METHODOLOGY AND APPLICATIONS OF STOCHASTIC FRONTIER ANALYSIS

Andrea Furková

STRUCTURE OF THE PRESENTATION

Part 1

Theory: Illustration the basics of Stochastic Frontier Analysis (SFA)

- Concept of efficiency
- Estimation
- Identification of sources of inefficiency

Part 2

An applications of SFA

- Theory usually presents the producers as successful optimizers. They maximize production, minimize cost, and maximize profits.

- Conventional econometric techniques build on this base to estimate production/cost/profit function parameters using regression techniques where deviations of observed choices from optimal ones are modeled as statistical noise.

Econometric estimation techniques should allow for the fact that deviations of observed choices from optimal ones are due to two factors:

- failure to optimize i.e., inefficiency
- due to random shocks.

Stochastic Frontier Analysis is one such technique to model producer behavior.

Methods based on efficient frontier

- Based on benchmarking, that is, a unit's performance is compared with a reference performance (so-called efficient frontier).

- Unit's inefficiency can results from technological deficiencies (technical inefficiency) or non-optimal allocation of resources into production (allocative inefficiency). Both technical and allocative inefficiencies are included in cost (economic) inefficiency.

Generally, there are two families of methods based on efficient frontier:

- **Non-parametric methods**, like Data Envelopment Analysis (DEA) or Free Disposal Hull (FDH). These methods originate from operations research and use linear programming to calculate an efficient deterministic frontier against which units are compared.

- **Parametric methods**, like Stochastic Frontier Analysis (SFA), Thick Frontier Approach (TFA) and Distribution Free Approach (DFA). Econometric theory is used to estimate pre-specified functional form and inefficiency is modeled as an additional stochastic term.

Stochastic Frontier Analysis – time invariant efficiency

- Based on econometric theory and pre-specified functional form is estimated and inefficiency is modeled as an additional stochastic term.
- **The Stochastic frontier production function model** (single Cobb-Douglas form for panel data):

$$\ln y_{it} = \beta_0 + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i \qquad i = 1,\ldots,N \quad t = 1,\ldots,T$$

$$\varepsilon_{it} = v_{it} - u_i \qquad\qquad u_i \geq 0$$

y_{it} - observed output quantities of the *i*-th unit in year *t*,

x_{it} - observed inputs quantities of the *i*-th unit in year *t*,

u_i - non negative time-invariant random variables capturing time-invariant technical inefficiency,

v_{it} - random variables of *i*-th unit in year *t* reflecting effect of statistical noise

- Frontier cost function: identifies the minimum costs at a given output level, input prices and existing production technology
- **Stochastic frontier cost function model** (single Cobb-Douglas form for panel data):

$$\ln C_{it} = \beta_0 + \beta_y \ln y_{it} + \sum_n \beta_n \ln w_{nit} + v_{it} + u_i \qquad u_i \geq 0$$

$$\varepsilon_{it} = v_{it} + u_i \qquad i = 1, \ldots, N \quad t = 1, \ldots, T$$

C_{it} - observed total costs of the *i*-th unit in year *t*,

y_{it} - a vector of outputs of the *i*-th unit in year *t*,

w_{it} - an input price vector of the *i*-th unit in year *t*,

u_i - time-invariant cost inefficiency,

v_{it} - random variables of *i*-th unit in year *t* reflecting effect of statistical noise

SPECIFICATION OF THE MODEL

1. Deterministic kernel of the model

   - Cobb-Douglas (in log form)
   - Translog (a flexible functional form)

2. Estimation of the model

   - Maximum likelihood estimation (ML)
   - Generalised Least Squares method (GLS)
   - Method of moments (MM)

3. Calculating unit´s specific efficiency

Maximum likelihood estimation (ML)

- Distribution assumptions:

  $u_i \sim \text{iidN}^+(0,\sigma_u^2)$ or $u_i \sim \text{iidN}^+(\mu,\sigma_u^2)$, $v_{it} \sim \text{iidN}(0,\sigma_v^2)$

- Maximization of the log likelihood function
- The individual estimates of the technical (cost) inefficiency:

  JLMS decomposition - the conditional distribution of u_i, conditional mean $E\left(u_i \middle| \varepsilon_{it}\right)$ or conditional modus $M\left(u_i \middle| \varepsilon_{it}\right)$ of this distribution can be used as a point estimator for u_i

- The individual estimates of the technical (cost) efficiency:

$$TE_i = \exp\left\{-E\left(u_i \mid \varepsilon_{it}\right)\right\}$$

Analyzing Efficiency Behavior

Two questions:

- What is the behavior of efficiencies over time? Are they increasing, decreasing or constant?
- What explains the variations in inefficiencies among units and across time?

Time behavior of inefficiencies

Assumption:

$$u_{it} = u_i + \beta_t$$

Model:

$$\ln y_{it} = \beta_{0t} + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it} \quad i = 1,\ldots,N \quad t = 1,\ldots,T$$

- Cornwell, Schmidt and Sickles, Lee and Schmidt, Kumbhakar, Battese and Coelli

Battese and Coelli proposed a simple model that can be used to estimate the time behavior of inefficiencies:

$$u_{it} = \exp\{-\eta(t-T)\}u_i$$

where $u_i \sim \mathrm{iidN}^+(\mu, \sigma_u^2)$ and η is a parameter to be estimated.

- Inefficiencies in periods prior to T depend on the parameter η.
- Efficiency behavior is monotonic and ordering of units in terms of inefficiencies time-invariant.
- Good for understanding aggregative behavior.

Explaining efficiency

Certain factors influence the environment in which production takes place e.g. degree of competitiveness, input and output quality, network characteristics, ownership form, regulation etc.

Two ways to handle them

- Include them as variables in the production process as control variables. Using this interpretation, these variables influence the structure of the technology by which conventional inputs are converted into outputs, but not efficiency.
- Associate variation in estimated efficiency with variation in the exogenous variables.

Battese a Coelli

Inefficiencies are assumed to be a function of a set of explanatory variables associated with inefficiency of units over time :

$$u_{it} = \mathbf{z}_{it}^T \delta + w_{it}$$

where

$\mathbf{z}_{it}$ - vector of variables which may influence the efficiency of units

δ - vector of unknown parameters to be estimated

w_{it} ~ iid N (0, σ_w^2) – random variables reflecting effect of statistical noise

u_{it} ~ iid $N^+(\mathbf{z}_{it}^T \delta, \sigma_u^2)$

EMPIRICAL APPLICATIONS

1. REGULATION OF DISTRIBUTION UTILITIES

- Benchmarking analysis can be used by regulator as an additional instrument to establish a larger informational basis for more effective price cap regulation
- Incentive price–cap regulation (X – efficiency factor)

Regulation of Slovak and Czech electricity distribution utilities

- Based on incentive price - cap regulation
- Main issue: How to set efficiency factor X, i.e. cost efficiency prediction

- **Data**: Panel data set (55 observations) for 3 Slovak and 8 Czech electricity distribution utilities over the 2000 – 2004 period

Cost function specifications (Cobb-Douglas form) :

Time invariant cost efficiency (estimation method ML, GLS):

$$\ln\left(\frac{C}{P_P}\right)_{it} = \beta_0 + \beta_Y \ln Y_{it} + \beta_K \ln\left(\frac{P_K}{P_P}\right)_{it} + \beta_L \ln\left(\frac{P_L}{P_P}\right)_{it} + \beta_{CUD} \ln CUD_{it} + v_{it} + u_i$$

$$i = 1,\dots,N \quad t = 1,\dots,T$$

Time varying cost efficiency (estimation method ML):

$$\ln\left(\frac{C}{P_P}\right)_{it} = \beta_0 + \beta_Y \ln Y_{it} + \beta_K \ln\left(\frac{P_K}{P_P}\right)_{it} + \beta_L \ln\left(\frac{P_L}{P_P}\right)_{it} + \beta_{CUD} \ln CUD_{it} + v_{it} + u_{it}$$

$$i = 1,\dots,N \quad t = 1,\dots,T$$

where *C* represents total costs, *Y* is the output, *PK*, *PL*, *PP* are the prices of capital, labor and input purchased energy respectively, *CUD* is customer density

