DISCUSSION PAPER SERIES

DP11675

PREFERENCE ESTIMATION WITH UNOBSERVED CHOICE SET HETEROGENEITY USING SUFFICIENT SETS

Gregory S. Crawford, Rachel Griffith and Alessandro Iaria

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

PREFERENCE ESTIMATION WITH UNOBSERVED CHOICE SET HETEROGENEITY USING SUFFICIENT SETS

Gregory S. Crawford, Rachel Griffith and Alessandro Iaria

Discussion Paper DP11675
Published 01 December 2016
Submitted 23 October 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PREFERENCE ESTIMATION WITH UNOBSERVED CHOICE SET HETEROGENEITY USING SUFFICIENT SETS

Abstract

This paper introduces "sufficient sets" and shows how they can address unobserved choice set heterogeneity. Motivated by economic theory, a sufficient set is a combination of a consumer's observed choices that is a subset of their true but unobserved choice set. We introduce a Sufficient Set logit that "differences out" consumers' true choice sets, permitting consistent parameter estimation, and show how to use sufficient sets to semi-parametrically estimate preferences and lessen the computational burden of methods that integrate over the distribution of choice sets. We illustrate our ideas using Monte Carlo simulations and by estimating demand for chocolate in the UK.

JEL Classification: N/A

Keywords: unobserved choice set heterogeneity, discrete choice demand estimation, sufficient sets, attention, search, endogenous product choice

Gregory S. Crawford - gregory.crawford@econ.uzh.ch
University of Zurich and CEPR

Rachel Griffith - rgriffith@ifs.org.uk
IFS and University of Manchester and CEPR

Alessandro Iaria - alessandro.iaria@ensae.fr
CREST, ENSAE, University Paris-Saclay and IFS

Acknowledgements
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).

Preference Estimation with Unobserved Choice Set Heterogeneity using Sufficient Sets*

Gregory S. Crawford† Rachel Griffith‡ Alessandro Iaria§

October 12, 2017

Abstract

This paper introduces "sufficient sets" and shows how they can address unobserved choice set heterogeneity. Motivated by economic theory, a sufficient set is a combination of a consumer's observed choices that is a subset of their true but unobserved choice set. We introduce a Sufficient Set logit that "differences out" consumers' true choice sets, permitting consistent parameter estimation, and show how to use sufficient sets to semi-parametrically estimate preferences and lessen the computational burden of methods that integrate over the distribution of choice sets. We illustrate our ideas using Monte Carlo simulations and by estimating demand for chocolate in the UK.

*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

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

§Dept. of Economics, University of Bristol, alessandro.iaria@bristol.ac.uk

1 Introduction

There are many reasons why consumer choice sets may be heterogeneous and unobserved by econometricians, including limited consumer attention, search, or endogenous product choices by firms. Failing to account for unobserved choice set heterogeneity will generally cause estimates of preference parameters to be inconsistent. The dominant approach to tackle this problem empirically is to integrate over the distribution of unobserved choice sets (Manski (1977), Goeree (2008), Van Nierop et al. (2010)). This approach requires additional information about the distribution of choice sets or functional form assumptions on the choice set formation process and suffers from a curse of dimensionality in estimation.

In this paper, we introduce the concept of "sufficient sets" and show how they can help address the problem of preference estimation in the presence of unobserved choice set heterogeneity. A sufficient set is a combination of a consumer's observed choices that is a subset of their true but unobserved choice set. We illustrate how the choice of sufficient sets can be linked to economic theory and show how sufficient sets can help address the problem of unobserved choice set heterogeneity in three ways. First, we propose a class of conditional logit models that provide consistent estimates of preference parameters by conditioning individuals' choice probabilities on sufficient sets. In this context, a sufficient set represents a sufficient statistic for each individual's true but unobserved choice set. It prevents violations of the Independence of Irrelevant Alternatives (IIA) property in models with Gumbel errors which would otherwise cause inconsistency by "differencing out" consumers' true choice sets from the likelihood function. Second, we show how sufficient sets can be used in semi-parametric models, in the context of Fox (2007)'s Pairwise Maximum Score Estimator, by identifying subsets of consumers' choices on which to base preference estimation in the presence of unobserved choice set heterogeneity. Finally, sufficient sets can be used in models that "integrate out" unobserved choice set heterogeneity by introducing restrictions that can substantially reduce the curse of dimensionality inherent in the estimation of such models, regardless of their maintained distributional assumptions.

We begin our analysis by introducing the concept of a sufficient set in simple logit models with i.i.d. Gumbel errors. We show that including alternatives in estimation that are not in a consumer's true choice set introduces individual-time-specific fixed effects that typically lead to inconsistent parameter estimates. Relying on results from McFadden (1978), we show that a Sufficient Set logit that conditions on a sufficient set of each consumer's true but unobserved choice set yields consistent parameter estimates. It does not, however, allow us to point-identify functions of preferences that depend on the distribution of choice sets in the population of consumers. We therefore show how to use sufficient sets in combination with assumptions about supersets of consumers' choice sets to bound predicted choice probabilities, elasticities, and consumer surplus. We also propose Hausman specification tests to compare alternative sufficient sets within a particular application.

In cases where the assumptions underlying the Sufficient Set logit aren't attractive to researchers, we demonstrate how to semi-parametrically estimate preferences in the presence of heterogeneous unobserved choice sets by using the Pairwise Maximum Score estimator of Fox (2007). Fox (2007) shows how to extend the Maximum Score estimator originally developed by Manski (1975) to consistently estimate preferences from subsets of the true choice sets. Our sufficient sets can provide the choice subsets his method requires in order to estimate preferences in environments with unobserved choice set heterogeneity, a heretofore untried application.

We contribute to the sizeable empirical literature built on Manski (1977) that specifies particular models of choice set formation and "integrates out" heterogeneous unobserved choice sets (Chiang et al. (1998), Goeree (2008), Van Nierop et al. (2010), Draganska and Klapper (2011), and Conlon and Mortimer (2013)). These require additional information with respect to standard choice data, for example information that predicts choice set variation. In a recent paper, Abaluck and Adams (2017) provide identification results for this class of models, showing that whenever some alternatives are not in the choice sets of some individuals, the discrete-choice analogy of Slutsky symmetry will be violated and that one can therefore use deviations from Slutsky symmetry to separately identify consideration from preferences given consideration. Our approach is complementary to this literature, both as a means to informally test the maintained assumptions of a particular model (e.g. by comparing preference estimates out of both approaches) as well as to reduce the computational cost of this approach by integrating over all combinations of choices in each consumer's sufficient set rather than combinations of all choices.

Critically, the economic characteristics of the choice environment can inform the choice of sufficient sets. For example, settings characterized by non-sequential or fixed-sample search (e.g., Morgan and Manning (1985)) imply a choice environment that is stable for any individual consumer over time, but (possibly) varying across consumers. These models imply sufficient sets based on the collection of products chosen by a consumer over time, what we call a Full Purchase History sufficient set. Settings characterized by sequential search (e.g., Caplin et al. (2011)), limited attention (e.g., Eliaz and Spiegler (2011), or consumer focus (e.g., Kőszegi and Szeidl (2013)) imply choice environments that evolve over time, and suggest sufficient sets based on the accumulation of products chosen by a consumer in the past, what we call a Past Purchase History sufficient set. Cross-sectional settings where a group of individuals face a common choice environment, as for example in the analyses of Currie et al. (2010) of whether greater availability of fast food increases obesity or Gaynor et al. (2016) of hospital choice by doctors, suggest sufficient sets made of the collection of products observed to be chosen by individuals in the relevant group, what we call an Inter-Personal sufficient set. These are but a few examples of sufficient sets; they are neither mutually exclusive nor exhaustive. Sufficient sets can be combined or devised to reflect a large range of choice environments.

We demonstrate the usefulness of our approach in both Monte Carlo exercises and an empirical illustration. The illustration estimates price and advertising sensitivity for demand on-the-go for chocolate among adult women in the UK. Advertising can affect both a consumer's valuation of a specific product and the likelihood that the product enters the consumer's choice set. We first estimate a multinomial logit model from an enlarged choice set including all chocolate products available in the type of outlet in which the consumer is shopping (what we call the universal choice set). We then compare the estimates from this model to estimates from a Sufficient Set logit that conditions on each consumer's past purchase history at the type of outlet from which they are currently purchasing, a generalization of one of our proposed sufficient sets. The estimates of the impact of advertising on purchase probabilities differ significantly in the two models; our interpretation is that with the universal choice set we are incorrectly including alternatives in consumers' choice sets that were not in fact present, and that this biases upwards the estimated advertising sensitivity. These results are consistent with models of imperfect consumer attention like that in Eliaz and Spiegler (2011). We find that estimated price sensitivities are biased towards zero and that estimated willingness-to-pay for individual products are on the whole upwardly biased.

Our work relates to several literatures in economics and marketing. In addition to the empirical literature that integrates over unobserved choice sets summarized above, there is a fast-growing theoretical literature in which limited attention is used to rationalize apparently incongruent consumer and firm behaviors that motivates our interest in developing empirical approaches that accommodate these theories. These include, for example, consumer attention as in Eliaz and Spiegler (2011), rational inattention as in Gabaix (2014) and Matejka and McKay (2015), search as in Janssen and Moraga-González (2004), screening rules as in Gilbride and Allenby (2004), models of salience as in Bordalo et al. (2014), and focus as in Kőszegi and Szeidl (2013). In our empirical application, we develop sufficient sets and estimate preferences inspired by the work of Eliaz and Spiegler (2011); similar applications of these other theories are also possible using our approach.

Our contribution relates to the literature on the identification of preferences and consideration sets uniquely from choice data. In addition to Abaluck and Adams (2017), summarized above, Masatlioglu et al. (2012) describe whether and how preferences and attention (or consideration) can be separately identified when both are non-stochastic, while choices and choice sets are observed. 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. Our work is complementary to these papers that focus on *identification*, in that we provide a tractable empirical approach for the consistent *estimation* of consumer preferences when choice sets are heterogeneous and unobserved.

Finally, our paper relates to recent work by Lu (2016). 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 probability computed on a subset of her true choice set, and below by a probability computed on a superset of her true choice set. This allows him to set-identify preferences by making assumptions, or by having additional data, on supersets and subsets of individuals' unobserved choice sets. Our approaches are complementary. As Lu (2016) does not specify how to construct valid choice subsets, our sufficient sets can be combined with his approach. More broadly, Lu (2016)'s method requires weaker assumptions on the model's error structure than our Sufficient Set logit, but may lack power if his choice subsets or supersets are far from consumers' true choice sets and face computational difficulties as the scale of the application grows.

The rest of the paper is structured as follows. In section 2 we introduce sufficient sets and show how they can be used both to consistently estimate preferences without specifying the choice set formation process as well as lessen the computational burden of estimators that do specify the choice process and integrate out over the distribution of unobserved choice sets. In section 3 we provide examples of sufficient sets motivated by economic theories of search, inattention, and focus and show how they apply in both panel and cross-sectional data settings. In section 4 we present a number of results for the Sufficient Set logit, the application of sufficient sets when unobserved preferences are distributed i.i.d. Gumbel. We present Monte Carlo results on the performance of sufficient sets, show how to test between alternative sufficient sets, and describe how we can point-identify preferences and bound choice probabilities, elasticities, and consumer surplus. In section 5, we present an application that looks at the impact of price and advertising on demand on-the-go for 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 Estimating Preferences with Unobserved Choice Set Heterogeneity using Sufficient Sets

To build intuition we start by showing that unobserved choice set heterogeneity leads to violations of the Independence of Irrelevant Alternatives (IIA) property in models with Gumbel errors and therefore to inconsistent estimators. We introduce the concept of "sufficient set," which is any combination of a consumer's observed choices that is a subset of their true but unobserved choice set. We propose a "Sufficient Set logit" that condition on sufficient sets and show how this "differences out" consumers' true choice sets, permitting consistent parameter estimation. Finally, we discuss how sufficient sets can be used in the context of the Pairwise Maximum Score Estimator (Fox (2007)) and simplify the computational burden of methods based on Manski (1977) that "integrate out" consumers unobserved choice sets.

2.1 MNL Model and Bias from Unobserved Choice Set Heterogeneity

Let there be $i = 1, \dots, I$ *(types of) individuals* and for each of them $t = 1, \dots, T$ different *choice situations*. For example, this could be one person (i) observed to make several separate decisions over time ($t = 1, \dots, T$), or several people ($t = 1, \dots, T$) of the same type (i) each making a separate decision at a single point in time. Denote i's "sequence" of choices by $Y_i = (Y_{i1}, \dots, Y_{iT})$. For simplicity, we assume to observe exactly T choice situations for each i, but this is inessential.

We consider a situation in which i is matched to her choice set $CS_{it}^{\star}$ in choice situation t, but this choice set is unobserved to the researcher. We are interested in the estimation consequences of mistakenly imputing to i in t an incorrect choice set, S_{it}, such that $CS_{it}^{*} \neq S_{it}$.

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

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

Given θ and a specific match with a set of possible choice sequences, $\mathcal{CS}_i^{\star} = c$, each individual i is observed to make a sequence of choices $Y_i = j = (j_1, ..., j_T)$. We assume that the conditional indirect utility of alternative j_t in choice situation t for individual i is

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

where X_{ij_tt} is a vector of observable characteristics, and ϵ_{ij_tt} is the portion of i's utility that is unobserved to the econometrician. 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:

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

The first term in (2.2) is the probability of choosing j solely due to preferences given the sequence of choice sets i is matched to, and the second is the probability that the individual is matched to that sequence of choice sets.

We make the following important assumption, which implies that $\Pr\left[Y_i = j \middle| \mathcal{CS}_i^{\star} = c, \theta\right]$ is multinomial logit for any c.

