

# Nonparametric Identification of Differentiated Products Demand Using Micro Data\*

Steven T. Berry  
*Yale University*

Philip A. Haile  
*Yale University*

April 13, 2022

## Abstract

We examine identification of differentiated products demand when one has “micro data” linking individual consumers’ characteristics and choices. 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 product characteristics are endogenous. We discuss implications of these results for applied work.

\*Early versions of this work were presented in the working paper “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” first circulated in 2007 and superseded by the present paper. We thank Jesse Shapiro, Suk Joon, Son and numerous seminar participants for helpful comments. Jaewon Lee provided capable research assistance.

# 1 Introduction

Systems of demand for differentiated products are central to many questions in economics. In practice it is common to estimate demand using panel data on the characteristics and choices of many individual consumers within each market. This is often referred to as “micro data,” in contrast to another common case in which only market-level outcomes are observed. 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 the advantages of micro data? What specific kinds of variation within and across markets are helpful, and how?

In this paper we consider nonparametric identification of demand, focusing on the particular benefits of micro data. We consider a nonparametric consumer-level demand model that substantially generalizes parametric models following Berry, Levinsohn, and Pakes (1995) that are widely used in practice. A key benefit of micro data is that unobservables at the level of the product  $\times$  market remain fixed as consumer attributes and choices vary within a given market. This clean “within” variation can be combined with cross-market variation in choice characteristics, market characteristics, prices, and instruments for prices, to yield identification. Compared to settings with market-level data, this can allow both a more general model and substantially reduced demands on instrumental variables.

We focus exclusively on identification. The celebrated “credibility revolution” in applied microeconomics 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 special challenges that are absent in much of empirical economics (Berry and Haile (2021)). Nonetheless, we show that these same types of variation allow identification of demand systems exhibiting rich consumer heterogeneity and endogeneity. Of course, 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. Formal identification results can also clarify which maintained assumptions may be most difficult to relax, reveal the essential sources of variation in the data, offer assurance that robustness analysis is possible, and potentially lead to new (parametric or nonparametric) estimation approaches.

Our results also provide insights that can inform applied practice. Our most important message is that micro data has a high marginal value. 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. With market-level data, nonparametric identification typically requires instruments for all quantities and prices (Berry and Haile (2014, 2016, 2021)). With micro data, we find that the only essential instruments are those for prices. This cuts the number of required instruments in half and avoids the necessary reliance on so-called “BLP instruments” (characteristics of competing products). The availability of micro data also opens the possible use of classes of instruments that are often unavailable in the case of market-level data.

Another important finding for applied work concerns the need for cross-market variation: micro data from a single market does not suffice for identification in our model. We discuss the distinct roles of within-market and cross-market variation, including why the latter is needed. An implication is that existing studies using consumer-level data within just one market may be relying on functional form restrictions that can escape notice as assumptions of necessity rather than convenience. These restrictions may include strong assumptions on how observed consumer attributes change demand.

We also show that it is sometimes possible to identify the *ceteris paribus* effects of prices on quantities demanded (critically, e.g., own- and cross-price demand elasticities) even when observed non-price product characteristics are endogenous and not instrumented.<sup>1</sup> This requires that instruments for prices remain valid when conditioning on the endogenous observed product characteristics, and we illustrate through simple causal graphs how different cases do or do not satisfy this requirement. Potentially 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 cases in which features of demand of primary interest can be identified despite these concerns.

Several key aspects of our model and setting are worth emphasizing to clarify our contributions. First, as in the large empirical literature building on Berry (1994) and Berry, Levinsohn, and Pakes (1995, 2004), we emphasize the role of latent market-level demand shocks (unobservables at the level of the product  $\times$  market) that result in the econometric endogeneity of prices. Explicitly accounting for these demand shocks is essential to the identification of policy-relevant features such as demand elasticities and equilibrium counterfactuals.<sup>2</sup> This drives our focus on market-level endogeneity and differentiates our work from much of the prior research on identification of demand or dis-

---

<sup>1</sup>This is related to well-known results regarding endogenous controls in regression models.

<sup>2</sup>See section 2 of Berry and Haile (2021) for more extensive discussion of this sometimes underappreciated point.

crete choice models.<sup>3</sup>

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, and the combination of within- and cross-market variation is essential to the reduction in instrumental variables needs discussed above. This panel structure also allows us to avoid any restriction on the way product-level observables enter demand. These features contrast with those of Berry and Haile (2014), who previously considered identification from market-level data in nonparametric models generalizing that of Berry, Levinsohn, and Pakes (1995). In that case, the demand system (especially if combined with supply) closely resembles a nonparametric simultaneous equations model, as studied by, e.g., Matzkin (2008, 2015), Benkard and Berry (2006), Blundell, Kristensen, and Matzkin (2013, 2020), and Berry and Haile (2018). However, the panel structure essential to the present paper is absent in that prior work.

Third, our model avoids requiring consumer-level observables that can be linked exclusively to the desirability of specific products. Such a requirement (often in combination with large support conditions) is widely used in “special regressor” approaches to identification of consumer-level discrete choice models,<sup>4</sup> 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.<sup>5</sup> Although voter-specific measures like age, income, gender, race, and education 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, as in the large applied literature following Berry, Levinsohn, and Pakes (1995, 2004),

---

<sup>3</sup>This includes prior work on identification (with micro data) of discrete choice models allowing endogeneity but specifying only a composite “error” term for each choice, representing all unobservables (and their interactions with observables) at the level of the product, consumer, or market. See, e.g., Lewbel (2000).

<sup>4</sup>See the review by Lewbel (2014) and references therein. A very early version of this paper, (Berry and Haile (2010)), featured an example of this sort of identification approach in the panel context considered here. One common idea for an exclusive consumer-choice interaction is the distance to each of  $J$  choices in geographic space. 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.

<sup>5</sup>Advertising, 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).

this is not essential. The primitive feature of interest in our analysis is an “invertible” demand function mapping observables (at the level of market, products, and consumer) and demand shocks (at the level of product  $\times$  market) 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. The second is injectivity of the mappings that link observed consumer attributes to choice probabilities.<sup>6</sup> Below we connect these requirements to the literature through some familiar examples.

Finally, we require sufficient variation in the consumer observables to satisfy a “common choice probability” condition that we believe is new to the literature. Given values of the non-price observables at the product or market level, this condition requires that there exist some point  $s^*$  (unknown *a priori*) in the probability simplex such that in every market one can obtain  $s^*$  as the conditional choice probability vector by conditioning on the “right” set of consumer observables for that market. This requires that the number of observed consumer attributes be at least as large as the number of products, and that they have sufficient independent variation. However, this requirement contrasts with a standard “large support” condition, which would imply that *every* point  $s$  within the simplex is a common choice probability. We require only one such point (which may differ with the non-price product/market observables) and allow a broad range of cases where choice probabilities are never close to one or zero.<sup>7</sup> An attractive feature of the common choice probability condition is that is verifiable; i.e., its satisfaction or failure is identified.

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

---

<sup>6</sup>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.

<sup>7</sup>In the voting example, our condition would require a vote share vector—say 0.4 for R and 0.4 for D (the remainder abstaining)—such that in every market (e.g., metro area) with the same pair of candidates and equal values of any metro-level observables, there is a combination of individual-level sociodemographic measures that generates this conditional vote share. The level of education etc. required to match the given vote share might be higher (and perhaps income lower, etc.) in an unobservably conservative market.

available mode-specific commute times and other factors. This example illustrates the defining characteristic of the type of micro data considered here: consumer-specific observables that alter the relative attractiveness of different options. Consumer distances to different options have been used in a number of applications, including those involving demand for hospitals, retail outlets, residential 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 a household’s 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 examples include product-specific advertising exposure (Akerberg (2003)), consumer-newspaper ideological match (Gentzkow and Shapiro (2010)), and the match between household demographics and those of a school or neighborhood (Bayer, Ferreira, and McMillan (2007), Hom (2018)).<sup>8</sup> An important feature of these 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 parametric examples from the empirical literature. We present our identification results in section 4. In section 5 we discuss potential instruments, variations, and extensions, including models of continuous demand and additional structure that would allow identification when the available micro data or instruments have lower dimension. We discuss other key implications for applied work in section 6 and conclude in section 7. An appendix provides further discussion of instruments, using simple causal graphs.

## 2 Model and Features of Interest

### 2.1 Setup

We consider multinomial choice among  $J$  goods (or “products”) and an outside option (“good 0”) by consumers  $i$  in “markets”  $t$ . A market is defined formally

---

<sup>8</sup>Here we cite only a small representative handful of papers out of a selection that spans many topics and many years. See also the examples in section 3.

by:<sup>9</sup>

- a price vector  $P_t = (P_{1t}, \dots, P_{Jt})$ ;
- a set of additional observables  $X_t$ ;
- a vector  $\Xi_t = (\Xi_{1t}, \dots, \Xi_{Jt})$  of unobservables;<sup>10</sup>
- a distribution  $F_{YZ}(\cdot; t)$  of consumer observables  $(Y_{it}, Z_{it}) \in \mathbb{R}^H \times \mathbb{R}^J$ ,  $H \geq 0$ , with support  $\Omega(X_t)$ ;

The variables  $P_t$ ,  $X_t$ , and  $\Xi_t$  are common to all consumers in a given market. 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  $\Xi_t$ . In fact, we will see below that identification of demand elasticities and other key features of demand can often be obtained without requiring such an assumption (or additional instruments for  $X_t$ ).<sup>11</sup>

Although  $X_t$  will typically include observable product characteristics, it may also include other factors defining markets. For example, consumers might be partitioned into “markets” based on a combination of geography, time, product availability, and average demographics included in  $X_t$ . In contrast, observables varying across consumers within a market are represented by  $Y_{it}$  and  $Z_{it}$ . We make a distinction between  $Y_{it}$  and  $Z_{it}$  in order to isolate our requirements on consumer-level data. Key conditions, made precise below, are that consumer observables have dimension of at least  $J$  (hence,  $Z_{it} \in \mathbb{R}^J$ ) and that changes in  $Z_{it}$  alter the relative attractiveness of different goods. We do not require the additional consumer observables  $Y_{it}$ ; however, we can accommodate them in an unrestricted way, and conditioning on an appropriate value of  $Y_{it}$  can weaken some assumptions.

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.

---

<sup>9</sup>In practice, markets are typically defined by time period or geography. We condition on a fixed number of goods without loss. With additional assumptions, variation in the number of goods available can be valuable and data from markets with  $J$  available goods could be used to predict outcomes in markets with more or fewer goods.

<sup>10</sup>For clarity we write random variables in uppercase and their realizations lowercase. Note that  $\Xi$  is the uppercase form of the standard notation  $\xi$  for unobservables at the product  $\times$  market level.

<sup>11</sup>Alternatively, when instruments are available for endogenous components of  $X_t$ , our results generalize immediately by expanding  $P_t$  to include these endogenous characteristics.

Nor will we require independence (full, conditional, or mean independence) between  $(Y_{it}, Z_{it})$  and  $\Xi_t$ .

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

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

Let  $\mathcal{C}$  denote the support of  $C_{it}$ . The most basic primitive characterizing consumer behavior in this setting is a distribution of decision rules for each  $c_{it} \in \mathcal{C}$ .<sup>12</sup> 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).<sup>13</sup>

## 2.2 Demand and Conditional Demand

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

$$\delta(C_{it}) = (\delta_1(C_{it}), \dots, \delta_J(C_{it})) = E[Q_{it} | C_{it}]. \quad (1)$$

Given the measure of consumers in each choice environment, the mapping  $\delta$  fully characterizes consumer demand. We therefore consider identification of the demand mapping  $\delta$  on  $\mathcal{C}$ .

---

<sup>12</sup>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 of choice behavior. 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 but suggests use of control function methods for identification/estimation of demand. As discussed in Berry and Haile (2021), control function methods are valid only under strong functional form restrictions (Blundell and Matzkin (2014)), which are violated even in standard parametric specifications of demand for differentiated products.

