choice set formation process, which are commonly employed in empirical work, that guarantee the separate identification of consideration from utility. Our work is complementary to these papers focusing on *identification*, in that we provide a tractable empirical approach for the consistent *estimation* of consumer preferences when choice sets are heterogeneous and unobserved.

Our contribution relates to two empirical literatures in economics and marketing focused on demand estimation. The first is the discrete-choice demand literature which has developed widely-used methods for estimating demand using either aggregate or individual-level data (Berry et al. (1995), Nevo (2001), Train (2009)), but which does not specifically address unobserved choice set heterogeneity, instead assuming that all consumers have access to the same choice set, often consisting of all the products in the market or the highest market share products. The second is the empirical literature built on Manski (1977) that specifies particular models of choice set formation and estimates a consumer's *unconditional* purchase probability for product j as the sum across all possible choice sets including j of the purchase probability for j given that choice set times the probability of that choice set (Chiang et al. (1998), Goeree (2008), Van Nierop et al. (2010), Draganska and Klapper (2011), and Conlon and Mortimer (2013)). This requires additional information with respect to standard choice data, for example knowledge of consumers' actual choice sets or information that predicts choice set variation. Our approach is complementary to both of these literatures. When researchers have individual-level data and additional information on the choice process is not available, authors can use our approach; when this information is available, authors can use our approach to test the assumptions maintained in their preferred approach.

The closest papers to ours are Lu (2016) and Fox (2007). Lu (2016) proposes a method for demand estimation under unobserved choice set heterogeneity that also does not require the specification of the choice set formation process. His insight is that an individual's true choice probability must be bounded above by a choice probability computed on a subset of the true choice set, and below by a choice probability computed on a superset of the true choice set. This allows him to set-identify preferences by making assumptions, or by having additional data, on supersets and subsets of the unobserved choice sets faced by each individual in each choice situation. Our approach is complementary. Lu (2016) does not specify how to construct valid choice subsets, but researchers could rely on the sufficient sets we introduce in this paper. More broadly, Lu's methods requires weaker assumptions than the logit family on the nature of unobserved preference heterogeneity, but may lack power and/or face computational difficulties as the scale of the application grows. By contrast, our approach is fast and can be conveniently implemented using standard econometric software irrespectively of the scale of the application. Fox (2007) proposes a method for the estimation of preference parameters from restricted choice sets, and in a short extension also shows that, if products with higher systematic utilities are not only more likely to be chosen when available, but also more likely to be included in individuals' choice sets, his estimator is also able to deal with unobserved choice set heterogeneity. Compared to our approach Fox's method allows more flexibility in preferences, but requires more structure on the relationship between preferences and the choice set formation process, and is also more difficult to implement.

The rest of the paper is structured as follows. In section 2.2 we describe the problems that arise from unobserved choice set heterogeneity. We show that estimators of demand parameters that mistakenly impute to consumers alternatives not originally available in their true choice sets will be biased. We quantify this bias in Monte Carlo simulations. In section 2.3, we derive a class of conditional logit estimators that are based on "sufficient sets" of consumer choices. We show how the estimator can accommodate unobserved preference heterogeneity, some examples of sufficient sets (section 3.1), how to test between alternative sufficient sets (section 3.2), and how we can point-identify preferences and bound choice probabilities, elasticities, and consumer surplus (section 4). In section 5, we discuss some examples from the broad class of models that generate choice data compatible with the estimator we propose in this paper, including models of limited consumer attention, search, and endogenous product choice. In section 6, we present an illustrative example that looks at the impact of price and advertising on demand for on-the-go chocolate purchases in the UK. A final section concludes and discusses possible extensions and several appendices provide additional detail, auxiliary findings, and further examples.

2 Bias from Unobserved Choice Set Heterogeneity and Proposed Solution

In this section we introduce the primitives of the choice environment and characterize the mechanism by which unobserved choice sets introduce an estimation bias. Building on this characterization, we then propose a simple solution to the problem of estimating demand with unobserved choice sets.

2.1 Preference Fundamentals

Let there be $i = 1, \ldots, I$ *(types of) decision makers* and for each type $t = 1, \ldots, T$ different *choice opportunities*. For example, this could be one individual (i) observed to make several separate decisions over time ($t = 1, \ldots, T$), or several individuals ($t = 1, \ldots, T$) of the same type (i) each making a separate decision at a single point in time. Denote i's "sequence" of choices by $Y_i = (Y_{i1}, \ldots, Y_{iT})$. While sequence suggests choices that vary over time, we emphasize here that a sequence could be a group of decisions made by the same type of decision maker at a single point in time. For simplicity, we assume to observe exactly T choice situations for each i, but this is trivially relaxed.

We consider a situation in which i is matched to (unobserved) choice set $CS^{\star}_{it}$ in choice situation t. We are interested in the estimation consequences of mistakenly imputing to i in t an incorrect choice set, E_{it}, such that $CS^{\star}_{it} \neq E_{it}$.

Denote by $\times$ the cartesian product and let i's set of possible choice sequences be given by $\mathcal{CS}^{\star}_i = \times_{t=1}^{T} CS^{\star}_{it}$. By construction, any observed choice sequence Y_i must belong to $\mathcal{CS}^{\star}_i$. Similarly, let the incorrect set of choice sequences be denoted by $\mathcal{E}_i = \times_{t=1}^{T} E_{it}$.

Let preferences be defined by a parameter vector θ and the probability with which i is matched to a given set of possible choice sequences, $\mathcal{CS}^{\star}_i = c$, be given by $\Pr\left[\mathcal{CS}^{\star}_i = c \middle| \gamma\right]$, with γ a vector of parameters governing this process. In principle, γ could include some or all of the parameters that are in θ and, as described in detail in Section 5, could be a result of limited consumer attention, consumer search behavior, or stategic decision-making by firms.

Given θ and a specific match with a set of possible choice sequences, $\mathcal{CS}^{\star}_i = c$, each consumer type i is observed to make a sequence of choices Y_i. We assume that the conditional indirect utility of alternative j in choice situation t for consumer type i is

$$U_{ijt} = V\left(X_{ijt}, \theta\right) + \epsilon_{ijt}, \tag{2.1}$$

where $X_{ijt}]$ is a vector of observable characteristics, while ϵ_{ijt} is a portion of i's utility that is unobserved to the econometrician. For expositional reasons we start by considering the function $V()$ to be common across i, and in section XXX we introduce unobserved preference heterogeneity, and show that we can accomodate flexible forms of unobserved preference heterogeneity such as nested logit, some forms of mixed logit, or multinomial logit with individual-alternative-specific fixed effects.

The probability that i is matched to the set of possible choice sequences $\mathcal{CS}_i^\star = c$ *and* makes a sequence of choices $Y_i = j$ is then

$$\Pr\left[Y_i = j, \mathcal{CS}_i^\star = c \middle| \theta, \gamma\right] \quad = \Pr\left[Y_i = j \middle| \mathcal{CS}_i^\star = c, \theta\right] \Pr\left[\mathcal{CS}_i^\star = c \middle| \gamma\right]. \tag{2.2}$$

Equation (2.2) is quite general and captures two features of behavior typical of demand. The first term embodies preferences given the choice set a consumer is matched to, and the second the probability that the consumer is matched to that choice set. In addition, $\Pr\left[\mathcal{CS}_i^\star = c \middle| \gamma\right]$ can take any form and be a function of any element of $\Pr\left[Y_i = j \middle| \mathcal{CS}_i^\star = c, \theta\right]$.

We next make the following important assumption:

Assumption 1. *Conditional on all $V(X_{ijt}, \theta)$'s and on $\mathcal{CS}_i^\star = c$, ϵ_{ijt} from (2.1) is distributed i.i.d. Gumbel.*

Assumption 1 implies that $\Pr\left[Y_i = j \middle| \mathcal{CS}_i^\star = c, \theta\right]$ from (2.2) is logit for any c. This assumption is ubiquitous in empirical work, both among papers in the discrete-choice demand estimation literature,[1] and among papers that focus on demand estimation with unobserved choice sets.[2] This is because Assumption 1 allows very general matching processes, $\Pr\left[\mathcal{CS}_i^\star = c \middle| \gamma\right]$, as long as choice probabilities belong to the logit family given any such process. For example, this framework accommodates models in which firms select the products to sell or in which consumers search for products on the basis of both observable characteristics and expectations over unobservable characteristics. Various models of unobserved preference heterogeneity in $V(X_{ijt}, \theta)$ fit within this framework, as we will discuss in section (2.4) (e.g., nested logit, fixed effect logit, and mixed logit with discrete distribution of random coefficients). A useful implication of Assumption 1 is that *conditional* Maximum Likelihood estimators of θ can be constructed from $\Pr[Y_i = j | \mathcal{CS}_i^\star = c, \theta]$, since this conditional probability retains a logit form.[3]

Assumption 1 would be violated if the matching of individuals to choice sets depended on the *realization* of the unobservables, ϵ_{ijt}'s. To our knowledge, the only paper in the applied literature that analyzes choices in a revealed preference framework that allows for such an environment is De los

[1]For example, (Berry et al., 1995, p864-868) and (Berry et al., 2004, p76)

[2]For example, (Draganska and Klapper, 2011, p660), (Draganska et al., 2009, p110), and (Goeree, 2008, p1025).

[3]Given Assumption 1, failing to control for the choice set matching process, $\Pr\left[\mathcal{CS}_i^\star = c \middle| \gamma\right]$, only causes losses of efficiency in the resulting *conditional* Maximum Likelihood estimator with respect to the *joint* Maximum Likelihood estimator derived from (2.2).

Santos et al. (2012), who estimate a non-sequential search model for the online purchase of books where the selection of websites a consumer visits depends on her realized epsilon draws, ϵ_{ijt}.[4]

In section 5, we discuss several broad classes of economic models that generate choice data that are compatible with Assumption 1.

2.2 Unobserved Choice Set Heterogeneity and IIA Violations

Using Assumption 1, i's conditional probability of selecting choice sequence $Y_i = j$, given their set of possible choice sequences, $\mathcal{CS}_i^\star = c = \times_{t=1}^T c_t$, is:

$$\Pr\left[Y_i = j \middle| \mathcal{CS}_i^\star = c, \theta\right] = \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in CS_{it}^\star = c_t} \exp\left(V\left(X_{ijt}, \theta\right)\right)}. \tag{2.3}$$

Because $\mathcal{CS}_i^\star = c$ is unobserved, suppose that the researcher instead specifies the likelihood function to be used in estimation on the basis of the conditional probability of i choosing Y_i from choice set $\mathcal{E}_i = e = \times_{t=1}^T e_t$:

$$\Pr\left[Y_i = j \middle| \mathcal{E}_i = e, \theta\right] = \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ijt}, \theta\right)\right)}{\sum_{m \in E_{it} = e_t} \exp\left(V\left(X_{ijt}, \theta\right)\right)}, \tag{2.4}$$

where the difference between equations (2.3) and (2.4) lies in the summations in their denominators.

Proposition 1. Given true model (2.3), the likelihood function obtained from model (2.4) will mistakenly ignore a sequence of (i,t)-specific fixed effects if and only if at least one choice set $E_{it} = e_t$ of the sequence $\mathcal{E}_i = e$ includes at least one alternative not originally available in $CS_{it}^\star = c_t$.

Proof. See Appendix A.

Proposition 1 shows that, if the data were generated by model (2.3), then choice model (2.4), used in estimation, will differ from the *true* choice model conditional on $\mathcal{E}_i = e$ (model (A.1) from Appendix A) if and only if $\mathcal{E}_i = e$ contains choice sequences not originally included in $\mathcal{CS}_i^\star = c$. Choice set *restriction* is fine (see McFadden (1978)), while choice set *enlargement* is an issue. Specifically, choice set enlargement is a problem because, *even* in logit models, it represents a violation of the IIA property.

[4] However, they show that the economic consequences for their results of this assumption relative to the more standard Assumption 1 are modest (see Table 9, columns 2 and 3 and Table 12 in their paper).

There is no easy solution to this problem among methods commonly used by applied researchers. Alternative-specific constants or random coefficients will not control for the fixed effects because they are *individual-time* specific and correlated with X_{ijt} (see numerator of equation (A.2) in Appendix A). What can work in principle is the approach of Manski (1977), which models the unconditional probability i chooses j by integrating over all possible choice sets that include j. This is akin to treating unobserved choice set heterogeneity in a manner analogous to unobserved preference heterogeneity. Indeed, this is the general approach taken by much of the applied literature (e.g. Goeree (2008), Van Nierop et al. (2010)).

There are two issues, however, that may constrain the use of this approach in practice. First, because the set of all possible choice sets grows exponentially in the number of products, integrating them out poses a large computational burden. Second, partly to alleviate this computational burden and partly to aid identification, researchers often assume strong functional form restrictions on the choice set generating process, $\Pr[\mathcal{CS}_i^\star = c|\gamma]$. Appendix B provides further details about these problems and illustrates them in an application to the model of Goeree (2008).

2.3 Proposed Solution: Difference Out Choice Sets

Most of the existing literature deals with the problem of unobserved choice set heterogeneity by *integrating out* over the distribution of choice sets. This is the favored route whenever additional information on the choice set generation process is available and the computational burden implied by the estimation of the full model can be overcome. However, Proposition 1 shows that the econometric issue introduced by unobserved choice set heterogeneity can be characterized as a violation of the IIA property, *even* in simple logit models. We therefore suggest to directly prevent IIA violations by *differencing out* unobserved choice sets. Without having access to additional data, this can be achieved by using i's *observed* choice sequence $Y_i = j$ to construct sufficient statistics for the *unobserved* $\mathcal{CS}_i^\star = c$. Specifically, consider any correspondence $f(Y_i)$ that satisfies the following property.

Condition 1. Given any choice sequence $Y_i \in \mathcal{CS}_i^\star$, the correspondence f is such that $f(Y_i) \subseteq \mathcal{CS}_i^\star$.

In what follows, we show that if Assumption 1 and Condition 1 hold, then $f(Y_i)$ will be a sufficient statistic for $\mathcal{CS}_i^\star$ or, equivalently, a model conditional on $f(Y_i)$ will be guaranteed to satisfy the IIA property even if choice sets are *unobserved*. For this reason, we refer to any f that satisfies Condition 1 as to a *sufficient set*.

Proposition 2. Suppose that Assumption 1 and Condition 1 hold. Then, for every consumer type i and choice sequence $Y_i = j$ such that $f(j) = r$:

$$\Pr[Y_i = j | f(Y_i) = r, \theta] = \frac{\prod_{t=1}^{T} \exp(V(X_{ijt}, \theta))}{\sum_{k:f(k)=r} \prod_{t=1}^{T} \exp(V(X_{ijt}, \theta))} \tag{2.5}$$

and θ can be consistently estimated by the conditional Maximum Likelihood Estimator derived from $\Pr[Y_i = j | f(Y_i) = r, \theta]$.

Proof. See Appendix D.

Proposition 2 implies that, in the presence of unobserved choice set heterogeneity, parameters θ can be consistently estimated by Maximum Likelihood on the basis of the conditional logit model $\Pr[Y_i = j | f(Y_i) = r, \theta]$. Note in particular that equation (2.5) does not depend on i's unobserved sequence of possible choices, $\mathcal{CS}_i^* = c$. Whenever $\mathcal{CS}_i^\star = c$ is observable, the econometrician can easily detect appropriate subsets of $CS_{it}^\star = c_t$ for any i in t, and rely on McFadden (1978) to consistently estimate θ. However, whenever $\mathcal{CS}_i^\star = c$ is unobservable, the econometrician needs to be careful in constructing the set of choice sequences $f(Y_i)$. Indeed, if $f(Y_i)$ is *not* a subset of $\mathcal{CS}_i^\star = c$, then Condition 1 does not hold and we can use Proposition 1 to show inconsistency of any estimator based on $\Pr[Y_i = j | f(Y_i) = r, \theta]$: the econometrician is again mistakenly enlarging i's set of potential choice sequences and violating the IIA property.

Proposition 2 is quite general and works for any f that generates subsets of $\mathcal{CS}_i^\star$, it does not require additional assumptions to work. Even though later we will make assumptions about the stability or evolution of choice sets to illustrate a natural way of constructing f's that satisfy Condition 1, it is important to stress that these represent *a* set of sufficient conditions that imply Proposition 2, but, crucially, they are neither necessary nor the *minimal* sufficient conditions for the result to hold. There are many other assumptions implying Proposition 2 that could be more appropriate in some applications and not involve choice set stability or evolution (for example, as would be the case in a model with stockouts).

2.4 Unobserved Preference Heterogeneity

Before introducing specific examples of sufficient sets, we note that our approach can handle some forms of unobserved preference heterogeneity where the systematic utility $V_{it}(X_{ijt}, \theta)$ is indexed by i (beyond its dependence on the observed X_{ijt}). In particular, Proposition 2 can be extended to accomodate unobserved preference heterogeneity such as nested logit, RG: REWRITE FE logit with $V_{it}(X_{ijt}, \theta) = \delta_{ij} + V_{jt}(X_{ijt}, \theta)$, and mixed logit with discrete distribution of random coefficients or latent classes. In this section we discuss the case of the mixed logit and in the next the case of the FE logit. Results for the nested logit are available from the authors on request.

Suppose that each i has systematic utilities of the form $V_{jt}(X_{it}, \theta_i)$ with individual-specific preferences θ_i, which are distributed in the population according to $p(\theta_i = \theta|\psi)$. By conditioning the choice probability of sequence $Y_i = j$ on the sufficient set $f(Y_i) = r$, with $f(Y_i) = r \subseteq \mathcal{CS}_i^\star = c$, we obtain:[5]

$$
\begin{aligned}
&\Pr[Y_i = j | f(Y_i) = r, \psi] \\
&= \int_\theta \Pr[Y_i = j, \theta_i = \theta | f(Y_i) = r, \psi] d\theta \\
&= \int_\theta \Pr[Y_i = j | f(Y_i) = r, \theta_i = \theta] p(\theta_i = \theta | f(Y_i) = r, \psi) d\theta \\
&= \int_\theta \frac{\prod_{t=1}^T \exp(V_{jt}(X_{it}, \theta))}{\sum_{k: f(k) = r} \prod_{t=1}^T \exp(V_{kt}(X_{it}, \theta))} p(\theta_i = \theta | f(Y_i) = r, \psi) d\theta.
\end{aligned}
\tag{2.6}
$$

The first equality of (2.6) follows from the law of total probability and the second from the definition of conditional probability. The third equality follows from applying Proposition 2 to each conditional logit $\Pr[Y_i = j|f(Y_i) = r, \theta_i = \theta]$ of the mixture, which can then be expressed as in (2.5). Note that here, differently from the standard case of the mixed logit model, the distribution of random coefficients used to integrate out unobserved preference heterogeneity is *conditional* on the realization of the sufficient set $f(Y_i) = r$.

