13 Endogeneity

13.1 Overview

So far we have assumed that the explanatory variables that enter a discrete choice model are independent of the unobserved factors. In many situations, however, the explanatory variables are endogenous, that is, are correlated or otherwise not independent of the unobserved factors. Examples include the following:

1. *Unobserved attributes of a product can affect its price.*
   In modeling consumers' choices among products, it might be impossible to measure all of the relevant attributes of the various products. In the case of cars, for example, the researcher can obtain information about the fuel efficiency, length, width, horsepower, weight, and many other attributes of each car that is offered by manufacturers, but attributes such as comfort, beauty of the design, smoothness of the ride, handling in curves, expected resale value, and prestige cannot be measured directly. Yet the price of the product can be expected to reflect these unobserved (i.e., unmeasured) attributes. There are two reasons why price is affected. First, insofar as the unobserved attributes are costly for the manufacturer, the price of the product can be expected to reflect these costs. Second, insofar as the unobserved attributes affect demand for the product, a price that is determined by the interaction of demand and supply can be expected to reflect these differences in demand. The end result is that price is correlated with unobserved attributes, rather than being independent as we have assumed so far in this book.
2. *Marketing efforts can be related to prices.*
   Advertising and sales promotions, such as coupons and discounts, are ubiquitous. Often the marketing practices of firms create a correlation between the price of products, which the researcher observes, and nonprice promotional activities, which the researcher generally cannot measure directly. The

correlation can go in either direction. A product might be promoted through an advertising blitz coupled with discounts. The advertising is then negatively correlated with price: greater advertising occurs simultaneously with lower prices. Alternatively, firms can raise the price of their products to pay for the advertising, which creates a positive correlation. In either case, the price of the product is no longer independent of the unobserved factors affecting consumers' choices.

3. *Interrelated choices of decision makers.*
   In many situations, the observed factors that affect one choice that a person makes are determined by another choice by that person. Travel mode and housing location are an important example. In a standard mode choice model for the commute to work, the observed explanatory variables are usually the cost and time of travel from home to work on each mode (car, bus, and train). However, people who tend to like traveling on public transit (or dislike it less than the average person) might also tend to buy or rent homes that are near public transit. Travel time by transit is therefore lower for these people than for people who locate further from transit. Stated in terms of the observed and unobserved factors in the mode choice model, unobserved attitudes toward public transit, which affect mode choice but cannot be measured completely by the researcher, are (negatively) correlated with the observed time of travel by public transit.

In situations such as these, estimation without regard to the correlation between observed and unobserved factors is inconsistent. The direction of bias can often be determined logically. For example, if desirable unobserved attributes are positively correlated with price, then estimation without regard to this correlation will result in an estimated price coefficient that is biased downward in magnitude. The reason is clear: since higher prices are associated with desirable attributes, consumers avoid the higher-priced products *less* than they would if the higher prices occurred without any compensating change in unobserved attributes. Essentially, the estimated price coefficient picks up both the price effect (which is negative) and the effect of the desirable unobserved attributes (which is positive), with the latter muting the former. A similar bias, though in the opposite direction, occurs if marketing consists of advertising coupled with price discounts. The increased demand for marketed products comes from both the lower price and the nonprice advertising. The estimated price coefficient picks up both effects, which is greater than the impact of lower prices by themselves.

Several methods have been developed to estimate choice models in the presence of endogenous explanatory variables. In this chapter, we describe these methods, delineating the advantages and limitations of each approach. We first describe the BLP approach, developed by Berry, Levinsohn, and Pakes (hence the initials) through a series of publications. Berry (1994) pointed out that constants can be included in the choice model to capture the average effect of product attributes (both observed and unobserved). The estimated constants can then be regressed against the observed attributes in a linear regression, where endogeneity is handled in the usual way by instrumental variables estimation of the linear regression. Essentially, he showed that the endogeneity could be taken out of the choice model, which is inherently nonlinear, and put into a linear regression model, where endogeneity can be handled through standard instrumental variables estimation. To apply this method, it is often necessary to estimate a very large number of constants in the choice model, which can be difficult using standard gradient-based methods for maximization. To address this issue, BLP (1995) provided a procedure, called "the contraction," that facilitates estimation of these constants. These two early papers were based on aggregate models, that is, models estimated on aggregate share data. However, the concepts are applicable to individual-level choice data or a combination of aggregate and individual-level data, as utilized in a later paper by BLP (2004). Applications of the BLP approach include Nevo (2001), Petrin (2002), Goolsbee and Petrin (2004), Chintagunta, Dubé, and Goh (2005), and Train and Winston (2007), to name only a few. A Bayesian form of the procedure has been developed by Yang, Chen, and Allenby (2003) and Jiang, Manchanda, and Rossi (2007).

The second procedure that we describe is the control function approach. The concepts motivating this approach date back to Heckman (1978) and Hausman (1978), though the first use of the term "control function" seems to have been by Heckman and Robb (1985). Endogeneity arises when observed variables are correlated with unobserved factors. This correlation implies that the unobserved factors conditional on the observed variables do not have a zero mean, as is usually required for standard estimation. A control function is a variable that captures this conditional mean, essentially "controlling" for the correlation. Rivers and Voung (1988) adapted these ideas for handling endogeneity in a binary probit model with fixed coefficients, and Petrin and Train (2009) generalized the approach to multinomial choice models with random coefficients. The procedure is implemented in two steps. First, the endogenous explanatory variable (such as price) is regressed against exogenous variables. The estimated regression is used to create a new variable (the control function) that is entered into the choice model. The choice model is then estimated with the original variables plus the new one, accounting appropriately for the distribution of unobserved factors conditional on both this new and the original variables. Applications include Ferreira (2004) and Guervara and Ben-Akiva (2006).

The third procedure is a full maximum likelihood approach, as applied by Villas-Boas and Winer (1999) to a multinomial logit with fixed coefficients and generalized by Park and Gupta (forthcoming) to random coefficient choice models. The procedure is closely related to the control function approach, in that it accounts for the nonzero conditional mean of the unobserved factors. However, instead of implementing the two steps sequentially (i.e., estimate the regression model to create the control function and then estimate the choice model with this control function), the two steps are combined into a joint estimation criterion. Additional assumptions are required to allow the estimation to be performed simultaneously; however, the procedure is more efficient when those assumptions are met.

In the sections that follow, we discuss each of these procedures, followed by a case study using the BLP approach.

13.2 The BLP Approach

This procedure is most easily described for choice among products where the price of the product is endogenous. A set of products are sold in each of several markets, and each market contains numerous consumers. The attributes of the products vary over markets but not over consumers within each market (i.e., all consumers within a given market face the same products with the same attributes.) The definition of a market depends on the application. The market might be a geographical area, as in Goolsbee and Petrin's (2004) analysis of households' choice among TV options. In this application, the price of cable TV and the features offered by cable TV (such as number of channels) vary over cities, since cable franchises are granted by local governments. Alternatively, the market might be defined temporally, as in BLP's (1995, 2004) analysis of new vehicle demand, where each model-year consists of a set of makes and models of new vehicles with prices and other attributes. Each year constitutes a market in this application. If the analysis is over products whose attributes are the same for all the consumers, then there is only one market, and differentiating by market is unnecessary.

13.2.1. Specification

Let M be the number of markets and J_m be the number of options available to each consumer in market m. J_m is the number of products available in market m plus perhaps, depending on the analysis, the option of not buying any of the products, which is sometimes called "the outside good."[1] The price of product j in market m is denoted p_{jm}. Some of the nonprice attributes of the products are observed by the researcher and some are not. The observed nonprice attributes of product j in market m are denoted by vector x_{jm}. The unobserved attributes are denoted collectively as ξ_{jm}, whose precise meaning is discussed in greater detail later.

The utility that consumer n in market m obtains from product j depends on observed and unobserved attributes of the product. Assume that utility takes the form

$$U_{njm} = V(p_{jm}, x_{jm}, s_n, \beta_n) + \xi_{jm} + \varepsilon_{njm},$$

where s_n is a vector of demographic characteristics of the consumer, $V(\cdot)$ is a function of the observed variables and the tastes of the consumer as represented by β_n, and ε_{njm} is iid extreme value. Note that ξ_{jm} enters the same for all consumers; in this setup, therefore, ξ_{jm} represents the average, or common, utility that consumers obtain from the unobserved attributes of product j in market m.

