Efficiency & Productivity Analysis for Incentive Regulation

Tom Weyman-Jones,

Julia Boucinha, Celia Godhino, Catarina Feteira Inacio, Pedro Martins, Jon Latore

This paper reflects work of academic + regulatory + company economists on yardstick competition in practice

Loughborough University

Motivation: principal-agent regulation

- Single risk averse agent payment
  - if too large for principal, then X-inefficient equilibrium (Leibenstein, J. Farrell)
  - Under yardstick competition, lower payment to reduce inefficiency
- Yardstick competition, or benchmarking used by utility regulators when the network service is provided by many companies.

Questions about regulation in practice

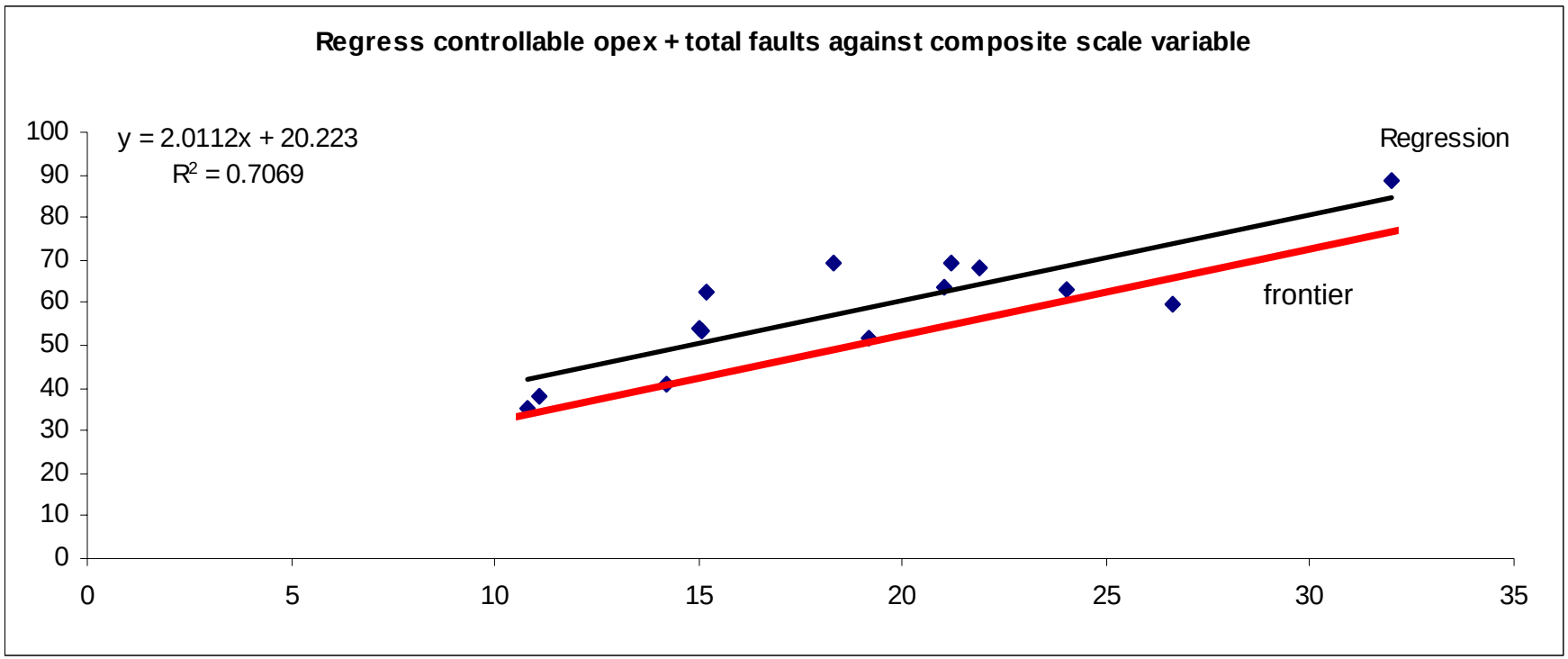
- Data tightly clustered, or highly variable?
- Widely separated data on costs = high degree of inefficiency?
- Preserve regulatory incentives with few data points?
- Is Stochastic Frontier Analysis the solution?

