# Econometrics 1
# Lecture 12: Simultaneous Equation Models
ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

# Introduction

This lecture studies the estimation of simultaneous equation models. We only study two-equation models.

The two-equation models we consider are of the form:

$$y_1 = \beta_{10} + \alpha_1 y_2 + \beta_{11} z_{11} + ... + \beta_{1k_1} z_{1k_1} + u_1 \quad (1)$$

$$y_2 = \beta_{20} + \alpha_2 y_1 + \beta_{21} z_{21} + ... + \beta_{2k_2} z_{2k_2} + u_2 \quad (2)$$

where $y_1$ and $y_2$ are the endogenous variables, $u_1$ and $u_2$ the error terms, $z_{11}$,..., $z_{1k_1}$ the $k_1$ exogenous regressors of the first equation, $z_{21}$,..., $z_{2k_2}$ the $k_2$ exogenous regressors of the second equation (here exogeneity means that all $z$s are uncorrelated with $u_1$ and $u_2$), $\beta_{10}$ and $\beta_{20}$ the intercepts, $\alpha_1$ the causal effect of $y_2$ on $y_1$, $\alpha_2$ the causal effect of $y_1$ on $y_2$, and all other $\beta$s measure the causal effects of the associated exogenous regressors.

These two equations constitute a simultaneous equations model (SEM).

# Introduction (cntd)

The particularity of this model is that the endogenous variables $y_1$ and $y_2$ are simultaneously determined.

There are many practical examples of SEMs:

- $y_1$ the murder rate in a city and $y_2$ the size of the city police force
- $y_1$ the regional crime rate and $y_2$ the number of prisoners in the region
- $y_1$ the labor supply (number of hours worked) of married women and $y_2$ their hourly wage rate.
- $y_1$ the demand for a product and $y_2$ the supply.

When the endogenous variables are jointly determined, $y_2$ is generally correlated with $u_1$ in model (1) because $y_2$ is itself a function of $y_1$. Similarly, $y_1$ is generally correlated with $u_2$ in model (2) for an analogous reason.

OLS estimation of models (1) or (2) therefore lead to biased estimators.

# Demand and supply: bias of OLS estimator

To see this, consider the example of a supply-and-demand model.

The following demand equation specifies the quantity of milk that consumers in a region wish to buy given the price:

$$Q = \alpha_d P + u_d$$

The supply equation specifies the quantity of milk the farmers wish to produce given the price:

$$P = \alpha_s Q + u_s$$

For simplicity, this SEM has no constants nor any exogenous variable. The aim is to estimate the parameters $\alpha_d$ and $\alpha_s$ using a sample of $n$ regions.

## Demand and supply: bias of OLS estimator (cntd)

It is important to note that for each region we only observe the equilibrium quantity and equilibrium price (quantity and price such that demand equals supply).

The equilibrium quantity satisfies

$$\begin{aligned} Q &= \alpha_d P + u_d = \alpha_d(\alpha_s Q + u_s) + u_d \\ (1 - \alpha_s\alpha_d)Q &= \alpha_d u_s + u_d \\ Q &= (\alpha_d u_s + u_d)/(1 - \alpha_s\alpha_d) \end{aligned}$$

It also follows that the equilibrium price equals

$$\begin{aligned} P &= \alpha_s Q + u_s = \alpha_s(\alpha_d u_s + u_d)/(1 - \alpha_s\alpha_d) + u_s \\ &= (\alpha_s u_d + u_s)/(1 - \alpha_s\alpha_d) \end{aligned}$$

# Demand and supply: bias of OLS estimator (cntd)

So when we regress equilibrium quantities on equilibrium prices we regress

$$Q = (\alpha_d u_s + u_d)/(1 - \alpha_s \alpha_d)$$

on

$$P = (\alpha_s u_d + u_s)/(1 - \alpha_s \alpha_d)$$

The OLS estimator of the slope parameter is

$$\hat{\alpha} = \frac{\sum_{i=1}^n P_i Q_i}{\sum_{i=1}^n P_i^2} \xrightarrow{p} \frac{E(P_i Q_i)}{E(P_i^2)} = \frac{\alpha_d \sigma_s^2 + \alpha_s \sigma_d^2 + (1 + \alpha_d \alpha_s)\sigma_{sd}}{\sigma_s^2 + \alpha_s^2 \sigma_d^2 + 2\alpha_s \sigma_{sd}}$$

where $\sigma_s^2 = E(u_s^2)$, $\sigma_d^2 = E(u_d^2)$, and $\sigma_{sd} = E(u_s u_d)$.

