discussed earlier. Second, we present various ways in which the supply side has been modeled in the literature.

In some of the applications we will combine the demand model with a supply model to estimate the parameters of both demand and supply and test the supply model, as we saw in the motivating example discussed in Section 2. There are several ways to combine demand and supply. First, we could use equation (2.3) to recover marginal costs without assuming a parametric functional form for costs. Given demand estimates, a model of pricing and observed prices, we can back out the marginal costs that make the first-order condition for prices hold exactly for each observation in the data. Note, that the model will perfectly fit the data and therefore we cannot test it, unless we bring in additional information, such as information about marginal costs, and therefore markups.

Second, we can parameterize the marginal cost function and estimate its parameters, potentially jointly with the demand equation using the demand parameters estimated in a first stage. The advantage of estimating the marginal cost function is that it allows us to extrapolate to counterfactual situations not observed in the data. It will also allow us to test among models of competition, as we saw in Section 2, and estimate parameters associated with the pricing model.

5.1 The Workhorse Model of Horizontal Competition

The workhorse supply model in the study of differentiated-products industries is the static pricing model, described in Section 2.1.1.[42] This model delivers a pricing equation given in equation (2.2), which can be used in a few different ways that we discuss below.

Berry et al. (1995) is a seminal paper in the development of equilibrium models of demand and supply in differentiated-products industries, and where much of the demand modeling discussed above was first developed. The authors were interested in understanding the impact of a voluntary export restraints placed on exports of automobiles from Japan to the United States. To study this question they developed and estimated a model of demand and supply in Berry et al. (1995) and applied it to this question in Berry et al. (1999).

The indirect utility in BLP is given by

$$u_{ijt} = x_{jt}\beta_i + \alpha \ln(y_i - p_{jt}) + \xi_{jt} + \varepsilon_{ijt}$$

---

[42]We will refer to this model interchangeably as the workhorse model and Nash-Bertrand model.

where y_i is income and all other variables are as previously defined. Note, that income enters the indirect utility non-linearly and therefore will not cancel, as we discussed on page 18. For estimation they assume that income is distributed log normal and estimate the parameters of the distribution directly from income data. Consumer heterogeneity is described by

$$\beta_i^{(k)} = \beta_0^{(k)} + \beta_\nu^{(k)} \nu_i^{(k)} \qquad \nu_i^{(k)} \sim N(0,1).$$

Finally, the utility from the outside option is given by

$$u_{i0t} = \alpha \ln(y_i) + \varepsilon_{i0t}.$$

They estimate the model using 20 years of annual national data on the sales of automobiles in the United States. The model is estimated using both demand and supply-side moments. The demand-side IVs are the ones we described in Section 4.3.1. They also add supply-side moments as we described in Section 4.3.2.

The paper delivers two sets of important results. First, the paper demonstrates the importance of controlling for the endogeneity of price in the aggregate demand equation derived from discrete choice. Nowadays this is taken as obvious – not surprisingly since the importance of accounting for the endogeneity of price has been confirmed repeatedly in numerous industries – but at the time it was questioned as empirically relevant. For example, Table 3 of Berry et al. (1995) shows that a Logit demand model estimated without correcting for endogeneity of price yields a large number of inelastic demand curves (1494/2217=67.3%), which is inconsistent with static profit maximization. Once they instrument for price that number drops below 1%.

Second, the paper demonstrates the ability of the demand model to yield reasonable substitution patterns. BLP present these in Tables 5-7 . For example, in Table 7 the authors compare the diversion to the outside good, i.e., the fraction of consumers who substitute to the outside good in response to a price increase as a fraction of all those who substitute away from a product, implied by the Logit model and the more flexible random coefficients model. For the Logit model, as expected, the diversion is roughly 90% for all cars, which is roughly the share of the outside good. This number is high in absolute value, but maybe more important is that it is roughly constant: if the price of a BMW 735i increases, consumers are equally likely to divert to the no-purchase option as consumers of a Mazda 323. This seems unreasonable (and is totally driven by the Logit assumptions.) The outside option includes all the choices that do not involve purchasing a new car,

such as buying a used car, not replacing an existing car or delaying purchase. Intuitively, consumers who purchase a car are more likely to buy another car than switch to the no-purchase decision in response to a price increase, compared to the average consumer. Furthermore, consumers who purchased a lower priced car are more likely to switch to the no-purchase option in response to a price increase. In contrast to the Logit model, the full model can capture these effects: the overall substitution to the outside good is lower and the more expensive the car the lower the number. For example, the diversion for Mazda 323 is roughly 27% while for BMW 735i is 10%.

In large part, the paper is able to deliver statistically significant estimates of the variation in random coefficients because the authors impose the supply/pricing equation in estimation. As we noted earlier, the supply equation, jointly with a functional form for marginal costs that ensures that these costs will be non-negative, puts significant restrictions on the demand estimates. For example, demand estimates that imply inelastic demand also imply negative marginal cost under many models of pricing. A functional form that imposes (a reasonable assumption) that all marginal cost are positive would prevent this from happening. If we believe strongly in the supply model it is efficient to impose it in estimation. However, as we discussed in Section 2, and as we discuss below, a large motivation for estimating demand and supply models is precisely to test the supply model rather than assume it.

The paper abstracts from some elements of the car industry. For example, cars are a durable good. Yet the demand system is static: consumers are not forward-looking in the sense that they anticipate future needs, nor do they take into account whether they own a car and if so which one, when they make a purchase decision. On the supply side, the model assumes that manufacturers set a uniform price to consumers: dealers play no role and there is no price discrimination. In reality, we know that cars are sold through dealers and this market exhibits significant variation across consumers in the price they pay. Furthermore, the pricing model is static while in reality prices might reflect inventory considerations, generating brand loyalty or other dynamic effects. For many questions it is fine to abstract away from these issues, yet for some questions these issues might be quite important. Indeed, the paper inspired a large literature that relaxed some of these modeling assumptions on both the demand and supply side. We discuss some of the papers in this literature in the rest of this chapter and many others are discussed in other chapters in this Handbook.

5.2 Distinguishing Between Models of Competition

In this subsection we look to expand the supply, or pricing, models we consider. We will have a dual goal of coming up with potentially more general, or more flexible, models that would allow us to explain different patterns of pricing (and markups), as well as testing the workhorse, Nash-Bertrand model, and finding ways to distinguish it from alternative models. For the most part, the alternative models will focus on models that are more "collusive" and therefore tend to imply higher markups. We also use the discussion below as an opportunity to discuss several empirical implementations of the demand model.

A natural place to start thinking about expanding beyond the workhorse model is to consider testing of the model. Testing will generally follow the ideas presented in Section 2. As we saw there, one way to test the model is to informally compare predictions of prices and costs from different models to patterns we see in the data, even if only at an aggregated level. Nevo (2001) does precisely that. He studies pricing in the ready-to-eat cereal industry. The industry, at the time he studies it, was characterized by high concentration (the top 3 firms had approximately 75% share, and the top 6 approximately 90%), "high" price-cost margins (approximated to be around 45%), large advertising to sales ratios (roughly 13%) and numerous introductions of brands (67 new brands introduced by top 6 firms in 1980's). These facts were used to claim that this is a perfect example of an industry where firms collude on pricing but compete on advertising and brand introductions.[43] The paper asks if prices observed in this industry, and the margins that were approximated, are consistent with collusive pricing? Specifically, the paper notes that seemingly high margins can be due to product differentiation and multi-product pricing (of substitute products) and are not necessarily indicative of collusion.

