Identification of Demand Parameters Using Ordinary Least Squares

Alexander MacKay
Harvard University*

Nathan H. Miller
Georgetown University†

First Draft: March 20, 2017
This Draft: August 11, 2017

Abstract

One of the main challenges of demand estimation is the problem of price endogeneity. We address this problem by pairing the standard model of supply-side profit-maximization with the known functional form of demand. These supply-side restrictions allow us to interpret the coefficient from an ordinary least squares regression of (transformed) quantity on price as a function of demand parameters. When combined with minimal assumptions about the correlations of unobservable shocks, the profit-maximization conditions are sufficient to identify the key parameters of several models commonly used in demand estimation.

*Harvard University, Harvard Business School. Email: amackay@hbs.edu.
†Georgetown University, McDonough School of Business. Email: nathan.miller@georgetown.edu

1 Introduction

One of the main challenges of demand estimation is the problem of price endogeneity. When firms adjust prices due to unobserved demand shocks, equilibrium variation in prices and quantities does not represent the causal demand curve, and an ordinary least squares (OLS) regression produces biased estimates. This has long been recognized in the economics literature (Working, 1927). Typically, OLS estimates are dismissed as uninformative, and the identification challenge is cast as a problem of finding a valid instrument.

We reconsider whether equilibrium variation in prices and quantities can identify demand parameters when supply-side behavior is subject to the typical assumptions used in empirical work. With profit-maximizing behavior and a surprisingly weak condition, which we detail below, we show that the OLS bias can be quantified as a function of the data. OLS estimates are informative; they capture a blend of the causal demand curve and the endogenous response by firms. When interpreted through the lens of a sufficiently tight model, the causal component can be recovered without the need for an instrument.

Consider the simple case of a monopolist with a constant marginal cost technology and a linear demand schedule. Equilibrium variation is introduced through demand and cost shocks that are independent and unobservable to the econometrician. We prove that a consistent estimate of the price parameter, β, is given by

$$\hat{\beta} = -\sqrt{\left(\hat{\beta}^{OLS}\right)^2 + \frac{\hat{Cov}(\hat{\xi}^{OLS}, q)}{\hat{Var}(p)}}$$

where $\hat{\beta}^{OLS}$ is the price coefficient from an OLS regression of quantities on prices, and $\hat{\xi}^{OLS}$ is a vector of the OLS residuals. The information provided by OLS regression is sufficient for the consistent identification of the structural parameter. In Section 2, we prove this result, and we use numerical simulations to illustrate how a model of supply and demand allows for the recovery of causal parameters from the familiar "cloud" of price-quantity pairs.

We develop an expression for the OLS bias under more general supply and demand conditions, which we provide in Section 3. The demand systems we consider include random coefficients logit (e.g., Berry et al. (1995); Nevo (2001)) and other workhorse models as special cases. Using our formulation of the bias, we construct a consistent three-step estimator from the OLS estimate, OLS residuals, and residualized prices. We explore alternative identification approaches under related supply-side assumptions, including another novel approach, which we call "indirect" instrumental variables. In simulations, we find that the three-step estimator outperforms traditional two-stage least squares in small samples.

Invoking the classic adage, there is no free lunch in economics. Point identification requires the econometrician to understand the relationship between unobserved demand and cost shocks in the model. We derive our main results under an *uncorrelatedness assumption*, where the covariance between these shocks equals zero. This provides a moment that can be exploited in estimation. The identifying power of covariance restrictions in linear systems of equations was recognized in early work at the Cowles Foundation (e.g., Koopmans et al. (1950)). In more recent research, Hausman and Taylor (1983) show that if the error term in one linear equation can be recovered with a first-stage regression, then under uncorrelatedness it can serve as an "unobserved instrument" that enables estimation of a second linear equation; Matzkin (2004) extends that approach to nonlinear models. Our methodology is similar in many respects, but we do not require supply to be estimated before demand. This allows the pricing function to depend directly on demand parameters, as is the case under typical assumptions.

Uncorrelatedness is a reasonable assumption for many settings, especially if panel data allow the econometrician to employ fixed effects that absorb product- and time-specific characteristics. That said, it is clear that uncorrelatedness is not universally applicable. When econometric corrections are not sufficient to motivate uncorrelatedness, a simple modification to our estimator allows for consistent estimation given any correlation value. This modification can be used to set identify the price parameter even with no a priori information about the correlation, as the data and model together rule out certain values. When further restricted by priors, such as information about the sign of the correlation, the bounds on the price parameter may be sufficient to answer relevant policy questions.

We consider the generalizability of our estimator to alternative assumptions in Section 4. We describe how to implement our method with standard discrete choice models and a richer class of demand systems that includes constant elasticity demand. In more complex models, we show how to first account for non-linear parameters when using our estimator to identify the price coefficient. For arbitrarily complex demand systems, a method-of-moments estimator based on analogous conditions can be implemented. On the supply side, we extend the analysis to likely use cases, including multi-product firms, nonlinear marginal cost schedules, and the alternative paradigms of Cournot competition and perfect competition.

Finally, we apply our methodology to estimate demand in the contexts studied by two recently published articles. In the first application, we estimate demand for a model of Cournot competition in the Portland cement industry studied by Fowlie et al. (2016). The three-step estimator obtains a mean price elasticity of -1.19, which is statistically indistinguishable from the elasticity of -1.22 we obtain relying on the instrumental variables (IV) used by the original authors. In the second application, we follow the analysis of Aguirregabiria and Ho (2012) to estimate demand in the U.S. airline industry. In contrast to the first application, the empirical airline model employs nested logit demand and multi-product firms with Bertrand pricing. In our preferred three-step specification, we find a price parameter very close to what we obtain with our replication of the IV estimate (-0.182

compared to -0.189). Interestingly, we find a nesting parameter that is significantly smaller (0.53 compared to 0.82), which implies that the exclusion restrictions used in IV and our uncorrelatedness assumptions are incompatible in this context. We demonstrate how to incorporate additional moments when there are non-price endogenous covariates, such as the within-group share in nested logit demand.

We hope that this methodology generates richer empirical research in areas where valid instruments are difficult to obtain. That said, our methodology could also provide a useful complement to the IV approach in many applications. Because the uncorrelatedness assumption can be recast as a moment restriction, combining it with IV restrictions allows for specification tests that otherwise would be unavailable to the econometrician. When an IV approach is possible, our method may be used to provide an overidentification restriction for testing, as the IV approach is agnostic about supply-side behavior. In combination, our supply-side restrictions and the IV conditions provide power to reject the functional form of demand or the assumed competitive environment.

Though we focus our results on specific, widely-used assumptions about demand and supply, we view our method as not particular to these assumptions. Rather, the main insight is that information about supply-side behavior can be modeled to adjust the observed relationships between quantity and price. Price can be decomposed into marginal cost and a markup; our method provides a direct way to correct for endogeneity arising from the latter component. In a more general sense, this insight lies in a long lineage of control function estimation procedures (e.g., Heckman (1979)). Our particular approach benefits from the ubiquity of profit-maximization assumptions in microeconomics.

Our paper contributes to a rich literature of within-the-data solutions to price endogeneity. When no cost shifters or natural experiments are available, researchers typically rely on instruments constructed from other (demand-side) variables in the data. The most common approaches use functions of the characteristics of other products within a market (Berry et al., 1995; Berry and Haile, 2014; Gandhi and Houde, 2015) or the prices of the same good in other markets (Hausman, 1996; Nevo, 2001). Like our method, these approaches require assumptions about the correlation structure of unobservables. A recent paper that leverages supply-side modeling is Byrne et al. (2016), which provides an identification strategy when supplemental data on costs and market size are available.[1]

[1] The authors assume that quantities, input prices, and total costs are sufficient to pin down marginal costs and also excludable from marginal revenues, after conditioning on shares. They show how to identify demand parameters by mapping a function of these variables to the marginal revenues implied by profit maximization.

2 Motivating Example: Monopoly with Linear Demand

We introduce the uncorrelatedness assumption using a model of monopoly pricing, in the spirit of Rosse (1970). In each market t, the monopolist faces a downward-sloping linear demand schedule, $q_t = \alpha + \beta p_t + \xi_t$, where q_t and p_t denote quantity and price, respectively, and ξ_t is a demand shock with $E[\xi] = 0$. Its marginal costs are given by the function $mc_t = \gamma + \eta_t$, where γ is some constant and η_t is a cost shock. Prices are set to maximize profit. The econometrician observes vectors of prices, $p = [p_1, p_2, \dots]'$, and quantities, $q = [q_1, q_2, \dots]'$. The markets can be conceptualized as being geographically and temporally distinct. We focus our discussion on identification of the price parameter, β. Once it is recovered, the other parameters can be expressed as linear functions of the data.

It well known that an OLS regression of q on p obtains an estimate of β that is biased if the monopolist's price reflects the unobservable demand shock, as is the case here given the assumption of profit maximization. Formally,

$$\hat{\beta}^{OLS} = \frac{\hat{Cov}(p,q)}{\hat{Var}(p)} \quad \xrightarrow{p} \quad \beta + \frac{Cov(\xi,p)}{Var(p)} \tag{1}$$

The conventional wisdom in most applications is that $|\hat{\beta}^{OLS}| < |\beta|$ because $Cov(\xi,p) > 0$, and this is easily confirmed in our motivating example. Thus, the empirical literature often provides OLS estimates as a benchmark against which to evaluate IV estimates, but treats OLS as uninformative about the true parameter (e.g., Berry et al. (1995); Nevo (2001)).

This approach can underrepresent the informational content of OLS. Often, maintained assumptions on the supply-side inform the covariance of price and the demand shock. In our motivating example, consider that the monopolist's profit-maximization conditions, $p_t + (\frac{dq}{dp})^{-1} q_t = \gamma + \eta_t$, can be solved to yield price as a linear function of the demand and cost shocks: $p_t = \frac{1}{2}(-\frac{\alpha}{\beta} - \frac{\xi_t}{\beta} + \gamma + \eta_t)$. Indeed, if the covariance between the shocks is known, then this equilibrium condition is sufficient to fully resolve the bias present in the OLS estimate without reliance on instruments. We now provide the uncorrelatedness assumption.

Assumption 1 (Uncorrelatedness): $Cov(\xi, \eta) = 0$.

Uncorrelatedness is a reasonable assumption for many settings, especially if panel data are used to absorb product-specific and time-specific factors. Weaker assumptions, such as $Cov(\xi, \eta) \geq 0$, are insufficient for point identification but nonetheless allow the econometrician to place bounds on the price coefficient. For now, we defer discussion along these lines, and we develop our main results under the uncorrelatedness assumption as they apply to the motivating example.

Proposition 1. *Let the OLS estimates of (α, β) be $(\hat{\alpha}^{OLS}, \hat{\beta}^{OLS})$ with probability limits $(\alpha^{OLS}, \beta^{OLS})$, and denote the residuals at the limiting values as $\xi_t^{OLS} = q_t - \alpha^{OLS} - \beta^{OLS} p_t$. The proba-*

bility limit of the OLS estimate can be expressed as a function of the true price parameter, the residuals from the OLS regression, prices, and quantities:

$$\beta^{OLS} \equiv plim\left(\hat{\beta}^{OLS}\right) = \beta - \frac{1}{\beta + \frac{Cov(p,q)}{Var(p)}} \frac{Cov(\xi^{OLS}, q)}{Var(p)} \tag{2}$$

Proof: We provide the proofs in this section for illustrative purposes; all subsequent proofs are confined to the appendix. Reformulate equation (1) as follows:

$$\beta^{OLS} = \beta + \frac{Cov(\xi, \eta - \frac{1}{\beta}q)}{Var(p)} = \beta - \frac{1}{\beta}\frac{Cov(\xi, q)}{Var(p)}$$

The first equality holds due to the first order condition $p = \gamma + \eta_t - \frac{1}{\beta}q$. The second equality holds due to the uncorrelatedness assumption. The structure of the model also allows for us to solve for $Cov(\xi, q)$:

$$\begin{aligned} Cov(\xi, q) &= Cov(\xi^{OLS} - (\beta - \beta^{OLS})p, q) \\ &= Cov(\xi^{OLS}, q) - (\beta - \beta^{OLS})Cov(p, q) \\ &= Cov(\xi^{OLS}, q) - \frac{1}{\beta}\frac{Cov(\xi, q)}{Var(p)}Cov(p, q) \end{aligned}$$

Collecting terms and rearranging implies

$$\frac{1}{\beta}Cov(\xi, q) = \frac{1}{\beta + \frac{Cov(p,q)}{Var(p)}} Cov(\xi^{OLS}, q)$$

Plugging into the reformulation of equation (1) obtains the proposition. QED.