<sup>13</sup>For many purposes, one need not take 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.

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

$$\bar{\delta}(Z_{it}, Y_{it}, P_t; t) = \delta(Z_{it}, Y_{it}, P_t, x_t, \xi_t)$$

on

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

for each market  $t$ . The function  $\bar{\delta}(Z_{it}, Y_{it}, P_t; t)$  is simply the demand function  $\delta$  when  $(X_t, \Xi_t)$  are fixed at the values  $(x_t, \xi_t)$  realized in market  $t$ . Because  $\Xi_t$  is unobserved and prices are fixed within each market, identification of  $\bar{\delta}(Z_{it}, Y_{it}, P_t; t)$  is nontrivial. However, this mapping fully characterizes the responses of demand (at all combinations of  $(Z_{it}, Y_{it})$ ) to counterfactual *ceteris paribus* price variation, holding  $X_t$  and  $\Xi_t$  fixed at their realized values in market  $t$ . Thus, knowledge  $\bar{\delta}(\cdot; t)$  for each market  $t$  suffices for many purposes motivating demand estimation in practice.

Notably,  $\bar{\delta}(\cdot; t)$  fully determines the own- and cross-price demand elasticities for all goods in market  $t$ . One implication is that  $\bar{\delta}(\cdot; t)$  is the feature of  $\delta$  needed to discriminate between alternative models of firm competition (e.g., Berry and Haile (2014), Backus, Conlon, and Sinkinson (2021), Duarte, Magnolfi, Sølvesten, and Sullivan (2021)). And, given an assumed model of supply,  $\bar{\delta}(\cdot; t)$  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 the equilibrium “unilateral effects” of a merger (e.g., Nevo (2000a), Miller and Sheu (2021)). Furthermore,  $\bar{\delta}(\cdot; t)$  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  $\bar{\delta}(\cdot; t)$  are defined with fixed values of  $(X_t, \Xi_t)$ , they do not suffice for answering all questions—in particular, those requiring knowledge of *ceteris paribus* effects of  $X_t$  on demand.<sup>14</sup> However, by avoiding the need to separate the effects of  $X_t$  and  $\Xi_t$  on demand, identification of  $\bar{\delta}(\cdot; t)$  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.

---

<sup>14</sup>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  $\bar{\delta}(\cdot; t)$  in all markets suffices in a large fraction of applications, knowledge of  $\delta$  is required for others.

## 2.3 Core Assumptions

So far we have implicitly made two significant assumptions: (i) unobservables at the market level can be represented by a  $J$ -vector  $\Xi_t$ ; (ii) conditional on  $X_t$ , the support of  $(Y_{it}, Z_{it})$  is the same in all markets. The first is standard but important. The second seems mild for many applications and can be relaxed at the cost of more cumbersome exposition. Our results will also rely on the following key structure.

**Assumption 1** (Index).  $s(C_{it}) = \sigma(\gamma(Z_{it}, Y_{it}, X_t, \Xi_t), Y_{it}, P_t, X_t)$ , with  $\gamma(Z_{it}, Y_{it}, X_t, \Xi_t) = (\gamma_1(Z_{it}, Y_{it}, X_t, \Xi_t), \dots, \gamma_J(Z_{it}, Y_{it}, X_t, \Xi_t)) \in \mathbb{R}^J$ .

**Assumption 2** (Invertible Demand).  $\sigma(\cdot, Y_{it}, P_t, X_t)$  is injective on the support of  $\gamma(Z_{it}, Y_{it}, X_t, \Xi_t) | (Y_{it}, P_t, X_t)$ .

**Assumption 3** (Injective Index).  $\gamma(\cdot, Y_{it}, X_t, \Xi_t)$  is injective on the support of  $Z_{it} | (Y_{it}, X_t)$ .

**Assumption 4** (Separable Index).  $\gamma_j(Z_{it}, Y_{it}, X_t, \Xi_t) = \Gamma_j(Z_{it}, Y_{it}, X_t) + \Xi_{jt}$  for all  $j$ .

Assumption 1 requires that, given  $(Y_{it}, P_t, X_t)$ ,  $Z_{it}$  and  $\Xi_t$  affect choices only through indices  $(\gamma_1(Z_{it}, Y_{it}, X_t, \Xi_t), \dots, \gamma_J(Z_{it}, Y_{it}, X_t, \Xi_t))$  that exclude  $P_t$ . This is a type of weak separability assumption. Observe that  $X_t$  and  $Y_{it}$  can affect demand both directly and through the indices, and that the indices themselves enter the function  $\sigma$  in fully flexible form. As we illustrate below, this index structure generalizes standard specifications used in practice. Assumption 2 further requires that the choice probability function  $\sigma$  be “invertible” with respect to the index vector—that, holding  $(Y_{it}, P_t, X_t)$  fixed, distinct index vectors map to distinct choice probabilities. This is not without loss, and in general injectivity requires that  $\sigma$  map to interior values, i.e., that  $\sigma_j(C_{it}) > 0$  for all  $j$  and  $C_{it} \in \mathcal{C}$ . Berry, Gandhi, and Haile (2013) provide sufficient conditions for invertibility and point out that these are natural in discrete choice settings when each  $\gamma_j(Z_{it}, Y_{it}, X_t, \Xi_t)$  can be interpreted as a (here, consumer-specific) quality index for good  $j$ . Assumption 3 requires injectivity of the index function  $\gamma$  with respect to the vector  $Z_{it}$ . This generalizes common utility-based specifications in which each  $Z_{ijt}$  is assumed to affect only the conditional indirect utility of good  $j$  and to do so monotonically. For example, if each index were a linear function of the  $J$  components of  $Z_{it}$ , Assumption 3 would require the matrix of coefficients to be full rank.<sup>15</sup> Assumption 4 requires the indices  $\gamma_j(Z_{it}, Y_{it}, X_t, \Xi_t)$  to take an

---

<sup>15</sup>General sufficient conditions for injectivity can be found in, e.g., Palais (1959), Gale and Nikaido (1965), Parthasarathy (1983), and Berry, Gandhi, and Haile (2013).

additively separable structure.<sup>16</sup>

## 2.4 A Useful Representation of the Index

Although the formulation of our index vector  $\gamma(Z_{it}, Y_{it}, X_t, \Xi_t)$  above maximizes clarity about our core assumptions, for the study of identification it will be convenient to define

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

and

$$h(X_t, \Xi_t) = \Xi_t - E[\Xi_t | X_t], \quad (2)$$

so that

$$\gamma(Z_{it}, Y_{it}, X_t, \Xi_t) = g(Z_{it}, Y_{it}, X_t) + h(X_t, \Xi_t). \quad (3)$$

Observe that

$$E[h(X_t, \Xi_t) | X_t] = 0 \quad (4)$$

by construction.

With this notation, we have

$$s(Z_{it}, Y_{it}, P_t, X_t, \Xi_t) = \sigma(g(Z_{it}, Y_{it}, X_t) + h(X_t, \Xi_t), Y_{it}, P_t, X_t) \quad (5)$$

and

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

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

## 2.5 Technical Conditions

Let  $\mathcal{X}$  denote the support of  $X_t$ . For  $x \in \mathcal{X}$ , let  $\mathcal{Y}(x)$  denote the support of  $Y_{it} | \{X_t = x\}$  and, for  $y \in \mathcal{Y}(x)$ , let  $\mathcal{Z}(y, x) \subset \mathbb{R}^J$  denote the support of  $Z_{it} | \{Y_{it} = y, X_t = x\}$ . In parts (i)–(iii) of Assumption 5 we assume conditions permitting our applications of calculus and continuity arguments below. Part (iv) strengthens the injectivity requirements of Assumptions 2 and 3 slightly

---

<sup>16</sup>One key role of additive separability here is to allow use of standard IV conditions—those needed for identification of separable nonparametric regression models—where instruments are required. In the context of market-level data, Berry and Haile (2014) include results that allow relaxation of additive separability by strengthening IV conditions or other assumptions. See also Matzkin (2015) and Blundell, Kristensen, and Matzkin (2020). None of these covers the panel structure (consumers within markets) of the micro data setting we consider here.

by requiring that the Jacobian matrices  $\partial g(z, y, x)/\partial z$  and  $\partial \sigma(\gamma, y, p, x)/\partial \gamma$  be nonsingular almost surely.<sup>17</sup>

**Assumption 5** (Technical Conditions). *For all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}(x)$ ,*

- (i)  $\mathcal{Z}(y, x)$  is open and connected;
- (ii)  $g(z, y, x)$  is uniformly continuous in  $z$  on  $\mathcal{Z}(y, x)$  and continuously differentiable with respect to  $z$  on  $\mathcal{Z}(y, x)$ ;
- (iii)  $\sigma(\gamma, y, p, x)$  is continuously differentiable with respect to  $\gamma$  for all  $(\gamma, p) \in \text{supp}(\gamma(Z_{it}, Y_{it}, X_t, \Xi_t), P_t) | \{Y_{it} = y, X_t = x\}$ ; and
- (iv)  $\partial g(z, y, x)/\partial z$  and  $\partial \sigma(\gamma, y, p, x)/\partial \gamma$  are nonsingular almost surely on  $\mathcal{Z}(y, x)$  and  $\text{supp}(\gamma(Z_{it}, Y_{it}, X_t, \Xi_t), Y_{it}, P_t, X_t) | \{Y_{it} = y, X_t = x\}$ , respectively.

## 2.6 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  $E[\Xi_t] = 0$  without loss. The second reflects the fact that any injective transformation of the index vector  $\gamma(Z_{it}, Y_{it}, X_t, \Xi_t)$  can be reversed by appropriate modification of the function  $\sigma$ . For example, take arbitrary  $A(X_t) : \mathcal{X} \rightarrow \mathbb{R}^J$  and  $B(X_t) : \mathcal{X} \rightarrow \mathbb{R}^{J \times J}$  ( $B(x)$  invertible at all  $x$ ). By letting

$$\begin{aligned} \tilde{\gamma}(Z_{it}, Y_{it}, X_t, \Xi_t) &= A(X_t) + B(X_t)\gamma(Z_{it}, Y_{it}, X_t, \Xi_t) \\ \tilde{\sigma}(\tilde{\gamma}(Z_{it}, Y_{it}, X_t, \Xi_t), Y_{it}, P_t, X_t) &= \sigma(B(X_t)^{-1}(\tilde{\gamma}(Z_{it}, Y_{it}, X_t, \Xi_t) - A(X_t)), Y_{it}, P_t, X_t) \end{aligned}$$

one obtains a new representation of the same distribution of decision rules (and thus same demand), 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.<sup>18</sup>

---

<sup>17</sup>Although we state Assumption 5 with the quantifier “for all  $y \in \mathcal{Y}(x)$ ,” our arguments require these properties only at the arbitrary point  $y^0(x)$  selected below for each  $x \in \mathcal{X}$ . Given parts (i) and (ii), the injectivity of  $g(\cdot, Y_{it}, X_t)$  required by Assumption 3 implies (by invariance of domain) that the image  $g(\mathcal{O}, y, x)$  of any open set  $\mathcal{O} \subseteq \mathcal{Z}(y, x)$  is open. An implication is that even without part (iv) there could be no nonempty open set  $\mathcal{O} \in \mathcal{Z}(y, x)$  on which  $\partial g(z, y, x)/\partial z$  was singular, as  $g(\mathcal{O}, y, x)$  would then be a nonempty open subset of  $\mathbb{R}^J$ , contradicting Sard’s theorem. A similar observation applies to  $\partial \sigma(\gamma, y, p, x)/\partial \gamma$ .