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

Assumption 1 allows for general matching processes: $\Pr[\mathcal{CS}_i^\star = c|\gamma]$ can take any form and be a function of any element of the multinomial logit model $\Pr[Y_i = j|\mathcal{CS}_i^\star = c, \theta]$. 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, but rules out the matching of individuals to choice sets depending on the *realization* of the unobservables, ϵ_{ij_tt}'s.[1]

We use Assumption 1 to build up intuition about the econometric problem introduced by unobserved choice set heterogeneity (Proposition 1 below) and in the first of our three proposed solutions (subsection 2.2). In subsection 2.3, we relax Assumption 1 and discuss our ideas in the context of semi-parametric discrete-choice models that can be estimated by the Pairwise Maximum Score estimator proposed by Fox (2007). The ideas presented in subsection 2.4 are not specific to any particular distributional assumption.

An 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 is multinomial logit.[2] 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:

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

Since $\mathcal{CS}_i^\star = c$ is unobserved, suppose that the researcher instead specifies the likelihood function to be used in estimation based on the conditional probability of i choosing $Y_i = j$ from some other set of choice sequences $\mathcal{S}_i = s = \times_{t=1}^T s_t$. Researchers often specify $\mathcal{S}_i$ to be common across i and given by, for example, all those products above a certain market share threshold or a given number of products with the largest market shares. In such cases, the potentially misspecified model is:

$$\Pr\left[Y_i = j|\, \mathcal{S}_i = s, \theta\right] = \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ij_tt}, \theta\right)\right)}{\sum_{k \in S_{it} = s_t} \exp\left(V\left(X_{ik_tt}, \theta\right)\right)}, \tag{2.4}$$

[1] To our knowledge, the only paper in the applied literature that allows for this possibility is De los 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 ϵ_{ij_tt}'s. 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).

[2] Given Assumption 1, if γ shares some common element with θ (as is likely), failing to control for the choice set matching process $\Pr\left[\mathcal{CS}_i^\star = c|\, \gamma\right]$ only causes a loss of efficiency in the resulting conditional Maximum Likelihood estimator relative to a joint Maximum Likelihood estimator derived from Equation (2.2).

where the difference between equations (2.3) and (2.4) lies in the terms included in the summations in the denominator of each.

Proposition 1. Given the 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 $S_{it} = s_t$ of the sequence $\mathcal{S}_i = s$ includes at least one alternative not originally available in $CS^{\star}_{it} = c_t$.

Proof. Individual i's *true* choice probability conditional on $\mathcal{S}_i = s$, given $\mathcal{CS}^{\star}_i = c$, is:

$$\begin{aligned}
\Pr\left[Y_i = j \middle| \mathcal{S}_i = s, \mathcal{CS}^{\star}_i = c, \theta\right] &= \prod_{t=1}^{T} \frac{\Pr[Y_{it} = j_t | CS^{\star}_{it} = c_t, \theta]}{\sum_{r_t \in s_t \cap c_t} \Pr[Y_{it} = r_t | CS^{\star}_{it} = c_t, \theta] + \sum_{k_t \in s_t \setminus c_t} \Pr[Y_{it} = k_t | CS^{\star}_{it} = c_t, \theta]} \\
&= \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ij_t t}, \theta\right)\right)}{\sum_{r_t \in S_{it} = s_t \cap CS^{\star}_{it} = c_t} \exp\left(V\left(X_{ir_t t}, \theta\right)\right)} \\
&= \Pr\left[Y_i = j \middle| \mathcal{S}_i = s \cap \mathcal{CS}^{\star}_i = c, \theta\right],
\end{aligned} \tag{2.5}$$

where the denominator in the first line decomposes s_t into those alternatives that are in c_t ($r_t \in s_t \cap c_t$) and those that are not ($k_t \in s_t \setminus c_t$). The second equality obtains as the probability i selects an alternative not in her choice set is zero, $\Pr[Y_{it} = k_t | CS^{\star}_{it} = c_t, \theta] = 0$ for all $k_t \notin CS^{\star}_{it} = c_t$. In other words, since Assumption 1 implies the IIA property only when the choice set assumed by the researcher is a *subset* of i's true choice set, $\mathcal{S}_i = s \subseteq \mathcal{CS}^{\star}_i = c$, Equation (2.5) is not guaranteed to equal (2.4). By expressing (2.5) in terms of (2.4), we obtain:

$$\begin{aligned}
\Pr\left[Y_i = j \middle| \mathcal{S}_i = s \cap \mathcal{CS}^{\star}_i = c, \theta\right] &= \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ij_t t}, \theta\right)\right)}{\sum_{r_t \in s_t \cap c_t} \exp\left(V\left(X_{ir_t t}, \theta\right)\right)} \times \frac{\sum_{m_t \in s_t} \exp\left(V\left(X_{im_t t}, \theta\right)\right)}{\sum_{m_t \in s_t} \exp\left(V\left(X_{im_t t}, \theta\right)\right)} \\
&= \prod_{t=1}^{T} \frac{\exp\left(V\left(X_{ij_t t}, \theta\right) - \ln\left(\frac{\sum_{r_t \in s_t \cap c_t} \exp\left(V\left(X_{ir_t t}, \theta\right)\right)}{\sum_{m_t \in s_t} \exp\left(V\left(X_{im_t t}, \theta\right)\right)}\right)\right)}{\sum_{m_t \in S_{it} = s_t} \exp\left(V\left(X_{im_t t}, \theta\right)\right)} \\
&= \prod_{t=1}^{T} \frac{\exp(V\left(X_{ij_t t}, \theta\right) - \ln(\pi_{it}(\theta)))}{\sum_{m_t \in S_{it} = s_t} \exp\left(V\left(X_{im_t t}, \theta\right)\right)}.
\end{aligned} \tag{2.6}$$

IF part. Suppose $s_t \cap c_t \subset s_t$ for some t's (i.e., s_t includes choices not in c_t), then $\ln(\pi_{it}) < 0$ for those t's and models (2.4) and (2.5) 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. Suppose instead $s_t \subseteq c_t$ for all t's (i.e., s_t is a weak subset of the true c_t), then $\ln(\pi_{it}) = 0$ for all t's and (2.5) equals (2.4). Consequently, the model used in estimation will correspond to the true conditional choice model. ■

Discussion. The $\ln(\pi_{it}(\theta))$'s are (i, t)-specific fixed effects that cause inconsistency in estimation. $\pi_{it}(\theta)$ measures the probability that individual i would choose one of the alternatives in her true choice set when faced with the set of products assumed by the researcher. If $s_t \cap c_t \subset s_t$, i.e. if s_t includes alternatives not in c_t, this probability will be strictly less than one, and smaller (and thus the likely inconsistency greater) the more likely it is that i would have preferred one of the products mistakenly included in s_t.

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 multinomial logit models where IIA would normally hold. This violation, if not prevented, introduces individual-time-specific fixed effects that are functions of the rest of the model and that lead to the inconsistency of the estimator of θ. In Appendix A, we explore the severity of this econometric bias in a Monte Carlo experiment and demonstrate that it can be substantial.

The inconsistency induced by incorrectly including in individuals' choice sets alternatives among which they did not choose cannot be resolved using econometric methods commonly used by applied researchers. Alternative-specific constants or random coefficients will not control for the individual-time-specific fixed effects, because they are *individual- and time-* specific and their distribution is a function of all of the observables. What can work in principle is the approach of Manski (1977), which models the unconditional probability i chooses j by integrating out over *all* possible unobserved 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) and Van Nierop et al. (2010)).

There are two issues, however, that may constrain the use of the Manski approach in practice. First, researchers often make strong functional form assumptions about the choice set generating process $\Pr[\mathcal{CS}_i^\star = c|\gamma]$, which may be incorrect and lead to the inconsistency of the estimator of θ. Second, because the set of all possible choice sets grows exponentially in the number of alternatives, integrating out over all of them poses a large computational burden.[3]

Faced with the problem of inconsistency caused by unobserved choice set heterogeneity and the challenges posed by "integrating them out," we propose two solutions. First, we propose to "difference them out" instead, both in multinomial logit and semi-parametric models. These are discussed in sections 2.2 and 2.3 below. Second, we discuss how the concept of sufficient set proposed in this paper can alleviate the curse of dimensionality inherent in the Manski (1977)'s approach, facilitating its practical implementation. This is discussed in section 2.4.

2.2 A Sufficient Set logit that "Differences Out" Unobserved Choice Set Heterogeneity

We propose to directly prevent IIA violations by *differencing out* unobserved choice sets. This can be achieved by using i's observed choice sequence, $Y_i = j$, to construct *subsets* of i's true but unobserved choice set, $\mathcal{CS}_i^\star = c$, without requiring additional data. We call such subsets *"sufficient sets."*

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. It is for this reason that we call any f that satisfies Condition 1 a sufficient set.

[3]In subsection 2.4, we provide more details about this approach.

Proposition 2 (The Sufficient Set logit). Suppose that Assumption 1 and Condition 1 hold. Then, for every individual i and choice sequence $Y_i = j$ such that $f(j) = r$:

$$\Pr[Y_i = j | f(Y_i) = r, \theta] = \frac{\prod_{t=1}^{T} \exp(V(X_{ij_t t}, \theta))}{\sum_{k: f(k)=r} \prod_{t=1}^{T} \exp(V(X_{ik_t t}, \theta))} \tag{2.7}$$

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 B.

The proof of Proposition 2, presented in Appendix B, shows that if Condition 1 is satisfied, conditioning on the sufficient set, $f(Y_i) = r$, allows all terms that depend on i's unobserved sequence of choice sets, $CS_i^* = c$, to cancel from the likelihood function. We call this property the "differencing out" of unobserved choice sets and we call the resulting model the Sufficient Set logit.

Proposition 2 implies that, in the presence of unobserved choice set heterogeneity, the parameters θ can be consistently estimated by Maximum Likelihood on the basis of the Sufficient Set logit, $\Pr[Y_i = j | f(Y_i) = r, \theta]$. In essence, one can estimate preferences based only on the variation in characteristics of those products in i's sufficient set, not her full (but unobserved) sequence of choice sets. This is evident as equation (2.7) does not depend on i's unobserved sequence of possible choice sets, $\mathcal{CS}_i^* = c$. Whenever $\mathcal{CS}_i^\star = c$ is observed, 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 unobserved, the econometrician needs to be careful in constructing sufficient sets of choice sequences for each i, $f(Y_i)$. 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 estimators based on equation (2.7): the econometrician is again mistakenly enlarging i's set of potential choice sequences, violating the IIA property, and (unintentionally) introducing $\ln(\pi_{it}(\theta))$ terms which cause inconsistency in the estimation of θ.

Proposition 2 is quite general and works for any f that generates subsets of $\mathcal{CS}_i^\star$. In section 3 we discuss how different economic theories of consumer choice behavior naturally lead to sufficient sets f that satisfy Condition 1. While these examples of sufficient sets are suggestive, we note here that they represent *a* set of sufficient conditions that imply Proposition 2, but they are neither necessary nor the minimal sufficient conditions for the result to hold. In subsection 4.2 we present Hausman tests to help researchers choose among different sufficient sets.

The next result introduces a restriction on $f(Y_i)$ that greatly simplifies the practical implementation of the Maximum Likelihood estimator of the Sufficient Set logit $\Pr[Y_i = j|f(Y_i) = r, \theta]$.

Proposition 3. Suppose that Assumption 1 and Condition 1 hold. Then, the T random variables $Y_i = (Y_{i1}, \ldots, Y_{iT})\,|(f(Y_i) = r, \theta)$ are *conditionally independent* if and only if $f(Y_i) = \times_{t=1}^{T} f_t(Y_i)$.

Proof. See Appendix C.

Proposition 3 implies that the Sufficient Set logit (2.7) over *sequences of choices* can be equivalently expressed as the product of T separate t-specific multinomial logits over *alternatives* if and only if the sufficient set (over *sequences of choices*) $f(Y_i)$ can be expressed as the cartesian product of T separate t-specific sufficient sets (over *alternatives*). More specifically, if and only if $f(Y_i) = \times_{t=1}^{T} f_t(Y_i)$:

$$\Pr[Y_i = j|f(Y_i) = r, \theta] = \prod_{t=1}^{T} \Pr[Y_{it} = j_t|f(Y_i) = r, \theta] = \prod_{t=1}^{T} \Pr[Y_{it} = j_t|f_t(Y_i) = r_t, \theta]$$

where each $\Pr[Y_{it} = j_t|f_t(Y_i) = r_t, \theta]$ is a multinomial logit model over the set of alternatives in r_t. To see why this results in convenient estimators, suppose that the econometrician specifies a $f(Y_i)$ such that in each $t = 1, ..., 10$ an individual can choose one out of $J_t = J = 5$ different alternatives. It follows that $f(Y_i)$ contains 5^{10} possible sequences of choices of length $T = 10$. The resulting Sufficient Set logit (2.7) the econometrician would have to estimate would be a multinomial logit model with a summation over 5^{10} addends in the denominator. However, since $f(Y_i)$ can be obtained as the cartesian product of $T = 10$ separate t-specific sets each containing $J = 5$ alternatives, Proposition 3 guarantees that this Sufficient Set logit can equivalently be expressed as the product of $T = 10$ multinomial logit models each with a summation over $J = 5$ addends in the denominator. The examples of sufficient sets that we will propose in section 3 all satisfy Proposition 3, giving rise to computationally simple estimators. As we will illustrate later, a famous exception to Proposition 3 is the Choice Permutations sufficient set first proposed by Chamberlain (1980), which defines—in our terminology—the sufficient set as the collection of all permutations of the observed $Y_i = j$ and which results in Chamberlain's fixed effect logit model.