The proposition makes clear that, among the objects that characterize the probability limit of the OLS estimate, only the true price parameter itself does not have a well understood sample analog. Because the probability limit itself can be estimated consistently, this raises the possibility that the true price parameter can be recovered from the data. Indeed, a closer inspection of equation (2) reveals that β solves a simple quadratic equation.

Proposition 2. *The price parameter, β, is the lower root of the quadratic equation*

$$\beta^2 + \beta\left(\frac{Cov(p, q)}{Var(p)} - \beta^{OLS}\right) - \frac{Cov(\xi^{OLS}, q)}{Var(p)} - \frac{Cov(p, q)}{Var(p)}\beta^{OLS} = 0$$

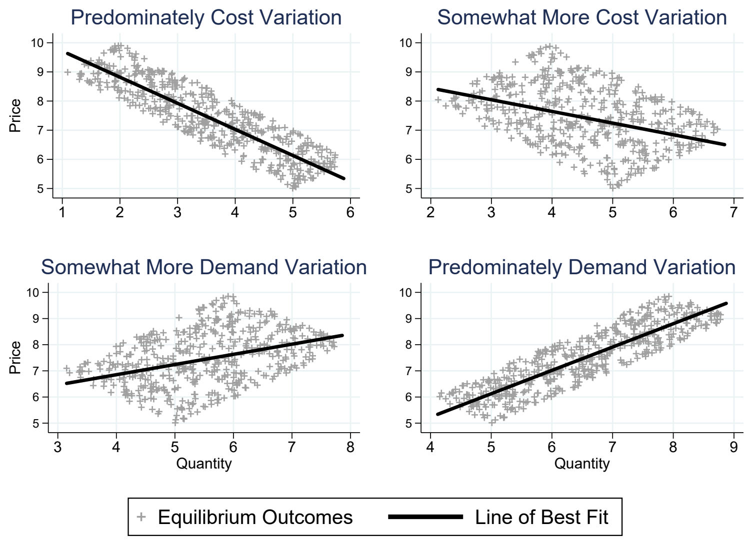

Figure 1: Price and Quantity Variation in Monopoly Model

so that a consistent estimate of β is given by

$$\hat{\beta} = -\sqrt{\left(\hat{\beta}^{OLS}\right)^2 + \frac{\hat{Cov}(\hat{\xi}^{OLS}, q)}{\hat{Var}(p)}}. \tag{3}$$

Proof: The quadratic equation is obtained as a re-expression of equation (2). An application of the quadratic formula provides the following roots:

$$\frac{-\left(\frac{Cov(p,q)}{Var(p)} - \beta^{OLS}\right) \pm \sqrt{\left(\frac{Cov(p,q)}{Var(p)} - \beta^{OLS}\right)^2 + 4\left(\frac{Cov(\xi^{OLS},q)}{Var(p)} + \frac{Cov(p,q)}{Var(p)}\beta^{OLS}\right)}}{2}.$$

In the univariate case, $\frac{Cov(p,q)}{Var(p)} = \beta^{OLS}$, which cancels out terms and obtains the probability limit analog of equation (3). As the sample estimates of covariance terms are consistent for the limits, we obtain a consistent estimate. The upper root is positive and therefore can be ruled out as an estimate of β.

Thus, the OLS residuals contain the information necessary to construct a consistent estimate of β. The simplicity of equation (3) arises from the linearity of the model. As will be shown in the next sections, Proposition 1 and Proposition 2 can be generalized to many oligopoly settings with nonlinear demand systems and marginal cost schedules, but the required equations become somewhat more complicated.

Before turning to the more general model, we provide a simple numerical example to

Table 1: Numerical Illustration for the Monopoly Model

	(1)	(2)	(3)	(4)
$\hat{\beta}^{OLS}$	-0.87	-0.38	0.37	0.87
$\hat{Cov}(\hat{\xi}^{OLS}, q)$	0.31	0.90	0.90	0.31
$\hat{Var}(p)$	1.44	1.11	1.11	1.43
$\hat{Cov}(\hat{\xi}^{OLS}, q)/\hat{Var}(p)$	0.22	0.81	0.81	0.22
$\hat{\beta}$	-1.00	-0.98	-0.97	-0.99

Notes: The table provides sample estimates based on numerically generated data that conform to the motivating example of monopoly pricing. The demand curve is $q_t = 10 - p_t + \xi_t$ and marginal costs are $mc_t = \eta_t$, so that $(\beta_0, \beta, \gamma_0)$ = $(10, -1, 0)$. In column (1), $\xi \sim U(0, 2)$ and $\eta \sim U(0, 8)$. In column (2), $\xi \sim U(0, 4)$ and $\eta \sim U(0, 6)$. In column (3), $\xi \sim U(0, 6)$ and $\eta \sim U(0, 4)$. In column (4), $\xi \sim U(0, 8)$ and $\eta \sim U(0, 2)$. Thus, the support of the cost shocks are largest (smallest) relative to the support of the demand shocks in the left-most (right-most) column. The lower root provides a consistent estimate of $\beta = -1.00$.

help fix ideas. Let demand be given by $q_t = 10 - p_t + \xi_t$ and let marginal cost be $mc_t = \eta_t$, so that $(\alpha, \beta, \gamma) = (10, -1, 0)$. Let the demand and cost shocks have independent uniform distributions. The monopolist sets price to maximize profit. As is well known, if both cost and demand variation is present then equilibrium outcomes provide a "cloud" of data points that do not necessarily correspond to the demand curve. To illustrate this, we consider four cases with varying degrees of cost and demand variation. In case (1), $\xi \sim U(0, 2)$ and $\eta \sim U(0, 8)$. In case (2), $\xi \sim U(0, 4)$ and $\eta \sim U(0, 6)$. In case (3), $\xi \sim U(0, 6)$ and $\eta \sim U(0, 4)$. In case (4), $\xi \sim U(0, 8)$ and $\eta \sim U(0, 2)$. We randomly take 1,000 draws for each case and calculate the equilibrium prices and quantities.

The data are plotted in Figure 1 along with OLS fits. The experiment represents the classic identification problem of demand estimation: the empirical relationship between equilibrium prices and quantities can be quite misleading about the slope of the demand function. However, Proposition 2 states that the structure of the model together with the OLS estimates allow for consistent estimates of the price parameter. Table 1 provides sample analogs for the objects that appear in equation (3). The OLS estimates, $\hat{\beta}^{OLS}$, are negative when the cost shocks are relatively more important and positive when the demand shocks are relatively more important, as is revealed in the scatterplots. By contrast, $\frac{\hat{Cov}(\hat{\xi}^{OLS}, q)}{\hat{Var}(p)}$ is larger if the cost and demand shocks have relatively more similar support. The lower roots of the quadratic equation are -1.00, -0.98, -0.97, and -0.99, all of which are close to the true price parameter of -1.00.

3 Methodology

3.1 Data-Generating Process

We now provide a more general formulation of the methodology. There are $j = 1, 2, \ldots, J$ single-product firms that set prices in each of $t = 1, 2, \ldots, T$ markets, subject to downward sloping demands. The econometrician observes vectors of prices, $p_t = [p_{1t}, p_{2t}, \ldots, p_{Jt}]'$, and quantities, $q = [q_{1t}, q_{2t}, \ldots, p_{Jt}]'$, corresponding to each market t, as well as a matrix of covariates $X_t = [x_{1t} \; x_{2t} \; \ldots \; x_{Jt}]$. We assume that the covariates are exogenous in the sense that $E[X\xi] = E[X\eta] = 0$, where ξ_{jt} and η_{jt} are demand and cost shocks that are common knowledge among firms but unobserved by the econometrician.[2]

We develop results using the following assumptions on demand and supply:

Assumption 2 (Demand): *The demand schedule of each firm is determined by the following semi-linear form:*

$$h(q_{jt}; w_{jt}) = \beta p_{jt} + x_{jt}'\alpha + \xi_{jt} \tag{4}$$

where $h(q_{jt}; .)$ increases monotonically in its argument, i.e., $h'(q_{jt}; \cdot) > 0$, and where w_{jt} is a vector of observables and parameters that allow the semi-linear relationship.

Assumption 3 (Supply): *Each firm sets price to maximize its profit in each market, taking as given the prices of other firms, with knowledge of the demand schedule equation* (4) *and the linear constant marginal cost schedule*

$$mc_{jt} = x_{jt}'\gamma + \eta_{jt}. \tag{5}$$

The demand assumption restricts attention to systems for which, after a transformation of quantities, there is additive separability in prices, covariates, and the demand shock. It is sufficiently flexible to nest monopolistic competition with linear demands (e.g., as in the motivating example) and the discrete choice demand models that support much of the empirical research in industrial organization. For example, with a logit demand system $h(q_{jt}; w_{jt}) \equiv \log(q_{jt}/q_{0t})$, where q_{0t} is the quantity of an outside good, and with a random coefficients logit demand system $h(q_{jt}; w_{jt})$ can be computed numerically via the contraction mapping of Berry et al. (1995). We develop these connections in Section 4.1. The supply assumption restricts attention to constant marginal cost schedules (e.g., Berry et al. (1995); Nevo (2001); Miller and Weinberg (2017)). Other assumptions, including alternative equilibrium concepts (e.g., Cournot), multi-product firms, and non-constant costs can be accommodated with more complicated notation that we defer to Section 4.2.

[2]Another information environment consistent with this methodology is one in which each demand shock has a common component and an independent private component, and firms commit to prices before observing those of their rivals. In estimation, the common components may be captured by fixed effects, and then the rival firms' prices may then be included directly in X, as they will be orthogonal to the private demand shock for each firm.

Assumptions 2 and 3 generate first order conditions that must hold in equilibrium for each product j:

$$p_{jt} = mc_{jt} - \frac{1}{\beta} h'(q_{jt}; w_{jt}) q_{jt}. \tag{6}$$

The first order conditions clarify that prices are additively separable in marginal costs and a markup term, where the markup term depends on the unobserved demand shock. Throughout the paper, we assume that markets are "in equilibrium," in that prices are linked to an optimality condition that can be modeled by the researcher. In principle, markets with multiple equilibria do not alter our results as long as the selected equilibrium can be identified by the econometrician.

3.2 Three-Step Estimation

Assumptions 1, 2, and 3 are sufficient to identify the linear parameters of the model (α, β, γ). In this section, we develop a three-step estimator based on OLS that performs well in small samples. Because (α, γ) can be recovered trivially if β is known, we focus solely on the price parameter. In principle, the model could feature additional parameters in w_{jt} that result in the semi-linear transformation (e.g., random coefficient logit). These parameters are not identified by our supply-side restrictions, and, to focus efforts, we assume they are known to the econometrician.[3]

The OLS estimate of the price parameter is obtained by a regression of $h(q_{jt}; \cdot)$ on p. The probability limit of the estimator contains the standard bias term:

$$\hat{\beta}^{OLS} \equiv \frac{\hat{Cov}(p^*, h(q))}{\hat{Var}(p^*)} \quad \xrightarrow{p} \quad \beta + \frac{Cov(p^*, \xi)}{Var(p^*)}$$

where $p^* = [I - x(x'x)^{-1}x']p$ is a vector of residuals from a regression of p on x. Our first general result is that the OLS bias can be expressed as a function of observables:

Proposition 3. *Under assumptions 1 through 3, the probability limit of the OLS estimate can be written as a function of the true price parameter, the residuals from the OLS regression, prices, and quantities:*

$$\beta^{OLS} \equiv plim\left(\hat{\beta}^{OLS}\right) = \beta - \frac{1}{\beta + \frac{Cov(p^*, h'(q)q)}{Var(p^*)}} \frac{Cov\left(\hat{\xi}^{OLS}, h'(q)q\right)}{Var(p^*)} \tag{7}$$

Proof. See appendix.

[3] An alternative and perhaps more realistic interpretation is that the econometrician is considering a candidate vector of nonlinear parameters, and wishes to determine the values that the linear parameters must take to rationalize the data. This alternative interpretation would apply in the nested fixed point estimation routine of Berry et al. (1995) and Nevo (2001) for the random coefficients logit demand system.

Thus, with the exception of the unknown parameter β, all of the components that characterize the probability limit of the OLS estimate can be directly estimated from the data. We use this formula to solve for the price parameter. The following proposition establishes identification and consistency:

Proposition 4. *Under assumptions 1 through 3, a consistent estimate of the price parameter β is given by*

$$\hat{\beta}^{\text{3-Step}} = \frac{1}{2}\left(\hat{\beta}^{OLS} - \frac{\hat{Cov}\,(p^*, h'(q)q)}{\hat{Var}(p^*)} - \sqrt{\left(\hat{\beta}^{OLS} + \frac{\hat{Cov}\,(p^*, h'(q)q)}{\hat{Var}(p^*)}\right)^2 + 4\frac{\hat{Cov}\left(\hat{\xi}^{OLS}, h'(q)q\right)}{\hat{Var}(p^*)}}\right) \tag{8}$$

if the following auxiliary condition holds: $\beta < \frac{Cov(p^*,\xi)}{Var(p^*)} - \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}$.