<sup>18</sup>Like location and scale normalizations of utility functions, our normalizations place no restriction on the demand function  $s$  or the conditional demand functions  $\bar{s}(\cdot; t)$ . However, our example illustrates an inherent ambiguity in the interpretation of how a given variable alters preferences. For example, in terms of consumer behavior (e.g., demand), there is no difference between a change in  $Z_{ijt}$  (all else fixed) that makes good  $j$  more desirable and a change in  $Z_{ijt}$  that makes all other goods (including the outside good) less desirable. In practice, this ambiguity is often resolved with *a priori* exclusion assumptions—e.g., an assumption that  $Z_{ijt}$  affects only the utility obtained from good  $j$ . Such assumptions could only aid identification. See, for example, section 5.2.1 below.

To do this, for each  $x$  we take an arbitrary  $(z^0(x), y^0(x))$  from the support of  $(Z_{it}, Y_{it}) | \{X_t = x\}$ . We then select the representation of demand in which

$$E [\gamma (z^0 (X_t), y^0 (X_t), X_t, \Xi_t) | X_t = x] = 0 \quad \forall x \quad (7)$$

and

$$\left[ \frac{\partial g (z^0 (x), y^0 (x), x)}{\partial z} \right] = I \quad \forall x, \quad (8)$$

where  $I$  denotes the  $J$ -dimensional identity matrix. Observe that (3), (4), and (7) together imply

$$g (z^0 (x), y^0 (x), x) = 0 \quad \forall x. \quad (9)$$

In the example above this choice of normalization is equivalent to taking

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

and

$$A(x) = -B(x)g (z^0(x), y^0(x), x)$$

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

### 3 Parametric Examples from the Literature

The empirical literature includes many examples of parametric specifications that are special cases of our model. Discrete choice demand models are frequently formulated using a random coefficients random utility specification such as

$$u_{ijt} = x_{jt}\beta_{ijt} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}, \quad (10)$$

where  $u_{ijt}$  represents individual  $i$ 's conditional indirect utility from choice  $j$  in market  $t$ . The additive  $\epsilon_{ijt}$  is typically specified as a draw from a type-1 extreme value or normal distribution, yielding a mixed multinomial logit or probit model. Components  $k$  of the random coefficient vector  $\beta_{ijt}$  are often specified as

$$\beta_{ijt}^{(k)} = \beta_{0j}^{(k)} + \sum_{\ell=1}^L \beta_{zj}^{(k,\ell)} z_{ilt} + \beta_{\nu j}^{(k)} \nu_{it}^{(k)}, \quad (11)$$

where each  $z_{ilt}$  represents an observable characteristic of individual  $i$ , and each  $\nu_{it}^{(k)}$  is a random variable with a pre-specified distribution. Often, the coefficient on price is also specified as varying with some observed consumer characteristics  $y_{it}$ , such as income. A typical specification of  $\alpha_{it}$  takes the form

$$\ln(\alpha_{it}) = \alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}. \quad (12)$$

With (11) and (12), we can rewrite (10) as

$$u_{ijt} = g_j(z_{it}, x_t) + \xi_{jt} + \mu_{ijt}, \quad (13)$$

where

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

$$\mu_{ijt} = \sum_k x_{jt}^{(k)} \left( \beta_{0j}^{(k)} + \beta_{\nu_j}^{(k)} \nu_{it}^{(k)} \right) - p_{jt} \exp(\alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}) + \epsilon_{ijt}. \quad (14)$$

Observe that all effects of  $z_{it}$  and  $\xi_t$  operate through indices

$$\gamma_j(z_{it}, x_t, \xi_t) = g_j(z_{it}, x_t) + \xi_{jt} \quad j = 1, \dots, J,$$

satisfying our Assumptions 1 and 4. It is easy to show that the resulting choice probabilities satisfy Berry, Gandhi and Haile’s (2013) “connected substitutes” condition with respect to the vector of indices  $(\gamma_1(z_{it}, x_t, \xi_t), \dots, \gamma_J(z_{it}, x_t, \xi_t))$ ; therefore, the injectivity of demand required by Assumption 2 holds. Our assumptions require  $L \geq J$ .<sup>19</sup> Injectivity of  $g(z_{it}, x_t) = (g_1(z_{it}, x_t), \dots, g_J(z_{it}, x_t))$  in  $z_{it}$  (Assumption 3) might then be assumed as a primitive condition of the model or derived from other conditions, as in the example we discuss below.<sup>20</sup>

Of course, our model does not rely on the linear structure of this example, on any parametric distributional assumptions, or on a representation of demand through random utility maximization. But this example connects our model to a large number of applications and shows one way that the individual-level observables  $z_{it}$  can interact with product characteristics to generate preference heterogeneity across consumers facing the same choice set (i.e., where all  $x_{jt}, p_{jt}$  and  $\xi_{jt}$  are fixed). Note that this standard specification lacks features sometimes relied on in results showing identification of discrete choice models: in addition to 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 (13) and any of the observables  $z_{it}, x_t, p_t$ .<sup>21</sup>

---

<sup>19</sup>If  $L > J$ , we can combine the “extra” components of  $Z_{it}$  with income to redefine the partition of consumer observables as  $(Y_{it}, Z_{it})$  with  $Z_{it} \in \mathbb{R}^J$ . More generally, income and any extra components of  $Z_{it}$  may affect both the index (reintroducing  $Y_{it}$  as an argument of  $g$ ) and the coefficients on  $(X_t, P_t)$ .

<sup>20</sup>Although we discuss only the core assumptions, the technical conditions of Assumption 5 can be confirmed for these examples as long as  $\text{supp } Z_{it} | \{Y_{it}, X_t\}$  is open and connected.

<sup>21</sup>This is true even without the demand shocks  $\xi_{jt}$ . For example, the individual “taste shock” vectors  $\nu_{it}$  and  $\epsilon_{it}$  are typically assumed independent across  $i$  and  $t$ ; however,  $x_{jt}$  and  $p_{jt}$  enter the composite error  $\mu_{ijt}$ . Likewise,  $z_{it}$  may be correlated with  $y_{it}$ , which enters  $\mu_{ijt}$ . Furthermore,  $x_{jt}$  and  $p_{jt}$  may be correlated with changes in the distribution of  $z_{it}$  across markets, introducing variation in this distribution with  $\mu_{ijt}$

To see another way that our index structure arises in practice, consider Ho’s (2009) model of demand for health insurance. Each consumer  $i$  in market  $t$  considers  $J$  insurance plans as well as the outside option of remaining uninsured. Each consumer has a vector of observable characteristics  $d_{it}$  (used below to define  $z_{it}$ ).<sup>22</sup> Let  $n_{jt}$  denote the set of hospitals in plan  $j$ ’s network, along with their characteristics (e.g., location and the availability of speciality services like cardiac care). Each insurance plan is associated with its network  $n_{jt}$ , an annual premium  $p_{jt}$ , additional observed plan characteristics  $x_{jt}$  (e.g., the size of its physician network), and an unobservable  $\xi_{jt}$ .

A consumer’s insurance plan demand depends on her particular likelihood of having of each type of hospital need (diagnosis), as well as how her preferences over hospital characteristics will vary with the type of need. This gives each consumer  $i$  an expected utility  $EU(n_{jt}, d_{it})$  for the option to use plan  $j$ ’s hospital network. Ho derives this expected utility from auxiliary data on hospital choice (see also Ho (2006) and Ho and Lee (2016)). From the perspective of identification, this yields a known functional form for the consumer-specific measures

$$z_{ijt} \equiv EU(n_{jt}, d_{it})$$

entering consumer  $i$ ’s conditional indirect utilities

$$u_{ijt} = \lambda z_{ijt} + x_{jt}\beta - \alpha(y_{it})p_{jt} + \xi_{jt} + \epsilon_{ijt} \quad (15)$$

for each plan  $j$ . Here  $y_{it} \in d_{it}$  represents the consumer’s income. Ho assumes each  $\epsilon_{ijt}$  is an independent draw from a type-1 extreme value distribution, yielding a multinomial logit model.

Observe that in this example Ho combines data on the characteristics of consumers and choices with additional modeling to derive a scalar  $z_{ijt}$  that exclusively affects only the utility of choice  $j$ .<sup>23</sup> In this case, the injectivity of the index vector  $\gamma(z_{it}, \xi_t)$  required by our Assumption 3 holds as long as  $\lambda \neq 0$ . Satisfaction of our remaining assumptions follows as in the previous example.

## 4 Identification

We consider identification of the demand system

$$\mathcal{D}(Z_{it}, Y_{it}, P_t, X_t, \Xi_t)$$

---

<sup>22</sup>Ho’s data include measures of individual age, gender, income, home location, employment status, and industry of employment.

<sup>23</sup>Our model would also allow the possibility that the function  $EU$  here is not learned from auxiliary data; in that case our  $g_j(z_{it}, y_{it}, x_t)$  would play the role of  $EU(n_{jt}, d_{it})$ , with  $z_{it} = d_{it}$  and  $n_{jt} \subset x_{jt}$ .

and the conditional demand systems

$$\bar{s}(Z_{it}, Y_{it}, P_t; t)$$

from observation of the choice decisions of the population of consumers  $i$  in a population of markets  $t$ . The observables comprise  $Z_{it}, Y_{it}, P_t, X_t, Q_{it}$ , and a vector of instruments  $W_t$  discussed below. These observables imply observability of choice probabilities conditional on  $(Z_{it}, Y_{it}, P_t, X_t)$  in each market  $t$ .

Because our arguments do not require variation in  $Y_{it}$ , in much of what follows we will fix  $Y_{it}$  (conditional on  $X_t$ ) at  $y^0(X_{it})$ . We proceed in three steps. First, in section 4.1 we present lemmas demonstrating identification of the function  $g(\cdot, y^0(x), x)$  at each  $x \in \mathcal{X}$ . Second, 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 a given conditional choice probability 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  $\bar{s}(\cdot; t)$  in each market straightforward. Finally, in section 4.3 we show that  $s$  is 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

$$\mathcal{S}(p, x, \xi) = \sigma(g(\mathcal{Z}(y^0(x), x), y^0(x), x) + h(x, \xi), y^0(x), p, x)). \quad (16)$$

Thus,  $\mathcal{S}(p, x, \xi)$  denotes the support of choice probabilities in any market  $t$  for which  $P_t = p, X_t = x$ , and  $\Xi_t = \xi$  (holding  $Y_{it} = y^0(x)$ ).<sup>24</sup> By Assumptions 2 and 3, for each  $s \in \mathcal{S}(p, x, \xi)$  there must be a unique  $z^* \in \mathcal{Z}(y^0(x), x)$  such that  $\sigma(g(z^*, y^0(x), x) + h(x, \xi), y^0(x), p, x) = s$ . So for  $(p, x, \xi) \in \text{supp}(P_t, X_t, \Xi_t)$  and  $s \in \mathcal{S}(p, x, \xi)$ , we define the function

$$z^*(s; p, x, \xi)$$

implicitly by

$$\sigma(g(z^*(s; p, x, \xi), y^0(x), x) + h(x, \xi), y^0(x), p, x) = s. \quad (17)$$

This definition leads 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  $z^*(s; p_t, x_t, \xi_t)$  for

---

<sup>24</sup>Because  $\mathcal{Z}(y^0(x), x)$  is open, continuity and injectivity of  $\sigma$  with respect to the index and of the index with respect to  $Z_{it}$  imply (by invariance of domain) that  $\mathcal{S}(p, x, \xi)$  is open.

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

$$g(z^*(s; p, x, \xi), y^0(x), x) + h(x, \xi) = \sigma^{-1}(s; y^0(x), p, x) \quad (18)$$

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

## 4.1 Key Lemmas

Let  $\|\cdot\|$  denote the Euclidean norm. We will require the following nondegeneracy condition.