2.3 Beyond Gumbel Errors: Pairwise Maximum Score Estimator

In this subsection we relax Assumption 1 and discuss a more general solution to the estimation of preferences in the presence of unobserved choice set heterogeneity using sufficient sets in the context of semi-parametric models that can be estimated by the Pairwise Maximum Score Estimator (PMS) proposed by Fox (2007).

McFadden (1978) shows that multinomial logit models can be consistently estimated by a conditional Maximum Likelihood (ML) Estimator using subsets of consumers' true choice sets. We relied on McFadden (1978) in our proof of Proposition 2 presented in Appendix B. Specifically, the sufficient sets defined by Condition 1 are designed to guarantee the applicability of McFadden (1978) in multinomial logit models with *unobserved* choice sets. Building on Manski (1975), Fox (2007) extends McFadden (1978) beyond the i.i.d. Gumbel errors case to the case of semi-parametric models.[4] Fox (2007) shows that semi-parametric discrete-choice models can be consistently estimated by a Pairwise Maximum Score (PMS) estimator using subsets of consumers' true choice sets. Much like Section 2.2 showed how to apply sufficient sets to estimate preference parameters in a conditional logit model with unobserved choice sets, we propose here a class of sufficient sets that guarantees the applicability of Fox (2007) in semi-parametric discrete-choice models with unobserved choice sets.

Assumption 2. *Suppose that $V(X_{ikt}, \theta) = X_{ikt}\theta$ and that X_{it} is the matrix stacking all the X_{ikt}'s for $k \in CS^{\star}_{it} = c_{it}$. For any given (i,t) and $k, k' \in CS^{\star}_{it} = c_{it}$, $X_{ikt}\theta > X_{ik't}\theta$ if and only if $\Pr[Y_{it} = k | X_{it}, CS^{\star}_{it} = c_{it}, \theta] > \Pr[Y_{it} = k' | X_{it}, CS^{\star}_{it} = c_{it}, \theta]$.*

Assumption 2 states that the alternatives belonging to (i,t)'s true but unobserved choice set with higher systematic utilities are more likely to be chosen. Note that, differently from parametric models such as the multinomial logit or probit, Assumption 2 does not impose that the distribution of ϵ_{ikt} is the same across individuals or even that the distribution is the same across the choice situations of the same individual (e.g., ϵ_{ikt} could be distributed Laplace while $\epsilon_{ikt'}$ normal). However, Assumption 2 does constrain the joint distribution of ϵ_{ikt} across alternatives for any given (i,t). Goeree et al.

[4] Another notable extension of McFadden (1978) is Bierlaire et al. (2008). They propose conditional ML estimators that allow for the consistent estimation of discrete-choice models with Generalized Extreme Value errors from subsets of the true choice sets. To the best of our knowledge, results of this kind are still not available for the mixed logit or random coefficient logit model.

(2005) show that a sufficient condition for Assumption 2 is that, for any (i,t), the joint density of the errors across alternatives is *exchangeable*.[5] See Fox (2007) for more details about Assumption 2.

The implementation of the PMS estimator requires an additional condition on consumers' sufficient sets.

Condition 2. Suppose that $f(Y_i) = \times_{t=1}^{T} f_t(Y_i) = \times_{t=1}^{T} f_{it}$ and that there is a non-empty set N of (i,t)'s with $|N| = n \leq I \cdot T$ for which $K = \cap_{(i,t)\in N} f_{it}$ contains at least two alternatives, $|K| \geq 2$.

Condition 2 imposes two restrictions. First, it requires the sufficient set over sequences $f(Y_i)$ to be the cartesian product of t-specific sufficient sets f_{it}'s over alternatives. Second, it requires that there is a set of (i,t)'s whose sufficient sets f_{it}'s contain the same two or more alternatives. The PMS Estimator makes pairwise comparisons of alternatives belonging to some subset K for all those (i,t)'s that are known to have originally made choices from some choice set $CS_{it}^{\star}$ such that $K \subseteq CS_{it}^{\star}$. Condition 2 uses sufficient sets to construct a K guaranteed to be strictly included in the true but unobserved choice set $CS_{it}^{\star}$ of every (i,t) belonging to N.[6]

With a slight abuse of notation, we re-label the alternatives in subset K so that $K = \{1, ..., k, ..., K\}$. The PMS Estimator using choice-based data on the subset K of alternatives is the parameter vector $\hat{\theta}_n^K$ that maximizes:

$$Q_n^K(\theta) = \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \frac{1}{n} \sum_{(i,t)\in N} (1[Y_{it} = k] \cdot 1[X_{ikt}\theta > X_{ik't}\theta] + 1[Y_{it} = k'] \cdot 1[X_{ik't}\theta > X_{ikt}\theta]). \quad (2.8)$$

Theorem 1 [Fox (2007), p.1011]. Under Assumption 2, Condition 1, Condition 2, and technical Assumptions 3 and 4 (see Fox (2007), p.1009 and p.1011 for statements of these), the Pairwise Maximum Score estimator $\hat{\theta}_n^K$ is consistent for θ.

Theorem 1 states that semi-parametric discrete-choice models characterized by Assumption 2 and

[5] Despite the flexibility, there are popular models among applied researchers that violate Assumption 2, such as the mixed logit or random coefficient logit model.

[6] As we will see below, the sizes of N and K will directly affect the number of pairwise comparisons, i.e. observations, used to construct the PMS estimator.

unobserved choice set heterogeneity can be consistently estimated by the PMS estimator whenever the econometrician can construct sufficient sets that satisfy Conditions 1 and 2.

Discussion. In this and the previous subsection, we have introduced two different estimators that rely on sufficient sets to estimate preferences in the presence of unobserved choice sets. Each has its advantages and disadvantages. The primary advantage of Fox (2007)'s Pairwise Maximum Score estimator is that it allows for flexible distributions of unobserved preferences (as long as they satisfy Assumption 2). By contrast, the Maximum Likelihood estimator of Sufficient Set logit from Proposition 2 usually requires unobserved preferences to be distributed i.i.d. Gumbel.[7]

Despite (or because of) its greater flexibility and robustness in the estimation of the preference parameters θ, however, Fox (2007)'s Pairwise Maximum Score Estimator does not allow for the calculation of some objects typically of interest to applied researchers. Researchers using either method can evaluate the willingness to pay for alternatives' attributes (see, e.g., Bajari et al. (2008) or section 5.3), but knowledge of $\hat{\theta}_n^K$ does not allow for the evaluation of predicted choice probabilities, marginal effects, price elasticities, and consumer surplus. As we will discuss in subsection 4.3 below, the Sufficient Set logit permits for the construction of lower and upper bounds on such objects of interest.

In addition, the practical implementation of the PMS estimator is more involved than the estimation of the Sufficient Set logit. The objective function $Q_n^K(\theta)$ of the PMS estimator is a step-function always involving multiple global maxima, so that its (proper) maximization requires the use of global search routines which are not always readily accessible to applied researchers. Moreover, because the PMS estimator is cube-root consistent (rather than square-root consistent), the computation of valid standard errors requires specialized subsampling (see Delgado et al. (2001)) or bootstrapping (see Cattaneo et al. (2017)) procedures.[8] By contrast, the Sufficient Set logit can be estimated with standard Stata commands for the estimation of multinomial logit models, especially whenever Proposition 3 holds and $f(Y_i) = \times_{t=1}^{T} f_t(Y_i)$.[9]

[7] As we will see in the next section, Proposition 2 applies almost directly to the case of unobserved preference heterogeneity of the form $V_i(X_{ij_tt}, \theta) = \delta_{ij_t} + X_{ij_tt}\beta$, the so-called fixed effect logit model. In addition, under some mild additional conditions, Proposition 2 can be extended to the case of the nested logit model (results not presented but available on request).

[8] To alleviate these implementation problems, Jeremy Fox and David Santiago share on their personal webpages Mathematica codes for the implementation of the PMS estimator.

[9] In those cases in which $f(Y_i)$ cannot be expressed as $\times_{t=1}^{T} f_t(Y_i)$, the econometrician can directly apply McFadden (1978) and estimate θ from subsets of the *observed* but potentially huge sufficient sets $f(Y_i)$. As an example,

2.4 The Use of Sufficient Sets when "Integrating Out" Unobserved Choice Sets

So far we have discussed the use of sufficient sets to specify *conditional* discrete-choice models that "difference out" unobserved choice sets. Here we illustrate how sufficient sets can also be used to simplify the estimation of *unconditional* discrete-choice models that "integrate out" unobserved choice sets.

As originally discussed by Manski (1977), the *unconditional* probability of i selecting choice sequence j can be written as:

$$\Pr[Y_i = j|\theta, \gamma] = \sum_{c \in C} \Pr[Y_i = j|\mathcal{CS}_i^\star = c, \theta] \times \Pr[\mathcal{CS}_i^\star = c|\gamma], \tag{2.9}$$

where C is the collection of sets of possible choice sequences to which individual i can be matched.

Until recently, it was believed that the identification of the Manski model relied on the availability of auxiliary data about $\Pr[\mathcal{CS}_i^\star = c|\gamma]$ (e.g., Roberts and Lattin (1991)) and/or the availability of instruments that exclusively affected consideration, $\Pr[\mathcal{CS}_i^\star = c|\gamma]$ (e.g., Goeree (2008)). In a recent paper, however, Abaluck and Adams (2017) present identification results for this model that do not require the availability of such auxiliary data.[10]

Even when model (2.9) is correctly specified and identified, however, its estimation can prove difficult. In particular, estimation of equation (2.9) is likely to suffer from a curse of dimensionality because the number of elements in C grows exponentially in the number of choice sequences J available to consumers (i.e. $|C| = 2^J - 1$). A direct consequence of this curse of dimensionality is that, unless the researcher makes strong functional form assumptions on $\Pr[\mathcal{CS}_i^\star = c|\gamma]$, the model can be estimated only when J is very small (e.g., Abaluck and Adams (2017) limit their estimations to the case of J=10). In what follows, we show how to rely on sufficient sets to provide additional restrictions on model (2.9) that can significantly ease the curse of dimensionality and make its estimation more tractable for any given J (or even make it possible if J is large). If Condition 1 is satisfied, $f(Y_i) = r \subseteq \mathcal{CS}_i^\star = c$, where c is the *true* sequence of choice sets to which i is matched. It therefore follows that any set of choice

D'Haultfœuille and Iaria (2016) implement this idea in the context of Chamberlain (1980)'s fixed effect logit model. Matlab codes are available on the authors' personal webpages.

[10] The intuition of their argument is that whenever some alternatives are not in the choice sets of some individuals, the discrete-choice analogy of Slutsky symmetry will be violated. They show that one can therefore use deviations from Slutsky symmetry to separately identify consideration from preferences given consideration.

sequences $c' \in C$ such that $f(Y_i) = r \nsubseteq \mathcal{CS}_i^\star = c'$ cannot be the set of choice sequences to which i is matched, so that $\Pr[\mathcal{CS}_i^\star = c'|\gamma]$ must be zero. In other words, the researcher knows that i's true but unobserved choice set must contain *all* the choice sequences in the sufficient set, and consequently any candidate choice set that does not include even just one of these choice sequences can be removed from the collection of *possible* sets of choice sequences C. Let $C_{f(Y_i)} = \{c|f(Y_i) = r \subseteq \mathcal{CS}_i^\star = c\}$ be the collection of choice sets consistent with $f(Y_i) = r$. Then model (2.9) simplifies to:

$$\Pr[Y_i = j|\theta, \gamma] = \sum_{c \in C_{f(Y_i)}} \Pr[Y_i = j|\mathcal{CS}_i^\star = c, \theta] \times \Pr[\mathcal{CS}_i^\star = c|\gamma]. \tag{2.10}$$

where the only difference with Equation (2.9) is in the terms included in the summation. Note that $C_{f(Y_i)}$ will typically be substantially smaller than the unrestricted C. For example, suppose that there are four choice sequences a, b, c, and d. Any individual i will have a sufficient set of one of four possible sizes: $|f(Y_i)| = 1$ (e.g., $f(Y_i) = \{a\}$, $f(Y_i) = \{b\}$, etc.), $|f(Y_i)| = 2$ (e.g., $f(Y_i) = \{a, b\}$, $f(Y_i) = \{b, c\}$, etc.), $|f(Y_i)| = 3$ (e.g., $f(Y_i) = \{a, b, c\}$, $f(Y_i) = \{b, c, d\}$, etc.), or $|f(Y_i)| = 4$ (i.e., $f(Y_i) = \{a, b, c, d\}$). The collection C, the power set of $\{a, b, c, d\}$, will then contain $2^4 - 1 = 15$ possible (non-empty) choice sets. However, $C_{f(Y_i)}$ will only contain: 8 choice sets if $|f(Y_i)| = 1$, 4 choice sets if $|f(Y_i)| = 2$, 2 choice sets if $|f(Y_i)| = 3$, and 1 choice set if $f(Y_i) = \{a, b, c, d\}$. Importantly, note that this use of the sufficient sets is quite general and does not rely on the specific functional form assumptions made by the researcher in specifying model (2.9).

3 Sufficient Sets and Their Economic Foundations

A primary contribution of this paper is to show how to consistently estimate preferences in the presence of unobserved choice sets by using sufficient sets of consumers' choices. In this section, we describe how choice environments that have been analyzed in a wide variety of economics literatures map into particular sufficient sets. We consider both panel data settings, where individuals are observed making several decisions over time, as well as cross-sectional settings, where several different individuals face the same choice environment at a point in time. The sufficient sets introduced here are neither mutually exclusive nor exhaustive: they can be combined or devised to reflect a large range of choice environments.