The basic issue that motivates the estimation approach is the endogeneity of price. In particular, the price of each product depends in general on all its attributes, both those that are observed by the researcher and those that are not measured by the researcher but nevertheless affect the demand and/or costs for the product. As a result, price p_{jm} depends on ξ_{jm}.

Suppose now that we were to estimate this model without regard for this endogeneity. The choice model would include price p_{jm} and the observed attributes x_{jm} as explanatory variables. The unobserved portion of utility, conditional on β_n, would be $\varepsilon^*_{njm} = \xi_{jm} + \varepsilon_{njm}$, which includes the average utility from unobserved attributes. However, since

[1] If the outside good is included then the model can be used to predict the total demand for products under changed conditions. If the outside good is not included, then the analysis examines the choice of consumers among products conditional on their buying one of the products. The model can be used to predict changes in shares among those consumers who originally purchased the products, but cannot be used to predict changes in total demand, since it does not include changes in the number of consumers who decided not to buy any of the products. If the outside good is included, its price is usually considered to be zero.

p_{jm} depends on ξ_{jm}, this unobserved component, ε^*_{njm}, is not independent of p_{jm}. To the contrary, one would expect a positive correlation, with more desirable unobserved attributes being associated with higher prices.

The BLP approach to this problem is to move ξ_{jm} into the observed portion of utility. This is accomplished by introducing a constant for each product in each market. Let $\bar{V}$ be the portion of $V(\cdot)$ that varies over products and markets, but is the same for all consumers. Let $\tilde{V}$ be the portion that varies over consumers as well as markets and products. Then $V(\cdot) = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n)$, where $\bar{\beta}$ are parameters that are the same for all consumers and $\tilde{\beta}_n$ are parameters that vary over consumers. Note that $\bar{V}$ does not depend on s_n since it is constant over consumers.[2] It is most natural to think of $\bar{V}$ as representing the average V in the population; however, it need not. All that is required is that $\bar{V}$ is constant over consumers. Variation in utility from observed attributes around this constant is captured by $\tilde{V}$, which can depend on observed demographics and on coefficients that vary randomly. Utility is then

$$U_{njm} = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \xi_{jm} + \varepsilon_{njm}.$$

Rearranging the terms, we have

$$U_{njm} = [\bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm}] + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \varepsilon_{njm}.$$

Note that the term in brackets does not vary over consumers. It is constant for each product in each market. Denote this constant as

$$\delta_{jm} = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm} \tag{13.1}$$

and substitute it into utility

$$U_{njm} = \delta_{jm} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \varepsilon_{njm}. \tag{13.2}$$

A choice model based on this utility specification does not entail any endogeneity. A constant is included for each product in each market, which absorbs ξ_{jm}. The remaining unobserved portion of utility, ε_{njm}, is independent of the explanatory variables. The constants are estimated along with the other parameters of the model. Essentially, the term that caused the endogeneity, namely, ξ_{jm}, has been subsumed into the product-market constant such that it is no longer part of the unobserved component of utility.

[2] It is possible for $\bar{V}$ to include aggregate demographics, such as average income for the market, which varies over markets but not over consumers in each market. However, we abstract from this possibility in our notation.

The choice model is completed by specifying how $\tilde{\beta}_n$ varies over consumers. Denote the density of $\tilde{\beta}_n$ as $f(\tilde{\beta}_n \mid \theta)$, where θ are parameters of this distribution representing, for example, the variance of coefficients around the common values. Given that ε_{njm} is iid extreme value, the choice probability is a mixed logit:

$$(13.3) \quad P_{nim} = \int \left[\frac{e^{\delta_{im} + \tilde{V}(p_{im}, x_{im}, s_n, \tilde{\beta}_n)}}{\sum_j e^{\delta_{jm} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n)}} \right] f(\tilde{\beta}_n \mid \theta)\, d\tilde{\beta}_n.$$

Usually, $\tilde{V}$ is linear with coefficients $\tilde{\beta}_n$ and explanatory variables that are the observed attributes, p_{jm} and x_{jm}, interacted perhaps with demographics, s_n. Other distributions can be specified for ε_{njm}; for example, Goolsbee and Petrin (2004) assume that ε_{njm} is jointly normal over products, such that the choice probability is a probit. Also, if information is available on the consumers' ranking or partial ranking of the alternatives, then the probability of the ranking is specified analogously; for example, Berry, Levinsohn, and Pakes (2004) and Train and Winston (2007) had data on the vehicle that each consumer bought as well as their second-choice vehicle and represented this partial ranking by inserting the exploded logit formula from Section 7.3 inside the brackets of Equation (13.3) instead of the standard logit formula.

Estimation of the choice model in (13.3) provides estimates of the constants and the distribution of tastes. However, it does not provide estimates of the parameters that enter the part of utility that is constant over consumers; that is, it does not provide estimates of $\bar{\beta}$ in $\bar{V}$. These parameters enter the definition of the constants in equation (13.1), which constitutes a regression model that can be used to estimate the average tastes. It is customary to express $\bar{V}$ as linear in parameters, such that (13.1) becomes

$$(13.4) \quad \delta_{jm} = \bar{\beta}' \bar{V}(p_{jm}, x_{jm}) + \xi_{jm},$$

where $\bar{V}(\cdot)$ is a vector-valued function of the observed attributes. A regression can be estimated where the dependent variable is the constant for each product in each market and the explanatory variables are the price and other observed attributes of the product. The error term for this regression is ξ_{jm}, which is correlated with price. However, procedures for handling endogeneity in linear regression models are well developed and are described in any standard econometrics textbook. In particular, regression (13.4) is estimated by instrumental variables rather than ordinary least squares. All that is required for this estimation is that the researcher has, or can calculate, some additional exogenous variables that are used as instruments in lieu of the endogenous price. The selection of instruments is discussed as part of the estimation procedure given later; however, first we need to address an important issue that has been implicit in our discussion so far, namely, how to handle the fact that there might be (and usually are) a very large number of constants to be estimated, one for each product in each market.

13.2.2. The Contraction

As described earlier, the constants $\delta_{jm}\ \forall j, m$ are estimated along with the other parameters of the choice model. When there are numerous products and/or markets, estimation of this large number of constants can be difficult or infeasible numerically if one tries to estimate them in the standard way. For example, for vehicle choice, there are more than 200 makes and models of new vehicles each year, requiring the estimation of more than 200 constants for each year of data. With 5 years of data, more than a thousand constants would need to be estimated. If the procedures in Chapter 8 were used for such a model, each iteration would entail calculating the gradient with respect to, say, 1,000 + parameters and inverting a 1,000 + by 1,000 + Hessian; also, numerous iterations would be required since the search is over a 1,000 + dimensional parameter space.

Luckily, we do not need to estimate the constants in the standard way. BLP provided an algorithm for estimating them quickly, within the iterative process for the other parameters. This procedure rests on the realization that constants determine predicted market shares for each product and therefore can be set such that the predicted shares equal actual shares. To be precise, let S_{jm} be the share of consumers in market m who choose product j. For a correctly specified model, the *predicted* shares in each market should equal these actual shares (at least asymptotically). We can find the constants that enforce this equality, that is, that cause the model to predict shares that match the actual shares. Let the constants be collected into a vector $\delta = \langle \delta_{jm}\ \forall j, m\rangle$. The predicted shares are $\hat{S}_{jm}(\delta) = \sum_n P_{njm}/N_m$, where the summation is over the N_m sampled consumers in market m. These predicted shares are expressed as a function of the constants δ because the constants affect the choice probabilities which in turn affect the predicted shares.

Recall that in Section 2.8 an iterative procedure was described for recalibrating constants in a model so that the predicted shares equal the actual shares. Starting with any given values of the constants, labeled $\delta^t_{jm}\ \forall j, m$, the constants are adjusted iteratively by the formula

$$\delta^{t+1}_{jm} = \delta^t_{jm} + \ln\left(\frac{S_{jm}}{\hat{S}_{jm}(\delta^t)}\right).$$

This adjustment process moves each constant in the "right" direction, in the following sense. If, at the current value of the constant, the actual share for a product exceeds the predicted share, then the ratio of actual to predicted (i.e., $S_{jm}/\hat{S}_{jm}(\delta^t)$) is greater than 1 and $\ln(\cdot)$ is positive. In this case, the constant is adjusted upward, to raise the predicted share. When the actual share is below the predicted share, the ratio is below 1 and $\ln(\cdot) < 0$, such that the constant is adjusted downward. The adjustment is repeated iteratively until predicted shares equal actual shares (within a tolerance) for all products in all markets.