**Assumption 6** (Nondegeneracy). *For each  $x \in \mathcal{X}$ , there exists  $p \in \text{supp} P_t | \{X_t = x\}$  such that  $\text{supp} \Xi_t | \{P_t = p, X_t = x\}$  contains an open subset of  $\mathbb{R}^J$ .*

Assumption 6 requires continuously distributed  $\Xi_t$  but is otherwise mild. It rules out trivial cases in which conditioning on  $(P_t, X_t)$  indirectly fixes  $\Xi_t$  as well. This nondegeneracy condition is implied 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  $\Xi_t$ . A key implication, exploited to prove Lemma 1 below, follows from the definition (2): for each  $x \in \mathcal{X}$  there exist  $\epsilon > 0$  and  $p \in \text{supp} P_t | \{X_t = x\}$  such that for any  $d \in \mathbb{R}^J$  satisfying  $\|d\| < \epsilon$ ,  $\text{supp} \Xi_t | \{P_t = p, X_t = x\}$  contains vectors  $\xi$  and  $\xi'$  satisfying  $h(x, \xi) - h(x, \xi') = d$ .

**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  $\mathcal{Z}(y^0(x), x)$  satisfying  $\|z' - z\| < \Delta$ , there exist a choice probability vector  $s$  and vectors  $\xi$  and  $\xi'$  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  $(\Delta, p)$  are identified.*

*Proof.* Fix a value of  $x \in \mathcal{X}$ . By Assumption 6, there exist  $p \in \text{supp} P_t | \{X_t = x\}$  and  $\epsilon > 0$  such that for any  $z$  and  $z'$  in  $\mathcal{Z}(y^0(x), x)$  for which

$$\|g(z', y^0(x), x) - g(z, y^0(x), x)\| < \epsilon, \quad (19)$$

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

$$h(x, \xi) - h(x, \xi') = g(z', y^0(x), x) - g(z, y^0(x), x),$$

i.e.,  $\gamma(z', y^0(x), x, \xi') = \gamma(z, y^0(x), x, \xi)$ . Taking

$$s = \sigma(\gamma(z', y^0(x), x, \xi'), y^0(x), p, x) = \sigma(\gamma(z, y^0(x), x, \xi), y^0(x), p, x),$$

the definition (17) implies that

$$z = z^*(s; p, x, \xi) \quad \text{and} \quad z' = z^*(s; p, x, \xi'). \quad (20)$$

By uniform continuity of  $g(\cdot, y^0(x), x)$ , there exists  $\Delta > 0$  such that (19) holds whenever

$$\|z' - z\| < \Delta. \quad (21)$$

Because satisfaction of (21) and (20) is observable, all  $(\Delta, p)$  allowing satisfaction of these conditions are identified.  $\square$

With this result in hand, we can use equation (18) to relate partial derivatives of  $g(z, y(x), x)$  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.

**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  $z, z' \in \mathcal{Z}(y^0(x), x)$  satisfying (21) the matrix  $\left[\frac{\partial g(z, y^0(x), x)}{\partial z}\right]^{-1} \left[\frac{\partial g(z', y^0(x), x)}{\partial z}\right]$  is identified.*

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

$$z = z^*(s; p, x, \xi_t) \neq z' = z^*(s; p, x, \xi_{t'}), \quad (22)$$

revealing that  $\xi_t \neq \xi_{t'}$ . Lemma 1 ensures that such  $t, t'$ , and  $s$  exist for all  $z, z' \in \mathcal{Z}(y^0(x), x)$  satisfying (21). And although  $\xi_t$  and  $\xi_{t'}$  are latent, the identities of markets  $t$  and  $t'$  satisfying (22) are observed, as are the associated values of  $s$ ,  $z^*(s; p, x, \xi_t)$ , and  $z^*(s; p, x, \xi_{t'})$ . Differentiating (18) with respect to the vector  $s$  within these two markets, we obtain

$$\frac{\partial g(z, y^0(x), x)}{\partial z} \frac{\partial z^*(s; p, x, \xi_t)}{\partial s} = \frac{\partial \sigma^{-1}(s; y^0(x), p, x)}{\partial s}$$

and

$$\frac{\partial g(z', y^0(x), x)}{\partial z} \frac{\partial z^*(s; p, x, \xi_{t'})}{\partial s} = \frac{\partial \sigma^{-1}(s; y^0(x), p, x)}{\partial s}.$$

Thus, recalling Assumption 5, for almost all such  $(z, z')$  we have

$$\left[\frac{\partial g(z', y^0(x), x)}{\partial z}\right]^{-1} \frac{\partial g(z, y^0(x), x)}{\partial z} = \frac{\partial z^*(s; p, x, \xi_{t'})}{\partial s} \left[\frac{\partial z^*(s; p, x, \xi_t)}{\partial s}\right]^{-1}.$$

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

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

in Lemma 2 to the known (normalized) value of the matrix  $\left[\frac{\partial g(z, y^0(x), x)}{\partial z}\right]$  at  $z = z^0(x)$ .

**Lemma 3.** *Under Assumptions 1–6,  $g(\cdot, y^0(x), x)$  is identified on  $\mathcal{Z}(y^0(x), x)$  for all  $x \in \mathcal{X}$ .*

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

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

By construction, all  $z$  and  $z'$  in any given set  $\mathcal{B}_\tau$  satisfy (21). So by Lemma 2, the value of

$$\left[\partial g(z, y^0(x), x)/\partial z\right]^{-1} \left[\partial g(z', y^0(x), x)/\partial z\right] \quad (23)$$

is known for almost all  $z$  and  $z'$  in any set  $\mathcal{B}_\tau$ . Because  $\cup_{\tau \in \mathbb{Z}^J} \mathcal{B}_\tau$  forms an open cover of  $\mathcal{Z}(y^0(x), x)$ , given any  $z \in \mathcal{Z}(y^0(x), x)$  there exists a simple chain of open sets  $\mathcal{B}_\tau$  in  $\mathcal{Z}(y^0(x), x)$  linking the point  $z^0(x)$  to  $z$ .<sup>25</sup> Thus,

$$\left[\partial g(z, y^0(x), x)/\partial z\right]^{-1} \left[\partial g(z^0(x), y^0(x), x)/\partial z\right]$$

is known for almost all  $z \in \mathcal{Z}(y^0(x), x)$ . With the normalization (8) and the continuity of  $\partial g(z, y^0(x), x)/\partial z$  with respect to  $z$ , the result then follows from the fundamental theorem of calculus for line integrals and the boundary condition (9).  $\square$

Before moving to identification of conditional demand, we pause to point out that our constructive identification of  $g(\cdot, y^0(x), x)$  used only a single price vector  $p$  at each value of  $x$ —that required by Assumption 6. In typical models of supply this condition would hold for almost all price vectors in the support of  $P_t|\{X_t = x\}$ . In addition to providing falsifiable restrictions, this indicates a form of redundancy that would typically be exploited by estimators used in practice. Similarly, our proof of Lemma 3 used, for each  $z \in \mathcal{Z}(y^0(x), x)$ , only one of infinitely many paths between  $z^0$  and  $z$ ; integrating along any such path must yield the same function  $g(\cdot, y^0(x), x)$  at each  $x$ .

## 4.2 Identification of Conditional Demand

We demonstrate identification of the conditional demand functions  $\bar{y}(\cdot; t)$  under two additional conditions. The first is a requirement of sufficient variation in the consumer-level observables  $Z_{it}$ .

<sup>25</sup>See, e.g., van Mill (2002, Lemma 1.5.21).

**Assumption 7** (Common Choice Probability). *For each  $x \in \mathcal{X}$ , there exists a choice probability vector  $s^*(x)$  such that  $s^*(x) \in \mathcal{S}(p, x, \xi)$  for all  $(p, \xi) \in \text{supp}(P_t, \Xi_t) | \{X_t = x\}$ .*

Assumption 7 requires that, at each  $x \in \mathcal{X}$ , there exist some choice probability vector  $s^*(x)$  that is common to all markets—that

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

be nonempty. The nondegeneracy of each set  $\mathcal{S}(p_t, x_t, \xi_t)$  (recall (16)) reflects variation in  $Z_{it}$  across its support. Assumption 7 requires enough variation in  $Z_{it}$  that for some  $s^*(x)$  we have  $s^*(x) \in \mathcal{S}(p_t, x, \xi_t)$  for all  $(p_t, \xi_t)$  in their support conditional on  $X_t = x$ .

The strength of this assumption depends on the joint support of  $(P_t, \Xi_t)$  given  $\{X_t = x\}$  and on the relative impacts of  $(Z_{it}, \Xi_t, P_t)$  on choice behavior. Observe that  $P_{jt}$  and  $\Xi_{jt}$  typically will have opposing impacts and will be positively dependent conditional on  $X_t$  under equilibrium pricing behavior; thus, large support for  $g(Z_{it}, y^0(x), x)$  may not be required even if  $\Xi_t$  were to have large support. Indeed, we can contrast our assumption with a requirement of special regressors with large support: the latter would imply that *every* interior choice probability vector  $s$  is a common choice probability for all  $x$ ; we require only a single common choice probability at each  $x$ . Note also that, because choice probabilities conditional on  $(Z_{it}, Y_{it})$  are observable in all markets, Assumption 7 is verifiable.<sup>26</sup> This is important on its own. And, because the choice of each  $y^0(x)$  is arbitrary, it implies that we require only existence (for each  $x$ ) of one such  $y^0(x) \in \mathcal{Y}$  such that Assumption 7 holds.<sup>27</sup> Finally, an important observation for what follows is that the values of any common choice probability vectors  $s^*(x)$  may be treated as known.

Our second requirement is existence of instruments for prices satisfying the standard nonparametric IV conditions.

**Assumption 8** (Instruments for Prices).

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

Part (i) of Assumption 8 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

<sup>26</sup>See Berry and Haile (2018) for a formal definition of verifiability.

<sup>27</sup>When more than one such value  $y^0(x)$  exists, or when there is more than one common choice probability vector  $s^*$ , this introduces additional falsifiable restrictions.

that  $E[h(X_t, \Xi_t)|X_t] = 0$  by construction; thus part (i) implies

$$E[h_j(X_t, \Xi_{jt})|X_t, W_t] = 0 \quad \text{a.s. for all } j. \quad (24)$$

This is true regardless of whether  $X_t$  itself is exogenous. Of course, one must be cautious about satisfaction of part (i) when  $X_t$  is thought to be endogenous. In general, candidate instruments that are properly excludable unconditionally may not be so conditional on an endogenous control. We discuss this further below and devote the appendix to a detailed discussion of when standard instruments for prices will (or will not) satisfy the exclusion requirement when  $X_t$  is endogenous. Part (ii) is a standard completeness condition—the nonparametric analog of the classic rank condition for linear regression. For example, Newey and Powell (2003) have shown that under mean independence (the analog of (24) here), completeness is necessary and sufficient for identification in separable nonparametric regression. The following result demonstrates that, given existence of a common choice probability vector  $s^*$ , 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 1–8, the scalar  $h_j(x_t, \xi_{jt})$  is identified for all  $j$  and  $t$ .*

*Proof.* Taking  $x = x_t, p = p_t, \xi = \xi_t$  and  $s = s^*(x_t)$  in equation (18), we have

$$g(z^*(s^*(x_t); p_t, x_t, \xi_t), y^0(x_t), x_t) = \sigma^{-1}(s^*(x_t); y^0(x_t), p_t, x_t) - h(x_t, \xi_t).$$

Thus, for all  $t$  and each  $j = 1, \dots, J$ ,

$$g\left(z^*(s^*(x_t); p_t, x_t, \xi_t), y^0(x_t), x_t\right) = f_j(x_t, p_t) - e_{jt} \quad (25)$$

where  $f_j(x_t, p_t) \equiv \sigma_j^{-1}(s^*(x_t); y^0(x_t), p_t, x_t)$  and  $e_{jt} \equiv h_j(x_t, \xi_{jt})$ . By Lemma 3 the left side of (25) is known (recall that the values of each  $z^*(s^*(x_t); p_t, x_t, \xi_t)$  are observable, even though the value of each  $\xi_t$  is not). Thus, for each  $j$  this equation takes the form of a separable nonparametric regression model. Given Assumption 8, identification of each function  $f_j$  follows immediately from the identification result of Newey and Powell (2003). This implies identification of each  $e_{jt}$  (i.e.,  $h_j(x_t, \xi_{jt})$ ) as well.  $\square$

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

**Theorem 1.** *Under Assumptions 1–8,  $\bar{\alpha}(\cdot; t)$  is identified on  $\mathcal{C}(x_t, \xi_t)$  for all  $t$ .*

*Proof.* Recall that

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

Because  $Q_{it}, Z_{it}, Y_{it}, P_t, X_t$  are observed and each  $h(x_t, \xi_t)$  is known, the result follows.  $\square$

We emphasize that although the conditional demand functions  $\bar{s}(\cdot; t)$  are indexed by  $t$ , this merely stands in for the values of  $X_t$  and  $h(X_t, \Xi_t)$ . Within a single market, there is no price variation. However, Lemma 4 allows us to utilize information from all markets with same values of  $X_t$  and  $h(X_t, \Xi_t)$  to reveal how price variation affects demand at all  $(Z_{it}, Y_{it}, P_t, X_t, h(X_t, \Xi_t))$  in their joint support.