3.1 Sufficient Sets for Economic Environments Characterized by Stable Choice Sets

Fixed-Sample Search. In an influential paper in the search literature, Morgan and Manning (1985) present general results on the existence and properties of expected-utility maximizing search rules for dynamic search problems in which individuals may choose both the number of periods in which samples of alternatives are searched and the size of the sample searched in each period. Individuals conduct searches over T periods. In a non-sequential or fixed-sample search strategy, individuals do all their search in the first period, construct a choice set, and then make sequences of T choices from this choice set. The authors show that if individuals have "full recall," i.e. once alternatives are searched, the individual does not forget they exist until period T, and "no lost alternatives," i.e. alternatives are not removed from the market until period T, then a fixed-sample search strategy is optimal if either the marginal cost of searching or the individual's discount factor is sufficiently high. Intuitively, fixed-sample search strategies are appealing when individuals find it more advantageous to gather information quickly, because the search results are observed with some delay and the cost of waiting is high.

Fixed-sample search strategies have been studied in both product and labor markets. For example, Janssen and Moraga-González (2004) study oligopolistic markets characterized by consumers who engage in costly fixed-sample searches for the best prices, while De los Santos et al. (2012) find evidence in support of fixed-sample search strategies in the online market for books from data on the web browsing and purchasing behavior of a large panel of consumers.

Whenever a fixed-sample search strategy is optimal, individuals' unobserved choice sets will be stable across the T choice situations for each i, but potentially different across i's: i.e., $CS^{\star}_{it} = CS^{\star}_{i}$, $\forall\ t$. In this context, any alternative purchased by i in any t, Y_{it}, is guaranteed to belong to $CS^{\star}_{i}$.

Referral networks. Another context in which the assumption of stable choice sets plausibly applies is for specialists' referral networks. For example, Gaynor et al. (2016) study the impact of a regulatory policy that expanded the set of hospitals to which a physician could refer a patient in the English National Health Service (NHS). They consider the post-reform period and assume that in this period, a physician's choice of referral hospitals is unconstrained, showing that this allows them to identify preference parameters. They then use these and apply them to the pre-reform period to learn about the impact of the pre-reform constraint on doctors' referral choices.

The authors assume that physicians' referral networks are stable over time. As such, and similar to the example of fixed-sample search above, $CS^{\star}_{it} = CS^{\star}_{i}$, $\forall\ t$ and any hospital chosen by i in any t, Y_{it}, is guaranteed to belong to $CS^{\star}_{i}$.

School Choice. In many countries, the set of public schools that parents can consider to send their children for T years is constrained by the neighbourhood where they live and by various allocation mechanisms such as the Gale-Shapley deferred acceptance mechanism (see Abdulkadiroglu and Sönmez (2003)). Usually, the set of schools in any neighbourhood does not evolve rapidly, and those households that do not change neighbourhood over the T years will face a stable set of schools for their children. In this context, Walters (2014) investigates the demand for charter middle schools in Boston, while Fack et al. (2015) study the demand for high schools in the southern district of Paris. Similar to the two previous examples above, the set of schools faced by household i in academic year t can be assumed to be stable $\forall\ t$, $CS^{\star}_{it} = CS^{\star}_{i}$, and any school chosen by i for their children in any t, Y_{it}, is guaranteed to belong to $CS^{\star}_{i}$.

3.1.1 Sufficient Sets Consistent with Stable Choice Sets

Full Purchase History Sufficient Set. Consistent with the three examples above, suppose that individuals' choice sets are potentially heterogeneous across i's but stable over the T choice situations, $CS^{\star}_{it} = CS^{\star}_{i}$. Let $H_i = \bigcup_{t=1}^{T} \{Y_{it}\} \subseteq CS^{\star}_{i}$ be the collection of all the alternatives that individual i is observed to choose in any of the T choice situations. We define the Full Purchase History (FPH) sufficient set as $f_{FPH}(Y_i) = (H_i)^T$, the set of choice sequences given by the cartesian product of H_i in each of the T choice situations. Note that $f_{FPH}(Y_i)$ satisfies the requirements of Proposition 3 and thus gives rise to Sufficient Set logit models that are very simple to estimate.

Practically, the FPH sufficient set assumes that an individual that chooses a collection of alternatives H_i over a given sequence of T choice situations is familiar with all of those alternatives in all of those choice situations. The intuition of our approach, evident in the denominator of Equation (2.7) for the case of Sufficient Set logit models, is that the researcher exploits variation in the characteristics of these alternatives over choice situations, and the differences in choices made by each i over choice situations, to estimate preference parameters θ in the absence of information about i's true choice set, $CS_i^\star$.

Further pursuing this point, note that each individual may have considered other alternatives beyond those included in her FPH sufficient set, $f_{FPH}(Y_i)$, that were ultimately not chosen. All that is required by our estimators is that a sufficient set is a *subset* of an individual's true choice set (as stated by Condition 1 for conditional logit models and by Conditions 1 and 2 for semi-parametric models). Note also that, to simplify exposition, we have assumed that choice sets are stable across all T choice situations of each individual, but this is not necessary. The estimators proposed in this paper work whenever we observe at least two (different) purchase decisions by the same individual from the same choice set. More generally, i's full choice 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. In Section 4.2 below, we describe how to form Hausman tests to check the length of the sequence over which choice sets are plausibly stable.

Choice Permutations Sufficient Set. In addition to the FPH sufficient set, the assumption of stable choice sets also underpins another sufficient set: namely, the sufficient set proposed by Chamberlain (1980) for the classic Fixed Effect logit model (FE logit). In a model with systematic utilities given by $V_i(X_{ij_tt}, \theta) = \delta_{ij_t} + X_{ij_tt}\beta$, Chamberlain (1980) shows that β can be consistently estimated by the ML estimator of a Sufficient Set logit model with sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$: the set of all possible permutations of observed choice sequence Y_i.[11] As such, we call this the Choice Permutations sufficient set.

Matejka and McKay (2015) propose a choice model with rational inattention in which the reduced form choice probabilities take the form of a FE logit model (see Theorem 1, p.282). In this model, individuals have priors about the indirect utilities associated with consuming any alternative in $CS_i^\star$ and, prior to choosing an alternative, can decide to gather more precise payoff-relevant information at a cost. In the reduced form, the individual-alternative specific fixed effects δ_{ij_t}'s are not functions of i's preferences, but rather of the cost of gathering information and of i's priors.[12] As a special case, the

[11] For example, if i is observed to choose alternatives 1, 3, and 5 in choice situations 1, 2, and 3, the $f_{CP}(Y_i)$ sufficient set will be the collection of all possible permutations of sequence (1, 3, 5), i.e. (1, 3, 5), (1, 5, 3), (3, 1, 5), (3, 5, 1), etc.

[12] Recall j indexes the full sequence of i's choices and j_t indexes the choice made in the t^{th} choice occasion.

reduced form choice probabilities simplify to a multinomial logit model when i's priors are completely uninformative (i.e., each alternative in $CS_i^\star$ is perceived to generate the same level of indirect utility).

Chamberlain (1980)'s expressed motivation for the sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$ was to difference out the fixed effects (δ_{ij_t}) from each individual's systematic utility. But his assumption of choice set stability also implies that $f_{CP}(Y_i) \subseteq \mathcal{CS}_i^\star$. As such, by the same arguments developed in Proposition 2, sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$ will not only accommodate unobserved preference heterogeneity in the form of individual-alternative specific fixed effects, but also unobserved choice set heterogeneity.[13]

While this is a significant benefit, the FE logit also comes with meaningful costs. To obtain predicted choice probabilities and their functions, such as elasticities, researchers typically need to be able to identify the whole vector of preference parameters θ. The FE logit does not, however, usually allow the identification of i's fixed effects δ_{ij_t}'s, but only those elements of β associated with time-varying observables. This can limit its usefulness to applied researchers. By contrast, the Sufficient Set logit model obtained from $f_{FPH}(Y_i)$, while relying on the same assumption of choice set stability, typically allows the identification of the entire parameter vector θ.

Note that the Choice Permutations sufficient set, $f_{CP}(Y_i)$, *cannot* be expressed as the cartesian product of t-specific sufficient sets, and consequently does not satisfy the requirements of Proposition 3. For those cases where T is large and/or there is substantial heterogeneity in the alternatives chosen across the T choice situations, the computational burden implied by the FE logit can be considerable. D'Haultfœuille and Iaria (2016) show how to ease this computational burden by applying the insights of McFadden (1978) to the estimation of β from (uniformly) random subsets of $f_{CP}(Y_i)$.

3.2 Sufficient Sets for Economic Environments Characterized by Growing Choice Sets

Sequential Search. Morgan and Manning (1985) show that, in the general context described in subsection 3.1, if the assumptions ensuring "full recall' and "no lost alternatives" hold, then any sequential search strategy over T periods will imply choice sets that are weakly growing over time, so that $CS_{it}^\star \subseteq CS_{it+1}^\star$.

[13]The argument is essentially identical, except for the different systematic utilities that in Chamberlain (1980) have individual-alternative specific coefficients. By replacing $V(X_{ij_tt}, \theta)$ with $\delta_{ij_t} + X_{ij_tt}\beta$ and $f(Y_i)$ with $\mathcal{P}(Y_i)$ in the proof of Proposition 2, Chamberlain (1980)'s result follows.

Caplin and Dean (2011) propose two models of sequential search: the alternative-based search (ABS) model and the reservation-based search (RBS) model. The ABS model "captures the process of sequential search with [full] recall, in which the [consumer] evaluates an ever-expanding set of objects, choosing at all times the best object thus far identified" (Caplin and Dean (2011), p.23). This model provides the micro foundations underlying some of the functional form restrictions used by Goeree (2008), Manzini and Mariotti (2014), and Abaluck and Adams (2017) to aid the identification of Manski (1977)'s model. By contrast, the RBS model is a formalization of Simon (1955)'s satisficing model. In related work, Caplin et al. (2011) find experimental evidence in support of this model.

In these settings, a sequential search strategy implies that products that we have observed the individual purchasing in the (recent) past are in their choice set and can be used to form a sufficient set. We describe such a sufficient set after introducing other examples.

Choice by Iterative Search. Masatlioglu and Nakajima (2013) propose another dynamic search framework that they call Choice by Iterative Search (CIS). In each period t, the history of search until $t-1$, also called the "status quo" (i.e., the set of alternatives searched so far), and the feasible set of alternatives both can affect the evolution of the choice set $CS^{\star}_{it}$ in arbitrary ways. However, if assumptions analogous to "full recall" and "no lost alternatives" hold, then $CS^{\star}_{it}$ will coincide with next period's status quo and the sequence of status quos will weakly grow over time, so that $CS^{\star}_{it} \subseteq CS^{\star}_{it+1}$.

Several models in the fast-growing literature on limited attention build on the CIS framework. An example is Eliaz and Spiegler (2011), who study a setting in which consumers have a singleton status quo, i.e. a choice set only including one product (possibly different across consumers), and firms seek to use marketing devices, e.g. advertising, to include their products in consumers' choice sets.[14] Consumer preferences are themselves unaffected by such advertising; it only influences the products consumers' include in their choice sets. While Eliaz and Spiegler (2011) only consider a static environment, one could imagine a dynamic extension within the CIS framework, in which multiple firms compete in each period with advertising to encourage consumers to consider their products and status quos evolve over time as in Masatlioglu and Nakajima (2013). In such a setting, again the alternatives

[14] Other relevant examples of the application of the CIS framework are Ho et al. (2015) and Heiss et al. (2016). Both papers study the Medicare Part D program and document that consumers switch health plans infrequently and search imperfectly, possibly because of high search costs.

a consumer has purchased in the (recent) past will be in their choice set and can be used to form a sufficient set.[15]

Focus. Another example related to the CIS framework is the work of Kőszegi and Szeidl (2013), who analyze the impact of "focus" on consumer choice. They provide numerous examples of consumers focusing on one of a choice's (possibly many) attributes, leading them to select an alternative that exceeds others in this attribute, even if a comparison of the alternatives across all attributes would lead to a different choice.[16]

Formally they model this as consumers making a choice from a subset of an individual's full choice set. They motivate this subset using conjuntive and disjunctive screening rules like those from Gilbride and Allenby (2004) (Kőszegi and Szeidl, 2013, p.61). They then apply their model of consumer focus to intertemporal choice (e.g. consumption-savings decisions).

As they describe themselves, "Formally, there are T periods and in period t, a consumer makes a choice x_t from the deterministic ... consideration set $X_t(h_{t-1})$, where $h_{t-1} = (x_1, \ldots, x_{t-1})$ is the history of choices up to period $t-1$." While in a consumption-savings environment, past decisions limit current choices by means of a summary statistic (e.g. how much income a consumer has in period t), in a repeat-purchase environment (e.g. retail purchases of household goods), h_{t-1} would consist of the history of that consumer's previous purchase decisions, a fact that can be used to form a sufficient set as we describe next.

3.2.1 Sufficient Sets that Allow for Growing Choice Sets

The Past Purchase History Sufficient Set. The three examples above suggest the following sufficient set. Let $H_{it} = \bigcup_{b=1}^{t} \{Y_{ib}\} \subseteq CS_{it}^{\star}$ be the collection of all the alternatives that individual i is observed to choose between choice situation 1 and t. We define the Past Purchase History (PPH) sufficient set as $f_{PPH}(Y_i) = \times_{t=1}^{T} H_{it}$, the cartesian product of H_{it} between choice situation 1 and T.[17] Note

[15] This is the framework we study in our empirical application in section 5.

[16] For example, a person comparing the quality of life in California with that in the Midwest may focus more on climate than other aspects of life satisfaction in which the two areas are more similar (e.g. crime rate, availability of public goods, etc.), and therefore be too likely to believe that California is a better place to live.

[17] An analytically related sufficient set to f_{PPH} is the sufficient set compatible with choice sets that are weakly *shrinking*, rather than growing, over choice situations. This can be obtained by just "turning around" the choice situations of each choice sequence, so to have them re-ordered from T to 1, and then by applying the same definition of f_{PPH} to the re-ordered choice sequences.

that, similarly to the Full Purchase History sufficient set, $f_{PPH}(Y_i)$ also satisfies the requirements of Proposition 3 and thus gives rise to Sufficient Set logit models that are very simple to estimate.

