Nonparametric Identification of Differentiated Products Demand Using Micro Data*

Steven T. Berry
Yale University

Philip A. Haile
Yale University

May 14, 2024

Abstract

We examine identification of differentiated products demand when one has "micro data" linking the characteristics and choices of individual consumers. Our model nests standard specifications featuring rich observed and unobserved consumer heterogeneity as well as product/market-level unobservables that introduce the problem of econometric endogeneity. Previous work establishes identification of such models using market-level data and instruments for all prices and quantities. Micro data provides a panel structure that facilitates richer demand specifications and reduces requirements on both the number and types of instrumental variables. We address identification of demand in the standard case in which non-price product characteristics are assumed exogenous, but also cover identification of demand elasticities and other key features when these product characteristics are endogenous and not instrumented. We discuss implications of these results for applied work.

Keywords: demand, nonparametric identification, differentiated products

*Early versions of this work were circulated under the title "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers." We thank Jesse Shapiro, Paulo Somaini, Suk Joon Son, and numerous seminar participants for helpful comments. Miho Hong and Jaewon Lee provided capable research assistance. We thank the referees and Co-Editor (Kate Ho) for helpful comments.

1 Introduction

Demand systems for differentiated products are central to many questions in economics. In practice it is common to estimate demand using data on the characteristics and choices of many individual consumers within each market. This setting is often referred to as "micro data," in contrast to another common case in which only market-level outcomes are observed.[1] At an intuitive level, the panel structure of micro data seems to offer more information than market-level data alone. But in what precise sense does micro data help? How significant are its advantages? What specific kinds of variation within and across markets are helpful, and how?

This paper explores these questions by examining nonparametric identification in a model substantially generalizing standard demand models used in a large literature building on Berry, Levinsohn, and Pakes (1995, 2004). Micro data provides a panel structure, with many consumers in each of many markets. A key benefit is that unobservables at the level of the product$\times$market remain fixed as consumers' attributes and choices (quantities demanded) vary within a given market. We show that identification can be obtained by combining this clean "within" variation with cross-market variation in choice characteristics, market characteristics, prices, and instruments for prices (only). Compared to settings with only market-level data, this both allows a more general demand model and substantially reduces instrumental variables requirements.

Although we focus exclusively on identification, our aim is to inform the practice and evaluation of empirical work. The celebrated "credibility revolution" in applied economics has redoubled attention to identification obtained through quasi-experimental variation, such as that arising through instrumental variables, geographic boundaries, or repeated observations within a single economic unit. Identification of demand presents challenges that are absent in much of empirical economics (see, e.g., Berry and Haile (2021)). Nonetheless, we show that these same types of variation allow identification of demand systems exhibiting rich consumer heterogeneity and endogeneity. Nonparametric identification results do not eliminate concerns about the impact of parametric assumptions relied on in practice. However, they address the important question of whether such assumptions can be viewed properly as finite-sample approximations rather than essential maintained hypotheses. Identification results can also clarify which assumptions may be most difficult to relax, reveal essential sources of variation, point to specific roles that functional forms may play in practice, offer assurance that robustness analysis is possible, and potentially lead to new (parametric or nonparametric) estimation approaches.

[1] See Berry and Haile (2021) for a discussion of other forms of data, including consumer panels and hybrids such as that in Petrin (2002).

Our most important message for applied work is that micro data has a high marginal value over market-level data alone. Availability of instrumental variables is the most important and challenging requirement for identification of demand, and micro data can substantially reduce both the number and types of instruments needed. Berry and Haile (2014) showed that with market-level data, nonparametric identification typically requires instruments for all quantities and prices. There, the so-called "BLP instruments" (i.e., exogenous characteristics of competing products) play a crucial role as instruments for quantities.[2] In contrast, here we find that with sufficiently rich micro data the only essential instruments are those for prices. This cuts the number of required instruments in half and avoids the necessary reliance on BLP instruments. This in turn permits a more flexible model of how non-price product characteristics affect demand and avoids the necessity that at least some such characteristics be exogenous. Micro data also opens the possible use of additional classes of instruments.

We also show that it is often possible to identify the *ceteris paribus* effects of prices on quantities demanded—critically, e.g., own- and cross-price demand elasticities—when observed non-price characteristics of products/markets are endogenous and not themselves instrumented. This requires that instruments for prices remain valid when conditioning on the endogenous non-price observables.[3] We show that standard instruments (or variations thereon) can satisfy this requirement under many models of endogeneity. Our analysis of instruments for this case makes elementary use of causal graphs, which provide attractive tools for evaluating the necessary exclusion condition. Endogenous product characteristics are an important concern in the applied literature on differentiated products, and it can be difficult to find instruments for all such characteristics. Thus, our findings expand the range of applications in which primary features of interest can be identified despite these concerns.

Our model and setting incorporate several key features. First, as in the large empirical literature building on Berry (1994) and Berry, Levinsohn, and Pakes (1995, 2004), we model market-level demand shocks (unobservables at the level of the product$\times$market) that result in the econometric endogeneity of prices. These shocks make identification of demand nontrivial and create a (perhaps counterintuitive) need for non-price sources of variation. Accounting for these demand shocks is essential to the identification of policy-relevant features such as demand elasticities and equilibrium counterfactuals. This drives our focus on market-level endogeneity and cross-market data, differentiating

[2]Instruments for quantities are what have sometimes been referred to informally as instruments for the "nonlinear parameters" in the applied literature using random coefficients discrete choice models. Berry and Haile (2021) provide additional discussion.

[3]This is related to well-known results regarding endogenous controls in regression models.

our work from much of the prior research on the identification of choice models with micro data.[4] To our knowledge, the examination of identification with market-level data in Berry and Haile (2014) offers the only prior nonparametric identification results applicable to the workhorse models of the large empirical literature that motivates our work.

Second, the panel structure of consumers-within-markets is essential to the questions we ask. It is what distinguishes micro data from market-level data. This panel structure is responsible for the reduction in the number of needed instrumental variables, as well as the elimination of restrictions on the way product-level observables enter demand. These features contrast with the setting and model in Berry and Haile (2014). There the demand system (especially if combined with a model of supply) also connects to nonparametric simultaneous equations models, as studied by, e.g., Benkard and Berry (2006), Matzkin (2008, 2015), Blundell, Kristensen, and Matzkin (2013, 2020), and Berry and Haile (2018). However, the panel structure essential to the present paper is absent in all of that prior work.

Third, our model avoids requirements that consumer-level observables be exogenous, that they have large support, or that certain consumer observables be linked exclusively to the desirability of specific products. The last of those requirements is widely used (often in combination with large support and exogeneity assumptions) in "special regressor" approaches to identification of consumer-level discrete choice models,[5] but is often difficult to motivate in practice. More natural are situations in which multiple consumer-level observables interact to alter tastes for all goods. As a simple example—one illustrating a broader interpretation of "demand"—consider a discrete choice model of expressive voting in a two-party ("R" vs. "D") election, applied to survey data matching individual reported votes to voter sociodemographics.[6] Although voter-specific measures like age, income, gender, race, and educa-

[4]This includes prior work on discrete choice models allowing market-level demand shocks only though composite error terms—one for each choice—representing all latent heterogeneity (e.g., Lewbel (2000)). Explicit modeling of demand shocks also makes clear that the functional form assumptions permitting application of control function approaches (Blundell and Matzkin (2014)) generally fail, even in standard parametric models. See Berry and Haile (2021) for further discussion of the challenges created by these demand shocks.

[5]See the review by Lewbel (2014) and references therein. A very early version of this paper (Berry and Haile (2010)) featured an example of such an approach. In practice, geographic distances are often modeled as providing consumer-level variation exclusive to each product. But even these are inherently restricted to lie on a 2-dimensional surface in $\mathbb{R}_+^J$, since the underlying consumer heterogeneity reflects only consumer locations.

[6]Candidate spending, rather than price, often plays the role of the endogenous choice characteristic—one whose effects are sometimes of primary interest. See, e.g., Gerber (1998) and Gordon and Hartmann (2013).

tion may provide rich variation in preferences between the two parties (and the outside option to abstain), no such measure is naturally associated exclusively with the attractiveness of a single option.

Fourth, although we initially emphasize discrete choice demand, this is not essential. The structural feature of interest in our analysis is a demand function mapping observables (at the level of market, products, and consumer) and a vector of market-level demand shocks to expected quantities demanded. This can allow continuous demand as well as departures from common assumptions regarding consumers' full information or rationality.

Of course, our results do require some structure, including conditions on sources of variation. In addition to instruments for prices satisfying standard conditions, we rely on three important assumptions. One is a nonparametric index restriction on the way market-level demand shocks and some observed consumer attributes enter the model.[7] The second is injectivity of the mappings that link observed consumer attributes to choice probabilities. Below we connect these requirements to canonical specifications from the literature. Finally, we require at least as many observed consumer attributes as goods.

This final requirement is tightly related to our focus on the extent to which micro data can fully eliminate the need to instrument for the endogenous quantities in a flexible nonparametric demand model that, e.g., places few restrictions on substitution patterns or how price effects vary across products. In practice, some settings—particularly those with a large number of goods—may lack the dimension of variation we require for the most flexible models. This motivates our exploration (in the Supplemental Appendix) of trade-offs between common modeling assumptions and the types/dimension of variation sufficient for identification. For example, we discuss semiparametric restrictions that can allow identification (even with a large choice set) with a single consumer-level observed attribute having only binary support.

Our results are relevant to a large empirical literature exploiting micro data to estimate demand. A classic example is McFadden's study of transportation demand (McFadden, Talvitie, and Associates (1977)), where each consumer's preferences over different modes of transport are affected by her available mode-specific commute times and other factors. This example illustrates an essential feature of the type of micro data considered here: consumer-specific observables that alter the relative attractiveness of different choices. Consumer distances to different options have been used in a number of applications, including those involving demand for hospitals, retail outlets, residential

[7]In contrast to related conditions in Berry and Haile (2014, 2018) or Matzkin (2008), here each index depends on observed consumer attributes rather than observed product characteristics (which are fixed within markets), and there is no requirement that these observables be exogenous.

locations, or schools, as in the examples of Capps, Dranove, and Satterthwaite (2003), Burda, Harding, and Hausman (2015), Bayer, Keohane, and Timmins (2009), and Neilson (2021). More broadly, observable consumer-level attributes that shift tastes for products might include income, sociodemographic measures, or other proxies for idiosyncratic preferences. For example, income and family size have been modeled as shifting preferences for cars (Goldberg (1995), Petrin (2002)); race, education, and birth state have been modeled as shifting preferences for residential location (Diamond (2016)). Other prominent examples include applications to demand for grocery products (Ackerberg (2003)), newspapers (Gentzkow and Shapiro (2010)), neighborhoods (Bayer, Ferreira, and McMillan (2007)), and schools (Hom (2018)).[8] An important feature of many examples, reflected by our model, is that the typical consumer-level observable cannot be tied exclusively to a single good.

In what follows, section 2 sets up our model of multinomial choice demand. Section 3 connects this model to random coefficients random utility specifications widely used in practice. We present our identification results in section 4. We discuss some key implications for applied work in section 5 before concluding in section 6. Appendix A examines the proper excludability of standard instruments for prices when non-price observables are endogenous and uninstrumented. A Supplemental Appendix discusses variations on our baseline model, including continuous demand and examples in which key assumptions can be relaxed by strengthening others.

2 Model and Features of Interest

We consider choice among J goods/products and an outside option (indexed as good 0) by consumers i in "markets" t. Formally, a market is defined by:

- a price vector $P_t = (P_{1t}, \ldots, P_{Jt})$;
- a set of additional observables X_t;
- a vector $\Xi_t = (\Xi_{1t}, \ldots, \Xi_{Jt})$ of unobservables;
- a distribution $F_Z(\cdot; t)$ of consumer observables $Z_{it} \in \mathbb{R}^H$, $H \geq J$, with support $\mathcal{Z}(X_t)$.

For clarity, we write random variables (all of which have t among their subscript indices) in uppercase and their realizations in lowercase. The variables

[8] Here we cite only a small representative handful of papers out of a selection that spans many topics and many years. Recent development of commercial consumer-level data sets suggests the potential for micro data to play an even larger role in the future.

(P_t, X_t, Ξ_t) are common to all consumers in a given market.[9] We distinguish between P_t and X_t due to the particular interest in how demand responds to prices and the typical focus on endogeneity of prices. However, we have not yet made the standard assumption that X_t is exogenous—e.g., independent or mean independent of the demand shocks Ξ_t. We will see below that identification of demand elasticities and other key features of demand can often be obtained without such an assumption (or additional instruments for X_t).[10]

Although X_t will typically include observable product characteristics (possibly including characteristics of the outside option), it may also include other factors defining markets.[11] For example, consumers might be partitioned into "markets" based on a combination of geography, time, product availability, and demographics (e.g., binned age or income) included in X_t. In contrast, observables varying across consumers within a market are represented by Z_{it}. Key conditions, made precise below, are that consumer observables provide variation of dimension at least J (hence $H \geq J$), with J components of Z_{it} entering the model through an index structure. When $H > J$, the "extra" consumer observables can be treated fully flexibly; thus, without further loss, we henceforth assume $H = J$.[12] We emphasize that although our requirements on Z_{it} permit the case in which each component Z_{ijt} exclusively affects the attractiveness of good j, we will not require this. Nor will we require independence (full, conditional, or mean independence) between Z_{it} and Ξ_t.

The choice environment of consumer i in market t is then represented by

$$C_{it} = (Z_{it}, P_t, X_t, \Xi_t)\,.$$

Let $\mathcal{C}$ denote the support of C_{it}. A basic primitive characterizing consumer

[9]The assumption that all consumers in a market face the same prices and product characteristics is standard and may influence how markets are defined in practice. This rules out some forms of price discrimination.

[10]Alternatively, when instruments are available for endogenous components of X_t, our results generalize immediately by expanding P_t to include these endogenous characteristics.

[11] X_t could also include product fixed effects. Such fixed effects generally do not address the endogeneity challenges central to identification of demand (see, e.g., Berry and Haile (2021)). With additional assumptions, cross-market variation in the number of goods available could be valuable; e.g., data from markets with J available goods could be used to predict outcomes in markets with more or fewer goods.