Current practice

- UK regulators use of deterministic frontier analysis: Data envelopment analysis (DEA) or Corrected Ordinary Least Squares (COLS) regression.
- Stochastic frontier analysis, SFA, allows for inefficiency and measurement error.
- Large sample needed, small sample properties unknown - in UK networks, no large samples

Some UK experience

- 2004 Electricity Distribution price control, 14 companies, 1 year: 2002



Loughborough University

P_0 calculation

- Costs are: *OPEX* + D* + [(wacc)RAV*]*
- * = *benchmarked figure*
- Initial price calculation:

$$P_0 = \frac{PV\left(OPEX + D + (wacc)RAV\right)}{PV\left[(1 + RPI - X)Q\right]}$$

- price control: $P_1 = P_0(1 + RPI - X)$

Disincentives to participate?

- COLS is a point estimator of the frontier like all of the yardstick estimators in the theoretical Shleifer model
- Causes the regulated firms to question yardstick competition.
- Paradox: Regressions with bad fit (widely dispersed residuals) will result in a large shift of the fitted line to make the frontier - harsh price cap

Hypothesis tests

- Contrary to UK regulators, DFA is not the default to SFA, DFA is nested in SFA

$$H_0 : \gamma = \sigma_u^2 / \left(\sigma_u^2 + \sigma_v^2\right) = 1 \Rightarrow DFA$$

$$H_1 : \gamma = \sigma_u^2 / \left(\sigma_u^2 + \sigma_v^2\right) < 1 \Rightarrow SFA$$

$$ln\,C = ln\,f(x) + v + u$$

$$v \sim iidN\left(0, \sigma_v^2\right)$$

$$u \sim iidN^{+}\left(\mu, \sigma_u^2\right)$$

$$H_0 : \sigma_u^2 = 0 \Rightarrow OLS$$

$$H_1 : \sigma_u^2 > 0 \Rightarrow SFA$$

Solutions?

- Expand the sample to international data to fit SFA models - multiple equilibria
- Relax the degree of shift of the deterministic frontier by bargaining
- Develop a decision-making procedure that incorporates the variation in the data to relax the implied inefficiency scores from the deterministic approach.

Wider questions?

- Use interval estimates because of quality of the principal's data?
- What does agency theory require from EPA?
- What does EPA imply about agency theory?
- Properties of interval estimators?

