DIW BERLIN

cepe
Centre for Energy Policy and Economics
Department of Management, Technology and Economics

International Benchmarking of German and Swiss Urban Public Transport Sectors using Panel Data

By Astrid Cullmann, Mehdi Farsi, Massimo Filippini

DIW Berlin and ETH Zürich (CEPE)

INFRADAY 2008, TU-Berlin

Agenda

1. Issues and Motivation
2. Methods
3. Empirical Application
4. Results
5. Conclusions

Literature

Issues and Motivation

- **European countries changed the regulatory approach in their public transport sectors**

1) Competitive Tendering
2) Incentive regulation instruments

→ In order to improve the efficiency and quality

- **Such instruments are usually based on the results obtained from a benchmarking analysis, using production, distance or cost frontier models**

- **Regulators sometimes interested in performing an international benchmarking**

Issues and Motivation

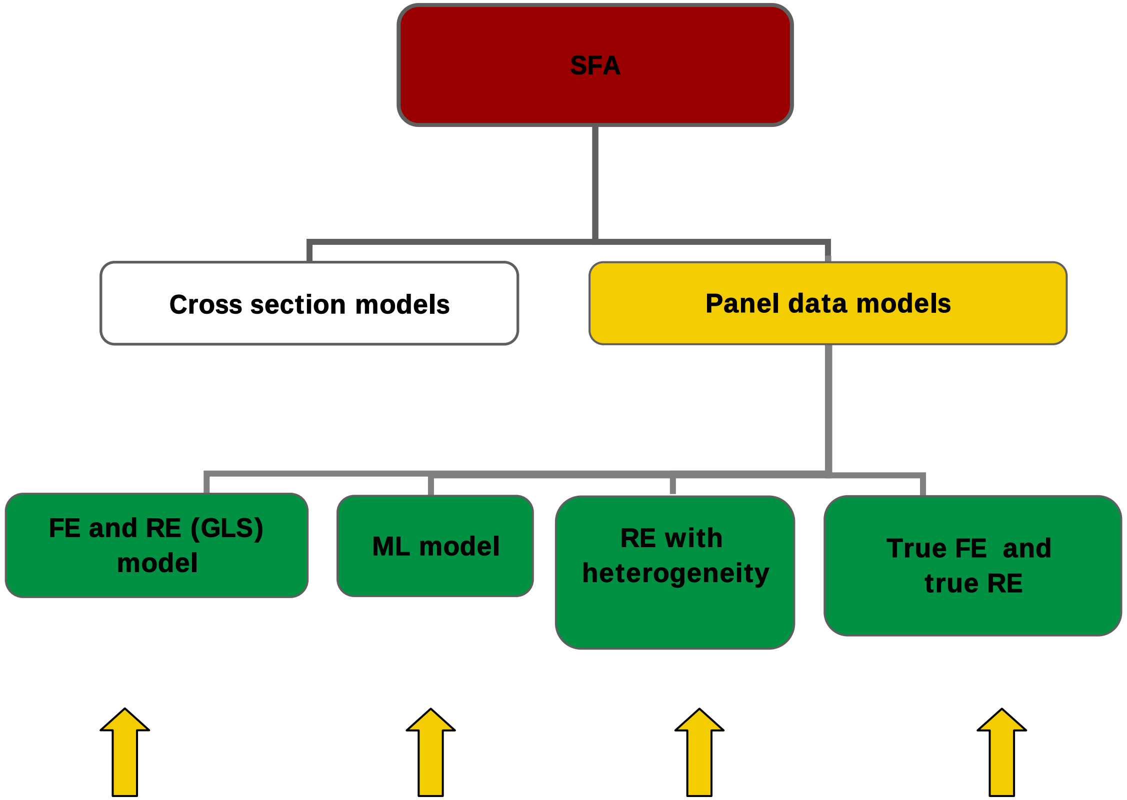
Implication

- **Reliability of efficiency estimates is crucial for an effective implementation**
- **Empirical evidence: estimates are sensitive to the adopted benchmarking approach.**

Problems

- **High level of output heterogeneity (multi-utilities)**
- **Vehicles, shape of networks, organization, coordination, density of stops, different environmental characteristics**
- **Information is not available for all output characteristics**
- **Omitted from the model specifications**
- **Unobserved firm-specific heterogeneity becomes more serious in cross-country, comparative efficiency analyses.**

Panel Data Models and Unobserved Heterogeneity



Schmidt and Sickles (1984)

Cornvell, Schmidt, Sickles (1990)

Pitt and Lee (1981)

Battese and Coelli (1992)

Kumbhakar (1993)

Heshmati and Kumbhakar(1994)

Greene (2005a, b)

Farsi, Filippini, Greene (2005, 2006)

Farsi, Filippini, Kuenzle (2006)

Unobserved firm-specific heterogeneity can be taken into account with conventional fixed or random effects in a panel data model.