So the OLS estimator is not a consistent estimator of $\alpha_s$ or $\alpha_d$. This illustrates the simultaneity bias of the OLS estimator.

Using only the data on equilibrium prices we cannot estimate the demand or supply equation: in fact, we have no way to tell them apart!

# Demand and supply: supply shifter

Suppose now that we have an exogenous variable $z_1$ that affects the supply function (2) but not the demand equation (1). $z_1$ can be for instance the price of cattle feed.

The SEM becomes

$$Q = \alpha_d P + u_d \tag{3}$$
$$P = \alpha_s Q + \beta_s z_1 + u_s \tag{4}$$

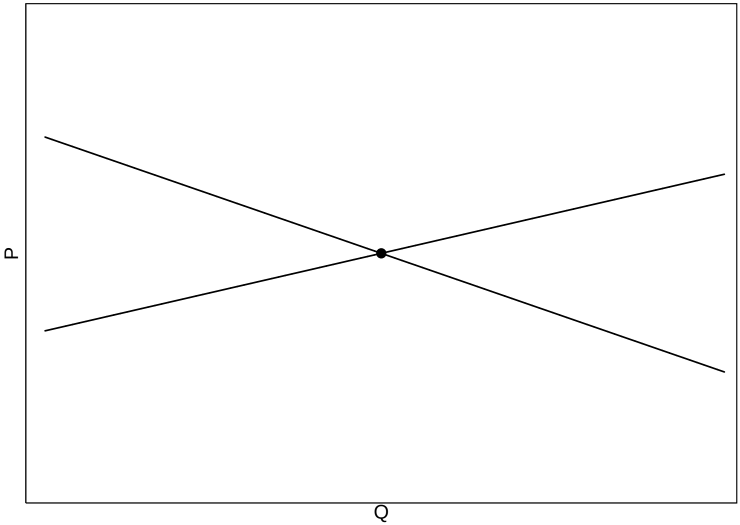
It turns out that $z_1$ allows us to identify the parameter $\alpha_d$ in the demand equation (but not the slope parameter in the supply equation nor the coefficient on $z_1$).

Intuition: the observed variable $z_1$ shifts the supply equation up and own, but does not affect the demand equation.

Given different values of $z_1$, we can trace out the demand equation, as illustrated in the next graphs.