Examples of Point Estimators

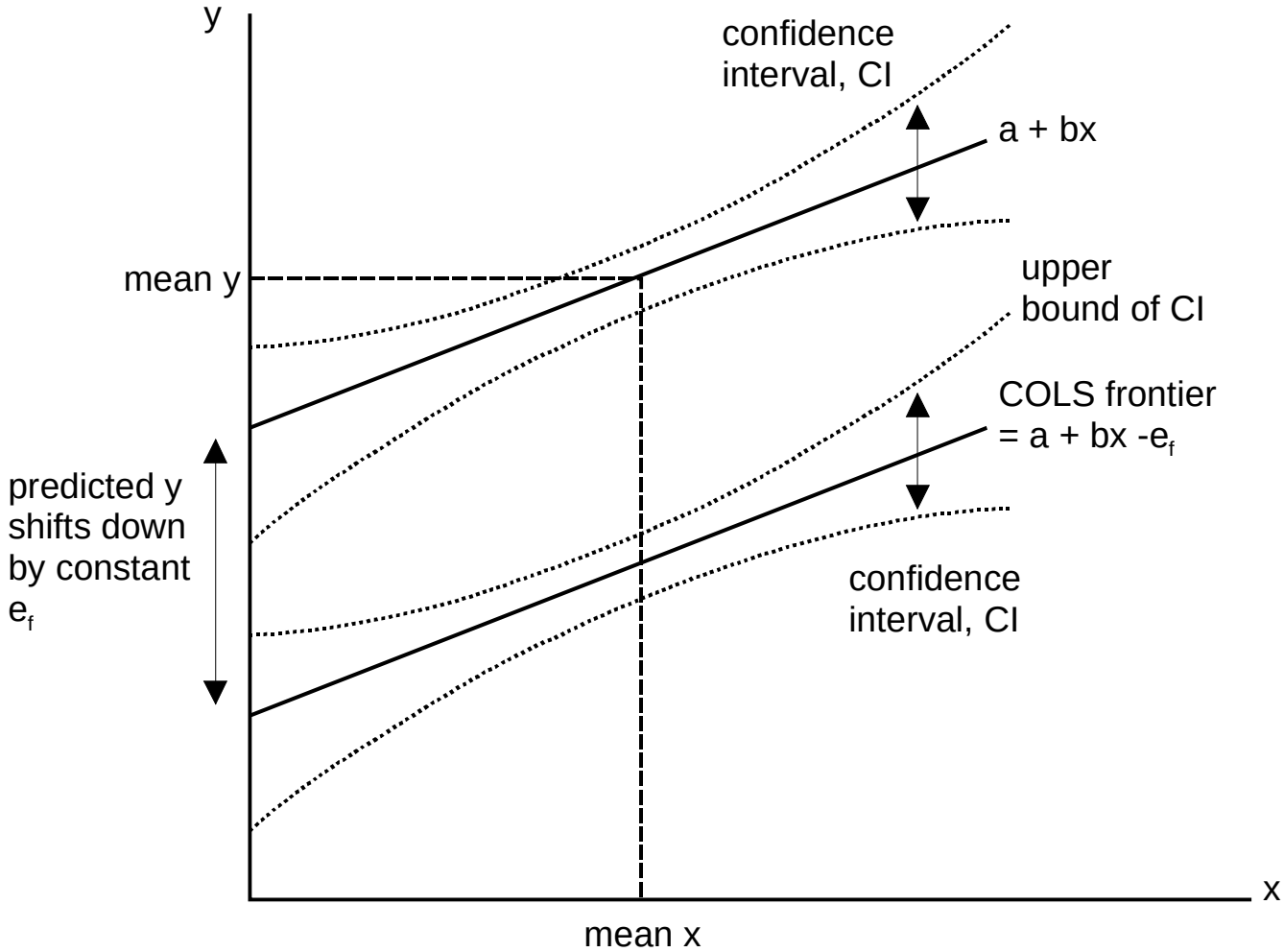
- Shleifer (1985): YC uses sample mean (agents are risk neutral, stochastic error not explicitly modelled)
- UK Water & Electricity used COLS regression
  - point estimate with no stochastic error
  - caused major transactions costs in participation
  - led to judicial reviews
- COLS/DEA widely used in Europe

Point & Interval Estimators

- Sample mean, min, are point estimators
- Agents claim sample variance reflects:
  - stochastic error (sampling error, measurement error, omitted variables)
  - as well as inefficiency
- Interval estimators:
  - depend on the variance of the principal's data
  - depend on the confidence level adopted

Example of an interval benchmark

- Upper bound of COLS prediction interval



Loughborough University

COLS upper bound

- Prediction interval for COLS frontier

$$(a - e_f + bx) \pm t_{½} \sqrt{s^2 \left(1 + (1/n) + \left((x - \bar{x})^2 / \sum (x - \bar{x})^2\right)\right)}$$

- generally:

$$\hat{c}_i - e_f \pm t_{n-(k+1)} \sqrt{est.var(c_i - \hat{c}_i)}$$

$$var(c_i - \hat{c}_i) = \mathrm{x}_i' \left(\sigma^2 (\mathrm{X}'\mathrm{X})^{-1}\right) \mathrm{x}_i + \sigma^2$$

Loughborough University

Incentives in interval benchmarking

- Helps the companies if the fit is weak and the confidence level is required to be high, while providing the regulator with an incentive to improve the sample size and reduce the unexplained variation.