Practically, the PPH sufficient set assumes that an individual has in their choice set the set of alternatives chosen between some beginning period and the current one. Importantly, alternatives are only assumed to be in i's choice set after they are observed to have been chosen. As with the FPH sufficient set, the intuition is to exploit the variation in the characteristics of only these alternatives over time. As for the FPH sufficient set, any individual i's full choice sequence $(Y_{i1}, ..., Y_{iT})$ can be divided into sub-sequences of length of at least two. The assumption here allows for the possibility of choice sets that are growing within each sub-sequence, but with potentially different choice sets between the different sub-sequences for the same individual. In Section 4.2 below, we describe how to form Hausman tests to check the length of the sequence over which choice sets are allowed to weakly grow.

3.3 Cross-Sectional Data: Inter-Personal Sufficient Set

Each of the examples above assumed a panel of individuals making decisions in multiple choice situations. Here we illustrate how sufficient sets can also be constructed in cross-sectional environments to a group of individuals, each making a separate purchase decision at a single point in time from the same choice set.

As a motivating example, consider the question of whether greater availability of fast food outlets causes obesity as in Currie et al. (2010). One would like to be able to identify whether it is the availability of fast food outlets that leads to increased consumption or whether preferences are the driving factor. The authors collected precise geographic data on the location of fast food outlets and where children live and attend school and examined the effect of the presence of a fast food restaurant within 0.1, 0.25, and 0.5 miles of the school attended by the student. While this is reasonable, this definition did not exploit the location of each child's home and, even if it did, one cannot be sure of exactly which outlets lie within individual children's choice sets. By contrast, if the authors were willing to assume that all children living on the same street *and* attending the same school faced the same choice set, they could conclude that all such outlets were in the choice set for all such children and this could form the basis for a sufficient set in our approach.

More generally, to apply our method in a cross-sectional environment, we call each i a "consumer type" and each t one of the T individuals of that type.[18] In a cross-sectional environment, our method relies on the 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 we assume that each individual could have purchased the product bought by the other at the same price, i.e. $p_{ijt} = p_{ijt'} = p_{ij}$ and $p_{ikt} = p_{ikt'} = p_{ik}$. If instead 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 needs to be able to compute $d_{ijt'}$ for any other individual t' of type i.

The assumption of an identical choice set across the T individuals of the same consumer type i gives rise to the Inter-Personal sufficient set: $f_{IP}(Y_i) = (H_i)^T$, where $H_i = \bigcup_{t=1}^{T} \{Y_{it}\} \subseteq CS^{\star}_{it}$.[19] The sufficient set $f_{IP}(Y_i)$ imputes to each individual t the collection of all the alternatives observed to be chosen by any of the T individuals of consumer type i. As for $f_{FPH}(Y_i)$ and $f_{PPH}(Y_i)$, note that $f_{IP}(Y_i)$ also satisfies the requirements of Proposition 3 and thus gives rise to Sufficient Set logit models that are very simple to estimate.

3.4 Other sufficient sets

Combining sufficient sets. The sufficient sets listed above are neither mutually exclusive nor exhaustive. Regarding the first point, if one had panel data and a stable choice environment (suggesting the application of the FPH sufficient set) and some subset of the individuals faced the same choice set in each period (suggesting the application of the IP sufficient set), one could combine the sufficient sets into a Inter-Personal Full Purchase History sufficient set. This is also more generally true: if multiple sufficient set definitions apply, then one can use the intersection of each sufficient set to form a new, composite, sufficient set. In the empirical application in section 5, we propose one such composite sufficient set, a store-type-specific Past Purchase History (i.e., each i has a separate PPH sufficient

[18] This is without loss of generality. We could allow each type to have a different number of individuals, T_i, but this would only complicate notation and provide no deeper insights into the underlying mechanism at work.

[19] Note that this definition of $f_{IP}(Y_i)$ is identical to that for the Full Purchase History sufficient set, but because the underlying economic environments are so different (e.g. i is an individual and t is a time period in $f_{FPH}(Y_i)$, while i is a consumer type and t is an individual in $f_{FPH}(Y_i)$), we prefer to define the two separately.

set that is specific to each of the store-types in our data).

Other sufficient sets. It is critically important to note that all of the examples provided in this section merely *illustrate* sufficient sets, focusing on those sufficient sets that are most likely to be broadly applicable by applied researchers. In general, *any set* that somehow combines i's choices across t's can in principle serve as a sufficient set.

Typically, there may be some choice sets that vary in predictable ways due to idiosyncracies of someone's choice environment. For example, an individual might choose from a different set of alternatives 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 set by conditioning on past purchases made in that state (e.g. lunch choices made on past Mondays instead of lunch choices made more generally).

More generally, information available to the researcher can and should be incorporated into the definition of sufficient sets. So, generalizing Eliaz and Spiegler (2011) and considering how to measure the impact of an in-store advertising campaign for a new product, if one had information about both each consumer's past purchase history as well as their exposure to and awareness of the advertising campaign for the new product *and* one was willing to make the assumption that exposure and awareness to the campaign meant the new product was in a consumer's choice set, one could use each consumer's past purchase history to form an (initial) sufficient set for that consumer and then augment it with the new product for those consumers known to have been exposed to the advertisement.

4 Sufficient Sets: Performance, Testing, and Identification

4.1 Monte Carlo Evidence On Conditional Logit Models

Table 4.1 reports the results of Monte Carlo simulations evaluating the performance of four conditional logit models in the presence of unobserved choice set heterogeneity. The first column in the table reports results showing the bias in a multinomial logit model from incorrectly assuming that all individuals have access to the full (universal) choice set. The second column estimates the true multiniomial logit model, i.e. the model that correctly assigns the choice set facing each individual in each choice situation. There is of course never estimation bias in this case. The results of Full

Purchase History (FPH) and Choice Permutation (CP) logits are reported in the third and fourth columns. The economic environment being analyzed is described in the notes to the table.

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. Overall, 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 when relying on either of these two sufficient sets for estimation. Appendix A presents further evidence of the magnitudes of biases that can arise when researchers assume choice sets greater than that available to consumers in their data.

Table 4.1: *Increasing the share of individuals with restricted choice set*

	Universal logit		True logit		FPH logit		CP 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 individuals 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 $V(X_{ij_tt}, \theta) = \delta_{j_t} + X_{ij_tt}\beta$*, and the unobserved portion of utility,* ϵ_{ij_tt}*, is distributed i.i.d. Gumbel. In the baseline specification,* X_{ij_tt} *is drawn from a normal distribution with mean 0 and variance 5,* $\delta_{j_t} = 0$ *for all* j_t*'s, and* $\beta = 2$*. We simulate 20 replications. To speed up computations, the CP 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).)*

4.2 Specification Tests for Sufficient Sets

There are many possible sufficient sets compatible with Conditions 1 and 2. In the context of the Maximum Likelihood estimator, 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 how to test some of 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 or grow; and *(b)* the presence of individual-alternative-specific fixed effects in the systematic utilities.

4.2.1 Theoretical Foundations

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

Using this result, we propose Hausman specification tests between conditional logit models based on different sufficient sets, i.e. different ways of constructing a correspondence f that satisfies Condition 1. These comparisons enable us to test for some forms of choice set stability and unobserved preference heterogeneity.

Proposition 4. Suppose that Assumption 1 holds, and that sufficient sets f_L and f_Z satisfy Condition 1, that $f_Z(Y_i) \subset f_L(Y_i)$, $Y_i \in \mathcal{CS}_i^\star = c$, and that $i = 1, \ldots, I$. Define $l_L(\theta)$ and $l_Z(\theta)$ as the log-likelihood functions corresponding to the Sufficient Set logit models conditional on $f_L(Y_i)$ and on $f_Z(Y_i)$, with $\widehat{\theta}_L$ and $\widehat{\theta}_Z$ the corresponding 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 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 D.1.

Proposition 4 hinges on the *Factorization Theorem* proposed by Paul Ruud. For more details, see Ruud (1984) and Hausman and Ruud (1987). The Likelihood Ratio statistic, LR, proposed in result (3b) allows us to compare different Sufficient Set logit 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, $l_L\left(\widehat{\theta}_L\right)$.[20] Even though LR requires the computation of a third estimator, $\widehat{\theta}_\triangle$, it is simpler to implement than other Hausman 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 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 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.

4.2.2 Application to Sufficient Sets

The examples of Choice Permutation (CP), Full Purchase History (FPH), and Past Purchase History (PPH) sufficient sets discussed earlier rely on the following economic assumptions:

- f_{CP}: 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_t}.

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

- f_{FPH}: Choice set stability across T choice situations and no unobserved preference heterogeneity.
- f_{PPH}: Choice set evolution in the form of growing choice sets ~~entry-but-no-exit or exit-but-no-entry~~ across T choice situations and no unobserved preference heterogeneity.

Proposition 4 leads to statistics that can be used to test these economic assumptions by comparing the estimates of Sufficient Set logit models based on different sufficient sets (different f's). The first possibility is to compare f_{CP}, f_{FPH}, and f_{PPH} for choice sequences of constant length T. In fact, for choice sequences of a given length T, both the CP and PPH sufficient sets are subsets of the FPH sufficient set, $f_{CP}(Y_i) \subset f_{FPH}(Y_i)$ and $f_{PPH}(Y_i) \subset f_{FPH}(Y_i)$ for any $Y_i \in \mathcal{CS}_i^\star = c$. Given Proposition 4, these relationships among sufficient sets allow us to construct Hausman 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_{CP}, 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_{CP}'s, f_{CP}^Z (separately from 1 to T^Z and from $T^Z + 1$ to T^L) and f_{CP}^L (from 1 to T^L) such that $f_{CP}^Z(Y_i) \subset f_{CP}^L(Y_i)$ 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 for sub-sequences of length at least two, this method of making comparisons allows us to test for very general forms of choice set stability or evolution. In Appendix D.2 we discuss each of these test statistics in more detail.

4.3 Point-Identification of Preferences and Bounds on Choice Probabilities, Elasticities, and Consumer Surplus

Our focus to this point has been the consistent estimation of preference parameters, θ. The three models presented above differ in the extent to which they allow us to construct estimates of functions of θ, for example choice probabilities and elasticities. At one extreme is the Pairwise Maximum Score estimator discussed in Section 2.3. The Pairwise Maximum Score estimator can point-identify the preference parameters θ, as well as simple functions of them (e.g. willingness-to-pay), but cannot tell us about about functions of θ that involve knowledge of the distribution of the unobserved portion of utility.[21] At the other extreme are models based on Manski (1977) that integrate out unobserved choice

[21]See Fox (2007) and Bajari et al. (2008) for more details on this point.

sets discussed in Section 2.4. These approaches involve specifying a model of choice set formation that reveals the distribution of choice sets in the population of consumers, so they allow us to point-identify all functions of θ that depend on this distribution.

The Sufficient Set logit is an intermediate case. In this section, we describe parameters and 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 E.

For expositional simplicity, suppose that the systematic utilities take the form:

$$V(X_{ij_tt}, \theta) = \delta_{j_t} + X_{ij_tt}\beta + \alpha p_{j_tt},$$

where $\theta = [\delta_1, ..., \delta_J, \beta, \alpha]$ and p_{j_tt} is the price of alternative j_t.

We can point-identify the vector of preference parameters θ from the Sufficient Set logit model $\Pr[Y_i = j | f(Y_i) = r_i, \theta]$.[22] We can similarly point-identify simple functions of θ. For example, we are often interested in willingness-to-pay (WTP) for product characteristic k, $X^k_{ij_tt}$. By Roy's Identity, this can be computed as:

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

Other 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 when one relies on strong assumptions about the choice set formation process. The probability with which i chooses alternative j_t given choice set $CS^\star_{it} = c_{it}$ is

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

if $j_t \in CS^\star_{it} = c_{it}$ and zero otherwise. This choice probability depends on i's (unobserved) choice set, $CS^\star_{it}$. Suppose that we observe a superset Q_{it} of the true but unobserved choice set, so that $CS^\star_{it} \subseteq Q_{it}$. This could be, for example, the collection of all alternatives observed to be chosen by any

[22] Note that here, differently from most other parts in the paper, 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 E.

i in choice situation t. It follows that, even if we do not directly observe $CS_{it}^{\star} = c_{it}$, $f_t(Y_i) \subseteq CS_{it}^{\star}$ and $CS_{it}^{\star} \subseteq Q_{it}$. We can therefore use these conditions to bound the true but unobserved denominator of the Sufficient Set logit choice probabilities for any $X_{it} = [X_{i1t}, p_{1t}, ..., X_{iJt}, p_{Jt}]$ and θ:

$$\sum_{m \in f_t(Y_i) = r_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}) \le \sum_{m \in CS_{it}^{\star} = c_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}) \le \sum_{m \in Q_{it} = q_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}). \tag{4.3}$$

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

$$Pr_{ij_t t}^{Q}(\theta) \le Pr_{ij_t t}^{CS^{\star}}(\theta) \le Pr_{ij_t t}^{f}(\theta). \tag{4.4}$$

That is to say, the true choice probability with which i chooses j_t in t is bounded from below by the same probability assuming i chooses from some superset of the unobserved choice set, $Q_{it} = q_{it}$, and from above by the same probability assuming i chooses from just their sufficient set, $f_t(Y_i) = r_{it}$. In Appendix E, we expand on this idea and discuss how to bound average predicted choice probabilities and individual 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:

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

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

$$W_{it}(X_{it}|\theta, f_t(Y_i) = r_{it}) \le W_{it}(X_{it}|\theta, CS_{it}^{\star} = c_{it}) \le W_{it}(X_{it}|\theta, Q_t = q_{it}). \tag{4.6}$$

5 Empirical Illustration

To illustrate how our ideas can be applied in practice, we present an empirical illustration. In Section 3.2, we discussed 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.