To separate the effect of collusion Nevo estimates a brand-level demand system, he then computes price-cost margins implied by different pricing models and chooses the pricing model that cannot reject the approximated margins of 45%. He finds that the "high" margins are consistent with Nash-Bertrand pricing by multi-product firms and therefore one does not need to rely on collusion to explain the "high" margins. In other words, the data cannot reject the Nash-Bertrand model.

The demand model he uses follows equation (3.3). He estimates the model using scanner data for the top 25 brands of cereal. The data are aggregated to the MSA-quarter

---

[43]For example, Scherer (1982) argues that "...the cereal industry's conduct fits well the model of price competition-avoiding, non-price competition-prone oligopoly" (p. 189)

level and he aggregates different SKUs to a brand. This results in 1,124 markets and 27,862 brand-quarter-MSA observations. He defines prices as total revenues divided by the number of servings, which he defines using the suggested servings size. Characteristics come from nutritional information (e.g., fat or sugar content), segment information used by the industry, and subjective information (e.g., he defines a "mushy" dummy variable). He estimates demand using the BLP algorithm and prices in other cities as IVs. Unlike BLP he estimates demand without imposing any supply-side moments, which has the advantage of yielding consistent estimates even if we are unsure about the supply model. He is able to do so because he has more markets than BLP and more variation in demographics across markets.

For the supply side he computes three models: single product, multi product and collusion, following the models discussed in Section 2.1.1. The markups implied by different supply and demand models are presented in Table 8 in the original paper. Using the random coefficients demand estimates he finds that the current ownership of the top 25 brands predicts an average margin of 42%, while joint ownership of these brands predicts a margin of 73% (with a confidence interval between 62% and 97%). By comparing these results to the approximation of the margins, which he estimates at 45%, he rejects the null of perfect collusion, or joint profit maximization of the top 25 brands, but cannot reject the null of Nash-Bertrand pricing.

Like BLP, this paper also raises several questions regarding the modeling assumptions. First, a common concern is whether demand is really discrete. In micro data, we often observe consumers purchasing more than one box of cereal, at times even different brands on the same trip. How do we reconcile this pattern with a discrete choice? One option is to think of the choice as happening at the time of consumption: at that point it seems more reasonable to assume a discrete choice. Thinking of the model this way ignores the two-step process, where the consumer first decides what to purchase at the store and then what to consume from the brands available at home , but with aggregate data it is unclear we can separate this process. In Section 6.1.1 we will discuss models that tackle the modeling and estimation of multiple choices.

The paper does not explicitly model retailer behavior despite using retailer prices to study manufacturer competition. This is consistent with retail margins going into the cost manufacturers pay to get the product on the shelf. This assumption, however, is not consistent with a strategic retailer who will change their margin in response to the

manufacturer pricing. In the next subsection, we will discuss how to add retailers to the model.

Finally, much of the price variation at the store level is coming from "sales", or temporary price reductions. This creates an incentive for consumers to purchase the products when the prices are low and consume them later. Follow-up work by Hendel and Nevo (2006a,b), which we discuss in Section 6.2.1, follows up on this issue.

One of the common uses of the supply model is to simulate the effects of a proposed merger. The idea was discussed by Berry and Pakes (1993) and Werden and Froeb (1994), and implemented empirically by Nevo (2000a). The basic idea is as follows. Using pre-merger data one can estimate demand and recover marginal costs. The marginal cost can be recovered as in BLP, by parameterizing the cost function and estimating it jointly with demand. Alternatively, marginal cost can be recovered, without making any parametric assumptions on the cost function, by "inverting" the pricing equation (2.2) such that

$$\hat{mc} = p - \hat{\Omega}^{-1}q, \tag{5.1}$$

where $\hat{\Omega}^{-1}$ is computed using the demand estimates and q are observed quantities. One then uses these estimates to simulate the effect of the merger by changing the ownership structure defined in equation (2.1). In general, holding costs constant a merger between substitutes products will lead to higher prices, and a merger between complements will lead to lower prices. The real issue is not the direction of the effect, but the magnitude: if the products are closer substitutes the effect will be larger. The reason to estimate demand is to be able to quantify the effect. Furthermore, one could also use the model to simulate the impact of various efficiencies such as reductions in marginal costs, or improvements in qualities. This approach has found support in the academic literature and in practice.[44]

Simulating the effects of mergers can be the object of interest, but can also serve as a way to test the model. An early attempt at this was provided by Peters (2006) who used comparisons between predicted and actual outcomes of airlines mergers to test the demand model. Miller and Weinberg (2017) study the beer industry and use a joint venture, which they treat as a merger, to test the Nash-Bertrand model. Their study is motivated by a desire to understand the price effects after the 2008 Miller-Coors joint venture (JV). The JV significantly increased concentration in the industry. After it, ABI (the producer

---

[44]For a discussion of the use in policy see Griffith and Nevo (2019), and for discussion of the use of merger simulation in a specific merger see Bayot et al. (2018).

of Budweiser) had a 37% market share, Miller-Coors had 29%, Modelo (the producer of Corona) had 9%, and Heineken had 5%.

They document, in Figure 1 (and Tables 2 and 3) of their paper, that post-JV (i) the price of Miller-Coors products increased; (2) the price of ABI increased almost one-to-one with the Miller-Coors price; and (3) the prices of Modelo and Heineken did not increase. The direction of the first fact is not surprising, since, as we discussed above, this is what a standard Nash-Bertrand model would predict. The magnitude might be higher than expected, but we cannot measure that without knowing the cross-price elasticities between the products. The fact that ABI prices increase should also not be surprising since prices are strategic complements in a pricing game for many (but not all) demand models. In these cases, the increase in the price of Miller-Coors would lead to an increase in the ABI price. What is surprising is that the ABI price increase is similar in magnitude to the Miller-Coors price increase. The paper sets to see if the demand estimates can reconcile these increases with what a competitive pricing model would predict.

The demand model they estimate is a random coefficients Nested Logit model, where all the products are in one nest and the outside good is in another. They estimate the model using monthly (or quarterly) scanner data for 13 brands in 39 different distinct regions, and the BLP nested fixed point algorithm discussed above. For IVs they use the distance between the brewery and the region, an indicator variable equal to one for ABI and Miller-Coors products after the merger, the number of products, mean income (by region) interacted with product characteristics. Their estimates (Table 4 in the paper) find statistically significant heterogeneity in preferences, but economically fairly small. The implied elasticities (presented in Table 5) show a slight variation from the pattern implied by the model with no heterogeneity: substitution between the inside goods is roughly proportional to share.

