Water Utility Benchmarking: Measurement, Methodologies, and Performance Incentives

Sanford V. Berg

Distinguished Service Professor—Economics

Director of Water Studies, PURC

(With assistance from Rui Cunha Marques)

Some Thoughts

Spare us from cowardice that shrinks from new truths;

Spare us from laziness that is content with half-truths; and

Spare us from arrogance in thinking that we know all truth.

Public Utility Research Center

Research Expanding the body of knowledge in public utility regulation, market reform, and infrastructure operations (e.g. benchmarking studies of Peru, Uganda, Brazil and Central America)

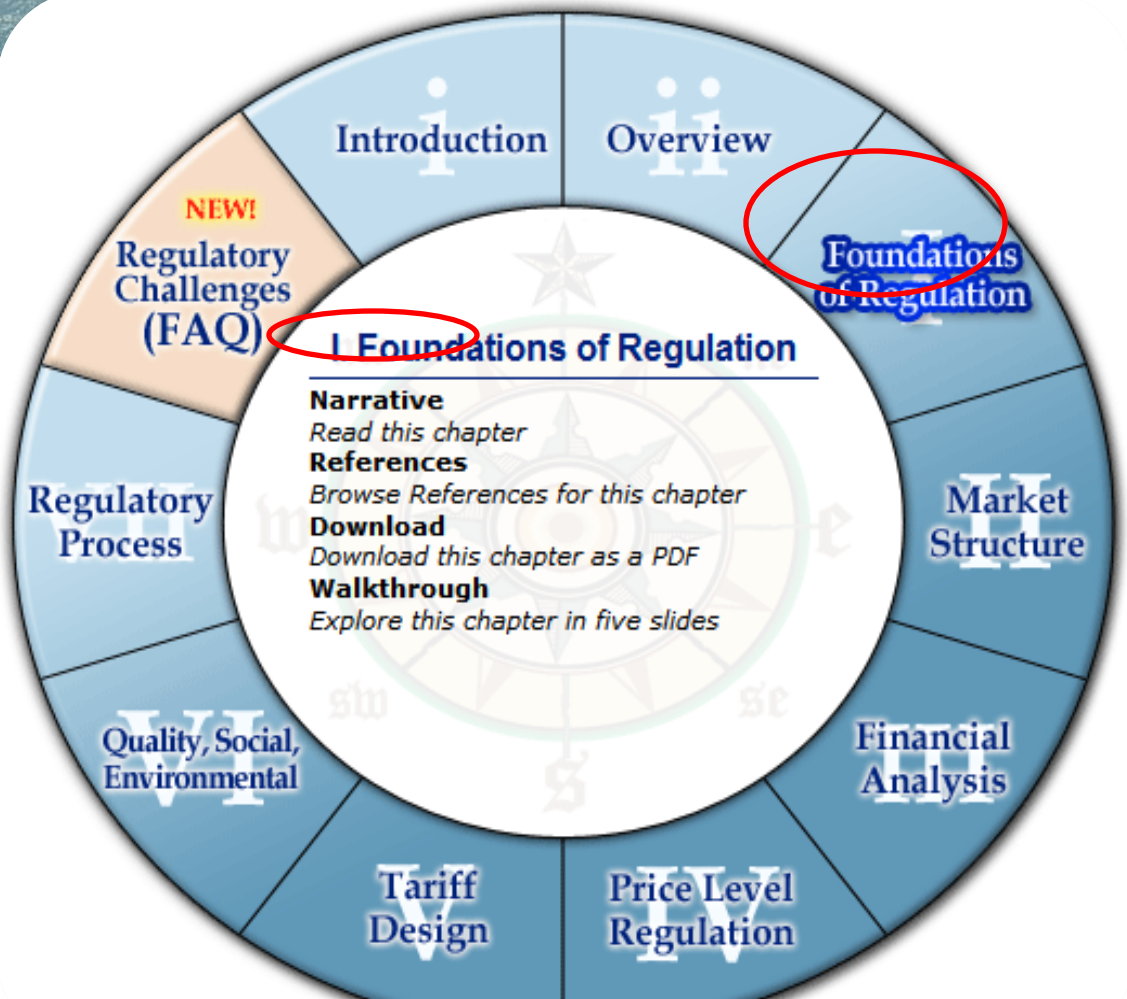
Education Teaching the principles and practices that support effective utility policy and regulation (e.g. *PURC/World Bank International Training Program on Utility Regulation and Strategy*, January 2010)

Service Engaging in outreach activities that provide ongoing professional development and promote improved regulatory policy and infrastructure management (30 in-country trainings and university collaborations)

Foundations of Regulation

The Body of Knowledge on Infrastructure Regulation is divided into seven main sections. Chapter I introduces the general concepts presented in the Body of Knowledge.

www.regulationbodyofknowledge.org



Narratives summarize key topics

Additional Resources: Glossary in Spanish and four other languages

Frequently Asked Questions, including

•Social pricing to promote access

•Management and regulation of State-owned Enterprises

500 PDFs as References

Self-paced Quizzes (for capacity building and classrooms)

Context for Managers and Regulators

Significant change is a continuing condition: the future holds more **unknowns** than it does certainties.

Task: develop fresh perspectives and knowledge about the future, trusting the wisdom of the past.

Civil Society demands performance improvements (e.g. Unaccounted for water).

(Mark Jamison and Araceli Castaneda, 2009)

1. *Next* practices, not best practices

Best practice is about imitation (following in someone else's footsteps). A focus on next practice is needed when we are going into areas where no one has gone before.

Implication: Consider Metric Benchmarking

Heifetz, Grashow, and Linsky. 2009. "Leadership in a (Permanent) Crisis." Harvard Business Review

2. Focus on *Why* rather than on What

When we ask ourselves "What should we do next?" we emphasize practice. But the practice needs a foundation, basic principles, and values.

Ask "***Why*** have certain practices been successful or unsuccessful?" so that we analyze our underlying priorities and our context.

We learn, keep what is important, and discard what holds us back.

Collins, 2009. How the Mighty Fall and Why Some Companies Never Give in

3. *Leadership* vs. *Leading*

A leader provides direction (when the right direction is already known).

Leadership mobilizes people to tackle difficult and often ambiguous problems and circumstances.

Decision-makers currently face uncertainty and ambiguity. They need to exercise leadership.

(Heifetz, Ronald A. 1994. Leadership Without Easy Answers, p. 15)

Benchmarking: Can an Index Capture Complexity?

A single index of utility performance will be neither comprehensive nor fully diagnostic.

- Physician can have information on a patient's temperature, pulse, height and weight.
- Patient is in trouble: dangerous fever and/or is significantly overweight.
- Blood tests provide more detailed information
- Diagnosing and treating mental health issues would require other diagnostics and treatments . . . Still, temperature and weight provide useful information.