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 that 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)).

At any one point in time, there are up to 250 products in the on-the-go chocolate market from which a consumer is potentially able to choose in the two years that we analyze (2010-11). In such a choice environment it is unlikely that a consumer will spend the time to consider each one, and collecting information on which products the consumer considered (for example, using eye-tracking technologies) is expensive. We compare results from estimation based on two different assumptions on choice sets. First, we assume that each consumer considers all of the products that are available in the type of store in which they are currently shopping; we call this the "store-type specific Universal choice set." Second, we assume that each consumer considers (at minimum) the products that are available in the type of store in which they are currently shopping *and* that the consumer has purchased in the past (in any store type); we call this the "store-type specific Past Purchase History (PPH) sufficient set."

5.1 Model

We adapt the general model presented in Section 2.2 to the specifics of on-the-go chocolate demand. We assume that each consumer makes a purchase from their own choice set $CS^{\star}_{it}$. This set is not observed. It could include many or only a few of the products currently available in the market; it always includes the option not to purchase. We observe what product was purchased, the price paid, the type of store the product was purchased in, and what products are available in that type of store. See Appendix F for further details.

In a first stage, a consumer i is matched to their choice set $CS^{\star}_{it} = c_{it}$. In a second stage, consumer i chooses the utility-maximizing product from her choice set $CS^{\star}_{it} = c_{it}$. As described above, we assume a store-type specific Past Purhcase History (PPH) sufficient set, i.e. we assume that the consumer definitely considers the chocolate bars that are available in the type of store they are observed to be shopping in at t and that they have purchased in the past (in any store-type). Note that this sufficient set satisfies Proposition 3 and is therefore computationally easy to estimate.

The probability consumer i buys the sequence of products $j = (j_1, ..., j_t, ..., j_T)$ given her store-type specific PPH sufficient set is given by:

$$\Pr[Y_i = j | f(Y_i) = r_i, \theta] = \prod_{t=1}^{T} \frac{\exp(V(X_{ij_t t}, \theta))}{\sum_{k \in r_{it}} \exp(V(X_{ikt}, \theta))}. \tag{5.1}$$

where $f(Y_i) = \times_{t=1}^{T} f_t(Y_i) = \times_{t=1}^{T} r_{it}$ and $f_t(Y_i) = r_{it}$ is the set of chocolate bars belonging to individual i's store-type specific PPH sufficient set in week t. Utility for any chocolate bar j_t in week t is given by

$$U_{ij_t t} = V(X_{ij_t t}, \theta) + \epsilon_{ij_t t}, \tag{5.2}$$

with

$$V(X_{ij_t t}, \theta) = \delta_{j_t} + \alpha p_{oj_t t} + \beta \ln a_{(i)bt}, \tag{5.3}$$

where δ_{j_t} is a product-j_t-specific fixed effect, $p_{oj_t t}$ is the price of product j_t in store-type o in week t, and $\ln a_{(i)bt}$ is log advertising exposure to brand b (to which product j_t belongs) in week t. The price variable and two different measures of advertising exposure are defined in the next subsection.

The utility of the outside option of not purchasing a chocolate bar is given by

$$U_{i0t} = \delta_0 + \sum_m \tau_m + \epsilon_{i0t}, \tag{5.4}$$

where the τ_m's are month effects meant to capture seasonality and/or cyclicality in on-the-go chocolate demand.

For comparison, we also estimate the model assuming that the individual i considers all the chocolate bars S_{ot} that are available in the type of store o in which i is observed to be shopping in at t, whether they have purchased them in the past or not. We call this the (store-type specific) Universal choice set. The "Universal logit" is given by:

$$\Pr[Y_i = j | \mathcal{S}_o = s_o, \theta] = \prod_{t=1}^{T} \frac{\exp(V(X_{ij_t t}, \theta))}{\sum_{m \in S_{ot} = s_{ot}} \exp(V(X_{imt}, \theta))}. \tag{5.5}$$

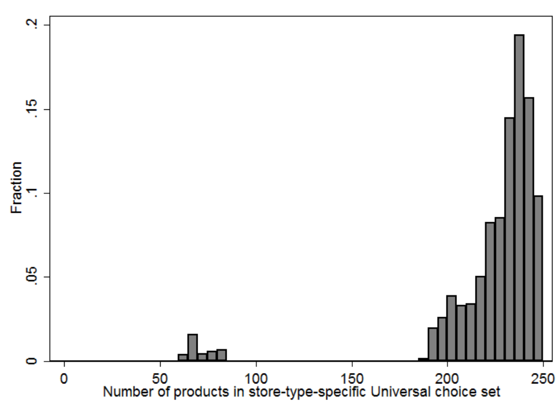
5.2 Data

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.[23] 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. The outside option, when a chocolate bar is not purchased, has a 74% market share. Prices are constructed at the level of the store-type o and week t.

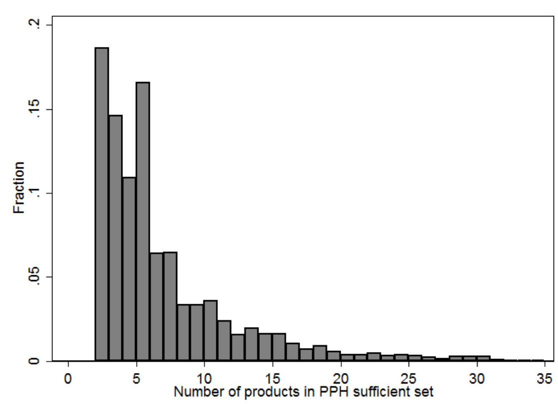
Figure 5.1 shows the distribution of the number of chocolate bars used in estimation. On the left-hand side, we show the distribution of the number of chocolate bars in the store-type specific Universal choice set across the 9,387 purchase occasions. The maximum is 250 chocolate bars; the small cluster between 50 and 100 is for purchase occasions where the individual chooses from a vending machine. On the right hand side, we show the distribution of the number of chocolate bars in each consumer's store-type specific PPH sufficient sets; these range from 2 to 35.

Figure 5.1: *The Number of Products Used in Estimation: store-type-specific Universal choice sets and Past Purchase History sufficient sets*

(a) Store-type specific Universal choice set



(b) Store-type specific Past Purchase History (PPH)



Note: The histograms shows the distribution of the number of products used in estimation across 9,387 purchase occasions in our two specifications. The left-hand panel shows variation in the number of products in the store-type specific Universal choice set. The right-hand panel shows variation in the number of products in each consumer's store-type specific Past Purchase History sufficient set.

[23] These data were 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.

We use two measures of advertising exposure. Both convert weekly advertising ("flows") into an advertising "stock;" advertising stocks are the depreciated accumulation 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 then depreciate and accumulate. We denote this brand-level stock variable, $stock_{bt}$. 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. As earlier, we depreciate and accumulate these values to define an individual-level stock variable, which we denote $stock_{ibt}$. $stock_{ibt}$ ranges from 0 for individuals that do not watch TV, or only watch advertising-free public TV (the BBC), to a maximum of 65 minutes of accumulated exposure to advertisements for a particular brand. The mean is 29 minutes of accumulated 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, $stock_{(i)bt}$, using the log inverse hyperbolic sine function, $\ln a_{(i)bt} = ln\left(stock_{(i)bt} + \sqrt{1 + stock^2_{(i)bt}}\right)$. Further details are available in Appendix F.

5.3 Estimation Results

Table 5.1 presents the estimated preference parameters for our two specifications. Columns (1) and (3) are estimated using the Universal choice set, and columns (2) and (4) are estimated using the store-type specific Past Purchase History (PPH) sufficient sets. Columns (1) and (2) use brand-level advertising stocks, while columns (3) and (4) use individual-level advertising stocks. Regardless of the advertising stock variable, the estimates based on the store-type specific PPH yield more elastic price responses and less elastic advertising responses.

The patterns that we find in this 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 store-type specific 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 store-type specific 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 store-type specific 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 store-type specific PPH logit.

Table 5.1: *Coefficient estimates*

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

Notes: Each column represents a separate regression. Columns labelled "Universal logit" indicate that estimate was based on the products in the store-type-specific Universal choice set. Columns labelled "PPH logit" indicate that estimate was based on variation in products in each consumer's store-type-specific Past Purchase History sufficient set. All columns include month dummies interacted with the outside option to control for seasonal and cyclical effects. Advertising is measured in seconds and then divided by 1000. We include 58 product-specific fixed effects, one for each product that has a market share greater than 0.1%, or is purchased at least 10 times).

We can use the estimated preference parameters to estimate the willingness-to-pay for advertising, something in which firms and advertising executives are likely to be interested, and the willingness-to-pay for single products, something in which firms and retailers are likely to be interested.

In our simple illustrative example, willingness-to-pay for (log) advertising is given by:

$$\widehat{WTP}_a = -\frac{\partial V_{ij_tt}/\partial \ln a_{(i)bt}}{\partial V_{ij_tt}/\partial p_{oj_tt}} = -\frac{\widehat{\beta}}{\widehat{\alpha}}. \tag{5.6}$$

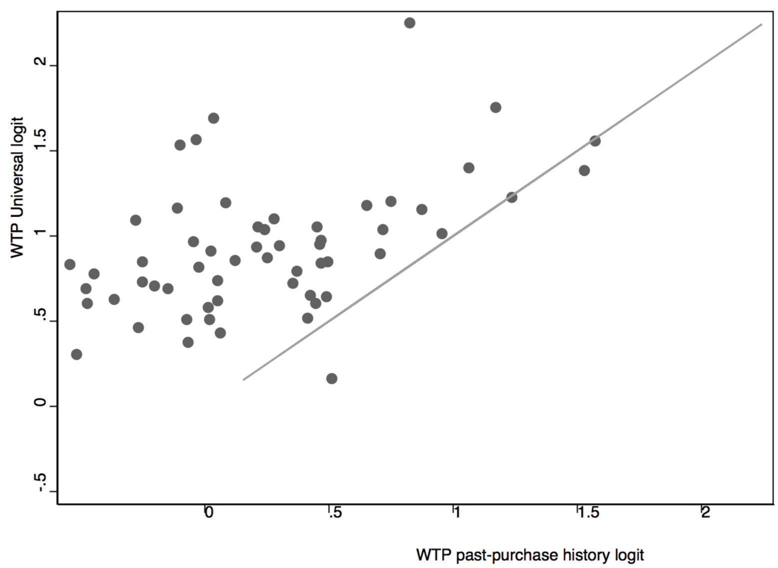
The parameter estimates in Table 5.1 indicate that there is significant bias in willingness-to-pay when estimating from the Universal choice set. The estimates obtained using the Universal logit in column (3) suggest that a one-standard deviation increase in the log advertising stock, $\ln a_{ibt}$, equal to 0.54 (or 54%), implies an increase in valuation of a product by 4.6 pence.[24] As the average price of a chocolate product is 74 pence, this is a 6.2% increase. By contrast, the estimates obtained using the

[24] $(-0.54 \times 0.159) \div -1.853 = 0.0463$.

store-type-specific 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 chocolate bars, given for each chocolate bar j_t by $\widehat{\delta}_{j_t}/\widehat{\alpha}$. Figure 5.2 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 store-type specific 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 chocolate bars with the Universal logit than with the store-type specific PPH logit. In most cases, the downward bias in the estimated price coefficient seen in Table 5.1 is more than the downward bias in the estimated product dummies.

Figure 5.2: *Willingness to pay for products*



Note: Each dot is a chocolate bar, 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 store-type specific PPH logit estimates in column (4) of Table 5.1. For each set of estimates, willingness-to-pay is calculated as $\hat{\delta}_{j_t}/\hat{\alpha}$.

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 preferences, and that most products are more desired than consumer preferences truly indicate.

6 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 econometric biases and we suggest two solutions based on the concept of "sufficient sets." First, when the Manski (1977)'s approach of "integrating out choice sets" is not feasible, we propose to use sufficient sets to "difference out choice sets," both in multinomial logit and semi-parametric models. Second, we discuss how sufficient sets can alleviate the curse of dimensionality inherent in the Manski approach, facilitating its practical implementation. We illustrate the method using an application estimating 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 demand. 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. Our approach can serve as a useful complement to this approach by both facilitating its practical implementation and 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 preference estimates, what can be learned about the distribution 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) address this topic.

References

Abaluck, Jason and Abi Adams, "What Do Consumers Consider Before They Choose: Identification from Asymmetric Demand Responses," *Working Paper*, 2017.

Abdulkadiroglu, Atila and Tayfun Sönmez, "School choice: A mechanism design approach," *The American Economic Review*, 2003, *93* (3), 729–747.

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

Bajari, Patrick, Jeremy T Fox, and Stephen P Ryan, "Evaluating wireless carrier consolidation using semiparametric demand estimation," *Quantitative Marketing and Economics*, 2008, *6* (4), 299.

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.

Caplin, Andrew and Mark Dean, "Search, choice, and revealed preference," *Theoretical Economics*, 2011, *6* (1), 19–48.

_ , _ , and Daniel Martin, "Search and satisficing," *The American Economic Review*, 2011, *101* (7), 2899–2922.

Cattaneo, Matias D, Michael Jansson, and Kenichi Nagasawa, "Bootstrap-Based Inference for Cube Root Consistent Estimators," *Working Paper*, 2017.

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.

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.

Delgado, Miguel A, Juan M Rodrı?guez-Poo, and Michael Wolf, "Subsampling inference in cube root asymptotics with an application to Manskis maximum score estimator," *Economics Letters*, 2001, *73* (2), 241–250.

