1 Why demand analysis/estimation?

- Estimation of demand functions is an important empirical endeavor. Why?
- Fundamental empircial question: how much market power do firms have?
  - Market power: ability to raise prices profitably. (What market power do price-taking firms have?)
  - Market power measured by markup: $\frac{p-mc}{p}$.
  - Problem: mc not observed!
  - Motivates empirical methodology in IO.
  - For example, you observe high prices in an industry. Is this due to market power, or due to high costs? Cannot answer this question directly, because we don't observe costs.
- *Indirect approach*: obtain estimate of firms' markups by estimating firms' demand functions.
- Intuition is most easily seen in monopoly example:
  - $\max_p pq(p) - C(q(p))$, where $q(p)$ is demand curve.
  - FOC: $q(p) + pq'(p) = C'(q(p))q'(p)$
  - At optimal price p^*, **Inverse Elasticity Property** holds:

$$(p^* - MC(q(p^*))) = -\frac{q(p^*)}{q'(p^*)}$$

    or

$$\frac{p^* - mc\,(q(p^*))}{p^*} = -\frac{1}{\epsilon(p^*)},$$

    where $\epsilon(p^*)$ is $q'(p^*)\dfrac{p^*}{q(p^*)}$, the price elasticity of demand.
  - Hence, if we can estimate $\epsilon(p^*)$, we can infer what the markup $\frac{p^*-mc(q(p^*))}{p^*}$ is, even when we don't observe the marginal cost $mc\,(q(p^*))$.

- Caveat: validity of exercise depends crucially on using the right supply-side model (in this case: monopoly without entry possibility).

  If costs were observed: markup could be estimated directly, and we could test for vaalidity of monopoly pricing model (ie. test whether markup= $\frac{-1}{\epsilon}$).

- Start by reviewing some econometrics. (No attempt to be exhaustive.)

2 Primer: Least-squares estimation

- Observe data points $\{y_i, x_i\}$ for $i = 1, \ldots n$. What is *linear* relationship between y and x?

- Graph. What linear function of x – that is, $\alpha + \beta x$ – fits y the best?

- Ordinary least squares (OLS) regression:

$$\min_{\alpha,\beta} \sum_i [y_i - \alpha - \beta x_i]^2.$$

- In multivariate case: X_i and $\vec{\beta}$ are both $K - dimensionalvectors$. Then

$$\min_{\alpha,\vec{\beta}} \sum_i [y_i - \alpha - X_i' \vec{\beta}]^2.$$

To analyze properties of OLS regression, consider a closely-related statistical problem of **Best Linear Prediction**,

Consider two random variables X and Y. What is the “best” predictor of Y, among all the possible linear functions of X?

“Best” linear predictor minimizes the mean squared error of prediction:

$$\min_{\alpha,\beta} E(Y - \alpha - \beta X)^2. \tag{1}$$

(Recall: expectation is linear operator, so that $E(A+B) = EA + EB$)

The first-order conditions are:

$$\text{For } \alpha\text{: } 2\alpha - 2EY + 2\beta EX = 0$$
$$\text{For } \beta\text{: } 2\beta EX^2 - 2EXY + 2\alpha EX = 0.$$

Solving:

$$\begin{aligned} \beta^* &= \frac{Cov(X,Y)}{VX} \\ \alpha^* &= EY - \beta^* EX \end{aligned} \tag{2}$$

where

$$Cov(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - EX \cdot EY$$

and

$$VX = E[(X-EX)^2] = E(X^2) - (EX)^2.$$

Additional implications of b.l.p.: Let $\hat{Y} \equiv \alpha^* + \beta^* X$ denote a "fitted value" of Y, and $U \equiv Y - \hat{Y}$ denote the "residual" or prediction error:

- $EU = 0$
- $V\hat{Y} = (\beta^*)^2 VX = (Cov(X,Y))^2/VX = \rho^2_{XY} VY$
- $VU = VY + (\beta^*)^2 VX - 2\beta^* Cov(X,Y) = VY - (Cov(X,Y))^2/VX = (1 - \rho^2_{XY})VY$

Hence, the b.l.p. accounts for a ρ^2_{XY} proportion of the variance in Y; in this sense, the correlation measures the linear relationship between Y and X.