Proof. See appendix.

From Proposition 4, we obtain a three-step method for a consistent estimate of the price parameter, β:

1. Regress $h(q;\cdot)$ on p and x using OLS.
2. Regress p on x using OLS and construct the residuals p^*.
3. Construct $\hat{\beta}^{\text{3-Step}}$ based on equation (8).

Some discussion of the auxiliary condition is warranted. If the model conforms to the conventional wisdom that $Cov(p^*, \xi) > 0$ then the condition holds unless prices are sufficiently negatively correlated with markups (recall from equation (6) that markups are $-\frac{1}{\beta}h'(q_{jt}; w_{jt})q_{jt}$). This can be verified in some models. For example, if demand is linear then $\frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)} = \frac{Cov(p^*,q)}{Var(p^*)} = \beta^{OLS}$ and $\frac{Cov(p^*,\xi)}{Var(p^*)} - \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)} = -\beta$. Because $\beta < 0$, $\beta < -\beta$ and the condition holds. The auxiliary condition can also be verified by checking both roots. If one is positive, then the other is the true parameter. One root is guaranteed to be positive if (residualized) prices are positively correlated with both the demand shocks and markups, which is the case in many models. If the auxiliary condition cannot be verified analytically and both roots are negative, then there are two candidate values for β, where the second is the upper root of the quadratic obtained from equation (8). In applications, if the roots diverge substantially then the upper root may be possible to rule out on basis of implying negative marginal costs.

3.3 Alternative Estimation Strategies

3.3.1 Method-of-Moments

The three-step estimator relies on the moment condition $E[\xi{\cdot}mc] = 0$, which holds given the maintained assumptions $Cov(\xi, \eta) = 0$, $E[X\xi] = 0$, and $E[\xi] = 0$. An alternative approach is to search numerically to find a price parameter, $\hat{\beta}$, such that the empirical moment $\frac{1}{T}\sum_t \xi_t(\hat{\beta}; w, X)' mc_t(\hat{\beta}; w, X) = 0$ holds, where $\xi(\hat{\beta}; w, X)$ and $mc(\hat{\beta}; w, X)$ are computed given the data and the price parameter using equations (4)-(6). Formally, this "method-of-moments" estimator is defined as

$$\hat{\beta}^{MM} = \arg\min_{\tilde{\beta}<0} \left[\frac{1}{T} \xi_t(\tilde{\beta}; w, X)' mc_t(\hat{\beta}; w, X) \right]^2 \tag{9}$$

The linear parameters (α, γ) can be obtained for any given $(\tilde{\beta}, w, X, q, p)$, and therefore can be concentrated out of the nonlinear optimization problem. The conditions for identification are identical to those of the three-step estimator. Indeed, in Monte Carlo experiments based on the data generating process outlined above, we find that the two estimators are equivalent to numerical precision.

The three-step estimator has distinct advantages over the method-of-moments approach. The analytic formula reduces the computational burden in estimation. This can be especially beneficial when nested inside of a nonlinear routine for other parameters, as multi-dimensional optimization can be particularly burdensome. Additionally, the three-step estimator reduces computational error by immediately obtaining the correct solution and rejecting cases when no solution is possible. A method-of-moments optimizer may get stuck at a local minimum, and it produces a solution even when the moment may not be set to zero (which should reject the model).

On the other hand, the numerical approach to estimation also offers some advantages. First, there are settings for which the three-step estimator does not generalize or is difficult to calculate. Examples include models in which demand is not semi-linear or firms price multiple products. (We discuss these and other generalizations in Section 4). The method-of-moments estimator handles such settings without difficulty, provided that marginal costs and the unobserved demand term can be recovered from the data for any candidate price parameter.

Second, in some settings one or more of the covariates may enter the cost function but not demand (e.g., $\alpha^{(k)} = 0$ and $\gamma^{(k)} \neq 0$ for some k). The three-step estimator requires orthogonality between the unobserved demand shock and *all* the regressors. This can be relaxed for regressors that enter only the cost function by basing estimation on the moment $E[\xi\eta] = 0$. The corresponding empirical moment is $\frac{1}{T}\sum_t \xi_t(\hat{\beta}; w, X)' \eta_t(\hat{\beta}; w, X) = 0$. In small samples, we find that this tends to increase bias and variance relative to the method-of-moments estimator defined by equation (9). We believe that this reflects that the moment $E[\xi_t' mc_t] = 0$ usually creates more identifying variation than the moment $E[\xi_t' \eta_t] = 0$ because, by construction, $mc_t = X_t\gamma + \eta_t$.

Third, the empirical moments defined above can be combined with more standard exclusion restrictions (e.g., $E[\xi Z] = 0$ for instruments Z), to construct a generalized method-of-moments estimator. This can improve efficiency and allows for specification tests that otherwise would be unavailable to the econometrician (e.g., Hausman (1978); Hansen (1982)).

3.3.2 Two-Step Estimation

In the presence of an additional restriction, we can produce a more precise estimator that can be calculated in one fewer step. When the observed cost and demand shifters are uncorrelated, there is no need to project the price on demand covariates when constructing a consistent estimate, and one can proceed immediately using the OLS regression. We formalize the additional restriction and the estimator below.

Assumption 4: *Let the parameters $\alpha^{(k)}$ and $\gamma^{(k)}$ correspond to the demand and supply coefficients for covariate k in X. For any two covariates k and l, $Cov(\alpha^{(k)}x^{(k)}, \gamma^{(l)}x^{(l)}) = 0$.*

Proposition 5. *Under assumptions 1 through 4, a consistent estimate of the price parameter β is given by*

$$\hat{\beta}^{\text{2-Step}} = \frac{1}{2}\left(\hat{\beta}^{OLS} - \frac{\hat{Cov}\left(p, h'(q)q\right)}{\hat{Var}(p)} - \sqrt{\left(\hat{\beta}^{OLS} + \frac{\hat{Cov}\left(p, h'(q)q\right)}{\hat{Var}(p)}\right)^2 + 4\frac{\hat{Cov}\left(\hat{\xi}^{OLS}, h'(q)q\right)}{\hat{Var}(p)}}\right) \tag{10}$$

when the auxiliary condition, $\beta < \frac{Cov(p^*,\xi)}{Var(p^*)}\frac{Var(p)}{Var(p^*)} - \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}$, holds.

The estimator can be expressed entirely in terms of the data, the OLS coefficient, and the OLS residuals. The first step is an OLS regression of $h(q;\cdot)$ on p and x, and the second step is the construction of the estimator as in equation (10). Thus, we eliminate the step of projecting p on x. This estimator will be consistent under the assumption that any covariate affecting demand does not covary with marginal cost. The auxiliary condition parallels that of the three-step estimator, and we expect that it will hold in typical cases.

3.3.3 Indirect Instruments

Our examination of the supply-side conditions suggests an alternative approach to identification. When there exists some covariate, $x^{(1)}$, that is uncorrelated with both the demand shocks and the marginal costs, we may obtain identification even without assumption 1 on uncorrelatedness. Formally, we state the alternative assumption and the estimator:

Assumption 5 (Indirect Instrument): *There exists an exogenous demand shifter $x^{(1)}$ for which $Cov(p, x^{(1)}) \neq 0$ and $\gamma^{(1)} = 0$. That is, it sastifies the exclusion restrictions $Cov(x^{(1)}, \xi) = 0$ and $Cov(x^{(1)}, mc) = 0$.*

Proposition 6. *Under assumptions 2, 3, and 5, a consistent estimate for β is given by*

$$\hat{\beta}^{IIV} = \frac{\hat{Cov}\left(\tilde{p}, -h'(q)q\right)}{\hat{Var}(\tilde{p})}$$

where $\tilde{p}$ is the predicted values of a regression of p on $x^{(1)}$.

We term this approach the method of indirect instruments because the intuition is similar to that of instrumental variables. The demand covariate $x^{(1)}$ satisfies a relevance condition and a pair of exclusion restrictions. Rather than instrument for price using $x^{(1)}$, which is not possible, the relationships implied by the model allow for the indirect identification of the price parameter. Though less efficient than our three-step estimator, it is obtained using weaker assumptions about the exogeneity of observables. The approach may be appealing even when other estimators are plausible, as separate estimates of β may be constructed using different covariates that satisfy both exclusion restrictions. The separate estimates can be used to test the validity of the assumptions.

3.4 Small-Sample Properties

We generate Monte Carlo results to examine the small sample properties of the estimators. We consider a monopolist that prices against a logit demand curve,

$$q_t = \frac{\exp\left(\beta p_t + x_t^{(1)}\alpha + \xi_t\right)}{1 + \exp\left(\beta p_t + x_t^{(1)}\alpha + \xi_t\right)}$$

given the constant marginal cost schedule $mc_t = x_t^{(2)}\gamma + \eta_t$. The standard transformation of quantity yields a semi-linear form consistent with equation (4).[4] We draw the exogenous data, $(x_t^{(1)}, x_t^{(2)}, \xi_t, \eta_j)$, from independent $U[0,1]$ distributions, set $(\beta, \alpha, \gamma) = (-0.5, 2, 2)$, and compute the equilibrium price and quantity for each draw of data. The mean price and margin are 4.54 and 0.68, respectively, and the mean price elasticity of demand is -1.52. We construct samples with 25, 50, 100, and 500 observations, drawing 1,000 simulations for each sample size. We estimate with OLS and 2SLS, using $x_t^{(2)}$ as an excluded instrument in the latter case. We also apply the alternative estimators developed above, all of which are consistent by construction.

Table 2 summarizes the results. Panel A shows that the average bias of the three-step estimator is less one percent, even for the smallest sample size ($N = 25$). The standard deviations among the sample estimates are provided in Panel B; with $N = 25$ the standard deviation is 0.103. Panel C shows that mean squared error (MSE) is 0.011 with $N = 25$ and is within rounding error of zero for the larger sample sizes. The bias, variance, and MSE of 2SLS are notably greater than those of the three-step estimator at each sample size. The differences reflect that 2SLS uses the moment $E[\xi_t' x_t^{(2)}] = 0$, whereas the three-step estimator uses $E[\xi_t' mc_t] = 0$. Because $mc_t = x_t^{(2)}\gamma + \eta_t$, the latter moment typically provides more identifying variation. This also highlights the conceptual trade-off between instrumental

[4] The form, which is well-known, is $\ln(s_t/(1-s_t)) = \beta p_t + x_t^{(1)}\alpha + \xi_t$.

Table 2: Small Sample Properties of Estimators

Panel A: Average Estimates (Truth is $\beta = -0.500$)						
Sample Size	3-Step	2SLS	OLS	MM	IIV	2-Step
25	-0.504	-0.512	-0.266	-0.511	-0.543	-0.500
50	-0.502	-0.504	-0.266	-0.506	-0.528	-0.501
100	-0.502	-0.505	-0.268	-0.502	-0.510	-0.501
500	-0.502	-0.503	-0.268	-0.501	-0.502	-0.501
Panel B: Standard Deviation of Estimates						
Sample Size	3-Step	2SLS	OLS	MM	IIV	2-Step
25	0.103	0.171	0.110	0.116	0.227	0.090
50	0.069	0.109	0.074	0.072	0.154	0.062
100	0.048	0.075	0.051	0.051	0.095	0.043
500	0.022	0.034	0.023	0.023	0.038	0.020
Panel B: Mean Squared Error						
Sample Size	3-Step	2SLS	OLS	MM	IIV	2-Step
25	0.011	0.029	0.067	0.014	0.053	0.008
50	0.005	0.012	0.060	0.005	0.024	0.004
100	0.002	0.006	0.056	0.003	0.009	0.002
500	0.000	0.001	0.055	0.001	0.001	0.000

Notes: The methods-of-moments ("MM") estimator reported in the fourth column is based on the moment $E[\xi_t' \eta_t] = 0$, and is computed using a one-dimensional grid search over the bounded parameter space $[-1, 0]$.

variables and three-step estimation. In three-step estimation, the cost-shifter $x^{(2)}$ *need not be observed by the econometrician* because it does not enter the demand equation. On the other hand, an additional assumption about pricing behavior is necessary.

In the last three columns, we report results for the three alternative estimators. The methods-of-moments ("MM") estimator is based on the moment $E[\xi_t' \eta_t] = 0$. It compares favorably to the three-step estimator, though its precision is a little worse because the moment contains less identifying variation than the moment $E[\xi_t' mc_t] = 0$ employed in three-step estimation. (Indeed, we replicate the three-step estimator to numerical precision if we implement method-of-moments using $E[\xi_t' mc_t] = 0$). The indirect instrumental variables ("IIV") estimator does not perform as well as the other estimators that require more restrictions, though the differences are small for larger samples. Finally, the 2-Step estimator is consistent because $x_t^{(1)}$ affects only demand, $x_t^{(2)}$ affects only marginal costs, and $Cov(x_t^{(1)}, x_t^{(2)}) = 0$. Its bias, variance, and MSE are smaller than those of the other estimators.