D'Haultfœuille, Xavier, Alessandro Iaria, and Guillaume Lecué, "Recovering the Distribution of Consideration Sets from Choice Data," 2016. Work in Progress.

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

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.

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

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond truth-telling: Preference estimation with centralized school choice," *Working Paper*, 2015.

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, J. K., C. A. Holt, and T. R. Palfrey, "Regular quantal response equilibrium," *Experimental Economics*, 2005, *8* (4), 347–367.

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.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou, "Inattention and switching costs as sources of inertia in medicare part d," 2016. Working paper.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton, "The impact of consumer inattention on insurer pricing in the Medicare Part D program," 2015. Working paper.

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

Janssen, Maarten CW and José Luis Moraga-González, "Strategic pricing, consumer search and the number of firms," *The Review of Economic Studies*, 2004, *71* (4), 1089–1118.

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.

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, "Maximum score estimation of the stochastic utility model of choice," *Journal of econometrics*, 1975, *3* (3), 205–228.

_ , "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 and Daisuke Nakajima, "Choice by iterative search," *Theoretical Economics*, 2013, *8* (3), 701–728.

_ , _ , 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.

Morgan, Peter and Richard Manning, "Optimal search," *Econometrica*, 1985, pp. 923–944.

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.

Simon, Herbert A, “A behavioral model of rational choice,” *The quarterly journal of economics*, 1955, *69* (1), 99–118.

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

Appendices

A Quantifying the Size of the Bias

Table A.1 provides Monte Carlo evidence in which we quantify the size of the bias from mistakenly attributing to individuals alternatives that were not available to them. The three panels describe the relative importance of different 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 imputes 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 individuals have the full choice set available to them. In the second panel, we show that the bias increases with the share of individuals 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 individual prefers the alternative 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 individuals that do not make choices from the full choice set from 10% to 50% increases the average bias from assuming all individuals have access to the full choice set from 11.2% to 36.0%, increasing the share of unavailable alternatives from 1 of 5 to 3 of 5 increases the average bias from 26.3% to 57.0%, and increasing the average utility gap between the unavailable first-best and chosen second-best alternative from approximately 10% to 70% increases the average bias from assuming the first-best option was available from 19.0% to 38.6%.

Table A.1: *The size of bias in the 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 number of alternatives 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 the differentiation of alternatives		
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 individuals 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 $V(X_{ij_tt}, \theta) = \delta_{j_t} + X_{ij_tt}\beta$, and the unobserved portion of utility, ϵ_{ij_tt}, is distributed i.i.d. Gumbel. In the baseline specification, X_{ij_tt} is drawn from a normal distribution with mean 0 and variance 5, $\delta_{j_t} = 0$ for all j_t's, and $\beta = 2$. We simulate 20 replications. In the final panel, 30% of individuals have their first-best choice removed.

B Proof of Proposition 2

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

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, $\sum_{v_t \in CS_{it}^\star = c_t} \exp(V(X_{iv_t t}, \theta))$ 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). ■

C Proof of Proposition 3

IF part. Suppose that $f(Y_i) = \times_{t=1}^{T} f_t(Y_i)$. Then we can re-write the denominator of conditional logit model (2.7), $\Pr[Y_i = j | f(Y_i) = r, \theta]$, as (omitting the r for simplicity):

$$\begin{aligned}
\sum_{(k_1,\ldots,k_T)\in f} \prod_{t=1}^{T} \exp(V_{ik_t t}) &= \sum_{(k_1,\ldots,k_T)\in f_1\times\cdots\times f_T} \prod_{t=1}^{T} \exp(V_{ik_t t}) \\
&= \left(\sum_{k_1\in f_1} \exp(V_{ik_1 1})\right) \sum_{(k_2,\ldots,k_T)\in f_2\times\cdots\times f_T} \prod_{t=2}^{T} \exp(V_{ik_t t}) \\
&= \left(\sum_{k_1\in f_1} \exp(V_{ik_1 1})\right)\left(\sum_{k_2\in f_2} \exp(V_{ik_2 2})\right) \sum_{(k_3,\ldots,k_T)\in f_3\times\cdots\times f_T} \prod_{t=3}^{T} \exp(V_{ik_t t}) \\
&\vdots \\
&= \prod_{t=1}^{T}\left(\sum_{k_t\in f_t} \exp(V_{ik_t t})\right),
\end{aligned} \tag{C.1}$$

which implies that:

$$
\begin{aligned}
\Pr\left[Y_i = j \middle| f\left(Y_i\right) = r, \theta\right] &= \frac{\prod_{t=1}^{T} \exp\left(V_{ij_t t}\right)}{\sum\limits_{(k_1,\ldots,k_T)\in f} \prod_{t=1}^{T} \exp\left(V_{ik_t t}\right)} \\
&= \frac{\prod_{t=1}^{T} \exp\left(V_{ij_t t}\right)}{\prod\limits_{t=1}^{T}\left(\sum\limits_{k_t\in f_t} \exp\left(V_{ik_t t}\right)\right)} \\
&= \prod_{t=1}^{T} \frac{\exp\left(V_{ij_t t}\right)}{\sum\limits_{k_t\in f_t} \exp\left(V_{ik_t t}\right)} \\
&= \prod_{t=1}^{T} \Pr\left[Y_{it} = j_t \middle| f_t\left(Y_i\right) = r_t, \theta\right].
\end{aligned}
\tag{C.2}
$$

To complete the proof, we are now going to show that $\Pr[Y_{it} = j_t|f(Y_i) = r, \theta] = \Pr[Y_{it} = j_t|f_t(Y_i) = r_t, \theta]$.

Define the set $M\left(\widetilde{j}_s\right) = \left\{(z_1, \ldots, z_s, \ldots, z_T) \middle| z \in f\left(Y_i\right) = r, z_s = \widetilde{j}_s\right\}$ as the collection of choice sequences that have alternative $\widetilde{j}_s$ in position s. It then follows that:

$$
\begin{aligned}
&\Pr\left[Y_{is} = \widetilde{j}_s \middle| f\left(Y_i\right) = r, \theta\right] \\
&= \sum_{j\in M\left(\widetilde{j}_s\right)} \Pr\left[Y_i = j \middle| f\left(Y_i\right) = r, \theta\right] \\
&= \sum_{j\in M\left(\widetilde{j}_s\right)} \frac{\prod_{t=1}^{T} \exp\left(V_{ij_t t}\right)}{\sum\limits_{(k_1,\ldots,k_T)\in f=r} \prod_{t=1}^{T} \exp\left(V_{ik_t t}\right)} \\
&= \left(\sum_{(k_1,\ldots,k_T)\in f=r} \prod_{t=1}^{T} \exp\left(V_{ik_t t}\right)\right)^{-1} \left(\sum_{j\in M\left(\widetilde{j}_s\right)} \prod_{t=1}^{T} \exp\left(V_{ij_t t}\right)\right).
\end{aligned}
\tag{C.3}
$$

Similarly to (C.1), $f\left(Y_i\right) = \times_{t=1}^{T} f_t\left(Y_i\right)$ implies that $M\left(\widetilde{j}_s\right) = f_1\left(Y_i\right) \times \cdots \times \left\{\widetilde{j}_s\right\} \times \cdots \times f_T\left(Y_i\right)$, and consequently that the numerator of (C.3) can be re-written as:

$$\sum_{j\in M(\widetilde{j}_s)} \prod_{t=1}^{T} \exp\left(V_{ij_tt}\right) = \sum_{(j_1,\ldots,j_s,\ldots,j_T)\in f_1\times\cdots\times\left\{\widetilde{j}_s\right\}\times\cdots\times f_T} \prod_{t=1}^{T} \exp\left(V_{ij_tt}\right)$$
$$= \exp\left(V_{i\widetilde{j}_ss}\right)\prod_{t\neq s}\left(\sum_{j_t\in f_t}\exp\left(V_{ij_tt}\right)\right). \tag{C.4}$$

Plugging (C.1) and (C.4) into (C.3), we obtain:

$$\Pr\left[Y_{is}=\widetilde{j}_s\middle| f\left(Y_i\right)=r,\theta\right]$$
$$=\sum_{j\in M(\widetilde{j}_s)} \Pr\left[Y_i=j\middle| f\left(Y_i\right)=r,\theta\right]$$
$$=\left(\prod_{t=1}^{T}\left(\sum_{k_t\in f_t}\exp\left(V_{ik_tt}\right)\right)\right)^{-1}\left(\exp\left(V_{i\widetilde{j}_ss}\right)\prod_{t\neq s}\left(\sum_{j_t\in f_t}\exp\left(V_{ij_tt}\right)\right)\right)$$
$$=\left(\left(\sum_{k_s\in f_s}\exp\left(V_{ik_ss}\right)\right)\prod_{t\neq s}\left(\sum_{k_t\in f_t}\exp\left(V_{ik_tt}\right)\right)\right)^{-1}\left(\exp\left(V_{i\widetilde{j}_ss}\right)\prod_{t\neq s}\left(\sum_{j_t\in f_t}\exp\left(V_{ij_tt}\right)\right)\right) \tag{C.5}$$
$$=\frac{\exp\left(V_{i\widetilde{j}_ss}\right)}{\sum_{k_s\in f_s}\exp\left(V_{ik_ss}\right)}$$
$$=\Pr\left[Y_{is}=\widetilde{j}_s\middle| f_s\left(Y_i\right)=r_s,\theta\right].$$

ONLY IF part. Consider two choice situations t and s. For these, define $f_t=\{j_t|\,(j_1,\ldots,j_t,\ldots,j_T)\in f\left(Y_i\right)=r\}$ and $f_s=\{j_s|\,(j_1,\ldots,j_s,\ldots,j_T)\in f\left(Y_i\right)=r\}$ as the collections of alternatives that appear in at least one sequence belonging to $f\left(Y_i\right)=r$ at positions t and s, respectively. Suppose $f\left(Y_i\right)\neq\times_{t=1}^{T}f_t\left(Y_i\right)$, then $\exists$ t and s such that $\left(\widetilde{j}_t\in f_t,\widetilde{j}_s\in f_s\right)$ and $\left(\widetilde{j}_1,\ldots,\widetilde{j}_t,\ldots,\widetilde{j}_s,\ldots,\widetilde{j}_T\right)\notin f\left(Y_i\right)=r$. It then follows that:

$$\Pr\left[Y_{it}=\widetilde{j}_t\middle| Y_{is}=\widetilde{j}_s,f\left(Y_i\right)=r,\theta\right]=0,$$

while, since $\widetilde{j}_t \in f_t$:

$$\Pr\left[Y_{it} = \widetilde{j}_t \middle| f(Y_i) = r, \theta\right] > 0.$$

This implies that Y_{it} and Y_{is} are not conditionally independent. ■

D Specification Testing Appendix

This Appendix provides a proof of Proposition 4 and an example of how the specification tests proposed in subsection 4.2 can work in practice.

D.1 Proof of Proposition 4

D.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 $f_Z(Y_i) = z \subset f_L(Y_i) = l$ as:

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

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

D.1.2 Proof of Result (2)

Given result (1) of Proposition 4, results (2a) and (2b) follow from the *Factorization Theorem* of (Ruud, 1984, result (1), p.24).

D.1.3 Proof of Result (3)

Given result (1) of Proposition 4, 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 $\text{Var}\left(\widehat{\theta}_L\right)\text{Var}\left(\widehat{\theta}_Z\right)^{-1}\widehat{\theta}_Z + \text{Var}\left(\widehat{\theta}_L\right)\text{Var}\left(\widehat{\theta}_\triangle\right)^{-1}\widehat{\theta}_\triangle$. This implies that $\text{Cov}\left(\widehat{\theta}_L, \widehat{\theta}_Z\right) = \text{Cov}\left(\text{Var}\left(\widehat{\theta}_L\right)\text{Var}\left(\widehat{\theta}_Z\right)^{-1}\widehat{\theta}_Z, \widehat{\theta}_Z\right) = \text{Var}\left(\widehat{\theta}_L\right)\text{Var}\left(\widehat{\theta}_Z\right)^{-1}\text{Var}\left(\widehat{\theta}_Z\right) = \text{Var}\left(\widehat{\theta}_L\right)$, where the first equality follows from result (2a). Consequently, $\text{Var}\left(\widehat{\theta}_L - \widehat{\theta}_Z\right) = \text{Var}\left(\widehat{\theta}_L\right) + \text{Var}\left(\widehat{\theta}_Z\right) - 2\text{Cov}\left(\widehat{\theta}_L, \widehat{\theta}_Z\right) = \text{Var}\left(\widehat{\theta}_Z\right) - \text{Var}\left(\widehat{\theta}_L\right)$. ■

D.2 Testing Procedures

The various sufficient sets introduced earlier rely on the following economic assumptions:

- f_{CP}: 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_t}.
- f_{FPH}: Choice set stability across T choice situations *and* no unobserved preference heterogeneity.
- f_{PPH}: Choice set evolution in the form of weakly growing choice sets (or, symmetrically, weakly shrinking choice sets) across T choice situations *and* no unobserved preference heterogeneity.

There are two possibilities for making comparisons across conditional logit models based on different sufficient sets f's. The first possibility is to compare f_{CP}, f_{FPH}, and f_{PPH} for choice sequences of constant length T. The second possibility is to fix a specific f, say f_{CP}, 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_{CP}'s, f^Z_{CP} and f^L_{CP} such that $f^Z_{CP}(Y_i) \subset f^L_{CP}(Y_i)$ for any $Y_i \in \mathcal{CS}^\star_i = c$. We will discuss each testing possibility in turn.