[5]Importantly, $\Pr[Y_i = j|f(Y_i) = r, \psi]$ is a very *different* object from $\Pr[Y_i = j|f(Y_i) = r, \theta]$ in equation (2.5): $\Pr[Y_i = j|f(Y_i) = r, \psi]$ is the integral over both the distribution of the Gumbel errors and the distribution of the random coefficients, while $\Pr[Y_i = j|f(Y_i) = r, \theta]$ is the integral only over the distritribution of the Gumbel errors for given realizaion θ of the random coefficients.

Equations (2.6) suggest a different approach for modeling unobserved preference heterogeneity compared to the existing literature, which commonly makes assumptions about the *unconditional* distribution of random coefficients in the population $p(\theta_i = \theta|\psi)$. Since $p(\theta_i = \theta|\psi) = \sum_r p(\theta_i = \theta|f(Y_i) = r, \psi) \Pr[f(Y_i) = r]$, parametric assumptions on $p(\theta_i = \theta|\psi)$ will not always translate into convenient restrictions on $p(\theta_i = \theta|f(Y_i) = r, \psi)$. For example, starting from an unconditional normal distribution, there is no reason to believe that the distribution of preferences conditional on a particular sufficient set will still be normal. Instead, we propose to directly make assumptions and estimate the conditional distribution of random coefficients $p(\theta_i = \theta|f(Y_i) = r, \psi)$.

There are two cases to consider depending on the richness of the available data. In the first case, suppose that there is a large number of individuals for each realization of $f(Y_i) = r$, so that one can reliably compute $\Pr[Y_i = j|f(Y_i) = r, \psi]$ for each $j \in f(Y_i) = r$. In this scenario, a simple model of unobserved preference heterogeneity compatible with (2.6) would be a mixed logit model with a *discrete* distribution of random coefficients: θ_i has a discrete support with $q = 1, \ldots, Q$ values and each individual belonging to group $f(Y_i) = r$ has preferences drawn from a categorical distribution $p(\theta_i = \theta_{[q]}|f(Y_i) = r, \psi) = \psi_q^r$, $q = 1, \ldots, Q$. In this case, model (2.6) simplifies to:

$$\Pr[Y_i = j|f(Y_i) = r, \psi] = \sum_{q=1}^{Q} \Pr[Y_i = j|f(Y_i) = r, \theta_i = \theta_{[q]}] \times \psi_q^r. \tag{2.7}$$

The Q values $\theta_{[q]}$ in the discrete support of θ_i could be assumed to be known a priori (i.e., and only the $R \times Q$ ψ_q^r conditional probabilities estimated), or both the Q values $\theta_{[q]}$ and the $R \times Q$ conditional probabilities ψ_q^r could be jointly estimated. Note that, in this case, assuming that $\psi_q^r = \psi_q$ for all r would amount to ignoring an endogeneity or selection problem. In this first case, after having obtained estimates $\widehat{p}(\theta_i = \theta_{[q]}|f(Y_i) = r, \psi) = \widehat{\psi}_q^r$, $q = 1, \ldots, Q$ and $r = 1, \ldots, R$, we can compute the estimated *unconditional* distribution of random coefficients in the population:

$$\widehat{p}(\theta_i = \theta_{[q]}|\psi) = \sum_{r=1}^{R} \widehat{\psi}_q^r \times \widehat{\Pr}\left[f(Y_i) = r\right], \tag{2.8}$$

where $\widehat{\Pr}\left[f(Y_i) = r\right]$ is the observed share of choice sequences that give rise to sufficient set $f(Y_i) = r$.

In the second case, there are too few individuals for any realization of the sufficient set, $f(Y_i) = r$, $r = 1, \ldots, R$, to estimate each one's distribution flexibly. In the extreme, there could be as many different sufficient sets as individuals, in which case the previous methods would not work since we would always have more parameters to estimate than observations. However, by relying on more parametric assumptions than in the first case, we can still obtain simple estimators. In a model of preference heterogeneity in which all individuals with $f(Y_i) = r$ have (common) preferences $\theta_i = \theta_r$, parametric assumptions about the distribution of θ_r *across* sufficient sets, $r = 1, \ldots, R$, would result in mixed logit models similar to those commonly used in applied work. For example, $p(\theta_r = \theta|\psi)$ could be assumed to be normal and ψ estimated by the simulated maximum likehood estimator.[6]

3 Sufficient Sets and Specification Tests

In this section, we first discuss some natural economic assumptions that lead to sufficent sets (i.e., correspondences f's compatible with Condition 1) and then present specification tests that can help the researcher to choose between candidate sufficient sets. We conclude the section with some Monte Carlo simulations to evaluate the practical performance of these sufficient sets. Again, we underline that the specific sufficient sets discussed here are a few simple examples, there are many other f's compatible with Condition 1 that could be more appropriate in different economic environments.

3.1 Sufficient Sets: Examples

Here we present some illustrative examples of assumptions on the stability or evolution of $CS^{\star}_{it}$ across the T choice situations that lead to sufficient sets. In Appendix E, we provide further detail.

3.1.1 Stable Choice Environments: Fixed Effect logit and Full Purchase History logit

Consider a situation in which i indexes individuals who make several purchase decisions $t = 1, \ldots, T$ at different points in time. We start with the simplest of examples: suppose that unobserved choice sets are *stable* across the T choice situations for each i (but potentially different across i's): $CS^{\star}_{it} = CS^{\star}_{i}$, for every t. To simplify exposition, we assume that choice sets are stable across *all* choice situations of each individual, but this is not essential. This method works whenever we observe *at least* two purchase decisions by the same individual from the same choice set. More generally, i's full choice

[6] While the current paper mainly focuses on the problem of unobserved choice set heterogeneity, the joint analysis of unobserved preference heterogeneity *and* unobserved choice set heterogeneity is work in progress by Xavier D'Haultfœuille and Alessandro Iaria.

sequence of length T can be divided into sub-sequences of length of at least two. Then, the assumption of stable choice sets implies fixed choice sets within each sub-sequence, but potentially different choice sets between the different sub-sequences of individual i.

Fixed Effect logit (FE logit) The Fixed Effect logit, or FE logit, is characterized by $V_{ijt}(X_{it}, \theta) = \delta_{ij} + V_{jt}(X_{it}, \theta)$. Then, given individual i's sequence of purchase decisions $Y_i = (Y_{i1} = j_1, \ldots, Y_{iT} = j_T)$, Chamberlain (1980) shows that θ can be consistently estimated from the conditional logit with sufficient set $f_{FE}(Y_i) = \mathcal{P}(Y_i)$: the set of all possible permutations of observed choice sequence $Y_i = (Y_{i1}, \ldots, Y_{iT})$. The validity of Chamberlain (1980)'s approach to differencing out fixed effects hinges on the assumption of choice set stability. In addition, note that choice set stability also implies that $\mathcal{P}(Y_i) \subseteq \mathcal{CS}_i^\star$: by the same arguments developed in Proposition 2, sufficient set $f_{FE}(Y_i) = \mathcal{P}(Y_i)$ will accomodate both unobserved preference heterogeneity in the form of individual-alternative specific fixed effects and unobserved choice set heterogeneity.[7]

Even though the FE logit addresses the problem of unobserved choice set heterogeneity whenever choice sets are stable across choice situations, it also comes with some meaningful costs. To obtain predicted purchase probabilities and their functions, such as elasticities, we usually need to be able to identify the whole vector of preference parameters θ. However, the FE logit does not usually allow for the identification of the entire vector θ. In particular, with linear-in-parameter systematic utility $V_{ijt}(X_{it}, \theta) = \delta_{ij} + X_{ijt}\theta$, only those elements of θ associated to time-varying observables will be identified.[8] In the next examples, we propose alternative sufficient sets that allow us both to difference out consumers' unobserved choice sets and to identify the entire preference vector θ.

Full Purchase History logit (FPH logit) An alternative to the FE logit that allows us to identify the whole vector θ relying on the same assumption of stable choice sets is the conditional logit obtained from sufficient set $f_{FPH}(Y_i) = (H_i)^T$, where $H_i = \bigcup_{t=1}^{T} \{Y_{it}\} \subseteq CS_{it}^\star$. The sufficient set $f_{FPH}(Y_i)$ imputes to each choice situation t the collection of all the products that individual i ever purchased in any of the T choice situations. We call the model implied by $f_{FPH}(Y_i)$ the Full Purchase History logit, or FPH logit. The relative advantage of the FPH logit over the FE logit of

[7]See D'Haultfœuille and Iaria (2016) for a discussion of methods to ease the computational burden associated with estimating the FE logit.

[8]If the vector X_{ijt} contains some element X_{ij}^1 that does not vary with t, so that $X_{ijt}\theta = X_{ij}^1\theta_1 + X_{ijt}^2\theta_2$, then the FE logit will difference out $X_{ij}^1\theta_1$ along with δ_{ij} and only θ_2 will be identified.

allowing for the identification of the full vector θ comes at the cost of not being robust to unobserved preference heterogeneity in the form of individual-alternative specific fixed effects. Differently from $f_{FE}(Y_i)$, sufficient set $f_{FPH}(Y_i)$ will only difference out the unobserved choice sets, and not both the choice sets and the δ_{ij}'s fixed effects.

3.1.2 Experience Goods: Past Purchase History logit (PPH logit)

Another example is to allow for the choice sets of each individual i to *expand* over *subsequent* choice situations $t = 1, \ldots, T$: $CS^{\star}_{it} \subseteq CS^{\star}_{it+1}$. Then the sufficient set is $f_{PPH}(Y_i) = \times_{t=1}^{T} H_{it}$, where $H_{it} = \bigcup_{b=1}^{t} \{Y_{ib}\} \subseteq CS^{\star}_{it}$. The sufficient set $f_{PPH}(Y_i)$ imputes to each choice situation t the collection of products that individual i ever purchased in any *past* choice situation up to t. We call the model implied by $f_{PPH}(Y_i)$ the Past Purchase History logit, or PPH logit.[9]

3.1.3 Common Choice Environments: Inter-Personal logit (IP logit)

With cross-sectional data, different *choice situations*, indexed by t, can be interpreted as different individuals of the same *consumer type*, indexed by i, each making a separate purchase decision at a single point in time from an *identical* choice set. In this context, the assumption of identical choice sets among the different T individuals of the same consumer type *is* the assumption of stable choice sets: $CS^{\star}_{it} = CS^{\star}_{i}$, for every i and t.

With cross-sectional data, the econometrician typically observes one purchase decision for each of the T individuals of consumer type i, $Y_i = (Y_{i1} = j_1, \ldots, Y_{iT} = j_T)$, and the observable characteristics of the chosen products, $(X_{ij_11}, \ldots, X_{ij_TT})$. In this situation we would rely on an assumption that the observable characteristics of any product j are the same for each of the T individuals of consumer type i, or that the econometrician knows how they change across t's. For example, if individual t and t' of consumer type i are observed purchasing, respectively, product j at price p_{ijt} and product k at price $p_{ikt'}$, then our method will rely on the assumption that each individual could have purchased the product bought by the other at exactly the same price, i.e. $p_{ijt} = p_{ijt'} = p_{ij}$ and $p_{ikt} = p_{ikt'} = p_{ik}$. If different products have observable characteristics that take different values for different t's of type i, say driving distance d_{ijt} between j and t in a model of supermarket choice, then the econometrician must be able to compute $d_{ijt'}$ for any other t' of type i.

[9] Symmetrically, we can also allow for choice sets that shrink over time.

The assumption of an identical choice set across the T individuals of the same consumer type i gives rise to a sufficient set similar in spirit to that of the Full Purchase History, $f_{FPH}(Y_i)$: $f_{IP}(Y_i) = (H_i)^T$, where $H_i = \bigcup_{t=1}^{T} \{Y_{it}\} \subseteq CS_{it}^{\star}$. The sufficient set $f_{IP}(Y_i)$ imputes to each individual t the collection of all the products purchased by any of the T individuals of consumer type i. Since this sufficient set entails comparisons of purchase behaviors between individuals, rather than within individual, we call the model implied by $f_{IP}(Y_i)$ the Inter-Personal logit, or IP logit.

3.1.4 Discussion

We reiterate here that the Full Purchase History, Past Purchase History, and Inter-Personal sufficient sets are merely examples: any set that is a subset of i's choices across t can serve as a sufficient set. For example, there may be some choice sets that vary in predictable ways due to idiosyncracies of this environment: an individual might choose from a different set of choices for lunch on Monday, when they are working at the office, than on Tuesdays when they are working from home. This could naturally be accommodated in the definition of sufficient sets by conditioning on past purchases made in that state (e.g. lunch choices made on past Mondays). Alternatively, information available to the researcher can be incorporated into the definition of sufficient sets, as for example accommodating stockouts when defining sufficient sets.

Note that we can combine any of our time-series sufficient sets (Full Purchase History, Past Purchase History, Monday Choice Sets) with the cross-sectional (Inter-Personal) sufficient set. For example, if we had confidence that the purchases of different individuals within a cross-section at a given point of time spanned the set of products available at that time, then we could use that information to limit the amount of time-series "extrapolation" between different cross-sections induced by the FPH or PPH sufficient sets.

3.2 Specification Tests for Sufficient Sets

There are a large number of possible sufficient sets. These lead to more or less robust and/or efficient estimators along the lines of Hausman and McFadden (1984) and can be used to form specification tests. In what follows, we discuss some approaches to testing the maintained assumptions implicit in several sufficient sets, in particular: (a) the length of the sequence of choice situations for which choice sets are stable, ever expand, or ever shrink; and (b) unobserved individual-product-specific preference heterogeneity.

3.2.1 Theoretical Foundations

The basis of our specification tests is a theoretical result that enables us to "rank" Maximum Likelihood Estimators (MLE) of logit models with different sufficient sets in terms of their efficiency, and to use that to form specification tests of some of the assumptions that lead to the sufficient sets discussed in section 3.1. Roughly, the MLE of a logit model with sufficient set f_L is more efficient than the MLE of a logit model with sufficient set $f_Z \subseteq f_L$. This result can be applied recursively, so that if two subsets of f_L are available, say f_Z and f_{XZ} with $f_{XZ} \subseteq f_Z$, then the efficiency rank of the three MLEs will be clear: $f_L \succeq f_Z \succeq f_{XZ}$.

Using this result, we propose Hausman–like specification tests between models based on different sufficient sets, i.e. different ways of constructing the correspondence f. These comparisons enable us to test for some forms of choice set stability and preference heterogeneity.

Proposition 3. Suppose that Assumption 1 holds, and that sufficient sets f_L and f_Z satisfy Condition 1 such that $f_Z(Y_i) \subseteq f_L(Y_i)$, $Y_i \in \mathcal{CS}_i^\star = c$ and $i = 1, \ldots, I$. Define $l_L(\theta)$ and $l_Z(\theta)$ as the log–likelihood functions corresponding to the logit models conditional on $f_L(Y_i)$ and on $f_Z(Y_i)$, with $\widehat{\theta}_L$ and $\widehat{\theta}_Z$ the corresponding Maximum Likelihood Estimators (MLEs). Then the following results hold:

1. The log–likelihood function $l_L(\theta)$ can be written as $l_L(\theta) = l_Z(\theta) + l_\triangle(\theta)$.
2. Provided that θ is identified in $l_\triangle(\theta)$, so that $\widehat{\theta}_\triangle$ is a well–defined MLE, then:
   (a) $\widehat{\theta}_Z$ and $\widehat{\theta}_\triangle$ are asymptotically independent, and
   (b) $\widehat{\theta}_L$ is more efficient than $\widehat{\theta}_Z$.
3. Given result (2), then:
   (a) All Hausman–like tests based on pairwise estimator comparisons among $\widehat{\theta}_L$, $\widehat{\theta}_Z$, and $\widehat{\theta}_\triangle$ are equivalent,

(b) The Likelihood Ratio statistic $LR = 2[l_Z(\widehat{\theta}_Z)+l_\triangle(\widehat{\theta}_\triangle)-l_L(\widehat{\theta}_L)]$ is asymptotically equivalent to the Hausman statistic comparing $\widehat{\theta}_L$ and $\widehat{\theta}_Z$, and

(c) $\text{Var}\left(\widehat{\theta}_L - \widehat{\theta}_Z\right) = \text{Var}\left(\widehat{\theta}_Z\right) - \text{Var}\left(\widehat{\theta}_L\right)$.

Proof. See Appendix F.1.

The Likelihood Ratio statistic, LR, proposed in result (3b) allows us to compare different models derived from alternative assumptions on sufficient sets. It consists of the difference between an unrestricted log–likelihood function, $l_Z\left(\widehat{\theta}_Z\right) + l_\triangle\left(\widehat{\theta}_\triangle\right)$, and a restricted one, $l_L\left(\widehat{\theta}_L\right)$.[10] Even though LR requires the computation of a third estimator, $\widehat{\theta}_\triangle$, it is simpler to implement than other Hausman–like statistics based on quadratic forms. For instance, the statistic LR is always non–negative, bypassing the practical inconvenience of some estimated covariance matrices that fail to be positive definite. In contrast to some other Hausman–like statistics, LR also makes very transparent the computation of the degrees of freedom of the corresponding χ^2 distribution: they equal the number of parameters in $\widehat{\theta}_L$. Result (3c) is of practical convenience, since it implies that the computation of $\text{Var}\left(\widehat{\theta}_L - \widehat{\theta}_Z\right)$, necessary for classical Hausman–like statistics, can proceed as in the standard case in which one of the compared estimators is fully efficient under the null hypothesis, even though no such efficiency assumption is required here. Proposition 3 hinges on the *Factorization Theorem* proposed by Paul Ruud. For more details see Ruud (1984) and Hausman and Ruud (1987).