Input prices

- labor price (*PL*) – the average annual salary of utility´ employees
- capital price (*PK*) – the ratio of capital expenses to the total installed capacity of the utility's transformers in MVA
- purchased energy price (*PP*) – average price of purchased energy from generator

Output variable:

- total output (*Y*) - measured as the total number of delivered electricity in MWh

2. COST EFFICIENCY ESTIMATION OF THE BANKING SECTOR

- The efficiency measuring and relative efficiency comparison of banks are crucial questions for analysts as well as for economic policy creators.

Methodology for analyzing banking efficiency

- **Financial ratios -** the standard technique of banking efficiency measuring

  Advantage: simplicity of understanding.

  Disadvantage: they fail to consider the multidimensional input and output process, and are unable to identify the best performers in a group of units.
- **Methods based on efficient frontier**

Model specification and Data

Data: unbalanced panel data set (66 observations) for 9 Slovak banks over the 2000 – 2007 period:

- Tatra banka, a.s., Všeobecná úverová banka, a.s., Slovenská sporiteľňa, a.s., Dexia banka Slovensko, a.s., OTP banka Slovensko, a.s., Istrobanka, a.s., Poštová banka, a.s., Unibanka, a.s., Ľudová banka, a.s.

Cost function specification (Cobb-Douglas form)

Time invariant cost efficiency (estimation method ML, GLS):

$$\ln\left(\frac{C}{P_F}\right)_{it} = \beta_0 + \beta_L \ln\left(\frac{P_L}{P_F}\right)_{it} + \beta_K \ln\left(\frac{P_K}{P_F}\right)_{it} + \beta_U \ln U_{it} + \beta_V \ln V_{it} + \beta_Z \ln Z_{it} + \beta_T T + v_{it} + u_i$$

$$i = 1,\ldots,N \quad t = 1,\ldots,T$$

Time varying cost efficiency (estimation method ML):

$$\ln\left(\frac{C}{P_F}\right)_{it} = \beta_0 + \beta_L \ln\left(\frac{P_L}{P_F}\right)_{it} + \beta_K \ln\left(\frac{P_K}{P_F}\right)_{it} + \beta_U \ln U_{it} + \beta_V \ln V_{it} + \beta_T T + v_{it} + u_{it}$$

where

$$u_{it} = \delta_0 + \delta_1 Z_{it} + w_{it} \qquad i = 1,\ldots,N \quad t = 1,\ldots,T$$

C represents total costs, *PK*, *PL*, *PF* are the prices of capital, labor and funds respectively. Three inputs (physical capital, labor, funds) are used to produce two outputs total loans *U* and total deposits *V* and both models include additional variables *Z* – equity capital and *T* – trend.

- total costs (*C*) - the sum of total expenses
- price of capital (*PK*) - the depreciation over fixed assets
- price of labor (*PL*) - the ratio of personnel expenses over total assets
- price of funds (*PF*) - the ratio of interest expenses over the sum of deposits
- equity capital (*Z*) - the amount of bank equity that reflects the size of banking operations

3. ANALYSIS OF REGIONAL COMPETITIVENESS

- Last few years economic policy making and research have shown increasing interest for regional competitiveness evaluation (economic efficiency of regions represents the basis of economic success for micro-economic level and also the competitiveness of the country)

Regional competitiveness evaluation methods:

Non existence of unique methodology for competitiveness evaluation

- **Specific economic indicators of efficiency**

  Indicator compares a concrete level of the value in the region with respect to its total level in the country
- **Methods based on efficient frontier**

Model specification and Data

The competitiveness of regions is compared through the estimated levels of technical efficiency as the efficiency we perceived as the “mirror” of the competitiveness.