[12]For example, one may interpret all that follows as applying conditional on each value of the "extra" consumer observables. Berry and Haile (2023) provide a more explicit and slightly more flexible treatment.

behavior in this setting is a distribution of decision rules for each $c_{it} \in \mathcal{C}$.[13] As usual, heterogeneity in decision rules (i.e., nondegeneracy of the distribution) within a given choice environment may reflect a variety of factors, including latent preference heterogeneity across consumers, shocks to individual preferences, latent variation in consideration sets, or stochastic elements of choice (e.g., optimization error).[14]

2.1 Demand and Conditional Demand

The choice made by consumer i is represented by $Q_{it} = (Q_{i1t}, \ldots, Q_{iJt})$, where Q_{ijt} denotes the quantity (here, 0 or 1) of good j purchased. Given C_{it}, a distribution of decision rules is fully characterized by the conditional cumulative joint distribution function $F_Q\left(q|C_{it}\right) = \mathbb{E}\left[1\left\{Q_{it} \leq q\right\}|C_{it}\right]$. In the case of discrete choice, this distribution can be represented without loss by the structural choice probabilities

$$\mathscr{s}\left(C_{it}\right) = \left(\mathscr{s}_1(C_{it}), \ldots, \mathscr{s}_J(C_{it})\right) = \mathbb{E}\left[Q_{it}|C_{it}\right]. \tag{1}$$

Given the measure of consumers in each choice environment, the mapping $\mathscr{s}$ fully characterizes consumer demand. We will therefore consider identification of the latent demand shocks and the *demand* mapping $\mathscr{s}$ on $\mathcal{C}$.

However, it is useful to also consider identification of the *conditional demand* functions

$$\bar{\mathscr{s}}\left(Z_{it}, P_t; t\right) \equiv \mathscr{s}\left(Z_{it}, P_t, x_t, \xi_t\right)$$

on

$$\mathcal{C}(x_t, \xi_t) = \text{supp }\left(Z_{it}, P_t\right) | \left\{X_t = x_t, \Xi_t = \xi_t\right\}$$

for each market t. The function $\bar{\mathscr{s}}(\cdot; t)$ is simply the demand function $\mathscr{s}$ when (X_t, Ξ_t) are fixed at the values (x_t, ξ_t) realized in market t. Because Ξ_t is

---

[13] Under additional conditions a distribution of decision rules can be represented as the result of utility maximization. See, e.g., Mas-Colell, Whinston, and Green (1995), Block and Marschak (1960), Falmagne (1978), and McFadden (2005). We will not require such conditions or consider a utility-based representation. A related issue is identification of welfare effects. Standard results allow construction of valid measures of aggregate welfare changes from a known demand system in the absence of income effects. Bhattacharya (2018) provides such results for discrete choice settings when income effects are present.

[14] For many purposes, one need not take a stand on the interpretation of this randomness, since the economic questions of interest involve changes to the arguments of demand functions, not to the functions themselves. This covers the canonical motivation for demand estimation: quantifying responses to *ceteris paribus* price changes. However, for some questions—e.g., those involving information interventions or requiring identification of cardinal utilities—the interpretation becomes important. See Barseghyan, Coughlin, Molinari, and Teitelbaum (2021) for a recent contribution on this topic.

unobserved and prices are fixed within each market, identification of $\bar{\mathfrak{s}}(\cdot;t)$ is nontrivial. However, this mapping fully characterizes the responses of demand (at all values of Z_{it}) to counterfactual *ceteris paribus* price variation, holding X_t and Ξ_t fixed at their realized values in market t. Thus, knowledge of $\bar{\mathfrak{s}}(\cdot;t)$ for each market t suffices for many purposes motivating demand estimation in practice.

Notably, $\bar{\mathfrak{s}}(\cdot;t)$ fully determines the own- and cross-price demand elasticities for all goods in market t. One implication is that $\bar{\mathfrak{s}}(\cdot;t)$ is the feature of $\mathfrak{s}$ needed to discriminate between alternative models of firm competition (e.g., Berry and Haile (2014), Backus, Conlon, and Sinkinson (2021), Duarte, Magnolfi, Sølvsten, and Sullivan (2023)). And, given a model of supply, $\bar{\mathfrak{s}}(\cdot;t)$ suffices to identify firm markups and marginal costs, following Berry, Levinsohn, and Pakes (1995) and Berry and Haile (2014); to decompose the sources of firms' market power, as in Nevo (2001); to determine equilibrium outcomes under a counterfactual tax, tariff, subsidy, or exchange rate (e.g., Anderson, de Palma, and Kreider (2001), Nakamura and Zerom (2010), Decarolis, Polyakova, and Ryan (2020)); or to determine equilibrium "unilateral effects" of a merger (e.g., Nevo (2000), Miller and Sheu (2021)). Furthermore, $\bar{\mathfrak{s}}(\cdot;t)$ alone determines the "diversion ratios" (e.g., Conlon and Mortimer (2021)) that often play a central role in the practice of antitrust merger review.

Of course, because the functions $\bar{\mathfrak{s}}(\cdot;t)$ are defined with fixed values of (X_t, Ξ_t), they do not suffice for answering all questions—in particular, those requiring knowledge of *ceteris paribus* effects of X_t on demand.[15] However, by avoiding the need to separate the effects of X_t and Ξ_t, identification of $\bar{\mathfrak{s}}(\cdot;t)$ in each market t can often be obtained without requiring exogeneity of X_t. This can be important when exogeneity is in doubt and one lacks the additional instruments that would allow treating endogenous elements of X_t as we treat prices P_t below.

2.2 Core Assumptions

So far we have made three significant assumptions:

(i) latent market-level heterogeneity can be represented by a J-vector Ξ_t;
(ii) conditional on X_t, the support of Z_{it} is the same in all markets; and
(iii) the consumer-level observables Z_{it} have dimension (of at least) J.

The first is important but standard (when market-level unobservables are ac-

[15] In some cases, such effects may be of direct interest—e.g., to infer willingness to pay for certain product features. In other cases, such effects are inputs to determination of demand under counterfactual product offerings or entry. Thus, while knowledge of $\bar{\mathfrak{s}}(\cdot;t)$ in all markets suffices in a large fraction of applications, knowledge of $\mathfrak{s}$ is required for others.

knowledged). The second assumption seems mild for most applications,[16] and it can be relaxed at the cost of more cumbersome exposition.[17]

The third assumption is both important and restrictive. Although some modern micro data sets can offer dozens or even hundreds of consumer-level observables, others offer a much smaller set of consumer-level measures. Thus, full satisfaction of this requirement will depend on the dimension of the choice set and the richness of the micro data. We focus on J-dimensional Z_{it} to explore the the extent to which micro data can fully eliminate the need to instrument for J endogenous quantities in a flexible nonparametric model. The Supplemental Appendix illustrates some of the ways this requirement on the dimension of Z_{it} can be relaxed. For example, we discuss semiparametric models in which a single binary consumer characteristic Z_{it} and a single excluded instrument for price can suffice.

We will also rely on Assumptions 1–5 below, where we let $\mathcal{X}$ and $\mathcal{PX}$ denote the supports of X_t and (P_t, X_t), respectively. Assumptions 1–3 introduce an index structure that serves a key element of our strategy for demonstrating the gains from micro data: inferring variation across markets in the demand shock vector Ξ_t using variation in the value of the vector Z_{it} required to produce particular choice probabilities.

Assumption 1 (Index). *$\mathcal{s}(C_{it}) = \sigma(\gamma(Z_{it}, X_t, \Xi_t), P_t, X_t)$, where $\gamma(Z_{it}, X_t, \Xi_t) = (\gamma_1(Z_{it}, X_t, \Xi_t), \ldots, \gamma_J(Z_{it}, X_t, \Xi_t)) \in \mathbb{R}^J$ and $\gamma_j(Z_{it}, X_t, \Xi_t) = \Gamma_j(Z_{it}, X_t) + \Xi_{jt}$ for all j.*

Assumption 2 (Invertible Demand). *For all $(p, x) \in \mathcal{PX}$, $\sigma(\cdot, p, x)$ is injective on the support of $\gamma(Z_{it}, x, \Xi_t)$.*

Assumption 3 (Injective Index). *For all $(x, \xi) \in supp(X_t, \Xi_t)$, $\gamma(\cdot, x, \xi)$ is injective on $\mathcal{Z}(x)$.*

Assumption 1 requires that Z_{it} and Ξ_t affect choices only through a J-vector of indices in which prices are excluded and each Ξ_{jt} is exclusive to the jth component, entering this component of the index in additively separable form.[18] This is an important nonparametric functional form restriction. A key implication is a sense in which variation in Z_{it} has comparable effects across markets: a unit change in $\Gamma_j(Z_{it}, X_t)$ has the same effect on demand

[16] Note that the distribution of Z_{it} over this support may vary freely across markets, including in ways that depend on Ξ_t.

[17] For example, our results go through without this assumption if we instead let each $\mathcal{Z}(x)$ be a nonempty set (assumed open and connected below) common to the support of Z_{it} in all markets for which $X_t = x$.

[18] We show in Berry and Haile (2023) that the separability requirement can be dropped by strengthening other conditions.

as a unit change in Ξ_{jt}.[19] We link this structure (and the remaining core assumptions below) to familiar specifications from the literature in section 3. In the Supplemental Appendix we discuss a more general model permitting the price of good j to enter the jth component of the index vector.

Assumption 2 requires that the choice probability function σ be "invertible" with respect to the index vector—that, holding (P_t, X_t) fixed, distinct index vectors map to distinct choice probabilities. Berry, Gandhi, and Haile (2013) provide sufficient conditions for invertibility of a demand system, which are natural here when each $\gamma_j(Z_{it}, X_t, \Xi_t)$ can be interpreted as a quality shifter for good j.[20]Assumption 3 requires injectivity of the index function γ with respect to the vector Z_{it}. This generalizes common utility-based specifications of demand while avoiding the common requirement that each Z_{ijt} affect the utility of good j exclusively.

In addition to the index structure, we assume that both Z_t and Ξ_t have uncountable connected support. This allows transparent exploration of the potential gains from micro data, using calculus and moment equalities. Assumption 5, in addition, rules out trivial cases in which conditioning on (P_t, X_t) indirectly fixes Ξ_t as well. Such cases are ruled out by standard models of supply, where prices respond to continuous cost shifters or markup shifters (observed or unobserved), allowing the same equilibrium price vector p to arise under different realizations of Ξ_t.

Assumption 4 (Support)**.** *For all $x \in \mathcal{X}$, $\mathcal{Z}(x)$ is open and connected.*

Assumption 5 (Nondegeneracy)**.** *For each $x \in \mathcal{X}$, there exists (possibly unknown) $p \in supp\, P_t|\{X_t = x\}$ such that $supp\, \Xi_t|\{(P_t = p, X_t = x)\}$ is open and connected.*

Assumption 4's requirement of continuous variation in Z_{it} is important to our arguments. Absent appropriate restrictions on other elements of the model, it may not be surprising that continuous variation would be required for nonparametric point identification. Indeed, below we will require continuously distributed instruments for prices as well. In practice, of course, one may often rely on at least some instruments or consumer-level observables with discrete (even binary) support. As in other types of empirical models, in such cases parametric forms used in estimation will typically fill the gaps left by the

---

[19] A sufficient condition is that this equivalence hold for the determination of Q_{it} at the individual consumer level. In typical linear random utility discrete choice models (see section 3) this would allow for a random coefficient on each index $\Gamma_j(Z_{it}, X_t) + \Xi_{jt}$ but would otherwise rule out random coefficients on Ξ_{jt} or on terms involving Z_{it}.

[20] Other general sufficient conditions for injectivity can be found in, e.g., Palais (1959), Gale and Nikaido (1965), and Parthasarathy (1983).

available variation. The Supplemental Appendix considers identification in an example with discrete Z_{it}.

2.3 A Useful Representation of the Index

We have thus far written the index vector

$$\gamma\left(Z_{it}, X_t, \Xi_t\right) \equiv \Gamma\left(Z_{it}, X_t\right) + \Xi_t$$

in a way that maximizes clarity about our core assumptions. To demonstrate identification while allowing for endogenous X_t, it is more convenient to define

$$g\left(Z_{it}, X_t\right) = \Gamma\left(Z_{it}, X_t\right) + E\left[\Xi_t | X_t\right]$$

and

$$h\left(X_t, \Xi_t\right) = \Xi_t - \mathbb{E}\left[\Xi_t | X_t\right], \tag{2}$$

so that the index can be rewritten as

$$\gamma\left(Z_{it}, X_t, \Xi_t\right) = g\left(Z_{it}, X_t\right) + h\left(X_t, \Xi_t\right). \tag{3}$$

Here we have simply let $g\left(Z_{it}, X_t\right)$ absorb the mean of Ξ_t conditional on X_t, leaving the residualized structural error vector $h\left(X_t, \Xi_t\right)$. Observe that

$$\mathbb{E}\left[h\left(X_t, \Xi_t\right) | X_t\right] = 0 \tag{4}$$

by construction.

With this notation, we have

$$s\left(C_{it}\right) = \sigma\left(g\left(Z_{it}, X_t\right) + h\left(X_t, \Xi_t\right), P_t, X_t\right) \tag{5}$$

and

$$\bar{s}\left(Z_{it}, P_t; t\right) = \sigma\left(g\left(Z_{it}, x_t\right) + h(x_t, \xi_t), P_t, x_t\right) \tag{6}$$

We henceforth work with this representation of the demand and conditional demand functions.

2.4 Technical Conditions

In parts (i) and (ii) of Assumption 6 below we assume smoothness conditions permitting our applications of calculus and continuity arguments below. Parts (iii) and (iv) strengthen the injectivity requirements of Assumptions 2 and 3 slightly by requiring that the Jacobian matrices $\partial g(z,x)/\partial z$ and $\partial \sigma(\gamma, p, x)/\partial \gamma$ be nonsingular almost surely.

