Productivity and Efficiency Analysis

3) Stochastic frontier analysis (SFA)

c) Application of SFA

Timo Kuosmanen

Aalto University School of Business

https://people.aalto.fi/timo.kuosmanen

Cost frontier model (2012-2015)

$$\ln x = \ln C(y_1, y_2, y_3) + \delta z + u + v$$

- x is the total cost
- C is the frontier cost function
- y_1 is the energy transmission (GWh of 0.4 kV equivalents)
- y_2 is the total length of the network (km)
- y_3 is the number of customers connected to the network
- z is the proportion of underground cables (%)
- u is an asymmetric inefficiency term
- v is a random noise term

Cost frontier model (2012-2015)

$$\ln x = \alpha + \sum_{k=1}^{3} \beta_k \ln y_k + \delta z + u + v$$

- *x* is the total cost
- β_k is the coefficient of output k = 1,2,3
- y_1 is the energy transmission (GWh of 0.4 kV equivalents)
- y_2 is the total length of the network (km)
- y_3 is the number of customers connected to the network
- *z* is the proportion of underground cables (%)
- *u* is an asymmetric inefficiency term
- *v* is a random noise term

Linear regression (Cobb-Douglas)

$R^2 = 0.983$

	Coeff.	*St.Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2,508	0,117	21,495	0,000	2,276	2,740
lnEnergy	0,577	0,029	20,164	0,000	0,520	0,634
lnLength	0,526	0,053	10,011	0,000	0,421	0,630
lnCustomer	-0,142	0,033	-4,258	0,000	-0,208	-0,076
Underground	0,418	0,136	3,071	0,003	0,147	0,689

Corrected OLS

$$Eff_i = \exp(e_i - max_h \{e_h\})$$

Average COLS efficiency: 67.5%

Standard deviation: 10.7%

Efficiency scores of the first 8 firms:

65,6 %
71,0 %
61,5 %
74,4 %
85,7 %
59,6 %
69,0 %
71,8 %

Linear regression (translog)

$R^2 = 0.990$ Wrong skewness: -0.318

	Coeff.	*St.Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4,271	0,789	5,411	0,000	2,700	5,843
lnEnergy	0,770	0,175	4,392	0,000	0,421	1,119
lnLength	0,021	0,291	0,073	0,942	-0,558	0,600
lnCustomer	-0,286	0,141	-2,032	0,046	-0,566	-0,006
lnEnergy^2	0,075	0,037	2,033	0,045	0,002	0,149
lnLength^2	0,349	0,075	4,656	0,000	0,200	0,498
lnCustomer^2	0,049	0,068	0,721	0,473	-0,087	0,186
lnEne*lnLen	-0,447	0,136	-3,301	0,001	-0,717	-0,178
lnEne*lnCust	0,196	0,086	2,283	0,025	0,025	0,366
lnLen*lnCus	-0,192	0,152	-1,265	0,210	-0,494	0,110
Underground	0,290	0,121	2,405	0,019	0,050	0,531

SFA: ML estimation in R (normal/half-normal)

```
## Error Components Frontier (see Battese & Coelli 1992)
## Inefficiency increases the endogenous variable (as in a cost function)
## The dependent variable is logged
## iteration failed
##
## final maximum likelihood estimates
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.644531   1.000000  2.6445  0.00818 **
## log(Energy)      0.634358   1.000000  0.6344  0.52585
## log(Length)      0.388445   1.000000  0.3884  0.69769
## log(Customers) -0.069658   1.000000 -0.0697  0.94447
## sigmaSq          0.028295   1.000000  0.0283  0.97743
## gamma            0.050000   1.000000  0.0500  0.96012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## log likelihood value: 33.7975
##
## cross-sectional data
## total number of observations = 89
##
## mean efficiency: 0.9706829
```

SFA: ML estimation in R (normal/half-normal)

$$\sigma^2 = \sigma_u^2 + \sigma_v^2,\ \lambda = \sigma_u/\sigma_v,$$

```
##   (Intercept)    log(Energy)    log(Length) log(Customers)        sigmaSq
##   2.644531475    0.634357818    0.388445347   -0.069657530    0.028295350
##         gamma       sigmaSqU       sigmaSqV          sigma         sigmaU
##   0.049999853    0.001414763    0.026880587    0.168212219    0.037613340
##        sigmaV       lambdaSq         lambda           varU            sdU
##   0.163953003    0.052631416    0.229415378    0.000514097    0.022673708
##      gammaVar
##   0.018766306
```

Firm-specific efficiency estimates (JLMS)

```
##     efficiency
## 1   0.9720703
## 2   0.9697788
## 3   0.9693864
## 4   0.9684544
## 5   0.9654828
## 6   0.9728781
## 7   0.9701348
## 8   0.9705680
## 9   0.9661850
## 10  0.9735195
## 11  0.9688796
## 12  0.9708250
## 13  0.9685312
## 14  0.9686269
## 15  0.9617849
## 16  0.9754922
## 17  0.9694592
## 18  0.9704523
## 19  0.9702723
## 20  0.9726286
## 21  0.9755673
## 22  0.9765358
## 23  0.9722030
## 24  0.9695855
## 25  0.9698858
## 26  0.9715944
## 27  0.9681850
## 28  0.9688548
## 29  0.9689977
## 30  0.9706774
## 31  0.9730620
## 32  0.9706478
## 33  0.9730504
## 34  0.9638306
## 35  0.9705309
## 36  0.9657774
## 37  0.9721934
## 38  0.9743651
## 39  0.9735792
## 40  0.9712379
## 41  0.9672871
## 42  0.9677890
## 43  0.9691040
## 44  0.9737874
## 45  0.9744659
## 46  0.9659315
## 47  0.9701100
## 48  0.9670671
## 49  0.9730074
## 50  0.9733094
## 51  0.9685789
## 52  0.9729340
## 53  0.9701759
## 54  0.9690873
## 55  0.9725704
## 56  0.9731197
## 57  0.9757323
## 58  0.9752666
## 59  0.9754625
## 60  0.9683200
```

Advantages of SFA

- Intimate connection to regression analysis: natural approach to applied economists

- Probabilistic modeling of inefficiency and noise
  - Hypothesis testing, confidence intervals

- Easily extends to panel data

Disadvantages of SFA

- Untestable parametric assumptions:
  - Functional form of the frontier (CD, translog)
  - Inefficiency distribution (half-normal, exponential,…)
  - Normally distributed noise
- Difficulty to impose shape constraints (monotonicity, convexity)
- Main weakness: multiple outputs

Next lesson

4) Unified approach: StoNED