Issues

1 Why do regulators and managers benchmark?

2 What are the available techniques?

3 Identifying strong and weak performance

4 Providing incentives

5 How should prices be set?

6 Concluding remarks

Why do regulators and managers benchmark?

The fewer the facts,
the stronger the opinion.

Reduce information asymmetry: get facts

- Utility Managers have (or should have) access to detailed information on opportunities for cost containment, service quality improvement, and network expansion.
- Oversight (regulatory) Institutions benefit from benchmarking.

Steps: Data collection, Model specification, Analysis, Identification of strong and weak performers.

OFWAT's Use of Benchmarking

- **OPEX:** Econometric models of to assess relative efficiency;
- Companies performance measurement against others;
- "Imitates" a competitive market;
- Supports company-specific efficiency targets;
- More demanding targets for the less efficient utilities;
- **CAPEX:** Comparison of standard costs used to challenge forecasts of capital work costs;
- Published models and data, based on standard definitions;
- Subject to challenge and review;
- Special factors taken into account.

Cost Example: Ofwat Water Service Model (Y = Opex*)

Y: Operating expenditures (less "exceptional items", rates, third party services, abstraction charges, pumping costs)

$$\text{Ln } Y = 3.57 + 0.471 \text{ Ln } X_1 + 0.468 \text{ Ln } X_2 - 1.575 \text{ Ln } X_3$$

X1 = water delivered in Ml/day

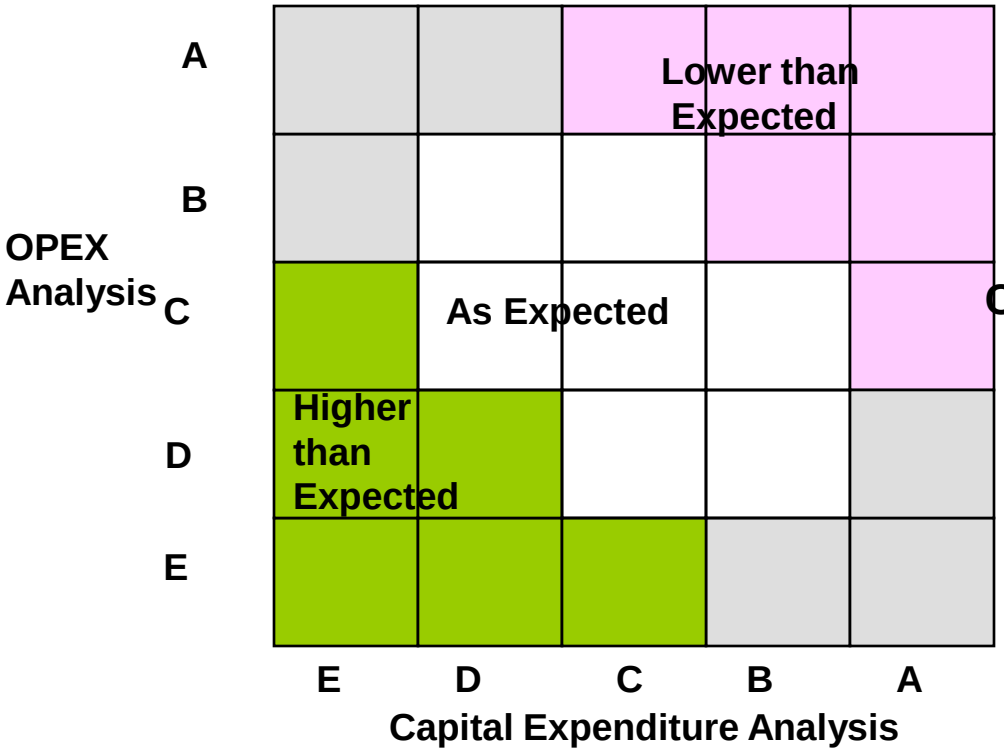
X2 = length of main in km

X3 = proportion of water delivered to measured non-households

(Chaplin, United Utilities)

OFWAT Trade-offs: Performance Benchmarking (X Factors)