3.2.2 Application to Sufficient Sets

The illustrative sufficient sets discussed in section 3.1 rely on the following economic assumptions:

- f_{FE}: Choice set stability across T choice situations and the possibility of having unobserved preference heterogeneity in the form of individual-alternative specific fixed effects, δ_{ij}.

- f_{FPH}: Choice set stability across T choice situations and no individual-alternative specific unobserved preference heterogeneity.

- f_{PPH}: Choice set evolution in the form of entry-but-no-exit or exit-but-no-entry across T choice situations and no individual-alternative specific unobserved preference heterogeneity.

[10] As developed more fully in Ruud (1984), this form is common to many econometric tests, including incremental over-identifying (or Sargan) tests (Arellano, 2003, Section 5.4.4) commonly used to investigate the validity of subsets of instruments.

Proposition 3 leads to statistics that can be used to test these economic assumptions by comparing the estimates obtained from models based on different sufficient sets (different f's). The first possibility is to compare f_{FE}, f_{FPH}, and f_{PPH} for choice sequences of constant length T. In fact, for choice sequences of a given length T, it can be shown that $f_{FE}(Y_i) \subseteq f_{FPH}(Y_i)$ and that $f_{PPH}(Y_i) \subseteq f_{FPH}(Y_i)$ for any $Y_i \in \mathcal{CS}_i^{\star} = c$. Given Proposition 3, these relationships among sufficient sets allow us to construct Hausman–like tests for the assumptions of choice set stability and for unobserved preference heterogeneity.

The second possibility for constructing test statistics is to fix a specific f, say f_{FE}, and to compare choice sequences with some of their *sub*–sequences: for example, the sequence $1, 2, \ldots, T^L$ can be split into two mutually exclusive sub–sequences $1, 2, \ldots, T^Z$ and $T^Z + 1, \ldots, T^L$, and this gives rise to different f_{FE}'s, f_{FE}^Z and f_{FE}^L such that $f_{FE}^Z(Y_i) \subseteq f_{FE}^L(Y_i)$ for any $Y_i \in \mathcal{CS}_i^{\star} = c$. The same holds for both f_{FPH} and f_{PPH}. As long as we can rely on the assumption of choice set stability or evolution of periods of length two, this method of making comparisons allows us to test for very general forms of choice set stability. In Appendix F.2 we discuss each of these test statistics in more detail.

3.3 How well do the sufficient set estimators perform?

Table 3.1 reports Monte Carlo simulations about the performance of the various sufficient sets against unobserved choice set heterogeneity.

The top panel shows the lack of bias in the absence of unobserved choice set heterogeneity. The following two panels show, in turn, the bias arising from first increasing the share of individuals with restricted choice sets and second increasing the severity of the restriction on choice sets. These results correspond to those shown in the second and third panels of Table C.1 in Appendix C. In each panel, the first column repeats the results from Table C.1 showing the bias from incorrectly assuming all consumers have access to the full (universal) choice set. The second column estimates the true model, i.e. the model that correctly assigns the choice set facing each individual in each choice situation. There is of course no estimation bias in this case.

The Full Purchase History (FPH) and Fixed Effects (FE) logits are reported in the third and fourth columns. The results show that there is significant bias when we incorrectly assume full choice sets (the first column), but that there is no average bias for estimation on either of these two sufficient sets.

Table 3.1: *Increasing the share of individuals with restricted choice set*

	Universal Logit		True Logit		FPH Logit		FE Logit	
	Bias (StdDev)	% Bias	Bias (StdDev)	% Bias	Bias (StdDev)	% Bias	Bias (StdDev)	% Bias
Baseline								
100% full choice set	0.005	0.3%	0.005	0.3%	0.011	0.6%	0.024	1.2%
	(0.032)		(0.032)		(0.030)		(0.085)	
Increasing share of individuals with constrained choice set								
10% constrained	-0.223	-11.2%	0.003	0.2%	0.008	0.4%	0.015	0.8%
	(0.021)		(0.028)		(0.027)		(0.080)	
30% constrained	-0.525	-26.3%	0.005	0.3%	0.012	0.6%	0.011	0.6%
	(0.013)		(0.028)		(0.029)		(0.087)	
50% constrained	-0.719	-36.0%	0.007	0.4%	0.014	0.7%	0.018	0.9%
	(0.007)		(0.027)		(0.027)		(0.078)	
Increasing share of products randomly removed from choice set								
30% have 4 of 5	-0.525	-26.3%	0.005	0.3%	0.012	0.6%	0.011	0.6%
	(0.013)		(0.028)		(0.029)		(0.087)	
30% have 3 of 5	-0.719	-36.0%	0.013	0.7%	0.02	1.0%	0.023	1.2%
	(0.007)		(0.027)		(0.027)		(0.068)	
30% have 2 of 5	-1.139	-57.0%	0.007	0.4%	0.009	0.5%	0.01	0.5%
	(0.003)		(0.027)		(0.027)		(0.067)	

We consider a population of 1,000 consumers making a sequence of choices over 10 choice situations. On each choice situation they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in equation (2.1). The systematic utility is linear with homogenous preferences, $V_{ijt}(X_{it}, \theta) = \delta_j + X_{jt}\beta$, and the unobserved portion of utility, ϵ_{ijt}, is distributed Gumbel. In the baseline specification, X_{jt} is drawn from a Normal distribution with mean 0 and variance 5, $\delta_j = 0$ for all j's, and $\beta = 2$. We simulate 20 replications. To speed up computations, the FE logit is estimated by sampling at random (uniformly), for each individual, 5000 permutations of the observed sequence of choices (see D'Haultfœuille and Iaria (2016).)

4 Discussion about Identification

In this section, we describe those parameters (or functions of parameters) that we can point-identify, and how we can use sufficient sets to derive bounds on several useful functions of these parameters. We summarize these results here and provide further detail in Appendix G. For expositional simplicity, suppose that the systematic utilities are linear in the parameters,

$$V_{jt}(X_{it}, \theta) = \delta_j + X_{ijt}\beta + \alpha p_{jt},$$

where $X_{it} = [X_{i1t}, p_{1t}, ..., X_{iJt}, p_{Jt}]$ and $\theta = [\delta_1, ..., \delta_J, \beta, \alpha]$.

As described in Sections 2.3 and 3.1, we can point-identify the vector of preference parameters θ from conditional logit model $\Pr[Y_i = j|f(Y_i) = r_i, \theta]$.[11] We can similarly point-identify simple functions of θ. For example, we are often interested in willingness-to-pay (WTP) for product characteristic k, X^k_{ijt}. By Roy's Identity, this can be computed as:

$$WTP_k = -\frac{\partial V_{ijt}/\partial X^k_{ijt}}{\partial V_{ijt}/\partial p_{jt}} = -\frac{\beta_k}{\alpha}. \tag{4.1}$$

Some outputs of economic interest, however, require information about the distribution of choice sets in the population for point identification. We cannot point-identify these functions, but we can place bounds on them. This makes clear that these functions are only point-identified relying on assumptions about the choice set formation process, while the bounds we specify require weaker assumptions.

The probability i purchases product j given choice set $CS^\star_{it} = c_{it}$ is

$$Pr^{CS^\star}_{ijt}(\theta) \equiv \Pr\left[Y_{it} = j|CS^\star_{it} = c_{it}, \theta\right] = \frac{\exp\left(\delta_j + X_{ijt}\beta + \alpha p_{jt}\right)}{\sum_{m \in c_{it}} \exp\left(\delta_m + X_{imt}\beta + \alpha p_{mt}\right)} \tag{4.2}$$

if $j \in CS^\star_{it} = c_{it}$ and zero otherwise. This choice probability depends on i's (unobserved) choice set, $CS^\star_{it}$.

Suppose that we observe a superset S_{it} of the true but unobserved choice set, so that $CS^\star_{it} \subseteq S_{it}$. This could be, for example, the collection of all products observed to be purchased by any i in choice situation t. It follows that, even if we do not directly observe $CS^\star_{it} = c_{it}$, $f_t\left(Y_i\right) \subseteq CS^\star_{it}$ and $CS^\star_{it} \subseteq S_{it}$. We can therefore use these conditions to bound the true but unobserved denominator of the logit choice probabilities for any X_{it} and θ:

$$\sum_{m \in f_t(Y_i) = r_{it}} \exp\left(\delta_m + X_{imt}\beta + \alpha p_{mt}\right) \le \sum_{m \in CS^\star_{it} = c_{it}} \exp\left(\delta_m + X_{imt}\beta + \alpha p_{mt}\right) \le \sum_{m \in S_{it} = s_{it}} \exp\left(\delta_m + X_{imt}\beta + \alpha p_{mt}\right). \tag{4.3}$$

It follows from (4.3) that for any $j \in f_t(Y_i) = r_{it}$:

$$Pr^S_{ijt}(\theta) \le Pr^{CS^\star}_{ijt}(\theta) \le Pr^f_{ijt}(\theta). \tag{4.4}$$

[11] Note that, differently from elsewhere in the paper, here we will keep track of the "i" subscript in the realizations of the sufficient sets, $f(Y_i) = r_i$, and of the choice sets, $CS^\star_{it} = c_{it}$. This is essential to avoid confusion when computing averages across individuals, as detailed in Appendix G.

That is to say, the true choice probability that i purchases j in t is bounded below by the same probability assuming i chooses from some superset of the unobserved choice set, $S_{it} = s_{it}$, and bounded above by the same probability assuming i chooses from just their sufficient set, $f_t(Y_i) = r_{it}$. In Appendix G, we expand on this idea and discuss the ability to bound average (across i) choice probabilities and household-level own- and cross-price elasticities.

The same bounds in Equation (4.3) imply the ability to bound consumer surplus. Let the true consumer surplus of individual i in t be:

$$W_{it}\left(X_{it}\middle|\,\theta, CS^{\star}_{it} = c_{it}\right) = \zeta + \frac{1}{\alpha}\ln\left(\sum_{m\in c_{it}} \exp\left(\delta_m + X_{imt}\beta + \alpha p_{mt}\right)\right). \tag{4.5}$$

where ζ is Euler's constant. Then, for any X_{it} and θ:

$$W_{it}\left(X_{it}\middle|\,\theta, f_t\left(Y_i\right) = r_{it}\right) \leq W_{it}\left(X_{it}\middle|\,\theta, CS^{\star}_{it} = c_{it}\right) \leq W_{it}\left(X_{it}\middle|\,\theta, S_t = s_{it}\right). \tag{4.6}$$

5 Models of Choice Set Formation Compatible with Our Approach

The estimator we propose in this paper allows for the choice set formation process, $\Pr\left[\mathcal{CS}^{\star}_i = c\middle|\,\gamma\right]$, to take any form (as long as Assumption 1 is satisfied). It can be a function of past, present, and future systematic utilities, it can depend on observable or unobservable variables outside of the choice model, and it can follow any distribution with or without closed-form expression. This generality makes our estimator compatible with a wide variety of economic situations that may lead to choice set heterogeneity in the data. We discuss a few broad classes of economic models that fit into the framework proposed by our estimator. This is not meant to be exhaustive; there are many other models to which our estimation method would also apply.

5.1 Limited Attention

There is a fast-growing literature in which limited attention is used to rationalize apparently incongruent consumer and firm behavior. Eliaz and Spiegler (2011) propose a model of boundedly rational consumers that make purchase decisions from subsets $CS^{\star}$ of the full set of available products in the market and explore how firms can use marketing devices, M, to manipulate these subsets. In a first stage, consumer i is matched to that choice set $CS_i^\star = c$ for which the reduced-form matching function $\phi(CS_i^\star = c, M_i)$ equals 1, while $\phi(CS_i^\star, M_i)$ equals 0 for any $CS_i^\star \neq c$. In a second stage, consumer i chooses her utility maximizing product from choice set $CS_i^\star = c$. In the language of equation (2.2), the choice set formation process specified by Eliaz and Spiegler (2011) can be expressed as:

$$\Pr[CS_i^\star = c|\gamma] = \Pr[\phi(CS_i^\star = c, M_i) = 1|\gamma].$$

In Eliaz and Spiegler (2011) the function $\phi(CS_i^\star, M_i)$ can take different forms in different contexts. For example, M_i can be the quantity of TV advertisements produced by the two market leaders, or the specific media to which consumer i is actually exposed, or any marketing device used by all the firms in the market.

In related work, Matejka and McKay (2015) model consumers that are inattentive, but *rationally* so. The authors assume consumers have imperfect information about payoffs but, before making a product choice, may gain information at a cost $\lambda_i > 0$. Using information theory to determine consumers' optimal information-processing strategy, they show that:

$$\Pr[Y_i = j|CS_i^\star = c, \theta] = \frac{\exp(V_{ij}(X_i, \theta) + \delta_{ij})/\lambda_i}{\sum_{m \in CS_i^\star = c} \exp(V_{im}(X_i, \theta) + \delta_{im})/\lambda_i},$$

where the true indirect utilities are the $V_{ij}(X_i, \theta)$'s, while the individual-alternative specific fixed effects δ_{ij}'s do not represent i's preferences, but are scaling factors that reflect the cost of acquiring information λ_i and i's prior beliefs about the realizations of the $V_{ij}(X_i, \theta)$'s. In practice, it may be infeasible for an econometrician to gather reliable information about these inherently unobservable objects. Our approach could be adapted to estimate both θ and features of the distribution of λ_i.

Goeree (2008) estimates a model of PC purchases in which consumers are subject to limited attention in the form of restricted choice sets, $CS_i^\star$'s. PC producers can influence with advertisements the probability that consumers consider their products for purchase. In particular, the author specifies the choice set formation process as:

$$\Pr[CS_i^\star = c|\gamma] = \prod_{l \in c} \frac{\exp(W_{il}(\gamma))}{1 + \exp(W_{il}(\gamma))} \prod_{k \notin c} \left(1 - \frac{\exp(W_{ik}(\gamma))}{1 + \exp(W_{ik}(\gamma))}\right),$$

where $W_{il}(\gamma)$ is a function of advertisement expenditures on PC model l, consumer i's advertisement exposure and demographics, and some parameters γ. In addition to various functional form restrictions, the identification of γ hinges on the availability of reliable data about advertising expenditures and exposure of different consumers to different ads based on how reading habits vary with consumer demographics. We discuss this model further in Appendix B.

5.2 Search

Since Stigler (1961)'s seminal work, models of costly search have been widely employed to explain imperfectly competitive outcomes in product and labor markets. De los Santos et al. (2012) use web browsing and online purchase data about books to test classical models of consumer search, i.e. fixed sample versus sequential search. The authors propose a model in which consumers are initially uncertain about the prices charged by J online retailers, but each price can be learned upon visiting the corresponding retailer's website. The authors specify the choice set formation process, or consumer i's probability of "sampling" the online retailers belonging to choice set $CS_i^\star = c$, as:

$$\Pr[CS_i^\star = c|\gamma] = \frac{\exp\left(\mathbb{E}\left[\max_{j\in c}\{U_{ij}\}|\gamma\right] - (\#c)\cdot s_i(\gamma)\right)}{\sum_{r\in\mathcal{J}}\exp\left(\mathbb{E}\left[\max_{k\in r}\{U_{ik}\}|\gamma\right] - (\#r)\cdot s_i(\gamma)\right)},$$

where U_{ij} is i's indirect utility of purchasing a book from online retailer j, $\#c$ is the number of online retailers in $CS_i^\star = c$, s_i is i's cost of searching for an additional online retailer, $\mathcal{J}$ is the set containing all the possible subsets of the J online retailers, and γ a vector of parameters. Similar specifications are common in the search literature (Roberts and Lattin (1991), Draganska and Klapper (2011), and Moraga-González et al. (2015)).

The estimation of this model requires detailed web browsing data, or more generally individual data about search behavior, and several further assumptions for the implementation of $\mathbb{E}\left[\max_{j\in c}\{U_{ij}\}|\gamma\right]$ and $s_i(\gamma)$. In addition, the power set $\mathcal{J}$, and consequently the denominator of the logit formula, has $2^J - 1$ elements: this translates into a curse of dimensionality in J, which is usually addressed by tightening the functional form restrictions (e.g. if one assumes away any form of unobserved heterogeneity and estimates a Logit model, one can rely on McFadden (1978) and estimate the model on random subsets of $\mathcal{J}$).

Dinerstein et al. (2014) use web browsing and online purchase data from eBay to estimate a model of consumer search and price competition, and study the effects of alternative online search designs. Similar to the last paper, Dinerstein et al. (2014) are able to observe in their detailed web browising data the choice set $CS_i^\star = c$ to which consumer i is matched prior to making her purchase decisions. The authors specify a choice set formation process in which each possible alternative j has a sampling weight of:

$$w_j = \exp\left[\gamma\left(\frac{p_j - \min_k\{p_k\}}{\sigma_p}\right)\right],$$

where p_j is the price of alternative j, $\min_k\{p_k\}$ is the minimum available price, σ_p is the price standard deviation, and γ is a parameter. Given sampling weights w_j's, consumer i's choice set $CS_i^\star = c$ is then constructed by sampling without replacement $\#c$ alternatives among the J available. This results in $\Pr[CS_i^\star = c|\gamma]$ following a Wallenius non-central hypergeometric distribution.

5.3 Screening Rules

Many theories of consumer behavior involve thresholds and discontinuities. For example, when buying a new car some car model may not be taken into consideration because its price exceeds the consumer's budget contraint. When going out for dinner the restaurants considered may be very different from those considered for lunch. Prospect theory, reference prices, and restricted choice sets all involve abrupt behavior changes under certain discrete circumstances. Gilbride and Allenby (2004) investigate consumers' use of *screening rules* as part of a discrete choice model. Alternatives that pass the screen are evaluated in a manner consistent with random utility theory. Alternatives that do not pass the screen have zero probability of being chosen.