For the remainder of this paper, we focus our attention on the three-step and method-of-moments estimators. Before we move on, we note the method of indirect instruments may be of particular interest to researchers in certain settings, as it performs reasonably well in our simulations and requires a different set of assumptions than those required for 2SLS. An indicator for the performance of the indirect instruments estimator is the variance of the predicted price, $\hat{Var}(\tilde{p})$. When this is small, the small-sample bias may be quite large.

3.5 Examining the Uncorrelatedness Assumption

There are a variety of reasons that marginal costs may be correlated with demand unobservables. For instance, unobserved quality is more costly to supply and also increases demand. Additionally, firms may be induced to produce further along an upward-sloping marginal cost schedule if demand is stronger. In many cases, econometric techniques, such as fixed effects methods, may capture the primary factors that drive this correlation, rendering the uncorrelatedness assumption more palatable.

To illustrate, consider the following generalized demand and cost functions:

$$\begin{aligned} h(q_{jt}; w_{jt}) &= \beta p_{jt} + x'_{jt}\alpha + D_j + F_t + E_{jt} \\ mc_{jt} &= g(q_{jt}, v_{jt}; \lambda) + x'_{jt}\gamma + U_j + V_t + W_{jt} \end{aligned}$$

Let the unobserved demand shocks be $\xi_{jt} = D_j + F_t + E_{jt}$ and the unobserved costs be $\eta_{jt} = U_j + V_t + W_{jt}$. Further assume that the $h(\cdot)$ and $g(\cdot)$ functions are known (up to parameters). If products with higher quality have higher marginal costs then $Cov(U_j, D_j) > 0$. With panel data, the econometrician can account for the relationship by estimating D_j for each firm. The residual $\xi^*_{jt} = \xi_{jt} - D_j$ is uncorrelated with U_j so that uncorrelatedness applies. Similarly, if costs are higher (or lower) in markets with high demand then F_t can be estimated with panel data such that uncorrelatedness applies. Uncorrelatedness may apply with quantity-dependent cost effects (e.g., due to capacity constraints) if the econometrician accounts for $g(q_{jt}, v_{jt}; \lambda)$ in estimation, a special case that we detail in Section 4.2.2.

Thus, if panel data permit the inclusion of product and market fixed effects then the remaining unobserved correlation, $Cov(W_{jt}, E_{jt})$, involves firm-specific demand and cost deviations within a market, and the assumption of uncorrelatedness may become palatable across a wide range of applications. This is not to argue that uncorrelatedness is universally valid. Even in the presence of product and market fixed effects, a number of mechanisms could create an empirical relationship between the unobserved demand and cost shocks. The assumption would be violated if some firms offer temporary per-unit incentives to sales representatives, or if market power in input markets creates a relationship between demand and input prices. If such mechanisms are left outside the model then the estimators derived above do not provide consistent estimates of demand. However, it still may be possible to sign the correlation between unobserved terms, in which case the bounds approach we discuss below may identify plausible ranges for the demand parameters.

3.6 Relaxing Uncorrelatedness and Constructing Bounds

It is straightforward to extend our methodology to applications where the uncorrelatedness condition does not hold. Under our supply and demand assumptions, what is required is a simple adjustment to the three-step calculation of β and an estimate of the covariance between the unobserved shocks.

Proposition 7. *Under assumptions 2 and 3, a consistent estimate of the price parameter β solves the following quadratic equation:*

$$\begin{aligned} 0 &= \beta^2 \\ &+ \left(\frac{\hat{Cov}(p^*, h'(q)q)}{\hat{Var}(p^*)} + \frac{\hat{Cov}(\xi, \eta)}{\hat{Var}(p^*)} - \hat{\beta}^{OLS} \right) \beta \\ &+ \left(-\hat{\beta}^{OLS} \frac{\hat{Cov}(p^*, h'(q)q)}{\hat{Var}(p^*)} - \frac{\hat{Cov}\left(\hat{\xi}^{OLS}, h'(q)q\right)}{\hat{Var}(p^*)} \right). \end{aligned}$$

where $\hat{\beta}^{OLS}$ is the OLS estimate and $\hat{\xi}^{OLS}$ is a vector containing the OLS residuals.

Proof. See appendix.

A simple corollary follows, which we leverage to construct bounds on the true parameter.

Corollary. *When $Cov(\xi, \eta)$ is identified, β is identified. The set of identified values has at most two points.*

Even with no a priori restrictions on the support of $Cov(\xi, \eta)$, the data will rule out covariance values above and below certain thresholds, as the resulting quadratic formula will have no solution for some values of $Cov(\xi, \eta)$. When paired with a priori information, one can trace out the implied values of β over the support, discarding any values that are ruled out by economic theory, such as those that result in a positive price elasticity or implausible marginal costs. This method may be used to construct the identified set for β under assumptions 2 and 3.

4 Generality and Extensions

4.1 Demand

The demand system of equation (4) is sufficiently flexible to nest monopolistic competition with linear demands (e.g., as in the motivating example) and the discrete choice demand models that support much of the empirical research in industrial organization. We illustrate with some typical examples:

1. *Nested logit demand:* Following the exposition of Cardell (1997), let the firms be grouped into $g = 0, 1, \ldots, G$ mutually exclusive and exhaustive sets, and denote the set of firms in group g as $\mathscr{J}_g$. An outside good, indexed by $j = 0$, is the only member of group 0. Then the left-hand-side of equation (4) takes the form

   $$h(q_{jt}; w_{jt}) \equiv \ln(q_{jt}/q_{0j}) - \sigma \ln(\overline{s}_{j|g,t})$$

   where $\overline{s}_{j|g,t} = \sum_{j \in \mathscr{J}_g} \frac{q_{jt}}{\sum_{j \in \mathscr{J}_g} q_{jt}}$ is the market share of firm j within its group. The parameter, $\sigma \in [0, 1)$, determines the extent to which consumers substitution disproportionately among firms within the same group. Our second application, developed in Section 5.2, examines this model. If the uncorrelatedness is combined with a supplemental moment, then the full set of parameters can be recovered.

2. *Random coefficients logit demand:* Modifying slightly the notation of Berry (1994), let the indirect utility that consumer $i = 1, \ldots, I$ receives from product j be

   $$u_{ij} = \beta p_j + x_j' \alpha + \xi_j + \left[\sum_k x_{jk} \sigma_k \zeta_{ik} \right] + \epsilon_{ij}$$

   where x_{jk} is the kth element of x_j, ζ_{ik} is a mean-zero consumer-specific demographic characteristic, and ϵ_{ij} is a logit error. We have suppressed market subscripts for notational simplicity. Decomposing the RHS of the indirect utility equation into $\delta_j = \beta p_j + x_j' \alpha + \xi_j$ and $\mu_{ij} = \sum_k x_{jk} \sigma_k \zeta_{ik}$, the probability that consumer i selects product j is given by the standard logit formula

   $$s_{ij} = \frac{\exp(\delta_j + \mu_{ij})}{\sum_k \exp(\delta_k + \mu_{ik})}$$

   Integrating yields the market shares: $s_j = \frac{1}{I} \sum_i s_{ij}$. Berry et al. (1995) prove that a contraction mapping recovers, for any candidate parameter vector $\tilde{\sigma}$, the vector $\delta(s, \tilde{\sigma})$ that equates these market shares to those observed in the data. This "mean valuation" is $h(s_j; \tilde{\sigma})$ in our notation. The three-step estimator can be applied to recover the price coefficient, again taking as given $\tilde{\sigma}$. This requires an expression for $h'(s, \tilde{\sigma})$, which takes the form

   $$h'(s_j; \tilde{\sigma}) = \frac{1}{\frac{1}{I} \sum_i s_{ij}(1 - s_{ij})}.$$

   Thus, uncorrelatedness assumption can recover the linear parameters given the candidate parameter vector $\tilde{\sigma}$. The identification of σ is a distinct issue that has received a great deal of attention from theoretical and applied research (e.g., Romeo (2010); Berry and Haile (2014); Gandhi and Houde (2015); Miller and Weinberg (2017)).

3. *Constant elasticity demand:* With a substitution of $f(p_{jt})$ for p_{jt} into equation (4), the constant elasticity of substitution (CES) demand model of Dixit and Stiglitz (1977) also can be incorporated:

$$\ln(q_{jt}/q_t) = \alpha + \beta \ln\left(\frac{p_{jt}}{\Pi_t}\right) + \xi_{jt}$$

   where q_t is an observed demand shifter, Π_t is a price index, and β provides the constant elasticity of demand. This model is often used in empirical research on international trade and firm productivity (e.g., De Loecker (2011); Doraszelski and Jaumandreeu (2013)). Due to the constant elasticity, profit-maximization generates $Cov(p, \xi) = 0$, and OLS produces unbiased estimates of the demand parameters. Indeed, this is an excellent illustration of our basic argument: so long as the data generating process is sufficiently well understood, it is possible to characterize the bias of OLS estimates. We opt to focus on semi-linear demand throughout this paper for analytical tractability.

Some demand systems are more difficult to reconcile with equation (4). Consider the linear demand system, $q_{jt} = \alpha_j + \sum_k \beta_{jk} p_k + \xi_{jt}$, which sometimes appears in identification proofs (e.g., Nevo (1998)) but is seldom applied empirically due to the large number of price coefficients. In principle, the system could be formulated such that $h(q_{jt}; w_{jt}) \equiv q_{jt} - \sum_{k \neq j} \beta_{jk} p_k$ and uncorrelatedness assumptions could be used to identify the β_{jj} and α_j coefficients. This would require, however, that the econometrician have other sources of identification for the β_{jk} ($j \neq k$) coefficients, which seems unlikely. The same problem arises with the almost ideal demand system of Deaton and Muellbauer (1980).

4.2 Supply

We now consider relaxing supply-side assumptions made in Section 3.2. We focus on three-step estimation, though, with the appropriate adjustments, the alternative estimators of Section 3.3 may be applied. First, we consider the case of multi-product ownership, which complicates the first-order condition faced by the firm. Under some restrictions about the transformation $h(q)$, it is straightforward to extend the three-step estimator. Second, we consider the case of non-constant marginal costs. Third, we illustrate how our method may be applied to alternative competitive environments by developing the estimator in the context of Cournot competition. Finally, we consider the classic case of perfect competition with linear supply and demand. We demonstrate how to apply extended models in the applications of Section 5.

4.2.1 Multi-Product Ownership

It is straightforward to extend the method-of-moments estimator to multi-product firms with demand for each product satisfying assumption 2:

$$h(q_{jt}; w_{jt}) = \beta p_{jt} + x'_{jt}\alpha + \xi_{jt}. \tag{11}$$

The only necessary adjustment is to the first order conditions, which must account for product ownership in the markup function (e.g., as in Berry et al. (1995) and Nevo (2001)). The following assumption provides a sufficient condition under to the three-step estimator also extends to multi-product ownership:

Assumption 6: *The derivatives of the transformation parameters with respect to contemporaneous prices, $\frac{dw_{kt}}{dp_{jt}}$, are linear in β.*

The assumption holds for several common demand systems, such as nested logit. Under this assumption, we define $\frac{dw_k}{dp_j} = \beta f_{kj}$. We omit the market subscript, t, to simplify notation. Let K^m denote the set of products owned by multi-product firm m. When the firm sets prices on each of its products to maximize joint profits, we obtain $|K^m|$ first-order conditions, which can be expressed as

$$\sum_{k \in K^m} (p_k - c_k) \frac{\partial q_k}{\partial p_j} = -q_j \ \ \forall j \in K^m.$$

For demand systems satisfying assumptions 2 and 6, $\frac{\partial q_j}{\partial p_j} = \beta \frac{1}{dh/dq_j}\left(1 - \frac{dh}{dw_j} f_{jj}\right)$ and $\frac{\partial q_k}{\partial p_j} = -\beta \frac{1}{dh/dq_k} \frac{dh}{dw_k} f_{kj}$ Therefore, the set of first-order conditions can be written as

$$\sum_{k \in K^m} (p_k - c_k) \left(\mathbf{1}[j = k] - \frac{dh/dw_k}{dh/dq_k} f_{kj} \right) = -\frac{1}{\beta} q_j \ \ \forall j \in K^m.$$

We can stack the first-order conditions, writing the LHS as the product of a vector of markups $(p_j - c_j)$ and a matrix A^m of loading components ($A^m_{i(j),i(k)} = \mathbf{1}[j = k] - \frac{1}{dh/dq_k} \frac{dh}{dw_k} f_{kj}$), where $i(\cdot)$ indexes products within a firm. We invert the loading matrix to solve for markups as function of the loading components and $-\frac{1}{\beta}\boldsymbol{q}^m$, where $\boldsymbol{q}^m$ is a vector of the multi-product firm's quantities. Optimal prices can be expressed as costs plus a markup, where the markup is determined by the inverse of A^m ($(A^m)^{-1} \equiv \Lambda^m$), quantities, and the price parameter:

$$p_j = c_j - \frac{1}{\beta} \left(\Lambda^m \boldsymbol{q}^m\right)_{i(j)}.$$

Here, $\left(\Lambda^m \boldsymbol{q}^m\right)_{i(j)}$ provides the entry corresponding to product j in the vector $\Lambda^m \boldsymbol{q}^m$. As the matrix Λ^m is not a function of the price parameter after conditioning on observables, this form of the first-order condition allows us to solve for β using a quadratic three-step solution analogous to that in equation (8).[5]

The quadratic takes the following form:

$$\begin{aligned} 0 &= \beta^2 \\ &+ \left(\frac{Cov(p^*, \tilde{h})}{Var(p^*)} - \hat{\beta}^{OLS} \right) \beta \\ &+ \left(-\frac{Cov(p^*, \tilde{h})}{Var(p^*)} \hat{\beta}^{OLS} - \frac{Cov(\hat{\xi}^{OLS}, \tilde{h})}{Var(p^*)} \right) \end{aligned}$$

where $\hat{\beta}^{OLS}$ is the OLS estimate, $\hat{\xi}^{OLS}$ is a vector of the OLS residuals, and $\tilde{h}_j \equiv (\Lambda^m \boldsymbol{q}^m)_{i(j)}$. The proof follows that of Proposition 4 using the modified first-order conditions, and so we offer the result as a corollary.