Assumption 6 (Technical Conditions). *For all $(p, x) \in \mathcal{PX}$, (i) $g(\cdot, x)$ is uniformly continuous and continuously differentiable; (ii) $\sigma(\cdot, p, x)$ is continuously differentiable; (iii) $\partial g(z,x)/\partial z$ is nonsingular a.s. on $\mathcal{Z}(x)$; and (iv) $\partial \sigma(\gamma, p, x)/\partial \gamma$ is nonsingular a.s. on supp $\gamma(Z_{it}, X_t, \Xi_t) | \{(P_t, X_t) = (p, x)\}$.*

2.5 Normalization

The model requires two types of normalizations before the identification question can be properly posed. The first reflects the fact that the latent demand shocks have no natural location. Thus, we set $\mathbb{E}[\Xi_t] = 0$ without loss. The second reflects the fact that arbitrary injective transformations of the index vector $\gamma(Z_{it}, X_t, \Xi_t)$ can be reversed by appropriate modification of the unknown function σ. For example, take arbitrary $A(X_t) : \mathcal{X} \to \mathbb{R}^J$ and $B(X_t) : \mathcal{X} \to \mathbb{R}^{J \times J}$, with $B(x)$ nonsingular at all x. By letting

$$\tilde{\gamma}(Z_{it}, X_t, \Xi_t) = A(X_t) + B(X_t)\gamma(Z_{it}, X_t, \Xi_t) \tag{7}$$

$$\tilde{\sigma}(\tilde{\gamma}(Z_{it}, X_t, \Xi_t), P_t, X_t) = \sigma\left(B(X_t)^{-1}(\tilde{\gamma}(Z_{it}, X_t, \Xi_t) - A(X_t)), P_t, X_t\right) \tag{8}$$

one obtains a new representation of the same map from $(Z_{it}, P_t, X_t, \Xi_t)$ to quantities demanded, the new one satisfying our assumptions whenever the original does. We must choose a single representation of demand before exploring whether the observables allow identification.[21]

To do this, for each $x \in \mathcal{X}$ we take an arbitrary $z^0(x) \in \mathcal{Z}(x)$ such that $\partial\bar{\mathfrak{s}}(z^0(x), p_t; t)/\partial z$ is nonsingular in some market t for which $X_t = x$, guaranteeing that $\frac{\partial g\left(z^0(x), x\right)}{\partial z}$ is nonsingular.[22] We then select the representation of demand in which

$$g\left(z^0(x), x\right) = 0 \tag{9}$$

and

$$\left.\frac{\partial g(z, x)}{\partial z}\right|_{z = z^0(x)} = I, \tag{10}$$

where I denotes the J-dimensional identity matrix. In the example above this choice of normalization is equivalent to taking

$$B(x) = \left.\left[\frac{\partial g(z, x)}{\partial z}\right]^{-1}\right|_{z = z^0(x)}$$

and

$$A(x) = -B(x)g\left(z^0(x), x\right)$$

at each x, then dropping the tildes from the transformed model.

---

[21]Like location and scale normalizations of utility functions, our normalizations place no restriction on the demand function $\mathscr{s}$ or the conditional demand functions $\bar{\mathfrak{s}}(\cdot; t)$. In practice, normalizations are often embedded in parametric functional forms and exclusivity assumptions. See the Supplemental Appendix.

[22]By Assumption 6, almost all points in $\mathcal{Z}(x)$ will satisfy this condition; and $\partial\bar{\mathfrak{s}}\left(z^0(x), p_t; t\right)/\partial z$ is observable in every market t for which $X_t = x$.

3 Example

The literature includes many examples of parametric special cases of our model. The canonical discrete choice demand model (e.g., Berry, Levinsohn, and Pakes (2004)) is derived from a random utility specification like

$$\begin{aligned} u_{ijt} &= x_{jt}\beta_{ijt} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt} \quad j = 1, \ldots, J \\ u_{i0t} &= \epsilon_{i0t}, \end{aligned} \tag{11}$$

where u_{ijt} is consumer i's conditional indirect utility from good j in market t. Here, consumer-level heterogeneity is reflected by the idiosyncratic taste shocks $\epsilon_{it} = (\epsilon_{i0t}, \ldots, \epsilon_{iJt})$ (e.g., draws from a normal or type-1 extreme value distribution) as well as the random coefficients α_{it} and $\beta_{ijt} \in \mathbb{R}^K$.[23] These random coefficients are often specified as

$$\alpha_{it} = \exp\left(\alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}\right), \tag{12}$$

$$\beta_{ijt}^{(k)} = \beta_{0j}^{(k)} + \sum_{\ell=1}^{L} \beta_{zj}^{(k,\ell)} z_{i\ell t} + \beta_{\nu j}^{(k)} \nu_{it}^{(k)} \quad k = 1, \ldots, K, \tag{13}$$

where each $z_{i\ell t}$ represents one of L observable characteristic of consumer (or household) i, y_{it} represents consumer/household income—a consumer-level observable beyond those in z_{it},[24] and each $\nu_{it} = (\nu_{it}^{(0)}, \ldots, \nu_{it}^{(K)})$ is a random vector drawn from a pre-specified distribution. The preference shocks $(\nu_{it}, \epsilon_{it})$ are assumed i.i.d. across consumers and markets.

With (12) and (13), we can rewrite (11) as

$$u_{ijt} = g_j\left(z_{it}, x_t\right) + \xi_{jt} + \mu_{ijt}, \tag{14}$$

where

$$g_j\left(z_{it}, x_t\right) = \sum_k x_{jt}^{(k)} \sum_{\ell=1}^{L} \beta_{zj}^{(k,\ell)} z_{i\ell t} \tag{15}$$

$$\mu_{ijt} = \sum_k x_{jt}^{(k)} \left(\beta_{0j}^{(k)} + \beta_{\nu j}^{(k)} \nu_{it}^{(k)}\right) - \exp\left(\alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}\right) p_{jt} + \epsilon_{ijt}. \tag{16}$$

[23] Random coefficients are popular in discrete choice models because they can allow demand systems whose substitution patterns are more flexible along certain dimensions. We focus directly on identification of a demand system with very flexible substitution patterns. Section 5.1 provides additional discussion.

[24] Recall that our analysis suppresses such "extra" consumer observables but treats them fully flexibly. Section S.1 of the Supplemental Appendix discusses a variation of our model allowing prices to interact with the J-dimensional Z_{it} itself.

Observe that all effects of z_{it} and ξ_t operate through indices

$$\gamma_j\left(z_{it}, x_t, \xi_t\right) = g_j\left(z_{it}, x_t\right) + \xi_{jt} \qquad j = 1, \ldots, J,$$

satisfying our Assumption 1. The injectivity of demand required by Assumption 2 can be confirmed following Berry (1994) under typical distributional assumptions, or Berry, Gandhi, and Haile (2013) under more general conditions. Assumption 3—invertibility of the linear mapping $g(z_{it}, x_t) = (g_1(z_{it}, x_t), \ldots, g_J(z_{it}, x_t))$ in a J-dimensional subvector of z_{it}—clearly requires $L \geq J$. The additional conditions required for the assumption might then be assumed directly or derived from other conditions.

This example connects our nonparametric model to a large number of applications. Of course, our model generalizes the example substantially. It does not require linear utility functions,[25] parametric distributional assumptions, or even a representation of demand through utility maximization. Even within the linear random coefficients random utility discrete choice paradigm, our model would allow the joint distribution of $(\nu_{it}, \epsilon_{it})$ to depend on $(g(z_{it}, x_t) + \xi_t, p_t, x_t)$. More generally, consumer heterogeneity in our model is not limited to a finite vector of shocks entering with particular functional forms: utilities could be specified as nonparametric random functions of $(g(z_{it}, x_t) + \xi_t, p_t, x_t)$.

Finally, observe that the example above lacks features sometimes relied on in results showing identification of discrete choice models: aside from the absence of individual characteristics that exclusively affect the utility from one choice j, this model does not exhibit independence between the "error term" $(\xi_{jt} + \mu_{ijt})$ in (14) and any of the observables z_{it}, p_t, x_t.[26]

4 Identification

We consider identification of the latent shocks ξ_t, the demand system $\mathfrak{s}(C_{it})$, and the conditional demand systems $\bar{\mathfrak{s}}(Z_{it}, P_t; t)$. The observables comprise the market index t, the variables $(Q_{it}, Z_{it}, P_t, X_t)$, and a vector of instruments W_t discussed below. As usual, to consider identification we treat the population joint distribution of the observables as known. Loosely, we may view this as the result of observing $(Q_{it}, Z_{it}, P_t, X_t)$ for many markets t and many consumers i

[25] Thus, it would allow generalization of the nonparametric random utility model in Allen and Rehbeck (2019) to incorporate market-level demand shocks, flexible heterogeneity in tastes for product characteristics, and nonadditive product-level taste shocks.

[26] The variables $(x_{jt}, p_{jt}, \xi_{jt})$ directly enter the composite error $\xi_{jt} + \mu_{ijt}$. Furthermore, x_{jt} and p_{jt} may be correlated with changes in the distribution of z_{it} across markets, introducing dependence between z_{it} and μ_{ijt}.

in each market. These observables imply observability of choice probabilities conditional on (Z_{it}, P_t, X_t) in each market t. Of course, this implies observability of market-level choice probabilities (market shares) as well. We will use the compact notation $s_t(z)$ to denote the observed choice probability in market t conditional on $Z_{it} = z$.

We proceed in three steps. First, in section 4.1 we demonstrate identification of the function $g(\cdot, x)$ at each $x \in \mathcal{X}$. In section 4.2 we use this result to link latent market-level variation in $h(X_t, \Xi_t)$ to variation in the observed value of Z_{it} required to produce certain conditional choice probabilities in each market. In particular, given instruments for prices, we show that the realized values $h(x_t, \xi_t)$ can be pinned down in every market, making identification of the conditional demand systems $\bar{\jmath}(\cdot; t)$ in each market straightforward. Finally, in section 4.3 we show that $\jmath$ and each ξ_t are also identified when one adds the usual assumption that X_t is exogenous. Thus, after the initial setup and lemmas, the main results themselves follow relatively easily.

Before proceeding, we provide some key definitions and observations. For $(p, x, \xi) \in \text{supp } (P_t, X_t, \Xi_t)$ let

$$\mathcal{S}\left(p, x, \xi\right) = \sigma\left(g\left(\mathcal{Z}(x), x\right) + h(x, \xi), p, x\right),$$

and

$$\mathcal{S}(p, x) = \cup_{\xi \in \text{supp } \Xi_t | \{P_t = p, X_t = x\}} \mathcal{S}(p, x, \xi).$$

Thus, $\mathcal{S}\left(p, x, \xi\right)$ denotes the support of choice probabilities within any market t for which $P_t = p, X_t = x$, and $\Xi_t = \xi$. $\mathcal{S}(p, x)$ denotes the support (within and across markets) conditional on $\{P_t = p, X_t = x\}$.[27]

By Assumptions 2 and 3, for each $s \in \mathcal{S}\left(p, x, \xi\right)$ there is a unique $z^* \in \mathcal{Z}(x)$ such that $\sigma\left(g\left(z^*, x\right) + h(x, \xi), p, x\right) = s$. So for $(p, x, \xi) \in \text{supp } (P_t, X_t, \Xi_t)$ and $s \in \mathcal{S}\left(p, x, \xi\right)$, we define the function $z^*\left(s; p, x, \xi\right)$ implicitly by

$$\sigma\left(g\big(z^*\left(s; p, x, \xi\right), x\big) + h(x, \xi), p, x\right) = s. \tag{17}$$

We also define the compact notation

$$z_t^*(s) \equiv z^*\left(s; p_t, x_t, \xi_t\right). \tag{18}$$

These definitions lead to two observations that play key roles in what follows. First, in each market t the set $\mathcal{S}(p_t, x_t, \xi_t)$ and the values of $z_t^*(s)$ for

[27] Because $\mathcal{Z}(x)$ is open, continuity and injectivity of σ with respect to the index vector and of the index function with respect to Z_{it} imply that $\mathcal{S}\left(p, x, \xi\right)$ (and, thus, $\mathcal{S}(p, x)$ as well) is open. Continuity of $\sigma(\cdot)$ with respect to the index vector and continuity of $g(\cdot)$ with respect to Z_{it} imply (recalling Assumption 4) that $\mathcal{S}(p, x)$ is connected.

all $s \in \mathcal{S}(p_t, x_t, \xi_t)$ are directly observed, even though the value of ξ_t is not. Second, by the invertibility of σ (Assumption 2), we have

$$g\left(z^*\left(s; p, x, \xi\right), x\right) + h(x, \xi) = \sigma^{-1}\left(s; p, x\right) \tag{19}$$

for all $(p, x, \xi) \in \text{supp}\,(P_t, X_t, \Xi_t)$ and $s \in \mathcal{S}(p, x, \xi)$.

4.1 Initial Steps

Let $||\cdot||$ denote the Euclidean norm. A key implication of our nondegeneracy condition (Assumption 5) follows from the definition (2): for each $x \in \mathcal{X}$ there exist $p \in \text{supp}\, P_t|\{X_t = x\}$ and $\epsilon > 0$ such that for any $d \in \mathbb{R}^J$ satisfying $||d|| < \epsilon$, $\text{supp}\, \Xi_t|\{P_t = p, X_t = x\}$ contains vectors ξ and ξ' satisfying $h(x, \xi) - h(x, \xi') = d$. This is exploited to prove Lemma 1, which demonstrates how local variation across markets in the latent Ξ_t will produce local variation in observed values of $z_t^*(s)$ for those markets.

Lemma 1. *Let Assumptions 1–6 hold. For each $x \in \mathcal{X}$ there exist $p \in \text{supp}\, P_t|\{X_t = x\}$ and $\Delta > 0$ such that for all z and z' in $\mathcal{Z}(x)$ satisfying*

$$||z' - z|| < \Delta, \tag{20}$$

there exist a choice probability vector s and vectors ξ and ξ' in $\text{supp}\, \Xi_t|\{P_t = p, X_t = x\}$ such that $z = z^*(s; p, x, \xi)$ and $z' = z^*(s; p, x, \xi')$. Furthermore, such (p, Δ) are identified.

