Instrumental Variables in a Market Model
Class Notes

Manuel Arellano

September 27, 2009

Model

- Consider the following market model. The demand equation is
$$q_d = \alpha + \beta p + u \tag{1}$$
where q_d is demand (for fish, say), p is price, and u is a preference shifter.

- The supply equation is
$$q_s = \delta + \psi p - \gamma z + \varepsilon \tag{2}$$
where q_s is supply, z is rain at sea, and ε is a production shifter.

- In equilibrium $q_d = q_s = q$.

- The model determines q and p whereas z, u, and ε are determined outside the model.

- So q and p are endogenous or internal to the model, whereas z, u, and ε are exogenous or external to the model.

Reduced form

- Now imagine realizations of these variables (for the same market over time or for a cross-section of markets).
- Values of z, u, and ε occur according to some probability distribution.
- Given these values, the model produces realizations of q and p given by the reduced form:

$$
\begin{aligned}
(\psi - \beta)\, q &= (\alpha\psi - \beta\delta) + \beta\gamma z + (\psi u - \beta\varepsilon) \\
(\psi - \beta)\, p &= (\alpha - \delta) + \gamma z + (u - \varepsilon)\,.
\end{aligned}
$$

Econometric problem

- Next, we formulate the following econometric problem:

1. We have data on p, q, and z but not on u and ε.
2. We believe model (1)-(2) is correct but we do not know the values of the parameters.
3. We want to use data and the model to infer the slope β of the demand equation.

Endogeneity

- In the econometric sense, we say that p is an *endogenous* explanatory variable in the demand equation if its *realized values* are correlated with those of the error term.
- Endogeneity in the econometric sense does not imply nor is it implied by endogeneity in the economic sense (of being a variable internally determined by the model) e.g. z is external but it would be endogenous in the supply equation if correlated to ε.
- The implication of endogeneity in the econometric sense is that the equation of interest, as it applies to realized values, is not a regression (which by construction would have lack of correlation between error and regressor).
- We do not really know empirically if the realized values of p are correlated with those of u because we do not have data on u, but the model lets us expect such correlation.
- If p and u were in fact uncorrelated then β would coincide with the OLS coefficient, but in general
$$\frac{Cov\left(p,q\right)}{Var\left(p\right)}=\beta+\frac{Cov\left(p,u\right)}{Var\left(p\right)}.$$
So our theory lets us suspect that as a measure of β, the quantity $Cov\left(p,q\right)/Var\left(p\right)$ has a bias.

Instrumental variables

- In the econometric sense, we say that z is exogenous in the demand equation if it is uncorrelated with the error term (again, we do not observe it, we assume it if we believe there is no association between variation in preferences for fish and rain at sea).
- Moreover, there is an *exclusion restriction* since the theory tells us that z has not effect on demand given u and p. In other words, if we write demand as
$$q = \alpha + \beta p + \varphi z + u,$$
the theory tells us that $\varphi = 0$.
- Given this exclusion, the orthogonality condition $Cov\left(z, u\right) = 0$, or equivalently
$$Cov\left(z, q\right) = \beta Cov\left(z, p\right),$$
implies that, as long as $Cov\left(z, p\right) \neq 0$, β is determined as a ratio of data covariances:
$$\beta = \frac{Cov\left(z, q\right)}{Cov\left(z, p\right)}.$$
- If so we say that β is *identified* in the econometric problem that we posed. Otherwise, if $Cov\left(z, p\right) = 0$ then β is not identified.
- Essentially $Cov\left(z, p\right) \neq 0$ if $\gamma \neq 0$. Thus, identification of demand depends on a property of supply.
- If $Cov\left(z, u\right) = 0$ (orthogonality) and $Cov\left(z, p\right) \neq 0$ (relevance) hold, z is a valid *instrumental variable*.