Also note that

$$\begin{aligned} Cov(\hat{Y}, U) &= Cov(\hat{Y}, Y - \hat{Y}) \\ &= E[(\hat{Y} - E\hat{Y})(Y - \hat{Y} - EY + E\hat{Y})] \\ &= E[(\hat{Y} - E\hat{Y})(Y - EY) - (\hat{Y} - E\hat{Y})(\hat{Y} - E\hat{Y})] \\ &= Cov(\hat{Y}, Y) - V\hat{Y} \\ &= E[(\alpha^* + \beta^* X - \alpha^* - \beta^* EX)(Y - EY)] - V\hat{Y} \qquad (3) \\ &= \beta^* E[(X - EX)(Y - EY)] - V\hat{Y} \\ &= \beta^* Cov(X, Y) - V\hat{Y} \\ &= Cov^2(X, Y)/VX - Cov^2(X, Y)/VX \\ &= 0. \end{aligned}$$

Hence, for any random variable X, the random variable Y can be written as the sum of a part which is a linear function of X, and a part which is uncorrelated with X.

Also,

$$Cov(X, U) = 0. \qquad (4)$$

Note: in practice, with a finite sample of Y, X, the minimization problem (1) is infeasible. In practice, we minimize the sample counterpart

$$\min_{\alpha,\beta} \sum_i (Y_i - \alpha - \beta X_i)^2 \qquad (5)$$

which is the objective function in ordinary least squares regression. The OLS values for α and β are the finite-sample versions of Eq. (2).

(In "sample" version, expectations are replaced by sample averages. eg. mean Ex is replaced by sample average from n observations $\bar{X}_n \equiv \frac{1}{n}\sum_i X_i$. Law of large numbers say this approximation should not be bad, especially for large n.)

■■■

Next we can see some intuition of least-squares regression. Assume that the "true" model describing the generation of the Y process is:

$$Y = \alpha + \beta X + \epsilon, \quad E\epsilon = 0. \tag{6}$$

What we mean by true model is that this is a causal model in the sense that a one-unit increase in X would raise Y by β units. (In the previous section, we just assume that Y, X move jointly together, so there is no sense in which changes in X "cause" changes in Y.)

Question: under what assumptions does doing least-squares on Y, X (as in Eqs. (1) or (5) above) recover the true model; ie. $\alpha^* = \alpha$, and $\beta^* = \beta$?

- For α^*:

$$\begin{aligned} \alpha^* &= EY - \beta^* EX \\ &= \alpha + \beta EX + E\epsilon - \beta^* EX \end{aligned}$$

  which is equal to α if $\beta = \beta^*$.

- For β^*:

$$\begin{aligned} \beta^* &= \frac{Cov(\alpha + \beta X + \epsilon, X)}{VarX} \\ &= \frac{1}{VarX} \cdot \{E[X(\alpha + \beta X + \epsilon)] - EX \cdot E[\alpha + \beta X + \epsilon]\} \\ &= \frac{1}{VarX} \cdot \left\{\alpha EX + \beta EX^2 + E[\epsilon X] - \alpha EX - \beta [EX]^2 - EXE\epsilon\right\} \\ &= \frac{1}{VarX} \cdot \left\{\beta[EX^2 - (EX)^2] + E[\epsilon X]\right\} \end{aligned}$$

which is equal to β if

$$E[\epsilon X] = 0. \tag{7}$$

This is an "exogeneity" assumption, that (roughly) X and the disturbance term ϵ are uncorrelated. Under this assumption, the best linear predictors from the infeasible problem (1)) coincide with the true values of α, β. Correspondingly, it turns out that the feasible finite-sample least-squares estimates from (5) are "good" (in some sense) estimators for α, β.

Note that the orthogonality condition (7) differs from the zero covariance property (4), which is a feature of the b.l.p.

When there is more than one X variable, then we use *multivariate regression*. In matrix notation, true model is:

$$Y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \epsilon_{n\times 1}.$$

The least-squares estimator for β is

$$\beta^{OLS} = (X'X)^{-1}X'Y.$$

Next we consider estimating demand functions, where exogeneity is usually violated.