Proof. See Appendix B. □

With this result in hand, Lemma 2 demonstrates that one can use equation (19) to relate partial derivatives of $g(z, x)$ at any point z to those at nearby points z' by examining the change in consumer characteristics required to create a given change in the vector of choice probabilities.[28]

Lemma 2. *Let Assumptions 1–6 hold. Then for every $x \in \mathcal{X}$ there exists a known $\Delta > 0$ such that for almost all z and z' in $\mathcal{Z}(x)$ satisfying (20), the matrix $\left[\frac{\partial g(z,x)}{\partial z}\right]^{-1}\left[\frac{\partial g(z',x)}{\partial z}\right]$ is identified.*

Proof. Given any $x \in \mathcal{X}$, take a (known) (p, Δ) as in Lemma 1. Consider markets t and t' in which $(P_t, X_t) = (P_{t'}, X_{t'}) = (p, x)$ but, for some choice probability vector $\hat{s}$,

$$z = z_t^*(\hat{s}) \neq z' = z_{t'}^*(\hat{s}), \tag{21}$$

[28] Abusing notation to simplify key expressions, below we write $\frac{\partial g(z,x)}{\partial z}$ to represent the Jacobian matrix $\left.\frac{\partial g(\hat{z},x)}{\partial \hat{z}}\right|_{\hat{z}=z}$. Similarly, we write $\frac{\partial g(z',x)}{\partial z}$ to represent $\left.\frac{\partial g(\hat{z},x)}{\partial \hat{z}}\right|_{\hat{z}=z'}$.

revealing (recall (18)) that $\xi_t \neq \xi_{t'}$. Lemma 1 ensures that such t, t', and $\hat{s}$ exist for all z and z' in $\mathcal{Z}(x)$ satisfying (20). And although ξ_t and $\xi_{t'}$ are latent, the identities of markets t and t' satisfying (21) are observed, as are the associated values of $\hat{s}$, $z_t^*(\hat{s})$, and $z_{t'}^*(\hat{s})$. Differentiating (19) with respect to the choice probability vector within these two markets, we obtain

$$\frac{\partial g\left(z,x\right)}{\partial z}\frac{\partial z_t^*\left(\hat{s}\right)}{\partial s} = \frac{\partial \sigma^{-1}\left(\hat{s};p,x\right)}{\partial s} \tag{22}$$

and

$$\frac{\partial g\left(z',x\right)}{\partial z}\frac{\partial z_{t'}^*\left(\hat{s}\right)}{\partial s} = \frac{\partial \sigma^{-1}\left(\hat{s};p,x\right)}{\partial s}. \tag{23}$$

Thus, recalling Assumption 6, for almost all such (z, z') we have

$$\left[\frac{\partial g\left(z',x\right)}{\partial z}\right]^{-1}\frac{\partial g\left(z,x\right)}{\partial z} = \frac{\partial z_{t'}^*\left(\hat{s}\right)}{\partial s}\left[\frac{\partial z_t^*\left(\hat{s}\right)}{\partial s}\right]^{-1}.$$

The matrices on the right-hand side are observed. $\square$

This leads us to the main result of this section, obtained by connecting (at each x) the matrix products $\left[\frac{\partial g(z,x)}{\partial z}\right]^{-1}\left[\frac{\partial g(z',x)}{\partial z}\right]$ identified in Lemma 2 to the known value of $\left[\frac{\partial g(z,x)}{\partial z}\right]$ at $z = z^0(x)$.

Lemma 3. *Under Assumptions 1–6, $g(\cdot, x)$ is identified on $\mathcal{Z}(x)$ for all $x \in \mathcal{X}$.*

Proof. For $\epsilon > 0$, let $\mathcal{B}\left(b, \epsilon\right)$ denote an open ball in $\mathbb{R}^J$ of radius ϵ, centered at a point b. Take any $x \in \mathcal{X}$ and associated $\Delta > 0$ as in Lemma 2. For each vector of integers $\iota \in \mathbb{Z}^J$, define the set

$$\mathcal{B}_\iota = \mathcal{Z}(x) \cap \mathcal{B}\left(z^0(x) + \frac{\iota\Delta}{2J}, \frac{\Delta}{2}\right).$$

By construction, all z and z' in any given set $\mathcal{B}_\iota$ satisfy (20). So by Lemma 2, the value of $\left[\partial g(z,x)/\partial z\right]^{-1}\left[\partial g(z',x)/\partial z\right]$ is known for almost all z and z' in every set $\mathcal{B}_\iota$. Because $\cup_{\iota\in\mathbb{Z}^J}\mathcal{B}_\iota$ forms an open cover of $\mathcal{Z}(x)$, given any $z \in \mathcal{Z}(x)$ there exists a simple chain of open sets $\mathcal{B}_\iota$ in $\mathcal{Z}(x)$ linking the point $z^0(x)$ to z (see, e.g., van Mill (2002, Lemma 1.5.21)). Thus, $\left[\partial g(z,x)/\partial z\right]^{-1}\left[\partial g(z^0(x),x)/\partial z\right]$ is known for almost all $z \in \mathcal{Z}(x)$. With the normalization (10) and continuity of $\partial g(z,x)/\partial z$ with respect to z, the result then follows from the fundamental theorem of calculus for line integrals and the boundary condition (9). $\square$

The following corollary documents implications of Lemma 3 that play an important role in the next steps of our argument.

Corollary 1. *Let Assumptions 1–6 hold. For all* $(p, x) \in \mathcal{PX}$*,*
(i) $\partial\sigma^{-1}(\cdot; p, x)/\partial s$ *is identified on* $\mathcal{S}(p, x)$*; and*
(ii) for all $s^0 \in \mathcal{S}(p, x)$ *and* $s^1 \in \mathcal{S}(p, x)$*,*

$$\Omega(p, x, s^1, s^0) \equiv \sigma^{-1}(s^1; p, x) - \sigma^{-1}(s^0; p, x) \tag{24}$$

is identified.

Proof. Take any $(p, x) \in \mathcal{PX}$. For every market t in which $(P_t, X_t) = (p, x)$, (18) and (19) imply

$$\frac{\partial g(z_t^*(s), x)}{\partial z}\frac{\partial z_t^*(s)}{\partial s} = \frac{\partial\sigma^{-1}(s; p, x)}{\partial s}$$

for all $s \in \mathcal{S}(p, x, \xi_t)$. For all such s, $\frac{\partial z_t^*(s)}{\partial s}$ is directly observed and, by Lemma 3, $\frac{\partial g(z_t^*(s),x)}{\partial z}$ is known. Part (i) follows. Because $\mathcal{S}(p, x)$ is open and connected (recall footnote 27), part (ii) follows from part (i) and the fundamental theorem of calculus for line integrals. □

4.2 Identification of Conditional Demand

We demonstrate identification of the conditional demand functions $\bar{\mathfrak{z}}(\cdot; t)$ under the additional assumption that there exist instruments W_t for prices satisfying the standard nonparametric IV conditions conditional on X_t.

Assumption 7 (Instruments for Prices)**.**
(i) $\mathbb{E}[h_j(X_t, \Xi_{jt})|X_t, W_t] = \mathbb{E}[h_j(X_t, \Xi_{jt})|X_t]$ *almost surely for all* $j = 1, \ldots, J$*;*
(ii) In the class of functions $\Psi(X_t, P_t)$ *with finite expectation,*
$\mathbb{E}[\Psi(X_t, P_t)|X_t, W_t] = 0$ *almost surely implies* $\Psi(X_t, P_t) = 0$ *almost surely.*

Part (i) of Assumption 7 is the exclusion restriction, requiring that variation in W_t not alter the mean of the latent $h(X_t, \Xi_t)$ conditional on X_t. Recall that $\mathbb{E}[h(X_t, \Xi_t)|X_t] = 0$ by construction; thus part (i) implies

$$\mathbb{E}[h_j(X_t, \Xi_{jt})|X_t, W_t] = 0 \quad \text{a.s. for all } j. \tag{25}$$

This is true regardless of whether X_t itself is exogenous. Of course, one must be particularly cautious about satisfaction of part (i) when X_t is thought to be endogenous. We discuss this further in section 5.3 and Appendix A. The relevance requirement, part (ii), is a standard completeness condition. This is the nonparametric analog of the rank condition required for identification of linear regression models. For example, Newey and Powell (2003) have shown that under mean independence (the analog of (25) here), completeness is necessary and sufficient for identification in separable nonparametric regression.[29] The following result demonstrates that the same instrumental variables conditions suffice here to allow identification of $h_j(x_t, \xi_{jt})$ for all j and t.

Lemma 4. *Under Assumptions 1–7, the scalar $h_j(x_t, \xi_{jt})$ is identified for all j and t.*

Proof. For each $(p, x) \in \mathcal{PX}$, let $s^0(p, x)$ be an arbitrary point in $\mathcal{S}(p, x)$. For each t, let $\tilde{z}_{it}$ be an arbitrary point in $\mathcal{Z}(x_t)$. By (19), in every market t we have

$$g_j\left(\tilde{z}_{it}, x_t\right) = \sigma_j^{-1}\left(s_t(\tilde{z}_{it}); p_t, x_t\right) - h_j(x_t, \xi_{jt}) \tag{26}$$

for each $j = 1, \ldots, J$. Using (24), rewrite the jth equation as

$$\mathscr{y}_{jt} = f_j(x_t, p_t) - e_{jt} \tag{27}$$

where

$$\begin{aligned} \mathscr{y}_{jt} &\equiv g_j\left(\tilde{z}_{it}, x_t\right) - \Omega_j\left(p_t, x_t, s_t\left(\tilde{z}_{it}\right), s^0\left(p_t, x_t\right)\right), \\ f_j(x_t, p_t) &\equiv \sigma_j^{-1}\left(s^0(p_t, x_t); p_t, x_t\right), \\ e_{jt} &\equiv h_j(x_t, \xi_{jt}). \end{aligned}$$

By Lemma 3 and Corollary 1, each $\mathscr{y}_{jt}$ on the LHS of (27) is known. Thus, (27) takes the form of a standard separable nonparametric regression model with regressors (X_t, P_t), an unknown regression function f_j, and additive structural errors E_{jt} (with realizations e_{jt}). By (25), we have $\mathbb{E}[E_{jt}|X_t, W_t] = 0$ almost surely. So under the completeness condition (part (ii) of Assumption 7) identification of f_j follows immediately from Proposition 2.1 in Newey and Powell (2003). This implies identification of each e_{jt} as well. □

Identification of the conditional demand functions $\bar{\mathscr{s}}(\cdot; t)$ now follows easily.

Theorem 1. *Under Assumptions 1–7, the conditional demand system $\bar{\mathscr{s}}(\cdot; t)$ is identified on $\mathcal{C}(x_t, \xi_t)$ for all t.*

Proof. Recall that

$$\begin{aligned} \bar{\mathscr{s}}(Z_{it}, P_t; t) &= \mathscr{s}(Z_{it}, P_t, x_t, \xi_t) \\ &= \sigma(g(Z_{it}, x_t) + h(x_t, \xi_t), P_t, x_t) \\ &= \mathbb{E}\left[Q_{it}|Z_{it}, P_t, x_t, h(x_t, \xi_t)\right]. \end{aligned}$$

For each t, $h(x_t, \xi_t)$ is now known. Q_{it}, Z_{it}, P_t, X_t are observed. □

---

[29] See also, e.g., Florens and Rolin (1990), Chernozhukov and Hansen (2005), and Severini and Tripathi (2006).

4.3 Identification of Demand

As discussed already, knowledge of the conditional demand functions suffices for a large fraction of the questions motivating demand estimation, but not all. In particular, it is not sufficient to answer questions concerning effects of X_t on demand or other counterfactual outcomes when X_t changes holding Ξ_t fixed. Addressing such questions will require separating the impacts of X_t and Ξ_t. This can be done by adding the standard assumption that X_t is exogenous.

Assumption 8 (Exogenous X_t)**.** $\mathbb{E}[\Xi_t|X_t] = 0$.

When Assumption 8 holds, the definition (2) implies

$$h(X_t, \Xi_t) = \Xi_t.$$

This has two important implications. First, when Assumption 8 is maintained, the IV exclusion condition (part (i) of Assumption 7) requires instruments W_t that are exogenous conditional on exogenous (rather than endogenous) X_t. Second, Lemma 4 now implies that each realization ξ_t of the demand shock vector is identified. Recalling that

$$\jmath(C_{it}) = \mathbb{E}\left[Q_{it}|Z_{it}, P_t, X_t, \Xi_t\right],$$

identification of $\jmath$ follows immediately from the facts that $(Q_{it}, Z_{it}, P_t, X_t)$ are observed and all realizations of Ξ_t are now known.

Theorem 2. *Under Assumptions 1–8, ξ_t is identified for all t, and the demand system $\jmath$ is identified on $\mathcal{C}$.*

Note that we have required excluded instruments only for prices. This contrasts with the setting of market-level data in Berry and Haile (2014), where instruments for quantities were also required for identification of demand. This difference is one of the gains from the micro data, and is reflected by the absence of choice probabilities (quantities) on the RHS of the "regression" equation (27)—something accomplished using the adjustment factors $\Omega_j(p_t, x_t, s_t(\tilde{z}_{it}), s^0(p_t, x_t))$ identified in Corollary 1.[30]

[30] In contrast to early drafts of this paper discussed in Berry and Haile (2016, 2021), identification of these adjustment factors avoids any need to assume that the support of Z_{it} is sufficiently large to ensure existence of choice probability vectors that are common to all markets with the same observables. Rather, the arbitrary $s^0(p, x)$ may be used in all markets t for which $P_t = p$ and $X_t = x$, regardless of whether $s^0(p, x) \in \mathcal{S}(p, x, \xi_t)$. Note that although the use of arbitrary $s^0(p, x)$ and $\tilde{z}_{it}$ in the proof of Lemma 4 might suggest overidentification in that step, this is not the case. Let any continuously differentiable functions $g(\cdot, x)$ and $z_t^*(\cdot)$ (for each t) be given. Then, by construction, any change in the choice of $\tilde{z}_{it}$ merely adds and subtracts the same constant from the LHS of (27) in each market t, leaving this regression equation unchanged. Similarly, any change in the choice of $s^0(p, x)$ merely adds the same function of (p_t, x_t) to each side of (27).

5 Lessons for Applied Work

Although the study of identification is formally a theoretical exercise, a primary motivation is to provide guidance for the practice and evaluation of applied work. Here we discuss some key messages.

5.1 The Incremental Value of Micro Data

The most important practical lesson from our results is that the marginal value of micro data is high. The specific benefits of micro data concern some of the most significant challenges to identification of demand when one has only market-level data: (i) the need to instrument for all prices and quantities, and (ii) the nonparametric functional form and exogeneity conditions that allow some of these IV requirements—in particular, the proper excludability of BLP instruments—to hold.