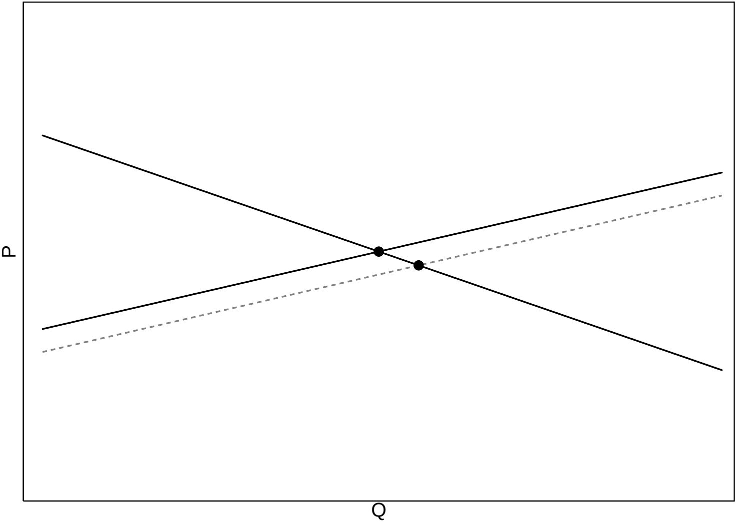
# Demand and supply

## Equilibrium



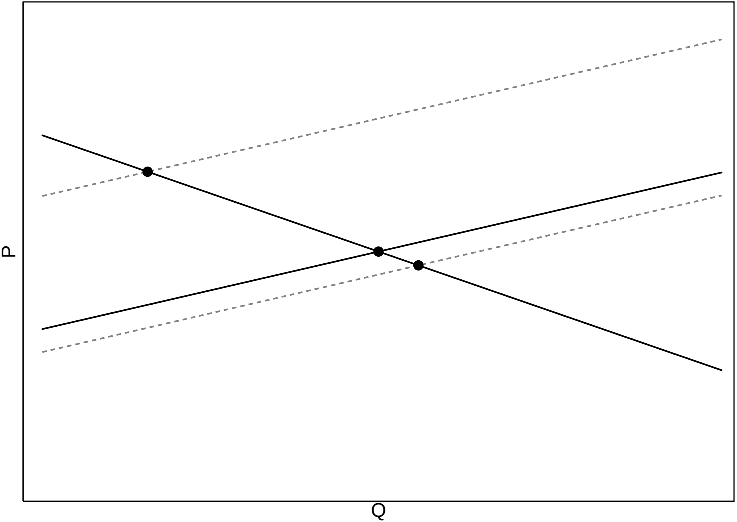
# Demand and supply

## Another equilibrium after shifting supply



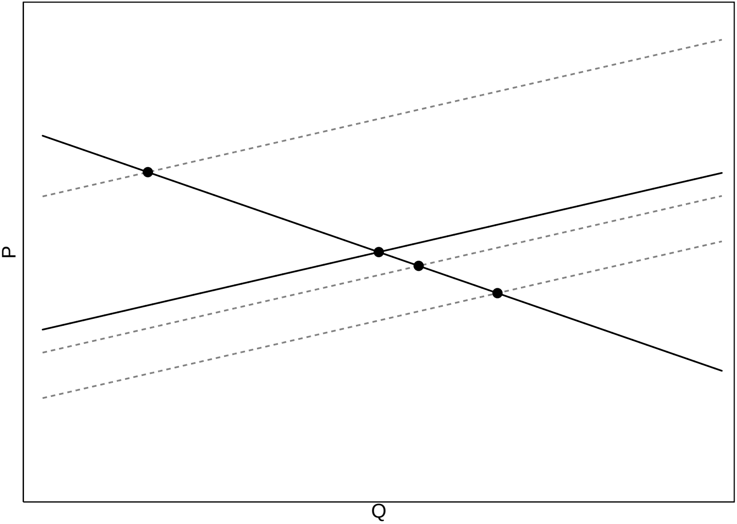
# Demand and supply

## Yet another equilibrium



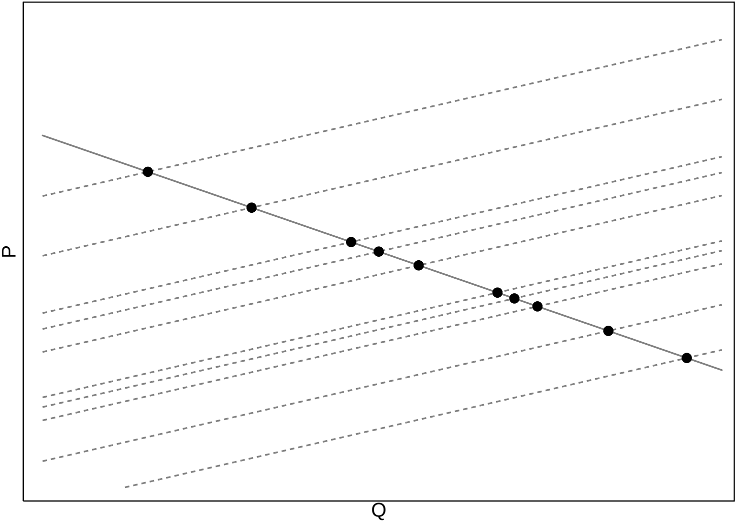
# Demand and supply

## And another one ...



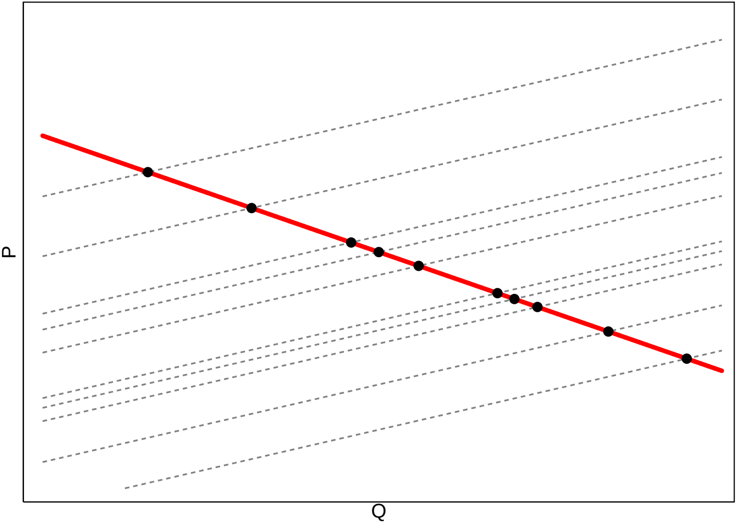
# Demand and supply

....



# Demand and supply

## The data now identify the demand equation!



# Demand and supply: estimator with supply shift

To see that the demand parameters are indeed identified and can be estimated consistently, substitute the demand curve (3) into the supply curve (4) and we get:

$$\begin{aligned} P &= \alpha_s Q + \beta_s z_1 + u_s = \alpha_s(\alpha_d P + u_d) + \beta_s z_1 + u_s \\ (1 - \alpha_d \alpha_s)P &= \beta_s z_1 + \alpha_s u_d + u_s \end{aligned}$$

which simplifies to the so-called reduced form supply equation (whereas (4) is the structural supply equation):