An example

- Forsund & Edvardsen (2001) collected input & output data on 122 electric utilities
- There are 3 inputs: opex, losses, capital costs
- There are 3 outputs: number of customers, lines, energy (mWh)
- Units are both physical and financial

Input Data Problems

- Financial data as inputs: assumes that the prices of inputs are the same for all DMUs - doubtful for an international sample
- Capital is measured by the value of assets, but this may be determined by regulators

Output Data problems

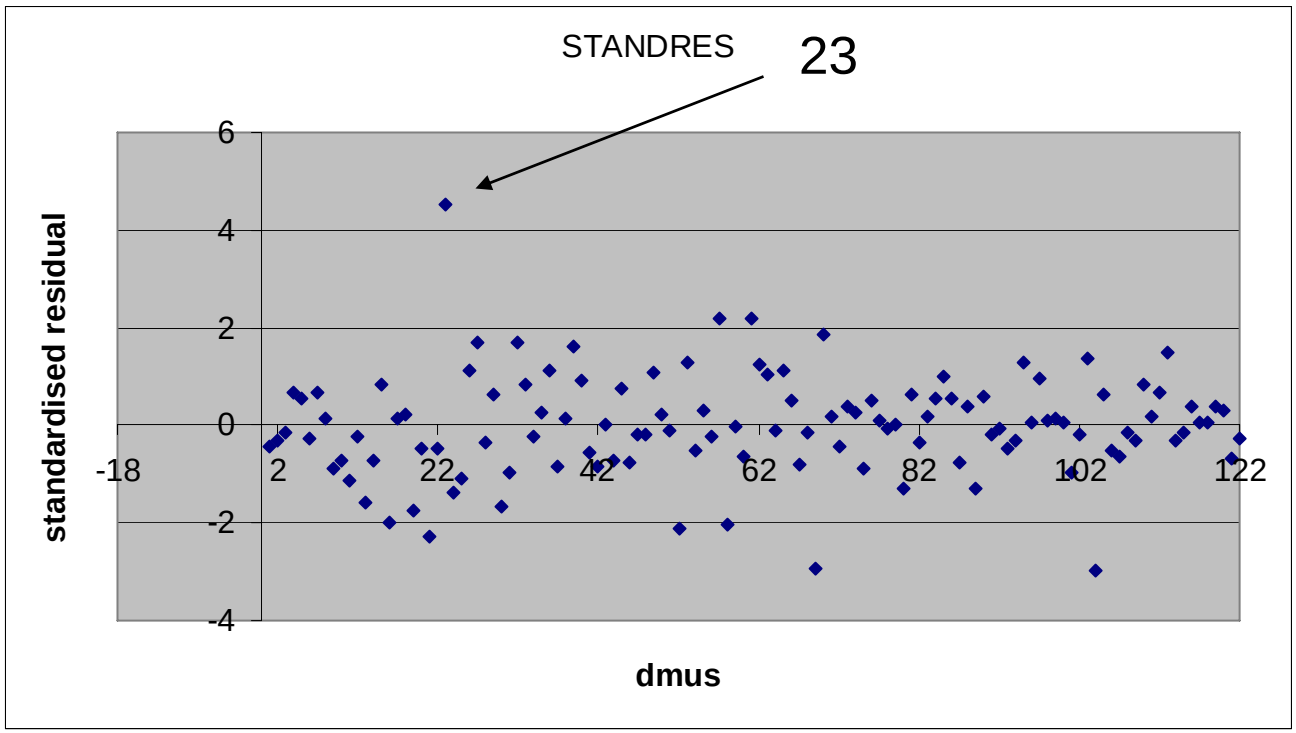
- Customers and energy delivered are standard output variables
- Lines: i.e. network length is a debatable output.
  - Network with longer lines has more difficulty in meeting customer needs
  - but network length has already appeared implicitly in the value of assets as an input
  - lines may be better as an uncontrolled environmental variable in DEA, but does not affect input distance function