A graphical representation

- Suppose that demand is inelastic ($\beta = 0$), $\psi = 1$, and $Cov(\varepsilon, u) = 0$. The model is

$$\begin{aligned} q &= \alpha + u \\ q &= \delta + p - \gamma z + \varepsilon. \end{aligned}$$

with reduced form

$$\begin{aligned} q &= \alpha + u \\ p &= (\alpha - \delta) + \gamma z + (u - \varepsilon). \end{aligned}$$

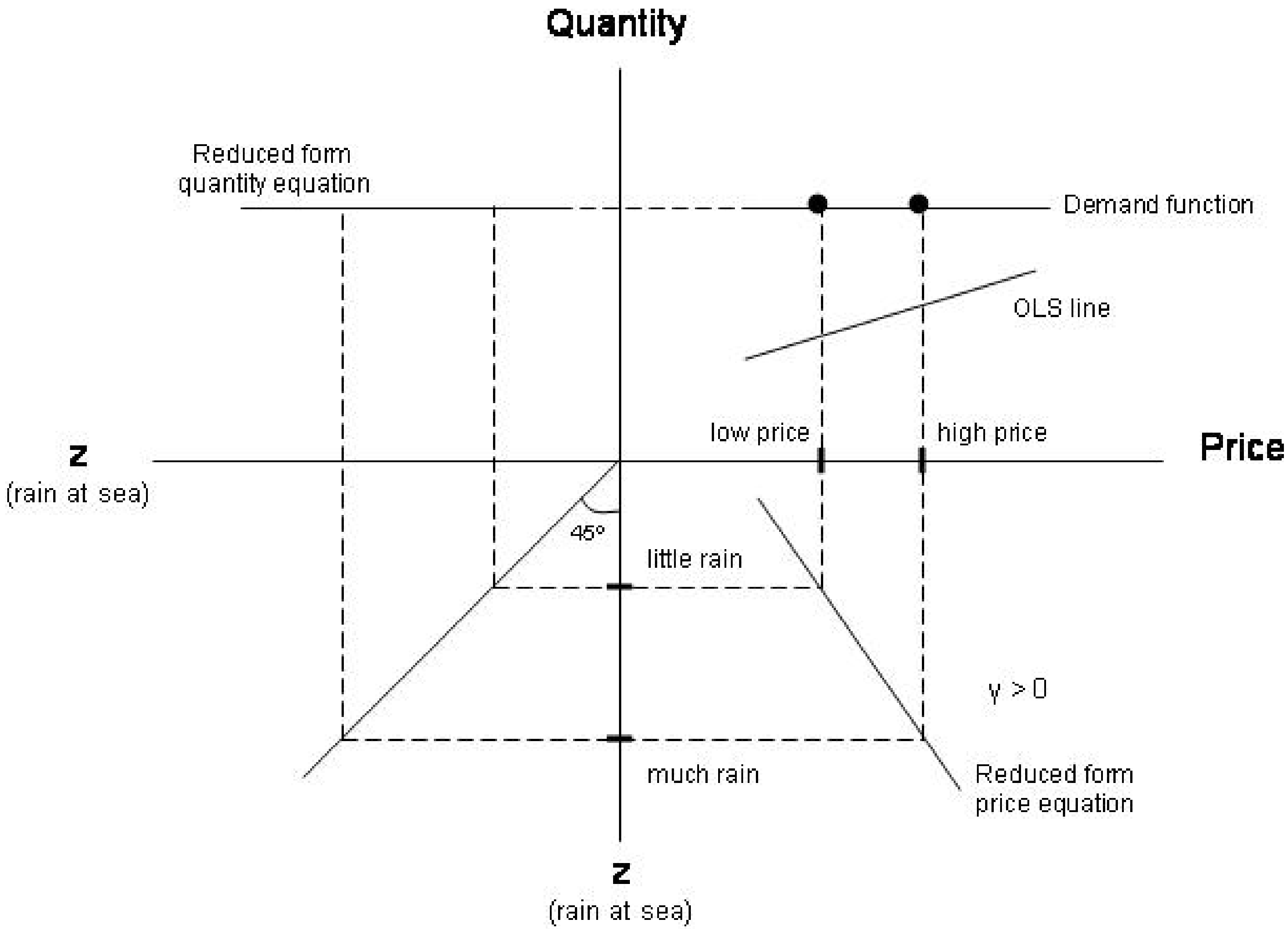
- The "first-stage regression" is a regression of p on z and has slope γ.
- The reduced-form quantity equation is a regression of q on z and has slope zero ($\beta\gamma$).
- The OLS regression of q on p has positive slope unless $Var(u) = 0$ (a perfect fit):