4.2.2 Non-Constant Marginal Costs

We turn now estimation of demand with firms that have non-constant marginal costs. We maintain assumptions 1 (uncorrelatedness) and 2 (demand), as well as a modified version of assumption 3 (supply) in which marginal costs depend directly on quantities:

$$mc_{jt} = x'_{jt}\gamma + g(q_{jt}; \lambda) + \eta_{jt} \tag{12}$$

The uncorrelatedness assumption alone is insufficient to identify the λ parameters, so for convenience we take $g(\cdot; \lambda)$ as known. In principle, λ could be estimated jointly with the demand parameters if an additional restriction is available (e.g., perhaps it is known that one or more demand covariates do not affect marginal cost). The OLS regression of $h(q_{jt}; w_{jt})$ on price and covariates yields a price coefficient with the following probability limit:

$$plim(\hat{\beta}^{OLS}) = \beta - \frac{1}{\beta} \frac{Cov(\xi, h'(q)q)}{Var(p^*)} + \frac{Cov(\xi, g(q))}{Var(p^*)} \tag{13}$$

Bias arises in the OLS coefficient due to markup adjustments, as before, and because unobserved changes in demand are correlated with changes in marginal costs through $g(q; \lambda)$. With the adjusted bias term, we develop a three-step estimator akin to that of Section 3.2:

Proposition 8. *Under assumptions 1-2 and a modified assumption 3 such marginal costs take the semi-linear form of equation 12, a consistent estimate of the price parameter β solves the*

[5] At this point, the reader may be wondering where the prices of other firms are captured under the adjusted first-order conditions for multi-product ownership. As is the case with single product firms, we expect prices of other firm's products to be included in w_j, which is appropriate under Bertrand price competition.

following quadratic equation:

$$\begin{aligned}
0 &= \left(1 - \frac{Cov(p^*, g(q))}{Var(p^*)}\right)\beta^2 \\
&+ \left(\frac{Cov(p^*, h'(q)q)}{Var(p^*)} - \hat{\beta}^{OLS} + \frac{Cov(p^*, g(q))}{Var(p^*)}\hat{\beta}^{OLS} + \frac{Cov(\hat{\xi}^{OLS}, g(q))}{Var(p^*)}\right)\beta \\
&+ \left(-\frac{Cov(p^*, h'(q)q)}{Var(p^*)}\hat{\beta}^{OLS} - \frac{Cov(\hat{\xi}^{OLS}, h'(q)q)}{Var(p^*)}\right)
\end{aligned}$$

where $\hat{\beta}^{OLS}$ is the OLS estimate and $\hat{\xi}^{OLS}$ is a vector containing the OLS residuals.

Proof. See appendix.

It is straightforward to extend the function $g(\cdot;\lambda)$ to include covariates that are not exogenous with respect to demand shocks, v_{jt}. As above, we would proceed as if $g(q_{jt}, v_{jt}; \lambda)$ is known in three-step estimation.

4.2.3 Cournot Competition

The estimators developed so far rely on an assumption of Bertrand competition. Yet the identification strategy we examine is substantially more general—what is required is that the model lays out how prices are set in the market. To illustrate, in this subsection we develop results for another workhorse model in empirical industrial organization, that of undifferentiated Cournot competition. Let there be a market-specific demand curve $h(Q; w) = \beta p + x'\gamma + \xi$, where we have omitted market subscripts for notational brevity. Total market quantity in the market is given by $Q = \sum_j q_j$, and p represents a price that is common to all firms in the market. The marginal costs of each firm are semi-linear, as specified in equation (12). These supply-side assumptions accord with those of recent empirical articles in which non-linear portion of the marginal cost function captures the effects of capacity constraints (e.g., Ryan (2012); Fowlie et al. (2016)).

Working with aggregated first order conditions, it is possible to show that the OLS regression of $h(Q; w_{jt})$ on price and covariates yields a price coefficient with the following probability limit:

$$plim(\hat{\beta}^{OLS}) = \beta - \frac{1}{\beta}\frac{1}{J}\frac{Cov(\xi, h'(Q)Q)}{Var(p^*)} + \frac{Cov(\xi, \overline{g})}{Var(p^*)} \tag{14}$$

where J is the number of firms in the market and $\overline{g} = \frac{1}{J}\sum_{j=1}^{J} g(q_j; \lambda)$ is the average contribution of $g(q, \lambda)$ to marginal costs. Bias arises due to markup adjustments and the correlation between unobserved demand and marginal costs generated through $g(q; \lambda)$. It is noteworthy that bias due to markup adjustments dissipates as the number of firms grows large. Thus, if marginal costs are constant then the OLS estimate is likely to be close to the population parameter in competitive markets.[6] Even with finite J and non-constant marginal costs, the method-of-moments estimator of Section 3.3.1 can be applied to obtain consistent estimates of β. Alternatively, a consistent estimator can be obtained as the solution to the following quadratic formula:

$$\begin{aligned}
0 &= \left(1 - \frac{Cov(p^*, \overline{g})}{Var(p^*)}\right)\beta^2 \\
&+ \left(\frac{1}{J}\frac{Cov(p^*, h'(Q)Q)}{Var(p^*)} - \hat{\beta}^{OLS} + \frac{Cov(p^*, \overline{g})}{Var(p^*)}\hat{\beta}^{OLS} + \frac{Cov(\hat{\xi}^{OLS}, \overline{g})}{Var(p^*)}\right)\beta \\
&+ \left(-\frac{1}{J}\frac{Cov(p^*, h'(Q)Q)}{Var(p^*)}\hat{\beta}^{OLS} - \frac{1}{J}\frac{Cov(\hat{\xi}^{OLS}, h'(Q)Q)}{Var(p^*)}\right)
\end{aligned}$$

The derivation of the quadratic tracks exactly the proof of Proposition 8 and so we offer the result as a simple corollary.

4.2.4 Perfect Competition

In our final supply-side extension, we consider the case of perfect competition with linear demand and supply curves. Let marginal costs be given by $mc = x'\gamma + \lambda q + \eta$. The first order condition for each firm is $P = \gamma + \lambda q + \eta$ so the firm-specific supply curve is $q^s = -\frac{\gamma}{\lambda} + \frac{1}{\lambda}p - \frac{\eta}{\lambda}$. This obtains the following market-level demand and supply curves:

$$\begin{aligned}
Q^D &= \beta p + x'\alpha + \xi \\
Q^S &= \frac{J}{\lambda}p - \frac{J}{\lambda}x'\gamma - \frac{J}{\lambda}\eta
\end{aligned}$$

The OLS estimation of demand yields a coefficient with the standard probability limit:

$$\beta^{OLS} \equiv plim(\hat{\beta}^{OLS}) = \beta + \frac{Cov(\xi, p^*)}{Var(p^*)}$$

Tracing the same sequence of steps provided in Section 2 for the monopoly model, it is possible to show that uncorrelatedness implies

$$Cov(\xi, q) = Cov(\xi^{OLS}, Q) + \frac{\lambda}{J}\frac{Cov(\xi, Q)}{Var(p^*)}Cov(p^*, Q)$$

[6]While this explicit analytical result is particular to the Cournot model, we find that a similar effect arises in Monte Carlo experiments based on Bertrand competition and logit demand.

where ξ^{OLS} is a vector of OLS residuals. Solving for $Cov(\xi, q)$ and plugging into the probability limit of the OLS estimator yields

$$\beta = \beta^{OLS} - \frac{1}{\frac{J}{\lambda} - \beta^{OLS}} Cov(\xi^{OLS}, Q) \tag{15}$$

Again, the model provides the necessary correction to the OLS estimator. Notably, the slope of the supply curve, $\frac{J}{\lambda}$, is an input to the bias correction. Thus, parallel with our results for oligopoly models, uncorrelatedness alone is insufficient for identification of the demand parameters in the presence of non-constant marginal costs. However, if one or more demand covariates can be excluded from the marginal cost function then joint estimation of demand and supply is possible even without an instrument for price in the demand equation.

5 Applications

5.1 The Portland Cement Industry

Our first application shows how the uncorrelatedness assumption can identify demand parameters in a model of Cournot competition. We work with the empirical example presented in Fowlie et al. (2016) ["FRR"] on the Portland cement industry. Firms produce a homogeneous product and compete in quantities. The marginal cost of firm j in market t is

$$mc_{jt}(q_{jt}; \gamma, \lambda) = \gamma + 2\lambda_2 1\{q_{jt}/k_{jt} > \lambda_1\}(q_{jt}/k_{jt} - \lambda_1) + \eta_{jt}$$

where k_{jt} and q_{jt}/k_{jt} are capacity and utilization, respectively. The marginal cost function is constant if utilization (q_{jt}/k_{jt} for capacity k_{jt}) is less then the threshold λ_1, and linearly increasing otherwise. FRR specify a constant demand curve:

$$\ln Q_t = \alpha_t + \beta \ln P_t + \xi_t$$

where α_t are market-specific intercepts, P_t is the uniform market price, and $Q_t = \sum_j q_{jt}$ is total market quantity. Among the estimators developed in Section 3, only the method-of-moments is applicable because constant elasticity demand does not comport with the functional form restrictions of equation (4). Thus, to more fully demonstrate the range of estimation strategies, we also estimate a logit demand curve:

$$\ln(Q_t) - \ln(M_t - Q_t) = \alpha_t^{alt} + \beta^{alt} P_t + \xi_t^{alt}$$

Table 3: Application to Portland Cement

Estimator:	3-Step	MM	2SLS	OLS	MM	2SLS	OLS
Demand:	logit	logit	logit	logit	constant elasticity	constant elasticity	constant elasticity
Mean Elasticity	-1.19	-1.07	-1.10	-0.48	-1.12	-1.22	-0.47
	(0.18)	(0.19)	(0.19)	(0.15)	(0.32)	(0.23)	(0.17)

Notes: We report the mean price elasticity of demand for comparability across columns. With constant elasticity demand, the mean elasticity equals the price coefficient, β. With logit demand, the market-specific elasticities are given by $\beta^{alt}P_t(1-Q_t/M_t)$, and these are averaged across the 520 observations. All regressions include region fixed effects. Bootstrapped standard errors are calculated based on 200 random samples constructing by drawing regions with replacement.

where M_t is the size of the market. The first order conditions track those of Section 4.2.3:

$$P_t - 2\lambda_2 1\{q_{jt}/k_j > \lambda_1\}(q_{jt}/k_j - \lambda_1) = \gamma - \left[\frac{\partial Q_t(P_t)}{\partial P_t}\right]^{-1} q_{jt} + \eta_{jt}$$

The sample includes 520 region-year observations over 1984-2009. We focus on estimating the price parameter, β. The nonlinear cost parameters (λ_1, λ_2) parameters are not identified from the uncorrelatedness assumption alone so we take as given the FRR estimates: $\lambda_1 = 0.869$ and $\lambda_2 = 803.65$. The remaining parameters—region fixed effects in the demand equation and an intercept in the cost equation—can be expressed as linear functions of data given the price and nonlinear cost parameters and thus are of limited interest. We estimate with OLS and 2SLS using the FRR instruments in the latter case: coal prices, natural gas prices, electricity prices, and wage rates. We also estimate without instruments using the method-of-moments and three-step estimators. With logit demand, we set the market size as three times the observed maximum quantity within the region, though results are robust to alternative market size assumptions.