Fit translog input distance function

- Input distance = 1/technical efficiency
- 3 inputs
- 3 outputs (or lines is exogenous characteristic)
- CRS or VRS?
- Check for economic properties before accepting efficiency scores

OLS regression

- Use to test for outliers, especially superefficiency:
- |standardised residual| > 2



Loughborough University

Full sample

- Three superefficient companies at least.
- Also several with large negative residuals.
- Several experiments we can try for elimination of outliers.
- Here, I will do 3 sets of results:
  - full sample (n = 122)
  - remove extreme efficient firm (n =121)
  - remove three firms with stand. Res. > 2 (n = 119)

SFA-MLE results

Review Table: Eliminate one outlier: normal exit from iterations

Variable	SFATBL1 Parameter	t-ratio
Constant	0.200139	5.16564
LXN1	0.312313	7.28249
LXN2	0.094784	1.90997
LY1	-0.24918	-5.74296
LY2	-0.54041	-13.4347
LY3	-0.19309	-3.36657
Second order terms	…..	…..
…..	…..	…..
Lambda	2.37778	2.00365
Sigma	0.17526	8.90181
Log-L	89.6511	

Review Table: Eliminate 3 outliers: iterations limit reached

Variable	SFATBL3 Parameter	t-ratio
Constant	0.242231	5.02348
LXN1	0.279638	4.72628
LXN2	0.072087	0.938249
LY1	-0.27098	-3.8027
LY2	-0.57681	-8.11813
LY3	-0.16942	-1.50817
Second order terms	….	….
….	….	….
Lambda	5559.91	0.004586
Sigma	0.200018	10.8374
Log-L	105.117	

- Full sample: wrong skew; inefficiency detected without superefficient firms
- all 1st order properties satisfied at sample means.

