# Additional Benchmarking Slides

Albert Banal-Estanol

# Stochastic Frontier Analysis (SFA)

- As in COLS, SFA requires specification of a…
  - production function based on input variables

- However, it…
  - does not assume that all errors are due to inefficiency
  - takes into account the possibility of measurement and specification errors
  - Needs to specify distribution of inefficiency and error

# Stochastic Frontier Analysis

- Specify (linear) relationship between output and costs

$$c_i = \beta_0 + \beta_y y_i + u_i + \varepsilon_i \ where$$

$c_i$ are the operating costs of plant (or firm) i

$u_i$ > 0 is unobserved firm-specific heterogeneity in "inefficiency"

$\varepsilon_i$ is unobserved measurement error

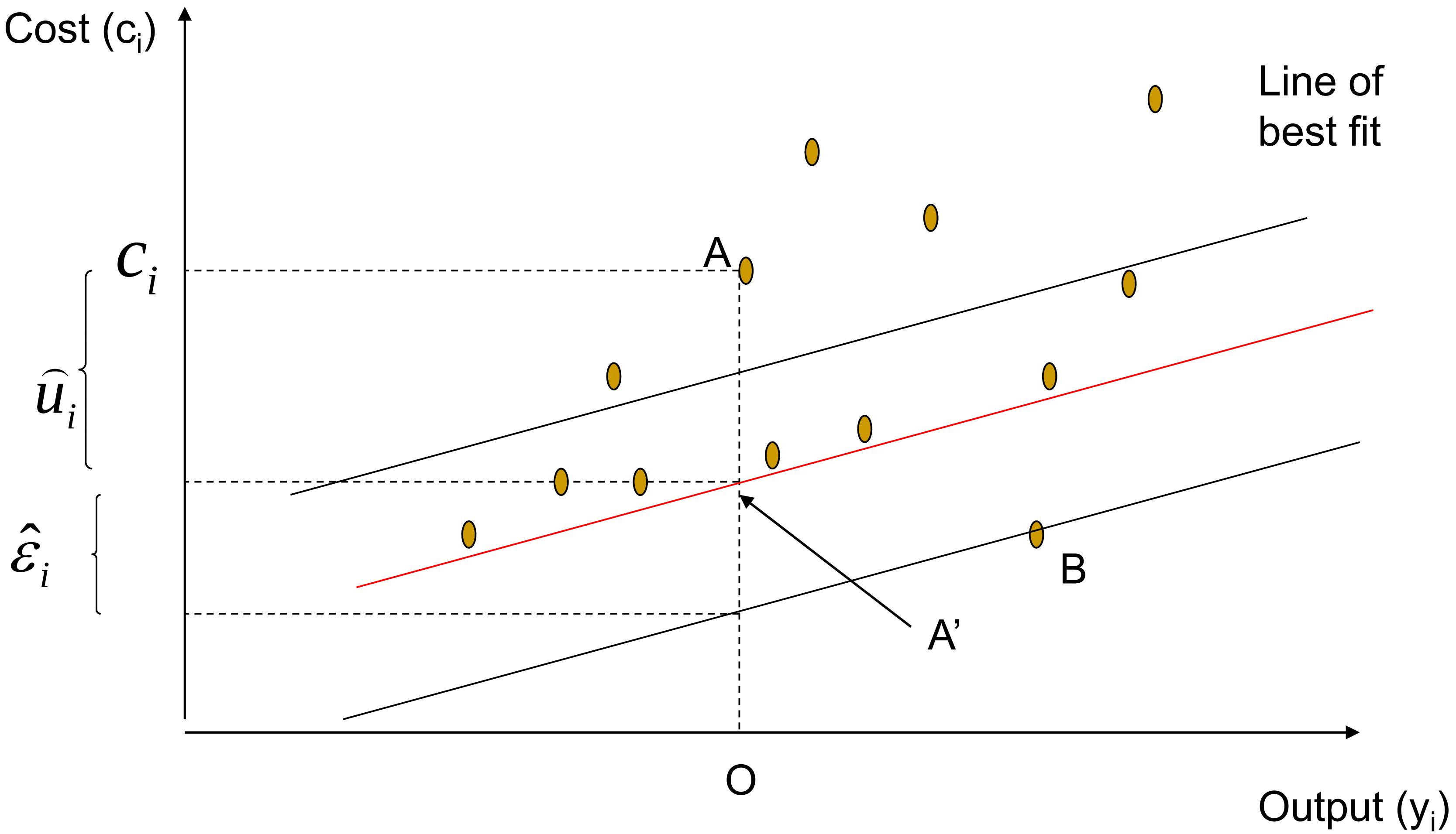
$\beta_o$ and $\beta_y$ are the parameters to be estimated