To estimate the supply side they specify marginal cost as a function of distance from the brewery, an indicator equal to 1 one for Miller-Coors products post-merger, and product region and period fixed effects. They use the model of Section 2.1.1 but slightly modify the definition of the ownership, given in equation (2.1). Specifically, they assume that the (j, k) element equals κ if products j and k are sold by ABI and Miller-Coors post-merger. This generates Nash–Bertrand competition in the post-merger periods if $\kappa = 0$ and joint profit maximization for ABI and Miller-Coors if $\kappa = 1$. Putting this together with equation

(4.12) yields the estimating equation

$$\boldsymbol{p}_t = \text{w}_t\boldsymbol{\gamma} + \Omega^{-1}(\kappa)\boldsymbol{q}(\boldsymbol{p_t}) + \boldsymbol{\omega}_t,$$

where variables are defined as in (4.12). They estimate the cost parameters and κ using GMM and assuming that ω_{jt} is mean independent of w_j and an indicator equal to one for ABI and Miller-Coors products post merger.[45]

They estimate that κ varies between 0.25 and 0.34, depending on the specification they use. All the specifications reject the null hypothesis that $\kappa = 0$, which is Nash-Bertrand pricing post merger. Interestingly, the model also rejects joint profit maximizing pricing, between ABI and Miller-Coors post merger. In other words, pricing is consistent with some increased coordination post merger but not joint profit maximization.

Ciliberto and Williams (2014) use a similar idea to study multi-market contact in the U.S. airline industry. They modify equation (2.1) by making the terms H_{jr} a function of a variable that measures the degree of multi-market contact between the airlines that produce products j and r. They conclude that airlines with a higher degree of multi market contact almost perfectly collude.

It is tempting to try to interpret intermediate values of κ. For example, one interpretation, which was popular for a while in the analysis of competition in homogeneous goods industries, views κ as a "conduct parameter" that captures beliefs about the equilibrium being played (Bresnahan, 1989). A slightly different interpretation views κ as an "as if" parameter. For example, if $\kappa = 0.5$ in a homogeneous good industry, the industry would be seen as being as competitive as an industry with 2 symmetric firms playing Cournot. Similar "as if" potential interpretations exist for industries with differentiated products. For example, building on Nevo (1998), Black et al. (2004) propose an interpretation of κ as a measure of cross ownership.[46]

These attempts to interpret intermediate values of κ have fallen mostly out of favor for several reasons. For example, Corts (1999) notes that if the true model implies variation in κ over time within each regime, the "as if" interpretation is problematic. Indeed, he shows via simulations that the estimated κ need not even be a good indicator of the relative competitiveness of industries. His main complaint is not about the idea of having

---

[45]Note that prices appear on both sides of the estimating equation, and therefore we need an (additional) IV, in this case an indicator equal to one for ABI and Miller-Coors products post merger.

[46]Backus et al. (2021) build on these ideas, and use data from the ready-to-eat cereal industry, to test recent claims that common ownership facilitates collusion.

a parameter κ but to the a-theoretical restrictions put on it. He calls for a specification of a (structural) model of collusion in order to justify restrictions on κ. Miller and Weinberg (2017) are partly immune to this critique because of their focus on testing whether κ is statistically different than zero as a test for post-merger Nash–Bertrand competition (since under the null the model is well specified).

Most of the above discussion of testing and distinguishing between models was somewhat informal. However, the ideas are can be formalized. Bresnahan (1982) and Lau (1982) discuss the use of rotation of the demand curve as an IV to estimate the model of competition in homogeneous goods industries. Berry and Haile (2014) show that the basic idea generalizes to models that study competition in differentiated products industries. Specifically they show that the model of competition can be identified using the supply-side conditional moments $E(\omega_{jt}|\boldsymbol{Z}_t) = 0$, as defined in (4.12). The intuition behind this is what we saw in the applications above: different models of competition have implications for patterns we observe in the data. The patterns can be either average markups (Nevo, 2001), variation in prices after a merger (Miller and Weinberg, 2017), or co-variation in prices and multi-market contact across markets Ciliberto and Williams (2014). The moment condition is a way to summarize this variation.

Casting the problem as a moment condition allows Berry and Haile (2014) to formally discuss identification. It also has the advantage of generalizing the idea of using a specific event, for example, a merger, to using variation like that discussed in Section 4.2.2, to distinguish models of competition. This raises the question of what IVs should be used. Active areas of research are what IVs have power to distinguish models of competition (Duarte et al., 2021) and somewhat related is what is the efficient ways to use these IVs (Backus et al., 2021).

5.3   Adding Retailers Into the Mix

One of the common complaints about some of the papers presented in the previous subsection is that they are are concerned with competition between manufacturers yet use retail data without explicitly modeling the retailers. This practice can be justified if we treat the retailers as passive and therefore part of the manufacturer costs of getting the products to market. In this subsection we discuss papers that allow retailers to make strategic decisions in response to price changes by the manufacturers.

Villas-Boas (2007) asks what model best describes the relationship between manufacturers and retailers in the yogurt industry. To answer this question she takes the idea of

using demand estimates jointly with a pricing model to recover cost, and adds to it an additional layer: her pricing models account for both retailer and manufacturer/wholesaler behavior. She computes markups/costs under different models. She then chooses the model that best describes the data.

To estimate demand she uses weekly scanner data for 43 products, produced by 5 manufacturers and sold by 3 retailers. She observes retail prices, advertising, aggregate quantity (by product-retailer), and product characteristics, but does not observe wholesale prices or contracts. She estimates a random coefficients Logit Model, where the choice is store-brand. In other words, consumers choose not just the brand but also where to purchase it.

She considers six supply models. In Model 1, manufacturers first set (linear) wholesale prices and then retailers, taking these wholesale prices as given, set retail prices. In each stage there is Nash-Bertrand pricing. This results in double marginalization. Formally, retailer r in market t maximizes profits given by

$$\pi_{rt} = \sum_{j \in \mathcal{J}_{rt}} [p_{jt} - p_{jt}^{\omega} - c_{jt}^r] q_{jt}(\boldsymbol{p}_t).$$

where p_{jt}^{ω} is the wholesale price paid by the retailer, c_{jt}^r is the retailer's marginal cost and $\mathcal{J}_{rt}$ is the set of products sold by the retailer. In her setting each "product" j is a brand-retailer combination and therefore this setting allows for different wholesale and retail prices to be charged for the same physical product. Rearranging the first-order conditions we can write

$$\boldsymbol{p}_t - \boldsymbol{p}_t^{\boldsymbol{\omega}} - \boldsymbol{c}_t^r = (\Omega^r)^{-1} \boldsymbol{q}(\boldsymbol{p}_t),$$

where Ω^r is a matrix with elements given by $\Omega_{jk}^r = -\partial q_k / \partial p_j \cdot H_{jk}^r$, where j indexes rows and k columns and H_{jk}^r is the retailer ownership structure, namely $H_{jk}^r = 1$ if both j and k are sold by r.

The manufacturers' problem is to maximize profits given by

$$\pi_{\omega t} = \sum_{j \in \mathcal{J}_{\omega t}} [p_{jt}^{\omega} - c_{jt}^{\omega}] q_{jt}(\boldsymbol{p}_t)$$

where c_{jt}^{ω} is the manufacturer's marginal cost and $\mathcal{J}_{\omega t}$ is the set of products produced by manufacturer ω. Rearranging yields the following first-order conditions

$$\boldsymbol{p}_t^{\boldsymbol{\omega}} - \boldsymbol{c}_t^{\boldsymbol{\omega}} = (\Omega^{\omega})^{-1} \boldsymbol{q}(\boldsymbol{p}_t),$$

where Ω^ω is a matrix with elements given by $\Omega^\omega_{jk} = -\partial q_k/\partial p^\omega_j \cdot H^\omega_{jk}$, where j indexes rows and k columns and H^ω_{jk} is the manufacturer ownership structure. Note, that $\partial q_k/\partial p^\omega_j$ can be computed using the derivatives of the retail prices with respect to wholesale prices (by totally differentiating the retailers first-order conditions with respect to wholesale prices) and the derivatives of quantity with respect to retail prices.