Estimation results are summarized in Table 3. For comparability between the logit (columns (1)-(4)) and constant elasticity (columns (5)-(7)) demand systems, we report the mean price elasticity of demand implied by the estimates rather than the price coefficient itself. As shown the three-step, method-of-moments, and 2SLS regression all obtain similar estimates of the mean elasticity, with a range across specifications of -1.22 to -1.07.[7] Thus, the different identifying restrictions appear to be consistent with each other. The OLS estimates, which we provide as a point of comparison, imply less elastic demand. Interestingly, if we estimate the logit system using uncorrelatedness under the alternative assumption of constant marginal costs (e.g., $\lambda_2 = 0$) we obtain an elasticity estimate of -0.79, near the

[7] FRR estimate the constant elasticity demand system using limited information maximum likelihood and obtain somewhat larger elasticities (in magnitude). Their main results, developed in a series of counterfactual simulations, are robust across a reasonably wide range of demand elasticities.

midpoint between the OLS and MM estimates in Table 3. Markup adjustments and non-constant marginal costs both contribute to the bias present in the OLS estimates.

5.2 Demand in the U.S. Airline Industry

Our second application estimates demand in the U.S. airline industry. We follow closely the analysis of Aguirregabiria and Ho (2012) ["AH"], which features Bertrand competition and nested logit demand.[8] The markets in question are directional round trips between origin and destination cities in a particular quarter.[9] Each product j is either a nonstop or a one-stop itinerary from a particular airline. Thus, airlines can have zero, one, or two products per market. The outside good is indexed as $j = 0$ and may be interpreted as alternative travel or no trip. A nesting parameter, σ, governs the correlations of unobserved consumer preferences across three product groups: nonstop flights, one-stop flights, and the outside good. This yields the demand relationship

$$\ln s_{jt} - \ln s_{0t} - \sigma \ln \overline{s}_{jt|g} = \beta p_{jt} + x_{jt}'\alpha + \xi_{jt} \tag{16}$$

where s_{jt} and $\overline{s}_{jt|g}$ are the market share and conditional market share of product j in market t, respectively. The covariates include an indicator for nonstop itineraries, the distance between the origin and destination cities, a measure of the airline's "hub sizes" at the origin and destination cities, airline fixed effects, and market fixed effects.[10] Marginal costs are constant and follow the linear structure $mc_{jt} = X_{jt}\gamma + \eta_{jt}$. The first order conditions are

$$p_t = X_t\gamma + \eta_t - \left[\Omega \circ \frac{\partial s_t(p_t)}{\partial p_t}\right]^{-1} s_t(p_t)$$

where the matrix Ω summarizes product ownership (e.g., as in Nevo (2001)), the operation $\circ$ is element-by-element multiplication, the matrix X_t includes the exogenous covariates, and the elements of η_t are unobserved cost shocks.

The data are drawn from the *Airline Origin and Destination Survey* (DB1B) survey, a ten percent sample of airline itineraries. Quarterly observations from the year 2004 are observed for each directional origin-destination pair. We estimate the nested logit demand system using the methodology in this paper, as we discuss below. As a baseline, we follow the AH specifications to estimate demand using OLS and 2SLS. The AH instruments are: the average hub-sizes of all other airlines on the route, the average hub-sizes of all legacy

[8]We thank Victor Aguirregabiria for providing the data. Our sample differs slightly from the one used in the AH publication so replication is not exact.

[9]Thus, "market" fixed effects account for location and period.

[10]In the demand equation, hub size of any given city-airline pair is the sum of population in other cities that the airline connects with direct itineraries from the city. In the supply equation, this is replaced with an analogous measure based on the number of connections rather than population. Market size, which determines the market share of the outside good, is equal to the total population in the origin and destination cities.

airlines on the route, the average value of the nonstop indicator for all the other carriers on the route, and an indicator for the presence of Southwest. The price coefficient and the nesting parameter are reported in the last two columns in Table 4. For all specifications, we construct standard errors via subsampling.[11]

The uncorrelatedness assumption, $Cov(\xi, \eta) = 0$, identifies (β, α, γ) for a given value of the σ parameter. In other words, $\beta^{\text{3-Step}}$ is a function of σ. As discussed in Section 4, the nested logit demand system requires a supplemental restriction to identify the nesting parameter, as the conditional market share is also endogenous. We proceed by picking the value of σ (and the corresponding three-step estimator) that minimizes a set of additional moments.

We provide our three-step estimates under alternative supplemental assumptions. The first uses the data structure employed in the AH analysis, in which some X covariates are in the marginal cost equation but are excluded from demand. We use the first-order condition to calculate mc_{jt} and calculate the cost residuals by regressing our estimates of mc_{jt} on the supply-side X covariates. We then minimize the correlation between η and ξ. The excluded supply-side covariates provide sufficient identifying variation as they are not used in construction of the three-step estimator. As a second approach, we use the exclusion restrictions implied by the AH instruments as our additional moments.

In the third specification, we consider an approach that is more easy to generalize to other settings, when the exclusion restrictions of the previous two approaches may not be palatable. We build off of the uncorrelatedness intuition to consider the case when unobserved demand and costs shocks are uncorrelated *across product groups within a market*. Let $\overline{\xi}_{gt} = \frac{1}{|g|}\sum_{j\in g}\xi_{jt}$ and $\overline{\eta}_{gt} = \frac{1}{|g|}\sum_{j\in g}\eta_{jt}$ be the mean demand and cost shocks within a group-market, after removing market (i.e., route-quarter) fixed effects. Collecting these values into the vectors $\overline{\xi}$ and $\overline{\eta}$, we impose the supplemental assumption $Cov(\overline{\xi}, \overline{\eta}) = 0$. To construct these group-market estimates, we use only the demand-side covariates to estimate η.[12] The results are reported in Table 4.

The first approach (3-Step-I), provides a price coefficient that is statistically identical to that of 2SLS (-0.182). However, it produces a nesting parameter that is significantly smaller (0.53 compared to 0.82).[13] Though we consider our estimator an alternative identification strategy that is not necessarily superior to instrumental variables, we note here that one of the challenges in nested logit estimation is finding credible instruments for the conditional share, as it is an outcome variable in the demand system. If uncorrelatedness holds in the data-generating process, then these results indicate that the 2SLS instruments are not

[11] We subsample 100 market-period observations out of the 11,474 market-periods in our data. This is analogous to clustering by market-period.

[12] Other supplemental assumptions are possible; we obtain identification if we assume that ξ and η are independent and implement with the moment restrictions $Cov(\xi^2, \eta) = 0$ and $Cov(\xi, \eta^2) = 0$.

[13] With the 2SLS estimates, $Cor(\xi, \eta) = 0.23$ and $Cor(\overline{\xi}, \overline{\eta}) = 0.09$. Ciliberto et al. (2016) estimate that $Cor(\xi, \eta) \in [0.38, 0.40]$ based on 2012 data.

Table 4: Application to U.S. Airlines

Parameter	3-Step-I	3-Step-II	3-Step-III	MM-I	MM-II	OLS	2SLS
β	-0.182	-0.107	-0.153	-0.106	-0.152	-0.106	-0.189
	(0.004)	(0.006)	(0.003)	(0.007)	(0.004)	(0.004)	(0.053)
σ	0.525	0.669	0.599	0.660	0.595	0.891	0.822
	(0.010)	(0.015)	(0.011)	(0.016)	(0.016)	(0.003)	(0.087)

Notes: We report the estimated price and nesting parameters for the nested logit demand system under seven specifications. The first three columns of results use the three-step methodology with different supplemental moments. The next two columns report method-of-moments analogs to 3-Step-II and 3-Step-III, relaxing an assumption used in three-Step estimation. The final two columns are OLS and 2SLS specifications that parallel the estimation of Aguirregabiria and Ho (2012). Standard errors are constructed via subsamples of 100 market-periods. There are 93,199 observations and 11,474 market-periods in the full sample.

sufficiently exogenous with respect to the conditional share. 3-Step-II and 3-Step-III also provide nesting parameters that are significantly smaller than the 2SLS estimate. Demand is less elastic with the three-step parameters compared to the 2SLS parameters. Though customers have essentially the same price sensitivity (in utility terms) under 3-Step-I and 2SLS, the smaller value for σ in the three-step estimation indicates that they consider flights within the same group (nonstop or one-stop) relatively more differentiated than would be implied by 2SLS. Whether these estimates are more in line with real-world behavior is outside the scope of this paper.

As a robustness check, we relax the assumption that all of the supply-side observables are uncorrelated with the unobserved demand shock. This invalidates the three-step estimator, but allows us to pursue a method-of-moments approach in which our baseline moment is the supplemental uncorrelatedness moment used in 3-Step-I. For additional moments, we use the AH instruments (MM-I) and the group uncorrelatedness assumptions (MM-II) discussed above. The alternative method-of-moments estimates are not significantly different from their three-step counterparts (3-Step-II and 3-Step-III). The AH instruments generate a price sensitivity in line with the OLS estimate, and the group moments lead to an estimate that lies between those generated by the other assumptions.

References

Aguirregabiria, V. and C.-Y. Ho (2012, May). A dynamic oligopoly game of the US airline industry: Estimation and policy experiments. *Journal of Econometrics 168*(1), 156–173.

Berry, S. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Berry, S., J. Levinsohn, and A. Pakes (1995, July). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2), 242–262.

Byrne, D. P., S. Imai, N. Jain, V. Sarafidis, and M. Hirukawa (2016). Identification and estimation of differentiated products models using market size and cost data.

Cardell, S. N. (1997). Variance components structures for the extreme-value and logistic distributions with applications to models of heterogeneity. *Econometric Theory 13*, 185–213.

Ciliberto, F., C. Murry, and E. Tamer (2016). Market structure and competition in airline markets.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(3), 1407–1451.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American Economic Review 70*(3), pp. 312–326.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review 67*(3), 297–308.

Doraszelski, U. and J. Jaumandreeu (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies 80*, 1338–1383.

Fowlie, M., M. Reguant, and S. P. Ryan (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy 124*(1), 249–302.

Gandhi, A. and J.-F. Houde (2015). Measuring substitution pattern in differentiated products industries.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods, Studies in Income and Wealth* Vol 58, ed. by T. Bresnahan and R. Gordon. Chicago: National Bureau of Economic Research.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251–1271.

Hausman, J. A. and W. E. Taylor (1983). Identification in linear simultaneous equations models with covariance restrictions: An instrumental variables interpretation. *Econometrica 51*(5), 1527–1549.

Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153.

Koopmans, T. C., H. Rubin, and R. B. Leipnik (1950). *Statistical Inference in Dynamic Economic Models (Cowles Commission Monograph 10)*, Chapter Measuring the Equation Systems of Dynamic Economics. New York: John Wiley & Sons.

Matzkin, R. L. (2004). Unobservable instruments. Unpublished manuscript.

Miller, N. and M. Weinberg (2017). Understanding the prices effects of the MillerCoors joint venture. *Econometrica*.

Nevo, A. (1998). Identification of the oligopoly solution concept in a differentiated-products industry. *Economics Letters 59*, 391–395.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Romeo, C. (2010). Filling out the instrument set in mixed logit demand systems for aggregate data. Available at SSRN: https://ssrn.com/abstract=1593590 or http://dx.doi.org/10.2139/ssrn.1593590.

Rosse, J. N. (1970, March). Estimating Cost Function Parameters Without Using Cost Data: Illustrated Methodology. *Econometrica 38*(2), 256.

Ryan, S. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica 80*(3), 1019–1062.

Working, E. J. (1927, February). What Do Statistical "Demand Curves" Show? *The Quarterly Journal of Economics 41*(2), 212.

A Proofs

Proof of Proposition 3 (Plim of OLS Estimator)

From the text, we have $\hat{\beta}^{OLS} \xrightarrow{p} \beta + \frac{Cov(p^*,\xi)}{Var(p^*)}$. Thus, we desire to show that $\frac{Cov(p^*,\xi)}{Var(p^*)} = -\frac{1}{\beta+\frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p^*)}$.

Given the demand assumptions, the firm's first-order condition is $p = c - \frac{1}{\beta}h'(q)q$. We can write $p = p^* + \hat{p}$, where $\hat{p}$ is the projection of p onto the exogenous demand variables, X. If we substitute the first-order condition into the numerator of interest, we obtain

$$\begin{aligned} Cov(p^*, \xi) &= Cov(c - \frac{1}{\beta}h'(q)q - \hat{p}, \xi) \\ &= -\frac{1}{\beta}Cov(\xi, h'(q)q), \end{aligned}$$

where the second line follows from the uncorrelatedness assumption (1) and the exogeneity assumption ($E[X\xi] = 0$).