The gains from micro data reflect the fact that consumer-level observables create within-market variation in consumers' choice problems. Such variation is similar in some ways to that which can be generated by instruments for quantities. In particular, it can pin down key aspects of consumer substitution patterns. From (5), we see that $\frac{\partial s_t(z_{it})}{\partial z_{it}} = \frac{\partial \sigma}{\partial \gamma} \frac{\partial g}{\partial z_{it}}$. Since $\frac{\partial s_t(z_{it})}{\partial z_{it}}$ is observed, identification of g (which we demonstrated without instruments) implies identification of the derivatives of demand with respect to the index vector γ (and, thus, with respect to the vector of demand shocks Ξ_t). In standard parametric models like the example of section 3, these substitution patterns—and those with respect to prices—are determined (conditional on observables) by the joint distribution of the random coefficients and product-level taste shocks. Our nonparametric model, of course, allows more flexible substitution patterns, and our results show that micro data allows their identification without any instruments—a stark contrast to the case of market-level data alone.

Critically, however, the reason micro-data variation is free from confounding effects of demand shocks is not an assumption of exogeneity—indeed, Z_{it} was *not* assumed to be independent or mean independent of Ξ_t. Rather, this follows from the fact that market-level demand shocks Ξ_t do not vary within a market. This has a strong connection to the "within" identification of slope parameters in panel data models with fixed effects.

Thus, researchers should prefer micro data and seek it out whenever possible. Collecting reliable micro data will sometimes be difficult, and in some cases only limited forms of micro data may be available. But even when the setting and assumptions permit use of BLP instruments—or when the micro data available are more limited than we have assumed here—variation from micro data can be a powerful addition. This message is consistent, for example, with the findings in the empirical literature (e.g., Petrin (2002), Berry, Levinsohn, and Pakes (2004)) that the addition of even limited forms of individual-level data can result in much more precise estimates than those obtained with market-level data alone.

5.2 The Necessity of Cross-Market Variation

Although within-market variation accounts for the advantages of micro data, cross-market variation remains essential. The proof of Lemma 2, for example, relied on variation in Ξ_t across markets in the key steps toward identification of g and, thus, of the substitution patterns discussed in section 5.1. More fundamentally, the demand system (5) depends on arguments (X_t, P_t, Ξ_t) that have no variation within a market. Thus, without strong additional restrictions, data from a single market cannot reveal anything about the effects of prices or product characteristics on demand.[31]

This observation serves as a caution. As a practical matter, a parametric specification of demand may allow estimation using data from only one market, exploiting a combination of functional form restrictions and cross-product variation in prices and product characteristics.[32] In some cases, as in the classic work of McFadden, Talvitie, and Associates (1977), only a single market is available for study. However, identification in such cases—indeed, the ability to rule out any arbitrary model of how prices (and other product characteristics) affect demand—will be possible only through additional *a priori* restrictions that could be relaxed in a multi-market setting.

5.3 A Focus on Instruments

Given sufficiently rich micro data, our main requirement for identification is a set of valid instruments for prices. Candidate instruments include most of those typically relied upon in the case of market-level data: these include cost shifters, proxies for cost shifters (e.g., "Hausman instruments"), and exogenous

---

[31]For example, with data from a single market τ, one could set $\Xi_\tau = 0$ without loss, assume that X_t has no effect on demand, and assume that demand follows the multinomial logit model with money-metric mean utilities $g_j(z_{it}) - p_{jt}$ for each good j. One can then fit the observed conditional choice probabilities $s_\tau(z_{i\tau})$ in market τ perfectly by setting $g(z_{i\tau}) = \sigma_{MNL}^{-1}(s_\tau(z_{i\tau})) + p_\tau$, where σ_{MNL}^{-1} is the inverse choice probability function for the multinomial logit. The fitted model's implications regarding effects of X_t and P_t on demand, of course, reflect only the *a priori* assumptions, and neither these nor other arbitrary assumptions can be ruled out using data from a single market.

[32] Absent additional restrictions on our nonparametric model, such within-market cross-product variation does not contribute to identification of demand.

shifters of market structure. Micro data can also make available a related category of candidate instruments: market-level observables (e.g., average demographics) that alter equilibrium markups, the "Waldfogel instruments."[33] With micro data, one can directly account for the impacts of individual-specific demographics, so it can be reasonable to assume that market-level demographics are properly excluded from the demand mapping $\mathfrak{s}$.

An important and subtle question is whether the required IV exclusion condition (part (i) of Assumption 7) will hold when X_t is endogenous. In Appendix A we find that, depending on the instrument and model of endogeneity, conditioning on endogenous X_t can (i) render otherwise-valid instruments for prices invalid; (ii) render otherwise-invalid instruments valid; or (iii) have no effect on instrument validity. In some cases where conditioning on endogenous X_t causes a violation of the exclusion condition, the problem can be overcome through natural timing assumptions. While encouraging, this analysis reveals that one must carefully examine the exclusion condition when X_t is endogenous. As Appendix A illustrates, causal graphs offer a tool for such examination that is simultaneously formal, transparent, and robust.

Absent from the discussion above are the BLP instruments. The characteristics X_{-jt} of goods competing with good j have direct effects on demand for good j and (in standard supply models) on good j's markup. The BLP instruments are thus relevant shifters of prices, and in our micro data setting they are not needed as exogenous shifters of quantities. However, the excludability of X_{-jt} requires not only their exogeneity but also a restriction on the way they enter demand (Berry and Haile (2014, 2021)). The Supplemental Appendix shows how adding such restrictions to our model can allow the use of BLP instruments for prices.

5.4 What Does Not Follow

Nonparametric identification results demonstrate a particular sense in which parametric assumptions are not essential. But this does not mean that parametric (or other) assumptions relied on in practice can be ignored. Functional form restrictions can constrain the answers to key questions, and empirical research involving demand estimation should continue to explore sensitivity to functional form choices. Likewise, it remains important to explore new (parametric, semiparametric, or nonparametric) estimation approaches. Our nonparametric identification results ensure that such explorations are possible and may even suggest new estimation strategies.

[33]See Waldfogel (2003) as well as Gentzkow and Shapiro (2010), Fan (2013), and Li, Hartmann, and Amano (2020).

We also emphasize that our sufficient conditions for nonparametric identification should not be viewed as necessary conditions for demand estimation in practice. Rather, they should guide our thinking about the strength of the available data and empirical results. Nonparametric identification of economic models (even regression models) relies on assumptions—index assumptions, separability assumptions, completeness conditions, support conditions, monotonicity conditions, or other shape restrictions—that will often (perhaps typically) fall short of full satisfaction in practice. Conditions for nonparametric identification are not a hurdle but an ideal—a point of reference that can guide our quest for and aid our assessment of the best available empirical evidence. With large choice sets, for example, it will often be difficult to fully satisfy our requirements of J excluded instruments for prices and J-dimensional consumer observables. The robustness analysis of the Supplemental Appendix is aimed to provide some insight about the trade-offs between these requirements and additional structure often employed in practice.

6 Conclusion

Since Berry, Levinsohn, and Pakes (1995), there has been an explosion of interest in empirical demand models that incorporate both flexible substitution patterns and explicit treatment of the demand shocks that must be accounted for to pin down policy-relevant features of demand. Understandably, this development has been accompanied by questions about identification of these models. Our results offer a reassurance that identification can be obtained from traditional sources of quasi-experimental variation in the form of instrumental variables and panel-style within-market variation. This reassurance is particularly important because of the wide relevance of these models to economic questions and the special identification challenges arising in the case of demand (see Berry and Haile (2021)).

Furthermore, identification of these models is not fragile. It does not rely on special regressors or "identification-at-infinity" arguments; it is not limited to particular settings (e.g., random utility discrete choice); one can substitute one type of variation for another (e.g., replacing instruments for quantities with micro-data variation), depending on the type of data available; and one can relax many key conditions by strengthening others. Thus, although this is a case where identification results come well after an extensive empirical literature has already developed, the nonparametric foundation for this literature is strong. Of course, not all applications will offer the combinations of rich micro data and instrumental variables permitting nonparametric identification. But even when such data limitations lead to greater reliance on functional form restrictions, our results shed light on the roles such assumptions will play.

Appendices

A Price Instruments When X_t is Endogenous

In section 5.3 we discussed several categories of instruments W_t commonly relied upon to provide exogenous variation in prices. Here we explore the question of when these instruments remain properly excluded conditional on observables X_t that are not (mean) independent of Ξ_t. Such instruments are needed for Theorem 1 to apply when Theorem 2 does not, allowing identification of conditional demand without exogeneity of X_t (or instruments for X_t). Unconditional independence is not sufficient (or necessary) for conditional independence. Indeed, in other contexts (e.g., linear regression) it is well known that conditioning on an endogenous "control" variable can lead to violation of independence conditions required for identification. In what follows we suppress the market subscripts t on the random variables X_t, P_t, W_t, Ξ_t, etc.

Our discussion will utilize graphical causal models, with the d-separation theorem providing the key criterion for assessing the independence between W and Ξ conditional on X.[34] Our use of these tools is elementary, and a graphical approach is not essential—causal graphs represent information implied by a fully specified economic model, standard change-of-variables formulas, and Bayes' rule. However, the graphical approach allows transparent treatment of many possible economic examples inducing a smaller number of canonical dependence structures. It also can be highly clarifying when one ventures beyond the simplest cases. Following the literature on graphical causal models, we focus on full conditional independence,

$$W \perp\!\!\!\perp \Xi \,|X, \tag{A.1}$$

which implies the conditional mean independence required by Theorem 1.

We first discuss several causal graphs (and motivating economic examples) that "work"—i.e., that imply (A.1). We then discuss the main type of structure that does not work—i.e., where (A.1) fails despite unconditional independence between W and Ξ. We will see that each type of instrument discussed in section 5.3 can remain valid under several models of endogenous X. Each type can also fail—in particular, when firms choose X in ways that depend on both W and Ξ (or their ancestors). However, in many of these situations, a natural timing assumption can yield a new set of valid instruments for prices.

[34] See, e.g., Pearl (2009) and Pearl, Glymour, and Jewell (2016), including references therein. Throughout we maintain the standard assumption that nodes in a causal directed acylic graph are independent of their nondescendants conditional on their parents.

A.1 Graphs that Work

A.1.1 Fully Exogenous Instruments

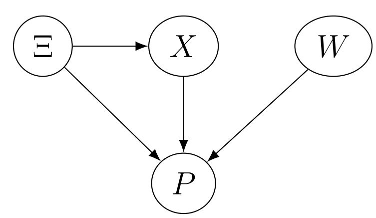
The simplest cases arise when the instruments W satisfy

$$W \perp\!\!\!\perp (X, \Xi). \tag{A.2}$$

The conditional independence condition (A.1) is then immediate, regardless of any dependence between X and Ξ. Although formal analysis is unnecessary in this case, it is also easily illustrated to build toward less obvious cases.

For example, suppose X is chosen by firms with knowledge of Ξ, so that X is endogenous in the same sense that prices are. Given (A.2), one obtains the causal graph shown in Figure 1.[35] The conditional independence condition (A.1) then can also be seen to follow immediately by the d-separation criterion. We would reach the same conclusion if the direction of causation between Ξ and X is reversed—e.g., if X is chosen without knowledge of Ξ but the distribution of Ξ changes with the choice of X. Taking the classic example of demand for automobiles, a manufacturer's choice to offer a fuel efficient hybrid sedan may imply a very different set of relevant unobserved characteristics than had a pickup truck or luxury SUV been offered instead.[36]

Figure 1



Most of the instrument types discussed in section 5.3 can satisfy (A.2). For example, W could represent exogenous (independent of Ξ) cost shifters such as input price shocks, realized after X is chosen and not affected by X. These might be shocks to import tariffs; shipping costs; retailer costs (e.g., rents, wages); demand shifters in other markets served by the same firms (if they face upward sloping marginal costs); or prices of manufacturing inputs. One can also obtain this structure when W represents exogenous shifters of markups. Mergers (full or partial) that are independent of Ξ and leave product

[35] We assume throughout that prices and quantities are not among the ancestors of (X, W, Ξ). This is implied by standard assumptions that consumers take X and Ξ as given, that W does not respond to prices or quantities, and that prices are chosen after X.

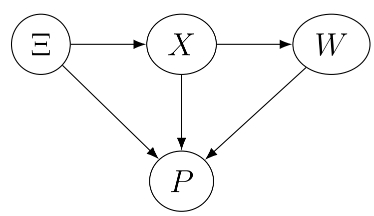
[36] A similar structure is obtained when dependence between Ξ and X reflects a common cause.

offerings unchanged offer one possibility. Another is cross-market variation in the distribution $F_Z(\cdot|t)$ (or other aggregate demographic measure at the market or regional level), as long as this variation is independent of Ξ (as required generally for the validity of Waldfogel instruments) and X.

A.1.2 Instruments Caused by X

Independence between X and W is not required. For example, consider the case in which X is chosen with knowledge of Ξ and the choice of X affects W. We then obtain a causal graph in Figure 2, where (A.1) is again easily confirmed by the d-separation criterion. This causal structure allows additional examples of cost shifters beyond those discussed above. For example, suppose X represents product characteristics affecting the level of labor skill (or quality of another input) required in production, while W is the producer's average wage. Alternatively, if producers have market power in input markets, input prices W would be affected by firms' choices of product characteristics X. Models in which the the direction of causation between X and Ξ in Figure 2 reverses will lead to the same conclusions.

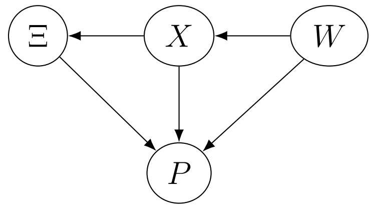
Figure 2



A.1.3 X Caused by Instruments

In some cases, the conditional independence condition (A.1) can hold even when X is affected by W. Consider the causal graph in Figure 3, where (A.1) is easily confirmed by d-separation. As an example motivating this structure, suppose W is a product-level cost of producing a product feature measured by X, X affects the nature of the unobservables Ξ, and X is chosen with knowledge of W but not Ξ.