3 Demand estimation

Linear demand-supply model:

$$\begin{aligned} \text{Demand:} \quad & q_t^d = \gamma_1 p_t + \mathbf{x'_{t1}}\beta_{\mathbf{1}} + u_{t1} \\ \text{Supply:} \quad & p_t = \gamma_2 q_t^s + \mathbf{x'_{t2}}\beta_{\mathbf{2}} + u_{t2} \\ \text{Equilibrium:} \quad & q_t^d = q_t^s \end{aligned}$$

Demand function summarizes consumer preferences; supply function summarizes firms' cost structure

First, focus on estimating demand function:

$$\text{Demand: } q_t = \gamma_1 p_t + \mathbf{x'_{t1}}\beta_{\mathbf{1}} + u_{t1}$$

If u_1 correlated with u_2, then p_t is endogenous in demand function: cannot estimate using OLS. Graph. Several estimation approaches.

1. Instrumental variable (IV) methods:

   - Assume there are instruments Z which satisfy certain properties
     (a) Uncorrelated with error term in demand equation: $E(u_1 Z) = 0$. **Exclusion** restriction. ("order condition")
     (b) Correlated with endogenous variable: $Cov(Z, p) \neq 0$. ("rank condition")
   - The x's are exogenous variables which can serve as instruments:
     (a) x_{t2} are *cost shifters*; affect production costs. Correlated with p_t but not with u_{t1}: use as instruments in demand function.
     (b) x_{t1} are *demand shifters*; affect willingness-to-pay, but not a firm's production costs. Correlated with q_t but not with u_{2t}: use as instruments in supply function.
   - Two-stage least squares:

     $$\beta^{2sls} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$$

     where $\hat{X} \equiv Z'(Z'Z)^{-1}(Z'X)$ are the predicted values of X from a least-squares regression of X on Z.

2. Maximum likelihood (more technical):

   - Likelihood function of the data is joint density of the endogenous variables $(\mathbf{q_t}, \mathbf{p_t})$ conditional on exogenous variables $(\mathbf{x_{t1}}, \mathbf{x_{t2}})$.
   - First, need to express endogenous variables in terms of exogenous variables:

     $$\begin{aligned} \text{Demand:} \quad & q_t = \gamma_1 p_t + \mathbf{x'_{t1}}\beta_1 + u_{t1} \\ \text{Supply:} \quad & p_t = \gamma_2 q_t + \mathbf{x'_{t2}}\beta_2 + u_{t2} \end{aligned}$$

     $$\Rightarrow \begin{pmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} x_{t1} \\ x_{t2} \end{pmatrix} + \begin{pmatrix} u_{t1} \\ u_{t2} \end{pmatrix}$$

     $$\Leftrightarrow \Gamma Y = BX + U \Leftrightarrow Y = \Gamma^{-1}BX + \Gamma^{-1}U$$

     This is called the "reduced form" representation of the demand-suppoly system.

- Assume that the unobservables $\{(u_{t1}, u_{t2})_{t=1}^{T}\}$ are distributed according to a density function $g(\cdots)$.

  Example: $(u_{t1}, u_{t2}) \sim$ i.i.d $N(0, \Sigma)$

  Then joint density of $\vec{u}$ is:

$$g(\vec{u}) = (2\pi)^{-1}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\vec{u})'\Sigma^{-1}(\vec{u})\right].$$

- Recall change of variables formula: if $Y = X/a$ and X has density function $g(X)$, then Y has density function

$$f(Y) = g(aY) \cdot a.$$

  Applying the multivariate version of this, we get

$$f(Y) = g(\Gamma Y - BX) * \mid \Gamma \mid \tag{8}$$

  Assuming you have T observations of (Y_t, X_t), then likelihood function is

$$L(Y|X) = \prod_{t=1}^{T} f(Y_t) = \mid \Gamma \mid^T \prod_{t=1}^{T} g(Y\Gamma - XB).$$

  Log-likelihood function is (ignoring the constant):

$$\log L(Y \mid X) \sim T \log \mid \Gamma \mid -\frac{T}{2} \log \mid \Sigma \mid - \sum_t \frac{1}{2} (\Gamma Y_t - BX_t)' \, \Sigma^{-1} \ (\Gamma Y_t - BX_t)$$

- Maximize this with respect to Γ, B, Σ to obtain maximum likelihood estimator.