This algorithm can be used to estimate the constants instead of estimating them by the usual gradient-based methods. The other parameters of the model are estimated by gradient-based methods, and at each trial value of these other parameters (i.e., each iteration in the search for the optimizing values of the other parameters), the constants are adjusted such that the predicted shares equal actual shares at this trial value. Essentially, the procedure that had been used for many years for postestimation recalibration of constants is used *during* estimation, at each iteration for the other parameters.

Berry (1994) showed that for any values of the other parameters in the choice model (i.e., of θ), there exists a unique set of constants at which predicted shares equal actual shares. Then BLP (1995) showed that the iterative adjustment process is a contraction, such that it is guaranteed to converge to that unique set of constants. When used in the context of estimation instead of postestimation calibration, the algorithm has come to be known as "the contraction."

A few additional notes are useful regarding the contraction. First, we defined the shares S_{jm} earlier as the "actual" shares. In practice, either aggregate market shares or sample shares can be used. In some situations, data on aggregate shares are not available or not reliable. Sample shares are consistent for market shares provided that the sampling is exogenous. Second, the procedure imposes a constraint or condition on estimation, namely, that predicted shares equal actual shares. As discussed in Section 3.7.1, maximum likelihood estimation of a standard logit model with alternative specific constants for each product in each market necessarily gives predicted shares that equal sample shares. The condition that predicted shares equal sample shares is therefore consistent with (or more precisely, is a feature of) maximum likelihood on a standard logit. However, for other models, including probit and mixed logit, the maximum likelihood estimator does not equate predicted shares with sample shares even when a full set of constants is included. The estimated constants that are obtained through the contraction are therefore not the maximum likelihood estimates. Nevertheless, since the condition holds asymptotically for a correctly specified model, imposing it seems reasonable.

13.2.3. Estimation by Maximum Simulated Likelihood and Instrumental Variables

There are several ways that the other parameters of the model (i.e, θ and $\bar{\beta}$) can be estimated. The easiest procedure to conceptualize is that used by Goolsbee and Petrin (2004) and Train and Winston (2007). In these studies, the choice model in equation (13.3) is estimated first, using maximum simulated likelihood (MSL) with the contraction. This step provides estimates of the parameters that enter equation (13.2), namely, the constants $\delta_{jm} \, \forall j, m$ and the parameters θ of the distribution of tastes around these constants. The contraction is used for the constants, such that the maximization of the log-likelihood function is only over θ.

To be more precise, since the choice probabilities depend on both δ and θ, this dependence can be denoted functionally as $P_{njm}(\delta, \theta)$. However, for any given value of θ, the constants δ are completely determined: they are the values that equate predicted and actual shares when this value of θ is used in the model. The calibrated constants can therefore be considered a function of θ, denoted $\delta(\theta)$. Substituting into the choice probability, the probability becomes a function of θ alone: $P_{njm}(\theta) = P_{njm}(\delta(\theta), \theta)$. The log-likelihood function is also defined as a function of θ: with i_n denoting n's chosen alternative, the log-likelihood function is $LL(\theta) = \sum_n ln P_{ni_n m}(\theta)$, where δ is recalculated appropriately for any θ. As such, the estimator is MSL subject to the constraint that predicted shares equal actual shares (either market or sample shares, whichever are used).[3]

Once the choice model is estimated, the estimated constants are used in the linear regression (13.4), which we repeat here for convenience:

$$\delta_{jm} = \bar{\beta}' \bar{v}(p_{jm}, x_{jm}) + \xi_{jm}.$$

The estimated constants from the choice model are the dependent variable in this regression, and the price and other observed attributes of the products are the explanatory variables. Since price is endogenous in this regression, it is estimated by instrumental variables instead of ordinary least squares. The instruments include the observed nonprice attributes

[3] From a programming perspective, the maximization entails iterations within iterations. The optimization procedure iterates over values of θ in searching for the maximum of the log-likelihood function. At each trial value of θ, the contraction iterates over values of the constants, adjusting them until predicted shares equal actual shares at that trial value of θ.

of the products, x_{jm}, plus at least one additional instrument in lieu of price. Denoting all the instruments as the vector z_{jm}, the instrumental variables estimator is the value of $\bar{\beta}$ that satisfies

$$\sum_j \sum_m [\hat{\delta}_{jm} - \bar{\beta}' \bar{v}(p_{jm}, x_{jm})] z_{jm} = 0,$$

where $\hat{\delta}_{jm}$ is the estimated constant from the choice model. We can rearrange this expression to give the estimator in closed form, like it is usually shown in regression textbooks and as given in Section 10.2.2:

$$\hat{\bar{\beta}} = (\sum_j \sum_m z_{jm} \bar{v}(p_{jm}, x_{jm})')^{-1} (\sum_j \sum_m z_{jm} \hat{\delta}_{jm}).$$

If the researcher chooses, efficiency can be enhanced by taking the covariance among the estimated constants into account, through generalized least squares (GLS); see, for example, Greene (2000), on GLS estimation of linear regression models.

The issue necessarily arises of what variables to use for instruments. It is customary, as already stated, to use the observed nonprice attributes as instruments under the assumption that they are exogenous.[4] BLP (1994) suggested instruments that are based on pricing concepts. In particular, each manufacturer will price each of its products in a way that takes consideration of substitution with its other products as well as substitution with other firms' products. For example, when a firm is considering a price increase for one of its products, consumers who will switch away from this product to another of the same firm's products do not represent as much of a loss (and might, depending on profit margins, represent a gain) as consumers who will switch to other firms' products. Based on these ideas, BLP proposed two instruments: the average nonprice attributes of other products by the same manufacturer, and the average nonprice attributes of other firms' products. For example, in the context of vehicle choice where, say, vehicle weight is an observed attribute, the two instruments for the Toyota Camry in a given year are (1) the average weight of all other makes and models of Toyota vehicles in that year, and (2) the average weight of all non-Toyota vehicles in that year.[5] Train and Winston (2007) used an extension of these instruments that reflects the extent to which each product differs from other

[4] This assumption is largely an expedient, since in general one would expect the unobserved attributes of a product to be related not only to price but also to the observed nonprice attributes. However, a model in which all observed attributes are treated as endogenous leaves little to use for instruments.

[5] Instead of averages, sums can be used, with the number of products by the same and by other firms (i.e., the denominators in the averages) also entering as instruments.

products by the same and different firms. In particular, they took the sum of squared differences between the product and each other product by the same and other firms. These two instruments for product j in market m are $z^1_{jm} = \sum_{k \in K_{jm}} (x_{jm} - x_{km})^2$, where K_{jm} is the set of products offered in market m by the firm that produced product j, and $z^2_{jm} = \sum_{k \in S_{jm}} (x_{jm} - x_{km})^2$, where S_{jm} is the set of products offered in market m by all firms except the firm that produced product j.

In other contexts, other instruments are appropriate. Goolsebee and Petrin (2004) examined households' choices among TV options, with each city constituting a market with different prices and nonprice attributes of cable and over-the-air TV. Following the practice suggested by Hausman (1997), they used the prices in other cities by the same company as instruments for each city. This instrument for city m is $z_{jm} = \sum_{m' \in K_m} p_{jm'}$, where K_m is the set of other cities that are served by the franchise operator in city m. The motivating concept is that the unobserved attributes of cable TV in a given city (such as the quality of programming for that city's franchise) are correlated with the price of cable TV in that city but are not correlated with the price of cable TV in other cities. They also included the city-imposed cable franchise fee (i.e., tax) and the population density of the city.