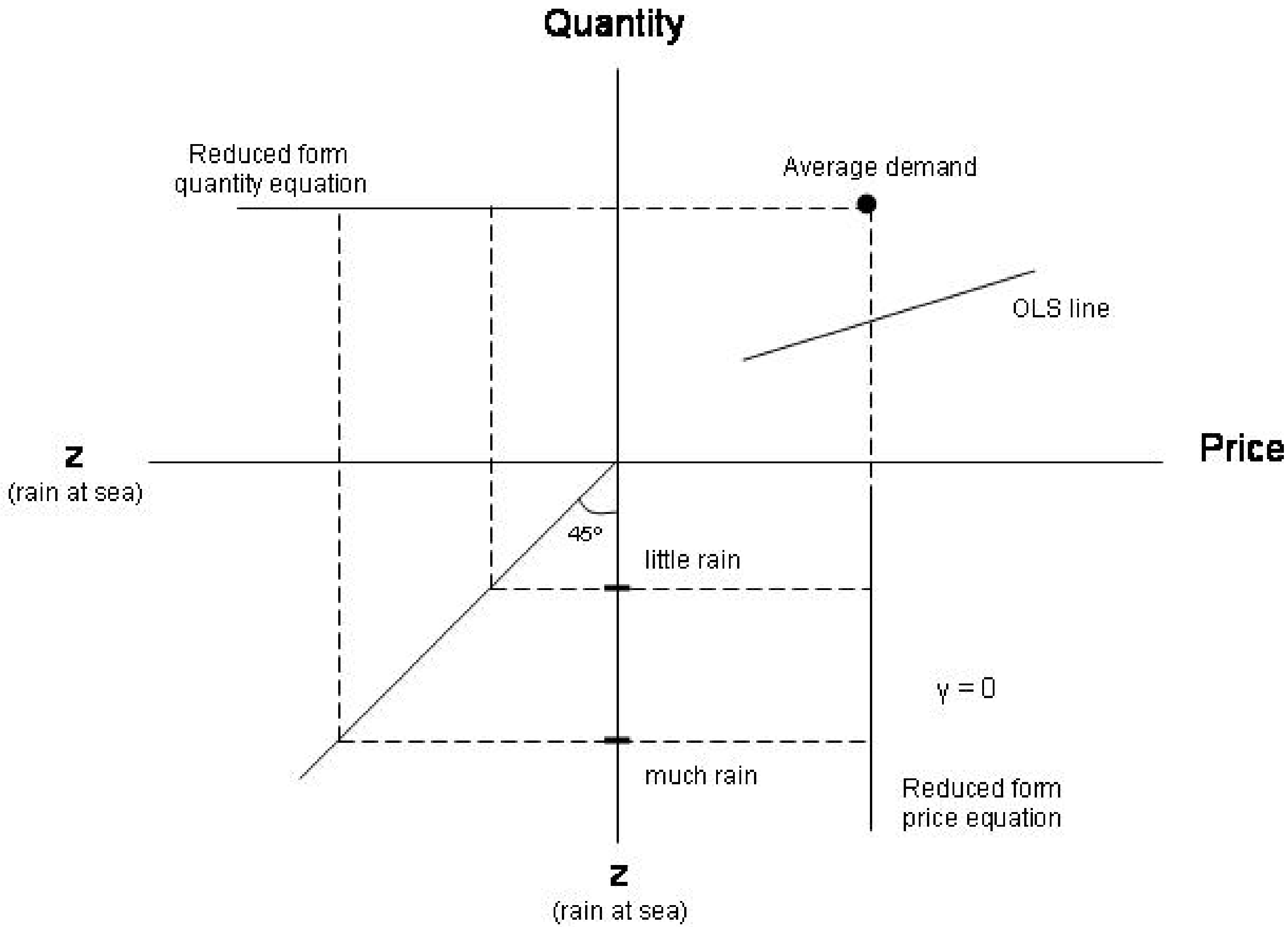
$$\frac{Cov(p, u)}{Var(p)} = \frac{Var(u)}{Var(u) + Var(\gamma z - \varepsilon)}.$$

- The relation between predicted q and p given z traces average demand for given p if $\gamma \neq 0$ (Figure 1):