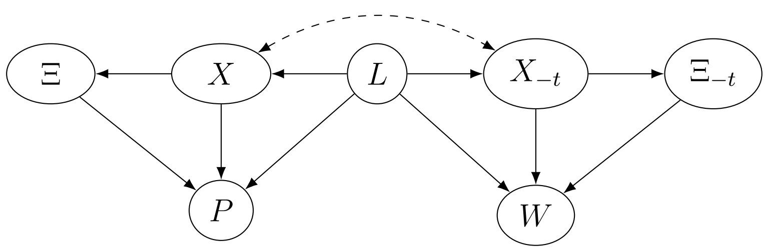
Figure 3



A.1.4 Hausman Instruments

When X is not independent of Ξ, Hausman instruments generally cannot yield the trivial case in which (A.2) holds. This is because prices set by a firm in some "other" market depend on the product characteristics in that market, and the firm's product characteristics are typically (highly) correlated across markets. Nonetheless, Hausman instruments can remain valid in some cases. Figure 4 illustrates one such case. Here L represents latent marginal cost shifters. We let X_{-t} and Ξ_{-t} represent the non-price observables and demand shocks of "other markets," both of which (along with L) affect the Hausman instruments W (prices in those markets).[37] The absence of an edge linking Ξ and Ξ_{-t} reflects an essential assumption justifying Hausman instruments in general (i.e., even when X is exogenous), as does the absence of an edge directly linking L and Ξ. Here W and Ξ are not independent. But the conditional independence condition (A.1) is satisfied.

Figure 4



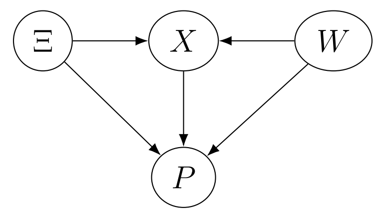
The causal structure of Figure 4 is consistent with a fully specified model in which X is chosen in each market with knowledge of the latent cost shocks but before Ξ is realized. The direction of causality between X and L (and similarly between X_{-t} and L) is not important to this conclusion. However, as demonstrated in the following section, the direction of causality between Ξ and X is typically critical in the case of Hausman instruments.

[37]Following standard convention, we use a dashed bidirectional edge to represent dependence between X and X_{-t} arising from unmodeled common causes.

A.2 Graphs that Don't Work: X is a Collider

The conditional independence condition (A.1) fails when both W and Ξ affect X. This is illustrated in Figure 5. Here X is a collider in the (undirected) path between W and Ξ. Thus, although Ξ and W are independent, (A.1) fails.

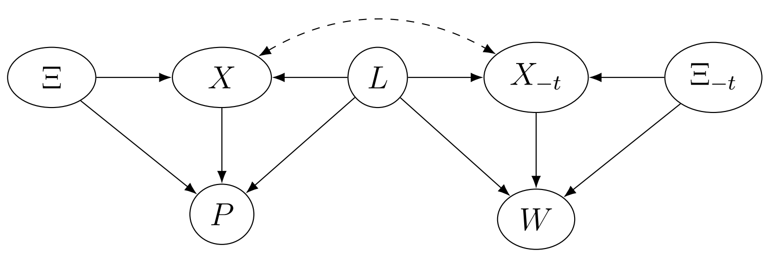
Figure 5



This structure arises when firms' choices of X depend on both W and Ξ, as will be typical if both Ξ and W are known by firms when choosing X. An example is when W is a cost shifter affecting firms' choices of X, the latter also chosen with knowledge of Ξ. Another example is when W is a market-level demographic measure or market structure measure (e.g., product ownership matrix) that, along with Ξ, influences firms' choices of X.[38]

As suggested already, we can also obtain this type of structure with Hausman instruments. Figure 6 illustrates. This graph represents models in which prices and product characteristics X are chosen with knowledge of the demand shocks Ξ and the latent cost shifters L. Here, X is a collider on an unblocked path between Ξ and W.

Figure 6



Thus, just as there are cases in which each type of instrument discussed in section 5.3 remains valid when conditioning on endogenous characteristics X, there are other important cases in which (A.1) will fail. In such situations, identification will require different instruments for prices. In many cases such

[38] A similar structure arises if the dependence between Ξ and X (or X and W) reflects a latent common cause.

instruments can be constructed under natural timing assumptions. This is a topic we take up in the final section of this appendix.[39]

A.3 Avoiding Colliders: Sequential Timing

The previous section describes a class of situations in which candidate instruments that would be properly excluded unconditional on X would fail to be properly excluded conditional on X. A leading case is that of cost shifters (e.g., input prices) that, along with Ξ (or its ancestors), partially determine firms' choices of product characteristics X. In such cases one may be able to obtain valid instruments by exploiting the (typical) sequential timing of a firm's decisions.

For example, physical characteristics of new automobiles sold in year τ will reflect design choices made well in advance—in particular, before the input costs for year-τ production are fully known. Pricing in year τ, on the other hand, will typically take place after those costs are known. Such timing is common to many markets. And, as in other contexts, temporal separation of observable choices can offer an identification strategy.[40] Here, for example, even if product characteristics are chosen in response to demand shocks and expected input costs, current-period *innovations* to input costs can offer candidate instruments for prices.

To illustrate, we introduce a time superscript τ to all random variables. Let M^τ denote a vector of period-τ input prices and suppose M^τ follows

$$M^\tau = \Phi(M^{\tau-1}) + W^\tau, \tag{A.3}$$

where Φ is a possibly unknown function and $W^\tau \perp\!\!\!\perp (\Xi^\tau, X^\tau, M^{\tau-1})$. Given observability of $(M^\tau, M^{\tau-1})$ in all markets, each vector of period-τ innovations W^τ is identified. Now suppose that X^τ is chosen by firms in period $\tau-1$, whereas prices for period τ are chosen in period τ. The causal graph in Figure 7 illustrates key features of such a model.[41]

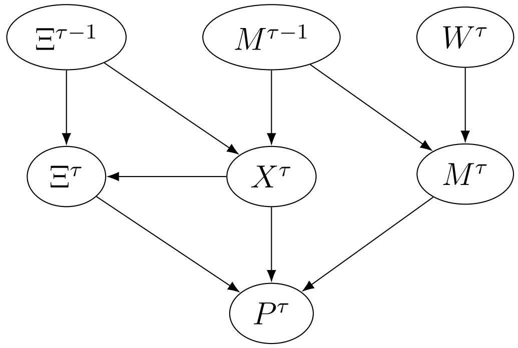
Here endogeneity of X^τ reflects its selection with knowledge of $\Xi^{\tau-1}$, the latter correlated with Ξ^τ. Neither the contemporaneous cost shifters M^τ nor the lagged cost shifters $M^{\tau-1}$ can serve as instruments for prices conditional on X^τ: X^τ would be a collider, as in the previous section. However, the period-τ innovation W^τ can serve as the instrument. Because W^τ alters period-

[39] We also note that when X is a collider, W provides a candidate instrument for X.

[40] Familiar examples in IO include strategies used by Olley and Pakes (1996), Ackerberg, Caves, and Frazer (2015), and others in the literature on estimation of production functions.

[41] The presence (or direction) of an edge from X^τ to Ξ^τ is not important to the argument. Likewise, although we show the case in which $\Xi^{\tau-1}$ is a cause of Ξ^τ, the same conclusion is reached if dependence between $\Xi^{\tau-1}$ and Ξ^τ reflects unmodeled common causes.

Figure 7



τ marginal cost, it is relevant for the determination of P^{τ}, conditional on X^{τ}. And, by the d-separation criterion, we see that W^{τ} is independent of Ξ^{τ} conditional on X^{τ}. Indeed, W^{τ} here is an example of a "fully exogenous instrument," as discussed in section A.1.1. Ultimately, the innovation W^{τ} is simply a cost shifter that is independent of all else. The important insight, however, is that natural timing assumptions can allow such fully independent cost shifters to be constructed from measures like input prices that themselves are not independent of Ξ^{τ} conditional on X^{τ}. Similar arguments can allow construction of valid instruments from observed markup shifters (e.g., market-level demographics) whose lagged values affect firms' choices of X. One may simply reinterpret M^{τ} above as a period-τ markup shifter.

B Proof of Lemma 1

Fix a value of $x \in \mathcal{X}$. By Assumption 5 and the definition (2), there exist $p \in \operatorname{supp} P_t | \{X_t = x\}$ and $\epsilon > 0$ such that for any z and z' in $\mathcal{Z}(x)$ for which

$$||g(z', x) - g(z, x)|| < \epsilon, \tag{B.1}$$

there exist ξ and ξ' in $\operatorname{supp} \Xi_t | \{P_t = p, X_t = x\}$ such that

$$h(x, \xi) - h(x, \xi') = g(z', x) - g(z, x),$$

i.e., $\gamma(z', x, \xi') = \gamma(z, x, \xi)$. Taking

$$s = \sigma(\gamma(z', x, \xi'), p, x) = \sigma(\gamma(z, x, \xi), p, x),$$

the definition (17) implies that

$$z = z^*(s; p, x, \xi) \quad \text{and} \quad z' = z^*(s; p, x, \xi').$$

By uniform continuity of $g(\cdot, x)$, there exists $\Delta > 0$ such that (B.1) holds whenever

$$||z' - z|| < \Delta. \tag{B.2}$$

To see that all such (p, Δ) are identified, recall (18) and observe that (p, Δ) meet the requirement if and only if for all z and z' in $\mathcal{Z}(x)$ that satisfy (B.2), there exist a choice probability vector s and markets t and t' such that

$$z = z_t^*(s) \quad \text{and} \quad z' = z_{t'}^*(s).$$

For any candidate (p, Δ), satisfaction of (B.2) is observable, as are the values of $z_\tau^*(s)$ for all $s \in \mathcal{S}(p, x, \xi_\tau)$ in every market τ such that $(P_\tau, X_\tau) = (p, x)$. $\square$

References

Ackerberg, D. (2003): "Advertising Learning and Customer Choice in Experience Good Markets: A Structural Empirical Examination," *International Economic Review*, 44, 1007–1040.

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," *American Economic Review*, 107, 633–89.

Allen, R. and J. Rehbeck (2019): "Identification with Additively Separable Heterogeneity," *Econometrica*, 87, 1021—1054.

Anderson, S. P., A. de Palma, and B. Kreider (2001): "Tax Incidence in Differentiated Product Oligopoly," *Journal of Public Economics*, 81, 173–192.

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common Ownership and Competition in the Ready-To-Eat Cereal Industry," working paper, NYU-Stern.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89, 2015–2048.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115, 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and Hedonic valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, 58, 1 – 14.

Benkard, L. and S. T. Berry (2006): "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on Brown (1983) and Roehrig (1988)," *Econometrica*, 74, 1429–1440.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 23, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60, 889–917.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T., A. Gandhi, and P. A. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81, 2087–2111.

Berry, S. T. and P. A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Discussion Paper 1718, Cowles Foundation, Yale University.

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

——— (2016): "Identification in Differentiated Products Markets," *Annual Review of Economics*, 8, 27–52.

——— (2018): "Nonparametric Identification of Simultaneous Equations Models with a Residual Index Structure," *Econometrica*, 86, 289–315.

——— (2021): "Foundations of Demand Estimation," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier.

——— (2023): "Nonparametric Identification of Differentiated Products Demand Using Micro Data," Cowles Foundation Discussion Paper 2357, Yale University.

Bhattacharya, D. (2018): "Empirical Welfare Analysis for Discrete Choice: Some General Results," *Quantitative Economics*, 9, 571–615.

Block, H. and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. by I. Olkin, S. Ghurye, W. Hoeffding, W. G. Mado, and H. B. Mann, Stanford University Press, 97–132.

Blundell, R., D. Kristensen, and R. L. Matzkin (2013): "Control Functions and Simultaneous Equations Methods," *American Economic Review, Papers and Proceedings*, 103, 563–569.

——— (2020): "Individual Counterfactuals with Multidimensional Unobserved Heterogeneity," Cemmap working paper cwp60/17, University College, London.

Blundell, R. and R. L. Matzkin (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, 5, 271–295.

Burda, M., M. Harding, and J. Hausman (2015): "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," *Journal of Applied Econometrics*, 30, 353–376.

Capps, C., D. Dranove, and M. Satterthwaite (2003): "Competition and Market Power in Option Demand Markets," *RAND Journal of Economics*, 34, 737–763.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Conlon, C. and J. H. Mortimer (2021): "Empirical Properties of Diversion Ratios," *RAND Journal of Economics*.

Decarolis, F., M. Polyakova, and S. P. Ryan (2020): "Subsidy Design in Privately Provided Social Insurance: Lessons from Medicare Part D," *Journal of Political Economy*, 128, 1712–1752.

Diamond, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000," *The American Economic Review*, 106, 479–524.

Duarte, M., L. Magnolfi, M. Sølvsten, and C. Sullivan (2023): "Testing Firm Conduct," Working paper, UW-Madision.

Falmagne, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the U.S. Daily Newspaper Market," *American Economic Review*, 103, 1598–1628.

Florens, J.-P., M. M. and J. Rolin (1990): *Elements of Bayesian Staistics*, Marcel Decker.

Gale, D. and H. Nikaido (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

Gentzkow, M. and J. Shapiro (2010): "What Drives Media Slant? Evidence from U.S. Newspapers," *Econometrica*, 78, 35–71.

Gerber, A. (1998): "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables," *The American Political Science Review*, 92, 401–411.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951.

Gordon, B. and W. R. Hartmann (2013): "Advertising Effects in Presidential Elections," *Marketing Science*, 32, 19–35.

Hom, M. (2018): "School Choice, Segregation and Access to Quality Schools: Evidence from Arizona," working paper, Yale University.

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145–177.

——— (2014): "An Overview of the Special Regressor Method," in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, ed. by J. S. Racine, L. Su, and A. Ullah, Oxford University Press, 38–62.

Li, X., W. Hartmann, and T. Amano (2020): "Preference Externality Estimators: A Comparison of Border Approaches and IVs," Working paper, Stanford University.

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

Matzkin, R. L. (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

——— (2015): "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1–66.

McFadden, D. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26, 245–264.

McFadden, D., A. Talvitie, and Associates (1977): *Demand Model Estimation and Validation*, Berkeley CA: Institute of Transportation Studies.

Miller, N. and G. Sheu (2021): "Quantitative Methods for Evaluating the Unilateral Effects of Mergers," *Review of Industrial Organization*, 58, 143–177.