Translog Input Distance Function

$$-\ln(x_K) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln y_m \ln y_n + \sum_{k=1}^{K-1} \beta_k \ln\left(\frac{x_k}{x_K}\right)$$
$$+\frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1} \beta_{kl} \ln\left(\frac{x_k}{x_K}\right)\ln\left(\frac{x_l}{x_K}\right) + \sum_{k=1}^{K-1}\sum_{m=1}^{M} \delta_{km} \ln\left(\frac{x_k}{x_K}\right)\ln y_m + \gamma_t T - \ln D_I$$

ln Di → **nonnegative variable which can be associated with technical inefficiency .**

- Given the stochastic error this model can be formulated in the common SFA form with the combined error term

True Random Effects Model

→ **Random-constant frontier model (conventional random-effects model with a skewed stochastic term)**

$$-\ln(x_{Kit}) = \alpha_i + \sum_{m=1}^{M} \alpha_m \ln y_{mit} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln y_{mit} \ln y_{nit} + \sum_{k=1}^{K-1} \beta_k \ln\left(\frac{x_{kit}}{x_{Kit}}\right)$$
$$+ \frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1} \beta_{kl} \ln\left(\frac{x_{kit}}{x_{Kit}}\right)\ln\left(\frac{x_{lit}}{x_{Kit}}\right) + \sum_{k=1}^{K-1}\sum_{m=1}^{M} \delta_{km} \ln\left(\frac{x_{kit}}{x_{Kit}}\right)\ln y_{mit} + \gamma_t T + v_{it} - u_{it}$$

ai **normal *i.i.d.* in random-effects framework**

uit* , *vit **half-normal variable representing inefficiency and a normal random variable that captures the statistical noise.**

Unobserved firm-specific heterogeneity is accounted for by individual effects → might be correlated with the explanatory variables, in which case the estimations might be affected by ‘heterogeneity bias.’

Random Coefficient Frontier Model

Random coefficients for some of the explanatory variables: Variation capture part of the correlation of the random intercept with the corresponding variables.

$$-\ln\left(x_{K_{it}}\right)=\alpha_i+\sum_{m=1}^{M}\alpha_{mi}\ln y_{mit}+\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\alpha_{mn}\ln y_{mit}\ln y_{nit}+\sum_{k=1}^{K-1}\beta_k\ln\left(\frac{x_{kit}}{x_{Kit}}\right)$$

$$+\frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1}\beta_{kl}\ln\left(\frac{x_{kit}}{x_{Kit}}\right)\ln\left(\frac{x_{lit}}{x_{Kit}}\right)+\sum_{k=1}^{K-1}\sum_{m=1}^{M}\delta_{km}\ln\left(\frac{x_{kit}}{x_{Kit}}\right)\ln y_{mit}+\gamma_{it}T+v_{it}-u_{it}$$

Intercept

Two output coefficients

Linear time trend

} **Random variables with a normal distribution.**

Different underlying production technologies

Different scale economies

Company specific technological progress

Model Specification

Variables
Inputs
Number of employees (X1)
Number of streetcars (x2)
Number of buses (X3)
Outputs
Seat-kilometers in streetcars (y1)
Seat-kilometers in buses (y2)
Structural Variable
Size of the operating area (z1)

Unbalanced panel with 56 multi-output local public transport companies

- **49 from Germany** (1994-2006), source: VDV statistics

- **7 from Switzerland** (1991-2003), source: Swiss Federal Statistical Office, annual reports

→ Supply oriented model

Estimation Results (TRE and RCM)

	Model 1 TRE	p-value	Model 2 RCM	p-value
Constant	-0.10	0.21	-0.13	0.00
ln(x2/x1)	0.19	0.00	0.19	0.00
ln(x3/x1)	0.34	0.00	0.31	0.00
ln(y1)	-0.29	0.00	-0.25	0.00
ln(y2)	-0.53	0.00	-0.49	0.00
Trend	0.02	0.00	0.03	0.00
ln(z1)	-0.09	0.00	-0.07	0.00