Data: balanced panel data set (280 observations) for 35 V4 (Visegrad four countries) NUTS2 (Nomenclature of Units for Territorial Statistics) regions observed over a period from 2001 to 2008 **:**

Slovakia - 4 NUTS2 regions

Czech Republic - 8 NUTS2 regions

Hungary - 7 NUTS2 regions

Poland -16 NUTS2 regions

Production function specifications (Cobb-Douglas form)

Time invariant technical efficiency (estimation method ML, GLS):

$$\ln(GDP_{it}) = \beta_0 + \beta_1 \ln(ER_{it}) + \beta_2 (\ln GFCF_{it}) + \beta_3 (\ln NI_{it}) + v_{it} - u_i$$

$$i = 1,\dots,N \quad t = 1,\dots,T$$

Time varying technical efficiency (estimation method ML):

$$\ln(GDP_{it}) = \beta_0 + \beta_1 (\ln ER_{it}) + \beta_2 \ln(GFCF_{it}) + \beta_3 (\ln NI_{it}) + v_{it} - u_{it}$$

where

$$i = 1,\dots,N \quad t = 1,\dots,T$$

$$u_{it} = \exp\{-\eta (t - T)\} u_i$$

Inputs:

- **Employment Rate** - *ER* (annually in %) - overall performance of the regional economy affects the number of people employed in various sectors
- **Gross Fixed Capital Formation** – *GFCF* (in % GDP) - efficiency in our model should demonstrate the ability of the regions to transform its capital for its further development.
- **Net Disposable Income of Households** - *NI* (per capita) - in terms of competitiveness the disposable income plays an important role, it directly reflects the purchasing power of the region.

Output:

- **Gross Domestic Product** – *GDP* (in purchasing power parity standards per capita) - the most important macroeconomic indicator.

4. MACROECONOMIC PERFORMANCE EVALUATION

- **Economic growth analysis of transition countries**

  Centrally planned economies are characterized by low economic efficiency and low productivity growth. Main issue: Evaluation the transition to market based economy.

- **Foreign Direct Investment and economic growth relationship**

  Conflicting predictions concerning the economic growth – FDI relation. According to the theory and empirical literature economic growth may induce FDI inflow, and FDI may also stimulate economic growth.

Model specification and Data

Data: unbalanced panel data set (266 observations) for 33 OECD countries observed over a period from 2002 to 2010

Production function specification (translog form)

Time varying technical efficiency (estimation method ML):

$$\ln Y_{it} = \beta_0 + \beta_1 \ln C_{it} + \beta_2 \ln L_{it} + (1/2)\left[\begin{array}{l}\beta_{11}(\ln C_{it})^2 + \beta_{22}(\ln L_{it})^2 \\ + \beta_{12}(\ln C_{it})(\ln L_{it})\end{array}\right] +$$

$$+ \beta_{13}(\ln C_{it})t + \beta_{23}(\ln L_{it})t + \beta_{33}t^2 + v_{it} - u_{it} \qquad i = 1,\ldots,N \quad t = 1,\ldots,T$$

where

$$u_{it} = \delta_1 \ln FDI_{it} + \delta_2 \ln Corruption_{it} + \delta_3 \ln Openness_{it} + w_{it}$$

Factor Inputs:

- Capital (*C*) expressed as Gross Capital Formation (in constant 2000 USD millions)
- Labor (*L*) labor defined as civilian labor force (in thousands)
- Interaction terms of explanatory variables
- Linear and non-linear time trends

Inefficiency variables:

- *FDI* (in USD millions)
- *CPI* - The Corruption Perceptions Index (*Corruption*) chosen to control for institutional inefficiency, index varies from 0 (highly corrupt) to 10 (highly clean)
- *Openness* - country's trade (sum of exports and imports in USD millions)

Output variable:

- Real GDP (*Y*) (in of millions of USD)

USEFULNESS OF SFA

- SFA produces efficiency estimates or efficiency scores of individual units. Thus one can identify those who need intervention and corrective measures.

- SFA provides a powerful tool for examining effects of intervention. For example, has efficiency of the banks changed after deregulation? Has this change varied across ownership groups?

- Since efficiency scores vary across units, they can be related to unit´s characteristics like size, ownership, location, etc. Thus one can identify source of inefficiency.