13.2.4. Estimation by GMM

As stated previously, several methods can be used for estimation, not just maximum likelihood. BLP (1995, 2004), Nevo (2001), and Petrin (2002) utilized a generalized method of moments (GMM) estimator. This procedure is a generalized version of the method of simulated moments (MSM) described in Chapter 10, augmented by moments for the regression equation. Moment conditions are created from the choice probabilities as

$$(13.5)\quad \sum_n \sum_j (d_{njm} - P_{njm}) z_{njm} = 0,$$

where d_{njm} is the dependent variable, which is 1 if consumer n in market m chooses alternative j and 0 otherwise, and z_{njm} is a vector of instruments that vary over products and markets (such as the observed nonprice attributes and functions of them) as well as over consumers in each market (such as demographic variables interacted with nonprice attributes.) In estimation, the exact probabilities are replaced with their simulated version. Note that this is the same formula as in Section 10.2.2, on MSM estimation of choice models. These moment conditions, when satisfied, imply that the observed mean of the instruments for the chosen alternatives, $\sum_n \sum_j d_{njm} z_{njm} / N_m$, is equal to the mean predicted by the model, $\sum_n \sum_j P_{njm} z_{njm} / N_m$. Note that, as described in Section 3.7, for a standard logit model, these moment conditions are the first-order condition for maximum likelihood estimation, where the instruments are the explanatory variables in the model. In other models, this moment condition is not the same as the first-order condition for maximum likelihood, such that there is a loss of efficiency. However, as described in Chapter 10 in comparing MSL with MSM, simulation of these moments is unbiased given an unbiased simulator of the choice probability, such that MSM is consistent for a fixed number of draws in simulation. In contrast, MSL is consistent only when the number of draws is considered to rise with sample size.

Moment conditions for the regression equation are created as

$$\sum_j \sum_m \xi_{jm} z_{jm} = 0,$$

where z_{jm} are instruments that vary over products and markets but not over people in each market (such as the observed nonprice attributes and functions of them). These moments can be rewritten to include the parameters of the model explicitly, assuming a linear specification for $\bar{V}$:

$$\sum_j \sum_m [\delta_{jm} - \bar{\beta}' \bar{v}(p_{jm}, x_{jm})] z_{jm} = 0. \tag{13.6}$$

As discussed in the previous section, these moments define the standard instrumental variables estimator of the regression coefficients.

The parameters of the system are $\bar{\beta}$, capturing elements of preferences that are the same for all consumers, and θ, representing variation in preferences over consumers. The constants δ_{jm} can also be considered parameters, since they are estimated along with the other parameters. Alternatively, they can be considered functions of the other parameters, as discussed in the previous subsection, calculated to equate predicted and actual shares at any given values of the other parameters. For the distinctions in the next paragraph, we consider the parameters to be $\bar{\beta}$ and θ, without the constants. Under this terminology, estimation proceeds as follows.

If the number of moment conditions in (13.5) and (13.6) combined is equal to the number of parameters (and have a solution), then the estimator is defined as the value of the parameters that satisfy all the moment conditions. The estimator is the MSM described in Chapter 10 augmented with additional moments for the regression. If the number of moment conditions exceeds the number of parameters, then no set of parameter values can satisfy all of them. In this case, the number of independent conditions is reduced through the use of a generalized method of moments (GMM).[6] The GMM estimator is described most easily by defining observation-specific moments:

$$g^1_{njm} = (d_{njm} - P_{njm})z_{njm},$$

which are the terms in (13.5), and

$$g^2_{jm} = [\delta_{jm} - \bar{\beta}'\bar{v}(p_{jm}, x_{jm})]z_{jm},$$

which are the terms in (13.6). Stack these two vectors into one, $g_{njm} = \langle g^1_{njm}, g^2_{jm} \rangle$, noting that the second set of moments is repeated for each consumer in market m. The moment conditions can then be written succinctly as $g = \sum_n \sum_j g_{njm} = 0$. The GMM estimator is the parameter value that minimizes the quadratic form $g'\Theta^{-1}g$, where Θ is a positive definite weighting matrix. The asymptotic covariance of g is the optimal weighting matrix, calculated as $\sum_n \sum_j g_{njm}g'_{njm}$. Ruud (2000, Chapter 21) provides a useful discussion of GMM estimators.

13.3 Supply Side

The supply side of a market can be important for several reasons. First, insofar as market prices are determined by the interaction of demand and supply, forecasting of market outcomes under changed conditions requires an understanding of supply as well as demand. An important example arises in the context of antitrust analysis of mergers. Prior to a merger, each firm sets the prices of its own product so as to maximize its own profits. When two firms merge, they set the prices of their products to maximize their joint profit, which can, and usually does, entail different prices than when the firms competed with each other. A central task of the Department of Justice and the Federal Trade Commission when deciding whether to approve or disallow a merger is to forecast the impact of the merger on prices. This task generally entails modeling the supply side, namely, the marginal costs and the pricing behavior of the firms, as well as the demand for each product as a function of price.

Second, firms in many contexts can be expected to set their prices not just at marginal cost or some fixed markup over marginal cost, but

[6] When simulated probabilities are used in the conditions, the procedure can perhaps more accurately be called a generalized method of simulated moments (GMSM) estimator. However, I have never seen this term used; instead, GMM is used to denote both simulated and unsimulated moments.

also on the basis of the demand for their product and the impact of price changes on their demand. In these contexts, the observed prices contain information about the demand for products and price elasticities. This information, if correctly extracted, can be used in the estimation of demand parameters. The researcher might, therefore, choose to examine the supply side as an aspect of demand estimation, even when the researcher's objectives do not entail forecasting the supply side *per se*.

In the paragraphs that follow, we describe several types of pricing behavior that can arise in markets and show how the specification of this behavior can be combined with the demand estimation discussed earlier. In each case, some assumptions about the behavior of firms are required. The researcher must decide, therefore, whether to incorporate the supply side into the analysis under these assumptions or estimate demand without the supply side, using the methods in the previous section. The researcher faces an inherent trade-off. Incorporation of the supply side has the potential to improve the estimates of demand and expand the use of the model. However, it entails assumptions about pricing behavior that might not hold, such that estimating demand without the supply side might be safer.

13.3.1. Marginal Cost

A basic tenet of economic theory is that prices depend on marginal cost (MC). The marginal cost of a product depends on its attributes, including the attributes that are observed by the researcher, x_{jm}, as well as those that are not, ξ_{jm}. Marginal cost also depends on input prices, such as rents and wages, and other "cost shifters." The researcher observes some of the cost shifters, labeled c_{jm}, and does not observe others. It is often assumed that marginal cost is separable in unobserved terms, such that it takes the form of a regression:

$$(13.7) \quad MC_{jm} = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm},$$

where $W(\cdot)$ is a function with parameters γ. The error term in this equation, μ_{jm}, depends in general on unobserved attributes ξ_{jm} and other unobserved cost shifters.

Marginal costs relate to prices differently, depending on the pricing mechanism that is operative in the market. We examine the prominent possibilities next.

13.3.2. MC Pricing

In perfectly competitive markets, the price of each product is driven down, in equilibrium, to its marginal cost. Pricing in a competitive market under separable marginal costs is therefore

$$p_{jm} = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm}. \tag{13.8}$$

This pricing equation can be used in conjunction with either of the two ways described earlier for estimating the demand system, each of which we discuss next.

MSL and IV with MC Pricing

Suppose that the method in Section 13.2.3 is used; that is, the choice model is estimated by maximum likelihood and then the estimated constants are regressed against product attributes using instrumental variables to account for the endogeneity of price. With this estimation, the variables that enter the MC pricing equation are the appropriate instruments to use in the instrumental variables. Any observed cost shifters c_{jm} serve as instruments along with the observed nonprice attributes x_{jm}. In this setup, the pricing equation simply provides information about the appropriate instruments. Since the pricing equation does not depend on demand parameters, it provides no information, other than the instruments, that can improve the estimation of demand parameters.

The researcher might want to estimate the pricing equation even though it does not include demand parameters. For example, the researcher might want to be able to forecast prices and demand under changes in cost shifters. In this case, equation (13.8) is estimated by ordinary least squares. The estimation of demand and supply then constitutes three steps: (1) estimate the choice model by MSL, (2) estimate the regression of constants on observed product attributes by instrumental variables, and (3) estimate the pricing equation by ordinary least squares. In fact, if the variables that enter the pricing equation are the only instruments, then the pricing equation is estimated implicitly as part of step 2 even if the researcher does not estimate the pricing equation explicitly in step 3. Recall from standard regression textbooks that instrumental variables estimation is equivalent to two-stage least squares. In our context, the instrumental variables in step 2 can be accomplished in two stages: (i) estimate the price equation by ordinary least squares and use the estimated coefficients to predict price, and then (ii) regress the constants on product attributes by ordinary least squares using the predicted price as an explanatory variable instead of the observed price. The overall estimation process is then as follows: (1) estimate the choice model, (2) estimate the pricing equation by ordinary least squares, and (3) estimate the equation for the constants by ordinary least squares using predicted prices instead of actual prices.