Change in Price Cap: CPI – X + K +Q

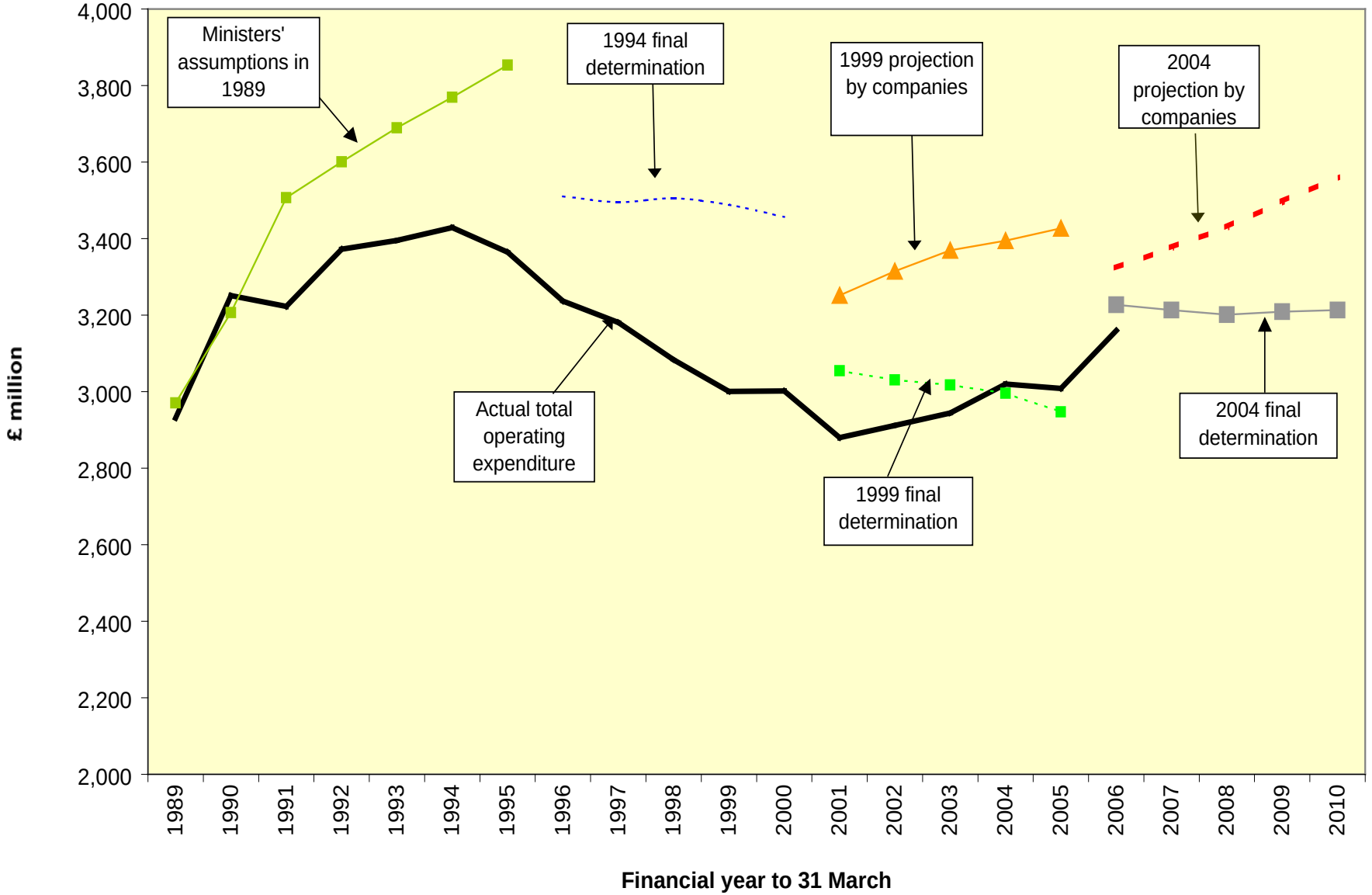


AA: clearly outstanding

CPI: Consumer Price Index
X: Productivity Adj.
K: Capacity Adjustment
Q: Environmental Adj.

A: excellent
E: very weak

Projections vs. Actual OPEX



Uganda NWSC	1998	2006
Service Coverage	48%	**Could a developing**
Unaccounted for water	51%	**nation apply principles**
Percent Metered	65%	**from the UK?**
Percent Connections Active	63%	
New Connections/ year	3,317	
Total Connections	50,826	
Turnover (Revenue)	US$ 11 million	

Uganda NWSC	1998	2006
Service Coverage	48%	**70%**
Unaccounted for water	51%	**29%**
Percent Metered	65%	**99.6%**
Percent Connections Active	63%	**94%**
New Connections/ year	3,317	**23,312**
Total Connections	50,826	**148,312**
Turnover (Revenue)	US$ 11 million	**US$ 34 million**

Incentives Based on Benchmarking

- Understanding past and current performance
- Establishing open communication between workers, managers & executives
- Developing achievable goals & sound strategies
- Increasing accountability for workers & managers
- Designing incentives & rewards to encourage efficiency **(bonuses of 50% if targets met)**
- Reviewing progress and readjusting strategies

"Using internal incentive contracts to improve water utility performance: the case of Uganda's NWSC," by Mugisha, Berg, & Muhairwe, *Water Policy*, 2007.

Okay, Benchmarking can Improve Performance.

What are the available techniques?

Classifications for Metric Benchmarking

- Total methods or partial methods (ratios);
- Frontier or non-frontier;
- Parametric or non-parametric;
- Stochastic or deterministic;

Other types of Benchmarking

- Process Benchmarking
- Customer Perception Surveys
- Model Company Comparisons (engineering models)

Practical issues: Data Availability

You Need adequate data of sufficient quality

'garbage in' → garbage out .

Consultants and academics recognize that:

"If you torture the data enough they will confess."

Decision-makers should note that not all elements that <u>can</u> be counted really "count":

"Make what is important measurable; do not make what is measured important."

Identifying strong and weak performance: Data from PURC Case

You have collected data:

twenty water distribution utilities.

1. Rankings: Identify the three best and the three worst performing firms.
2. Rationale: Explain why you placed these firms in these categories.
3. Robustness: How would you determine whether your rankings (or performance scores) are robust?

Find a partner and examine the data

Identify three best and three worst-performing water utilities in the sample

OPEX: operating expenditures

Number of full time Staff (Labor)

Quantity Delivered (Q)

Number of Customers

OPEX/unit delivered

% of water accounted for

Availability (hours per day the residence receives service)

Coverage (% of population receiving service)

Metered Ratio (% of customers with meters)

Quality (% water treated and tested)

Other ratios: Q/OPEX and Q/Labor

Votes Strong, Weak

1	11
2	12
3	13
4	14
5	15
6	16
7	17
8	18
9	19
10	20

Providing incentives

Rankings:

Many benefits:

- Very useful. They are effective in calling attention of stakeholders to issues!...
- Sunshine Regulatino (“name and shame”)
- Consequences for tariff-setting (carrot and stick approach).

Several shortcomings:

- Selection of performance measures (often partial ones);
- Weighting the indicators;
- Dealing with bad and missing data;
- Reliability;
- Stability (over time).

Robustness of rankings:

Empirical rules:

- Minimum and maximum scores;
- The same trend;
- Comparable means, standard deviations, and other distributional properties;
- Stability over time;
- Correlation with intuitive partial measures.

Statistical tests:

- Spearman and Pearson tests
- Kendall Tau when there are ties

Importance of **regulatory processes** utilizing **performance incentives:**

Advantages:

➔Offering strong incentives towards efficiency and innovation;

➔Fostering the sharing and transparency of information;

Problems:

➔ Omitted explanatory factors?

➔ Does the methodology affect the results?

➔ Confidence in the underlying Data?

➔ Is there trust among stakeholders?

Once we can rank performance,

How should prices be set?

- Problems of price cap regulation when compared with rate-of-return (attention to quality);
- The regulatory review period (incentive versus risk);
- X Factor for whole the sector or individual firm (depending on the activity regulated?)
- How can each firm get close to the frontier? What is a reasonable target?
- The annual target should reflect gradual (not instantaneous) attainment (e.g. trying to attain the fourth quartile (75%);

- The target should be gradually set (trying to attain the fourth quartile at least (75%);
- The business plan is essential. How is it evaluated and monitored?
- The dichotomy between CAPEX and OPEX
- Attention to quality of service? Penalties for not meeting targets?
- Does the regulatory method depend on ownership or on the financial health of the utilities?

Example of Computation of X Factor

Technical change - obtained by industry average TFP (Törnqvist) – 1.5% per year

Efficiency change obtained by DEA or SFA techniques
Firm A - 0.775

Assume a six-year regulatory period; ask firms to achieve 1.5% per year plus catch-up factor of 75% towards frontier

For firm A: X Factor = 1.5 + 0.75 (1-0.775)/6=
1.5 + 2.8 = 4.3% per year

Concluding remarks

Benchmarking is Part of the Tool-kit for Regulators and Managers

- The application of the techniques summarized here can improve service quality, expand networks, and optimize operations.
- Any benchmarking study will have limitations, but sound studies can be used to place the burden of proof on other parties who might argue that the analysis is incomplete or incorrect.
- Over time, data availability will improve and studies will be strengthened as professionals gain experience with these quantitative techniques.