Gilbride and Allenby (2004) specify a general choice set formation process:

$$\Pr[CS_i^\star = c|\gamma] = \Pr[1(X_{ik}, \gamma) = 1, 1(X_{il}, \gamma) = 0 | k \in c, l \notin c],$$

where $1(.)$ is the indicator function, X_{ik} denotes a generic argument capturing the decision rule applied to screen alternative k, and γ is a vector of parameters. This framework can accomodate a wide variety of screening rules, for instance:

- $1(X_{ik}, \gamma) = 1(V_{ik}(X_i, \theta) > \gamma)$, where the consumer discards from the choice set alternatives with low systematic utility (typical of rational inattention and search models)

- $1(X_{ik}, \gamma) = \prod_{m \in M} 1(X_{ikm} > \gamma_m)$, where the consumer considers alternative k only if every attribute $m \in M$ is acceptable.

- $1(X_{ik}, \gamma) = 1\left(\sum_{m \in M} 1(X_{ikm} > \gamma_m)\right)$, where the consumer considers alternative k only if at least one attribute $m \in M$ is acceptable.

In practice, it may be difficult for the econometrician to know a priori which of the several possible screening rules to implement, and consequently the risk of selecting the wrong screening model may be high. Moreover, screening rules introduce endogenous thresholds and discontinuities in the econometric model, which Gilbride and Allenby (2004) address with an involved Markov Chain Monte Carlo procedure that requires strong distributional assumptions. Our approach would allow estimation of θ without regard to the particular screening rule and could be compared to those from the full model to test its robustness.

5.4 Regulatory Restrictions on Choice Sets

Public reforms that expand the scope for individual choice are an increasingly common phenomenon, especially in public sectors such as health care and mandatory education. Gaynor et al. (2016) exploit a reform in the English National Health Service to evaluate the effect of expanding hospital choice on the quality of the service and, ultimately, on the health of patients. Prior to the reform, any patient i was matched to a constrained and unobserved choice set $CS_i^\star = c \subseteq H$, where H is the set of all possible hospitals. After the reform, all patients could choose their favorite hospital from the full set H. The constrained and unobserved set of hospitals $CS_i^\star = c$ was chosen, or "suggested," by the physician in charge of patient i. Hence, Gaynor et al. (2016) specify their pre-reform choice set formation process as a physician choice model:

$$\Pr[CS_i^\star = c|\gamma_i] = \Pr[W_{ij}^p(\mu) > \max_{l \in H}\{W_{il}^p(\mu)\} - \lambda_i, W_{ik}^p(\mu) \leq \max_{l \in H}\{W_{il}^p(\mu)\} - \lambda_i | j, k \in CS_i^\star = c],$$

where $W_{ij}^p(\mu)$ is the indirect utility of physician p for including hospital j in the choice set $CS_i^\star = c$ of patient i and $\gamma_i = (\mu, \lambda_i)$ is a vector of parameters, with μ being physician p's preferences, and

$\lambda_i \in [0, +\infty)$ representing the relative weight of patient versus physician preferences in the decision making process. Gaynor et al. (2016) do not assume $\Pr[CS_i^\star = c|\gamma_i]$ to have a closed-form solution and approximate it with simulation methods.

In an education context, Walters (2014) estimates a demand model for charter middle schools in Boston and quantifies the effectiveness of charter schools on student performance, controlling for endogenous self-selection of students into charter schools. Of all the charter school applicants, only a random subset is accepted for enrolment. Conditional on having applied, charter school acceptance is exogenously determined by a lottery; however, the initial decision to apply to the charter school is not, and the observed sample of charter school students may be be selected. Walters (2014) addresses the sample selection problem by specifying a joint model of charter application portfolio choices, lottery outcomes, and school attendance decisions. In this framework, the school decision process starts with student i's choice of a portafolio of applications to a subset of the J charter schools, $CS_i^\star = c$:

$$\Pr[CS_i^\star = c|\gamma = (\pi, \theta)] = \frac{\exp\left(\sum_{z \in \mathcal{J}} f(z|c, \pi) \cdot W_{iz}(\theta)\right)}{\sum_{r \in \mathcal{J}} \exp\left(\sum_{z \in \mathcal{J}} f(z|r, \pi) \cdot W_{iz}(\theta)\right)},$$

where $f(z|c, \pi)$ is the lottery probability of being accepted by a set z of charter schools conditional on having applied to a set c of schools, W_{iz} is student i's expected net benefit of receiving offers from the set z of charter schools, $\mathcal{J}$ is the set containing all the possible subsets of the J charter schools, and $\gamma = (\pi, \theta)$ is a vector of parameters. Because of its dependence on (nested) summations across all the elements of $\mathcal{J}$, this model suffers from a curse of dimensionality in J. For a similar example of a choice set generation process in the context of stochastic portfolio choice problems, see also Chade and Smith (2006).

5.5 Endogenous Product Choice

There is a growing empirical literature that treats product choices as endogenous (see Crawford (2012) for a survey): firms can ease price competition by choosing the set of products, and the varieties of these products, they offer in each market. The specification of models that allow firms to choose both the set of products they offer and their prices allows researchers to conduct more realistic evaluations of both firm strategy and public policy decisions.

Eizenberg (2014) investigates the welfare implications of the rapid innovation in computers' central processing unites (CPUs), and asks whether it results in inefficient elimination of basic PC products as technology evolves. The author specifies a demand for PCs in line with our Assumption 1 and a two-stage game played by PC makers. The two-stage game determines the choice set formation process $\Pr[CS_i^\star = c|\gamma]$. In the first stage, PC makers observe the realizations of shocks to fixed costs and then simultaneously choose which PC products to offer. In the second stage, for any $CS_i^\star = c$ chosen in the first stage, PC makers observe the realization of marginal cost and demand shocks *up to the logit components* and simultaneously choose prices. Importantly, in the first stage PC makers choose $CS_i^\star = c$ on the basis of *expected* demand and marginal costs, and also in the second stage price decisions are taken without knowing the exact realization of the demand logit errors, but only their expected values. These informational assumptions ensure our Assumption 1 holds, but still leave the estimation of $\Pr[CS_i^\star = c|\gamma]$ a challenging task. Eizenberg (2014) addresses the issues of sample selection and multiplicity of equilibria involved in the estimation of γ by relying on partial identification techniques.

Similarly to Eizenberg (2014), Draganska et al. (2009) estimate a two-stage game in which firms first choose product assortments and then compete on prices. The authors estimate their endogenous product choice model using data from the ice-cream market. Like Eizenberg (2014)'s, also Draganska et al. (2009)'s model relies on our Assumption 1. The authors make stronger assumptions than Eizenberg (2014) in their two-stage game (e.g. no sample selection), so that the choice set formation process $\Pr[CS_i^\star = c|\gamma]$ is point identified rather than only partially identified. Even in this simplified framework, the estimation of $\Pr[CS_i^\star = c|\gamma]$ is computationally challenging, and the empirical analysis is constrained to the endogenous choice of only one ice-cream flavor, vanilla.

6 Empirical Illustration

To illustrate how our estimator can be applied in practice, we present an empirical illustration. In Section 5, we discussed the literature on models of limited attention and the role that marketing expenditure can play at influencing consumers' choice sets (as in the models of Eliaz and Spiegler (2011) and Goeree (2008)). In this section, we estimate demand for chocolate by a sample of adult working-age women making decisions on-the-go, i.e. chocolate purchased outside of the home in small corner stores, vending machines, concession stands, and other outlets for immediate consumption.

There are over 600 products in the on-the-go chocolate market from which a consumer is able to choose. In such a choice environment it is unlikely that a consumer will spend the time to consider each one, and the costs of collecting information on which products the consumer considered (for example, from eye tracking technologies) is expensive. We are interested in estimating consumers willingness to pay for different brands and in how advertising might affect consumers choices. Advertising is important in the chocolate market, and there is intuitive appeal to the idea ads might play an important role in bringing products to consumers' attention (as in Eliaz and Spiegler (2011) and Goeree (2008)) as well as potentially entering their utility directly (as in Becker and Murphy (1993)). We estimate preference parameters while allowing for the possibility that consumers choose from a subset of products that is unobserved and potentially heterogenous using the Past-Purchase History logit.

6.1 Model

We adapt the general model presented in Section 2.1 to the specifics of on-the-go chocolate demand. We assume each consumer makes a purchase from a subset of products, $CS_{it}^{\star}$, available in the market. This set is not observed, it could include all products or only a few products, but it does include the option not to purchase at all times for all individuals.

In a first stage, a consumer i is matched to their choice set $CS_{it}^{\star} = c_{it}$. We assume that products that the consumer is observed having purchased in the past are in their choice set; this is the past-purchase history sufficient set discussed in Section 3.1.2. In a second stage, consumer i chooses the utility-maximizing product from her choice set $CS_{it}^{\star} = c_{it}$.

The probability consumer i buys product j in period t given her past purchase history is given by:

$$\Pr[Y_i = j | f(Y_i) = r_i, \theta] = \frac{\prod_{t=1}^{T} \exp(V_{jt}(X_{it}, \theta))}{\sum_{k: f(k) = r_i} \prod_{t=1}^{T} \exp(V_{kt}(X_{it}, \theta))}. \tag{6.1}$$

where $f_t(k) = r_{it}$ is the set of products in individual i's past purchase history and $f(k) = \times_{t=1}^{T} f_t(k) = r_i$ the corresponding set of sequences of products. Utility for the inside products, $j = 1, \ldots, J$ is given by

$$U_{ijt} = V_{jt}\left(X_{it}, \theta\right) + \epsilon_{it}, \tag{6.2}$$

with

$$V_{jt}\left(X_{it}, \theta\right) = \delta_j + \alpha p_{ojt} + \beta \ln a_{(i)bt}, \tag{6.3}$$

where δ_j is product j specific fixed effect, p_{ojt} is the price of product j in outlet type o in week t, and $\ln a_{(i)bt}$ is log advertising exposure to brand b (to which product j belongs) on day t. The price variable and two measures of advertising exposure are defined in the next subsection.

Utility to the outside good of not purchasing a chocolate bar is given by

$$U_{i0t} = \delta_0 + \sum_m \tau_m + \epsilon_{i0t}, \tag{6.4}$$

where τ_ms are month effects meant to capture seasonality and/or cyclicality in on-the-go chocolate demand. We also estimate the Universal (Choice Set) logit, where we assume that all individuals choose from all products in each period, given by

$$\begin{aligned}
\Pr[Y_i = j|\mathcal{S} = s, \theta] \quad &= \Pr[Y_i = j|\mathcal{CS}_i^\star = c_i, \theta] \Pr[Y_i \in \mathcal{CS}_i^\star = c_i|\mathcal{S} = s, \theta] \\
&= \prod_{t=1}^{T} \frac{\exp\left(V_{jt}\left(X_{it}, \theta\right)\right)}{\sum_{m \in S_t = s_t} \exp\left(V_{mt}\left(X_{it}, \theta\right)\right)}, \qquad (6.5) \\
&= \prod_{t=1}^{T} \frac{\exp\left(V_{jt}\left(X_{it}, \theta\right)\right)}{\sum_{m \in CS_{it}^\star = c_{it}} \exp\left(V_{mt}\left(X_{it}, \theta\right)\right)} \frac{\sum_{m \in CS_{it}^\star = c_{it}} \exp\left(V_{mt}\left(X_{it}, \theta\right)\right)}{\sum_{r \in S_t = s_t} \exp\left(V_{rt}\left(X_{it}, \theta\right)\right)},
\end{aligned}$$

6.2 Data

We use data on 297 working-age women (ages 19-59) without children from the Kantar Worldpanel on-the-go survey, collected from individuals who record purchases that they make on-the-go for immediate consumption.[12] We use information on 9,387 purchase occasions over the period 2010-2011. A purchase occasion is when the women are observed purchasing a snack of any form on-the-go. At any one point in time there are up to 250 different types of chocolate products available in the market.

[12] This dataset was used to analyze the effects of banning advertising in the market for junk foods in Dubois et al. (2016); we follow their lead in many aspects of our data construction.

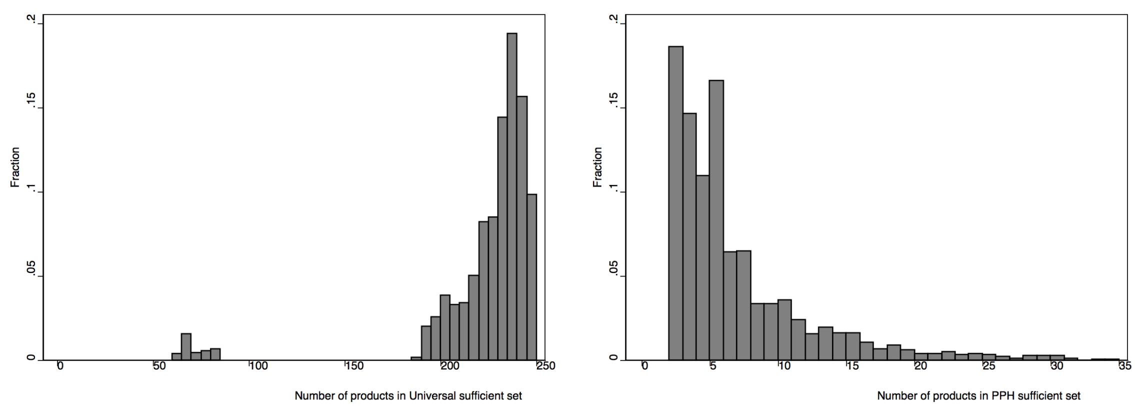
The outside option, when a chocolate bar is not purchased, has a 74% market share. Prices are constructed at the level of the outlet and week; we consider four types of outlets–large national chains, news agents, vending machines, and other outlets.

Figure 6.1 shows the distribution of the number of products in the sufficient sets. On the left hand side we show the number of products in the Universal set of products across the 9387 individual purchase occasions. The maximum is 250 products; the small cluster between 50 and 100 is purchase occasions where the individual chooses from a vending machine (these are all products sold in vending machines in that week in the UK). On the right hand side we show the distribution of the number of products in the past purchase history sufficient sets; these range from 2 to 35.

Figure 6.1: *Sufficient sets*

(a) Universal

(b) Past purchase history



Note: The histograms shows the distribution of the number of products in the sufficient set across the 9387 individual purchase occasions.

We use two measures of advertising exposure. Both convert weekly advertising ("flows") into an advertising "stock;" advertising stocks are the depreciated cumulation of the flows. The first measure is at the brand level and follows common practice in the empirical advertising literature; we use aggregate minutes of TV advertising aired during the week at the brand level to define advertising flows, which we denote s_{bt}. These data show that products that advertise often follow a pulsing strategy (as described in Dubé et al. (2005)) with short periods of high advertising followed by zero advertising; this also suggests that firms may be using these strategies to bring products to consumers' attention and into their choice sets.

Our second measure follows Goeree (2008) and measures advertising exposure at the individual level. We use detailed information about when individual ads were aired on television matched with self-reported viewing information to construct individual-level measures of exposure to brand advertising, which we denote $s_{ibt}.s_{ibt}$, ranges from 0 for individuals that do not watch TV, or only watch public TV (the BBC), to a maximum of 65 minutes of cumulated exposure to advertisements for a particular brand. The mean is 29 minutes of cumulated exposure.

For both measures of advertising exposure, we follow Dubé et al. (2005) and allow for diminishing returns to advertising by transforming the stock of advertising, $s_{(i)bt}$, using the log inverse hyperbolic sine function, $\ln a_{(i)bt} = ln\left(s_{(i)bt} + \sqrt{s_{(i)bt}^2 + 1}\right)$. Further details are available in Appendix H.

6.3 Estimation Results

Table 6.1 presents the estimated preference parameters for both the Universal logit and the Past Purchase History logit. Columns (1) and (2) use brand-level advertising stocks, while columns (3) and (4) use individual-level advertising stocks. Columns (1) and (3) are estimated using the Universal choice set (i.e. all products available in that month across all stores), and columns (2) and (4) are estimated using households' past purchase history (PPH) as sufficient sets. The estimates based on the past purchase history yields more elastic price responses and less elastic advertising responses.

The patterns that we find in this illustrative example are intuitive from both an economic and econometric perspective. In both specifications, our estimate of price sensitivity is greater (in absolute value) when using the PPH sufficient set. This is intuitive as if consumers are failing to consider some of the products in the universal choice set, then price variation over these products will be ignored, causing the Universal logit to attribute a lack of consumer response to a lack of price sensitivity.

By contrast, in both specifications we find that our estimates of advertising sensitivity are smaller when using the PPH sufficient set. As described above, the literature analyzing the economics of advertising has argued that advertising can both inform consumers about products' existence and so increase the likelihood that they are in consumers' choice sets, as well as directly influence consumer utility, shifting their preferences. The Universal logit can therefore be considered a "reduced form" that captures both of these effects, while the PPH logit, by focusing on those products for which consumer attention is by assumption already high, identifies the effects of advertising only through its influence on preferences. If this story is an accurate characterization of behavior in the on-the-go chocolate market, then one would expect to find, as we do, *smaller* estimated advertising sensitivity with the PPH logit.

Table 6.1: *Coefficient estimates*

	(1) Universal Logit	(2) PPH Logit	(3) Universal Logit	(4) PPH Logit
price, p_{ojt}	-1.900 (0.155)	-2.404 (0.145)	-1.853 (0.154)	-2.373 (0.144)
brand-level, a_{bt} advertising	0.040 (0.010)	0.024 (0.007)	—	—
individual, a_{ibt} advertising exposure	—	—	0.159 (0.042)	0.092 (0.045)
size	0.003 (0.001)	0.009 (0.001)	0.003 (0.001)	0.009 (0.001)
N	2,102,878	60,925	2,102,878	60,925
product effects	yes	yes	yes	yes

Notes: PPH: past-purchase history, U: universal choice set. All columns include month dummies interacted with the outside option, to control for seasonal and cyclical effects. Advertising is measures in seconds and the variable divided by 1000. We include 58 product effects, one for each product that has a market share of greater than 0.1%, or is purchased at least 10 times).