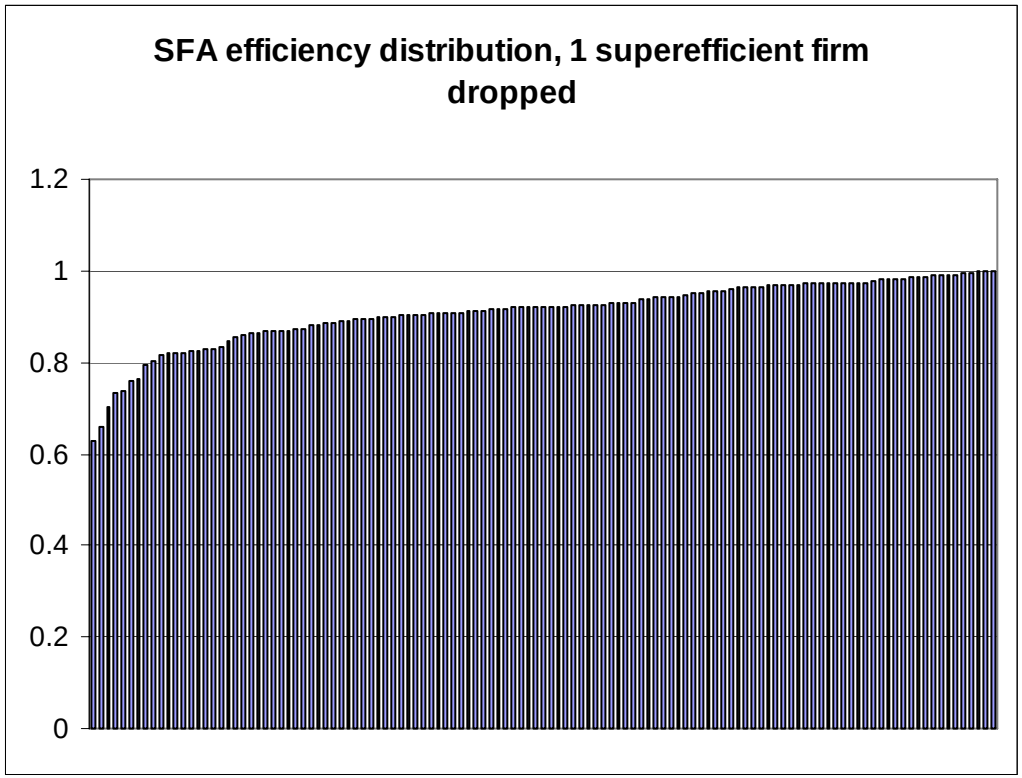
SFA Significance tests from Frontier

	coefficient	standard-error	t-ratio	
gamma	0.489	0.366	1.335	full sample
gamma	0.849	0.123	6.857	1 super eff dropped
gamma	0.999	0.145E-06	6867690.0	3 super eff dropped

- gamma = $\sigma^2_u/(\sigma^2_u + \sigma^2_v)$; H0: gamma = 0, i.e. $\sigma^2_u = 0$
- mu is restricted to be zero (half normal rather than truncated normal)
- LR test of the one-sided error, 5% significance critical $\chi^2 = 2.71$
- = 0.299 (full sample), fail to reject H0
- = 5.817 (1 super eff dropped), reject H0
- = 20.226 (3 super eff dropped), reject H0
- with number of restrictions = 1
- [note that this statistic has a mixed chi-square distribution]
- dropping 3 superefficient firms suggests no stochastic error (COLS) but we have not considered the very inefficient outliers: stand res < -2

SFA

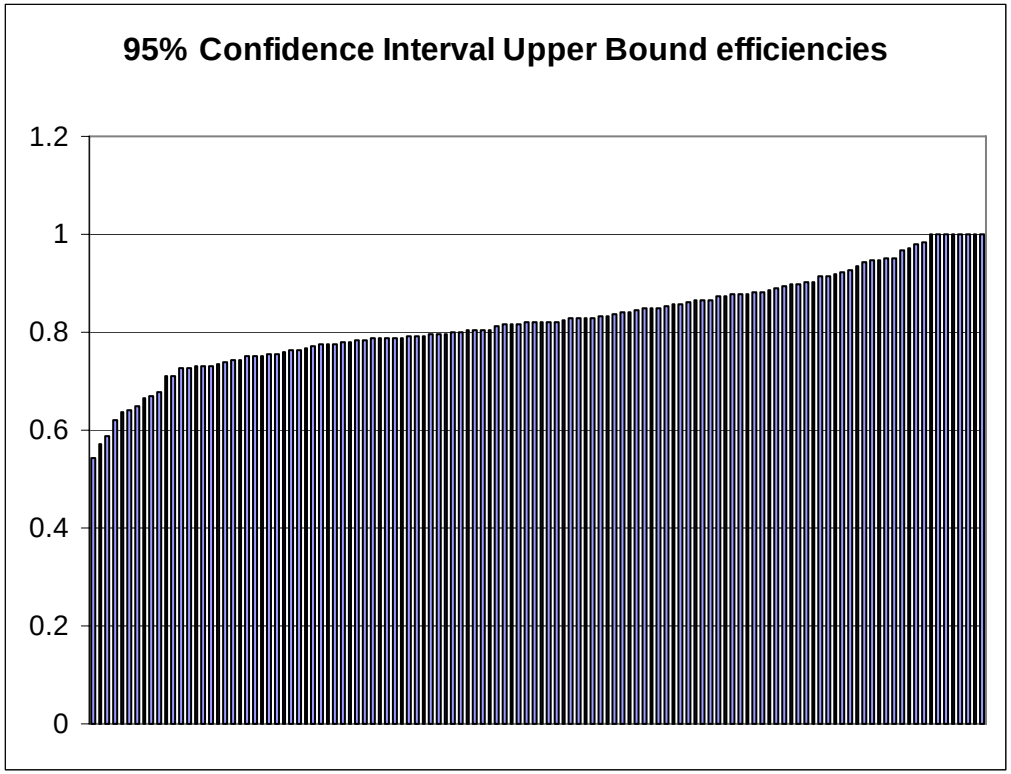
- Adjust sample of regulated companies to make SFA work