Improving Health: **"Do No Harm"**

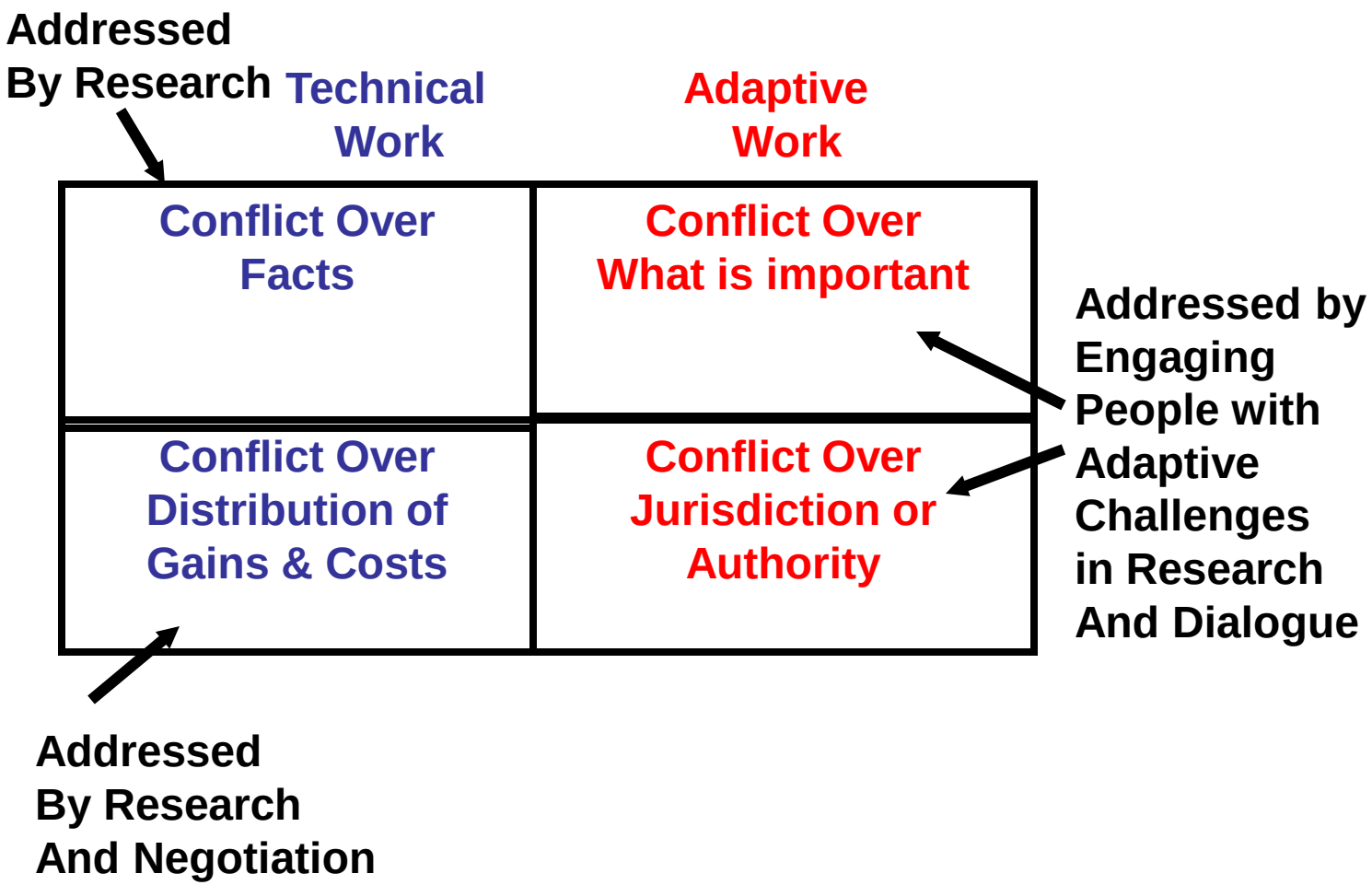
- Benchmarking specialists produce and critique studies that utilize various methodologies.
- Rankings can be manipulated by choice of variables, model specification, sample size, time frame, and outliers.
- Results can be misinterpreted & misused.
- High stakes are high, relative and absolute performance comparisons affect careers

Summing Up

Rankings can serve as catalysts for better stewardship of water and other resources.

If regulators cannot identify historical trends, determine today's baseline performance, and quantify relative performance across utilities, then as an Indian regulator said, they may as well be writing "pretty poetry".

Conflict Resolution Matrix



(From Mark Jamison)

Informal Survey

1. **Authority**: Is the regulatory agency used to help resolve key issues or is it by-passed? Are you in the middle of "turf wars"?

2. **Facts**: Has benchmarking been used to improve sector performance? Are contracts and targets based on reality?

3. **Values**: Does the regulatory agency help clarify how the targets reflect goals or stated political objectives?

4. **Special Interests**: Have regulatory decisions been inconsistent due to the influence of special interests? Are prices too low or too high? Costs?

Technical Appendix

- Practical Issues (inputs, outputs, explanatory factors)
- Global Productivity Measures
- DEA and SFA

Regulatory benchmarking has benefits:

- **Encourages the operators to be efficient** (both regarding the OPEX and CAPEX);
- Assures a **"fair" recovery of the costs** and enabling a **"fair" rate of return on investment;**
- Improves **information sharing** and **transparency**;

However, conduct studies with care:

- Method cannot be **subjective** nor **discretionary**;

Technical Issues

- **Heterogeneity** problems (comparing “apples with oranges”);
- Not all the kind of costs are known or included in the analysis (e.g. water resource availability in the future);
- The differences between the costs of the different peers might be due to inefficiency, to measurement errors, or to historical factors beyond current management’s control;
- The X Factor, **based on differences of efficiency** between regulated firms, might not capture all the explanatory factors.

Practical issues: Model Specification

Better simple models with robust analysis and consistency checks

than complex models with superficial analyses

Model specification:

- OPEX/CAPEX/TOPEX models;
- Physical or monetary units;
- The degrees of freedom;
- Production/cost models;
- Panel data/cross sectional data;
- Number of comparators/international comparators;
- Adjusting for environment;
- Outliers;
- Quality versus economical issues;

Practical issues: Inputs

What are the inputs of the network utilities?

- OPEX Operating Expenditures
- CAPEX Capital Expenditures (annual outlays on investments)
- CAPITAL ASSETS (cumulative investment less economic depreciation)

or

- TOPEX (Total Costs)

- Is this disaggregation enough? Or should it be divided into staff and other OPEX, recognizing outsourcing and other factors?

↳ Using just OPEX presents problems!... But, how should Capital be measured?

Practical issues: Measuring Capital Assets or CAPEX

- In physical units, as in network length?
- Other assets ? What about the quality or age of assets?
- Might rural utilities be penalized if density ignored?
- But isn't the task easier in monetary units?
- Book value, market value or replacement value?
- What about the historical subsidies and stranded costs: how are they taken into account?