$$P = \pi_s z_1 + \tilde{u}_s$$

with the

- reduced form parameter $\pi_s = \beta_s/(1 - \alpha_d \alpha_s)$, and
- the reduced form error $\tilde{u}_s = (\alpha_s u_d + u_s)/(1 - \alpha_d \alpha_s)$

The reduced form can be consistently estimated by OLS if $z_1$ is exogenous (then $z_1$ is not correlated with $u_d$ and $u_s$).

# Demand and supply: estimator with supply shift (cntd)

We can also substitute our supply equation into our demand equation and get:

$$\begin{aligned} Q &= \alpha_d P + u_d = \alpha_d(\alpha_s Q + \beta_s z_1 + u_s) + u_d \\ (1 - \alpha_d \alpha_s)Q &= \alpha_d \beta_s z_1 + \alpha_d u_s + u_d \end{aligned}$$

which simplifies to the reduced form demand equation

$$Q = \pi_d z_1 + \tilde{u}_d$$

with the

- reduced form parameter $\pi_d = \alpha_d \beta_s / (1 - \alpha_d \alpha_s)$, and
- the reduced form error $\tilde{u}_d = (\alpha_d u_s + u_d)/(1 - \alpha_d \alpha_s)$

This reduced form can also be consistently estimated by OLS if $z_1$ is exogenous.