To the complete the proof, we show that the following holds:

$$\frac{1}{\beta}\frac{Cov\left(\xi, h'(q)q\right)}{Var(p^*)} = \frac{1}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p^*)}.$$

The unobserved demand shock may be written as $\xi = h(q) - x\alpha - \beta p$. At the probability limit of the OLS estimator, we can construct the unobserved demand shock as $\xi = \xi^{OLS} + \left(\beta^{OLS} - \beta\right)p^*$.[14] From the prior step in this proof, $\beta^{OLS} - \beta = -\frac{1}{\beta}\frac{Cov\left(\xi,h'(q)q\right)}{Var(p^*)}$. Therefore, $\xi^{OLS} = \xi + \frac{1}{\beta}\frac{Cov\left(\xi,h'(q)q\right)}{Var(p^*)}p^*$. This implies

$$\begin{aligned} Cov\left(\xi^{OLS}, h'(q)q\right) &= \left(1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)}\right)Cov(\xi, h'(q)q) \\ \implies -\frac{1}{\beta}\frac{Cov\left(\xi, h'(q)q\right)}{Var(p^*)} &= -\frac{1}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p^*)} \end{aligned}$$

Thus, we have rewritten the OLS bias as a function of known or estimated quantities and the true parameter. QED.

Proof of Proposition 4 (Three-Step Estimator)

From equation (7), we can obtain a quadratic formula for β.

[14] For a proof, see a subsequent section in the Appendix.

$$\beta^2 + \beta\left(\frac{Cov\left(p^*, h'(q)q\right)}{Var(p^*)} - \beta^{OLS}\right) - \frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p^*)} - \frac{Cov\left(p^*, h'(q)q\right)}{Var(p^*)}\beta^{OLS} = 0$$

The roots of the quadratic are $\left(\beta, \frac{Cov(p^*,\xi)}{Var(p^*)} - \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}\right)$, which can be verified by the reader. Therefore, under the auxiliary condition $\beta < \frac{Cov(p^*,\xi)}{Var(p^*)} - \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}$, the lower root identifies the true parameter, β. As discussed in the text, the auxiliary condition will generally be the case, and can also be verified by checking both roots. If one is positive, then the other is the true parameter.

The formula provided in Proposition 4 is the sample analog of the lower root :

$$\frac{1}{2}\left(\hat{\beta}^{OLS} - \frac{\hat{Cov}\left(p^*, h'(q)q\right)}{\hat{Var}(p^*)} - \sqrt{\left(\hat{\beta}^{OLS} + \frac{\hat{Cov}\left(p^*, h'(q)q\right)}{\hat{Var}(p^*)}\right)^2 + 4\frac{\hat{Cov}\left(\hat{\xi}^{OLS}, h'(q)q\right)}{\hat{Var}(p^*)}}\right).$$

As the sample analog functions $\hat{Cov}$ and $\hat{Var}$ are consistent for Cov and Var, this formula provides a consistent estimator of the lower root. QED.

Proof of Proposition 5 (Two-Step Estimator)

Suppose that, in addition to assumptions 1-3, that marginal costs are uncorrelated with the exogenous demand factors (Assumption 4). Then, the expression $\frac{1}{\beta + \frac{Cov\left(p,h'(q)q\right)}{Var(p)}}\frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p)}$ is equal to $\frac{1}{\beta + \frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p^*)}$.

Assumption 4 implies $Cov(\hat{p}, c) = 0$, allowing us to obtain

$$\begin{aligned}
Cov(\hat{p}, \beta(\hat{p} + p^* - c)) &= \beta Var(\hat{p}) \\
Cov(p - p^*, \beta(\hat{p} + p^* - c)) &= \beta Var(p) - \beta Var(p^*) \\
Var(p)\beta + Cov\left(p, h'(q)q\right) &= Var(p^*)\beta + Cov\left(p^*, h'(q)q\right) \\
\left(\beta + \frac{Cov\left(p, h'(q)q\right)}{Var(p)}\right)\frac{1}{Var(p^*)} &= \left(\beta + \frac{Cov\left(p^*, h'(q)q\right)}{Var(p^*)}\right)\frac{1}{Var(p)} \\
\frac{1}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p^*)} &= \frac{1}{\beta + \frac{Cov(p,h'(q)q)}{Var(p)}}\frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p)}.
\end{aligned}$$

Therefore, the probability limit of the OLS estimator can be written as:

$$\text{plim}\hat{\beta}^{OLS} = \beta - \frac{1}{\beta + \frac{Cov(p,h'(q)q)}{Var(p)}}\frac{Cov\left(\xi^{OLS}, h'(q)q\right)}{Var(p)}.$$

The roots of the implied quadratic are:

$$\frac{1}{2}\left(\beta^{OLS}-\frac{Cov\left(p,h'(q)q\right)}{Var(p)}\pm\sqrt{\left(\beta^{OLS}+\frac{Cov\left(p,h'(q)q\right)}{Var(p)}\right)^2+4\frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p)}}\right)$$

which are equivalent to the pair $\left(\beta,\beta\left(1-\frac{Var(p^*)}{Var(p)}\right)+\frac{Cov(p^*,\xi)}{Var(p^*)}-\frac{Cov\left(p^*,h'(q)q\right)}{Var(p)}\right)$. Therefore, with the auxiliary condition $\beta<\frac{Cov(p^*,\xi)}{Var(p^*)}\frac{Var(p)}{Var(p^*)}-\frac{Cov\left(p^*,h'(q)q\right)}{Var(p^*)}$, the lower root is consistent for β. QED.

Proof of Proposition 6 (Indirect Instruments)

We desire to show $\beta=\frac{Cov\left(\tilde{p},-h'(q)q\right)}{Var(\tilde{p})}$, where $\tilde{p}$ is the residual of a regression of p on X_1.

The first-order condition gives $-h'(q)q=\beta\left(p-c\right)$. Therefore, the numerator is equal to

$$\begin{aligned}Cov\left(\tilde{p},\beta\left(p-c\right)\right)&=Cov\left(\tilde{p},\beta\left(\tilde{p}+\left(p-\tilde{p}\right)-c\right)\right)\\&=\beta\left(Var(\tilde{p})-Cov(\tilde{p},c)\right).\end{aligned}$$

Therefore, $\frac{Cov\left(\tilde{p},-h'(q)q\right)}{Var(\tilde{p})}=\beta\left(1-\frac{Cov(\tilde{p},c)}{Var(\tilde{p})}\right)$

Under the additional assumption 5, $\frac{Cov(\tilde{p},c)}{Var(\tilde{p})}=0$. By consistency of the sample covariance functions, we have consistency of the estimator. QED.

Proof of Proposition 7 (Relaxing Uncorrelatedness)

From the text, we have $\hat{\beta}^{OLS}\xrightarrow{p}\beta+\frac{Cov(p^*,\xi)}{Var(p^*)}$.

The general form for a firm's first-order condition is $p=c+\mu$, where c is the marginal cost and μ is the markup. We can write $p=p^*+\hat{p}$, where $\hat{p}$ is the projection of p onto the exogenous demand variables, X. By assumption, $c=X\gamma+\eta$.

If we substitute the first-order condition $p^*=X\gamma+\eta+\mu-\hat{p}$ into the bias term from the OLS regression, we obtain

$$\begin{aligned}\frac{Cov(p^*,\xi)}{Var(p^*)}&=\frac{Cov(\xi,X\gamma+\eta+\mu-\hat{p})}{Var(p^*)}\\&=\frac{Cov(\xi,\eta)}{Var(p^*)}+\frac{Cov(\xi,\mu)}{Var(p^*)}\end{aligned}$$

where the second line follows from the exogeneity assumption ($E[X\xi]=0$). Under our demand assumption, the unobserved demand shock may be written as $\xi=h(q)-x\alpha-\beta p$. At the probability limit of the OLS estimator, we can construct the unobserved demand shock as

$\xi = \xi^{OLS} + \left(\beta^{OLS} - \beta\right) p^*$.[15] From the prior step in this proof, $\beta^{OLS} - \beta = \frac{Cov(\xi,\eta)}{Var(p^*)} + \frac{Cov(\xi,\mu)}{Var(p^*)}$. Therefore, $\xi = \xi^{OLS} + \left(\frac{Cov(\eta,\xi)}{Var(p^*)} + \frac{Cov(\mu,\xi)}{Var(p^*)}\right) p^*$. This implies

$$\begin{aligned}
\frac{Cov\,(\xi,\mu)}{Var(p^*)} &= \frac{Cov\left(\xi^{OLS},\mu\right)}{Var(p^*)} + \left(\frac{Cov(\xi,\eta)}{Var(p^*)} + \frac{Cov\,(\xi,\mu)}{Var(p^*)}\right)\frac{Cov(p^*,\mu)}{Var(p^*)} \\
\frac{Cov\,(\xi,\mu)}{Var(p^*)}\left(1 - \frac{Cov(p^*,\mu)}{Var(p^*)}\right) &= \frac{Cov\left(\xi^{OLS},\mu\right)}{Var(p^*)} + \frac{Cov(\xi,\eta)}{Var(p^*)}\frac{Cov(p^*,\mu)}{Var(p^*)} \\
\frac{Cov\,(\xi,\mu)}{Var(p^*)} &= \frac{1}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},\mu\right)}{Var(p^*)} + \frac{1}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov(\xi,\eta)}{Var(p^*)}\frac{Cov(p^*,\mu)}{Var(p^*)}
\end{aligned}$$

When we substitute this expression in for β^{OLS}, we obtain

$$\begin{aligned}
\beta^{OLS} &= \beta + \frac{Cov(\xi,\eta)}{Var(p^*)} + \frac{1}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},\mu\right)}{Var(p^*)} + \frac{\frac{Cov(p^*,\mu)}{Var(p^*)}}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov(\xi,\eta)}{Var(p^*)} \\
\beta^{OLS} &= \beta + \frac{1}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},\mu\right)}{Var(p^*)} + \frac{1}{1 - \frac{Cov(p^*,\mu)}{Var(p^*)}}\frac{Cov(\xi,\eta)}{Var(p^*)}
\end{aligned}$$

Thus, we obtain an expression for the OLS estimator in terms of the OLS residuals, the residualized prices, the markup, and the correlation between unobserved demand and cost shocks. If the markup can be parameterized in terms of observables and the correlation in unobserved shocks can be calibrated, we have a method to estimate β from the OLS regression.

Under our supply and demand assumptions, $\mu = -\frac{1}{\beta}h'(q)q$. And

$$\beta^{OLS} = \beta - \frac{1}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p^*)} + \beta\frac{1}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)}}\frac{Cov(\xi,\eta)}{Var(p^*)}.$$

Thus, we obtain the following quadratic equation for the true parameter in terms of observables and the covariance between ξ and η:

$$\begin{aligned}
0 &= \beta^2 \\
&+ \left(\frac{Cov(p^*,h'(q)q)}{Var(p^*)} + \frac{Cov(\xi,\eta)}{Var(p^*)} - \beta^{OLS}\right)\beta \\
&+ \left(-\beta^{OLS}\frac{Cov(p^*,h'(q)q)}{Var(p^*)} - \frac{Cov\left(\xi^{OLS},h'(q)q\right)}{Var(p^*)}\right).
\end{aligned}$$

The quadratic provided in the main text is the sample analog of this equation, and all of the sample functions are consistent. QED.

[15] For a proof, see a subsequent section in the Appendix.

Proof of Proposition 8 (Non-Constant Marginal Costs)

Under the semi-linear marginal cost schedule of equation (12), the plim of the OLS estimator is equal to

$$\text{plim}\hat{\beta}^{OLS} = \beta + \frac{Cov(\xi, g(q))}{Var(p^*)} - \frac{1}{\beta}\frac{Cov\left(\xi, h'(q)q\right)}{Var(p^*)}.$$

This is obtain directly by plugging in the first–order condition for p: $Cov(p^*, \xi) = Cov(g(q) + \eta - \frac{1}{\beta}h'(q)q - \hat{p}, \xi) = Cov(\xi, g(q)) - \frac{1}{\beta}Cov(\xi, h'(q)q)$ under the assumptions.

Next, we re-express the terms including the unobserved demand shocks in in terms of OLS residuals.