In Model 2 the national brands are as in Model 1. However, the private labels are treated as a vertically integrated (the "manufacturer" sets the retail price for them).

In Model 3 she explores non-linear pricing. She consider two cases.[47] In case 1, manufacturers set the wholesale price equal to cost, $p^\omega_t = c^\omega_t$ and set fees F, aimed at extracting the retailers' profits. In this case, the retail price is

$$\boldsymbol{p}_t - \boldsymbol{c}^\omega_t - \boldsymbol{c}^r_t = (\Omega^r)^{-1}\boldsymbol{q}(\boldsymbol{p}_t)$$

In case 2, retailers' margins are set to zero, $p_{jt} = p^\omega_{jt} + c^r_{jt}$ and the retailers set a fee F to recover profits. Prices are set to maximize downstream profits for the manufacturers

$$\boldsymbol{p}_t - \boldsymbol{c}^\omega_t - \boldsymbol{c}^r_t = (\Omega^{\omega r})^{-1}\boldsymbol{q}(\boldsymbol{p}_t)$$

where $\Omega^{\omega r}$ is a matrix with elements given by $\Omega^{\omega r}_{jk} = -\partial q_k/\partial p_j \cdot H^\omega_{jk}$. The difference between the two cases is in the "ownership" structure used. In case 1 the pricing internalizes cross-price effects across brands *within* a store, while in case 2 the pricing internalizes cross-price effects within a brand *across* stores.

Models 4-6 allow for coordination. Model 4 is like Model 1 but with manufacturer-level (perfect) collusion. Model 5 is like Model 1 but retail-level (perfect) collusion. Finally, Model 6 offers a version of a vertically integrated monopolist.

Estimation of demand follows the standard approach, using as IVs cost data multiplied by a product fixed effect (i.e., the costs are allowed to impact each product differently). To choose between the supply models she uses two approaches. In the first she regresses

$$\boldsymbol{p_t} = \mathrm{w}_t\gamma + SIPCM_r(\Omega)\lambda_r + SIPCM_\omega(\Omega)\lambda_\omega + \boldsymbol{\omega}_t,$$

where w are cost variables, γ is a vector of coefficients, $SIPCM_r(\Omega)$ and $SIPCM_\omega(\Omega)$ retail and manufacturer price-cost margins, respectively, implied by the different scenarios

---

[47]See Bonnet and Dubois (2010) for the assumptions required to justify the first-order conditions that follow as coming from two-part tariff contracts.

above. She then tests if $\hat{\lambda}_r$ and $\hat{\lambda}_\omega$ are statistically different from 1.[48] The logic is that in a well specified model the observed prices should equal marginal cost plus the wholesale and retail margins, i.e., the coefficients on markups should equal 1.[49] Next, she uses a statistical test of non-nested alternatives, as in Bresnahan (1987) and Gasmi et al. (1992), using the procedure in Smith (1992).[50] Based on these she concludes that models that assume zero wholesale margins and in which retailers have pricing decisions best fit the data. She then discusses several contracts that are consistent with these outcomes.[51]

The paper shows that the logic of backing out markups and costs, presented in Section 2 can be extended to another (vertical) level of decisions. Indeed, the pricing equations she proposes are direct extensions of the pricing equation we discussed in the previous two subsections.

5.4 Models of Bargaining

Up to this point the supply models we focused on dealt with situations where one side of the market makes a take-it-or-leave-it offer. In the basic model, it was the manufacturers making a take-it-or-leave-it offer to consumers or to retailers. In the previous subsection we allowed retailers to make the offer to the manufacturers. In many real-world situations neither of these fits what actually happens because, for example, the parties, say a manufacturer and a retailer, negotiate the final outcome rather than one side dictating the terms of trade and the other side simply accepting or rejecting the offer. Even if there is no explicit negotiation, a bargaining model seems like a good way to model pricing that is between either extreme of prices set by upstream or downstream firms.

To fix ideas, we focus on a specific example, the study by Crawford and Yurukoglu (2012) of bargaining in TV markets. Crawford and Yurukoglu study the impact on consumer welfare of (un)bundling of TV channels offered in cable bundles. Generally cable companies offer consumers a bundle of channels and consumers cannot subscribe

---

[48]Note that like our discussion of the supply model in Miller and Weinberg (2017), the parameters entering the pricing equation might require IVs to identify them because of the fact that the markup terms $SIPCM_r$ and $SIPCM_\omega$ might also be functions of prices and shares and therefore correlated with the econometric error ω_t.

[49]Pakes (2017) conducts a similar exercise. Specifically, he uses demand estimates from Wollmann (2018) to construct the markup term. He then regresses the estimated markup on IVs and regresses the observed price on the observed cost determinants and this predicted markup. He finds that the estimated coefficient is not statistically different than 1.

[50]See Rivers and Vuong (2002) and Duarte et al. (2021) for additional discussion of non-nested testing.

[51]See Bonnet and Dubois (2010) for follow up work that examines these issues.

to channels à la carte. Some have suggested that consumers are hurt by this arrangement because they have to pay for channels that they do not want. Crawford and Yurukoglu (2012) ask: what are the (equilibrium) welfare effects of unbundling? Holding channel prices constant, consumers will be better off with unbundling because they could purchase "skinnier" bundles. However, Crawford and Yurukoglu point out that this argument ignores equilibrium effects. Therefore, to answer the question of whether consumers are better or worse off with unbundling, they need to model the effect of unbundling on input prices, namely the prices paid to the content providers. To do so they develop an empirical bargaining model based on the theoretical model of Horn and Wolinsky (1988a,b).

They model the negotiation between each conglomerate of channels (e.g., ABC Disney, which owns ESPN, ESPN2, Disney channel and other channels) and the operator/distributor (e.g. Comcast), as a bilateral bargaining problem over linear (input) prices for that pair. The Nash bargaining solution is the set of input prices that maximize the weighted product of the values to both parties from agreement (as a function of that price) relative to the values without agreement. In general, the Nash bargaining solutions are interdependent as the value from one agreement depends on what happens in the other agreements. For example, if negotiations break down with one conglomerate the value of another conglomerate to the distributor might increase. To make progress (Horn and Wolinsky, 1988a,b), make a Nash equilibrium like assumption: they assume that each negotiating pair takes the outcome of other negotiations as given.[52] This setup is often called "Nash-in-Nash" bargaining. As Crawford and Yurukoglu point out, this is a strong assumption, since the members of each pair also participate in other negotiations and realize that whether an agreement is reached, and what is that agreement, will impact other negotiations. Furthermore, because the negotiating parties participate in other negotiations they will have asymmetric information. Nevertheless, in order to make progress much of the empirical work has relied on this assumption.[53]

Formally, let p_{rj}^{ω} be the wholesale price, or input cost, paid by distributor r for channel j, which is owned by conglomerate f. Note that the distributors in this setting are equivalent to downstream retailers and the conglomerates are the equivalent of upstream

---

[52]In bilateral negotiation, Rubinstein (1982) and Binmore et al. (1986) show that the Nash bargaining solution corresponds to the unique subgame perfect equilibrium of an alternating offers non-cooperative game. Collard-Wexler et al. (2019) provide conditions such that the Horn and Wolinsky solution is the same as the unique perfect Bayesian equilibrium with passive beliefs of a specific simultaneous alternating offers game with multiple parties.