# Demand and supply: estimator with supply shift (cntd)

We have just shown that we can consistently estimate the reduced form parameters

- $\pi_s = \beta_s/(1-\alpha_d\alpha_s)$, and
- $\pi_d = \alpha_d\beta_s/(1-\alpha_d\alpha_s)$

But this means that we can also consistently estimate $\alpha_d$:

$$\hat{\alpha}_d \equiv \frac{\hat{\pi}_d}{\hat{\pi}_s} \xrightarrow{p} \frac{\alpha_d\beta_s/(1-\alpha_d\alpha_s)}{\beta_s/(1-\alpha_d\alpha_s)} = \alpha_d$$

This procedure to estimate the demand parameter $\alpha_d$ is in fact an application of 2SLS to our structural demand equation

$$Q = \alpha_d P + u_d$$

where the endogenous variable $P$ is instrumented by $z_1$. Indeed, in the first stage we estimate the reduced form supply equation

$$P = \pi_s z_1 + \tilde{u}_s$$

by OLS and get the estimator $\hat{\pi}_s$.

# Demand and supply: estimator with supply shift (cntd)

In the second stage we estimate the model

$$Q = \alpha_d \hat{\pi}_s z_1 + u_d + (\alpha_d P - \alpha_d \hat{\pi}_s z_1) = \alpha_d \hat{\pi}_s z_1 + error$$

by OLS, and get the 2SLS estimator $\hat{\alpha}_d$.

We have

$$\begin{aligned} \hat{\alpha}_d &= \frac{\sum_{i=1}^n \hat{\pi}_s z_{i1} Q_i}{\sum_{i=1}^n (\hat{\pi}_s z_{i1})^2} = \frac{1}{\hat{\pi}_s} \frac{\sum_{i=1}^n z_{i1} Q_i}{\sum_{i=1}^n z_{i1}^2} \\ &= \frac{\hat{\pi}_d}{\hat{\pi}_s} \end{aligned}$$

# Demand and supply: estimator with supply/demand shifts

The slope of the demand curve was identified thanks to an exogenous supply shifter ($z_1$)

We need an exogenous demand shifter $z_2$ to identify the parameters in the supply curve. $z_2$ can for example be the average income of consumers in the region.

The SEM now becomes:

$$\begin{aligned} Q &= \alpha_d P + \beta_d z_2 + u_d && (5) \\ P &= \alpha_s Q + \beta_s z_1 + u_s && (6) \end{aligned}$$

Our reduced form equations now become

$$Q = \underbrace{\frac{\alpha_d \beta_s}{1 - \alpha_s \alpha_d}}_{\pi_{d1}} z_1 + \underbrace{\frac{\beta_d}{1 - \alpha_s \alpha_d}}_{\pi_{d2}} z_2 + \tilde{u}_d \qquad (7)$$

## Demand and supply: estimator with supply/demand shifts (cntd)

and

$$P = \underbrace{\frac{\beta_s}{1-\alpha_s\alpha_d}}_{\pi_{s1}} z_1 + \underbrace{\frac{\alpha_s\beta_d}{1-\alpha_s\alpha_d}}_{\pi_{s2}} z_2 + \tilde{u}_s \tag{8}$$

The error terms $\tilde{u}_d$ and $\tilde{u}_s$ (defined as above) are not correlated with $z_1$ and $z_2$ and hence the OLS estimators $\hat{\pi}_{d1}$, $\hat{\pi}_{d2}$, $\hat{\pi}_{s1}$, and $\hat{\pi}_{s2}$ consistently estimate the reduced form parameters $\pi_{d1}$, $\pi_{d2}$, $\pi_{s1}$, and $\pi_{s2}$.

So $\alpha_d = \pi_{d1}/\pi_{s1}$ and $\alpha_s = \pi_{s2}/\pi_{d2}$ can be consistenly estimated by $\hat{\alpha}_d = \hat{\pi}_{d1}/\hat{\pi}_{s1}$ and $\hat{\alpha}_s = \hat{\pi}_{s2}/\hat{\pi}_{d2}$. (The estimators of the reduced form parameters also allow us to consistently estimate $\beta_d$ and $\beta_s$, the other two structural parameters).

The estimators $\hat{\alpha}_d$ and $\hat{\alpha}_s$ can again be interpreted as 2SLS estimators.

# Demand and supply: estimator with supply/demand shifts (cntd)

Consider for instance the 2SLS estimator of $\alpha_s$ in model (6).