We use the estimated preference parameters to explore the implications of our results. We look at the estimated willingness-to-pay for advertising, something in which firms and advertising executives are likely to be interested, and the willingness-to-pay for individual products, something in which firms and retailers are likely to be interested.

In our simple illustrative example preferences are homogeneous across households and so willingness-to-pay for (log) advertising is given by:

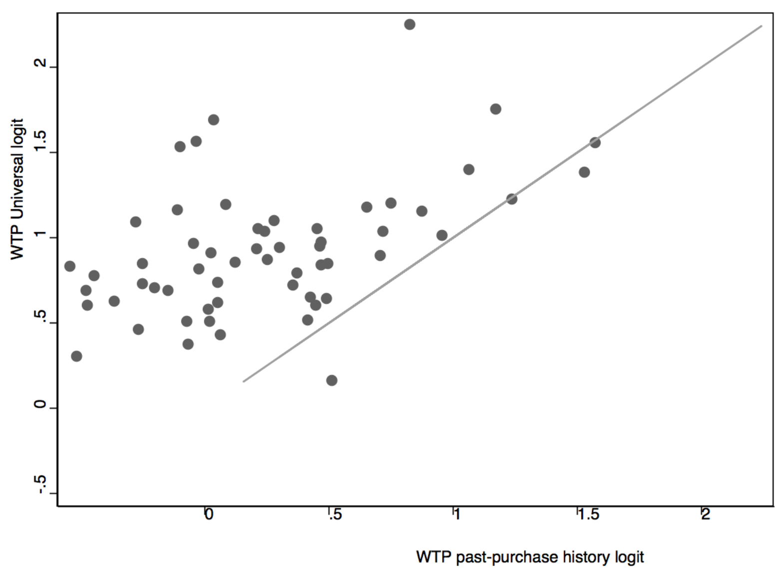
$$\widehat{WTP}_a = -\frac{\partial V_{ijt}/\partial \ln a_{(i)bt}}{\partial V_{ijt}/\partial p_{ojt}} = -\frac{\widehat{\beta}}{\widehat{\alpha}}. \tag{6.6}$$

The parameter estimates in Table 6.1 indicate that there is significant bias in willingness-to-pay when estimating from the Universal choice set. The estimates obtained using the Universal logit in column (3) suggest that a one-standard deviation increase in the log advertising stock, $\ln a_{ibt}$, equal to 0.54

(or 54%), implies an increase in valuation of a product by 4.6 pence.[13] As the average price of a chocolate product is 74 pence, this is a 6.2% increase. By contrast, the estimates obtained using the PPH Logit suggest a one-standard deviation increase in the log advertising stock increases the value of a product by less than half that: 2.1 pence, or a 2.8% increase.

We also look at willingness-to-pay for products, given for each product j by $\widehat{\delta}_j/\widehat{\alpha}$. Figure 6.2 plots willingness-to-pay for the 58 largest market share products. A data point in the figure represents the estimated willingness-to-pay for a product in both the Universal Logit (on the y-axis) and the PPH Logit (on the x-axis); the straight line is the 45-degree line indicating equal estimated willingness-to-pay from both models. The results show that the estimated willingness-to-pay is higher for most products with the Universal logit than with the PPH logit. In most cases, the downward bias in the estimated price coefficient seen in Table 6.1 is more than the downward bias in the estimated product dummies.

Figure 6.2: *Willingness to pay for products*



Note: each dot is a product, with the value on the y-axis indicating the estimated willingness-to-pay for the product using the Universal logit estimates in column (3) of Table 5.1 and the value on the x-axis indicating the estimated willingness-to-pay for the product using the PPH logit estimates in column (4) of Table 5.1. For each set of estimates, willingness-to-pay is calculated as $\hat{\delta}_j/\hat{\alpha}$.

13

-0.54*0.159/-1.853=0.0463.

This illustration in the on-the-go chocolate market in the UK shows how failing to account for unobserved choice set heterogeneity can significantly bias preference estimates and the economic inferences that might arise from them in discrete choice demand estimation. From a business strategy perspective, failing to account for unobserved choice set heterogeneity would lead a researcher to conclude that consumers are less sensitive to price, that advertising has a greater impact on demand, and that most products are more desired than consumer preferences truly indicate.

7 Conclusion

In this paper we have addressed the consequences of unobserved choice set heterogeneity on discrete choice demand estimation. We show that unobserved choice set heterogeneity causes bias and propose an estimation method based on the logit family of preferences and "sufficient sets" of consumer choices to consistently estimate preferences regardless of the form of the choice set formation process. Our solution can be applied to both cross-section and panel data and we show how the assumptions underlying particular sufficient sets can be tested. We illustrate the method using an application to on-the-go demand for chocolate in the UK and show that consumers' price sensitivity is biased towards zero and consumers' advertising sensitivity is biased away from zero, with important implications for firms' strategic decision-making.

These results show that treating carefully consumers' choice sets is critically important in the estimation of discrete choice demand models. We see two very promising directions for future research. The first is to further model consumers' process of choice set formation and estimate models that predict both choice sets and choices. This is indeed necessary in order to obtain point estimates on choice probabilities, market shares, and elasticities. Increasingly available individual-level data on both consumer choices and the sequence of decisions that precede these choices make such efforts increasingly viable. As noted above, our approach can serve as a useful complement to this approach by providing an alternative set of assumptions under which preferences can be estimated, providing a useful specification test for any particular model of choice set formation.

The second is to extend the methods presented here to recover information about distributions of choice sets without modeling how they are formed. While relying on subsets of choices in consumers' sufficient sets permits consistent preference estimation, it doesn't exploit all the potential information in the data. Given consistent information of preferences, what can be learned about distributions of choice sets from variation in the data that reflects the effects of both preferences and choice set formation? In follow up work, D'Haultfœuille et al. (2016) have begun to address this topic.

References

Abaluck, Jason and Abi Adams, "Discrete Choice Models with Consideration Sets: Identification from Asymmetric Cross-Derivatives," 2016.

Arellano, Manuel, *Panel data econometrics*, Oxford University Press, 2003.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica*, 1995, pp. 841–890.

_ , _ , and _ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, pp. 68–105.

Bierlaire, Michel, Denis Bolduc, and Daniel McFadden, "The estimation of generalized extreme value models from choice-based samples," *Transportation Research Part B: Methodological*, 2008, *42* (4), 381–394.

Bordalo, P., N. Gennaioli, and A. Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2014, *121* (5), 803–843.

Chade, Hector and Lones Smith, "Simultaneous search," *Econometrica*, 2006, *74* (5), 1293–1307.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 1980, *47*, 225–238.

Chiang, Jeongwen, Siddhartha Chib, and Chakravarthi Narasimhan, "Markov chain Monte Carlo and models of consideration set and parameter heterogeneity," *Journal of Econometrics*, 1998, *89* (1), 223–248.

Conlon, Christopher T and Julie Holland Mortimer, "Demand estimation under incomplete product availability," *American Economic Journal-Microeconomics*, 2013, *5* (4), 1–30.

Crawford, Gregory S, "Endogenous product choice: A progress report," *International Journal of Industrial Organization*, 2012, *30* (3), 315–320.

Currie, Janet, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania, "The Effect of Fast Food Restaurants on Obesity and Weight Gain," *American Economic Journal: Economic Policy*, 2010, *2* (3), 32–63.

D'Haultfœuille, Xavier, Alessandro Iaria, and Guillaume Lecué, "Recovering the Distribution of Consideration Sets from Choice Data," 2016. Working Paper, CREST (Paris).

_ **and** _ , "A Convenient Method for the Estimation of the Multinomial Logit Model with Fixed Effects," *Economics Letters*, 2016, *141*, 77–79.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer price search and platform design in internet commerce," Technical Report, National Bureau of Economic Research 2014.

Draganska, M., M. Mazzeo, and K. Seim, "Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions," *Quantitative Marketing and Economics*, 2009, *7*, 105–146.

Draganska, Michaela and Daniel Klapper, "Choice set heterogeneity and the role of advertising: An analysis with micro and macro data," *Journal of Marketing Research*, 2011, *48* (4), 653–669.

Dubé, JP, G. Hitsch, and P Manchanda, "An empirical model of advertising dynamics," *Quantitative Marketing and Economics*, 2005, *3*, 107–144.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell, "The Effects of Banning Advertising in Junk Food Markets," *CEPR Discussion Papers*, 2016.

Eizenberg, A., "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 2014, *81*, 1003–1045.

Eliaz, Kfir and Ran Spiegler, "Consideration sets and competitive marketing," *The Review of Economic Studies*, 2011, *78* (1), 235–262.

Fox, Jeremy T, "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 2007, *38* (4), 1002–1019.

G., S. Becker and K. M. Murphy, "A Simple Theory of Advertising as a Good or Ban," *The Quarterly Journal of Economics*, 1993, *108* (4), 941–964.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

Gaynor, Martin, Carol Propper, and Stephan Seiler, "Free to Choose? Reform, Choice and Consideration Sets in the English National Health Service," *American Economic Review*, 2016, *forthcoming.*

Gilbride, Timothy J and Greg M Allenby, "A choice model with conjunctive, disjunctive, and compensatory screening rules," *Marketing Science*, 2004, *23* (3), 391–406.

Goeree, Michelle Sovinsky, "Limited Information and Advertising in the U.S. Personal Computer Industry," *Econometrica*, 2008, *76* (5), 1017–1074.

Hausman, Jerry A and Paul A Ruud, "Specifying and testing econometric models for rank-ordered data," *Journal of econometrics*, 1987, *34* (1), 83–104.

Hausman, Jerry and Daniel McFadden, "Specification tests for the multinomial logit model," *Econometrica*, 1984, pp. 1219–1240.

Imbens, Guido W and Charles F Manski, "Confidence intervals for partially identified parameters," *Econometrica*, 2004, pp. 1845–1857.

Kőszegi, Botond and Adam Szeidl, "A model of focusing in economic choice," *The Quarterly Journal of Economics*, 2013, *128* (1), 53–104.

Leicester, Andrew and Zoe Oldfield, "An analysis of consumer panel data," *IFS Working Papers W09/09*, 2009.

Lerman, S. and C. Manski, "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, 1981.

los Santos, Babur De, Ali Hortaçsu, and Matthijs R Wildenbeest, "Testing models of consumer search using data on web browsing and purchasing behavior," *The American Economic Review*, 2012, *102* (6), 2955–2980.

Lu, Zhentong, "Estimating Multinomial Choice Models with Unobserved Choice Sets," 2016. Working paper, University of Wisconsin-Madison.

Manski, Charles F, "The structure of random utility models," *Theory and decision*, 1977, *8* (3), 229–254.

Manzini, Paola and Marco Mariotti, "Stochastic choice and consideration sets," *Econometrica*, 2014, *82* (3), 1153–1176.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y Ozbay, "Revealed attention," *American Economic Review*, 2012, pp. 2183–2205.

Matejka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–98.

McFadden, Daniel, "Modeling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Vol. 1, North-Holland, 1978, pp. 75–96.

_ , Kenneth Train et al., "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 2000, *15* (5), 447–470.

Moraga-González, José Luis, Zsolt Sándor, and Matthijs R Wildenbeest, "Consumer Search and Prices in the Automobile Market," *Tinbergen Institute Discussion Paper*, 2015, pp. 1–59.

Nevo, Aviv, "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Nierop, Erjen Van, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses, "Retrieving unobserved consideration sets from household panel data," *Journal of Marketing Research*, 2010, *47* (1), 63–74.

Roberts, John H and James M Lattin, "Development and testing of a model of consideration set composition," *Journal of Marketing Research*, 1991, pp. 429–440.

Ruud, Paul A, "Tests of specification in econometrics," *Econometric Reviews*, 1984, *3* (2), 211–242.

Shapiro, B, "Positive Spillovesr and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants," 2015.

Stigler, George J, "The economics of information," *The journal of political economy*, 1961, pp. 213–225.

Train, K., *Discrete Choice Methods with Simulation*, Cambridge University Press, 2009.

Walters, Christopher R, "The demand for effective charter schools," Technical Report, National Bureau of Economic Research 2014.

Appendices

A Proof of Proposition 1

Note that, given $\mathcal{CS}_i^\star = c$, i's *true* choice probability conditional on $\mathcal{E}_i = e$ is:

$$\begin{aligned}
\Pr\left[Y_i = j \middle| \mathcal{E}_i = e, \mathcal{CS}_i^\star = c, \theta\right] &= \prod_{t=1}^{T} \frac{\Pr[Y_{it} = j_t | CS_{it}^\star = c_t, \theta]}{\sum_{r_t \in e_t \cap c_t} \Pr[Y_{it} = r_t | CS_{it}^\star = c_t, \theta] + \sum_{k_t \in e_t \setminus c_t} \Pr[Y_{it} = k_t | CS_{it}^\star = c_t, \theta]} \\
&= \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{it}, \theta\right)\right)}{\sum_{r_t \in E_{it} = e_t \cap CS_{it}^\star = c_t} \exp\left(V_{irt}\left(X_{it}, \theta\right)\right)} \\
&= \Pr\left[Y_i = j \middle| \mathcal{E}_i = e \cap \mathcal{CS}_i^\star = c, \theta\right],
\end{aligned} \tag{A.1}$$

since $e_t = (e_t \cap c_t) \cup (e_t \setminus c_t)$ and $\Pr[Y_{it} = k_t | CS_{it}^\star = c_t, \theta] = 0$ for all $k_t \notin CS_{it}^\star = c_t$. In other words, since Assumption 1 implies the IIA property only when $\mathcal{E}_i = e \subseteq \mathcal{CS}_i^\star = c$, (A.1) is not guaranteed to equal (2.4). By expressing (A.1) in terms of (2.4), we obtain:

$$\begin{aligned}
\Pr\left[Y_i = j \middle| \mathcal{E}_i = e \cap \mathcal{CS}_i^\star = c, \theta\right] &= \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{it}, \theta\right)\right)}{\sum_{r_t \in e_t \cap c_t} \exp\left(V_{irt}\left(X_{it}, \theta\right)\right)} \times \frac{\sum_{m_t \in e_t} \exp\left(V_{imt}\left(X_{it}, \theta\right)\right)}{\sum_{m_t \in e_t} \exp\left(V_{imt}\left(X_{it}, \theta\right)\right)} \\
&= \prod_{t=1}^{T} \frac{\exp\left(V_{ijt}\left(X_{it}, \theta\right) - \ln\left(\frac{\sum_{r_t \in e_t \cap c_t} \exp(V_{irt}(X_{it}, \theta))}{\sum_{m_t \in e_t} \exp(V_{imt}(X_{it}, \theta))}\right)\right)}{\sum_{m_t \in E_{it} = e_t} \exp\left(V_{imt}\left(X_{it}, \theta\right)\right)} \\
&= \prod_{t=1}^{T} \frac{\exp(V_{ijt}\left(X_{it}, \theta\right) - \ln(\pi_{it}))}{\sum_{m_t \in E_{it} = e_t} \exp\left(V_{imt}\left(X_{it}, \theta\right)\right)}.
\end{aligned} \tag{A.2}$$

IF part. Suppose $e_t \cap c_t \subset e_t$ for some t's, then $\ln(\pi_{it}) < 0$ for those t's and models (2.4) and (A.1) will differ. In this case, if estimation proceeds on the basis of model (2.4), the likelihood function will be mistakenly ignoring a sequence of up to T fixed effects for each i, $\ln(\pi_{it})$'s, which are functions of the rest of the model.

ONLY IF part. Differently, if $e_t \subseteq c_t$ for all t's, then $\ln(\pi_{it}) = 0$ for all t's and (A.1) equals (2.4). Consequently, the model used in estimation will correspond to the true conditional choice model. $\square$

B Why Existing Methods Do Not Solve the Problem of Unobserved Choice Set Heterogeneity

Several methods routinely implemented in the literature do not generally avoid bias from unobserved choice set heterogeneity. Econometricians are used to thinking that most choice-based sampling issues in logit models can be addressed by adding a full set of alternative-specific constants (see Lerman and Manski (1981) and Bierlaire et al. (2008)). This is not the case for our problem, because $\ln(\pi_{it})$ from model (A.2) is an individual-time specific term that only affects the numerator of the logit formula.

We also cannot treat the term causing the bias, $\ln(\pi_{it})$, as a random coefficient, because the estimation of models with random coefficients, such as the mixed logit model, relies on an independence assumption between the distribution of random coefficients and the observables of the model.[14] Equation (A.2) makes clear that, by construction, the distribution of $\ln(\pi_{it})$ is a function of the X_{it}. Furthermore, we cannot treat the $\ln(\pi_{it})$'s as unobserved fixed effects and estimate them with individual-time specific dummy variables because of the incidental parameters problem.

The most popular approach used in the literature is to jointly model both the choice set formation process and the purchase decision given a choice set. As originally discussed by Manski (1977), the *unconditional* probability of i selecting choice sequence j can be written as:

$$\Pr[Y_i = j|\theta, \gamma] = \sum_{c \in C_i^\star} \Pr[Y_i = j|\mathcal{CS}_i^\star = c, \theta] \Pr[\mathcal{CS}_i^\star = c|\gamma], \tag{B.1}$$

where $C_i^\star$ is the collection of sets of possible choice sequences to which consumer type i can be matched. By having information on the matching process between consumer types and choice sets, researchers can integrate out unobserved choice set heterogeneity in a matter analogous to that routinely done with unobserved preference heterogeneity.

While the general Manski (1977) approach is common, there are two instances in which its application is unlikely to solve the problem. First, any *specific* model that integrates out choice sets is

[14]See (Berry et al., 1995, p865), (Berry et al., 2004, p76), (McFadden et al., 2000, p447), and (Nevo, 2001, p314).

likely to give rise to the same kinds of bias outlined earlier if the support of the collection of choice sets over which expectations are taken, $C_i^\star$, is misspecified. Intuitively, giving positive weight in the integral to choice sets that incorrectly include unconsidered alternatives causes bias. Second, even when $\Pr[Y_i = j|\theta, \gamma]$ is correctly specified, in practice its estimation may prove difficult. In particular, it is likely to suffer from a curse of dimensionality because the number of elements in $C_i^\star$ grows exponentially in the number of products J sold in the market.