The unobserved demand shock may be written as $\xi = h(q) - x\beta_x - \beta p$. The estimated residuals are given by $\xi^{OLS} = \xi + \left(\beta - \beta^{OLS}\right)p^*$. As $\beta - \beta^{OLS} = \frac{1}{\beta}\frac{Cov\left(\xi, h'(q)q\right)}{Var(p^*)} - \frac{Cov(\xi, g(q))}{Var(p^*)}$, we obtain $\xi^{OLS} = \xi + \left(\frac{1}{\beta}\frac{Cov\left(\xi, h'(q)q\right)}{Var(p^*)} - \frac{Cov(\xi, g(q))}{Var(p^*)}\right)p^*$. This implies

$$\begin{aligned} Cov\left(\xi^{OLS}, h'(q)q\right) &= \left(1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)}\right)Cov(\xi, h'(q)q) - \frac{Cov(p^*, h'(q)q)}{Var(p^*)}Cov(\xi, g(q)) \\ Cov\left(\xi^{OLS}, g(q)\right) &= \frac{1}{\beta}\frac{Cov(p^*, g(q))}{Var(p^*)}Cov\left(\xi, h'(q)q\right) + \left(1 - \frac{Cov(p^*, g(q))}{Var(p^*)}\right)Cov(\xi, g(q)) \end{aligned} \tag{17}$$

We write the system of equations in matrix form and invert to solve for the covariance terms that include the unobserved demand shock:

$$\begin{bmatrix} Cov(\xi, h'(q)q) \\ Cov(\xi, g(q)) \end{bmatrix} = \begin{bmatrix} 1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} & -\frac{Cov(p^*, h'(q)q)}{Var(p^*)} \\ \frac{1}{\beta}\frac{Cov(p^*, g(q))}{Var(p^*)} & 1 - \frac{Cov(p^*, g(q))}{Var(p^*)} \end{bmatrix}^{-1} \begin{bmatrix} Cov(\xi^{OLS}, h'(q)q) \\ Cov(\xi^{OLS}, g(q)) \end{bmatrix}$$

where

$$\begin{bmatrix} 1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} & -\frac{Cov(p^*, h'(q)q)}{Var(p^*)} \\ \frac{1}{\beta}\frac{Cov(p^*, g(q))}{Var(p^*)} & 1 - \frac{Cov(p^*, g(q))}{Var(p^*)} \end{bmatrix}^{-1} =$$

$$\frac{1}{1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} - \frac{Cov(p^*, g(q))}{Var(p^*)}} \begin{bmatrix} 1 - \frac{Cov(p^*, g(q))}{Var(p^*)} & \frac{Cov(p^*, h'(q)q)}{Var(p^*)} \\ -\frac{1}{\beta}\frac{Cov(p^*, g(q))}{Var(p^*)} & 1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} \end{bmatrix}.$$

Therefore, we obtain the relations

$$Cov(\xi, h'(q)q) = \frac{\left(1 - \frac{Cov(p^*,g(q))}{Var(p^*)}\right) Cov(\xi^{OLS}, h'(q)q) + \frac{Cov(p^*,h'(q)q)}{Var(p^*)} Cov(\xi^{OLS}, g(q))}{1 + \frac{1}{\beta}\frac{Cov(p^*,h'(q)q)}{Var(p^*)} - \frac{Cov(p^*,g(q))}{Var(p^*)}}$$

$$Cov(\xi, g(q)) = \frac{-\frac{1}{\beta}\frac{Cov(p^*,g(q))}{Var(p^*)} Cov(\xi^{OLS}, h'(q)q) + \left(1 + \frac{1}{\beta}\frac{Cov(p^*,h'(q)q)}{Var(p^*)}\right) Cov(\xi^{OLS}, g(q))}{1 + \frac{1}{\beta}\frac{Cov(p^*,h'(q)q)}{Var(p^*)} - \frac{Cov(p^*,g(q))}{Var(p^*)}}.$$

In terms of observables, we can substitute in for $Cov(\xi, g(q)) - \frac{1}{\beta} Cov\left(\xi, h'(q)q\right)$ in the plim of the OLS estimator and simplify:

$$\begin{aligned}
&\left(1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} - \frac{Cov(p^*, g(q))}{Var(p^*)}\right)\left(Cov(\xi, g(q)) - \frac{1}{\beta} Cov\left(\xi, h'(q)q\right)\right) \\
=& -\frac{1}{\beta}\frac{Cov(p^*, g(q))}{Var(p^*)} Cov(\xi^{OLS}, h'(q)q) + \left(1 + \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)}\right) Cov(\xi^{OLS}, g(q)) \\
&- \frac{1}{\beta}\left(1 - \frac{Cov(p^*, g(q))}{Var(p^*)}\right) Cov(\xi^{OLS}, h'(q)q) - \frac{1}{\beta}\frac{Cov(p^*, h'(q)q)}{Var(p^*)} Cov(\xi^{OLS}, g(q)) \\
=& Cov(\xi^{OLS}, g(q)) - \frac{1}{\beta} Cov(\xi^{OLS}, h'(q)q).
\end{aligned}$$

Thus, we obtain an expression for the probability limit of the OLS estimator,

$$\text{plim}\hat{\beta}^{OLS} = \beta - \frac{\frac{Cov(\xi^{OLS},h'(q)q)}{Var(p^*)} - \beta\frac{Cov(\xi^{OLS},g(q))}{Var(p^*)}}{\beta + \frac{Cov(p^*,h'(q)q)}{Var(p^*)} - \beta\frac{Cov(p^*,g(q))}{Var(p^*)}},$$

and the following quadratic β.

$$\begin{aligned}
0 =& \left(1 - \frac{Cov(p^*, g(q))}{Var(p^*)}\right)\beta^2 \\
&+ \left(\frac{Cov(p^*, h'(q)q)}{Var(p^*)} - \hat{\beta}^{OLS} + \frac{Cov(p^*, g(q))}{Var(p^*)}\hat{\beta}^{OLS} + \frac{Cov(\xi^{OLS}, g(q))}{Var(p^*)}\right)\beta \\
&+ \left(-\frac{Cov(p^*, h'(q)q)}{Var(p^*)}\hat{\beta}^{OLS} - \frac{Cov(\xi^{OLS}, h'(q)q)}{Var(p^*)}\right).
\end{aligned}$$

QED.

B A Consistent and Unbiased Estimate for ξ

The following proof shows a consistent and unbiased estimate for the unobserved term in a linear regression when one of the covariates is endogenous. Though demonstrated in the context of semi-linear demand, the proof also applies for any endogenous covariate, including when (transformed) quantity depends on a known transformation of price, as no supply-side assumptions are required. For example, we may replace p with $\ln p$ everywhere and obtain the same results.

We can construct both the true demand shock and the OLS residuals as:

$$\begin{aligned} \xi &= h(q) - \beta p - x'\alpha \\ \xi^{OLS} &= h(q) - \beta^{OLS} p - x'\alpha^{OLS} \end{aligned}$$

where this holds even in small samples. Without loss of generality, we assume $E[\xi] = 0$. The true demand shock is given by $\xi_0 = \xi^{OLS} + (\beta^{OLS} - \beta)p + x'(\alpha^{OLS} - \alpha)$. We desire to show that an alternative estimate of the demand shock, $\xi_1 = \xi^{OLS} + (\beta^{OLS} - \beta)p^*$, is consistent and unbiased. (This eliminates the need to estimate the true α parameters). It suffices to show that $(\beta^{OLS} - \beta)p^* \to (\beta^{OLS} - \beta)p + x'(\alpha^{OLS} - \alpha)$.

Consider the projection matrices

$$\begin{aligned} Q &= I - P(P'P)^{-1}P' \\ M &= I - X(X'X)^{-1}X', \end{aligned}$$

where P is an $N \times 1$ matrix of prices and X is the $N \times k$ matrix of covariates x.

Denote $Y \equiv h(q) = P\beta + X\alpha + \xi$. Our OLS estimators can be constructed by a residualized regression

$$\begin{aligned} \alpha^{OLS} &= \left((XQ)'QX\right)^{-1}(XQ)'Y \\ \beta^{OLS} &= \left((PM)'MP\right)^{-1}(PM)'Y. \end{aligned}$$

Therefore

$$\begin{aligned} \alpha^{OLS} &= \left(X'QX\right)^{-1}\left(X'QP\beta + X'QX\alpha + X'Q\xi\right) \\ &= \alpha + \left(X'QX\right)^{-1}X'Q\xi. \end{aligned}$$

Similarly,

$$\begin{aligned}\beta^{OLS} =& (P'MP)^{-1}(P'MP\beta + P'MX\alpha + P'M\xi)\\ =& \beta + (P'MP)^{-1}P'M\xi.\end{aligned}$$

We desire to show

$$P^*(\beta^{OLS} - \beta) \to P(\beta^{OLS} - \beta) + X(\alpha^{OLS} - \alpha).$$

Note that $P^* = MP$. Then

$$\begin{aligned}P^*(\beta^{OLS} - \beta) \to& P(\beta^{OLS} - \beta) + X(\alpha^{OLS} - \alpha)\\ MP(P'MP)^{-1}P'M\xi \to& P(P'MP)^{-1}P'M\xi\\ &+ X(X'QX)^{-1}X'Q\xi\\ -X(X'X)^{-1}X'P(P'MP)^{-1}P'M\xi \to& X(X'QX)^{-1}X'Q\xi\\ -X(X'X)^{-1}X'P(P'MP)^{-1}P'\left[I - X(X'X)^{-1}X'\right]\xi \to& X(X'QX)^{-1}X'\left[I - P(P'P)^{-1}P'\right]\xi\\ -X(X'X)^{-1}X'P(P'MP)^{-1}P'\xi \to& X(X'QX)^{-1}X'\xi\\ +X(X'X)^{-1}X'P(P'MP)^{-1}P'X(X'X)^{-1}X'\xi \quad & -X(X'QX)^{-1}X'P(P'P)^{-1}P'\xi.\end{aligned}$$

We will show that the following two relations hold, which proves consistency and completes the proof.

1. $X(X'X)^{-1}X'P(P'MP)^{-1}P'\xi = X(X'QX)^{-1}X'P(P'P)^{-1}P'\xi$
2. $X(X'X)^{-1}X'P(P'MP)^{-1}P'X(XX(X'QX)^{-1}X'X)^{-1}X'\xi \to X(X'QX)^{-1}X'\xi$

Part 1: Equivalence

It suffices to show that $X(X'X)^{-1}X'P(P'MP)^{-1} = X(X'QX)^{-1}X'P(P'P)^{-1}$.

$$
\begin{aligned}
X(X'X)^{-1}X'P\left(P'MP\right)^{-1} =& X\left(X'QX\right)^{-1}X'P(P'P)^{-1} \\
X(X'X)^{-1}X'P =& X\left(X'QX\right)^{-1}X'P(P'P)^{-1}\left(P'MP\right) \\
X(X'X)^{-1}X'P =& X\left(X'QX\right)^{-1}X'P(P'P)^{-1}\left(P'P\right) \\
& - X\left(X'QX\right)^{-1}X'P(P'P)^{-1}\left(P'X(X'X)^{-1}X'P\right) \\
X(X'X)^{-1}X'P =& X\left(X'QX\right)^{-1}X'P \\
& - X\left(X'QX\right)^{-1}X'\left[I-Q\right]X(X'X)^{-1}X'P \\
X(X'X)^{-1}X'P =& X\left(X'QX\right)^{-1}X'P \\
& - X\left(X'QX\right)^{-1}X'X(X'X)^{-1}X'P \\
& + X\left(X'QX\right)^{-1}X'QX(X'X)^{-1}X'P \\
X(X'X)^{-1}X'P =& X(X'X)^{-1}X'P
\end{aligned}
$$

QED.

Part 2: Consistency (and Unbiasedness)

Since $X(X'X)^{-1}X'P = X\left(X'QX\right)^{-1}X'P(P'P)^{-1}\left(P'MP\right)$, as shown above:

$$
\begin{aligned}
X(X'X)^{-1}X'P\left(P'MP\right)^{-1}P'X(X'X)^{-1}X'\xi &\to X\left(X'QX\right)^{-1}X'\xi \\
X\left(X'QX\right)^{-1}X'P(P'P)^{-1}P'X(X'X)^{-1}X'\xi &\to X\left(X'QX\right)^{-1}X'\xi \\
X\left(X'QX\right)^{-1}X'\left[I-Q\right]X(X'X)^{-1}X'\xi &\to X\left(X'QX\right)^{-1}X'\xi \\
X\left(X'QX\right)^{-1}X'X(X'X)^{-1}X'\xi &\to X\left(X'QX\right)^{-1}X'\xi \\
-X(X'X)^{-1}X'\xi & \\
X\left(X'QX\right)^{-1}X'\xi - X(X'X)^{-1}X'\xi &\to X\left(X'QX\right)^{-1}X'\xi \\
X(X'X)^{-1}X'\xi &\to 0.
\end{aligned}
$$

The last line, where the projection of ξ onto the exogenous covariates X converges to zero, holds by assumption. We say that two vectors converge if the mean absolute deviation goes to zero as the sample size gets large. QED.

Note that also $E[X(X'X)^{-1}X'\xi] = 0$, so that ξ_1 is both a consistent and unbiased estimate for ξ_0.