# Demand and Supply

Malgorzata Knauff

spring 2019/2020

# Outline:

# Market data

- Typically, in economics we do not have access to experimental data
- Moreover, our data is often not at the individual level
- Rather, the data we use is observational data
- Moreover, the unit of measurement is usually at the market level
- Suppose we collected data using the following method
  1. We randomly sampled $N$ markets (they can differ in geographic or time dimension)
  2. For each market $i$, we observe the price $P_i$ and quantity $Q_i$ of good $Q$ traded in that market, as well as the prevailing market income $Y_i$
  3. Data collection will result in the following sample of data $\{Q_i, P_i, Y_i : i = 1, 2, ...N\}$

# Problem with endogeneity

- With observational data at the market level, price is not an exogenous variable in the sample of data
- Instead, price is **endogenously** determined (determined within the system) in the market by the interaction of demand and supply
- The prices and quantities in our data set are not from a random sample
- Rather, they are equilibrium data
- To solve this **endogeneity issue**, we will need to take supply into consideration when we estimate demand
- Otherwise, our estimates of the demand parameters will be biased

# Modeling supply – economic theory

- Suppose that the industry for good $Q$ is perfectly competitive
- There are many firms, each with marginal costs given by $MC(Q) = \beta_0 + \beta_1 Q$
- The goal of each firm in the industry is to maximize profits

$$max\{(P - MC(Q))Q\}$$

  and so will set price equal to marginal cost
- So, the supply function is

$$P = \beta_0 + \beta_1 Q$$

# Modeling supply – econometrics

- The supply function from our economic theory describes average firm behavior – it is best thought of as model of expected supply
$$E[P_i] = \beta_0 + \beta_1 Q_i$$
- To translate this into an econometric model we introduce an IID *supply shock* $\varepsilon_i^S$, which captures how firm behavior may differ from the average
- Thus, the econometric model of market supply is

$$P_i = \beta_0 + \beta_1 Q_i + \varepsilon_i^S, \varepsilon_i^S \sim IID(0, \sigma^2)$$

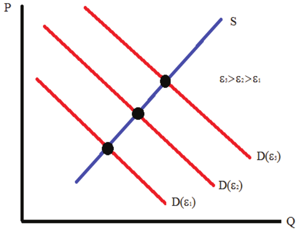
# Ignoring supply?

- What happens if we ignore supply?
- Suppose we plan to estimate demand using our observational data $\{Q_i, P_i, Y_i : i = 1, 2, ...N\}$ and the following econometric model of market demand

$$Q_i = \alpha_0 + \alpha_1 P_i + \alpha_2 Y_i + \varepsilon_i^D, \varepsilon_i^D \sim IID(0, \sigma^2)$$

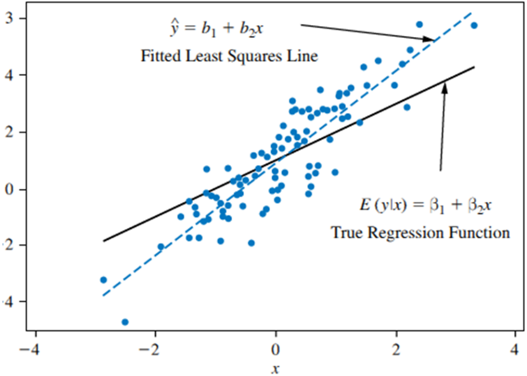
- Is the parameter vector $\alpha$ identified by OLS? The answer is no!
- This is because, with observational data, we have $cov[P_i \varepsilon_i^D] \neq 0$

# Problem: $cov[P_i \varepsilon_i^D] > 0$




- Large $\varepsilon_i^D$ will be associated with high $Q_i$ (demand)
- And, large $\varepsilon_i^D$ will also be associated with high $P_i$ (equilibrium)

- The combined effect is an above average $Q_i$
- However, since $\varepsilon_i^D$ is unobservable, the model wrongly attributes both of these effects all to $P_i$
- As a result, the parameter estimate for the effect of $P_i$ is too big
- That is, $\alpha_1$ will not be "negative enough", and maybe even "positive"




- The statistical consequences of correlation between $x$ and $\varepsilon$ is that the least squares estimator is biased and this bias will not disappear no matter how large the sample
- Consequently the least squares estimator is inconsistent when there is correlation between $x$ and $\varepsilon$

# Simultaneous equations models

- We will consider econometric models for data that are jointly determined by two or more economic relations
  - These simultaneous equations models differ from those previously studied because in each model there are two or more dependent variables rather than just one
  - Simultaneous equations models also differ from most of the econometric models we have considered so far, because they consist of a set of equations
- A very simple supply and demand model might look like:

$$Q = \alpha_1 P + \alpha_2 Y + \varepsilon_D \tag{1}$$

$$Q = \beta_1 P + \varepsilon_S \tag{2}$$

# Interdependence of demand and supply

- It takes two equations to describe the supply and demand equilibrium
  - The two equilibrium values, for price and quantity, $P^*$ and $Q^*$, respectively, are determined at the same time
  - In this model the variables $P$ and $Q$ are called endogenous variables because their values are determined within the system we have created
  - The endogenous variables $P$ and $Q$ are dependent variables and both are random variables
  - The income variable $Y$ has a value that is determined outside this system – such variables are said to be **exogenous**, and these variables are treated like usual explanatory variables