### 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  $\Xi_t$  fixed. Addressing such questions will require separating the impacts of  $X_t$  and  $\Xi_t$ . This can be done by adding the standard assumption that  $X_t$  is exogenous.

**Assumption 9** (Exogenous Product Characteristics).  $E[\Xi_t | X_t] = 0$ .

When Assumption 9 holds, the definition (2) implies

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

This has two important implications. First, when Assumption 9 is maintained, the IV exclusion condition (part (i) of Assumption 8) softens to require instruments  $W_t$  that are exogenous conditional on exogenous (rather than endogenous)  $X_t$ . Second, Lemma 4 now implies that each realization  $\xi_t$  of the demand shock vector is identified. Recalling that

$$s(Z_{it}, Y_{it}, P_t, X_t, \Xi_t) = E[Q_{it} | Z_{it}, Y_{it}, P_t, X_t, \Xi_t],$$

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

**Theorem 2.** *Under Assumptions 1–9,  $s$  is identified on  $\mathcal{C}$ .*

## 5 Discussion

The results above demonstrate nonparametric identification of demand (and conditional demand) using a combination of within-market and cross-market variation. Compared to a setting with market-level data, micro data can (i) permit demand specifications that condition on consumer-level observables, (ii) avoid the need to restrict how market/product observables  $X_t$  enter, and (iii) substantially reduce the reliance on instrumental variables.

The last of these may be especially important. The number of instruments needed is halved with micro data, and there is no need for the so-called “BLP instruments.” This softening of instrumental variables requirements is achieved because 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 (see Berry and Haile (2014, 2021)). However, the reason micro-data variation is free from confounding effects of variation in market-level demand shocks is not an assumed exclusion condition in the cross-section of markets but, rather, the fact that within a single market these shocks simply do not vary. Thus, our insights here have a connection to those underlying “within” identification of slope parameters in panel data models with fixed effects.

The most important message from these results is that identification of demand for differentiated products follows using the same sorts of quasi-experimental variation relied upon in simpler settings. Indeed, the exploitation of within-unit variation and instrumental variables are arguably the bread and butter of empirical economics. Of course, these conclusions lead to several questions about appropriate instruments, the potential for softening some conditions for identification by strengthening others, and extensions of our results to other types of demand models. We discuss these questions in the remainder of this section.

### 5.1 What Are Appropriate Instruments?

The fact that reliance on instruments is standard does not imply that instruments will always be available. Rather, this merely shifts discussion of identification largely to standard questions concerning the availability of suitable instruments. What are likely instruments in practice?

Candidate instruments for prices include most of those typically relied upon in the case of market-level data (see Berry and Haile (2016) for a more complete discussion of these candidate instruments). Classic instruments for prices are cost shifters that are excluded from the demand system and (mean-) independent of the demand shocks  $\Xi_t$ . When cost shifters are not observed, proxies for cost shifters may be available and can satisfy the required ex-

clusion conditions.<sup>28</sup> Exogenous shifters of market structure (e.g., exogenous merger activity or, in some cases, exogenous variation in common ownership) that affect prices through equilibrium markups can also serve as instruments.

Micro data can also result in availability of a related category of candidate instruments: market-level observables (e.g., market-level demographic measures) that alter equilibrium markups. Berry and Haile (2014, 2016) refer to these as “Waldfogel” instruments, after Waldfogel (2003).<sup>29</sup> When micro data are available, we can directly account for the impacts of individual-specific demographics, so it may be reasonable to assume that market-level demographics are excluded from the conditional demands we seek to identify. The requirement that these market-level measures be mean independent of the market-level demand shocks is a significant assumption, ruling out certain kinds of geographic sorting or peer effects, for example. But in many applications such an assumption may be natural.

The exclusion restriction that defines an appropriate instrument (part (i) of Assumption 8), requires  $W_t$  to be (mean) independent of the structural error conditional on  $X_t$ . This conditional independence assumption does not require exogenous product/market characteristics  $X_t$ , but neither does it allow all models with endogenous  $X_t$ . Making use of simple graphical causal models, Appendix A discusses a variety of cases in which instruments for prices remain valid under endogeneity of  $X_t$ . It also discusses the key case leading to a failure of the exclusion restriction: when  $X_t$  is chosen in response to both  $\Xi_t$  and  $W_t$ . In such cases,  $W_t$  could instead serve as an instrument for the endogenous components of  $X_t$ , but we would still need different instruments for prices. In some cases, such instruments may be obtained through natural timing assumptions—e.g., using only the current-period innovations to input costs as the instruments.

Absent from the discussion of candidate instruments above are the “BLP instruments”—characteristics of competing products. These play an essential role as instruments for quantities when one has only market-level data (Berry and Haile (2014)). The relevant exclusion condition in that case requires not only exogeneity of certain product characteristics, but also restrictions on the way they enter demand. Micro data makes it possible to avoid these requirements, although adding them can allow use of BLP instruments for prices (see section 5.2.1). Absent such additional assumptions, however, the BLP instruments are unavailable, even when  $X_t$  is assumed exogenous. This can be seen

---

<sup>28</sup>Examples of such proxies, plausibly exogenous in some applications, are so-called “Hausman instruments,” i.e., prices of the same good in other markets (e.g., Hausman, Leonard, and Zona (1994), Hausman (1996), or Nevo (2000, 2001)).

<sup>29</sup>See also Gentzkow and Shapiro (2010), Fan (2013), and Li, Hartmann, and Amano (2020).

in the key equation (25), where  $x_t$  appears on the right-hand side for each  $j$ . Each element of  $x_t$  “instruments for itself” in these equations, leaving no product characteristics excluded.

## 5.2 What About Stronger Functional Forms?

In practice, estimation in finite samples is almost always influenced by functional form assumptions—e.g., the choice of parametric structure, kernel functions, or sieve basis. Such functional forms enable interpolation, extrapolation, and bridging of gaps between the variation present in the sample and that needed for nonparametric point identification. A study of nonparametric identification can reveal whether functional form assumptions play a more essential role in one precise sense. One interpretation of our results is that only limited nonparametric structure is essential: beyond the nonparametric index structure, our main requirement for identification is adequate variation through  $(Z_t, W_t)$  of dimension equal to the dimension of the endogenous variables (prices and quantities).

But one can also ask how imposing additional structure on the demand model might allow relaxation of our identification requirements. Answers to this question may be of direct interest and can also suggest the sensitivity of identification to particular conditions. For example, we may feel more comfortable when we know that identification is robust in the sense that a relaxation of one condition for identification can be accommodated by strengthening another. A full exploration of these potential trade-offs describes an entire research agenda. But some examples can illustrate three directions one might go to enlarge the set of potential instruments, further reduce the number of required instruments, or reduce the required dimensionality of the micro data.

For simplicity, our discussion here will consider the typical case in which  $X_t$  is assumed exogenous, focusing then on identification of demand rather than conditional demand. Recall that in this case we have  $h(X_t, \Xi_t) = \Xi_t$ . Given our focus on the role of  $Z_{it}$ , for simplicity we will fix and suppress any additional consumer-level observables  $Y_{it}$  in what follows.

### 5.2.1 Strengthening the Index Structure

Our model avoided any restriction on the way the observables  $X_t$  enter demand. This contrasts with the structure used by Berry and Haile (2014) to consider identification with market-level data. There, for each good  $j$ , one element of  $X_{jt}$  was assumed to enter demand only through the  $j$ th element of the index vector. In practice, such an assumption is common. And adding such a restriction here can introduce another class of potential instruments:

the exogenous characteristics of competing goods, i.e., “BLP instruments.”<sup>30</sup>

To illustrate this as simply as possible, partition  $X_t$  as  $(X_t^{(1)}, X_t^{(2)})$ , where

$$X_t^{(1)} = \left( X_{1t}^{(1)}, \dots, X_{Jt}^{(1)} \right) \in \mathbb{R}^J.$$

Suppose demand takes the form

$$s(Z_{it}, P_t, X_t, \Xi_t) = \sigma \left( \gamma(Z_{it}, X_t, \Xi_t), P_t, X_t^{(2)} \right), \quad (26)$$

where for  $j = 1, \dots, J$

$$\gamma_j(Z_{it}, X_t, \Xi_t) = g_j(Z_{ijt}, X_t^{(2)}) + \eta_j(X_{jt}^{(1)}, X_t^{(2)}) + \Xi_{jt}. \quad (27)$$

Compared to our original specification, here we (a) restrict  $X_t^{(1)}$  to enter only through the index vector; (b) associate the  $j$ th components of  $Z_{it}$  and  $X_t^{(1)}$  exclusively with the  $j$ th element of the index vector; and (c) impose additive separability between  $Z_{ijt}$  and  $X_{jt}^{(1)}$  within each index.<sup>31</sup> This specification requires that each element of  $Z_{it}$  can be matched to an element of  $X_t^{(1)}$  that affects demand in a similar way. Many specifications in the literature satisfy this requirement, typically with additional restrictions such as linear substitution between  $Z_{ijt}$  and  $X_{jt}^{(1)}$ . We will also strengthen the common choice probability condition to require existence of a common choice probability vector  $s^*(X_t)$  that does not vary with  $X_t^{(1)}$ .<sup>32</sup> For simplicity we also assume that, for each  $x^{(2)}$  in the support of  $X_t^{(2)}$  there is some point  $z^0(x^{(2)})$  common to  $\mathcal{Z}(x, x^{(2)})$  for all  $x$  in the support of  $X_t^{(1)} \setminus \{X_t^{(2)} = x^{(2)}\}$ .

For the remainder of this section we will condition on  $X_t^{(2)}$  (treating it fully flexibly), suppress it from the notation, and let  $X_t$  represent  $X_t^{(1)}$ . Posing the identifications question here requires a different set of normalizations.<sup>33</sup>

<sup>30</sup>The key question is the proper excludability of these instruments, which in general requires more than mean independence between  $X_t^{(1)}$  and  $\Xi_t$ . The “relevance” of these measures as instruments for prices reflects the fact that in standard oligopoly models each good’s markup depends on the characteristics of all substitutes or complements.

<sup>31</sup>Exclusivity of  $X_{jt}^{(1)}$  to the index  $\gamma_j$  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. Part (c) substantially simplifies the exposition. As in our more general model, the elements of  $\gamma(Z_{it}, X_t, \Xi_t)$  need not be linked to particular goods.