In the first step we estimate model (7) by OLS and obtain the estimators $\hat{\pi}_{d1}$ and $\hat{\pi}_{d2}$. Take the predictor $\hat{Q} = \hat{\pi}_{d1} z_1 + \hat{\pi}_{d2} z_2$ as an instrument for $Q$.
In the second step we estimate model

$$P = \alpha_s \hat{Q} + \beta_s z_1 + error$$

by OLS. It can be verified that this 2SLS estimator of $\alpha_s$ indeed equals $\hat{\pi}_{s2}/\hat{\pi}_{d2}$.

# Demand and supply: underidentifcation

To identify $\alpha_s$ and $\alpha_d$ we made two exclusion restrictions

- $z_1$ affects $P$ but not $Q$
- $z_2$ affects $Q$ but not $P$

These exclusion restrictions are crucial for identification.

To see this, suppose, on the contrary, that $z_1$ and $z_2$ do enter both equations.

The SEM becomes:

$$\begin{aligned} Q &= \alpha_d P + \beta_{d1} z_1 + \beta_{d2} z_2 + u_d \\ P &= \alpha_s Q + \beta_{s1} z_1 + \beta_{s2} z_2 + u_s \end{aligned}$$

# Demand and supply: underidentifcation (cntd)

Then the reduced form equations become

$$
\begin{aligned}
Q &= \underbrace{\frac{\beta_{d1} + \alpha_d \beta_{s1}}{1 - \alpha_s \alpha_d}}_{\pi_{d1}} z_1 + \underbrace{\frac{\beta_{d2} + \alpha_d \beta_{s2}}{1 - \alpha_s \alpha_d}}_{\pi_{d2}} z_2 + \tilde{u}_d \\
P &= \underbrace{\frac{\beta_{s1} + \alpha_s \beta_{d1}}{1 - \alpha_s \alpha_d}}_{\pi_{s1}} z_1 + \underbrace{\frac{\beta_{s2} + \alpha_s \beta_{d2}}{1 - \alpha_s \alpha_d}}_{\pi_{s2}} z_2 + \tilde{u}_s
\end{aligned}
$$

and we lose identification (we have 6 structural parameters, but 4 reduced form parameters). Our system is <u>underidentified.</u>

# SEM: General setting

From the demand-and-supply model example we have learned two important things:

- The demand equation is identified if there is an exogenous variable that affects the supply equation but not the demand equation. Inversely, the supply equation is identified if there is an exogenous variable that shifts demand but not supply.
- The demand and supply functions can be estimated by 2SLS.

These conclusions can be transposed to the more general SEM defined by equations (1) and (2).

The parameters in this SEM are identified under a rank condition:

The parameters in equation (1) are identified if and only if equation (2) contains at least one exogenous variable (with a nonzero coefficient) that is excluded from (1). The rank condition for identification of (2) is analogous.

# SEM: General setting (cntd)

The <u>order condition</u> is necessary for the rank condition. The order condition for identification of equation (1) states that at least one exogenous variable is excluded from this equation. The rank condition requires more: at least one of the exogenous variables excluded from (1) should have a nonzero coefficient in (2). This guarantees that at least one of the exogenous variables omitted from (1) appears in the reduced form of $y_2$, so that these variables can be used as instruments for $y_2$.

The order condition for identification of equation (2) is analogous.

When the equations in a SEM are identified they can be estimated by 2SLS.

# SEM: General setting (cntd)

The parameters in (1) are estimated in two steps:

1. Estimate by OLS the parameters in the reduced form equation of $y_2$ ($y_2$ as a function of a constant and all exogenous variables). Get the predicted variable $\hat{y}_2$ from this regression.
2. Replace in the right hand side of equation (1) the explanatory variable $y_2$ by $\hat{y}_2$, and estimate the regression model that you get by OLS. The resulting estimator is the 2SLS estimator of the parameters in (1).

Estimation of the parameters in (2) is analogous.

# I understand/can apply...

- Simultaneity
- Structural form and reduced form
- Identification of a SEM
- Estimation of a SEM with 2SLS