Nakamura, E. and D. Zerom (2010): "Accounting for Incomplete Pass-Through," *Review of Economic Studies*, 77, 1192–1230.

Neilson, C. (2021): "Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students," Working paper, Princeton University.

Nevo, A. (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, 31, 395–421.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–42.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation in Nonparametric Models," *Econometrica*, 71, 1565–1578.

Olley, S. G. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–97.

Palais, R. S. (1959): "Natural Operations on Differential Forms," *Transactions of the American Mathematical Society*, 92, 125–141.

Parthasarathy, T. (1983): *On Global Univalence Theorems*, Springer.

Pearl, J. (2009): *Causality*, Cambridge University Press, second ed.

Pearl, J., M. Glymour, and N. Jewell (2016): *Causal Inference in Statistics: A Primer*, Wiley.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Severini, T. A. and G. Tripathi (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory*, 22, 258–278.

van Mill, J. (2002): *The Infinite-Dimensional Topology of Function Spaces*, Elsevier.

Waldfogel, J. (2003): "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets," *RAND Journal of Economics*, 34, 557–568.

Supplemental Appendix:
Extensions, Variations, and Robustness

In this Appendix we explore variations on our baseline model and associated identification conditions. We show that identification is robust in the sense that a relaxation of one condition assumed in the text can often be accommodated by strengthening another. An understanding of such trade-offs can be helpful to both producers and consumers of research relying on demand estimates. Although a full exploration of these trade-offs describes an entire research agenda, we illustrate some possibilities that relax key restrictions of our model, allow demand systems outside discrete choice settings, enlarge the set of potential instruments, reduce the number of required instruments, eliminate the need for continuous consumer-level observables, or reduce the required dimension of those observables. For simplicity, we focus here on the traditional case in which X_t is exogenous, recalling that in this case we have $h(X_t, \Xi_t) = \Xi_t$.

S.1 Prices in the Index

In the text we excluded prices from the index vector $\gamma\left(Z_{it}, X_t, \Xi_t\right)$. That did not rule out interactions between prices and individual-specific measures (e.g., income), but required that such measures be "extra" consumer-level observables beyond those in the J-vector Z_{it}. That requirement can be relaxed. A full investigation of identification in models allowing interactions between Z_{it} and P_t is beyond the scope of this project. However, here we discuss one class of fully nonparametric models permitting such interactions, demonstrating one direction in which our results can be extended.

S.1.1 Model and Normalizations

Suppose demand takes the form

$$\sigma\left(\gamma\left(Z_{it}, P_t, X_t, \Xi_t\right), P_t, X_t\right), \tag{S.3}$$

where, for each $j = 1, \ldots, J$,

$$\gamma_j\left(Z_{it}, P_t, X_t, \Xi_t\right) = g_j\left(Z_{it}, P_{jt}, X_t\right) + \Xi_{jt},$$

$$g_j\left(Z_{it}, P_{jt}, X_t\right) = \bar{g}_j\left(Z_{it}, X_t\right) + \tilde{g}_j(\tilde{Z}_{it}, P_{jt}, X_t), \tag{S.4}$$

and

$$\tilde{Z}_{it} \equiv (Z_{i2t}, \ldots, Z_{iJt}).$$

This specification imposes two restrictions on an otherwise fully-flexible index function $\gamma(Z_{it}, P_t, X_t, \Xi_t)$.[1] First, the price of good j affects only the index $\gamma_j(Z_{it}, P_t, X_t, \Xi_t)$ associated with good j. This is a natural restriction implied by standard specifications (see, e.g., section 3). Second, at least one element of Z_{it} is excluded from interacting with prices. This is an important restriction but also standard in practice. The key implication, exploited below, is that $\frac{\partial g(z,p)}{\partial z_1}$ does not vary with p.

Given the conditions in the text, identification in this model can be obtained under additional verifiable conditions. Here we sketch the argument. With exogenous X_t we can condition on X_t and drop it from the notation.[2]

This model requires a slightly different set of normalizations from those used in the text. Let z^0 denote an arbitrary point in $\mathcal{Z}$ for which $\partial s_t(z^0)/\partial z$ is nonsingular, and let $\tilde{z}^0 = (z_2^0, \ldots, z_J^0)$. Without loss of generality, for each $j = 1, \ldots, j$ we set

$$\begin{aligned} E[\Xi_{jt}] &= 0 && \text{(S.5)} \\ \tilde{g}_j\left(\tilde{z}^0, p_{jt}\right) &= 0 \quad \forall p_{jt} && \text{(S.6)} \\ \bar{g}_j\left(z^0\right) &= 0 && \text{(S.7)} \\ \frac{\partial \bar{g}_j\left(z^0\right)}{\partial z_1} &= 1. && \text{(S.8)} \end{aligned}$$

Equation (S.5) normalizes the location of Ξ_t, as in the text. The need for (S.6) reflects the fact that each price P_{jt} already appears in unrestricted form in the function σ. The role of prices in the index vector is to allow variation in price responses across consumers with different values of $\tilde{Z}_{it}$. Equation (S.6) defines $\tilde{z}^0$ as the (arbitrary) baseline value of $\tilde{Z}_{it}$ around which such variation is defined.[3] Given (S.5) and (S.6), (S.7) normalizes the location of each index γ_j, while (S.8) normalizes its scale (see the related discussion in the text).

---

[1] With no restriction on $\gamma(Z_{it}, P_t, X_t, \Xi_t)$, (S.3) would impose no restriction on demand as function of $(Z_{it}, P_t, X_t, \Xi_t)$, and there would be no role for the function σ in (S.3).

[2] Conditioning on X_t treats it fully flexibly, as the argument presented can be applied at each value of X_t.

[3] More formally, for any demand system $\sigma(\gamma(Z_{it}, P_t, \Xi_t), P_t)$ and any functions $\kappa_j(P_{jt})$ for each j, one can can define an equivalent representation of this demand as $\hat{\sigma}(\hat{\gamma}(Z_{it}, P_t, \Xi_t), P_t)$, where $\hat{\gamma}_j(Z_{it}, P_{jt}, \Xi_{jt}) \equiv \gamma_j(Z_{it}, P_{jt}, \Xi_{jt}) - \kappa_j(P_{jt})$ and $\hat{\sigma}(\hat{\gamma}, P_t) \equiv \sigma(\hat{\gamma} + \kappa(P_t), P_t)$. We select the representation with $\kappa_j(P_{jt}) = \tilde{g}_j\left(\tilde{z}_{it}^0, P_{jt}\right)$ for all j.

S.1.2 Identification: Sketch

Following the arguments in Lemmas 1–3 in the text, one can show that for every price p and any $z' \in \mathcal{Z}$ the value of

$$V\left(z', z^0, p\right) \equiv \left(\frac{\partial g(z^0, p)}{\partial z}\right)^{-1} \left(\frac{\partial g(z', p)}{\partial z}\right) \tag{S.9}$$

is identified, although the two matrices on the RHS are unknown. Thus, (S.9) provides a system of J^2 equations in the $2J^2$ elements of these matrices. Rewrite this system as

$$\frac{\partial g(z^0, p)}{\partial z} V(z', z^0, p) = \frac{\partial g(z', p)}{\partial z}. \tag{S.10}$$

These J^2 equations break naturally into J groups, each with form

$$\underbrace{\frac{\partial g_j(z^0, p)}{\partial z}}_{1\times J} \underbrace{V(z', z^0, p)}_{J\times J} = \underbrace{\frac{\partial g_j(z', p)}{\partial z}}_{1\times J}. \tag{S.11}$$

Take any $j \in \{1, \ldots, J\}$. A key observation is that one obtains a new system of J equations of the form (S.11) at every new price vector. By choosing prices carefully, this can provide new equations without new unknowns. Starting from an arbitrary price vector p and (S.11), any price vector $p^j \neq p$ for which $p^j_j = p_j$ yields

$$\frac{\partial g_j(z^0, p^j)}{\partial z} V(z', z^0, p^j) = \frac{\partial g_j(z', p^j)}{\partial z}. \tag{S.12}$$

Because g_j depends on the price vector only through p_j, we have $\frac{\partial g_j(z^0,p^j)}{\partial z} = \frac{\partial g_j(z^0,p)}{\partial z}$ and $\frac{\partial g_j(z',p^j)}{\partial z} = \frac{\partial g_j(z',p)}{\partial z}$. So we may rewrite (S.12) as

$$\frac{\partial g_j(z^0, p)}{\partial z} V(z', z^0, p^j) = \frac{\partial g_j(z', p)}{\partial z}. \tag{S.13}$$

Subtracting (S.13) from (S.11), we obtain

$$\frac{\partial g_j(z^0, p)}{\partial z} \Lambda\left(z', z^0, p, p^j\right) = 0 \tag{S.14}$$

where

$$\Lambda\left(z', z^0, p, p^j\right) \equiv V(z', z^0, p) - V(z', z^0, p^j). \tag{S.15}$$

Equation (S.14) is a homogeneous system of J linear equations in the J

components of $\frac{\partial g_j(z^0,p)}{\partial z}$. One of these elements is already known: by (S.4) and (S.8), $\frac{\partial g_j(z^0,p)}{\partial z_1} = 1$. One can solve (S.14) for the remaining elements under the following condition on $\Lambda(z', z^0, p, p^j)$.

Condition 1. *For some $k \in \{1, \ldots, J\}$, the submatrix of $\Lambda(z', z^0, p, p^j)$ obtained by dropping row j and column k is full rank.*

Example 1. *Consider the case of $J = 2$ and $j = 1$. Letting $d_{jk} = \partial g_j(z^0, p)/\partial z_k$, (S.14) takes the form (recalling the normalization $d_{11} = 1$)*

$$\begin{bmatrix} 1 & d_{12} \end{bmatrix} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = 0,$$

which we may rewrite as

$$\Lambda_{11} + d_{12}\Lambda_{21} = 0$$
$$\Lambda_{12} + d_{12}\Lambda_{22} = 0.$$

Either of these equations could be used to solve for the unknown d_{12}, although this requires that at least one of Λ_{12} and Λ_{22} be nonzero (Condition 1). If $\Lambda_{22} \neq 0$, for example, then $d_{12} = -\frac{\Lambda_{12}}{\Lambda_{22}}$.

Repeating the argument above for each j identifies the matrix $\frac{\partial g(z^0,p)}{\partial z}$. Plugging this into (S.10) (and recalling that $V(z', z^0, p)$ is known) then allows identification of $\frac{\partial g(z',p)}{\partial z}$ at every z'. Since the normalizations (S.6) and (S.7) imply the boundary condition $g(z^0, p) = 0$, we can integrate $\frac{\partial g(z,p)}{\partial z}$ from this point to identify $g(z, p)$ at all z. Repeating the entire argument at each p then identifies the function g. With g known, the arguments in the text (starting from Corollary 1) can be applied directly to show identification of demand.

Thus, our identification results extend when, in addition to the conditions in the text, for every price vector p and each $j = 1, \ldots, J$, there exists a pair (z', p^j) with $p^j_j = p_j$ and for which Condition 1 is satisfied. Because P_t is observed and Condition 1 is a property of identified objects, this requirement is verifiable.[4] Condition 1 requires that the price vector alter the derivative matrix $\frac{\partial g(z,p)}{\partial z}$, and differentially so at different prices p. This requires nonlinearity: it will fail, for example if g is everywhere linear in z at each p. However, one can confirm numerically that that Condition 1 holds in nonlinear examples—e.g., in specifications following Berry, Levinsohn, and Pakes (1995, 2004). An interesting question is whether there are useful sufficient conditions for Condition 1. That is a topic we leave to future work.

[4]See Berry and Haile (2018) for a formal definition of verifiability.

S.2 Strengthening the Index Structure

The model used by Berry and Haile (2014) to study identification with market-level data restricted the way some elements of X_t enter. Partitioning X_t as $(X_t^{(1)}, X_t^{(2)})$, where $X_t^{(1)} = \left(X_{1t}^{(1)}, \ldots, X_{Jt}^{(1)}\right) \in \mathbb{R}^J$, they assumed that each $X_{jt}^{(1)}$ affects demand only through the jth element of the index vector. This structure is common in specifications used in practice, and adding it here can allow the use of BLP instruments for prices.[5]

To illustrate, suppose demand takes the form

$$\mathcal{s}\left(Z_{it}, P_t, X_t, \Xi_t\right) = \sigma\left(\gamma\left(Z_{it}, X_t, \Xi_t\right), P_t, X_t^{(2)}\right), \tag{S.16}$$

where for each $j = 1, \ldots, J$

$$\gamma_j\left(Z_{it}, X_t, \Xi_t\right) = g_j(Z_{ijt}, X_t^{(2)}) + \eta_j(X_{jt}^{(1)}, X_t^{(2)}) + \Xi_{jt}, \tag{S.17}$$

with $\partial g_j(z, x^{(2)})/\partial z_j > 0$ for all $(z, x^{(2)})$. Compared to the model in the text, this introduces the exclusivity restriction on each $X_{jt}^{(1)}$, associates each Z_{ijt} exclusively with the jth element of the index as well, and imposes separability between Z_{ijt} and $X_{jt}^{(1)}$ within the index.[6] Many specifications in the literature satisfy these requirements, typically with additional restrictions such as linear substitution between Z_{ijt} and $X_{jt}^{(1)}$.

For the remainder of this section we condition on $X_t^{(2)}$ (treating it fully flexibly), suppress it from the notation, and let X_t represent $X_t^{(1)}$. For each $p \in \operatorname{supp} P_t$, define

$$\begin{aligned}
\mathcal{S}(p) &= \bigcup_{x \in \operatorname{supp} X_t | \{P_t = p\}} \mathcal{S}(p, x) \\
\underline{\mathcal{S}}(p) &= \bigcap_{x \in \operatorname{supp} X_t | \{P_t = p\}} \mathcal{S}(p, x) \\
\underline{\mathcal{Z}} &= \bigcap_{x \in \mathcal{X}} \mathcal{Z}(x).
\end{aligned}$$

We assume that $\underline{\mathcal{Z}}$ is nonempty, as is $\underline{\mathcal{S}}(p)$ for all $p \in \operatorname{supp} P_t$. Nonempty $\underline{\mathcal{S}}(p)$ requires that there exist $\underline{s}(p) \in \mathcal{S}(p)$ such that at each $x \in \operatorname{supp} X_t | \{P_t = p\}$

[5] As suggested in section 5.3, the key issue is proper excludability of these instruments, not their relevance.