GMM with MC Pricing

If GMM estimation is used, then the explanatory variables in the pricing equation become instruments $z_{jm} = \langle x_{jm}, c_{jm} \rangle$, which enter the moments given in equations (13.5) and (13.6). The pricing equation provides additional moment conditions based on these instruments:

$$\sum_j \sum_m (p_{jm} - W(x_{jm}, c_{jm}, \gamma))z_{jm} = 0.$$

The demand parameters can be estimated without these additional moments. Or these additional moments can be included in the GMM estimation, along with those defined in (13.5) and (13.6). The estimator is the value of the parameters of the demand model and pricing equation that minimizes $g'\Theta^{-1}g$, where now g includes the moments from the pricing equation and Θ includes their covariance.

13.3.3. Fixed Markup over Marginal Cost

Firms might price at some fixed markup over marginal cost, with this markup not depending on demand. This form of pricing is considered by business people to be ubiquitous. For example, Shim and Sudit (1995) report that this form of pricing is used by more than 80 percent of managers at manufacturing firms. Of course, it is difficult to know how to interpret managers' statements about their pricing, since each manager can think that they set their prices at a fixed markup over marginal cost and yet the size of the "fixed" markup can be found to vary over managers in relation to demand for the different manufacturers' products.

This form of pricing has the same implications for demand estimation as MC pricing. Price is $p_{jm} = kMC_{jm}$ for some constant k. The pricing equation is the same as earlier, equation (13.8), but with the coefficients now incorporating k. All other aspects of estimation, using either MSL and IV or GMM, are the same as with MC pricing.

13.3.4. Monopoly Pricing and Nash Equilibrium for Single-Product *Firms*

Consider a situation in which each product is offered by a separate firm. If there is only one product in the market, then the firm is a monopolist. The monopolist is assumed to set its price to maximize profits, holding all other prices in the economy fixed. If there are multiple products, then the firms are oligopolists. We assume that each oligopolist maximizes its profits, given the prices of the other firms. This assumption implies that each oligopolist utilizes the same profit-maximization condition as a monopolist, holding all other prices (including its rivals' prices) fixed. Nash equilibrium occurs when no firm is induced to change its price, given the prices of other firms.

Let p_j and q_j be the price and quantity of product j, where the subscript for markets is omitted for convenience since we are examining the behavior of firms in a given market. The firm's profits are $\pi_j = p_j q_j - TC(q_j)$, where TC is total cost. Profits are maximized when

$$\frac{d\pi_j}{dp_j} = 0$$

$$\frac{d(p_j q_j)}{dp_j} - MC_j \frac{dq_j}{dp_j} = 0$$

$$q_j + p_j \frac{dq_j}{dp_j} = MC_j \frac{dq_j}{dp_j}$$

$$p_j + q_j \left(\frac{dq_j}{dp_j}\right)^{-1} = MC_j \tag{13.9}$$

$$p_j + p_j \frac{q_j}{p_j}\left(\frac{dq_j}{dp_j}\right)^{-1} = MC_j$$

$$p_j + (p_j/e_j) = MC_j,$$

where MC_j is the marginal cost of product j and e_j is the elasticity of demand for product j with respect to its price. This elasticity depends on all prices, because the elasticity for a product is different at different prices (and hence different quantities) for that product as well as other products. Note that the elasticity is negative, which implies that $p_j + p_j/e_j$ is less than p_j such that price is above marginal cost, as expected.[7]

[7] The condition can be rearranged to take the form that is often used for monopolists and one-product oligopolists: $(p_j - MC_j)/p_j = -1/e_j$, where the markup as a percent of price is inversely related to the magnitude of the elasticity.

Substituting the specification in (13.7) for MC, and adding subscripts for markets, we have

$$(13.10)\quad p_{jm} + (p_{jm}/e_{jm}) = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm},$$

which is the same as the pricing equation under MC pricing, (13.8), except that the left-hand side has (p_{jm}/e_{jm}) added to price.

The marginal cost parameters can be estimated after the demand parameters. In the context of MSL and IV, the process consists of the same three steps as under MC pricing: (1) estimate the choice model by MSL, (2) estimate the regression of constants on price and other attributes by instrumental variables, and (3) estimate the pricing equation by ordinary least squares. The only change is that now the dependent variable in the pricing equation is not price itself, but $p_{jm} + (p_{jm}/e_{jm})$. This term includes the elasticity of demand, which is not observed directly. However, given estimated parameters of the demand model from steps 1 and 2, the elasticity of demand can be calculated and used for step 3.

The fact that the pricing equation includes the elasticity of demand implies that observed prices contain information about demand parameters. The sequential estimation procedure just described does not utilize this information. In the context the of GMM, utilizing this extra information is straightforward, at least conceptually. Additional moment conditions are defined from the pricing equation as

$$\sum_j \sum_m (p_{jm} + (p_{jm}/e_{jm}) - W(x_{jm}, c_{jm}, \gamma))z_{jm} = 0.$$

The only difference from the procedure used under MC pricing is that now the moment conditions include the additional term (p_{jm}/e_{jm}). At each trial value of the parameters in the GMM estimation, the elasticity is calculated and inserted into this moment.

13.3.5. Monopoly Pricing and Nash Equilibrium for Multiproduct *Firms*

We now generalize the analysis of the previous section to allow each firm to sell more than one product in a market. If there is only one firm that offers all the products in the market, that firm is a multiproduct monopolist. Otherwise, the market is an oligopoly with multiproduct firms. The market for new vehicles is a prominent example, where each manufacturer, such as Toyota, offers numerous makes and models of vehicles. The pricing rule differs from the situation with only one product per firm because now each firm must consider the impact of its price for one product on the demand for its other products. We make the standard assumption, as before, that each firm prices so as to maximize profits, given the prices of other firms' products.

Consider a firm that offers a set K of products. The firm's profits are $\pi = \sum_{j \in K} p_j q_j - TC(q_j \, \forall j \in K)$. The firm chooses the price of product $j \in K$ that maximizes its profits:

$$d\pi/dp_j = 0$$
$$q_j + \sum_{k \in K} (p_k - MC_k)(dq_k/dp_j) = 0.$$

Analogous conditions apply simultaneously to all the firm's products. The conditions for all the firm's products can be expressed in a simple form by use of matrix notation. Let the prices of the K products be stacked in vector p, quantities in vector q, and marginal costs in vector MC. Define a $K \times K$ matrix of derivatives of demand with respect to price, D, where the (i, j)th element is (dq_j/dp_i). Then the profit-maximizing prices for the firm's products satisfy

$$q + D(p - MC) = 0$$
$$D^{-1}q + p - MC = 0$$
$$p + D^{-1}q = MC.$$

Note that this last equation is just a generalized version of the one-product monopolist's rule given in (13.9). It is handled the same way in estimation. With MSL and IV estimation, it is estimated after the demand model, using the demand parameters to calculate D^{-1}. With GMM, the pricing equation is included as an extra moment condition, calculating D^{-1} at each trial value of the parameters.

13.4 Control Functions

The BLP approach is not always applicable. If observed shares for some products in some markets are zero, then the BLP approach cannot be implemented, since the constants for these product markets are not identified. (Any finite constant gives a strictly positive predicted share, which exceeds the actual share of zero.) An example is Martin's (2008) study of consumers' choice between incandescent and compact fluorescent lightbulbs (CFLs), where advertising and promotions occurred on a weekly basis and varied over stores, and yet it was common for a store not to sell any CFLs in a given week. Endogeneity can also arise over the decision makers themselves rather than over markets (i.e., groups of decision makers), such that the endogeneity is not absorbed into product-market constants. For example, suppose people who like public transit choose to live near transit such that transit time in their mode choice is endogenous. Constants cannot be estimated for each decision maker, since the constants would be infinity (for the chosen alternative) and negative infinity (for the nonchosen alternatives), perfectly predicting the choices and leaving no information for estimation of parameters. Even when the BLP approach can be implemented, the researcher might want to avoid the complication of applying the contraction that is usually needed in the BLP approach.

Alternatives that have been implemented include the control function approach, which we discuss in this section, and a more highly specified version of the control function approach that utilizes full-information maximum likelihood, which we discuss in the following section.

The setup for the control function approach quite closely follows the specification of simultaneous equation regression models. However, since choice models are nonlinear, an additional layer of complication must be addressed, and this complication can restrict the applicability of the method more than would be immediately expected given the analogy to linear models.