- Assume distribution of unobserved variables:

$u_i$ follows truncated normal distribution

$\varepsilon_i$ follows normal distribution

# Separating inefficiency and error



# Steps to follow

- To understand separation noise vs inefficiency:
  - Recognise presence of noise in OLS framework
  - Look at the properties of mixed distributions

# Recognising the presence of noise in OLS

- Suppose that you use COLS but your true model is

$$c_i = \beta_0 + \beta_y y_i + u_i + \varepsilon_i \ where$$

$u_i$ is an "inefficiency" term

$\varepsilon_i$ is unobserved measurement error

both are normally distributed with zero mean, and uncorrelated

- Which means that OLS estimates consistently:

$$\hat{c}_i = \beta_0 + \beta_y \hat{y}_i + e_i$$

and $e_i$ is a good estimate of $u_i + \varepsilon_i$

- How can we estimate the inefficiency?
  - We could say that a fraction (e.g. 50%) of error is inefficiency

# Relative variances

- We know that:

$$\sigma_e = \sigma_u + \sigma_\varepsilon$$

… and if we assume that both of them are equal, i.e.

$$\gamma = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2) = 0.5$$

… then our best estimate of u is half of e

- More generally, if both are normal, best estimate is

$$\hat{u} = e \times \sigma_u^2 / \sigma_e^2 = e \times \gamma$$

- But how do we know the relative size of variances?

# Stochastic frontier

- Suppose error components distributed differently

  $\varepsilon_i$ is still normal with mean zero as before

  $u_i$ is asymmetrically distributed, e.g. halfnormal (as an alternative the exponential) (both of them have a lower bound of zero)

- Mixing the two distributions with different weights:

  To which mixture the errors are closest to? (see next slide)

