Firm Heterogeneity and its Implications for Efficiency Measurement

Antonio Álvarez
University of Oviedo
&
European Centre for Soft Computing

Firm heterogeneity (I)

- Definition
  - Characteristics of the individuals (firms, regions, persons,…) that are not measured in the sample
    - Unobserved heterogeneity
- Examples
  - Input quality in production functions
    - Genetic level of the herds
    - Land fragmentation
    - Management

Firm heterogeneity (II)

- Consequences of unobserved heterogeneity
  - If not accounted for, may cause biased estimates
    - Griliches (1957)
  - Difficult to separate from inefficiency
    - Stigler (1966)

Firm heterogeneity (III)

- What about heterogeneity in the stochastic part?
  - One may believe that there is heterogeneity in the mean or the variance of U_{it}

$$u_{it} \approx N^{+}(\mu_i, \sigma_i^2)$$

  - Solution: make them a function of observed variables

$$\mu_i = \mu \exp(q_{it}\delta) \quad \sigma_i = \sigma_u \exp(q_{it}\delta)$$

    - Alvarez, Amsler, Orea and Schmidt (2006)

Review of Relevant Literature

Literature (I)

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics 126 (2005) 269–303

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Reconsidering heterogeneity in panel data estimators of the stochastic frontier model

William Greene*

Department of Economics, Stern School of Business, New York University, 44 West 4th Street, Mec 7-80, New York, NY 10012, USA

Available online 1 July 2004

"This paper examines several extensions of the stochastic frontier that account for unmeasured heterogeneity as well as firm inefficiency"

Literature (II)

Journal of Productivity Analysis, 21, 277–296, 2004

Estimation of Technical Inefficiencies with Heterogeneous Technologies

HO-CHUAN (RIVER) HUANG river@mail.tku.edu.tw
Department of Banking and Finance, Tamkang University, Tamsui, Taipei, Taiwan 251

"The paper proposes a SF model with random coefficients to distinguish technical inefficiency from technological differences across firms"

Literature (III)

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. 17: 127–147 (2002)
Published online 4 March 2002 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.637

STOCHASTIC FRONTIER MODELS WITH RANDOM COEFFICIENTS

EFTHYMIOS G. TSIONAS*

Department of Economics, Athens University of Economics and Business, 104 34 Athens, Greece

"The paper proposes a model to separate technical inefficiency from technological differences across firms"

Literature (IV)

Empirical Economics (2004) 29:169–183
DOI 10.1007/s00181-003-0184-2

EMPIRICAL ECONOMICS

Efficiency measurement using a latent class stochastic frontier model

Luis Orea[1], Subal C. Kumbhakar[2]

[1] Department of Economics, University of Oviedo, Avda. Del Cristo, s/n 33071 Oviedo, Spain (e-mail: lorea@correo.uniovi.es)
[2] Department of Economics, State University of New York, Binghamton, NY 13902-6000, USA (e-mail: kkar@binghamton.edu)

"Unobserved differences in technologies may be inappropriately labeled as inefficiency if variations in technology are not taken into account"

Literature (V)

Economics Letters 36 (1991) 43–48
North-Holland

Estimation of technical inefficiency in panel data models with firm- and time-specific effects

Subal C. Kumbhakar
University of Texas, Austin, TX 78712, USA

Received 30 July 1990
Accepted 10 January 1991

This paper uses a panel data model and develops methods to estimate production function parameters and technical inefficiency. The model distinguishes technical inefficiency from firm- and time-specific effects. Determinants of technical inefficiency is introduced by allowing its mean to be a function of exogenous variables. Both random and fixed effects treatment of firm- and time-specific effects are considered in estimation.

Summary of literature

- Many estimation techniques to separate firm heterogeneity and inefficiency
  - Stochastic Frontiers with Fixed Effects
    - Heshmati and Kumbhakar (1994)
    - Greene (2005)
  - Random Coefficient Models
    - Tsionas (2002), Huang (2004)
    - Alvarez, Arias and Greene (2005)
  - Latent Class Models
    - Orea and Kumbhakar (2004)
  - Local Maximum Likelihood
    - Kumbhakar, Simar, Park and Tsionas (2007)

Objectives of the Talk

- Propose a theoretical framework to understand the relationship between unobserved heterogeneity and inefficiency
- Review some techniques to deal with unobserved heterogeneity
- Study the implications for efficiency analysis

Theoretical Framework

Modelling firm heterogeneity

- Starting point: there is an unmeasured variable (Z)