Practical issues: Outputs

The most used outputs are:

☞ Volume of energy/water/… delivered (sold);

☞ Number of customers (connections).

↳ What is the importance of the percentage of industrial customers?

↳ What features of rural utilities influence costs?

↳ Why not consider the network length or the capacity of network length available to the customers?

↳ Is there data on quality of output? What about 'bad' outputs (environmental damages)? What about financial or resource sustainability?

Practical issues: Explanatory Factors

EXAMPLES from the water sector:

- ✓ Weather;
- ✓ Assets' age;
- ✓ % of non-residential customers;
- ✓ Water source;
- ✓ Availability of water resources and its quality;
- ✓ Topography;
- ✓ Peak factor;
- ✓ Customer density;
- ✓ Kind of soil;
- ✓ Local regulations and environment policies;
- ✓ Ownership;
- ✓ Regulation, …

➢ They influence efficiency significantly. A particular variable may not be statistically significant but it can influence softly the utilities efficiency or, in particular, a single utility's ranking can be dramatically affected.

Practical Issues: External Factors

Thus, efficiency is influenced by a set of external factors not controlled by the utility managers, such as:

- ✓ Market structure factors (scale, scope and density economies);
- ✓ Historical factors (past investments interfere with CAPEX/OPEX …);
- ✓ Social factors (% of industrial customers, bigger customers, consuming habits, peak factor/density economies, GDP,…);
- ✓ Environmental factors (weather, …);
- ✓ Regulatory factors (regulation, prices policies, taxes, demand policies, …);
- ✓ Local factors (topography, availability of resources, …).

Global productivity measures: Outputs/Inputs

Total factor productivity (TFP):

$$TFP = \frac{\text{Weighted Sum of Outputs}}{\text{Weighted Sum of Inputs}} = \frac{\sum_{i=1}^{M} a_i \cdot y_i}{\sum_{j=1}^{N} b_j \cdot x_j}$$

- Linear relationship between *inputs* and *outputs*;
- Constant weights for all the elements being compared;
- Different results can be reached according to the composition of weights adopted.

Productivity change measures:

$$TFP_{t,\,t+1} = \frac{TFP_{t+1}}{TFP_t} = \frac{f\left(y_{t+1},\ y_t\right)}{g\left(x_{t+1},\ x_t\right)}$$

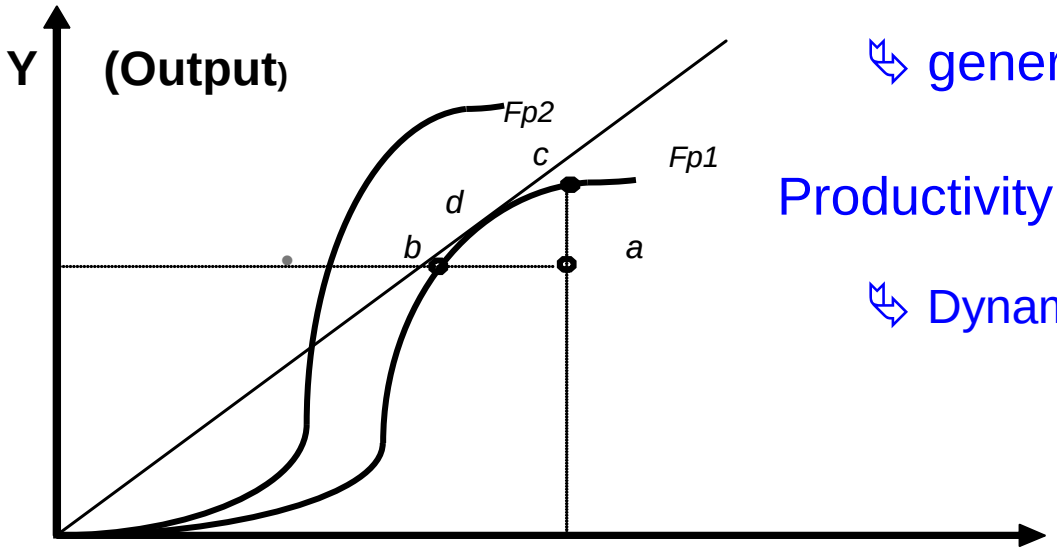
- The inputs and outputs choice and the way they can be aggregated characterize the different indexes than can be built;
- They are non-parametric and non-frontier measures of performance evaluation, called Index Numbers.

Definitions

The concept of efficiency is different from that of productivity;

Productivity is the ratio between the products (outputs) and the factors (inputs) used.

Efficiency analysis involves establishing a standard and determining how close the firm comes to meeting that standard. The operation at different scales or distinct operational environments leads to different productivities.