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

For choice sequences of a given length T, $f_{CP}(Y_i) \subseteq f_{FPH}(Y_i)$ and $f_{PPH}(Y_i) \subseteq f_{FPH}(Y_i)$ for any $Y_i \in \mathcal{CS}_i^\star = c$. Suppose $Y_i = (1,3)$. Then $f_{CP}(1,3) = \mathcal{P}(1,3) = \{(1,3),(3,1)\}$, $f_{FPH}(1,3) = \{1,3\} \times \{1,3\} = \{(1,1),(3,3),(1,3),(3,1)\}$, and $f_{PPH}(1,3) = \{1\} \times \{1,3\} = \{(1,1),(1,3)\}$. Note that there is no clear "inclusion" relationship between $f_{CP}(Y_i)$ and $f_{PPH}(Y_i)$.

Given the results from Proposition 4, the above relationships among sufficient sets lead to two possible classes of tests. The first is about choice set stability and the second about the presence of unobserved preference heterogeneity.

D.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 set 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 4 tells us that $\widehat{\theta}_{FPH}$ would be more efficient than $\widehat{\theta}_{PPH}$. On the other hand, if unobserved choice sets were growing over choice situations, 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 $LR = 2\left[l_{PPH}\left(\widehat{\theta}_{PPH}\right) + l_{\triangle}\left(\widehat{\theta}_{\triangle}\right) - l_{FPH}\left(\widehat{\theta}_{FPH}\right)\right]$.

D.2.3 Preference Homogeneity

The sufficient sets f_{FPH} and f_{CP} 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_{CP} allows for individual-alternative 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 $\widehat{\theta}_{CP}$, but Proposition 4 tells us that $\widehat{\theta}_{FPH}$ would be more efficient than $\widehat{\theta}_{CP}$. On the other hand, if unobserved preference heterogeneity were present in a form encompassed by individual-alternative specific fixed effects, then only $\widehat{\theta}_{CP}$ 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 $LR = 2\left[l_{CP}\left(\widehat{\theta}_{CP}\right) + l_{\triangle}\left(\widehat{\theta}_{\triangle}\right) - l_{FPH}\left(\widehat{\theta}_{FPH}\right)\right]$.

D.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 $f^Z_{CP}(Y_i) \subset f^L_{CP}(Y_i)$ 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.

D.2.5 Choice Set Stability: f_{CP} Example

In this section we will show with an example why $f^Z_{CP}(Y_i) \subset f^L_{CP}(Y_i)$ 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 individual i is observed to make the choice sequence $Y_i = (j_1, j_2, j_3, j_4) = (3, 5, 5, 4)$.[25] By considering the observed choice sequence "at once," $Y_i = (3, 5, 5, 4)$ can be re–ordered in 12 different choice sequences.[26] Collect these sequences into the set $f^L_{CP}(Y_i) = l$. Assume that $V_i(X_{ij_tt}, \theta) = \delta_{ij_t} + X_{ij_tt}\beta$. Then, i's likelihood contribution given $f^L_{CP}(Y_i) = l$ is:

$$\begin{aligned}
&\Pr\left[Y_i = (3,5,5,4) \middle| f^L_{CP}(Y_i) = l, \beta\right] \\
&= \frac{\exp\left(\left(X_{i31} + X_{i52} + X_{i53} + X_{i44}\right)\beta\right)}{\sum\limits_{(j_1,j_2,j_3,j_4)\in f^L_{CP}(j_1,j_2,j_3,j_4)=l} \exp\left(\left(X_{ij_11} + X_{ij_22} + X_{ij_33} + X_{ij_44}\right)\beta\right)}.
\end{aligned} \tag{D.1}$$

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 $f^Z_{CP}(Y_{i1}) = \{(3, 5), (5, 3)\}$ and $f^Z_{CP}(Y_{i3}) = \{(5, 4), (4, 5)\}$. Then, i's likelihood contribution given $f^Z_{CP}(Y_{i1}) = z_1$ and $f^Z_{CP}(Y_{i3}) = z_3$ is:

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

[26] 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).

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

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

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

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

$$\Pr_i\left[Y_i=(3,5,5,4)\middle|\, f_{CP}^{Z}(Y_i)=z,\beta\right]=$$

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

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

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 $l_{CP}^{Z}(\beta)$, while $l_{CP}^{L}(\beta)$ would still use the whole sequence $Y_i=(3,4,5,5)$. More precisely, if $Y_i=(3,4,5,5)$ were observed, then $f_{CP}^{L}(3,4,5,5)=f_{CP}^{L}(3,5,5,4)=l$ would still contain the same 12 choice sequences, while model (D.2) would collapse to:

$$
\Pr\left[Y_i = (3,4,5,5) \middle| f^Z_{CP}(Y_{i1}) = h_1, f^Z_{CP}(Y_{i3}) = h_3, \beta\right]
$$

$$
= \frac{\exp\left(\left(X_{i31} + X_{i42}\right)\beta\right)}{\exp\left(\left(X_{i31} + X_{i42}\right)\beta\right) + \exp\left(\left(X_{i41} + X_{i32}\right)\beta\right)} \times
$$

$$
\times \frac{\exp\left(\left(X_{i53} + X_{i54}\right)\beta\right)}{\exp\left(\left(X_{i53} + X_{i54}\right)\beta\right)} \tag{D.4}
$$

$$
= \frac{\exp\left(\left(X_{i31} + X_{i42} + X_{i53} + X_{i54}\right)\beta\right)}{\exp\left(\left(X_{i31} + X_{i42} + X_{i53} + X_{i54}\right)\beta\right) + \exp\left(\left(X_{i41} + X_{i32} + X_{i53} + X_{i54}\right)\beta\right)}
$$

$$
= \Pr\left[Y_i = (3,4,5,5) \middle| f^Z_{CP}(Y_i) = h, \beta\right],
$$

which is also equivalent to $\Pr\left[Y_{i1} = (3,4) \middle| f^Z_{CP}(Y_{i1}) = h_1, \beta\right]$. In this case, then, $f^Z_{CP}(Y_{i1}) = h_1 \subset f^Z_{CP}(Y_i) = z \subset f^L_{CP}(Y_i) = l$. By Proposition 4, 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 (D.1) requires stronger assumptions than model (D.3) for its consistent estimation. Consistent estimation of model (D.1) requires that alternatives $\{3,4,5\} \subseteq CS^\star_{it} = c_t$, $t = 1,2,3,4$. However, consistent estimation of model (D.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 (D.1) would not be consistent, while estimation of model (D.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 (D.1) and model (D.3) would be consistent. However, estimation of model (D.1) would be more efficient than estimation of model (D.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 (D.3) would be consistent. It follows that, under the maintained assumption of unobserved preference heterogeneity in a form encompassed by individual-alternative specific fixed effects, a test for H_0: *(choice set stability in* $1,2,3$ *and* 4*)* is $LR = 2\left[l^Z_{CP}\left(\widehat{\beta}^Z_{PPH}\right) + l_\triangle\left(\widehat{\beta}_\triangle\right) - l^L_{CP}\left(\widehat{\beta}^L_{CP}\right)\right]$.

E Bounding Choice Probabilities and Elasticities in Conditional Logit Models

As summarized in subsection 4.3, we are often interested in functions of parameters, for example, willingness to pay, elasticities, consumer surplus 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-identify individual specific choice probabilities and hence average choice probabilities. Still, under some weak assumptions, we can "bound" them. Also, given the convenient relationship between individual specific elasticities and individual specific choice probabilities in multinomial logit models, we can as well bound individual specific elasticities and their averages.[27]

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

$$V(X_{ij_tt}, \theta) = \delta_{j_t} + X_{ij_tt}\beta. \tag{E.1}$$

Denote by $Y_{it} = j_t$ whether product j_t is chosen by individual i in choice situation t, and by $f_t\left(Y_i\right) = r_{it}$ whether i's sufficient set collects alternatives r_{it} in choice situation t.

From the main text, we have that $f_t\left(Y_i\right) \subseteq CS^{\star}_{it}$ and that $CS^{\star}_{it} \subseteq Q_{it}$. In what follows we use these conditions to bound the true but unobserved denominator of the conditional logit choice probabilities for any X_{imt}, δ_m, and β:

$$\sum_{m \in f_t(Y_i) = r_{it}} \exp\left(\delta_m + X_{imt}\beta\right) \leq \sum_{m \in CS^{\star}_{it} = c_{it}} \exp\left(\delta_m + X_{imt}\beta\right) \leq \sum_{m \in Q_{it} = q_{it}} \exp\left(\delta_m + X_{imt}\beta\right). \tag{E.2}$$

For brevity, denote $Pr^{Q}_{ij_tt}(\theta) = \Pr[Y_{it} = j_t | Q_{it} = q_{it}, \theta]$, $Pr^{CS^{\star}}_{ij_tt}(\theta) = \Pr[Y_{it} = j_t | CS^{\star}_{it} = c_{it}, \theta]$, and $Pr^{f}_{ij_tt}(\theta) = \Pr[Y_{it} = j_t | f_t(Y_i) = r_{it}, \theta]$. It follows from (E.2) that for any $j_t \in f_t(Y_i) = r_{it}$:

[27] As explained in the main text of the paper, we can construct these bounds for conditional logit models but cannot for semiparametric models estimated by Pairwise Maximum Score Estimator. In the latter class of models we can point-identify preference parameters and willingness-to-pay for alternatives' characteristics, but it is not clear how to bound any parameter that requires knowledge of the distribution of the unobserved portion of utility.

$$Pr^{Q}_{ij_tt}(\theta) \leq Pr^{CS^\star}_{ij_tt}(\theta) \leq Pr^{f}_{ij_tt}(\theta). \tag{E.3}$$

Observe that $Pr^{f}_{ij_tt}(\theta)$ 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 $j_t \in q_{it}$ but $j_t \notin r_{it}$, $Pr^{f}_{ij_tt}(\theta)$ will not be a valid upper bound for $Pr^{CS^\star}_{ij_tt}(\theta)$: 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 $Pr^{CS^\star}_{ij_tt}(\theta) > 0$. Similarly, among the $j_t \in q_{it}$ that $j_t \notin r_{it}$, there can be some $j_t \notin c_{it}$. But for those $j_t \in q_{it}$ that $j_t \notin c_{it}$, $Pr^{Q}_{ij_tt}(\theta) > Pr^{CS^\star}_{ij_tt}(\theta) = 0$: $Pr^{Q}_{ij_tt}(\theta)$ will not be a valid lower bound for $Pr^{CS^\star}_{ij_tt}(\theta)$. It is then unclear how to bound $Pr^{CS^\star}_{ij_tt}(\theta)$ for those $j_t \in q_{it}$ but $j_t \notin r_{it}$. However, it is always possible to construct bounds for the probability with which i would choose j_t if indeed j_t were to be *added* to their true but unobserved choice set, $CS^\star_{it} \cup \{j_t\} = c_{it} \cup \{j_t\}$:

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

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

$$Pr^{Q\cup j}_{ij_tt}(\theta) \leq Pr^{CS^\star\cup j}_{ij_tt}(\theta) \leq Pr^{f\cup j}_{ij_tt}(\theta). \tag{E.5}$$

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

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

With linear indirect utilities (E.1), individual i's own- and cross-price elasticities are:

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

where $p_{j_t t}$ is j_t's price in choice situation t and β_p is the price coefficient. As (E.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 $Pr_{ij_t t}^{CS^\star \cup j}(\theta)$, $\forall j_t \in CS_{it}^\star \cup \{j_t\} = c_{it} \cup \{j_t\}$.

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

$$
\begin{aligned}
\underbrace{\beta_p p_{j_t t}(1 - Pr_{ij_t t}^{f \cup j}(\theta))}_{\text{Lower (in abs. value) Bound}} \le \xi_{it}^{jj}(X_{it},\theta) \le \underbrace{\beta_p p_{j_t t}(1 - Pr_{ij_t t}^{Q \cup j}(\theta))}_{\text{Upper (in abs. value) Bound}} \\
\underbrace{-\beta_p p_{k_t t} Pr_{ik_t t}^{Q \cup j}(\theta)}_{\text{Lower Bound}} \le \xi_{it}^{jk}(X_{it},\theta) \le \underbrace{-\beta_p p_{k_t t} Pr_{ik_t t}^{f \cup j}(\theta)}_{\text{Upper Bound}}.
\end{aligned}
\tag{E.8}
$$

E.1 Confidence Intervals for Elasticity Bounds in Conditional Logit Models

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 (E.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 (E.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:

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

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

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

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

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

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

In the extreme case in which $f_t(Y_i)\cup\{j_t\}=r_{it}\cup\{j_t\}=Q_{it}\cup\{j_t\}=q_{it}\cup\{j_t\}$, $\underline{\xi_{it}^{jk}}(X_{it},\theta)=\overline{\xi_{it}^{jk}}(X_{it},\theta)$ for any X_{it} and θ, and (E.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 (E.11) that overcomes this problem.) However, note that *(a)* both $f_t(Y_i)\cup\{j_t\}=r_{it}\cup\{j_t\}$ and $Q_{it}\cup\{j_t\}=q_{it}\cup\{j_t\}$ are always perfectly observed by the econometrician, so that the appropriate $CI_{1-\alpha}$ can always be implemented and that *(b)* in our empirical application $f_t(Y_i)\cup\{j_t\}\subset Q_{it}\cup\{j_t\}$ for every i and t.

F Data Appendix

In Section 5 we present an illustrative empirical example; here we describe the data used in that empirical example in greater detail.

F.1 Purchase data

We use 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 record 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 of their own TV viewing. We use information on 9387 purchase occasions over the period 2010-2011. A purchase occasion is when the woman 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 with a market share of 1%. There are multiple products per brand (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).

Consumers purchase products in different outlets. 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). We assume that the outlet that we observe the individual shopping in is chosen before, and independently from, the choice of a specific chocolate product (or none at all). Prices are measured on each individual transaction; we aggregate them to the level of the outlet and week. 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).

F.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:

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

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:

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

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,[28] 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:

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

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

[28] 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.

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, 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:

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

We define the *flow* as:

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

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.