- Two possible effects:

$$\ln y_{it} = \alpha + \beta \ln x_{it} + \gamma z_{it} + \gamma_x z_{it} \ln x_{it} + v_{it}$$

$$v_{it} \approx N(0, \sigma_v^2)$$

- If $Z_{it}=Z_i$, the model becomes

$$\ln y_{it} = \alpha_i + \beta \ln x_{it} + \gamma z_i + \gamma_x z_i \ln x_{it} + v_{it}$$

Special case: No interactions

- If $\gamma_x = 0$ the unobserved heterogeneity can be modelled as an individual effect

  - Fixed Effects

$$\ln y_{it} = \alpha_i + \beta \ln x_{it} + v_{it}$$

$$\alpha_i = \alpha + \gamma z_i$$

  - Random Effects

$$\ln y_{it} = \alpha + \beta \ln x_{it} + v_{it} + u_i$$

$$u_i = \gamma z_i$$

Stochastic Frontier

- The frontier will be defined by the firm with the largest Z (Z*)
  - "Average" function

$$\ln y_{it} = \alpha + \beta \ln x_{it} + \gamma z_i + \gamma_x z_i \ln x_{it} + v_{it}$$

  - Frontier

$$\ln y_{it}^* = \alpha + \beta \ln x_{it} + \gamma z_i^* + \gamma_x z_i^* \ln x_{it} + v_{it}$$

Efficiency and Firm heterogeneity

- Technical efficiency is the ratio of observed output to frontier output

$$LnTE_{it} = \ln y_{it} - \ln y_{it}^{*} = \gamma(z_i - z_i^{*}) + \gamma_x \ln x_{it}(z_i - z_i^{*})$$

- Implications
  - TE is time-varying even if Z is time invariant
  - If two firms use different inputs with the same level of TE, an increase in TE may require different increase in the level of management for each firm.

Separating Firm Heterogeneity from Inefficiency

Individual effects

- Fixed Effects $$\ln y_{it} = \alpha_i + \beta \ln x_{it} + v_{it}$$
  - Estimation by OLS with firm dummies
- Random Effects $$\ln y_{it} = \alpha + \beta \ln x_{it} + v_{it} + u_i$$
  - Estimation by GLS
- Heterogeneity is confounded with (time invariant) technical efficiency

Stochastic Frontier with Fixed Effects

$$\ln y_{it} = \alpha_i + \beta x_{it} + v_{it} - u_{it}$$

- Polachek and Yoon (1987)
- Kumbhakar (1991)
- Kumbhakar and Hjalmarson (1993)
- Greene (2005):Estimation by “brute force” ML

- α_i captures time-invariant heterogeneity
  - Technological differences
  - “Persistent” inefficiency
- U_{it} captures time-varying heterogeneity
  - Time-varying inefficiency (catch-up)
  - Time-varying technological differences (sector composition)

SF Random Coefficient Models

$$\ln y_{it} = \alpha + \beta_i \ln x_{it} + v_{it} - u_{it}$$

- Tsionas (2002), Huang (2004)

- Assumption: the parameters are random variables

$$\beta_i = \bar{\beta} + \omega_i$$

- Estimation: Bayesian techniques

SF Random Coefficient Models (II)

$$\ln y_{it} = \alpha_i + \beta_i \ln x_{it} + v_{it} - u_{it}$$

- Alvarez, Arias and Greene (2005)
- Estimation: Simulated Maximum Likelihood
  - (Limdep 9)

$$\ln y_{it} = \ln y_{it}^{*} - u_{it} = \alpha + \beta \ln x_{it} + \gamma z_i^{*} + \gamma_x z_i^{*} \ln x_{it} + v_{it} - u_{it}$$

$$\ln y_{it} = (\alpha + \gamma z_i^{*}) + (\beta + \gamma_x z_i^{*}) \ln x_{it} + v_{it} - u_{it}$$