[53]See Lee and Fong (2013), Ho and Lee (2019), Ghili (ming) and Liebman (2018) for alternatives.

manufacturers. Let $\boldsymbol{p}^\omega$ denote the vector of all wholesale prices, and let $\boldsymbol{p}^\omega_{rf}$ be the vector of wholesale prices paid for the channels of conglomerate f. The conglomerate and the distributor negotiate over the vector of wholesale prices, $\boldsymbol{p}^\omega_{rf}$. If negotiations break down the distributor will not have access to any of the conglomerate channels. The Nash bargaining solution determines the prices $\boldsymbol{p}^\omega_{rf}$, taking as given all other prices $\boldsymbol{p}^\omega_{-rf}$, to maximize distributor r's and conglomerate f's Nash Product, defined by

$$NP_{rf}(\boldsymbol{p}^\omega_{rf}; \boldsymbol{p}^\omega_{-rf}) = [\pi_r(\boldsymbol{p}^\omega_{rf}; \boldsymbol{p}^\omega_{-rf}) - \pi_r(\infty; \boldsymbol{p}^\omega_{-rf})]^{\zeta_{rf}} [\pi_f(\boldsymbol{p}^\omega_{rf}; \boldsymbol{p}^\omega_{-rf}) - \pi_f(\infty; \boldsymbol{p}^\omega_{-rf})]^{(1-\zeta_{rf})}, \quad (5.2)$$

where ζ_{rf} is a parameter that measures the (relative) bargaining power of r when bargaining with f, and π_r and π_f are the profit functions of r and f, respectively, when an agreement is reached and when it is not (denoted by $\boldsymbol{p}^\omega_{rf} = \infty$). Note, that the infinite price in the case of disagreement reflects that the distributor will not have access to any of the conglomerate channels. These profit functions are determined endogenously by parts of the model explained below.

Two key determinants of the bargaining outcome are the bargaining power, captured by ζ_{rf}, and the bargaining leverage, which is the loss from disagreeing relative to reaching an agreement. The bargaining parameter varies between zero, where f has all the bargaining power and the solution is equivalent to Nash-Bertrand pricing by the upstream providers, and one, where r has all the bargaining power and can impose its will (subject to a participation constraint by f). Values between the two extremes allow for different relative bargaining power.[54]

The leverage of the parties impacts the solution in a similar way: everything else equal the bargaining solution is closer to the optimal solution of the party with the greater leverage, namely the party with less to lose from a disagreement. For example, if distributor r is negotiating with two different conglomerates f and f' and the only difference is that

$$\pi_f(\boldsymbol{p}^\omega_{rf}; \boldsymbol{p}^\omega_{-rf}) - \pi_f(\infty; \boldsymbol{p}^\omega_{-rf}) > \pi_{f'}(\boldsymbol{p}^\omega_{rf'}; \boldsymbol{p}^\omega_{-rf'}) - \pi_{f'}(\infty; \boldsymbol{p}^\omega_{-rf'}).$$

Namely, that relative to f', f has more to gain from an agreement, or more to lose from a disagreement, then f' has more leverage and will obtain an outcome that is more favorable from its prospective.

---

[54]To formally see these claims, note that the maximization problem is equivalent to maximizing the log of the Nash product. Taking the first-order condition of the log of the Nash product and slightly rearranging we see that the resulting first-order condition is just a weighted average of the maximization problems of r and f, with the weights being a product of the relative bargaining leverage and bargaining power.

To compute the profit functions of both the conglomerate and the distributors Crawford and Yurukoglu estimate a viewership model and bundle (subscription) choice model. They then back out the implied input cost from the bundle pricing equation and use it to fit a parametric cost function for the distributors. The details of their modeling (and estimation) are a bit different from, for example, BLP, but the principle is the same: the pricing equation and demand estimates are used to back out input costs or estimate an input cost function. These input costs can then be used, the same way prices were used in the basic supply model, to either recover costs (here these are upstream costs) or parameters of the supply model (here bargaining parameters).

In their application, Crawford and Yurukoglu assume that upstream marginal costs are zero and therefore they estimate the bargaining parameters. They do so by choosing the bargaining parameters, ζ_{rf} that minimize the difference between the programming costs they recovered from the pricing and demand equations and those predicted by the bargaining model. One way to implement this idea is to write

$$\hat{p}_{rj}^{\omega} = p_{rj}^{\omega*}(\zeta_{rf(j)}) + \epsilon_{rj} \tag{5.3}$$

where $\hat{p}_{rj}^{\omega}$ is the wholesale price backed out from the pricing equation, $p_{rj}^{\omega*}(\zeta_{rf(j)})$ is the wholesale price that maximizes the Nash product defined in (5.2) and ϵ_{rj} is measurement error, or an "add-on" error term. The parameters ζ can be estimated using, for example, non-linear least squares.

Note, that the bargaining parameters vary by conglomerate and not channel and therefore in principle are of lower dimension than the wholesale prices. In practice, Crawford and Yurukoglu (2012) do not have enough variation to meaningfully estimate channel-level wholesale prices and instead estimate these at the conglomerate-level. This is of little economic significance because the way they set up the bargaining model all that matters is the total payment to each conglomerate. Furthermore, for computation reasons they solve the model for a market with "typical" distributors. The bottom line is that they have the same number of parameters as observations in the above equation and therefore can compute the set of bargaining parameters that set the backed out input cost exactly equal to those predicted by the bargaining model. This directly parallels the non-parametric approach to recovering cost discussed in the Introduction to this section. In this approach some view the bargaining parameters as the "error terms."

In other settings (Grennan, 2013; Bagwell et al., 2020), the number of observations in (5.3) is larger than the number of bargaining parameters and then those are estimated,

for example, by non-linear least squares.[55] In principle, the bargaining parameters can be estimated jointly with the pricing and demand parameters, but computationally it is easier to estimate them separately.

Using the estimated parameters Crawford and Yurukoglu are able to simulate the model both with bundling, as in the data, and when the consumers are offered channels à la carte. They find that equilibrium effects are significant: the counterfactuals suggest that à la carte offering of channels will increase input prices enough to offset consumer benefits from being able to choose only the channels they value most. Thus, unbundling will decrease welfare.

It is possible to set up, at least for some bargaining problems, a "structural" error term that follows a bit more closely the logic of the model discussed in Sections 5.1-5.3. This approach especially makes sense when the upstream costs are unknown and need to be estimated: it seems natural to assume that the error term consists of unobserved variation in marginal costs. Gowrisankaran et al. (2015) do precisely that.[56] They study the effects of mergers when prices are negotiated using data from negotiations between hospitals and health insurance companies, or managed care organizations (MCOs), to demonstrate the effect. Following Town and Vistnes (2001) and Capps et al. (2003) they set up the bargaining problems as follows. Let r denote the MCO, and f denote a system of hospitals that includes hospitals $j \in \mathcal{J}_f$. The negotiation is over the price that the MCO will pay the hospital if one of its enrollees receives care in that hospital.