Efficiency

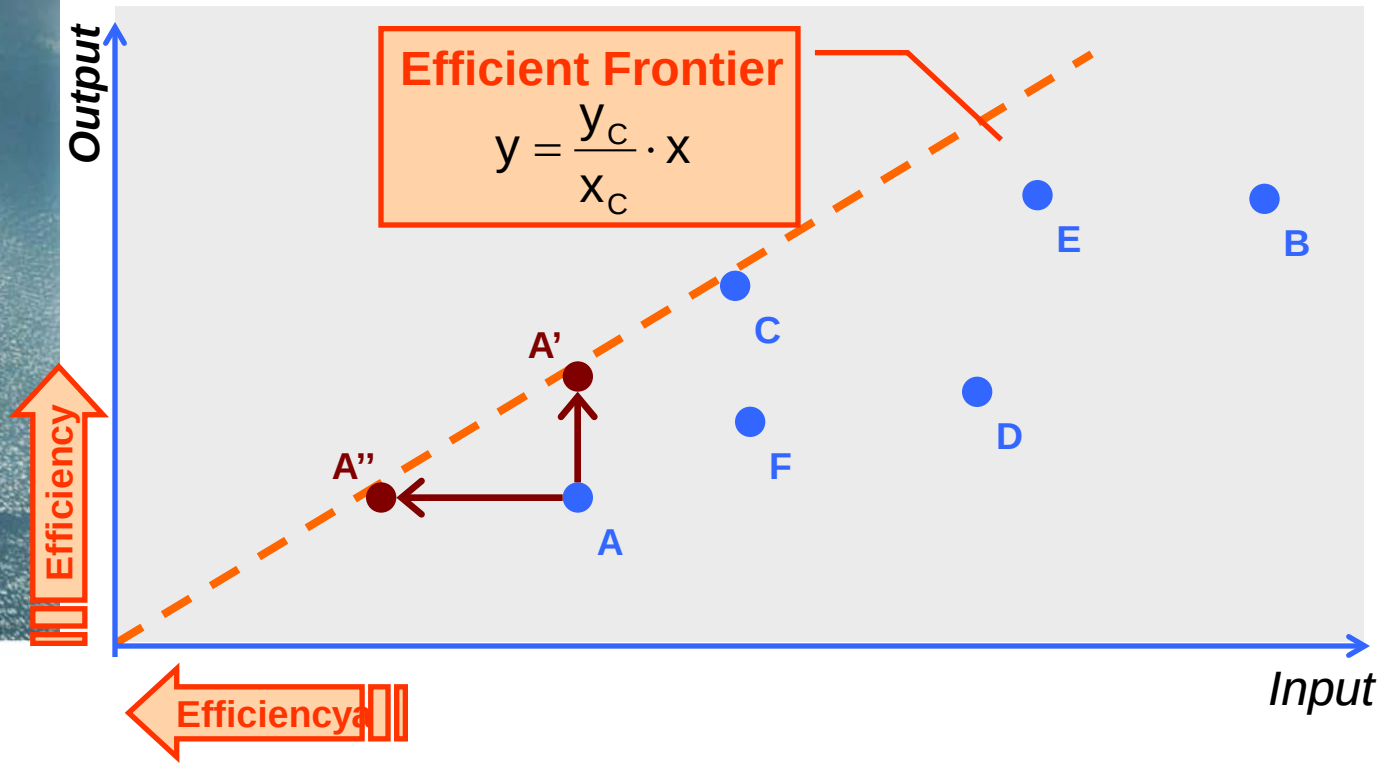
↳ generally Static

Productivity

↳ Dynamic

Data Envelopment Analysis (DEA)

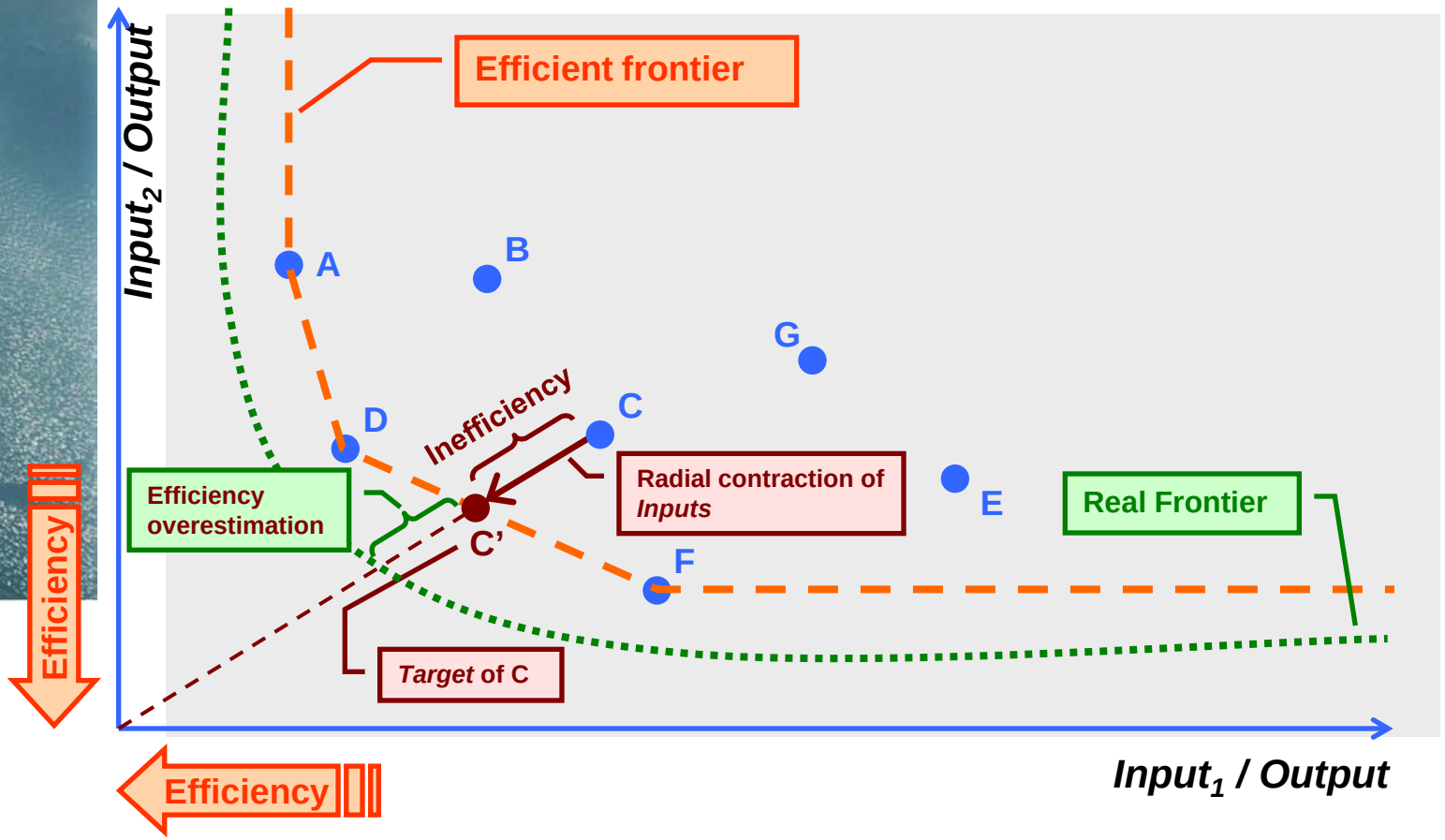
DEA Technique with 1 Input and 1 Output:



- **A** can raise the number of output produced and consume the same quantity of input, until A' (Output orientation);
- **A** can reduce the consume of input, for the same level of production of output, until A'' (Input orientation);
- Segment A'A'' represents the possibilities that **A** has to improve.

Data Envelopment Analysis (DEA)