$$\ln y_{it} = \delta_i + \delta_{xi} \ln x_{it} + v_{it} - u_{it}$$

SF Latent Class Models

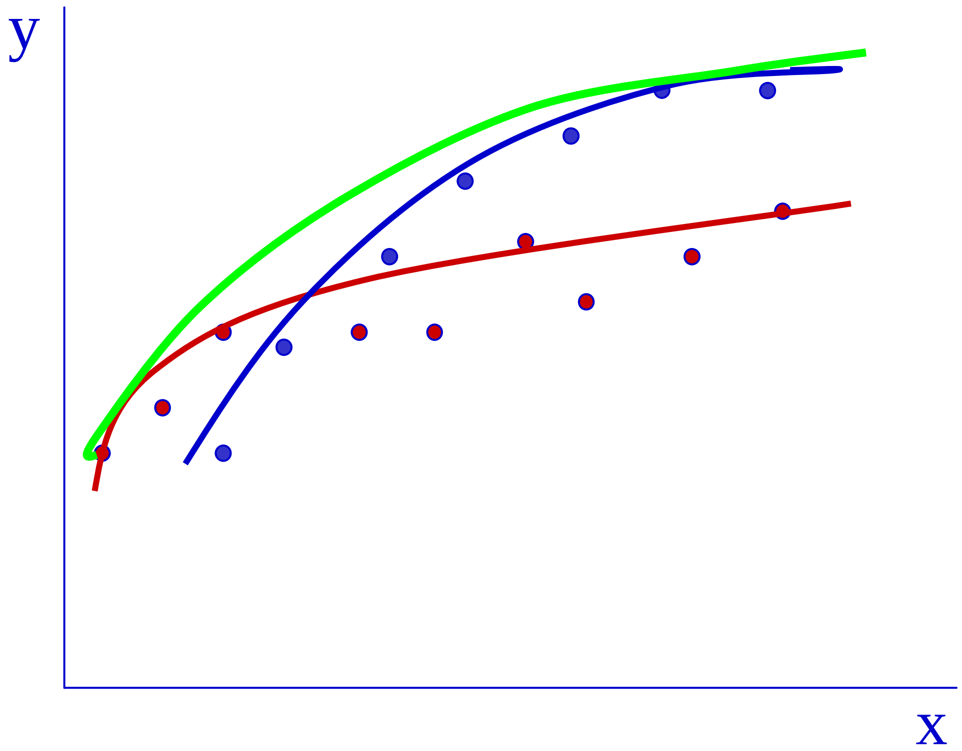
$$y_{it} = \alpha\big|_j + \beta\big|_j x_{it} + v_{it}\big|_j - u_{it}\big|_j$$

- Orea and Kumbhakar (2004)
- Alvarez and del Corral (2008)

- Estimation by ML

- Number of classes is unknown

Different Groups (Classes)



Application
Separating firm heterogeneity from inefficiency in regional production functions

Data

- Panel of 50 Spanish provinces (1985-1999)

- Output: GVA
- Inputs:
  - Private capital (K)
  - Labor (L)
  - Human Capital (HC)
  - Public Capital (G)

Empirical Model

- Functional form: Cobb-Douglas
  - Neutral Technical Change

- Stochastic Frontier with Fixed Effects (SFFE)

$$\ln y_{it} = \alpha_i + \sum_j \beta_j \ln x_{jit} + \delta_t t + \frac{1}{2}\delta_{tt} t^2 + v_{it} - u_{it}$$

  - V_{it} is assumed to be $N(0,\sigma_v)$
  - U_{it} is assumed to follow a half-normal distribution: $N^+(0,\sigma_u)$

Estimation

- Kumbhakar and Hjalmarsson (1993)
  - Estimation by GLS

- Greene (2002)
  - Estimation by ML

$$\log L = \sum_{i=1}^{N}\sum_{t=1}^{T}\log\left[\frac{1}{\Phi(0)}\,\Phi\left(-\lambda\left(\frac{y_{it}-\alpha_i-\boldsymbol{\beta}'\mathbf{x}_{it}}{\sigma}\right)\right)\phi\left(\frac{y_{it}-\alpha_i-\boldsymbol{\beta}'\mathbf{x}_{it}}{\sigma}\right)\right].$$

  - "Maximization of the unconditional log likelihood function can, in fact, be done by 'brute force' even in the presence of possibly thousands of nuisance parameters by using Newton's method and some well known results from matrix algebra"

Comparing inefficiency

	SFFE $(\alpha_i+V_{it}-U_{it})$	Pooled SF $(V_{it}-U_{it})$
Mean Inefficiency	0.08	0.09