<sup>32</sup>Formally, we assume that for each  $x^{(2)} \in \text{supp } X_t^{(2)}$ , there exists a choice probability vector  $s^*(x^{(2)})$  such that for all  $x^{(1)} \in \text{supp } X_t^{(1)} \setminus \{X_t^{(2)} = x^{(2)}\}$ ,  $s^*(x^{(2)}) \in \mathcal{S}(p, (x^{(1)}, x^{(2)}), \xi)$  for all  $(p, \xi) \in \text{supp } (P_t, \Xi_t) \setminus \{X_t = (x^{(1)}, x^{(2)})\}$ .

<sup>33</sup>Those in section 2.6 do not respect the exclusivity restrictions imposed here and, therefore, cannot be assumed without loss in this case.

These are standard location and scale normalizations. First, because adding a constant  $\kappa_j$  to  $g_j$  and subtracting the same constant from  $\eta_j$  would leave the model unchanged, we take an arbitrary  $x^0 \in \mathcal{X}$  and set

$$\eta_j(x_j^0) = 0 \quad \forall j. \quad (28)$$

Even with (28) (and our maintained  $E[\Xi_t] = 0$ ), it remains true that any linear (or other injective) transformation of the index  $\gamma_j$  could offset by an appropriate adjustment to the function  $\sigma$ , yielding multiple representations of the same demand system (recall the related observation in section 2.6). Thus, without loss, we normalize the location and scale of each index  $\gamma_j$  by setting

$$\begin{aligned} g_j(z_j^0) &= 0 \\ \frac{\partial g_j(z_j^0)}{\partial z_j} &= 1. \end{aligned}$$

The arguments in Lemmas 1–3 now demonstrate identification of each function  $g_j$ . At the common choice probability vector  $s^*$ , the inverted demand system takes the form of equations

$$g_j(z_j^*(s^*)) + \eta_j(x_{jt}) + \xi_{jt} = \sigma_j^{-1}(s^*; p_t)$$

for each  $j$ . Writing the  $j$ th equation as

$$g_j(z_j^*(s^*)) = -\eta_j(x_{jt}) + \sigma_j^{-1}(s^*; p_t) - \xi_{jt}, \quad (29)$$

we obtain a nonparametric regression equation with RHS variables  $x_{jt}$  and  $p_t$ . Here  $x_{-jt}$  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 Waldfoel instrument—would yield enough instruments to obtain identification of the unknown RHS functions and the “residuals”  $\xi_{jt}$ .<sup>34</sup> As before, 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  $j^{\text{th}}$  index. This can lead to a regression equation (the analog of (29)) of the form

$$g_j(z_j^*(s^*)) = -\eta_j(x_{jt}, p_{jt}) + \sigma_j^{-1}(s^*) - \xi_{jt}.$$

Now only one instrument for price is necessary. For example, the BLP instruments can overidentify demand.

---

<sup>34</sup>Here the separability in  $X_{jt}$  provides a falsifiable restriction.

### 5.2.2 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

$$U_{ijt} = g_j(Z_{ijt}) + \Xi_{jt} + \mathcal{E}_{ijt},$$

where  $\mathcal{E}_{ijt}$  is a scalar random variable whose nonparametric distribution depends on  $X_{jt}$  and  $P_{jt}$  (equation (14) gives a parametric example). In this case, our Lemma 3 demonstrates identification of each function  $g_j(\cdot)$  up to a normalization of utilities.

If one is willing to add the assumption of independence between  $Z_{ijt}$  and  $\mathcal{E}_{ijt}$ , this turns  $g_j(Z_{ijt})$  into a known special regressor. Under a further (typically very restrictive) large support assumption on  $g_j(Z_j)$ , a standard argument demonstrates identification of the marginal distribution of  $(\Xi_{jt} + \mathcal{E}_{ijt})|(X_t, P_t)$ . This is not sufficient to identify demand. However, one can use these marginal distributions to define a nonparametric IV regression equation for each choice  $j$ , where the LHS is a conditional mean and  $\Xi_{jt}$  appears on the RHS as an additive structural error.<sup>35</sup> In each of these equations the prices and characteristics of goods  $k \neq j$  are excluded. Identification of these equations identifies all demand shocks, and identification of demand then follows as in Theorem 2. Thus, in this framework one needs only one instrument for price, and exogenous characteristics of competing goods (BLP IVs) would be available as instruments.

### 5.2.3 A Semiparametric Model

Moving further in the direction of parametric models commonly used in practice can reduce both the required dimensionality of consumer attributes and the number of required instruments. As one example, consider a semi-parametric nested logit model. We condition on (and suppress from the notation)  $X_t$ ,<sup>36</sup> and consider a semiparametric nested logit model where inverse demand in market  $t$ , given  $z_{it}$ , is

$$g_j(z_{it}) + \xi_{jt} = \ln(s_{jt}(z_{it})/s_{0t}(z_{it})) - \theta \ln(s_{j/n,t}(z_{it})) + \alpha p_{jt}. \quad (30)$$

---

<sup>35</sup>See our earlier working paper, Berry and Haile (2010).

<sup>36</sup>By conditioning on  $X_t$ , we permit it to enter the model fully flexibly. Let conditional indirect utilities take the form

$$u_{ijt} = u(x_t, g_j(z_{it}, x_t) + \xi_{jt} - \alpha(x_t) p_{jt} + \mu_{ijt}(x_t)),$$

where  $u$  is strictly increasing in its second argument,  $\alpha(x_t)$  is arbitrary, and  $\mu_{ijt}(x_t)$  is a stochastic component taking the standard composite nested-logit form at each  $x_t$ . The identification argument sketched here may be repeated at each  $x_t$ .

Here  $s_{jt}(z_{it})$  denotes good  $j$ 's observed choice probability in market  $t$  conditional on  $z_{it}$ , and  $s_{j/n,t}(z_{it})$  denotes its within-nest conditional choice probability. The scalar  $\theta$  denotes the usual “nesting parameter.” Here we allow  $z_{it}$  to have fewer than  $J$  elements.

As with the standard representation of most parametric models of inverse demand, the nested logit model embeds normalizations of the indices and demand function analogous to our choices of  $A(x)$  and  $B(x)$  in section 2.6. However, we must still normalize the location of either  $\Xi_{jt}$  or  $g_j$  for each  $j$  to pose the identification question. Here we will set  $g_j(z^0) = 0$  for all  $j$ , breaking with our prior convention by leaving each  $E[\Xi_{jt}]$  free.

Take any market  $t$  and any  $z \in \mathcal{Z}$ . Differentiating (30) with respect to one (possibly, the only) element of  $z_{it}$ —say  $z_{it}^{(1)}$ —at the point  $z$  yields

$$\frac{\partial g_j(z)}{\partial z^{(1)}} = \frac{\partial \ln s_{jt}(z)}{\partial z^{(1)}} - \frac{\partial \ln s_{0t}(z)}{\partial z^{(1)}} - \theta \frac{\partial \ln s_{j/nt}(z)}{\partial z^{(1)}}. \quad (31)$$

In this equation,  $\frac{\partial g_j(z)}{\partial z^{(1)}}$  and  $\theta$  are the only unknowns. Moving to another market  $t'$ , we can obtain a second equation of the same form in which the left-hand side is identical to that in (31). Equating the right-hand sides yields

$$\frac{\partial \ln s_{jt}(z)}{\partial z^{(1)}} - \frac{\partial \ln s_{0t}(z)}{\partial z^{(1)}} - \theta \frac{\partial \ln s_{j/nt}(z)}{\partial z^{(1)}} = \frac{\partial \ln s_{jt'}(z)}{\partial z^{(1)}} - \frac{\partial \ln s_{0t'}(z)}{\partial z^{(1)}} - \theta \frac{\partial \ln s_{j/nt'}(z)}{\partial z^{(1)}}.$$

Thus, we can solve for  $\theta$  as long as

$$\frac{\partial \ln s_{j/nt}(z)}{\partial z^{(1)}} \neq \frac{\partial \ln s_{j/nt'}(z)}{\partial z^{(1)}},$$

a condition that will typically hold when  $\xi_{t'} \neq \xi_t$  or  $p_{t'} \neq p_t$ , and which is directly observed. With  $\theta$  known, we then identify (indeed, over-identify) all derivatives of  $g_j(z)$  from (31), yielding identification of the function  $g$  as in Lemma 3. Identification of the remaining parameter  $\alpha$  can then be obtained from the “regression” equation

$$g_j(z_{it}) = \ln(s_{jt}(z_{it})/s_{0t}(z_{it})) - \theta \ln(s_{j/n,t}(z_{it})) + \alpha p_{jt} - \xi_{jt}, \quad (32)$$

obtained from (30), using a single excluded instrument—e.g., an excluded exogenous market-level cost shifter or markup shifter that affects all prices. The constant recovered in this regression represents  $E[\Xi_{jt}]$ .

Although this example involves a model that is more flexible than nested logit models typically estimated in practice, it moves a considerable distance from our fully nonparametric model. But this example makes clear that additional structure can further reduce the dimension of the required exogenous variation. Indeed, here we can obtain identification with a single instrument

and a scalar individual-level observable  $z_{it}$ . This compares to the usual requirement of two instruments for the fully parametric nested logit when one has only market-level data (see Berry (1994)). Other semiparametric models may offer more intermediate points in the set of feasible trade-offs between the flexibility of the model and the dimension of exogenous variation needed for identification.

### 5.3 What about Continuous Demand Systems?

Although we have focused on the case in which the consumer-level quantities  $Q_{ijt}$  are those arising from a discrete choice model, nothing in our proofs requires this. In other settings, the demand function  $\nu$  defined in (1) may simply be reinterpreted as the expected vector of quantities demanded conditional on  $(X_t, P_t, \Xi_t, Z_{it}, Y_{it})$ .<sup>37</sup> Applying our results to continuous demand is therefore just a matter of verifying the suitability of our assumptions.<sup>38</sup>

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 reintroduce  $Y_{it}$  to denote consumer  $i$ ’s income, measured in units of the numeraire good 0. Each consumer  $i$  in market  $t$  has utility over consumption vectors  $q \in \mathbb{R}_+^{J+1}$  given by

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

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

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

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

$$q_{ijt} = \frac{y_{it} \exp(g_j(z_{it}, x_t) + \xi_{jt} + x_{jt}\beta_{it} - \alpha \ln(p_{jt}))}{1 + \left[ \sum_{k=1}^J \exp(g_k(z_{it}, x_t) + \xi_{kt} + x_{kt}\beta_{it} - \alpha \ln(p_{kt})) \right]}, \quad (33)$$

---

<sup>37</sup>Note that the demand faced by firms in market  $t$  is the expectation of this expected demand over the joint distribution of  $(Z_{it}, Y_{it})$  in the market.

<sup>38</sup>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.

where  $\alpha = 1/(1 - \rho)$ . Equation (33) resembles a choice probability for a random coefficients logit model, although the quantities  $q_{it}$  here take on continuous values and do not sum to one. It is easy to show that our Assumptions 1–4 are satisfied for the expected CES demand functions, which take the form

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

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

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

## 6 Lessons for Applied Work

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

### 6.1 The Incremental Value of Micro Data

The most important practical lesson from our results is that the marginal value of micro data is high. It is not surprising that a setting allowing one to exploit variation both across markets and within markets is more informative than one with only cross-market variation. But 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 these IV requirements to be satisfied. We have shown that adding micro data can eliminate the need to instrument for quantities and, therefore, the necessary reliance on BLP instruments. This, in turn, avoids the need for any restriction on the way observables at the level of the product and market enter the model. Furthermore, our results on identification of conditional demand imply that one can often obtain price elasticities without any exogenous product characteristics, much less the use of such characteristics as instruments.

These are significant advantages. Researchers should, therefore, not only prefer micro data, but should seek it out whenever possible. Of course, even when the setting and assumptions permit use of BLP instruments—or when the micro data available are more limited than we have assumed to explore fully nonparametric identification—variation from micro data can be powerful. This

message is consistent, for example, with the findings in the empirical literature (e.g., Petrin (2002)) that the addition of even limited micro data often results in much more precise estimates than those obtained with market-level data alone.