Example of the CCR model as *Input oriented*

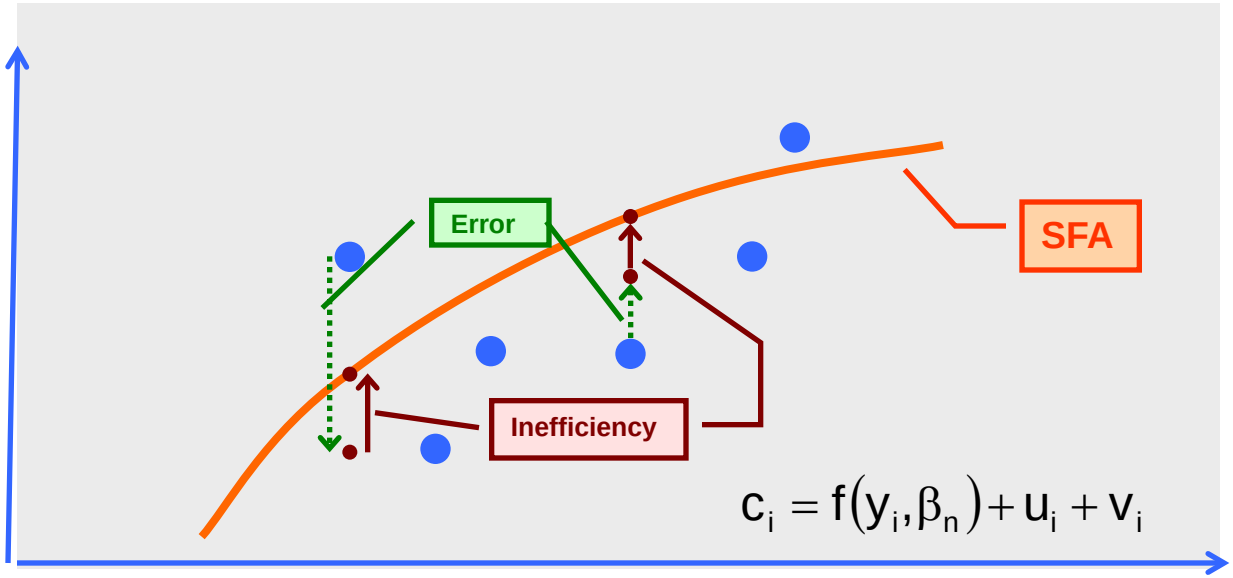


- Operators **D** and **F** are comparison peers of **C**;
- The projection of **C** in the frontier (*target*) is a l**inea**r combination of **D** and **F**.

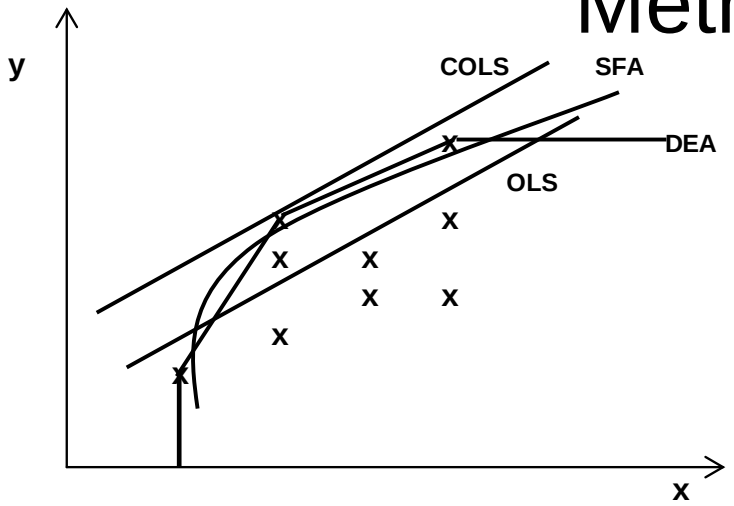
Stochastic frontiers (SFA)

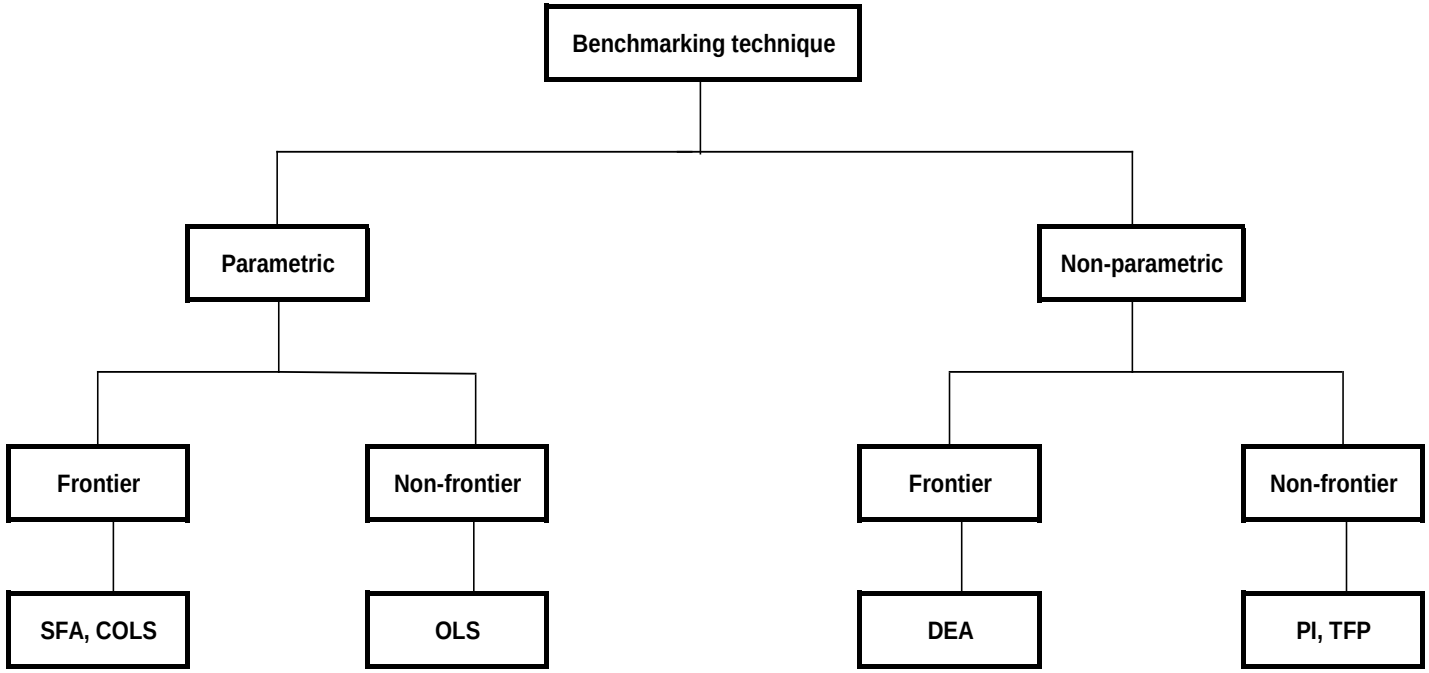
Characterization:

- An alternative methodology involving the estimation of parameters
- The determination of the frontier is based on the maximum likelihood method;
- Differs from other methodologies by separating the statistical error term from the inefficiency term.