Corr $(U_{it}$*_SFFE,*U_{it}*_PSF)= 0.50*

Comparing fixed effects

	Min	Max
SFFE	7.60	7.93
Within	10.14	12.58

Corr (FE_SFFE,FE_Within)= 0.19

Ranking of Fixed Effects

Province	Within	SFFE	Province	SFFE	Within
Madrid	1	23	Rioja	1	34
Barcelona	2	46	Las Palmas	2	13
Valencia	3	6	Baleares	3	6
Alicante	4	30	Salamanca	4	37
Vizcaya	5	48	Tenerife	5	16
Baleares	6	3	Valencia	6	3
Sevilla	7	20	Huelva	7	33
Zaragoza	8	26	Jaén	8	26
Málaga	9	18	Almería	9	32
Asturias	10	36	Cadiz	10	15

Main findings

- The models with U_{it} yield similar results, which in turn are very different from the FE model

- The estimated fixed effects in the FE and SFFE models are very different

Application
Identifying different technologies: extensive vs intensive dairy farms

Dairy Farming in Spain

- Recent trends
  - Large reduction in the number of farms
    - 73% reduction during 1990-2004
  - Quota System
    - Since 1991
  - Farms have grown
    - Average quota almost doubled in last seven years
  - Change in the production system
    - Many farms have adopted more intensive systems

Intensive vs Extensive Systems

- Characteristics of intensive systems
  - Farms produce more liters of milk per hectare of land

- How?
  - More cows per hectare of land
  - Higher use of concentrates per cow
  - Higher genetic level of the herds
    - Unobservable!!!

Objectives

- Are there differences in technological characteristics between extensive and intensive farms?
  - H0: Intensive farms have higher returns to scale
    - They have grown more than extensive farms

- Are there differences in technical efficiency?
  - H0: Intensive farms produce closer to their frontier
    - We consider that the intensive system is “easier” to manage

Latent Class Stochastic Frontiers

- Latent Class Stochastic Frontier Model

$$\ln y_{it} = f(x_{it})\big|_j + v_{it}\big|_j - u_{it}\big|_j$$

  - Likelihood function

$$\log LF = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{J} P_{ij} \prod_{t=1}^{T} LF_{ijt} \right)$$

  - Probabilities

$$P_{ij} = \frac{\exp(\delta_j q_i)}{\sum_{j=1}^{J} \exp(\delta_j q_i)}$$

  - Number of classes

Data

- Panel data set
  - 169 dairy farms
  - 6 years (1999-2004)

- Output
  - Milk liters

- Inputs
  - Cows, Feed (kg.), Labor (worker equivalents), Land (hectares), crop expenses (euros)

Empirical Model

- Translog stochastic production frontier

- Control variables
  - Time dummies
  - Location dummies

- Separating variables
  - Cows per hectare of land
  - Feed per cow

$$\ln y_{it} = \beta_0\big|_j + \sum_{l=1}^{L} \beta_l\big|_j \ln x_{lit} + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L} \beta_{lk}\big|_j \ln x_{lit} \ln x_{kit} + \sum_{t=2000}^{t=2004} \lambda_t\big|_j D_t + \sum_{z=1}^{z=6} \gamma_z\big|_j DLOC_{zi} + v_{it}\big|_j - u_{it}\big|_j$$

Estimation results

	'Pooled' Stochastic Frontier	Latent Class Model	
		Extensive Group	Intensive Group
Frontier			
Constant	**12.598*** **	**12.449*** **	**12.656*** **
Cows	**0.476*** **	**0.472*** **	**0.684*** **
Feed	**0.425*** **	**0.228*** **	**0.325*** **
Land	**0.006**	**0.027**	**0.024**
Farm$	**0.126*** **	**0.088*** **	**0.056*** **

*, **, *** indicate significance at the 10%, 5% and 1% levels

Characteristics of the Systems

	Extensive	Intensive
Farms	53	77
Milk (liters)	256,130	383,395
Cows	39	46
Land (ha.)	20	19
Milk per hectare	13,588	20,013
Cows per hectare	2.10	2.45
Milk per cow	6,522	8,130
Feed per cow	3,239	3,747
Milk per feed	2.07	2.23

Scale Elasticity in the LCM

Extensive	Intensive
0.945	1.052

Intensive farms have higher scale elasticity than extensive farms

Technical Efficiency

	Extensive	Intensive
Pooled	0.871	0.928
LCM	0.946	0.967

Discussion

- The results of the LCM help to explain two empirical facts
  - Farms grow despite the decline in the price of milk
  - Large farms buy quota from small farms
    - The marginal value of quota is price minus marginal cost

Conclusions

- It is important to model unobserved heterogeneity

- Some new techniques provide an interesting framework to control for firm heterogeneity