# Endogeneity problem

- The fact that $P$ is an endogenous variable on the right-hand side of the supply and demand equations means that we have an explanatory variable that is random
- This is contrary to the usual assumption of *fixed explanatory variables*
- The problem is that the endogenous regressor $P$ is correlated with the random errors, $\varepsilon_d$ and $\varepsilon_s$, which has a devastating impact on our usual least squares estimation procedure, making the least squares estimator biased and inconsistent

- The two structural equations Eqs. 1 and 2 can be solved to express the endogenous variables $P$ and $Q$ as functions of the exogenous variable $Y$
- This reformulation of the model is called the reduced form of the structural equation system

- To solve for $P$, set $Q$ in the demand and supply equations to be equal:

$$\beta_1 P + \varepsilon_S = \alpha_1 P + \alpha_2 Y + \varepsilon_D$$

- Solve for $P$:

$$P = \frac{\alpha_2}{\beta_1 - \alpha_1} Y + \frac{\varepsilon_D - \varepsilon_S}{\beta_1 - \alpha_1} = \pi_1 Y + \nu_1 \tag{3}$$

- Solving for Q:

$$
\begin{aligned}
Q = \beta_1 P + \varepsilon_S &= \beta_1 \left[ \frac{\alpha_2}{\beta_1 - \alpha_1} Y + \frac{\varepsilon_D - \varepsilon_S}{\beta_1 - \alpha_1} \right] + \varepsilon_S \\
&= \frac{\beta_1 \alpha_2}{\beta_1 - \alpha_1} Y + \frac{\beta_1 \varepsilon_D - \alpha_1 \varepsilon_S}{\beta_1 - \alpha_1} \\
&= \pi_2 Y + \nu_2
\end{aligned}
\tag{4}
$$

  - The parameters $\pi_1$ and $\pi_2$ in Eqs. 3 and 4 are called reduced-form parameters. The error terms $\nu_1$ and $\nu_2$ are called reduced-form errors

# $\pi$'s versus $\alpha$'s and $\beta$'s

- We can consistently estimate parameters of Eqs. 4 and 3, but we are interested in parameters of Eqs. 1 and 2
  - The parameters of the demand equation, $\alpha_1$ and $\alpha_2$, cannot be consistently estimated by any estimation method
  - The slope of the supply equation, $\beta_1$, can be consistently estimated
- It is the absence of variables in one equation that are present in another equation that makes parameter estimation possible

# Necessary Condition For Identification

| |
|---|
| In a system of $M$ simultaneous equations, which jointly determine the values of $M$ endogenous variables, at least $M - 1$ variables must be absent from an equation for estimation of its parameters to be possible |
| When estimation of an equation's parameters is possible, then the equation is said to be identified, and its parameters can be estimated consistently. If fewer than $M - 1$ variables are omitted from an equation, then it is said to be unidentified, and its parameters cannot be consistently estimated |

- The identification condition must be checked before trying to estimate an equation
- If an equation is not identified, then changing the model must be considered before it is estimated

# 2SLS

- The most widely used method for estimating the parameters of an identified structural equation is called two-stage least squares
- This is often abbreviated as 2SLS (or TSLS)
- The name comes from the fact that it can be calculated using two least squares regressions

# Stage 1

- Consider the supply equation discussed previously. We cannot apply the usual least squares procedure to estimate $\beta_1$ in this equation because the endogenous variable $P$ on the right-hand side of the equation is correlated with the error term $\varepsilon_S$
- The reduced-form model is:

$$P = E[P] + \nu_1 = \pi_1 Y + \nu_1 \tag{5}$$

- Suppose we know $\pi_1$.Then through substitution:

$$Q = \beta_1[E[P] + \nu_1] + \varepsilon_S = \beta_1 E[P] + (\beta_1 \nu_1 + \varepsilon_S) \tag{6}$$

# Stage 2

- We can estimate $\pi_1$ using $\hat{\pi_1}$ from the reduced-form equation for $P$
- A consistent estimator for $E[P]$ is:

$$\hat{P} = \hat{\pi_1} Y$$

- Then

$$Q = \beta_1 \hat{P} + \varepsilon_* \tag{7}$$

- Estimating Eq. 7 by least squares generates the so-called two-stage least squares estimator of $\beta_1$, which is consistent and normally distributed in large samples

# Summary of 2SLS

- The two stages of the estimation procedure are:
  - Least squares estimation of the reduced-form equation for $P$ and the calculation of its predicted value
  - Least squares estimation of the structural equation in which the right-hand-side endogenous variable $P$ is replaced by its predicted value

# Properties of 2SLS estimators

- The 2SLS estimator is a biased estimator, but it is consistent
- In large samples the 2SLS estimator is approximately normally distributed
- The variances and covariances of the 2SLS estimator are unknown in small samples, but for large samples we have expressions for them that we can use as approximations
- If you obtain 2SLS estimates by applying two least squares regressions using ordinary least squares regression software, the standard errors and t-values reported in the second regression are not correct for the 2SLS estimator