Loughborough University

SFA & CIUB

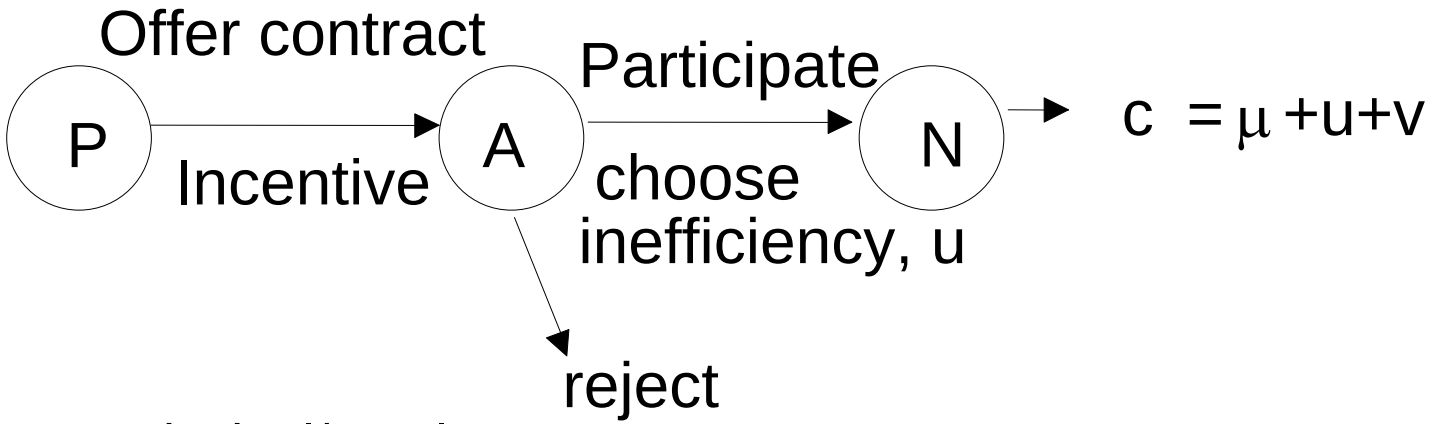
- Or, the CIUB can be incentive compatible when SFA is rejected but COLS is too harsh



Loughborough University

Regulatory Review Process

- Principal-Agent Hidden Action Game



P: principal/regulator
A: agent/firm
N: nature/realisation of random variable, v

Regulator's problem

μ: the frontier level of cost ($\mu = \alpha + \beta x$ in a regression framework)
u: the agent's chosen level of effort or inefficiency
v: stochastic error – e.g. measurement error, sampling error, omitted variables

In performance measurement with two comparators,

$c_i = \mu + u_i + v_i$ and $c_j = \mu + u_j + v_j$

No stochastic error: DEA/COLS - not a true agency problem

costs vary only with different levels of inefficiency.

$c_i = \mu + u_i$

Correlation of observed cost and inefficiency = 1.

$r_{cu} = 1$

Eric Rasmusen (2001)

"unobservability of effort is not a problem … if the contract can be conditioned on something which is observable and perfectly correlated with effort. The true agency problem occurs when that correlation breaks down"

Inefficiency + stochastic error defines true agency problem

correlation between observed cost and inefficiency with stochastic error:

$$c_i = \mu + u_i + v_i$$

$$\Rightarrow r_{cu} = \sqrt{\frac{\operatorname{var} u}{\operatorname{var} v + \operatorname{var} u}} < 1 \quad and \quad r_{cu} \to 0 \; as \; \operatorname{var} u \to 0$$

$H_0 : r_{cu} = 0$ against: $H_1 : r_{cu} > 0$

Rejecting H_0 in favour of H_1 means significant levels of inefficiency in the performance data.