In this setting the MCO is the equivalent of the retailer or distributor, and the hospital system is the equivalent of the manufacturer or conglomerate. The Nash product for this problem follows equation (5.2), with a small difference that the profit functions, and therefore the Nash product, are a function of wholesale prices and the "network" (i.e., which MCO-hospital system pairs that reach an agreement).[57] Specifically, hospital system, f profit from an agreement relative to disagreement is given by

$$\sum_{j \in \mathcal{J}_f} q_{rj}(\mathcal{N}, \boldsymbol{p}^\omega)[p_{rj}^\omega - mc_{rj}], \tag{5.4}$$

---

[55]For consistency of the estimates, generally, the error term in (5.3) cannot appear inside $\boldsymbol{p}^{\omega*}(\cdot)$ and therefore needs to be an "add-on" error term that is usually interpreted as "measurement error".

[56]Grennan (2013) also estimates costs parameters, but his econometric error term is in the bargaining parameter.

[57]This is not a meaningful distinction because an infinite wholesale price can denote disagreement. The main reason for the change in notation is to be more explicit about networks, which are of greater importance in health care where not all MCOs contract with all hospital systems.

where p_{rj}^ω is the price r pays f if one of its enrollees is treated in hospital j, $\boldsymbol{p}^\omega$ is the collection of all prices, $\mathcal{N}$ is a description of the network (i.e., which hospitals have contracts), $q_{rj}(\mathcal{N}, \boldsymbol{p}^\omega)$ are the patients that go to hospital j given the network and the vector of prices, mc_{rj} is the marginal cost of treating those patients (which are assumed (i) constant in quantity; and (ii) to vary by MCO to allow for different composition of enrollees.) This formulation assumes that patients do not go out of network and will not switch insurance providers in the case of disagreement.

On the MCO side, the value of an agreement is given by $\pi_r(\mathcal{N}_r, \boldsymbol{p}^\omega) - \pi_r(\mathcal{N}_r \setminus \mathcal{J}_f, \boldsymbol{p}^\omega)$. $\pi_r(\cdot)$ is the value to the MCO of having a network (with and without system f). Note, that as before, a breakdown in negotiation means the whole system leaves the network. The paper builds $\pi_r(\cdot)$ from micro-foundation based on consumers' willingness-to-pay for an insurance product that has a wider network.

They show that one can write the first-order conditions of the bargaining problem as

$$(1 - \zeta_{rf}) \frac{q_{rj} + \sum_{k \in \mathcal{J}_f} \frac{\partial q_{rk}}{\partial p_{rj}^\omega} [p_{rk}^\omega - mc_{rk}]}{\sum_{k \in \mathcal{J}_f} q_{rk} [p_{rk}^\omega - mc_{rk}]} = -\zeta_{rf} \underbrace{\frac{\overbrace{\frac{\partial \pi_r(\mathcal{N}_r, \boldsymbol{p}^\omega)}{\partial p_{rj}^\omega}}^{A}}{\pi_r(\mathcal{N}_r, \boldsymbol{p}^\omega) - \pi_r(\mathcal{N}_r \setminus \mathcal{J}_f, \boldsymbol{p}^\omega)}}_{B}. \tag{5.5}$$

The assumption of constant marginal costs implies that the first-order conditions in equation (5.5) are separable across MCOs and therefore one can rearrange the joint system of $\#(\mathcal{J}_f)$ first-order conditions from (5.5) to write

$$\boldsymbol{q} + \Omega(\boldsymbol{p} - \boldsymbol{mc}) = -\Lambda(\boldsymbol{p} - \boldsymbol{mc}) \tag{5.6}$$

where Ω and Λ are both $\#(\mathcal{J}_f) \times \#(\mathcal{J}_f)$ size matrices, with elements $\Omega_{jk} = \frac{\partial q_{rk}}{\partial p_{rj}^\omega}$ and $\Lambda_{jk} = \frac{\zeta_{rf}}{1 - \zeta_{rf}} \frac{A}{B} q_{rk}$. Solving for the equilibrium prices yields

$$\boldsymbol{p} = \boldsymbol{mc} - (\Omega + \Lambda)^{-1} \boldsymbol{q}, \tag{5.7}$$

where $\boldsymbol{p}$, $\boldsymbol{mc}$ and $\boldsymbol{q}$ denote the price, marginal cost and quantity vectors respectively for an MCO r across the different hospitals. Equation (5.7), which characterizes the equilibrium prices, has a form almost identical to standard pricing games, but differs in the inclusion of Λ. One case where $\Lambda = 0$ – and hence there is differentiated-products Nash-

66

Bertrand pricing with individual prices for each MCO – is where hospitals have all the bargaining weight, $\zeta_{rf} = 0, \forall f$.

Importantly, equation (5.7) shows that, as with static differentiated-products pricing models, we can back out implied marginal costs for the bargaining model as a closed-form function of prices, quantities and derivatives, given MCO and patient incentives. Furthermore, one can combine equation (5.7) with an assumption about cost, as in Section 4.3.2,

$$mc_{rj} = \mathbf{w}_{rj}\boldsymbol{\gamma} + \omega_{rj}$$

to form a basis for estimation where ω_{rj} is the structural error term and can be used to form a GMM objective function, as in equation (4.12).

The above discussion makes it clear that bargaining models can be specified in a variety of ways that impact formulation as well as identification and estimation. Our discussion is just an introduction to the topic, which is further explored by Lee et al. (2021) in this Handbook.

6 Extensions of the Demand Model

In this section we discuss some of the extensions of the discrete choice demand model discussed in Section 3. We start by outlining some extensions to the static model and then discuss models of dynamic demand.

6.1 Extensions to the Static Demand Model

A reasonably large literature has explored several extensions to the basic discrete choice models discussed in Section 3. The motivation for these extensions varies. Some extensions are motivated by a desire to better marry the model with patterns observed in micro data, and potentially explore the biases that would arise if we ignore these differences. Other extensions are needed to address specific questions that the basic model could not address. In many cases, the extensions are estimated using individual data, and therefore might not generalize to the various applications previously discussed.

Here we focus on two such extensions: multiple goods and a more general characteristics model. There are other extensions that have been considered in the literature. For example, Dubin and McFadden (1984) and Hanemann (1984) propose a model where consumers make a discrete choice, say choose an appliance, followed by a continuous

choice, for example, how much to use that appliance. The two decisions are obviously interlinked. These ideas have been used more recently in looking, for example, at store choice (Smith, 2004; Thomassen et al., 2017) and in choice of internet service plan and usage (Malone et al., 2020; Nevo et al., 2016).

Other areas that have received considerable attention are models where consumers do not know the characteristics of the products and have to learn them through search. See Honka et al. (2019) for a survey of this literature.

6.1.1 Multiple goods

A common complaint about discrete choice models is that actual purchases, when they are observed, are not of a single product or a single unit. For example, when studying the purchase of cereal we might observe consumers buying, on a single shopping trip, several boxes of the same cereal brand or buying several brands. One way to rationalize the multiple choices is to assume a single purchase instance, which is what we observe, is an aggregation over several consumption instances, which is what we are modeling. For example, a consumer shopping in a store might be buying for several household members for a week. If we assume that there are 3 household members, each making a daily consumption choice then we can view the single observed purchase as aggregation over 21 consumption choice instances. This can help rationalize purchase of multiple units, as well as different brands, and is probably sufficient when working with aggregate data where the individual choices are not observed anyway and we are already aggregating across consumers. However, this explanation is somewhat unappealing when working with individual-level data, in part because it assumes the choices across days are independent. Therefore, a somewhat better alternative is to more explicitly build the purchase model from underlying choices.