Standard Deviation for random parameters (a)	Model 2 RCM	p-value
Constant	0.62	0.00
ln(y1)	0.11	0.00
ln(y2)	0.07	0.00
Trend	0.01	0.00

Variation across companies

→ different economies of scale and density

→ different technological changes

RC model can improve the estimates

Efficiency Analysis

Summary Statistics of Efficiency Estimates

	Number of Observation	Mean	Std Dev	Min	Median	Max
Model 1 True Random Effects Model	707	0.905	0.061	0.469	0.921	0.987
Model 2 Random Coefficient Model	707	0.926	0.052	0.585	0.942	0.994

For Comparison Pitt and Lee 1981

Pitt and Lee (1981): mean (0.78) median (0.76))

Battese and Coelli (1992): mean (0.80)

→ Any unobserved, time-invariant, firm-specific heterogeneity is considered as inefficiency.

Correlation and Rank Correlation

Kruskall Wallis Test Statistics

	Model I	Model II
German (616) vs Swiss (91) Companies	0.004	1.101

- **Kruskal-Wallis test → no significant difference between Swiss and German transit companies.**
- **Kruskal-Wallis test for Pitt and Lee Model and Battese and Coelli Model → significant difference between Swiss and German transit companies.**

Conclusions

• **Issue of unobserved heterogeneity, a concern especially for international benchmarking, can be handled using panel data models that can account for stochastic variation in the model's parameters.**

• **The input distance function is a legitimate modeling concept that can be used to estimate the technical efficiency and its variations over time.**

• **When used in the random-coefficient framework, the model allows a better control for the latent variations across companies regarding technological characteristics as well as temporal changes and technical progress.**

Back up

Summary Statistics for Germany

Variable	Obs	Mean	Std. Dev.	Min	Max
Germany					
Inhabitants	616	366709	295151	40800	1642000
Number of employees	616	978	893	30	3996
Network length tram in Km	616	49	41	3	155
Network length bus in Km	616	465	364	5	2653
Number trams	616	118	124	2	755
Number busses	616	135	100	2	470
Tram km in 1000 Mm	616	5664	6412	61	34363
Bus km in 1000 Km	616	7211	5709	86	28519
Seat km tram in 1000 Km	616	964943	1200087	5000	6187000
Seat km bus in 1000 Km	616	584293	463609	4000	2303000
Area in km2	616	171	77	21	405

Summary Statistics for Switzerland

Variable	Obs	Mean	Std. Dev.	Min	Max
Switzerland					
Inhabitants	91	285215	117492	76381	421802
Number of employees	91	953	711	76	2798
Network length tram in Km	91	32	30	8	110
Network length bus in Km	91	139	94	42	362
Number trams	91	128	136	12	432
Number busses	91	167	105	30	314
Tram km in 1000 Mm	91	6111	6916	398	20518
Bus km in 1000 Km	91	8121	5729	1525	18438
Seat km tram in 1000 Km	91	847835	923549	37387	2926006
Seat km bus in 1000 Km	91	974580	722588	121443	2283553
Area in km2	91	169	63	90	275

Distance Function

- **Capture multi-output context without specification of a behavioral assumption**
- **Definition on the output set with input minimization:**

$$d_i(x, y) = \max\left\{\rho : (x/\rho) \in L(y)\right\}$$

- **Restrictions for homogeneity of degree +1 in inputs and symmetry:**
- → **Imposing homogeneity by dividing the specification of the translog function by an arbitrarily chosen input**

$$\sum_{k=1}^{K} \beta_k = 1, \ \sum_{l=1}^{K} \beta_{kl} = 0 \ \text{ and } \ \sum_{k=1}^{K} \delta_{km} = 0$$

$$d_i(x / X_K, y) = d_i(x, y) / X_K$$

Econometric Specification

	Model 1 **True RE Greene 2004,2005**	***Model 2*** **Random Coefficient Model**
Firm-specific component α_i	$\alpha_i \sim N(0,\sigma_\alpha^2)$	$\alpha_i \sim N(0,\sigma_\alpha^2)$ $\beta_i \sim N(\mu_\beta,\sigma_\beta^2)$
Random Error ε_{it}	$\varepsilon_{it} = v_{it} - u_{it}$ $u_{it} \sim N^+(0,\sigma_u^2)$ $v_{it} \sim N(0,\sigma_v^2)$	$\varepsilon_{it} = v_{it} - u_{it}$ $u_{it} \sim N^+(0,\sigma_u^2)$ $v_{it} \sim N(0,\sigma_v^2)$

Capturing unobserved Heterogeneity