Nash equilibrium in Yardstick Competition

Assume: “agents are alike”. They will choose very similar levels of effort or efficiency. This time we look at the correlation of their observed costs:

$$r_{c_i c_j} = \frac{\operatorname{cov} u_i u_j}{\sqrt{\operatorname{var} u_i + \operatorname{var} v_i}\sqrt{\operatorname{var} u_j + \operatorname{var} v_j}} > 0$$

Gangepain & Ivaldi (2002)

with a high powered incentive contract the values of the explanatory variables, i.e. input prices and outputs, are endogenised, so they become correlated with the effort level chosen by the firm. Since this effort level forms part of the inefficiency component of the SFA model, we can summarise by the result:

$r_{xu} \neq 0$

$\Rightarrow$ inconsistent estimates and difficulty in identifying the inefficiency arising from low effort.

SFA different from game theory

SFA model assumes zero correlation between both the inefficiency and stochastic error components

Otherwise: error terms correlated with each other in cross section samples.

SFA models different, non-correlated, inefficiency outcomes, or multiple equilibria & assumes independence of composed error term

$$r_{c_i c_j} = \frac{\operatorname{cov} u_i u_j}{\sqrt{\operatorname{var} u_i + \operatorname{var} v_i}\sqrt{\operatorname{var} u_j + \operatorname{var} v_j}} = 0 \qquad r_{xu} = 0$$

Shleifer model

- For a unique, symmetric, efficient, Nash equilibrium in an *n*-agent regulatory game. Shleifer (1985) proposition 1:
- fall in *j's* cost should not raise *i's* price cap

$$p_i = f\left(c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n\right)$$

$$f\left(a, \ldots, a\right) = a$$

$$\frac{\partial f}{\partial c_j} \geq 0, \text{ for any } j$$

Loughborough University

Some point estimates work

- *point estimator* of the average of the observed company costs

$$p_i = \bar{c}_i \quad \text{where} \quad \bar{c}_i = \frac{\sum_{j \neq i} c_j}{n-1} \qquad \frac{\partial p_i}{\partial c_j} = \frac{1}{n-1} \quad \text{any } j \neq i$$

- regression estimator does not

$$p_i = \bar{c}_i + b\left(x_i - \bar{x}_i\right)$$

$$\frac{\partial p_i}{\partial c_j} = \frac{1}{n-1} + \frac{\left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_i\right)}{\sum_{k \neq i}\left(x_k - \bar{x}_i\right)^2} = ?$$

COLS type methods

- COLS point estimator

$$p_i = \bar{c}_i + b_i\left(x_i - \bar{x}_i\right) + \min\left(c_j - \hat{c}_j\right)$$

- COLS interval estimator

$$p_i = \bar{c}_i + b_i\left(x_i - \bar{x}_i\right) + \min\left(c_j - \hat{c}_i\right) + t * \sqrt{est\,\mathrm{var}\left(c_i - \hat{c}_i\right)}$$

$$\frac{\partial CI_i}{\partial c_j} = \left(\left(\frac{1}{n-1}\right) + \frac{\left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_i\right)}{\sum_{k \neq i}\left(x_k - \bar{x}_i\right)^2}\right) + \frac{t * \left(c_j - \hat{c}_j\right) M^{1.5}}{(n-3)s_i}$$

$$M = \left(1 + \left(1/(n-1)\right) + \left(\left(x_j - \bar{x}_i\right)^2 \Big/ \left(\sum_{k \neq i}^{n}\left(x_k - \bar{x}_i\right)^2\right)\right)\right)$$

Loughborough University

Summarise

- True agency problem represented by SFA
- But SFA assumptions and Agency theory predictions do not coincide
- 1 sided error demonstrates regulatory failure
- If 1 sided error fails, how can agency problem be resolved?
- Confidence interval benchmarking?
- Shleifer properties are sample dependent.