[6] Exclusivity of $X_{jt}^{(1)}$ to the index γ_j is essential to the point we illustrate here, and this is most natural when exclusivity of each Z_{ijt} differentiates the elements of the index vector. The assumed separability simplifies the analysis.

there is a combination of Z_{it} and Ξ_t in their support conditional on $\{P_t = p, X_t = x\}$ that will map to the choice probability $\underline{s}(p)$. Nonempty $\underline{\mathcal{Z}}$ requires that there exist at least one value of Z_{it} that is present in all markets.

With this more restrictive model we must revisit the necessary normalizations. First, because adding a constant κ_j to g_j and subtracting the same constant from η_j would leave the demand function unchanged, we take an arbitrary $x^0 \in \operatorname{supp} X_t$ and set

$$\eta_j(x_j^0) = 0 \quad \forall j \tag{S.18}$$

without loss. Even with (S.18) (and our maintained $\mathbb{E}[\Xi_t] = 0$), it remains true that linear transformations of each index function γ_j could be offset by an appropriate adjustment to the function σ, yielding multiple representations of the same demand system (recall the related observation in section 2.5). Thus, without loss, we normalize the location and scale of each index by taking an arbitrary $z^0 \in \underline{\mathcal{Z}}$ and setting $g_j\left(z_j^0\right) = 0$ and $\frac{\partial g_j\left(z_j^0\right)}{\partial z_j} = 1$ for all j.

The arguments in Lemmas 1–3 will now demonstrate identification of each function g_j. Likewise, for each price vector p and arbitrary s^0 and s^1 in $\mathcal{S}(p)$, the arguments in Corollary 1 imply identification of

$$\Omega(s^1, s^0, p) \equiv \sigma^{-1}(s^1; p) - \sigma^{-1}(s^0; p).$$

Taking an arbitrary $\tilde{z}_{it} \in \mathcal{Z}(x_t)$ for each market t, the inverted demand system (cf. equation (19)) in each market takes the form

$$g_j\left(\tilde{z}_{ijt}\right) + \eta_j\left(x_{jt}\right) + \xi_{jt} = \sigma_j^{-1}\left(s_t(\tilde{z}_{it}); p_t\right) \quad j = 1, \ldots, J.$$

Taking an arbitrary $s^0(p) \in \mathcal{S}(p)$ at each price vector p, we can rewrite the jth equation as

$$g_j\left(\tilde{z}_{ijt}\right) - \Omega(s_t(\tilde{z}_{it}), s^0(p_t), p_t) = -\eta_j\left(x_{jt}\right) + \sigma_j^{-1}\left(s^0(p_t); p_t\right) - \xi_{jt}. \tag{S.19}$$

Because the LHS is known, this takes the form of a nonparametric regression equation with RHS variables x_{jt} and p_t. In this equation x_{-jt} is excluded, offering $J-1$ potential instruments for the endogenous prices p_t. Thus, one additional instrument—e.g., a scalar market-level cost shifter or Waldfogel instrument—could yield enough instruments to obtain identification of the unknown RHS functions and the "residuals" ξ_{jt}. Once these demand shocks are identified, identification of demand follows immediately.

Many variations on this structure are possible. For example, as in many empirical specifications, one might assume that p_{jt} enters demand only through the j^{th} index. Strengthening the assumption of nonempty $\underline{\mathcal{S}}(p)$ to require nonempty $\bigcap_{(p,x)\in\mathcal{PX}} S(p,x)$, this can lead to a regression equation (the analog of (S.19)) of the form

$$g_j\left(\tilde{z}_{ijt}\right) - \Omega(s_t(\tilde{z}_{it}), s^0) = -\eta_j\left(x_{jt}, p_{jt}\right) + \sigma_j^{-1}\left(s^0\right) - \xi_{jt},$$

where s^0 is an arbitrary point in $\bigcap_{(p,x)\in\mathcal{PX}} S(p,x)$ and the LHS is known. Now only one instrument for price is necessary. For example, the BLP instruments can overidentify demand.

S.3 A Nonparametric Special Regressor

A different approach is to assume that the demand system of interest is generated by a random utility model with conditional indirect utilities of the form

$$U_{ijt} = g_j(Z_{ijt}) + \Xi_{jt} + \mu_{ijt}, \tag{S.20}$$

where μ_{ijt} is a scalar random variable whose nonparametric distribution depends on X_{jt} and P_{jt} (equation (16) gives a parametric example). In this case, our Lemma 3 demonstrates identification of each function $g_j(\cdot)$ up to a normalization of utilities. Under the assumptions of Theorems 1 and 2, conditional demand and demand are identified as the main body of the text.

In the special case of equation (S.20), there is an alternate route to identification. Adding the assumption of independence between the vector μ_{it} and (Ξ_t, Z_{it}) then turns each $g_j(Z_{ijt})$ into a known special regressor. Under a further (and typically very strong) large support assumption on $g(Z_{it})$, standard arguments lead to identification of the marginal distribution of $(\xi_{jt} + \mu_{ijt})|(X_t, P_t)$ for each j in each market t. One can then use these marginal distributions to define a cross-market nonparametric IV regression equation for each choice j, where the LHS is a conditional mean and Ξ_{jt} appears on the RHS as an additive structural error.[7] In each of these equations the prices and characteristics of goods $k \neq j$ are excluded. Identification of the regression functions identifies all demand shocks, and identification of demand then follows as in Theorem 2. Thus, here one needs only one instrument for price, and exogenous characteristics of competing goods (BLP IVs) would be available as instruments.

S.4 Semiparametric Models

The previous example considered Z_{it} with large support. One can instead move in the opposite direction to consider Z_{it} with more limited dimension

[7]See our early working paper, Berry and Haile (2010).

and support than required in the text. We do so here by considering semiparametric specifications of inverse demand that generalize parametric models commonly used in practice. We focus on the case of a one-dimensional binary Z_{it}, taking values 0 and 1.

Consider a semiparametric nested logit model where the inverted demand system in each market takes the form

$$g_j(z_{it})+\xi_{jt} = \ln(s_{jt}(z_{it})/s_{0t}(z_{it}))-\theta\ln(s_{j/n,t}(z_{it}))+\alpha p_{jt} \quad j=1,\ldots,J. \tag{S.21}$$

Here we have conditioned on X_t (treating it fully flexibly) and suppressed it from the notation. On the RHS, $s_{jt}(z_{it})$ denotes good j's (observable) choice probability in market t conditional on z_{it}, with $s_{j/n,t}(z_{it})$ denoting the within-nest conditional choice probability. The scalar θ denotes the usual "nesting parameter."

The nested logit model embeds normalizations of the indices and demand function analogous to our choices of $A(x)$ and $B(x)$ in section 2.5. However, we must still normalize the location of either Ξ_{jt} or g_j for each j to pose the identification question. Here we set $g_j(0)=0$ for all j, breaking with our prior convention by leaving each $\mathbb{E}[\Xi_{jt}]$ free.

Here (S.21) implies the two equations

$$g_j(1)+\xi_{jt} = \ln(s_{jt}(1)/s_{0t}(1))-\theta\ln(s_{j/n,t}(1))+\alpha p_{jt} \tag{S.22}$$
$$g_j(0)+\xi_{jt} = \ln(s_{jt}(0)/s_{0t}(0))-\theta\ln(s_{j/n,t}(0))+\alpha p_{jt} \tag{S.23}$$

for every product j and market t. Differencing these equations in one market, we obtain

$$g_j(1) = \ln\left(\frac{s_{jt}(1)}{s_{0t}(1)}\right)-\ln\left(\frac{s_{jt}(0)}{s_{0t}(0)}\right)-\theta\left[\ln(s_{j/n,t}(1))-\ln(s_{j/n,t}(0))\right] \tag{S.24}$$

for $j=1,\ldots,J$. This is J equations in $J+1$ unknowns: θ and $g_1(1),\ldots,g_J(1)$.

Move now to a different market t', where the observed choice probabilities are different (perhaps because $\xi_t\neq\xi_{t'}$). For this market, (S.24) takes the form

$$g_j(1) = \ln\left(\frac{s_{jt'}(1)}{s_{0t'}(1)}\right)-\ln\left(\frac{s_{jt'}(0)}{s_{0t'}(0)}\right)-\theta\left[\ln(s_{j/n,t'}(1))-\ln(s_{j/n,t'}(0))\right] \tag{S.25}$$

for $j=1,\ldots,J$. This provides J new equations with no new unknowns. Given minimal variation in choice probabilities across markets, ensuring that

$$\ln(s_{j/n,t}(1))-\ln(s_{j/n,t}(0))\neq\ln(s_{j/n,t'}(1))-\ln(s_{j/n,t'}(0)) \tag{S.26}$$

for at least one good j, one can then solve for θ and $g_1(1),\ldots,g_J(1)$. Identification of the remaining parameter α can then be obtained from the "regression" equation (obtained from (S.21))

$$\ln(s_{jt}(z_{it})/s_{0t}(z_{it})) - g_j(z_{it}) - \theta \ln(s_{j/n,t}(z_{it}) = -\alpha p_{jt} - \xi_{jt} \tag{S.27}$$

using a single excluded instrument for price—e.g., an excluded exogenous market-level cost shifter or markup shifter that affects all prices. This compares to the usual requirement of two instruments in the fully parametric nested logit when one has only market-level data (see Berry (1994)). Thus, as in the fully nonparametric case, micro data cuts the number of required instruments by half. The intercept in this regression equation picks up the mean of the unobservable Ξ_{jt}.

Observe that here the argument proceeds in two steps, mirroring those in the main text. We first use a combination of within- and cross-market variation to uncover the function g and consumer substitution patterns (determined here by the parameter θ). We then use cross-market instrumental variables restrictions to separate the roles of prices (and other market-level factors) from the effects of the demand shocks.

Of course, in the first step (S.26) may typically hold for all j, implying overidentification and suggesting the potential to introduce a more flexible specification of the inverse demand mapping—e.g., adding nests or BLP-style random coefficients. Furthermore, there is no reason to limit attention to just two markets in the first step: each additional market adds new equations but no new unknowns.[8] Although this discussion is informal, it suggests the potential to obtain identification of semiparametric demand models with flexible substitution patterns, using (along with instrument(s) for prices) consumer-level characteristics with substantially more limited dimension and support than we required for the fully nonparametric model in the text.

S.5 Beyond Discrete Choice

Although the text emphasizes the case in which the consumer-level quantities Q_{ijt} take the particular form implied by a discrete choice model, nothing in our proofs requires this. In other settings, the demand function $\mathcal{s}$ defined in (1) may simply be reinterpreted as the expected vector of quantities demanded conditional on $(Z_{it}, P_t, X_t, \Xi_t)$.[9] Applying our results to continuous demand is

[8] It is easy to see how the example here generalizes if we allow Z_{it} to have more than two points of support. With K_Z points of support, differencing the analogs of (S.22) and (S.23) for one market yields $(K_Z - 1) \times J$ equations in $1 + (K_Z - 1) \times J$ unknowns. Each new market adds $(K_Z - 1) \times J$ equations and no new unknowns in the first step of the argument.

[9] Note that the demand faced by firms in market t is the expectation (over the distribution of consumer-level observables in the market) of this expected demand.

therefore just a matter of verifying the suitability of our assumptions.[10]

As one possibility, consider a "mixed CES" model of continuous choice, similar to the model in Adao, Costinot, and Donaldson (2017), with $J+1$ products. Here we introduce the notation Y_{it} for the observed income of consumer i in market t, measured in units of the numeraire good 0. For this example we treat Y_{it} as an additional consumer-level observable, beyond the J-dimensional Z_{it} assumed to obey our index restrictions. We again focus on the case in which X_t is exogenous.

Each consumer i in market t has utility over consumption vectors $q \in \mathbb{R}_+^{J+1}$ given by

$$u\left(q; z_{it}, p_t, x_t, \xi_t\right) = \left(\sum_{j=0}^{J} \phi_{ijt} q_j^{\rho}\right)^{1/\rho},$$

where $\rho \in (0,1)$ is a parameter and each ϕ_{ijt} represents idiosyncratic preferences of consumer i. Normalizing $\phi_{i0t} = 1$, let

$$\phi_{ijt} = \exp\left[(1-\rho)\left(g_j\left(z_{it}, x_t\right) + \xi_{jt} + x_{jt}\beta_{it}\right)\right], \; j = 1, \ldots, J,$$

where β_{it} is a random vector (with distribution F_β) representing consumer-level preferences for product characteristics. With $p_{0t} = 1$, familiar CES algebra shows that Marshallian demands are

$$q_{ijt} = \frac{y_{it} \exp\left(g_j\left(z_{it}, x_t\right) + \xi_{jt} + x_{jt}\beta_{it} - \alpha \ln(p_{jt})\right)}{1 + \left[\sum_{k=1}^{J} \exp\left(g_k\left(z_{it}, x_t\right) + \xi_{kt} + x_{kt}\beta_{it} - \alpha\rho \ln(p_{kt})\right)\right]}, \tag{S.28}$$

where $\alpha = 1/(1-\rho)$. It is easy to show that our Assumptions 1–3 are satisfied for the expected demand functions

$$\sigma_t(g(z_{it}, x_t) + \xi_t, y_{it}, x_t, p_t) = \mathbb{E}\left[Q_{it} | z_{it}, y_{it}, p_t, x_t, \xi_t\right],$$

where the jth component of $\mathbb{E}[Q_{it}|z_{it}, y_{it}, x_t, p_t, \xi_t]$ is

$$\int \frac{y_{it} \exp\left(g_j\left(z_{it}, x_t\right) + \xi_{jt} + x_{jt}\beta_{it} - \alpha \ln(p_{jt})\right)}{1 + \left[\sum_{k=1}^{J} \exp\left(g_k\left(z_{it}, x_t\right) + \xi_{kt} + x_{kt}\beta_{it} - \alpha\rho \ln(p_{kt})\right)\right]} \, dF_\beta(\beta_{it}).$$

[10] Berry, Gandhi, and Haile (2013) describe a broad class of continuous choice models that can satisfy the key injectivity property of Assumption 2. These can include mixed continuous/discrete settings, where individual consumers may purchase zero or any positive quantity of each good.