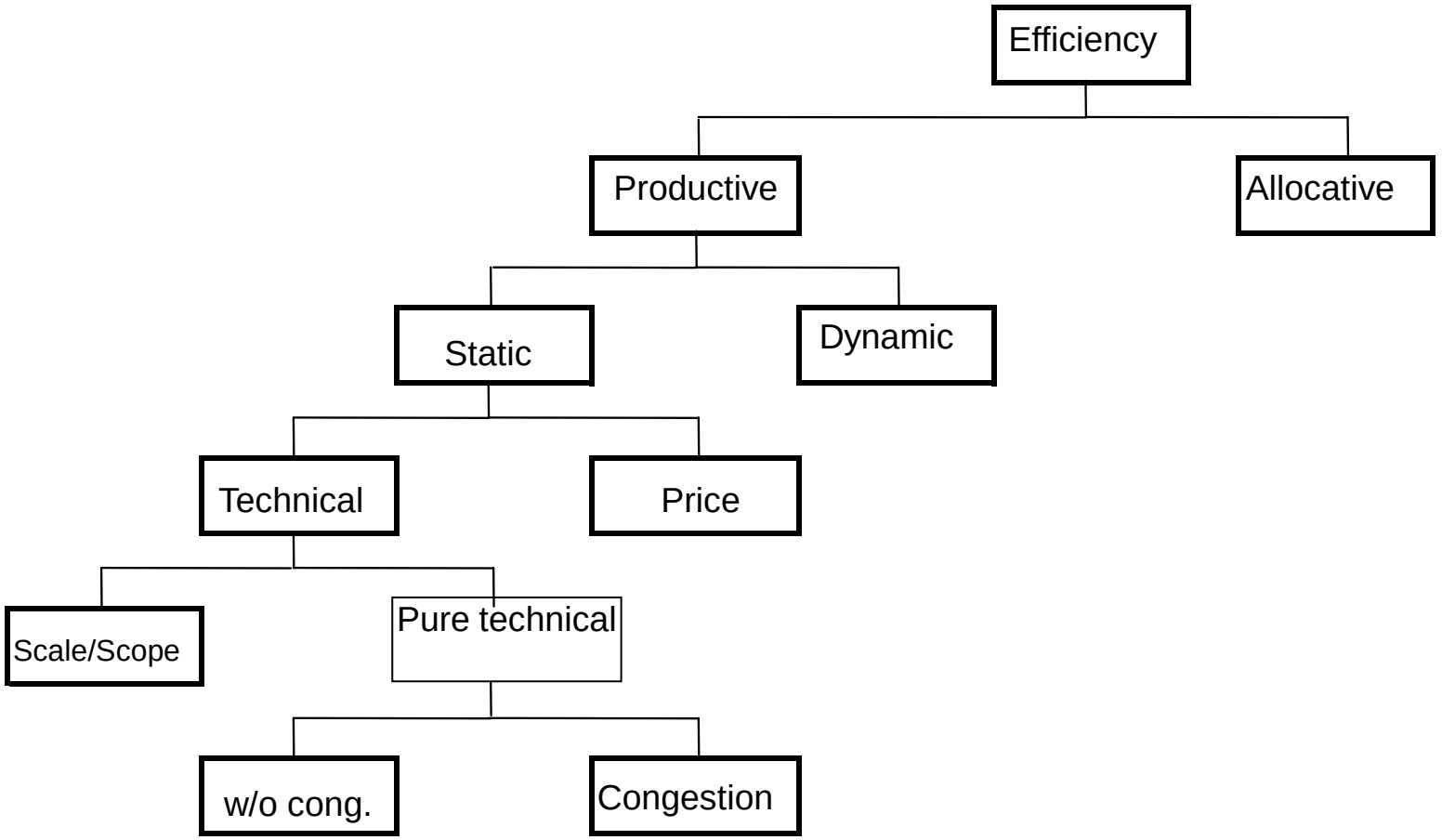


Methods


y
COLS
SFA
DEA
OLS
x


Benchmarking technique
Parametric
Non-parametric
Frontier
Non-frontier
Frontier
Non-frontier
SFA, COLS
OLS
DEA
PI, TFP

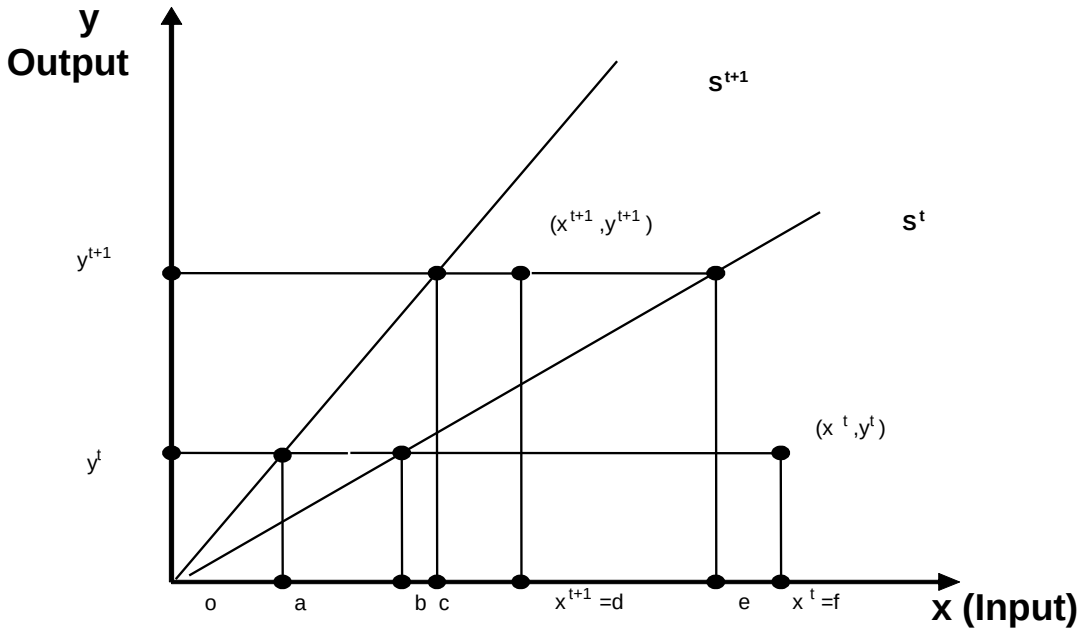
Efficiency Classifications



The problem of X Factor Computation:

X Factor decomposition = Technical change + efficiency change

- Technical change (the efficient frontier shift) associated with the productivity of the sector (or the economy in general);
- Efficiency change is related to movement towards the efficient frontier (the catch-up factor).



Features of Major Metric Techniques

Features	DEA	SFA	COLS	OLS
Specification of the functional form	No	Yes	Yes	Yes
Integration of multiple inputs and outputs	Yes	Difficult	Difficult	Difficult
Identification of best practices	Yes	No	Yes	No
Detail of efficiency measures	High	Low	Low	Low
Statistical inference	Difficult	Yes	Yes	Yes
Adjusting for operational environment	More difficult	Multico-linearity	Multico-linearity	Multico--linearity
Accounting for noise	No	Yes	No	No
Sensitivity to outliers	Very	Sensitive	Very	Little