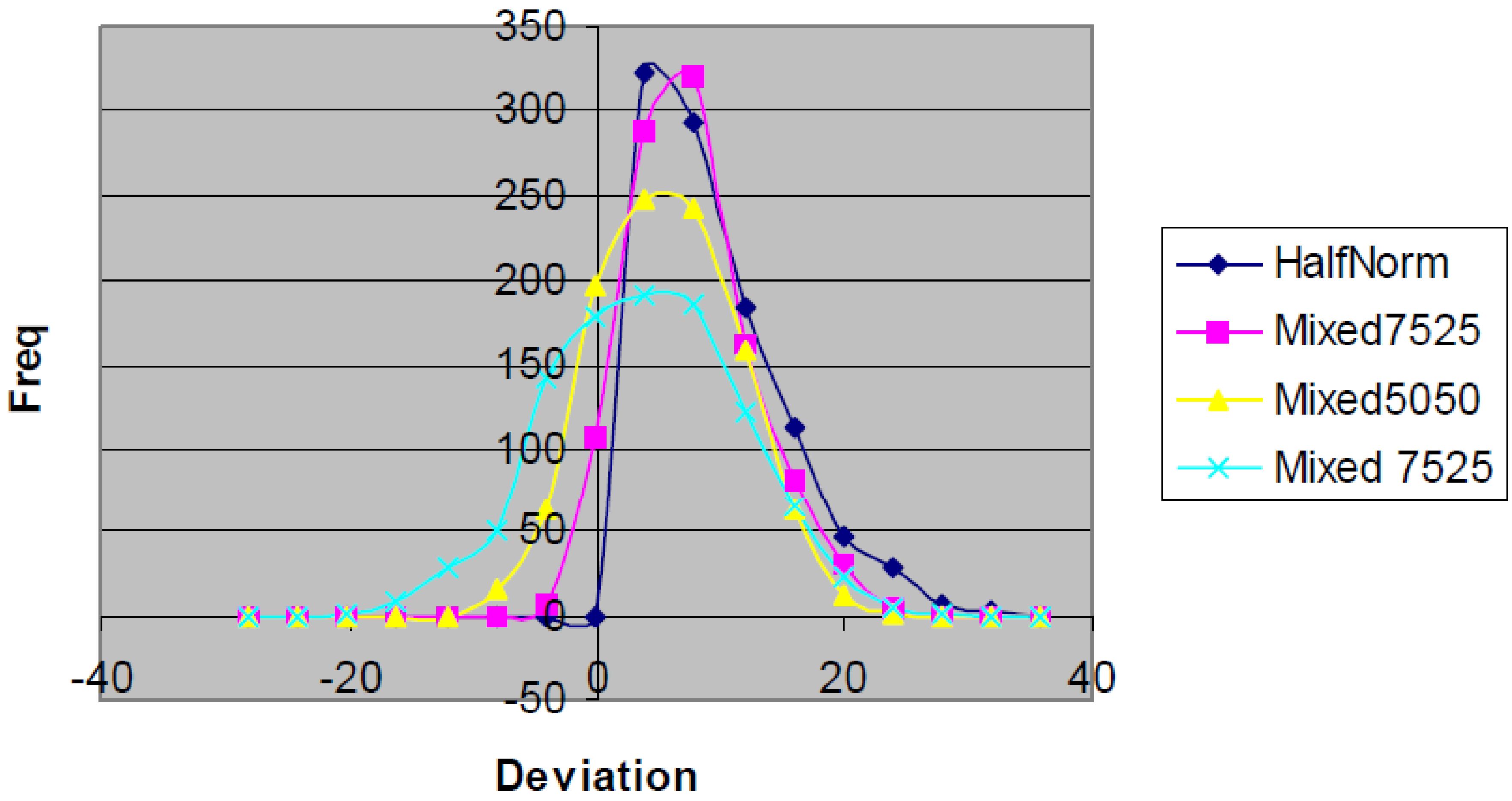
  More formally, measure skewness of the errors:

$$skewness = n\sum_i \left(x_i - \overline{x}\right)^3 / \sigma_x^3 (n-1)(n-2)$$

  where n is the number of observations,

  which varies from 0 of the full normal to 1 of the half normal

## Mixtures of normal and half-normal



# Stochastic frontier analysis(2)

- Getting at the relative variance:

i. Uses skewness of distribution of residuals:
   - To get a start on relative variance of two distributions

ii. Then, it calculates ML estimates of the model:
   - from which it can get estimates of relative variance

iii. And so on,
   - until it converges to a consistent set of parameters

- Splitting the overall residual e into u and ε:
  - Mapping from size of residual to estimated inefficiency
  - Two observations with same residual given same inefficiency

# Problems with Stochastic Frontiers

- Assumptions about error distribution
  - Results vary according for half normal or exponential (see Greene p 396).
  - Maybe use more general error distributions such as gamma distribution or truncated normal
- If inefficiencies are symmetrically distributed..
  - conclusion will be that there is no inefficiency
  - estimates of inefficiency biased down
  - See: Kennedy, J and Smith, A.S.J. (2002): Assessing the efficient cost of sustaining Britain's Rail Network