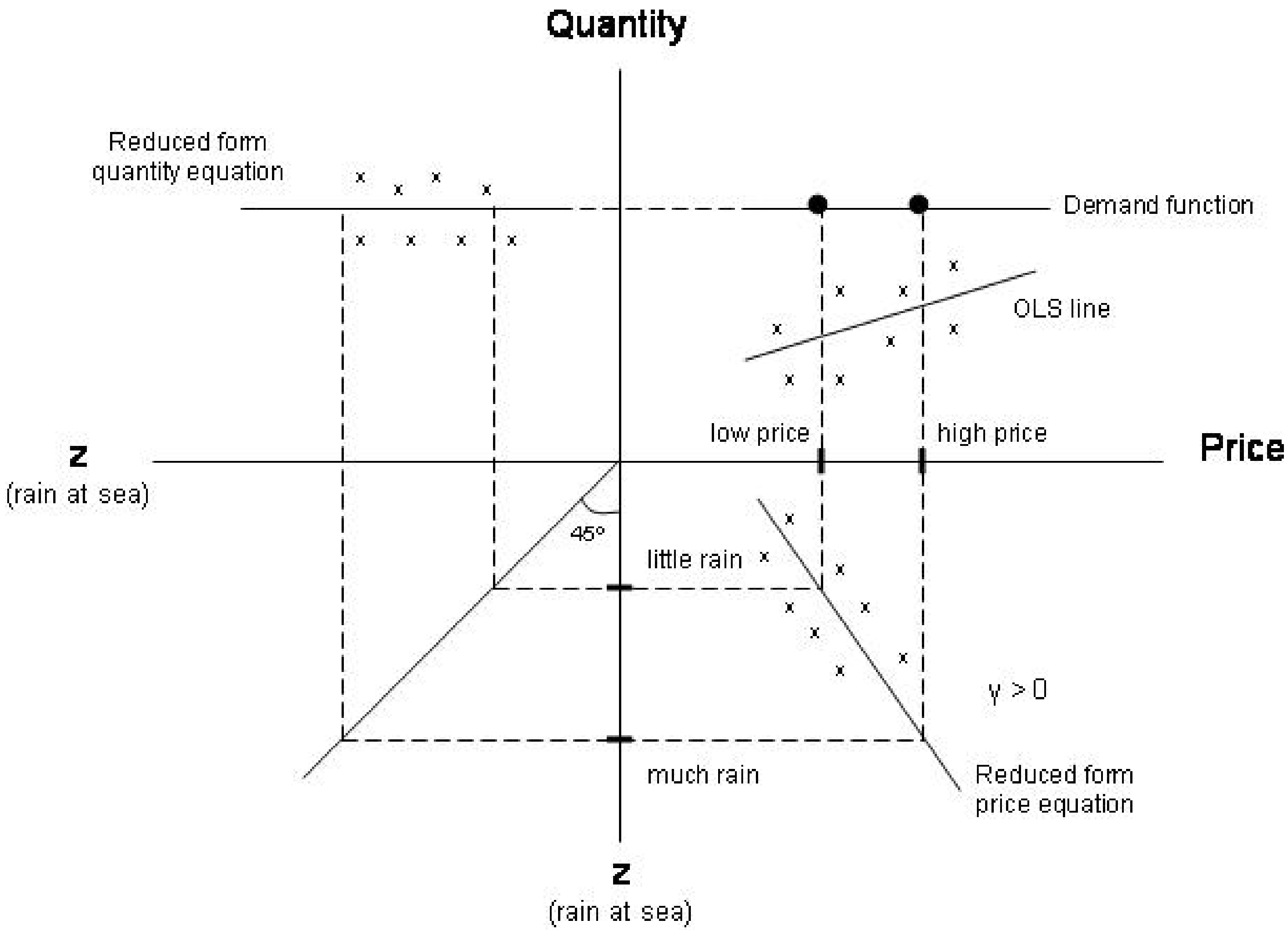
$$E^*(q \mid z) = \alpha + \beta E^*(p \mid z).$$

- If $\gamma = 0$, $E^*(p \mid z) = E(p)$ and $E^*(q \mid z) = \alpha + \beta E(p)$ for all z, so demand is not identified (Figure 2).
- Data points cluster along regression lines but not along the demand function (Fig. 3).


Quantity
Reduced form quantity equation
Demand function
OLS line
low price
high price
Price
z
(rain at sea)
45°
little rain
γ > 0
much rain
Reduced form price equation
z
(rain at sea)



2.Underidentification



3.Scatters of data points