We illustrate these two problems in the context of Goeree (2008), one of the most widely cited examples of Manski (1977)'s approach. In our notation, equation (3) in Goeree (2008) can be written as:

$$\Pr[Y_{it} = j_t|\theta, \gamma] = \underbrace{\sum_{c\in C^j} \frac{\exp\left(V_{ij_tt}(X_{it}, \theta)\right)}{\sum_{r\in c}\exp\left(V_{irt}(X_{it}, \theta)\right)}}_{\Pr[Y_{it} = j_t|CS_{it}^\star = c, \theta]} \underbrace{\left[\prod_{l\in c}\phi_{ilt}(\gamma)\prod_{k\notin c}\left(1-\phi_{ikt}(\gamma)\right)\right]}_{\Pr_{it}\left[CS_{it}^\star = c|\gamma\right]}, \tag{B.2}$$

where C^j is the collection of *all* the choice sets that include product j. This specification relies on:

- Our Assumption 1, with $\Pr[Y_{it} = j_t|CS_{it}^\star = c, \theta] = \dfrac{\exp\left(V_{ij_tt}(X_{it}, \theta)\right)}{\sum_{r\in c}\exp\left(V_{irt}(X_{it}, \theta)\right)}$ and
- The assumption that consideration of each product is *independent* from the consideration of the other products: $\Pr_{it}\left[CS_{it}^\star = c|\gamma\right] = \prod_{l\in c}\phi_{ilt}(\gamma)\prod_{k\notin c}\left(1-\phi_{ikt}(\gamma)\right)$.

Even given this second assumption, in Goeree's application the total number of PCs is still very large (over 2,000), so the non-parametric estimation of all the ϕ's is not feasible. Consequently, Goeree (2008) further assumes that

$$\phi_{ilt}(\gamma) = \frac{\exp\left(W_{ilt}(\gamma)\right)}{1+\exp\left(W_{ilt}(\gamma)\right)}. \tag{B.3}$$

This implies that *every* $c \in C^j$ will have a strictly positive probability in the distribution of choice sets for each (i,t), so that $\Pr_{it}\left[CS_{it}^\star = c|\gamma\right] > 0$ for every (i,t). However, it may be that for some (i,t) combination $\Pr_{it}\left[CS_{it}^\star = c|\gamma\right] = 0$ for some c, i.e. c is not in the *support* of the choice set distribution to which individual i can be matched to in period t:

$$\Pr_{it}\left[CS_{it}^\star = c|\gamma\right] = \begin{cases} \prod\limits_{l\in c}\phi_{ilt}(\gamma)\prod\limits_{k\notin c}\left(1-\phi_{ikt}(\gamma)\right) & \text{if } c\in C_{it}^{j\star} \\ 0 & \text{if } c\in C^j\setminus C_{it}^{j\star}, \end{cases} \tag{B.4}$$

where $C_{it}^{j\star}$ is the collection of choice sets to which individual i can possibly be matched to in period t. Since $C_{it}^{j\star}$ is typically unobserved and heterogeneous across (i,t) combinations, model (B.2) and (B.3) will suffer from *support misspecification* whenever there exists even one observation where the true collection of unobserved choice sets to which an individual can actually be matched is restricted, i.e. $\exists\,(i,j,t)$ combination such that $C_{it}^{j\star} \subset C^j$. With support misspecification, any estimator of model (B.2) and (B.3) will be inconsistent.

To be clear, computing expectations over the power set of the universal set, as with C^j in (B.2), would not be a problem if we could estimate a truly flexible specification for ϕ that was able to accomodate $\Pr_{it}\left[CS_{it}^{\star} = c|\gamma\right] = 0$ whenever necessary. The problem arises because we are not usually able to estimate a truly flexible model for ϕ, and we need to make additional assumptions along the lines of (B.3).

C Quantifying the Size of the Bias

Table C.1 provides monte carlo evidence that quantifies the size of the bias from mistakenly attributing to consumers choices that were not available to them. The three panels describe the relative importance of alternative features of the choice environment on the size of the bias arising from unobserved choice set heterogeneity. In each panel, a different specification is described. We report the average and standard deviation of the bias (across 20 replications) in the estimated coefficient that arises if the researcher uses the full choice set instead of the true (heterogeneous and unobserved) choice set. This percentage bias is reported in the last column in the table.

In the first panel we report the (lack of) bias in the baseline model where all consumers have the full choice set available to them. In the second panel we show that the bias increases with the share of consumers facing constrained choice sets. In the third panel we show that the bias increases with the extent of the constraint in choice sets. In the final panel we show that the bias increases the more the consumer prefers the option that is not included in their true choice set (but is mistakenly included in the universal choice set used in estimation).

The results are intuitive and show the consequences of failing to account for unobserved choice set heterogeneity. The size of the bias can be substantial. Increasing the share of consumers that have missing elements in the choice set from 10-50% increases the average bias from assuming all consumers have access to the full choice set from 11.2-36.0%, increasing the share of unavailable products from 1 of 5 to 3 of 5 increases the average bias from 26.3-57.0%, and increasing the average utility gap between the unavailable first-best and chosen second-best alternative from approximately 10-70% increases the average bias from assuming the first-best option was available from 19.0-38.6%.

Table C.1: *The size of bias in Universal Logit*

	Bias (StdDev)	% Bias
Baseline		
100% of consumers have full choice set	0.005	0.3%
	(0.032)	
Increasing the share of individuals with constrained choice sets		
90% have full choice set, 10% choose from 4 out of 5	-0.223	-11.2%
	(0.021)	
70% have full choice set, 30% choose from 4 out of 5	-0.525	-26.3%
	(0.013)	
50% have full choice set, 50% choose from 4 out of 5	-0.719	-36.0%
	(0.007)	
Increasing the share of products randomly removed from choice sets		
30% have 4 out of 5 available	-0.525	-26.3%
	(0.013)	
30% have 3 out of 5 available	-0.839	-42.0%
	(0.007)	
30% have 2 out of 5 available	-1.139	-57.0%
	(0.003)	
Increasing product differentiation		
First-best choice is slightly preferred	-0.379	-19.0%
$(V_1 - V_2)/V_1 \simeq 10\%$, $\sigma_X^2 = 1.75$	(0.016)	
First-best choice is preferred	-0.471	-23.6%
$(V_1 - V_2)/V_1 \simeq 30\%$, $\sigma_X^2 = 2.5$	(0.012)	
First-best choice is strongly preferred	-0.578	-28.9%
$(V_1 - V_2)/V_1 \simeq 50\%$, $\sigma_X^2 = 3.5$	(0.012)	
First-best choice is very strongly preferred	-0.771	-38.6%
$(V_1 - V_2)/V_1 \simeq 70\%$, $\sigma_X^2 = 6$	(0.009)	

We consider a population of 1,000 consumers making a sequence of choices over 10 choice situations. On each choice situation they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in equation (2.1). The systematic utility is linear with homogenous preferences, $V_{ijt}(X_{it}, \theta) = \delta_j + X_{jt}\beta$, and the unobserved portion of utility, ϵ_{ijt}, is distributed Gumbel. In the baseline specification, X_{jt} is drawn from a Normal distribution with mean 0 and variance 5, $\delta_j = 0$ for all j's, and $\beta = 2$. We simulate 20 replications. In the final panel, 30% of consumers have their first-best choice removed.

D Proof of Proposition 2

$$
\begin{aligned}
&\Pr[Y_i = j | f(Y_i) = r, \theta] \\
&= \Pr[Y_i = j | f(Y_i) = r, \mathcal{CS}_i^\star = c, \theta] \\
&= \frac{\Pr[f(Y_i) = r, Y_i = j, \mathcal{CS}_i^\star = c | \theta, \gamma]}{\Pr[f(Y_i) = r, \mathcal{CS}_i^\star = c | \theta, \gamma]} \\
&= \frac{\Pr[f(Y_i) = r | Y_i = j, \mathcal{CS}_i^\star = c, \theta] \Pr[Y_i = j | \mathcal{CS}_i^\star = c, \theta] \Pr[\mathcal{CS}_i^\star = c | \gamma]}{\Pr[f(Y_i) = r | \mathcal{CS}_i^\star = c, \theta] \Pr[\mathcal{CS}_i^\star = c | \gamma]} \\
&= \frac{\Pr[f(Y_i) = r | Y_i = j, \mathcal{CS}_i^\star = c, \theta] \Pr[Y_i = j | \mathcal{CS}_i^\star = c, \theta]}{\sum\limits_{k \in \mathcal{U}} \Pr[f(Y_i) = r | Y_i = k, \mathcal{CS}_i^\star = c, \theta] \Pr[Y_i = k | \mathcal{CS}_i^\star = c, \theta]} \qquad \text{(D.1)} \\
&= \frac{\prod\limits_{t=1}^{T} \frac{\exp(V_{jt}(X_{it}, \theta))}{\sum_{v \in CS_{it}^\star = c_t} \exp(V_{vt}(X_{it}, \theta))}}{\sum\limits_{k: f(k) = r} \prod\limits_{t=1}^{T} \frac{\exp(V_{kt}(X_{it}, \theta))}{\sum_{v \in CS_{it}^\star = c_t} \exp(V_{vt}(X_{it}, \theta))}} \\
&= \frac{\prod_{t=1}^{T} \exp(V_{jt}(X_{it}, \theta))}{\sum_{k: f(k) = r} \prod_{t=1}^{T} \exp(V_{kt}(X_{it}, \theta))}
\end{aligned}
$$

Assumption 1 and Condition 1 imply the IIA property, and the first equality follows from its definition. The second and third equalities follow from the definition of conditional probability, while the fourth follows from the law of total probability. In the fourth equality, $\mathcal{U}$ is the universal set of *all* choice sequences. The fifth equality follows from $\Pr[f(Y_i) = r | Y_i = k, \mathcal{CS}_i^\star = c, \theta]$ being 1 for any k such that $f(k) = r$ and 0 otherwise. This is the case since, conditional on a realization of Y_i, $Y_i = k$, $f(Y_i)$ is not a random set: $f(k)$ is either r with probability one, or different from r with probability one. In the last equality, $\sum_{v \in CS_{it}^\star = c_t} \exp(V_{vt}(X_{it}, \theta))$ cancels out. Finally, consistency of the conditional Maximum Likelihood estimator derived from $\Pr[Y_i = j | f(Y_i) = r, \theta]$ follows from McFadden (1978). $\square$

E A Simple Example Demonstrating Sufficient Sets

In Section 3.1, we described alternative sufficient sets that can underpin our estimation procedure. This Appendix provides a simple example that illustrates a few different sufficient sets and their implied choice probabilities, as well as describes how each can accommodate different economic environments.

Suppose there are three products, $j = 1, 2$, and 3, and two choice situations, $t = 1$ and 2. Consider a consumer type i with true but unobserved choice set containing all products in both choice situations, $CS_{i1}^{\star} = CS_{i2}^{\star} = \{1, 2, 3\}$. In this case, $\mathcal{CS}_i^{\star} = \{1, 2, 3\} \times \{1, 2, 3\} = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3,$ Assume consumer type i's indirect utility of choosing product j in choice situation t is $U_{ijt} = \delta_j + X_{ijt}\beta + \epsilon_{ijt}$, where $\theta = [\delta_1, \delta_2, \delta_3, \beta^{'}]^{'}$. Finally, i's observed choice sequence is $j = 1$ in the first choice situation and $j = 3$ in the second, so that $Y_i = (1, 3)$.

- *Fixed Effect (FE) logit*: The sufficient set that gives rise to the FE logit is made of all permutations of the observed choice sequence, $\mathcal{P}(Y_i)$. Since $Y_i = (1, 3)$, $f_{FE}(1, 3) = \mathcal{P}(1, 3) = \{(1, 3), (3, 1)\}$. Given this, the contribution to the likelihood function for consumer type i is:

  $$\Pr[Y_i = (1, 3)|\beta, \{(1, 3), (3, 1)\}] = \frac{\exp(X_{i11}\beta)\exp(X_{i32}\beta)}{\exp(X_{i11}\beta)\exp(X_{i32}\beta) + \exp(X_{i31}\beta)\exp(X_{i12}\beta)}.$$

  A drawback of the FE logit is that it does not allow the identification of the parameters of time-invariant regressors. This is evident in the example, as one cannot estimate the product-specific constants, δ_j.

- *Full Purchase History (FPH) and Inter-Personal (IP) logit*: In both the FPH and the IP logits, the sufficient set is made of the choice sequences compatible with products $j = 1$ and $j = 3$ being available in both choice situations, $H_i \times H_i$, where $H_i = \bigcup_{t=1}^{2} \{Y_{it}\}$. Thus, $f_{FPH}(1, 3) = H_i \times H_i = \{(1, 1), (3, 3), (1, 3), (3, 1)\}$. In this case, i's contribution to the likelihood function of the FPH logit is:

$$\Pr[Y_i = (1,3)|\theta, \{(1,1),(1,3),(3,1),(3,3)\}]$$

$$= \frac{\exp(\delta_1 + X_{i11}\beta)\exp(\delta_3 + X_{i32}\beta)}{\sum_{(j_1,j_2)\in\{(1,1),(1,3),(3,1),(3,3)\}} \exp(\delta_{j_1} + X_{ij_11}\beta)\exp(\delta_{j_2} + X_{ij_22}\beta)}$$

$$= \frac{\exp(\delta_1 + X_{i11}\beta)}{\exp(\delta_1 + X_{i11}\beta) + \exp(\delta_3 + X_{i31}\beta)} \times \frac{\exp(\delta_3 + X_{i32}\beta)}{\exp(\delta_1 + X_{i12}\beta) + \exp(\delta_3 + X_{i32}\beta)}.$$

While the FPH and the IP logit can be analytically expressed in a similar way, the economic assumptions they rely on are very different. With the FPH logit we have the same individual making several purchase decisions at different points in time, while in the IP logit there are different individuals of the same type each making a separate purchase decision at a specific point in time.

- *Past Purchase History logit*: In the PPH logit, the sufficient set—at any point in time—includes all choice sequences consistent with those products purchased up to that choice situation, $H_{i1} \times H_{i2}$, where $H_{it} = \bigcup_{b=1}^{t} \{Y_{ib}\}$. Thus, $f_{PPH}(1,3) = H_{i1} \times H_{i2} = \{(1,1),(1,3)\}$. In this case, consumer type i's contribution to the likelihood function is:

$$\Pr[Y_i = (1,3)|\theta, \{(1,1),(1,3)\}]$$

$$= \frac{\exp(\delta_1 + X_{i11}\beta)\exp(\delta_3 + X_{i32}\beta)}{\exp(\delta_1 + X_{i11}\beta)\exp(\delta_1 + X_{i12}\beta) + \exp(\delta_1 + X_{i11}\beta)\exp(\delta_3 + X_{i32}\beta)}$$

$$= 1 \times \frac{\exp(\delta_3 + X_{i32}\beta)}{\exp(\delta_1 + X_{i12}\beta) + \exp(\delta_3 + X_{i32}\beta)}.$$

F Specification Testing Appendix

This Appendix provides a proof of Proposition 3 and an example of how the specification tests proposed in Section 3.2 can work in practice.

F.1 Proof of Proposition 3

F.1.1 Proof of Result (1)

From Proposition 2, we can re-write for every i the probability of the observed choice sequence j given $f_Z(Y_i) = z \subseteq f_L(Y_i) = l$ as:

$$\begin{aligned}
\Pr\left[Y_i = j|, f_L(Y_i) = l, \theta\right] &= \frac{\prod_{t=1}^{T} \exp(V_{jt}(X_{it}, \theta))}{\sum_{k:f_L(k)=l} \prod_{t=1}^{T} \exp(V_{kt}(X_{it}, \theta))} \\
&= \Pr\left[Y_i = j|f_Z(Y_i) = z, \theta\right] \left(\frac{\Pr\left[Y_i = j|f_L(Y_i) = l, \theta\right]}{\Pr\left[Y_i = j|f_Z(Y_i) = z, \theta\right]}\right) \\
&= \frac{\prod_{t=1}^{T} \exp(V_{jt}(X_{it}, \theta))}{\sum_{q:f_Z(q)=z} \prod_{t=1}^{T} \exp(V_{qt}(X_{it}, \theta))} \frac{\sum_{q:f_Z(q)=z} \prod_{t=1}^{T} \exp(V_{qt}(X_{it}, \theta))}{\sum_{k:f_L(k)=l} \prod_{t=1}^{T} \exp(V_{kt}(X_{it}, \theta))}, \\
&= \Pr\left[Y_i = j|f_Z(Y_i) = z, \theta\right] \Pr\left[Y_i \in f_Z(Y_i) = z|f_L(Y_i) = l, \theta\right]
\end{aligned}$$

where $\Pr\left[Y_i \in f_Z(Y_i) = z|f_L(Y_i) = l, \theta\right]$ is the probability that a choice sequence belongs to the "smaller" set z relative to the "larger" set l. By multiplying $\Pr\left[Y_i = j|f_L(Y_i) = l, \theta\right]$ across all consumer types i's and by taking the logarithm of this likelihood function, result (1) follows with $l_{\triangle}(\theta) = \sum_{i=1}^{I} \ln\left(\Pr\left[Y_i \in f_Z(Y_i) = z|f_L(Y_i) = l, \theta\right]\right)$.

F.1.2 Proof of Result (2)

Given result (1) of Proposition 3, results (2a) and (2b) follow from the *Factorization Theorem* of (Ruud, 1984, result (1), p.24).

F.1.3 Proof of Result (3)

Given result (1) of Proposition 3, result (3a) follows from the *Factorization Theorem* by (Ruud, 1984, result (3), p.24), while result (3b) follows from (Ruud, 1984, pp.28-9).