## 6.2 The Necessity of Cross-Market Variation

Another important lesson from our work concerns the need for cross-market variation, even when one has micro data. Variation within a single market cannot suffice for identification, at least without additional assumptions.

Formally, our proofs relied on cross-market variation, even for identification of the function  $g(\cdot, y^0(x), x)$  (see section 4.1). But the necessity of cross-market variation is also easy to see. In a single market the observables consist of conditional choice probability vectors  $s(z_i, y_i)$  at all  $(z_i, y_i) \in \Omega$ —here we will suppress the index  $t$  as well as the observables  $(X_t, P_t)$ , since these have no variation in a single market. Consider an arbitrary (and, thus, typically misspecified) invertible parametric demand function  $\sigma(g(z_i, y_i) + \xi; \theta)$  that maps  $J$  indices  $g_j(z_i, y_i) + \xi_j$  to market shares. For concreteness, suppose this is a nested logit model with “mean utilities”  $g_j(z_i, y_i) + \xi_j$  and nesting parameter(s)  $\theta$ . By standard results (see Berry (1994)), given any value of  $\theta$ , this model can fit the data in the market perfectly by setting  $\xi_j = 0$  and

$$g(z_i, y_i) = \sigma^{-1}(s(z_i, y_i); \theta).$$

This yields a different function  $g$  for every candidate value of  $\theta$ , and no value of  $\theta$  can be ruled out. Thus, the observables from a single market cannot identify the nesting parameters in this semiparametric nested logit model, much less determine whether the nested logit structure is correct.

It is also easy to see here how having micro data in multiple markets can help. When the same demand model is assumed to apply to multiple markets, the same  $(g, \theta)$  pair must fit the data in each market. The resulting restrictions can rule out incorrect candidates for  $\theta$  and  $g$ , as we have seen in section 5.2.3.<sup>39</sup>

Of course, although we have suppressed the price vector  $P_t$  when talking about single market, the effects of price variation on quantities demanded are essential. Because price vectors are typically fixed within markets by definition, exogenous sources of cross-market price variation will be needed. Thus, even in the presence of micro data there are at least two reasons applied researchers should seek out data on multiple markets. First, a combination

---

<sup>39</sup>As suggested in section 5.2.3, with a scalar  $\theta$  defining substitution between products in response to changes in the index vector, two markets may suffice. Our Lemmas 1–3 show how the restrictions across many markets allow identification when these substitution patterns are nonparametric.

of within-market and cross-market variation is needed to identify flexibly-specified effects of consumer observables (including the function  $g$ ). Second, cross-market variation through instruments for prices is essential for learning how demand responds to price variation.

These observations also serve as a caution. As a practical matter, with a fully parametric specification of demand it will often be possible to estimate all parameters with data from only one market. And in some cases, only a single market is available for study. The classic work of McFadden et al. (1977) offers one example. However, identification in such cases will implicitly rely on functional form restrictions—restrictions that could be relaxed in a multi-market setting.

Our findings on the theoretical importance of cross-market variation can be linked to the practical findings of Berry, Levinsohn, and Pakes (2004), who reported that when using only consumer-level variation—no cross-market variation or “second choice” data<sup>40</sup>—their attempts to estimate random coefficients logit models failed due to a nearly flat objective function. They speculated (p. 90) that “in applications to other data sets, variation in the choice set (either over time or across markets) might provide the information necessary to estimate the random coefficients.” Our results provide a nonparametric confirmation of that conjecture, again pointing to the practical value of data that combine within- and cross-market variation.

### 6.3 What Does Not Follow

Although nonparametric identification results can offer important insights, they address a very specific question about what can be learned from data. A nonparametric identification result can 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. The choice of finite-sample approximation method can of course matter. In the case of demand estimation, functional form restrictions used in practice restrict the families of demand functions considered in a way that can constrain the answers to key questions. Thus, sensitivity of estimates (most importantly, estimates of the quantitative answers to the economic questions of ultimate interest) to functional form choices remains an important issue for empirical researchers to explore. 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.

---

<sup>40</sup>Berry and Haile (2021) discuss the close relationship between second-choice data and micro data from two markets.

We also emphasize that our sufficient conditions for nonparametric identification should not be viewed as necessary conditions, formally or informally, for demand estimation in practice. Nonparametric identification results should guide our thinking about the strength of the available data and empirical results. But it would be a mistake to view these as conditions that must be confirmed before proceeding with empirical work. Nonparametric identification of most models in economics (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.

## 7 Conclusion

Since Berry, Levinsohn, and Pakes (1995), there has been an explosion of interest in estimation of demand models that incorporate both flexible substitution patterns and explicit treatment of the demand shocks that introduce endogeneity/simultaneity. Understandably, this development has been accompanied by questions about what allows identification of these models. Our results, here and in Berry and Haile (2014), offer a reassurance that identification follows 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 depth of the identification challenge in the context of demand systems—notably, the fact that even purely exogenous variation in prices is generally not sufficient to identify price elasticities or other essential features of demand.<sup>41</sup>

Furthermore, identification of these models is not fragile. Identification does not rely on “identification-at-infinity” arguments; it is not limited to particular types of 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 some key conditions by strengthening others. Thus, although this is a case in which identification results come well after an extensive empirical literature has already developed, the nonparametric foundation for this literature is strong.

---

<sup>41</sup>See the discussion in Berry and Haile (2021).

# Appendices

## A Excludability Conditional on Endogenous Product Characteristics

In section 5.1 we discussed several categories of instruments  $W_t$  commonly relied upon to provide exogenous variation in prices. Here we discuss the question of when such instruments remain properly excluded even when conditioning on observables  $X_t$  that are not independent of  $\Xi_t$ . Such instruments are required for Theorem 1 to apply when Theorem 2 does not, allowing identification of conditional demand without requiring exogeneity of  $X_t$  (or instruments for  $X_t$ ).

In what follows we suppress the market subscripts  $t$  on the random variables  $X_t, W_t, \Xi_t$ , etc. Our discussion will utilize graphical causal models, with d-separation providing the key criterion for assessing the independence between  $W$  and  $\Xi$  conditional on  $X$ .<sup>42</sup> Our use of these tools is elementary. 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 of course implies the conditional mean independence required by Theorem 1.

Below we first discuss several causal graphs (and motivating economic examples) that “work”—i.e., that imply the conditional independence condition (A.1). We then discuss the main type of structure that does not work—i.e., in which (A.1) fails despite unconditional independence between  $W$  and  $\Xi$ . We will see that each type of instrument discussed in section 5.1 can remain valid under several models of endogenous  $X$ . However, each type of instrument can also fail; in particular, (A.1) will fail despite unconditional independence between  $W$  and  $\Xi$  when firms choose  $X$  in ways that depend on both  $W$  and  $\Xi$  (or their ancestors). However, in these situations, a natural timing assumption can often yield a new set of valid instruments for prices.

---

<sup>42</sup>See, e.g., Pearl (2009) and Pearl, Glymour, and Jewell (2016), including references therein. Throughout we maintain the standard assumption that nodes in the causal directed acyclic graphs 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  $\Xi$ .

For example, suppose  $X$  is chosen by firms with knowledge of  $\Xi$ , so that  $X$  is endogenous in the same sense that prices are. Given (A.2), one obtains the causal graph shown in Figure 1.<sup>43</sup> The conditional independence condition (A.1) then can also be seen to follow immediately by the d-separation criterion. We of course reach the same conclusion if the direction of causation between  $\Xi$  and  $X$  is reversed, as in Figure 2—e.g., if  $X$  is chosen without knowledge of  $\Xi$  but the distribution of  $\Xi$  changes with the choice of  $X$ . Taking the canonical example of demand for automobiles, a manufacturer’s choice to offer a fuel efficient hybrid sedan may imply a very different set of (or response to) relevant unobserved characteristics than had a pickup truck or luxury SUV been offered instead.

Figure 1

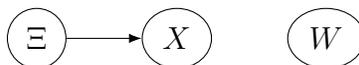
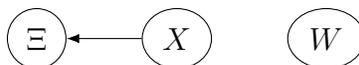


Figure 2

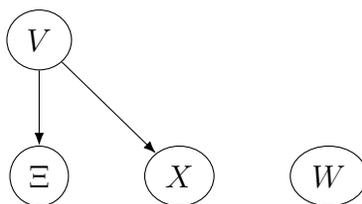


A similar structure is obtained when dependence between  $\Xi$  and  $X$  reflects a common cause (which could be latent). This is illustrated in Figure 3. For example, the common cause  $V$  might represent past values of demand shocks, which are predictive of the current shocks  $\Xi$  and are determinants of firms’ choices of product characteristics  $X$ .

---

<sup>43</sup>We assume throughout that prices and quantities are not among the ancestors of  $(X, W, \Xi)$  and, therefore, typically exclude them from the graphs without loss when examining the properties of the joint distribution of  $(X, W, \Xi)$ . Exclusion of prices and quantities from the ancestors of  $(X, W, \Xi)$  is implied by standard assumptions that consumers take  $X$  and  $\Xi$  as given when making purchase decisions, that  $W$  does not respond to prices or quantities, and that prices are not chosen before  $X$ . Note that in the case of simultaneously chosen  $X$  and  $P$ , the parents of  $X$  and  $P$  will include only previously determined variables (namely, those entering the reduced forms for  $X$  and  $P$ ); thus, neither  $X$  nor  $P$  will be an ancestor of the other. We include an example of simultaneous determination below.

Figure 3

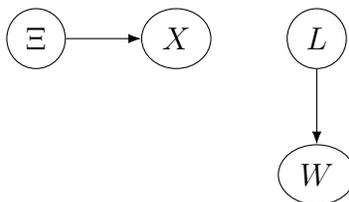


All the instrument types discussed in section 5.1 can satisfy (A.2). For example,  $W$  could represent cost shifters such as input price shocks, realized after  $X$  is chosen and not affected by  $X$ . These might be shocks to import tariffs; shipping costs; retailer costs (e.g., rents, wages); or 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  $\Xi$  and leave product offerings unchanged offer one possibility. Another is cross-market variation in the distribution  $F_{YZ}(\cdot|t)$  (or other aggregate demographic measure at the market or regional level), as long as this variation is independent of  $\Xi$  (as required generally for the validity of Waldfoegel instruments) and  $X$ .

Proxies for cost shifters can also satisfy (A.2). Proxies lead to slightly different causal graphs, as illustrated in Figure 4. There  $L$  represents latent cost shifters known by firms when setting prices but not affecting  $X$ . The instruments  $W$  are proxies for  $L$ . For example,  $W$  could represent Hausman instruments when we have both (a) the latent product-level cost shifters  $L$  are independent of  $X$ , and (b) for each market,  $\Xi$  and  $X$  are independent of prices in other markets. Condition (b) may often fail, since prices in other markets depend on the product characteristics in those other markets, and the characteristics of a given product will typically be highly correlated across markets. Fortunately, independence between  $X$  and  $W$  is not required, as we discuss next.

Figure 4



### A.1.2 Instruments Caused by $X$

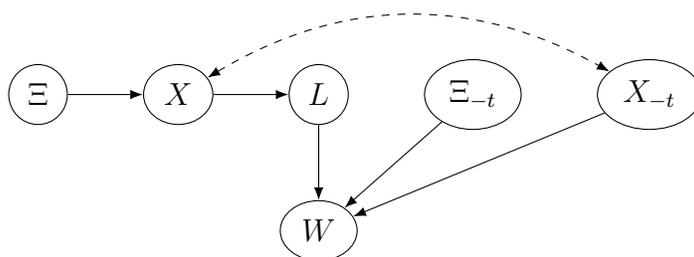
Independence between  $X$  and  $W$  is not required. For example, consider the case in which (i)  $X$  is chosen with knowledge of  $\Xi$  and (ii) the choice of  $X$  affects  $W$ . Then we can obtain the following causal graph, in which (A.1) again holds by the d-separation criterion.