Let the endogenous explanatory variable for decision maker n and alternative j be denoted as y_{nj}. We do not differentiate markets, as in the BLP approach, because we allow the possibility that the endogenous variables vary over decision makers rather than just over markets. The endogenous variable might be price, transit time, or whatever is relevant in a given application. In the sections that follow, we discuss issues that can arise when the endogenous variable is price.

The utility that consumer n obtains from product j is expressed as

$$(13.11)\quad U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj},$$

where x_{nj} are observed exogenous variables relating to person n and product j (including observed demographics). The unobserved term ε_{jm} is not independent of y_{nj} as required for standard estimation. Let the endogenous explanatory variable be expressed as a function of observed instruments and unobserved factors:

$$(13.12)\quad y_{nj} = W(z_{nj}, \gamma) + \mu_{nj},$$

where ε_{nj} and μ_{nj} are independent of z_{nj}, but μ_{nj} and ε_{nj} are correlated. The correlation between μ_{nj} and ε_{nj} implies that y_{nj} and ε_{nj} are correlated, which is the motivating concern. We assume for our initial discussion that μ_{nj} and ε_{nj} are independent for all $k \neq j$.

Consider now the distribution of ε_{nj} conditional on μ_{nj}. If this conditional distribution takes a convenient form, then a control function approach can be used to estimate the model. Decompose ε_{nj} into its mean conditional on μ_{nj} and deviations around this mean: $\varepsilon_{nj} = E(\varepsilon_{nj} \mid \mu_{nj}) + \tilde{\varepsilon}_{nj}$. By construction, the deviations are not correlated with μ_{nj} and therefore not correlated with y_{nj}. The conditional expectation is a function of μ_{nj} (and perhaps other variables); it is called the control function and denoted $CF(\mu_{nj}, \lambda)$, where λ are the parameters of this function. The simplest case is when $E(\varepsilon_{nj} \mid \mu_{nj}) = \lambda\mu_{nj}$, such that the control function is simply μ_{nj} times a coefficient to be estimated. We discuss motivations for various control functions later. Substituting the conditional mean and deviations into the utility equation, we have

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + CF(\mu_{nj}, \lambda) + \tilde{\varepsilon}_{nj}. \tag{13.13}$$

The choice probabilities are derived from the conditional distribution of the deviations $\tilde{\varepsilon}_{nj}$. Let $\tilde{\varepsilon}_n = \langle \tilde{\varepsilon}_{nj} \, \forall j \rangle$ and $\mu_n = \langle \mu_{nj} \, \forall j \rangle$. The conditional distribution of $\tilde{\varepsilon}_n$ is denoted $g(\tilde{\varepsilon}_n \mid \mu_n)$, and the distribution of β_n is $f(\beta_n \mid \theta)$. The choice probability is then

$$\begin{aligned} P_{nj} &= Prob(U_{nj} > U_{nk} \, \forall k \neq j) \\ &= \int \int I(V_{nj} + CF_{nj} + \tilde{\varepsilon}_{nj} > V_{nk} + CF_{nk} + \tilde{\varepsilon}_{nk} \, \forall k \neq j) \\ &\qquad g(\tilde{\varepsilon}_n \mid \mu_n) f(\beta_n \mid \theta) d\tilde{\varepsilon} d\beta_n, \end{aligned} \tag{13.14}$$

where the following abbreviations are used:

$$\begin{aligned} V_{nj} &= V(p_{nj}, x_{nj}, \beta_n) \\ CF_{nj} &= CF(\mu_{nj}, \lambda). \end{aligned}$$

This is a choice model just like any other, with the control function entering as an extra explanatory variable. Note that the inside integral is over the conditional distribution of $\tilde{\varepsilon}$ rather than the original ε. By construction, $\tilde{\varepsilon}$ is not correlated with the endogenous variable, while the original ε was correlated. Essentially, the part of ε that is correlated with y_{nj} is entered explicitly as an extra explanatory variable, namely, the control function, such that the remaining part is not correlated.

The model is estimated in two steps. First, equation (13.12) is estimated. This is a regression with the endogenous variable as the dependent variable and with exogenous instruments as explanatory variables. The residuals for this regression provide estimates of the μ_{nj}'s. These residuals are calculated as $\hat{\mu}_{nj} = y_{nj} - W(z_{nj}, \hat{\gamma})$ using the estimated parameters $\hat{\gamma}$. Second, the choice model is estimated with $\hat{\mu}_{nj}$ entering the control function. That is, the choice probabilities in (13.14) are estimated by maximum likelihood, with $\hat{\mu}_{nj}$ and/or a parametric function of it entering as extra explanatory variables.

The central issue with the control function approach is the specification of the control function and the conditional distribution of $\tilde{\varepsilon}_n$. In some situations, there are natural ways to specify these elements of the model. In other situations, it is difficult, or even impossible, to specify them in a way that meaningfully represents reality. The applicability of the approach depends on the researcher being able to meaningfully specify these terms.

Some examples of control function specifications are as follows:

1. Let

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}$$
$$y_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$$

and assume that ε_{nj} and μ_{nj} are jointly normal with zero mean and constant covariance matrix for all j. By the properties of normals, the expectation of ε_{nj} conditional on μ_{nj} is $\lambda\mu_{nj}$, where λ reflects the covariance, and deviations around the mean, $\tilde{\varepsilon}_{nj}$, are normal with constant variance. In this case, the control function is $CF(\mu_{nj}, \lambda) = \lambda\mu_{nj}$. Utility is

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \tilde{\varepsilon}_{nj}.$$

The choice model is probit with the residual from the price equation as an extra variable. Note however that the variance of $\tilde{\varepsilon}_{nj}$ differs from the variance of ε_{nj}, such that the scale in the estimated probit is different from the original scale. If β_n is random, then the model is a mixed probit.

2. Suppose ε_{nj} consists of a normally distributed part that is correlated with y_{nj} and a part that is iid extreme value. In particular, let

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon^1_{nj} + \varepsilon^2_{nj}$$
$$y_{nj} = W(z_{nj}, \gamma) + \mu_{nj},$$

where ε^1_{nj} and μ_{nj} are jointly normal and ε^2_{nj} is iid extreme value. The conditional distribution of ε^1_{nj} is, as in the previous example, normal with mean $\lambda\mu_{nj}$ and constant variance. However, the conditional distribution of ε^2_{nj} is the same as its unconditional distribution since ε^2_{nj} and μ_{nj} are independent. Utility becomes

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \tilde{\varepsilon}^1_{nj} + \varepsilon^2_{nj},$$

where $\tilde{\varepsilon}^1_{nj}$ is normal with zero mean and constant variance. This error component can be expressed as $\tilde{\varepsilon}^1_{nj} = \sigma\eta_{nj}$, where η_{nj}, is standard normal. Utility becomes

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \sigma\eta_{nj} + \varepsilon^2_{nj}.$$

The choice probability is a mixed logit, mixed over the normal error components $\sigma\eta_{nj}\ \forall j$ as well as the random elements of β_n. The standard deviation, σ, of the conditional errors is estimated, unlike in the previous example.

3. The notation can be generalized to allow for correlation between ε_{nj} and μ_{nk} for $k \neq j$. Let $\varepsilon_{nj} = \varepsilon^1_{nj} + \varepsilon^2_{nj}$ as in the previous example, except that now we assume that the stacked vectors ε^1_n and μ_n are jointly normal. Then ε^1_n conditional on μ_n is normal with mean $M\mu_n$ and variance Ω, where M and Ω are matrices of parameters. Stacking utilities and the explanatory functions, we have

$$U_n = V(y_n, x_n, \beta_n) + M\mu_n + L\eta_n + \varepsilon^2_n,$$

where L is the lower-triangular Choleski factor of Ω and η_n is a vector of iid standard normal deviates. Since the elements of $\tilde{\varepsilon}^2_n$ are iid extreme value, the model is mixed logit, mixed over error components η_n and the random elements of β_n. The residuals for all products enter the utility for each product.

13.4.1. Relation to Pricing Behavior

As stated previously, the primary limitation of the control function approach is the need to specify the control function and the conditional distribution of the new unobserved term $\tilde{\varepsilon}$. In some situations, the true conditional distribution is so complex that it cannot be derived and any specification using standard distributions, like normal, is necessarily incorrect. These issues are explained most readily in relation to the pricing behavior of firms, where price is the endogenous variable, but they can arise under any type of endogenous variable, depending on the way the endogenous variable is determined.