One way to model multiple choices is to redefine what we mean by an option. For example, assume that there are two brands, A and B, and consumers can buy no more than 2 units of each. This gives 9 options, ranging from not buying either brand to buying 2 of each brand. The utility from each option could in principle, be additive, in the utility from each brand separately or have interactions. For instance if buying one unit of each brand there could be an interaction term, which can be positive or negative, so that the utility is not just the sum of the utilities of buying each brand.

Gentzkow (2007) takes this approach when studying consumers' choice between print and online newspapers. He allows for purchase of more than one option and estimates

an interaction term in the utility function (to understand if print and online newspapers are complements of substitutes.) In his setting, consumers choose between the printed version of a newspaper, the online version, both, or neither. Thus, the choice is a discrete choice between bundles. He further assumes that each bundle receives an i.i.d. shock, as in the standard discrete choice model, and therefore he can estimate the model using standard methods. Note, that this assumption requires that any correlation between options, which include the same products, is captured through the rest of the utility terms. Furthermore, this approach is feasible when the number of choices is small, since the number of bundles increases exponentially with the number of products. For example, with $J = 20$ different options and even if the consumer can only choose at most one of each there are $2^{20} = 1,048,576$ different bundles available.

Another approach, proposed by Hendel (1999), and later also used by Dubé (2004) builds the demand explicitly from underlying tasks. Specifically, Hendel observes firms simultaneously buying several brands of computers and several units of each brand. To model this, he assumes that the firm has several tasks to do. For each task there is an optimal choice, but the observed purchases are an aggregation over several tasks. He does not allow for interaction in the utility from the different choices. He explains the purchase of several units, of the same computer, by a decreasing marginal utility from quantity, hence there is interaction in this dimension.

Fan (2013) studies mergers in the newspaper industry. She allows consumers to buy more than a single newspaper (but at most one copy of each) and introduces a parameter that measures the decrease in utility from a newspaper if it is bought "second". Therefore, like Gentzkow (2007) she allows for an interaction in utility, but at the same time, like Hendel (1999), does not treat the bundle as an independent option.

Nevo et al. (2005) study the decision of libraries to subscribe a subset of the 150 or so to Economics and Business journals that they observe. The libraries in their data subscribe to some subset of these 150 journals, but the subsets are not nested. If they were nested, we could model the choice of a bundle as a choice of how many journals to purchase. Instead, they model the choice problem as ranking journals by an index like that standard index in discrete choice models, for example the one given in equation (3.3). Unlike the standard discrete choice problem where the decision maker only chooses the top option, here the decision maker subscribes to journals following this ranking until a (budget or other) constraint is met. The utility from the journals does not interact (i.e., the utility does not depend on what other products are in the bundle), but the interaction is through

a (budget) constraint. They show how one could estimate this model using library-level subscriptions.

6.1.2 General Characteristics Demand Models

In Section 3.1, when we discussed the various demand models, we separated models in product space from models in characteristics space. For models in characteristics space we focused on discrete choice models. The idea of using characteristics to reduce the dimension of the estimation problem can be useful more generally. This turns out to be especially helpful when trying to estimate demand for more than a single product category. As such, this model offers an additional way to deal with multiple purchases.

Dubois et al. (2014) study such a problem.[58] They note, looking at household-level purchase data from France, the United Kingdom and the United States, differences across countries in the choices households make and in the prices and product offerings they face. These differences amount to large difference in the nutritional intake. For example, the average French consumer purchases roughly 200 calories less a day than the average American consumer. They ask the extent to which cross-country differences in purchases are attributable to differences in prices and the characteristics of products (the economic environment). To address this question they develop a model in characteristics space, based on Gorman (1980) and Lancaster (1966), that does not assume discrete choice. Their motivation is to use the richness of their data, which include disaggregated purchases, while still looking at the choice of a food "bundle" rather than narrowly defined products (e.g., soft drinks).

Specifically, a consumer i, with demographics D_i, chooses from J products, where product j is characterized by K characteristics $\{a_{j1}, .., a_{jK}\}$ and K is much smaller than J. These characteristics will in principle include both observed characteristics, like calories and protein in their example, and unobserved characteristics, which we previously denoted by ξ. The utility is given by $U(q_0, \boldsymbol{x}, \boldsymbol{q}; D_i)$ where q_0 is the amount of the numeraire good consumed, $\boldsymbol{x}$ is a $K \times 1$ vector of characteristics of food consumed by the consumer across all the products and $\boldsymbol{q}$ is a $J \times 1$ vector of the quantities purchased of all food products. Define the $J \times K$ matrix $\boldsymbol{A} \equiv \{a_{jk}\}_{j=1,..,J,k=1,..,K}$. This matrix will be

---

[58]See also Pinkse et al. (2002) who study competition between gas stations, and Pinkse and Slade (2004) who study the beer market. In both cases they model demand using a system in products space, and model the substitution matrix as a function of either physical distance or distance in attribute space. The resulting system is a statistical mapping between price and quantities, but is not derived from a well defined utility function nor is it guaranteed to be consistent with utility maximization.

used to transform the quantities of products purchased, given by $\boldsymbol{q}$ into the characteristics that they contain.[59] The household maximizes utility by choosing $\boldsymbol{q}$, subject to a budget constraint:

$$\max_{\boldsymbol{q}} \ U(q_0, \boldsymbol{x}, \boldsymbol{q}; D_i)$$

$$s.t. \ \ \sum_{j=0}^{J} q_j p_j \le y_i \ ; \ \ \ \boldsymbol{x} = \boldsymbol{A}'\boldsymbol{q}; \ \ q_j \ge 0, j = 0, 1..., J$$

where p_j is the price of one unit of product j, and y_i is the household's income.

Following standard arguments (and dropping the i subscripts) this can be written as

$$\max_{\boldsymbol{q}} \ U\left(\frac{y - \boldsymbol{p}'\boldsymbol{q}}{p_0}, \boldsymbol{A}'\boldsymbol{q}, \boldsymbol{q}\right)$$

$$s.t. \ \ q_j \ge 0.$$

Assuming that quantities $\{q_j\}_{j=0}^{J}$ are continuous, the first-order conditions are

$$\sum_{k=1}^{K} a_{jk} \frac{\partial U}{\partial x_k} - \frac{\partial U}{\partial q_0} \frac{p_j}{p_0} + \frac{\partial U}{\partial q_j} = 0 \ \ \text{if} \ \ q_j > 0.$$

This model nests various standard models. First, suppose the utility function is $U(q_0, \boldsymbol{x})$, which is the case in discrete choice models or in hedonic models (Court, 1939; Griliches, 1961; Rosen, 1974; Epple, 1987). Because the transformation from products to characteristics is linear and in this case $\partial U/\partial q_j = 0$, at most K of the J products would be purchased. If we restrict $q_j \in \{0, 1\}$ and $\sum_{j=1}^{J} q_j \le 1$, the model collapses to the standard discrete choice model. In general, the prediction that at most K products are purchased is a problem since we would like to consider cases where the number of products chosen is (much) greater than the number of observed characteristics.