Result (3c) can be proved as follows. (Ruud, 1984, result 2, p.24) shows that $\widehat{\theta}_L$ is asymptotically equivalent to $\mathrm{Var}\left(\widehat{\theta}_L\right) \mathrm{Var}\left(\widehat{\theta}_Z\right)^{-1} \widehat{\theta}_Z + \mathrm{Var}\left(\widehat{\theta}_L\right) \mathrm{Var}\left(\widehat{\theta}_{\triangle}\right)^{-1} \widehat{\theta}_{\triangle}$. This implies that $\mathrm{Cov}\left(\widehat{\theta}_L, \widehat{\theta}_Z\right) = \mathrm{Cov}\left(\mathrm{Var}\left(\widehat{\theta}_L\right) \mathrm{Var}\left(\widehat{\theta}_Z\right)^{-1} \widehat{\theta}_Z, \widehat{\theta}_Z\right) = \mathrm{Var}\left(\widehat{\theta}_L\right) \mathrm{Var}\left(\widehat{\theta}_Z\right)^{-1} \mathrm{Var}\left(\widehat{\theta}_Z\right) = \mathrm{Var}\left(\widehat{\theta}_L\right)$, where the first equality follows from result (2a). Consequently, $\mathrm{Var}\left(\widehat{\theta}_L-\widehat{\theta}_Z\right)=\mathrm{Var}\left(\widehat{\theta}_L\right)+\mathrm{Var}\left(\widehat{\theta}_Z\right)-2\mathrm{Cov}\left(\widehat{\theta}_L,\widehat{\theta}_Z\right)=\mathrm{Var}\left(\widehat{\theta}_Z\right)-\mathrm{Var}\left(\widehat{\theta}_L\right)$. □

F.2 Testing Procedures

The various sufficient sets introduced in section (3.1) rely on the following economic assumptions:

- f_{FE}: Choice set stability across T choice situations *and* the possibility of having unobserved preference heterogeneity in the form of individual-alternative specific fixed effects.
- f_{FPH}: Choice set stability across T choice situations *and* no unobserved preference heterogeneity.
- f_{PPH}: Choice set evolution in the form of entry-but-no-exit or exit-but-no-entry across T choice situations *and* no unobserved preference heterogeneity.

There are two possibilities for making comparisons across models based on different sufficient sets f's. The first possibility is to compare f_{FE}, f_{FPH}, and f_{PPH} for choice sequences of constant length T. The second possibility is to fix a specific f, say f_{FE}, and to compare choice sequences with some of their *sub*–sequences: for example, the sequence $1,2,\ldots,T^L$ can be split into two mutually exclusive sub–sequences $1,2,\ldots,T^Z$ and $T^Z+1,\ldots,T^L$, and this gives rise to different f_{FE}'s, f_{FE}^Z and f_{FE}^L such that $f_{FE}^Z\left(Y_i\right)\subseteq f_{FE}^L\left(Y_i\right)$ for any $Y_i\in\mathcal{CS}_i^\star=c$. We will discuss each testing possibility in turn.

F.2.1 Comparisons of Different f's with Constant T

For choice sequences of a given length T, it can be shown that $f_{FE}\left(Y_i\right)\subseteq f_{FPH}\left(Y_i\right)$ and that $f_{PPH}\left(Y_i\right)\subseteq f_{FPH}\left(Y_i\right)$ for any $Y_i\in\mathcal{CS}_i^\star=c$. These can be seen in our previous example from Appendix (E). Suppose $Y_i=(1,3)$. Then $f_{FE}\left(1,3\right)=\mathcal{P}\left(1,3\right)=\{(1,3),(3,1)\}$, $f_{FPH}\left(1,3\right)=\{1,3\}\times\{1,3\}=\{(1,1),(3,3),(1,3),(3,1)\}$, and $f_{PPH}\left(1,3\right)=\{1\}\times\{1,3\}=\{(1,1),(1,3)\}$. Note that there is no clear "inclusion" relationship between $f_{FE}\left(Y_i\right)$ and $f_{PPH}\left(Y_i\right)$.

Given the results from Proposition 3, the above relationships among sufficient sets lead to two possible classes of tests. The first is about choice set stability and the second about preference heterogeneity.

F.2.2 Choice Set Stability

f_{FPH} and f_{PPH} are both based on the same assumption of absence of unobserved preference heterogeneity. However, they rely on different assumptions regarding the evolution of choice sets across choice situations: f_{FPH} assumes that unobserved choice sets do not change along the whole choice sequence, while f_{PPH} allows for the entry of new alternatives in the unobserved choice while comparing choice situation t to $t+1$. On the one hand, if unobserved choice sets were stable, then both f's would give rise to consistent estimators $\widehat{\theta}_{FPH}$ and $\widehat{\theta}_{PPH}$, but result (2b) from Proposition 3 tells us that $\widehat{\theta}_{FPH}$ would be more efficient than $\widehat{\theta}_{PPH}$. On the other hand, if unobserved choice sets were subject to entry–but–no–exit of alternatives, then only $\widehat{\theta}_{PPH}$ would be consistent. It follows that, under the maintained assumption of no unobserved preference heterogeneity, a test for H_0: *(choice set stability in* $1, 2, \ldots, T$*)* is $LR = 2\left[l_{PPH}\left(\widehat{\theta}_{PPH}\right) + l_{\triangle}\left(\widehat{\theta}_{\triangle}\right) - l_{FPH}\left(\widehat{\theta}_{FPH}\right)\right]$.

F.2.3 Preference Homogeneity

The sufficient sets f_{FPH} and f_{FE} are both based on the same assumption of unobserved choice set stability in $1, 2, \ldots, T$. However, they rely on different assumptions regarding unobserved preference heterogeneity: f_{FPH} assumes that there is no unobserved preference heterogeneity, while f_{FE} allows for (i,j)–specific fixed effects. On the one hand, if there were no unobserved preference heterogeneity, then both f's would give rise to consistent estimators $\widehat{\theta}_{FPH}$ and $\widehat{\theta}_{FE}$, but Proposition 3 tells us that $\widehat{\theta}_{FPH}$ would be more efficient than $\widehat{\theta}_{FE}$. On the other hand, if unobserved preference heterogeneity were present in a form encompassed by (i,j)–specific fixed effects, then only $\widehat{\theta}_{FE}$ would be consistent. It follows that, under the maintained assumption of choice set stability in $1, 2, \ldots, T$, a test for H_0: *(preference homogeneity)* is $LR = 2\left[l_{FE}\left(\widehat{\theta}_{FE}\right) + l_{\triangle}\left(\widehat{\theta}_{\triangle}\right) - l_{FPH}\left(\widehat{\theta}_{FPH}\right)\right]$.

F.2.4 Comparisons of Same f with Different Choice Sub–Sequences

It is always possible to split choice sequences of length $1, 2, \ldots, T^L$ into two (or more) *mutually exclusive* sub–sequences $1, 2, \ldots, T^Z$ and $T^Z + 1, \ldots, T^L$. Then $f^Z_{FE}(Y_i) \subseteq f^L_{FE}(Y_i)$ for any $Y_i \in \mathcal{CS}^{\star}_i = c$. The same holds for both f_{FPH} and f_{PPH}. This method of making comparisons allows us to test for choice set stability in several alternative ways, but it does not enable us to test for preference homogeneity.

F.2.5 Choice Set Stability: $\boldsymbol{f_{FE}}$ Example

In this section we will show with an example why $f^Z_{FE}(Y_i) \subseteq f^L_{FE}(Y_i)$ for any $Y_i \in \mathcal{CS}^\star_i = c$ and, afterward, we will discuss how to use this fact to construct tests of choice set stability.

Suppose $J = 5$, $T^L = 4$, and that consumer type i is observed to make the choice sequence $Y_i = (j_1, j_2, j_3, j_4) = (3, 5, 5, 4)$.[15] By considering the observed choice sequence "at once," $Y_i = (3, 5, 5, 4)$ can be re–ordered in 12 different choice sequences.[16] Collect these sequences into the set $f^L_{FE}(Y_i) = l$. Assume that $V_{ijt}(X_{it}, \theta) = \delta_{ij} + X_{ijt}\beta$. Then, i's likelihood contribution given $f^L_{FE}(Y_i) = l$ is:

$$\Pr\left[Y_i = (3,5,5,4)\middle|\, f^L_{FE}(Y_i) = l, \beta\right]$$

$$= \frac{\exp\left(\left(X_{i31} + X_{i52} + X_{i53} + X_{i44}\right)\beta\right)}{\sum\limits_{(j_1,j_2,j_3,j_4)\in f^L_{FE}(j_1,j_2,j_3,j_4)=l} \exp\left(\left(X_{ij_11} + X_{ij_22} + X_{ij_33} + X_{ij_44}\right)\beta\right)}. \tag{F.1}$$

Differently, by splitting i's observed choice sequence into two mutually exclusive pairs of choices $Y_{i1} = (3, 5)$ and $Y_{i3} = (5, 4)$, we get $f^Z_{FE}(Y_{i1}) = \{(3,5),(5,3)\}$ and $f^Z_{FE}(Y_{i3}) = \{(5,4),(4,5)\}$. Then, i's likelihood contribution given $f^Z_{FE}(Y_{i1}) = z_1$ and $f^Z_{FE}(Y_{i3}) = z_3$ is:

$$\Pr\left[Y_i = (3,5,5,4)\middle|\, f^Z_{FE}(Y_{i1}) = z_1, f^Z_{FE}(Y_{i3}) = z_3, \beta\right]$$

$$= \frac{\exp\left(\left(X_{i31} + X_{i52}\right)\beta\right)}{\exp\left(\left(X_{i31} + X_{i52}\right)\beta\right) + \exp\left(\left(X_{i51} + X_{i32}\right)\beta\right)} \times \tag{F.2}$$

$$\times \frac{\exp\left(\left(X_{i53} + X_{i44}\right)\beta\right)}{\exp\left(\left(X_{i53} + X_{i44}\right)\beta\right) + \exp\left(\left(X_{i43} + X_{i54}\right)\beta\right)}.$$

By multiplying the binomial logits in (F.2), we get:

$$\Pr_i\left[Y_i = (3,5,5,4)\middle|\, f^Z_{FE}(Y_i) = z, \beta\right] =$$

$$\frac{\exp\left(\left(X_{i31} + X_{i52} + X_{i53} + X_{i44}\right)\beta\right)}{\sum\limits_{(j_1,j_2,j_3,j_4)\in f^Z_{FE}(j_1,j_2,j_3,j_4)=z} \exp\left(\left(X_{ij_11} + X_{ij_22} + X_{ij_33} + X_{ij_44}\right)\beta\right)}, \tag{F.3}$$

[15] Alternative one in the first choice situation, alternative three in the second choice situation, etc.

[16] These sequences are: $(3,5,5,4)$, $(5,3,5,4)$, $(5,5,3,4)$, $(5,5,4,3)$, $(4,3,5,5)$, $(3,4,5,5)$, $(3,5,4,5)$, $(5,3,4,5)$, $(5,4,3,5)$, $(5,4,5,3)$, $(4,5,3,5)$, and $(4,5,5,3)$.

where $f^Z_{FE}(Y_i) = z$ collects sequences: $(3,5,5,4)$, $(3,5,4,5)$, $(5,3,5,4)$, and $(5,3,4,5)$. Consequently $f^Z_{FE}(Y_i) = z \subseteq f^L_{FE}(Y_i) = l$. In this example, f^Z_{FE} only uses information about 4 of the 12 possible choice sequences in f^L_{FE}. This implies that if unobserved choice sets were stable, then estimator $\widehat{\beta}^L_{FE}$ would be more efficient than $\widehat{\beta}^Z_{FE}$.

Moreover, the FE logit estimated on choice sub–sequences may "discard" some choice situations: in the current example of sub–sequences of length two, whenever $j_t = j_{t+1}$ in $Y_{it} = (j_t, j_{t+1})$, then "fragment" Y_{it} of Y_i will not be used in estimation. For example, if i were observed to choose the sequence $Y_i = (3,4,5,5)$, then only $Y_{i1} = (3,4)$ would contribute to the likelihood function $l^Z_{FE}(\beta)$, while $l^L_{FE}(\beta)$ would still use the whole sequence $Y_i = (3,4,5,5)$. More precisely, if $Y_i = (3,4,5,5)$ were observed, then $f^L_{FE}(3,4,5,5) = f^L_{FE}(3,5,5,4) = l$ would still contain the same 12 choice sequences, while model (F.2) would collapse to:

$$\begin{aligned}
&\Pr\left[Y_i = (3,4,5,5) \middle| f^Z_{FE}(Y_{i1}) = h_1, f^Z_{FE}(Y_{i3}) = h_3, \beta\right] \\
&= \frac{\exp\left((X_{i31} + X_{i42})\beta\right)}{\exp\left((X_{i31} + X_{i42})\beta\right) + \exp\left((X_{i41} + X_{i32})\beta\right)} \times \\
&\times \frac{\exp\left((X_{i53} + X_{i54})\beta\right)}{\exp\left((X_{i53} + X_{i54})\beta\right)} \qquad \text{(F.4)} \\
&= \frac{\exp\left((X_{i31} + X_{i42} + X_{i53} + X_{i54})\beta\right)}{\exp\left((X_{i31} + X_{i42} + X_{i53} + X_{i54})\beta\right) + \exp\left((X_{i41} + X_{i32} + X_{i53} + X_{i54})\beta\right)} \\
&= \Pr\left[Y_i = (3,4,5,5) \middle| f^Z_{FE}(Y_i) = h, \beta\right],
\end{aligned}$$

which is also equivalent to $\Pr\left[Y_{i1} = (3,4) \middle| f^Z_{FE}(Y_{i1}) = h_1, \beta\right]$. In this case, then, $f^Z_{FE}(Y_{i1}) = h_1 \subseteq f^Z_{FE}(Y_i) = z \subseteq f^L_{FE}(Y_i) = l$. By Proposition 3, we can rank the corresponding estimators in terms of their relative efficiency. As a consequence, by splitting up choice sequences into mutually exclusive sub–sequences, we can face also this further loss of efficiency.

Model (F.1) requires stronger assumptions than model (F.3) for its consistent estimation. Consistent estimation of model (F.1) requires that alternatives $\{3,4,5\} \subseteq CS^\star_{it} = c_t$, $t = 1,2,3,4$. However, consistent estimation of model (F.3) only requires that $\{3,5\} \subseteq CS^\star_{it} = c_t$, $t = 1,2$ and that $\{4,5\} \subseteq CS^\star_{it} = c_t$, $t = 3,4$. In this example, if $4 \notin CS^\star_{it} = c_t$, $t = 1$ or 2, or $3 \notin CS^\star_{it} = c_t$, $t = 3$ or 4, then estimation of model (F.1) would not be consistent, while estimation of model (F.3) would.

These differences in consistency and relative efficiency suggest a Hausman–like test for unobserved choice set stability. If $\{3,4,5\} \subseteq CS^{\star}_{it} = c_t$, $t = 1,2,3,4$, then estimation of both model (F.1) and model (F.3) would be consistent. However, estimation of model (F.1) would be more efficient than estimation of model (F.3). If $4 \notin CS^{\star}_{it} = c_t$, $t = 1$ or 2 or $3 \notin CS^{\star}_{it} = c_t$, $t = 3$ or 4, then only estimation of model (F.3) would be consistent. It follows that, under the maintained assumption of unobserved preference heterogeneity in a form encompassed by (i,j)–specific fixed effects, a test for H_0: *(choice set stability in* $1,2,3$ *and* 4*)* is $LR = 2\left[l^{Z}_{FE}\left(\widehat{\beta}^{Z}_{PPH}\right) + l_{\triangle}\left(\widehat{\beta}_{\triangle}\right) - l^{L}_{FE}\left(\widehat{\beta}^{L}_{FE}\right)\right]$.

G Bounding Choice Probabilities and Elasticities

As summarized in Section 4, we are often interested in functions of parameters, for example, willingness to pay, elasticities, welfare or analysis of counterfactuals, such as evaluating the effects of a change in tax policy or a merger between manufacturers. This section describes how to construct bounds on these parameters within our framework.

We can obtain point estimates of preference parameters and so, for example, willingness to pay. However, without further information or restrictions on the true choice sets, we are not able to point-estimate consumer type-specific choice probabilities and hence average choice probabilities. Still, under some weak assumptions, we can "bound" them. Also, given the convenient relationship between consumer type-specific elasticities and consumer type-specific choice probabilities in logit models, we can as well bound consumer type-specific elasticities and their averages.

For expositional simplicity, suppose that indirect utilities are linear in parameters:

$$V_{jt}(X_{it}, \theta) = \delta_j + X_{ijt}\beta. \tag{G.1}$$

Denote by $Y_{it} = j_t$ whether product j_t is chosen by consumer type i in choice situation t, and by $f_t(Y_i) = r_{it}$ whether i's sufficient set collects products r_{it} in choice situation t.