Let the choice be among products and the endogenous variable y_{nj} be price p_{nj}, so that we can discuss alternative pricing behaviors and whether they can be accommodated in the control function approach. Consider first a situation where the control function can be readily applied: MC pricing. The utility that consumer n obtains from product j is

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \varepsilon^1_{nj} + \varepsilon^2_{nj},$$

where ε^1_{nj} is correlated with price and ε^2_{nj} is iid extreme value. For example, ε^1_{nj} might represent unobserved attributes of the product. Prices vary over people because different people are in different markets or prices are set separately for people or groups of people (e.g., to account for transportation costs to each customer). Suppose further that firms set prices at marginal cost, which is specified as $MC_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$, where z_{nj} are exogenous variables that affect marginal cost (including the observed attributes of the product) and μ_{nj} captures the effect of unobserved cost shifters (including the unobserved attributes of the product). The price equation is then

$$p_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$$

Assume that ε^1_{nj} and μ_{nj} are jointly normal with the same covariance matrix for all j. Correlation might arise, for example, because unobserved attributes affect utility as well as costs, thereby entering both ε^1_{nj} and μ_{nj}. As in the second example given earlier, utility becomes

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \sigma\eta_{nj} + \varepsilon^2_{nj},$$

where η_{nj} is iid standard normal. The model is estimated in two steps: First, the pricing equation is estimated and its residuals, $\hat{\mu}_{nj}$, are retained. Second, the choice model is estimated with these residuals entering as explanatory variables. The model is a mixed logit, mixed over the new error components η_{nj}. The same specification is applicable if firms price at a constant markup over marginal cost. With this pricing behavior, the pricing equation is sufficiently simple that reasonable assumptions on the observed terms in the model give a conditional distribution for unobserved terms in utility that can be derived and is convenient.

Consider, now, monopoly pricing or Nash equilibrium where price depends on the elasticity of demand as well as on marginal cost. As shown earlier in relation to the BLP approach, the pricing equation for a single-product monopolist or Nash oligopolists is

$$p_{nj} + (p_{nj}/e_{nj}) = MC$$
$$p_{nj} = -(p_{nj}/e_{nj}) + W(z_{nj}, \gamma) + \mu_{nj}.$$

Suppose now that ε^1_{nj} and μ_{nj} are jointly normal. The distribution of ε^1_{nj} conditional on μ_{nj} is still normal. However, μ_{nj} is no longer the only error in the pricing equation and so we cannot obtain an estimate of μ_{nj} to condition on. Unlike MC pricing, the unobserved component of demand, ε^1_{nj}, enters the pricing equation through the elasticity. The pricing equation includes two unobserved terms: μ_{nj} and a highly nonlinear transformation of ε^1_{nj} entering through e_{nj}.

If we rewrite the pricing equation in a form that can be estimated, with an observed part and a separable error, we would have

$$p_{nj} = Z_j(z_{nj}, \gamma) + u^*_{nj},$$

where z_n is a vector of all of the observed exogenous variables that affect marginal cost and the elasticity, $Z_j(\cdot)$ is a parametric function of these variables, and u^*_{nj} is the unobserved deviations of price around this function. We can estimate this equation and retain its residual, which is an estimate of u^*_{nj}. However, u^*_{nj} is not μ_{nj}; rather u^*_{nj} incorporates both μ_{nj} and the unobserved components of the elasticity-based markup p_{nj}/e_{nj}. The distribution of ε^1_{nj} conditional on this u^*_{nj} has not been derived, and may not be derivable. Given the way ε^1_{nj} enters the pricing equation through the elasticity, its conditional distribution is certainly not normal if its unconditional distribution is normal. In fact, its conditional distribution is not even independent of the exogenous variables.

Villas-Boas (2007) has proposed an alternative direction of derivation in this situation. He points out that the difficulty that we encountered above in specifying an appropriate control function arises from the assumption that marginal costs are separable in unobserved terms, rather than from the assumption that prices are related to elasticities. Let us assume instead that marginal cost is a general function of observed and unobserved terms: $MC = W^*(z_{nj}, \gamma, \mu_{nj})$. In many ways, the assumption of nonseparable unobserved terms is more realistic, since unobserved cost shifters can be expected, in general, to interact with observed cost shifters. Under this general cost function, Villas-Boas shows that for any specification of the control function and distribution of ε^1_{nj} conditional on this control function, there exists a marginal cost function $W^*(\cdot)$ and distribution of unobserved terms μ_{nj} and ε^1_{nj} that are consistent with them. This result implies that the researcher can apply the control function approach even when prices depend on elasticities, knowing that there is some marginal cost function and distribution of unobserved terms that make the approach consistent. Of course, this existence proof does not provide guidance on what control function and conditional distribution are most reasonable, which must still remain an important issue for the researcher.

13.5 Maximum Likelihood Approach

The maximum likelihood approach is similar to the control function approach except that the parameters of the model are estimated simultaneously rather than sequentially. As with the control function approach, utility is given in equation (13.13) and the endogenous explanatory variable is specified by (13.12). To allow more compact notation, stack each of the terms over alternatives, such that the two equations are

(13.15) $U_n = V(y_n, x_n, \beta_n) + \varepsilon_n$

(13.16) $y_n = W(z_n, \gamma) + \mu_n.$

Rather than specifying the conditional distribution of ε_n given μ_n, the researcher specifies their joint distribution, denoted $g(\varepsilon_n, \mu_n)$. From equation (13.16), μ_n can be expressed as a function of the data and parameters: $\mu_n = y_n - W(z_n, \gamma)$. So the joint distribution of ε_n and y_n is $g(\varepsilon_n, y_n - W(z_n, \gamma))$. Denote the chosen alternative as i. The probability of the observed data for person n is the probability that the endogenous explanatory variable takes the value y_n and that alternative i is chosen. Conditional on β_n, this probability is

$$P_n(\beta_n) = \int I(U_{ni} > U_{nj}\ \forall j \neq i) g(\varepsilon_n, y_n - W(z_n, \gamma)) d\varepsilon_n.$$

If β_n is random, then $P_n(\beta_n)$ is mixed over its distribution. The resulting probability P_n is inserted into the log-likelihood function: $LL = \sum_n ln(P_n)$. This LL is maximized over the parameters of the model. Instead of estimating (13.16) first and using the residuals in the choice probability, the parameters of (13.16) and the choice model are estimated simultaneously.

The third example of the control function approach (which is the most general of the examples) can be adapted to this maximum likelihood procedure. Stacked utility is

$$U_n = V(y_n, x_n, \beta_n) + \varepsilon_n^1 + \varepsilon_n^2$$

with the stacked endogenous variable:

$$y_n = W(z_n, \gamma) + \mu_n,$$

where $W(\cdot)$ is now vector valued. Each element of ε_n^2 is assumed to be iid extreme value. Assume that ε_n^1 and μ_n are jointly normal with zero mean and covariance Ω. Their density is denoted $\phi(\varepsilon_n^1, \mu_n \mid \Omega)$. The probability that enters the log-likelihood function for person n who chose alternative i is

$$P_n = \int\int \frac{e^{V(y_{ni}, x_{ni}, \beta_n) + \varepsilon_{ni}^1}}{\sum e^{V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}^1}} \phi(\varepsilon_n, (y_n - W(z_n, \gamma)) \mid \Omega) f(\beta_n \mid \theta) d\varepsilon_n d\beta_n.$$

This probability is inserted into the log-likelihood function, which is maximized with respect to γ (the parameters that relate the endogenous explanatory variable to instruments), θ (which describes the distribution of preferences affecting utility), and Ω (the covariance of the correlated unobserved terms ε_n^1 and μ_n.) Of course, in any particular application, restrictions can be placed on Ω to reduce the number of parameters. For example, Park and Gupta (forthcoming) assume that ε_{nj}^1 and μ_{nk} are not correlated for $k \neq j$.

The control function and maximum likelihood approaches provide a trade-off that is common in econometrics: generality versus efficiency. The maximum likelihood approach requires a specification of the joint distribution of ε_n and μ_n, while the control function requires a specification of the conditional distribution of ε_n given μ_n. Any joint distribution implies a particular conditional distribution, but any given conditional distribution does not necessarily imply a particular joint distribution. There may be numerous joint distributions that have the specified conditional distribution. The control function approach is therefore more general than the maximum likelihood approach: it is applicable under any joint distribution that is consistent with the specified conditional distribution. However, if the joint distribution can be correctly specified, the maximum likelihood approach is more efficient than the control function, simply by the fact that it is the maximum likelihood for all the parameters.