Alternatively, if the utility function is $U(q_0, \boldsymbol{q})$ then we can generate standard demand systems in product space. If we allow for a characteristic that is product-specific then a model in characteristics space is equivalent to a model in product space, as long as the characteristics do not vary over time or markets. Note, that we need more than just different values on a small number of unobserved characteristics, but a totally different characteristic that can *only* be obtained from each product. To better understand the role of the characteristics in this model we rewrite the first-order conditions for j such that

---

[59]This linear transformation is quite natural for characteristics like calories, sugar and protein, where it seems natural to aggregate across products. The linear transformation is less intuitive when it comes to characteristics like ξ in (3.3) that represent demand shocks more generally.

$q_j > 0$ as

$$\frac{\partial U/\partial q_j}{\partial U/\partial q_0} = \frac{p_j}{p_0} - \sum_{k=1}^{K} a_{jk} \frac{\partial U/\partial x_k}{\partial U/\partial q_0}.$$

Consider the case where characteristics do not enter the utility, i.e., $\partial U/\partial x_k = 0$. The first-order conditions, in this case $\frac{\partial U/\partial q_j}{\partial U/\partial q_0} = \frac{p_j}{p_0}$, which implicitly defines the demand correspondence. A similar idea applies in the above model. Demand depends on the *hedonic prices* of each good instead of prices. The hedonic prices, $\frac{p_j}{p_0} - \sum_{k=1}^{K} a_{jk} \frac{\partial U/\partial x_k}{\partial U/\partial q_0}$, depend on the marginal utility of the consumer from the characteristics. In other words, if two products have the same price but one has more of a characteristic, with a positive marginal utility, then the effective price to the consumer will be lower for the product with the higher value of the characteristic.

To estimate the model Dubois et al. (2014) focus on a particular functional form for utility and use rich household level data from France, the United Kingdom and the United States. They use the estimated parameters to decompose the cross-country differences into variation coming from differences in preferences and variation coming from the differences in the economic environment.[60]

6.2 Dynamic Demand

Up to this point the models we discussed were static. In this section we explore the implications of dynamic demand. Dynamics in demand can arise for a variety of reasons, including switching costs, learning, storable and durable products, non-separable (over time) utility, pricing (e.g., monthly usage caps) that creates dynamic linkage, and other reasons. Here we will mostly focus on storable and durable products. We will discuss patterns in the data that suggest that dynamic demand is relevant, and discuss the implications of ignoring dynamics. We will then review some of the main modeling and estimation challenges and solutions that have been offered in the literature.

6.2.1 Storable Products

Many of the products that have been studied using the methods presented earlier in this section are storable: in the sense that consumers can buy them in one period and consume in another. Furthermore, a typical pricing pattern in these markets involves short lived price reductions ("sales"), with a return to the regular price. This pattern of prices

---

[60]Allcott et al. (2019) use a similar demand system, estimated somewhat differently, to study food deserts in the U.S.

generates an incentive for consumers to purchase the product when the price is low and store it for future consumption. Boizot et al. (2001) and Pesendorfer (2002) were among the first to study the effects of temporary price reductions and storability in the economics literature.

The empirical literature has found evidence consistent with the theoretical idea that consumers have incentives to purchase not for immediate consumption but for inventory. For example, Pesendorfer (2002), and Hendel and Nevo (2006b, 2013) find that aggregate quantity sold depends on duration from previous sale, controlling for the current price. Hendel and Nevo (2006b) show that a household's likelihood of purchasing during a temporary price reduction is correlated with proxies of storage costs. They also show that when a household purchases during a temporary price reduction the duration to the next purchase is longer (consistent with the household buying for storage rather than immediate consumption), and households who purchase more often on sale also purchase less frequently overall (again consistent with the idea that these households store their purchases).

Stockpiling behavior has several implications for demand estimates and how they should be used. If consumers purchase for storage, and the evidence suggest that they do, then there is a difference between the short-run response to a temporary price change, and the long-run response to either a temporary or permanent price change. For most economic applications we care about long-run changes. In many data sets temporary price changes account for most of the observed variation in prices. Short-run responses to temporary price reductions, interpreted through a static model, overestimate the long-term own-price effects. Typically, there is a large response to a temporary price reduction, which in a static model is attributed to an increase in consumption (which in a static model equals purchase), and not to an increase in storage. Similarly a post price reduction dip, also often observed in the data, coincides with an increase in price, and is mis-attributed by a static model as a decline in consumption. At the same time, static estimation underestimates cross-price effects: the temporary price reduction diverts current sales from competing products, but it also diverts future sales (and past sales to the extent that the reduction was at least partially anticipated). A static model misses these additional effects and therefore underestimates the impact on other products.

The basic model of consumer stockpiling is an inventory model. Hendel and Nevo (2006b) propose such a model for a homogeneous good, which abstracts from product differentiation and assumes that purchases are of continuous quantities.

The per period utility consumer i obtains from consuming at time t is

$$u_i(c_t + \nu_t) + \alpha_i q_{0t}, \tag{6.1}$$

where c_t is the quantity consumed in t, ν_t is a shock to utility that changes the consumption needs of the consumer, and q_{0t} is the numeraire good consumed at time t. Facing random prices, p_t, in each period the consumer has to decide how much to buy, denoted by q_t, and how much to consume, denoted by c_t. The consumer's problem can therefore be represented as

$$V(\boldsymbol{s}_t) = max_{(c_t(\cdot), q_t(\cdot))} \sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E} \left[u(c_t + \nu_t) - C(i_t) + \alpha p_t q_t \mid \boldsymbol{s}_t \right] \tag{6.2}$$

$$s.t. \quad 0 \le i_t, \quad 0 \le c_t \quad 0 \le q_t \quad i_t = i_{t-1} + q_t - c_t,$$

where α is the marginal utility from income, δ is the discount factor, and $C(i_t)$ is the cost of storing inventory. The information set, or state space, at time t, $\boldsymbol{s}_t$, consists of the current inventory, i_t, current prices, and the current shock to utility from consumption, ν_t. Consumers face two sources of uncertainty: utility shocks and future prices. Hendel and Nevo assume that shocks to utility, ν_t, are i.i.d. over time. Prices are assumed to evolve according to a first-order Markov process and take on two states, sale and non-sale.

They show that within this setup the optimal consumer behavior is characterized by a trigger s, and a target inventory S. The target, S, is a decreasing function of current price. On the other hand, the trigger, s, which is the sum of the target and current consumption, depends on prices and the utility shock. They also show that the quantity purchased is a function of lagged inventory, the current prices and the current utility shock.

A key challenge is how to expand this model to differentiated products. In principle, this is simple: we can just make all the quantities in (6.1) vectors. This, of course, reintroduces the dimensionality problem discussed in Section 3, except now the problem is even worse and includes the well-known "curse of dimensionality" in dynamic problems (Rust, 1994). To deal with these issues, Hendel and Nevo (2006a) propose the following model.[61] Now consumer i can purchase one of $J + 1$ brands (including a no purchase decision), which come in different sizes, indexed by $q \in \{1, 2, ..., Q\}$. Let d_{jqt} equal to 1 if the consumer purchases brand j of size q at time t, and 0 otherwise. Consumers make a

---

[61]See Aguirregabiria and Nevo (2013) for a more detailed discussion.