Note that $f_t(Y_i) \subseteq CS^{\star}_{it}$ and that $CS^{\star}_{it} \subseteq S_{it}$. In what follows we use these conditions to bound the true but unobserved denominator of the logit choice probabilities for any X_{imt}, δ, and β:

$$\sum_{m \in f_t(Y_i) = r_{it}} \exp(\delta_m + X_{imt}\beta) \leq \sum_{m \in CS^{\star}_{it} = c_{it}} \exp(\delta_m + X_{imt}\beta) \leq \sum_{m \in S_{it} = s_{it}} \exp(\delta_m + X_{imt}\beta). \tag{G.2}$$

For brevity, denote $Pr^S_{ijt}(\theta) = \Pr[Y_{it} = j_t | S_{it} = s_{it}, \theta]$, $Pr^{CS^\star}_{ijt}(\theta) = \Pr[Y_{it} = j_t | CS^\star_{it} = c_{it}, \theta]$, and $Pr^f_{ijt}(\theta) = \Pr[Y_{it} = j_t | f_t(Y_i) = r_{it}, \theta]$. It follows from (G.2) that for any $j_t \in f_t(Y_i) = r_{it}$:

$$Pr^S_{ijt}(\theta) \leq Pr^{CS^\star}_{ijt}(\theta) \leq Pr^f_{ijt}(\theta). \tag{G.3}$$

Observe that $Pr^f_{ijt}(\theta) = \Pr[Y_{it} = j_t | f_t(Y_i) = r_{it}, \theta]$ takes the usual logit form whenever $j_t \in r_{it}$, but that it equals zero whenever $j_t \notin r_{it}$. Hence, for those $j_t \in s_{it}$ but $j_t \notin r_{it}$, $Pr^f_{ijt}(\theta)$ will not be a valid upper bound for $Pr^{CS^\star}_{ijt}(\theta)$: even if $j_t \notin r_{it}$, it can still be the case that $j_t \in CS^\star_{it} = c_{it}$ and so that $Pr^{CS^\star}_{ijt}(\theta) > 0$. Similarly, among the $j_t \in s_{it}$ that $j_t \notin r_{it}$, there can be some $j_t \notin c_{it}$. But for those $j_t \in s_{it}$ that $j_t \notin c_{it}$, $Pr^S_{ijt}(\theta) > Pr^{CS^\star}_{ijt}(\theta) = 0$: $Pr^S_{ijt}(\theta)$ will not be a valid lower bound for $Pr^{CS^\star}_{ijt}(\theta)$. It is then unclear how to bound $Pr^{CS^\star}_{ijt}(\theta)$ for those $j_t \in s_{it}$ but $j_t \notin r_{it}$. However, it is always possible to construct bounds for the probability with which i would purchase j_t if indeed j_t were to be *added* to their true but unobserved choice set, $CS^\star_{it} \cup \{j_t\} = c_{it} \cup \{j_t\}$:

$$Pr^{CS^\star \cup j}_{ijt}(\theta) = \Pr\left[Y_{it} = j_t | CS^\star_{it} \cup \{j_t\} = c_{it} \cup \{j_t\}, \theta\right] = \frac{\exp\left(\delta_{j_t} + X_{ij_tt}\beta\right)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp\left(\delta_m + X_{imt}\beta\right)}. \tag{G.4}$$

By defining $Pr^{S\cup j}_{ijt}(\theta)$ and $Pr^{f\cup j}_{ijt}(\theta)$ analogously, note that $Pr^{S\cup j}_{ijt}(\theta) = Pr^S_{ijt}(\theta)$, $Pr^{CS^\star\cup j}_{ijt}(\theta) = Pr^{CS^\star}_{ijt}(\theta)$, and $Pr^{f\cup j}_{ijt}(\theta) = Pr^f_{ijt}(\theta)$ for any $j_t \in r_{it}$, while $Pr^{S\cup j}_{ijt}(\theta) \leq Pr^{CS^\star\cup j}_{ijt}(\theta)$ and $Pr^{CS^\star\cup j}_{ijt}(\theta) \leq Pr^{f\cup j}_{ijt}(\theta)$ for any $j_t \notin r_{it}$. Using these facts, we can then complement condition (G.3) for those $j_t \notin r_{it}$ and propose choice probability bounds for all (i, j, t) combinations:

$$Pr^{S\cup j}_{ijt}(\theta) \leq Pr^{CS^\star\cup j}_{ijt}(\theta) \leq Pr^{f\cup j}_{ijt}(\theta). \tag{G.5}$$

Condition (G.5) can be used to construct bounds for functions of consumer type choice probabilities, such as *average* choice probabilities or elasticities. The average choice probability of product j_t for a certain group of consumer types $i = 1, \ldots, I_t$ can be bounded by:

$$I_t^{-1} \sum_{i=1}^{I_t} Pr^{S\cup j}_{ijt}(\theta) \leq I_t^{-1} \sum_{i=1}^{I_t} Pr^{CS^\star\cup j}_{ijt}(\theta) \leq I_t^{-1} \sum_{i=1}^{I_t} Pr^{f\cup j}_{ijt}(\theta). \tag{G.6}$$

With logit demand and linear indirect utilities (G.1), consumer type i's own- and cross-price elasticities are:

$$\begin{aligned}
\xi_{it}^{jj}(X_{it},\theta) &= \beta_p p_{j_t t}(1 - Pr_{ijt}^{CS^\star \cup j}(\theta)) \\
&= \beta_p p_{j_t t}\left(1 - \frac{\exp(\delta_{j_t} + X_{ij_t t}\beta)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp(\delta_m + X_{imt}\beta)}\right), \\
\xi_{it}^{jk}(X_{it},\theta) &= \beta_p p_{k_t t} Pr_{ikt}^{CS^\star \cup j}(\theta) \\
&= -\beta_p p_{k_t t}\left(\frac{\exp(\delta_{k_t} + X_{ik_t t}\beta)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp(\delta_m + X_{imt}\beta)}\right),
\end{aligned} \tag{G.7}$$

where $p_{j_t t}$ is j_t's price in choice situation t and β_p is the price coefficient. As (G.7) makes clear, even though we may have a consistent estimator of $\delta = [\delta_1, \ldots, \delta_j, \ldots, \delta_J]$ and β, we still do not know the exact $CS_{it}^\star \cup \{j_t\} = c_{it} \cup \{j_t\}$ for each i and t, and thus the true $Pr_{ijt}^{CS^\star \cup j}(\theta)$, $\forall j_t \in CS_{it}^\star \cup \{j_t\} = c_{it} \cup \{j_t\}$.

Given (G.7), (G.5), and $\beta_p < 0$, we obtain the following bounds on the elasticities for any j_t, k_t, X_{ijt}, δ, and β:

$$\begin{gathered}
\underbrace{\beta_p p_{j_t t}(1 - Pr_{ijt}^{f \cup j}(\theta))}_{\text{Lower (in abs. value) Bound}} \le \xi_{it}^{jj}(X_{it},\theta) \le \underbrace{\beta_p p_{j_t t}(1 - Pr_{ijt}^{S \cup j}(\theta))}_{\text{Upper (in abs value) Bound}} \\
\underbrace{-\beta_p p_{k_t t} Pr_{ikt}^{S \cup j}(\theta)}_{\text{Lower Bound}} \le \xi_{it}^{jk}(X_{it},\theta) \le \underbrace{-\beta_p p_{kt} Pr_{ikt}^{f \cup j}(\theta)}_{\text{Upper Bound}}
\end{gathered} \tag{G.8}$$

G.1 Confidence Intervals for Elasticity Bounds

We construct confidence intervals for these identification regions following Imbens and Manski (2004).

For notational simplicity, we limit our discussion to a single elasticity term $\xi_{it}^{jk}(X_{it},\theta)$, although the same ideas can be extended to the collection of all elasticities. Refer to the upper and lower bounds of $\xi_{it}^{jk}(X_{it},\theta)$ in (G.8) as to $\overline{\xi_{it}^{jk}}(X_{it},\theta)$ and $\underline{\xi_{it}^{jk}}(X_{it},\theta)$, respectively. Denote the elasticity *bounds* of $\xi_{it}^{jk}(X_{it},\theta)$ by the 2×1 vector $B\left(\xi_{it}^{jk}(X_{it},\theta)\right)=\left[\underline{\xi_{it}^{jk}}(X_{it},\theta),\overline{\xi_{it}^{jk}}(X_{it},\theta)\right]'$ and the corresponding elasticity *interval* from (G.8) by $IN\left(\xi_{it}^{jk}(X_{it},\theta)\right)$. Then, given X_{it} and our consistent $\widehat{\theta}$, we can estimate the elasticity bounds $B\left(\xi_{it}^{jk}(X_{it},\theta)\right)$ by $B\left(\xi_{it}^{jk}(X_{it},\widehat{\theta})\right)$. We derive the corresponding $100\,(1-\alpha)$ percent confidence interval $CI_{1-\alpha}$ from condition:

$$\inf_{\xi_{it}^{jk}\in IN\left(\xi_{it}^{jk}(X_{it},\theta)\right)}\left\{\lim_{I\to\infty}\Pr\left[\xi_{it}^{jk}\in CI_{1-\alpha}\right]\right\}\geq 1-\alpha. \tag{G.9}$$

Since our estimator is consistent and asymptotically normal, i.e., $\widehat{\theta}\sqrt{I}\xrightarrow{d}\mathcal{N}\left(\theta,V_\theta\right)$, by the delta-method:

$$B\left(\xi_{it}^{jk}(X_{it},\widehat{\theta})\right)\sqrt{I}\xrightarrow{d}\mathcal{N}\left(B\left(\xi_{it}^{jk}(X_{it},\theta)\right),\frac{\partial B\left(\xi_{it}^{jk}(X_{it},\theta)\right)}{\partial\theta'}V_\theta\frac{\partial B\left(\xi_{it}^{jk}(X_{it},\theta)\right)'}{\partial\theta'}\right). \tag{G.10}$$

Refer to the 2×2 asymptotic variance–covariance matrix of $B\left(\xi_{it}^{jk}(X_{it},\widehat{\theta})\right)$ as to $\varSigma_{B\left(\xi_{it}^{jk}\right)}$. It follows that, whenever $f_t\left(Y_i\right)\cup\{j_t\}=r_{it}\cup\{j_t\}$ is a strict subset of $S_{it}\cup\{j_t\}=s_{it}\cup\{j_t\}$, so that for any X_{it} and θ, $\underline{\xi_{it}^{jk}}(X_{it},\theta)<\overline{\xi_{it}^{jk}}(X_{it},\theta)$, condition (G.9) is satisfied by:

$$CI_{1-\alpha}=\left[\underline{\xi_{it}^{jk}}(X_{it},\widehat{\theta})-q_{1-\alpha}\sqrt{\varSigma^{11}_{B\left(\xi_{it}^{jk}\right)}},\overline{\xi_{it}^{jk}}(X_{it},\widehat{\theta})+q_{1-\alpha}\sqrt{\varSigma^{22}_{B\left(\xi_{it}^{jk}\right)}}\right], \tag{G.11}$$

where $q_{1-\alpha}$ is the $(1-\alpha)^{th}$ quantile of the standard normal distribution.

In the extreme case in which $f_t(Y_i)\cup\{j_t\}=r_{it}\cup\{j_t\}=S_{it}\cup\{j_t\}=s_{it}\cup\{j_t\}$, $\underline{\xi_{it}^{jk}}(X_{it},\theta)=\overline{\xi_{it}^{jk}}(X_{it},\theta)$ for any X_{it} and θ, and (G.11) is invalid. This is due to a discontinuity at $\underline{\xi_{it}^{jk}}(X_{it},\theta)=\overline{\xi_{it}^{jk}}(X_{it},\theta)$, since in that case the coverage of the interval is only $100\,(1-2\alpha)\,\%$ rather than the nominal $100\,(1-\alpha)\,\%$. (See Imbens and Manski (2004) for a modification of (G.11) that overcomes this problem.) However, note that (a) both $f_t(Y_i)\cup\{j_t\}=r_{it}\cup\{j_t\}$ and $S_{it}\cup\{j_t\}=s_{it}\cup\{j_t\}$ are always perfectly observed by the econometrician, so that the appropriate $CI_{1-\alpha}$ can always be implemented and (b) in our empirical application $f_t(Y_i)\cup\{j_t\}\subset S_{it}\cup\{j_t\}$ for every i and t.

H Data Appendix

In Section 6 we present an illustrative empirical example; here we describe the data used in that empirical example in greater detail.

H.1 Purchase data

We data from the Kantar Worldpanel (see Leicester and Oldfield (2009) and Dubois et al. (2016)). Kantar collects data on purchases made on-the-go from a random selection of individuals in the households that participate in the Worldpanel. The Kantar Worldpanel on-the-go survey is collected from individuals who recording purchases that they make on-the-go for immediate consumption using their mobile phone.

We use data on 297 working-age women (ages 19-59) without children who are the main shoppers in their household. This is a fairly homogeneous group of consumers for whom we have self-reported measures on their own TV viewing behaviour. We use information on 9387 purchase occasions over the period 2010-2011. A purchase occasion is when the women is observed purchases a snack of any form on-the-go.

At any one point in time there are up to 250 different types of chocolate products available in the market. The outside option, when a chocolate bar is not purchases, has a 74% market share. The three largest market share products are Cadbury Twirl, with a market share of 3.2%, a large KitKat, with a market share of 1.2% and Cadbury Crunchie wiht a market share of 1%. Brand is a more aggregate level than product, there are 192 brands, with Cadbury Twirl also being the largest brand (made up of 3 products), Cadbury Dairy Milk the second largest brand (made up of 51 different products), and KitKat the third largest brand (with 6 products).

Prices are measured on each individual transaction; we aggregate them to the level of the outlet and week. We consider the outlet that we observe the individual shopping in as chosen before, and independently from, choosing the specific chocolate product. We consider four types of outlets - large national chains (34% of sales), news agents (14% of sales), vending machines (4% of sales), and other types of small stores and outlets (49% of sales). 95% of prices range from 25 pence to £1.70, with a few exceptional items available at very low price (for example, single small Cadbury Creme Eggs for 15 pence) and a few large items (for example, a 400g Dairy Milk bar for 4.59).

H.2 Advertising data

We use two measures of advertising exposure. Both convert weekly advertising ("flows") into an advertising "stock"; advertising stocks are the depreciated cumulation of the flows. We use advertising data collected by AC Nielsen. The data contain aggregate advertising expenditure across all platforms (cinema, internet, billboards, press, radio and TV) and detailed disaggregate information for TV advertising. TV advertising is by far the most important form of advertising, accounting for 61.8% of total expenditure between 2009-2010, and we therefore focus on TV advertising expenditure. For each TV ad, we have information on the time the ad was aired, the brand that was advertised, the TV station, the duration of the ad, the cost of the ad, and the TV shows that immediately preceded and followed the ad.

The time path of advertising varies across brands, and all brands have some periods of zero advertising expenditure. These non-smooth strategies are rationalised in the model of Dubé et al. (2005) when the effectiveness of advertising can vary over time. This variation in the timing of adverts, coupled with variation in TV viewing behaviour, will generate considerable household level variation in exposure to brand level advertising.

The first measure we use at the brand level and is the aggregate minutes of TV adverts aired during the week at the brand level to define advertising flows. Denote the aggregate minutes of adverts for brand b in week t as s_{bt}, following Dubé et al. (2005) and Shapiro (2015), we convert this to a *stock* of advertising exposure as follows:

$$a_{bt} = \sum_{k=0}^{t} \eta^k s_{bt-k}, \tag{H.1}$$

and we set $\eta = 0.75$.

Length is measured in seconds; the average advert in the UK is 30 seconds. The mean of the stock variable a_{bt} is 1887 or 31.45 minutes.

Our second measure follows Goeree (2008) and Dubois et al. (2016) and measures advertising exposure at the individual level. We use detailed information about when individual adverts were aired on television matched with self-reported viewing information to construct individual level measures of exposure to brand advertising.

We combine the information on when ads were aired with information on households' TV viewing behaviour in order to get a household-level measure of exposure to each ad. We use data from the

Kantar media survey, an annual survey asking the main shopper in the household about their TV subscriptions and TV viewing behaviour. Households are asked "How often do you watch ...?" for 206 different TV shows, and can choose to answer Never, Hardly Ever, Sometimes or Regularly. At least one ad for chocolate is shown before, during, or after 112 of these shows (many of the shows with no chocolate advertising are on BBC channels, which are prohibited from showing ads). From this information we define the variable:

$$w_{is} = \begin{cases} 1 & i \text{ reports they ``regularly'' or ``sometimes'' watch show } s \\ 0 & \text{otherwise} \end{cases} \tag{H.2}$$

Households are also asked "How often do you watch ...?" 65 different TV channels and when they usually watch TV. In particular, for weekdays, Saturday, and Sunday and for 9 different time periods,[17] households are asked questions like "Do you watch live TV on Saturdays at breakfast time (6.00-9.30am)?" In each case, the household can answer Never, Hardly Ever, Sometimes or Regularly. We use this information, along with information on where the household lives (some TV channels are regional), to construct the variable:

$$w_{ikc} = \begin{cases} 1 & i \text{ says they ``regularly'' or ``sometimes'' watch on the day and time slot } k \\ & \qquad \text{and ``regularly'' or ``sometimes'' watch channel } c \\ & \qquad \text{and they live in the region in which } c \text{ is aired (or the channel is national)} \\ 0 & \text{otherwise} \end{cases} \tag{H.3}$$

We combine the data on household viewing behaviour with the detailed data on individual ads to create a household-specific measure of exposure to advertising. Variation in TV viewing behaviour creates considerable variation in the timing and extent of exposure an individual household has to ads of a specific brand. This leads to cross-household variation in advertising exposure that is plausibly unrelated to idiosyncratic shocks to chocolate products.

Denote by T_{bskct} the duration of time that an ad for brand b is shown during show s on day and time slot k on channel c during week t. From the viewing data, we construct an indicator variable of whether household i was likely to be watching channel c on day and time slot k during show s,

[17] Breakfast time 6.00am-9.30am, Morning 9.30am-12.00 noon, Lunchtime 12.00 noon-2.00pm, Early afternoon 2.00pm-4.00pm, Late afternoon 4.00pm-6.00pm, Early evening 6.00pm-8.00pm, Mid evening 8.00pm-10.30pm, Late evening 10.30-1.00am and Night time 1.00am-6.00am.

w_{iskc}. If show s is among the 206 specific shows households were asked for viewing information we set $w_{iskc} = w_{is}$, otherwise we set $w_{iskc} = w_{ikc}$. From this we define the household's total exposure to advertising of brand b during week t as:

$$s_{ibt} = \sum_{s,c,k} w_{iskc} T_{bskct}. \tag{H.4}$$

We define the *flow* as:

$$a_{ibt} = \sum_{k=0}^{t} \eta^k s_{ibt-k} \tag{H.5}$$

where $\eta = 0.75$

This stock is measured in seconds (and is divided by 1000 when includes in the regression). It is 0 for individuals that do not watch TV, or only watch public TV (the BBC), to a mean of 54 minutes of cumulated exposure to adverts for a particular brand.