Figure 5



The case represented by Figure 5 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  $X$ .

Figure 6

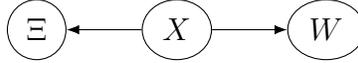


Allowing dependence between  $X$  and  $W$  also broadens the applicability of Hausman instruments (or other proxies for latent cost shocks) by allowing the latent costs shocks to depend on  $X$ . This is illustrated in Figure 6, where we have been explicit about the observed characteristics  $X_{-t}$  and demand shocks  $\Xi_{-t}$  of “other markets,” both of which affect the Hausman instruments  $W$ . Here we link  $X$  and  $X_{-t}$  with a dotted bidirectional edge to indicate (as usual) dependence through unmodeled common causes. Note that the absence of an edge linking  $\Xi$  and  $\Xi_{-t}$  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  $\Xi$ . Note that here  $W$  and  $\Xi$  are not independent, but the conditional independence condition (A.1) is satisfied, implying the exclusion condition needed for identification.

Reversing the direction of causation between  $X$  and  $\Xi$  in Figure 5 or Figure 6 leads to the same conclusion. For example, in the first case we obtain

structure in Figure 7, where the required conditional independence condition is again immediate. Examples generating this structure are similar to those just discussed (including the proxy variation), but with  $\Xi$  now representing market-level shocks whose distribution responds to firms' choices of  $X$ .

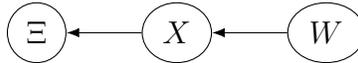
Figure 7



### A.1.3 X Caused by Instruments

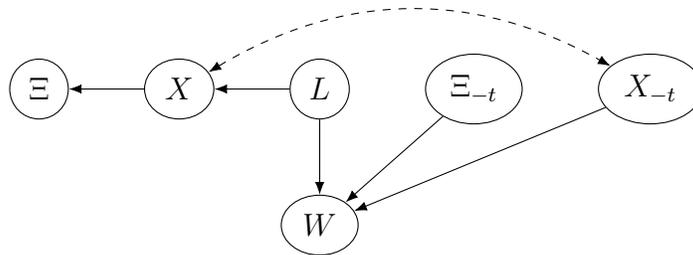
In some cases, the conditional independence condition (A.1) can hold even when  $X$  is affected by  $W$ . Consider the causal graph in Figure 8. As an example motivating this structure, suppose  $W$  is a product-level cost of producing a product feature measured by  $X$ , the latter chosen with knowledge of  $W$  but before  $\Xi$  (or any signal of its realization) is known.

Figure 8



If the product-level cost shifters are latent, proxies—e.g., the same firm's choice of price or product characteristics in other markets—could play the role of  $W$ , again assuming that  $X$  is chosen with knowledge of the cost shock but before  $\Xi$  is realized. As usual with such proxies (i.e., even when one assumes  $X$  to be exogenous), one must maintain an assumption that the demand shocks  $\Xi$  are independent across markets. Such examples are illustrated by Figure 9.

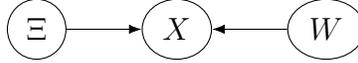
Figure 9



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

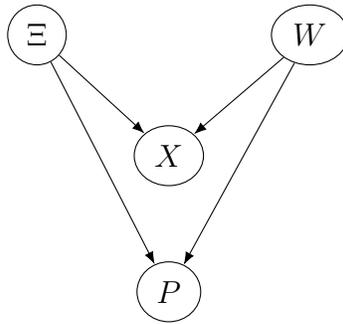
The conditional independence condition (A.1) fails when both  $W$  and  $\Xi$  affect  $X$ . This is illustrated in Figure 10. In this case,  $X$  is a collider in the (undirected) path between  $W$  and  $\Xi$ . Thus, although  $\Xi$  and  $W$  are independent, (A.1) fails.

Figure 10



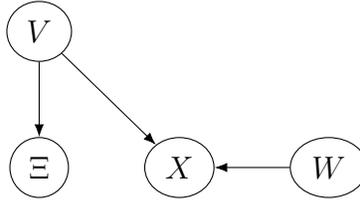
This structure arises whenever firms' choices of  $X$  depend on both  $W$  and  $\Xi$ . An example is when  $W$  is a cost shifter affecting firms' choices of  $X$ , the latter also chosen with knowledge of  $\Xi$ . Another example is when  $W$  is a market-level demographic measure or market structure measure (e.g., product ownership matrix) that, along with  $\Xi$ , influences firms' choices of  $X$ . As this discussion suggests, this structure arises in most cases in which  $X$  and  $P$  are chosen simultaneously, with knowledge of both  $\Xi$  and  $W$ . Indeed, Figure 11, which now includes  $P$ , provides the causal graph under such simultaneity. With this structure, Figure 10 is indeed the ancestral graph for  $\Xi, X, W$  (see also footnote 43).

Figure 11



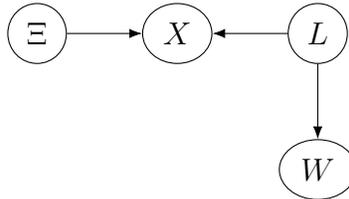
A similar structure arises when the dependence between  $\Xi$  and  $X$  reflects a common cause  $V$ , as in Figure 12. In this case,  $X$  is again a collider in the path between  $\Xi$  and  $W$ .

Figure 12



We can also obtain this type of structure if  $W$  is a proxy for a latent cost shifter affecting firms' choices of  $X$ . Figure 13 illustrates, letting  $L$  represent latent cost shifters known (along with  $\Xi$ ) by firms when choosing  $X$ , with  $W$  denoting a proxy for this shifter—e.g., prices or product characteristics of the same firm in other markets.

Figure 13



Thus, just as there are cases in which each type of instrument discussed in section 5.1 remains valid when conditioning on endogenous characteristics  $X$ , there are other important cases in which (A.1) will fail. Given  $W \perp\!\!\!\perp \Xi$ , the key threat to the conditional independence condition (A.1) is a case in which  $X$  responds both to the structural errors  $\Xi$  and to the candidate instrument  $W$  (or to the latent factor that  $W$  proxies). In such situations, identification will require different instruments for prices.

In many cases such instruments will be easily constructed under natural timing assumptions. This is a topic we take up in the final section of this appendix. We also note that when  $X$  is a collider,  $W$  provides a candidate instrument for  $X$ . Thus, the problem here can provide part of its own solution: when instruments for prices and the endogenous components of  $X_t$  are available, our results extend immediately by expanding  $P_t$  and  $W_t$  to include all endogenous variables and all instruments, respectively.

### A.3 Averting 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  $\Xi$ , partially determine firms' choices of product characteristics  $X$ . However, 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  $\tau$  will reflect design choices made well in advance—in particular, before the input costs for year- $\tau$  production are fully known. Pricing in year  $\tau$ , 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, the temporal separation of observable choices can offer an identification strategy.<sup>44</sup> 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 the time superscript  $\tau$  to all random variables. Let  $M^\tau$  denote a vector of time- $\tau$  input prices and suppose that  $M^\tau$  follows the stochastic process

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

where  $\Phi$  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  $W^\tau$  is identified. Now suppose that  $X^\tau$  is chosen by firms in period  $\tau - 1$ , whereas prices for period  $\tau$  are chosen at time  $\tau$ . The causal graph in Figure 14 illustrates key features of such a model.<sup>45</sup> For clarity, here we include the prices  $P^\tau$  in the graph.

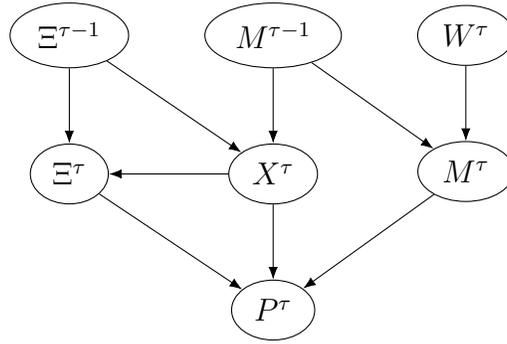
Here it is clear that neither the contemporaneous cost shifters  $M^\tau$  nor the lagged cost shifters  $M^{\tau-1}$  can serve as instruments for prices conditional on  $X^\tau$ :  $X^\tau$  would be a collider, exactly as in the previous section. However, the period- $\tau$  innovation  $W^\tau$  can do the job. Because  $W^\tau$  alters period- $\tau$  marginal cost, it is relevant for the determination of  $P^\tau$ , conditional on  $X^\tau$ . And, by the d-separation criterion, we see that  $W^\tau$  is independent of  $\Xi^\tau$  conditional on  $X^\tau$ . Indeed, this becomes particularly transparent if we again focus on the ancestral graph relevant for assessing this conditional independence, as in Figure 15. There we see that  $W^\tau$  is an example of a “fully exogenous instrument,” as discussed in section A.1.1. Ultimately, the innovation  $W^\tau$  is simply a cost shifter that is independent of all else. The important insight,

---

<sup>44</sup>Familiar examples in IO include strategies used by Olley and Pakes (1996), Akerberg, Caves, and Frazer (2006), and others in the literature on estimation of production functions.

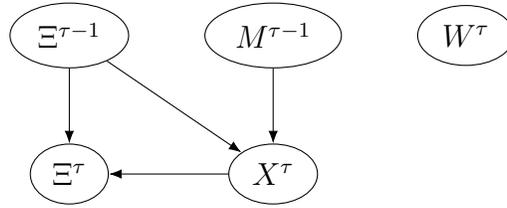
<sup>45</sup>The figure includes an edge from  $X^\tau$  to  $\Xi^\tau$ . The presence (or direction) of such an edge is not important to the argument here.

Figure 14



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  $\Xi^t$  conditional on  $X^t$ .

Figure 15



Similar arguments can allow construction of valid instruments from observed markup shifters (e.g., market-level demographics) whose lagged values affect firms' choices of  $X$ . Indeed, one may simply reinterpret  $M^t$  above as a time- $t$  markup shifter.

## 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 (2006): “Structural Identification of Production Functions,” working paper, Toronto.
- ADAO, R., A. COSTINOT, AND D. DONALDSON (2017): “Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade,” *American Economic Review*, 107, 633–89.
- 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.
- 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,” Tech. rep., CEMMaP Working Paper CWP60/17.
- 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. SATTERTHWAITTE (2003): “Competition and Market Power in Option Demand Markets,” *RAND Journal of Economics*, 34, 737–763.

- 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 (2021): “Testing Firm Conduct,” working paper, UW-Madison.
- 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.
- 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.
- HAUSMAN, J., G. LEONARD, AND J. ZONA (1994): “Competitive Analysis with Differentiated Products,” *Annales d’Economie et de Statistique*, 34, 159—180.
- HAUSMAN, J. A. (1996): “Valuation of New Goods under Perfect and Imperfect Competition,” in *The Economics of New Goods*, ed. by T. F. Bresnahan and R. J. Gordon, Chicago: University of Chicago Press, chap. 5, 209–248.

- HO, K. (2006): “The Welfare Effects of Restricted Hospital Choice in the U.S. Medical Care Market,” *Journal of Applied Econometrics*, 21, 1039–1079.
- (2009): “Insurer-Provider Networks in the Medical Care Market,” *American Economic Review*, 99, 393–430.
- HO, K. AND R. LEE (2016): “Insurer Competition in Health Care Markets,” *Econometrica*, forthcoming.
- HOM, M. (2018): “School Choice, Segregation and Access to Quality Schools: Evidence from Arizona,” Tech. rep., 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. (2000a): “Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry,” *RAND Journal of Economics*, 31, 395–421.
- (2000b): “A Practitioner’s Guide to Estimation of Random Coefficients Logit Models of Demand,” *Journal of Economics & Management Strategy*, 9, 513–548.
- (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, 69, 307–42.
- NEWKEY, 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,” *JPE*, 110, 705–729.
- 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.