13.6 Case Study: Consumers' Choice among New Vehicles

A useful illustration of the BLP approach is given by Train and Winston (2007). Their study examined the question of why the Big Three automakers have been losing market share. As part of the study, they estimated a model of buyers' choices among new vehicles. Since many attributes of new vehicles are unobserved by the researcher and yet affect price, it is expected that price is endogenous.

Estimation was performed on a sample of consumers who bought or leased new vehicles in the year 2000, along with data on the aggregate market shares for each make and model in that year. The analysis did not include an outside good, and as such, gives the demand conditional on new vehicle purchase. For each sampled buyer, the survey information included the make and model of the vehicle that the person bought plus a list of the vehicles that the person said they considered. The chosen and considered set were treated as a ranking, with the chosen vehicle first and the considered vehicles ranked in the order in which the person listed them. The choice model was specified as an "exploded logit" for the probability of the person's ranking, mixed over a distribution of random coefficients. See Section 7.3 for an extended discussion of this specification for rankings data. The choice model included constants

Table 13.1. *Mixed exploded logit model of new vehicle choice*

	Parameter	Standard Error
Price (MRSP) divided by respondents' income		
Mean coefficient	−1.6025	0.4260
Standard deviation of coefficient	0.8602	0.4143
Consumer Report's repair index, for women aged >30 years	0.3949	0.0588
Luxury or sports car, for lessors	0.6778	0.4803
Van, for households with an adolescent	3.2337	0.5018
SUV or station wagon, for households with an adolescent	2.0420	0.4765
ln(1 + number of dealers within 50 miles of person's home)	1.4307	0.2714
Horsepower: standard deviation of coefficient	0.0045	0.0072
Fuel consumption (1/mpg): standard deviation of coefficient	−102.15	20.181
Light truck, van, or pickup: standard deviation of coefficient	6.8505	2.5572
Number of previous consecutive GM purchases	0.3724	0.1471
Number of previous consecutive GM purchases, if rural	0.3304	0.2221
Number of previous consecutive Ford purchases	1.1822	0.1498
Number of previous consecutive Chrysler purchases	0.9652	0.2010
Number of previous consecutive Japanese purchases	0.7560	0.2255
Number of previous consecutive European purchases	1.7252	0.4657
Manufacturer loyalty, error component: standard deviation	0.3453	0.1712
SLL at convergence	−1994.93	

for each make and model of vehicle, as well as explanatory variables and random coefficients that capture variations in preferences over consumers. The analysis distinguished 200 makes and models and used the contraction to calculate the 199 constants (with one normalized to zero) within the maximum likelihood estimation of the other parameters. The estimated constants were then regressed on observed attributes of the vehicles, including price. Since price was considered endogenous, this regression was estimated by instrumental variables rather than ordinary least squares.

Table 13.1 gives the estimates of the parameters that relate to variation in preferences over consumers. These are the parameters that were estimated by maximum likelihood on the probability of each buyers' ranking of makes and models, using the contraction of the constants. The estimated coefficients have the following implications:

- Price divided by income enters as an explanatory variable, to capture the concept that higher-income households place less importance on price than households with lesser income. The variable is given a normally distributed random coefficient, whose mean and standard deviations are estimated. The estimated mean is negative, as expected, and the estimated standard deviation is fairly large and significant (with a t-statistic over 2), indicating considerable variation in response to price.

- The Consumer Report's repair index enters for women who are at least 30 years old, and not for men and younger women. This distinction was discovered through testing of alternative demographic variables interacted with the repair index.
- People who lease their vehicle are found to have a stronger preference for luxury and sports vehicles than people who buy.
- Households with adolescents are found to have a stronger preference for vans, SUVs, and station wagons than other households.
- Dealership locations are found to affect households' choices. One of the purposes of Train and Winston's analysis was to investigate the impact of dealership locations on vehicle choice, to see whether changes in dealership locations can explain part of the Big Three's loss in share. The demand model indicates that, as expected, the probability of buying a given make and model rises when there are more dealerships within a 50-mile radius of the household that sell that make and model.
- Horsepower, fuel consumption, and a truck dummy enter with random coefficients. These variables do not vary over consumers, only over makes and models. As a result, the mean coefficient times the variable is absorbed into the make/model constants, such that only the standard deviation is estimated in the choice model. To be precise, each of these variables enters utility as $\beta_n x_j$, with random β_n. Decompose β_n into its mean $\bar{\beta}$ and deviations $\tilde{\beta}_n$. The mean impact $\bar{\beta}x_j$ does not vary over consumers and becomes part of the constant for vehicle j. The deviations $\tilde{\beta}_n x_j$ enter the choice model separately from the constants, and the standard deviation of $\tilde{\beta}_n$ is estimated. The estimates imply considerable variation in preferences for horsepower, fuel efficiency, and trucks.
- The last set of variables captures the loyalty of consumers toward manufacturers. Each variable is specified as the number of past consecutive purchases from a given manufacturer (or group of manufacturers). The estimated coefficients imply that loyalty to European manufacturers is largest, followed by loyalty to Ford. Interestingly, rural households are found to be more loyal to GM than are urban households. These loyalty variables are a type of lagged dependent variable, which introduce econometric issues due to the possibility of serially correlated unobserved factors. Winston and Train discuss these issues and account for the serial correlation, at least partly, in their estimation procedure.

Table 13.2. *Regression of constants on vehicle attributes*

	Parameter	Standard Error
Price (MSRP, in thousands of dollars)	−0.0733	0.0192
Horsepower divided by weight (in tons)	0.0328	0.0117
Automatic transmission standard	0.6523	0.2807
Wheelbase (in inches)	0.0516	0.0127
Length minus wheelbase (in inches)	0.0278	0.0069
Fuel consumption (in gallons per mile)	−31.641	23.288
Luxury or sports car	−0.0686	0.2711
SUV or station wagon	0.7535	0.4253
Van	−1.1230	0.3748
Pickup truck	0.0747	0.4745
Chrysler	0.0228	0.2794
Ford	0.1941	0.2808
General Motors	0.3169	0.2292
European	2.4643	0.3424
Korean	0.7340	0.3910
Constant	−7.0318	1.4884
R-squared	0.394	

The choice model in Table 13.1 included constants for each make and model of vehicle. The estimated constants were regressed on vehicle attributes to estimate the aspects of demand that are common to all consumers. Table 13.2 gives the estimated coefficients of this regression, using instrumental variables to account for the endogeneity of price.

As expected, price enters with a negative coefficient, indicating that consumers dislike higher price, all else held equal (i.e., assuming there are no changes in a vehicle's attributes to compensate for the higher price). Interestingly, when the regression was estimated by ordinary least squares, ignoring endogeneity, the estimated price coefficient was considerably smaller: −0.0434 compared with the estimate of −0.0733 using instrumental variables. This direction of difference is expected, since a positive correlation of price with unobserved attributes creates a downward bias in the magnitude of the price coefficient. The size of the difference indicates the importance of accounting for endogeneity.

The other estimated coefficients have expected signs. Consumers are estimated to value additional power, as evidenced by the positive coefficient for the horsepower-to-weight ratio. Consumers also prefer to have an automatic transmission as standard equipment (holding the price of the vehicle constant). The size of the vehicle is measured by both its wheelbase and its length beyond the wheelbase. Both measures enter positively, and the wheelbase obtains a larger coefficient than the length beyond the wheelbase. This relation is expected, since the wheelbase is generally a better indicator of the interior space of the vehicle than the length beyond the wheelbase. The negative coefficient of fuel consumption implies that consumers prefer greater fuel efficiency, which reduces fuel consumption per mile.

Note that price enters both parts of the model: the regression in Table 13.2, which captures the impacts that are constant over consumers, and the choice model in Table 13.1, which capture impacts that differ over consumers. Taking both parts together, price enters utility with a coefficient: $-0.0773 - 1.602/\text{Income} + 0.860 * \eta/\text{Income}$, where η is a standard normal random term. The price coefficient has a constant component, a part that varies with the income of the household and a part that varies randomly over households with the same income.

The endogeneity of price has been appropriately handled in this case study by including constants in the choice model to absorb the unobserved attributes and then using instrumental variables when regressing the constants against price and other observed attributes, that is, by the BLP approach. A case study of the control function approach is provided by Petrin and Train (2009) and of the maximum likelihood approach by Park and Gupta (forthcoming).