# Modeling the effect of competition using robust conditional nonparametric frontiers: Evidence from U.S. manufacturing sector

Polemis, Michael and Stengos, Thanasis and Tzeremes, Nickolaos

University of Piraeus, University of Guelph, University of Thessaly

1 September 2018

# Modeling the effect of competition using robust conditional nonparametric frontiers: Evidence from U.S. manufacturing sector

Michael L. Polemis[a,b*], Thanasis Stengos[c], Nickolaos G. Tzeremes[d]

[a] *Department of Economics, University of Piraeus, Piraeus, Greece , Email: mpolemis@unipi.gr (*corresponding author)*
[b] *Hellenic Competition Commission, Athens, Greece.*
[c] *University of Guelph, Department of Economics, Ontario*
[d] *Laboratory of Economic Policy and Strategic Planning, Department of Economics, University of Thessaly, Greece.*

## Abstract

The study applies the probabilistic framework of nonparametric frontier estimation in order to model the effect of competitive conditions on sectors' production efficiency levels. We utilize conditional Order-m robust frontiers modeling the dynamic effects of competitive conditions on a sample of 462 U.S. 6-digit manufacturing sectors over the period 1958-2009. The results derived from the time-dependent robust conditional estimators unveil a non-linear relationship between market competition and productive efficiency. Our findings suggest that for higher competitive conditions the effect is positive up to a certain threshold point after which the effect becomes negative.

**Keywords:** Probabilistic frontier analysis; Conditional efficiency; Order-m estimators; U.S. manufacturing; Competition.

**JEL codes:** C14; L60; O14

## 1. Introduction

The link between competition and innovation (or productivity) has been thoroughly examined by many researchers (see for example Hashmi and Biesebroeck, 2016; Correa and Ornaghi, 2014; Hashmi, 2013; Aghion et al, 2005). However, the relationship between competition and productive efficiency although classical in the Industrial Organization (IO) literature has been nearly overlooked by the existing studies.

Most of the related studies rely on Data Envelopment Analysis (DEA) methods to empirically estimate the level of productive efficiency in an industry/sector (see among others Huang et al, 2014; Tran et al, 2018; López et al, 2018). Our study departs from this strand of literature since we use for the first time in the literature a flexible nonparametric partial frontier analysis to model sectors' productive efficiency which is able to capture possible non-linearities. In the applying methodological framework we take into account time effects and the effects generated by competition without imposing any restrictive assumptions on the statistical models describing the data generating process (Simar and Wilson, 2011; Daraio et al., 2018).

Our findings derived from the time-dependent robust conditional estimators uncover a non-linear relationship between market competition and productive efficiency. Moreover, we argue that for higher competitive conditions the effect is positive up to a certain threshold point after which the effect becomes negative.

The rest of this paper is as follows. Section 2 presents the data and discusses the methodology employed. The results of our analysis are presented in Section 3, while Section 4 performs the necessary robustness check to test for the validity of our findings. Finally, Section 5 concludes the paper.

## 2. Data and Methodology

This section describes the data and the methodology used in the empirical analysis. Our sample includes the US manufacturing industries broken down at a 6-digit level (462 DMUs) under NAICS classification over the period 1958-2009 (T = 52). Therefore, the sample accounts for a total balanced panel comprising of 24,024 observations.

Similarly to previous studies (Correa and Ornaghi, 2014; Polemis and Stengos, 2015) we use the following variables: a) Value added per sector as a proxy for total output (Q), b) Total employment per sector as a proxy for labour, c) Total capital stock per sector as a proxy for capital, d) Total cost of electricity and fuels per sector as a proxy for energy and e) Herfindahl-Hirschman Index (HHI) as a proxy for Product Market Competition (PMC). The HHI ranges from 0 (Perfect competition) to 10,000 (Monopoly). We mention though that since marginal cost is not observable the calculation of the Lerner Index (LI) was not feasible. Summary statistics are reported in the following table.

**< Insert Table 1 about here>**

Let a sectors' production function to be characterized by a set of inputs $x \in \mathbb{R}_{+}^{p}$ and by a set of outputs $y \in \mathbb{R}_{+}^{q}$. Then the vector of competitive conditions can be indicated as $W \in \mathcal{C} \subset \mathbb{R}^{r}$ and the production attainable set can be represented as:

$$P^{w} = \{(x, y) | W = w, x \text{ can produce } y\}. \tag{1}$$

Based on Daraio and Simar (2005) we have $P = \bigcup_{w \in \mathcal{C}} P^{w}$ so that we can have for all $W \in \mathcal{C}, P^{w} \subseteq P$ .

The output oriented efficiency measure of a sector operating at $(x_0, y_0)$ level can be defined as:

$$\vartheta(x_0, y_0) = \sup\{\vartheta > 0 | (x_0, \vartheta y_0) \in P\}. \tag{2}$$

Cazals et al. (2002) have shown that production process can be characterized from by the:

$$A(x, y) = \text{Prob}(X \leq x, Y \geq y). \tag{3}$$

Then the output oriented efficiency measure defined in (2) can be presented as:

$$\vartheta(x_0, y_0) = \sup\{\vartheta | A(x_0, \vartheta y_0) > 0\}. \tag{4}$$

Equation (3) can be decomposed to:

$$A(x, y) = \text{Prob}(Y \geq y | X \leq x)\text{Prob}(X \leq x) = \Phi_{Y|X}(y|x)F_x(x). \tag{5}$$

The conditional distribution can be represented as:

$$A(x, y|w) = \text{Prob}(Y \geq y | X \leq x, W = w)\text{Prob}(X \leq x | W = w)\Phi_{Y|X,W}(y|x, w)F_{X|W}(x|w), \tag{6}$$

Then by following the relative literature (Bădin et al. 2010; Mastromarco and Simar 2015) a sector's time-dependent conditional efficiency measure operating at level $(x_0, y_0)$ under the competitive conditions $W = w_0$, at a period $T = t_0$, can be expressed as:

$$\vartheta_t(x_0, y_0 | w_0) = \sup\{\vartheta > 0 | (x_0, \vartheta y_0) \in P_t^{w_0}\} =$$

$$\sup\{\vartheta > 0 | \Phi_{X,Y|W}^t(\vartheta y_0 | X \leq x_0, W = w_0, T = t_0) > 0\}. \tag{6}$$

For a given sectors' inputs $x$ in the interior of the support of $X$, consider $m, i.i.d.$ random variables $Y_i, i = 1, \ldots, m$ which have been generated by the conditional $q - variate$ distribution function $\Phi_{Y|X}(y|x_0) = \text{Prob}(Y \leq y_0 | X \leq x_0)$. Then we can define a random set as:

$$P_m(x_0) = \{(x\acute{}, y) \in \mathbb{R}_+^{p+q} | \acute{x} \leq x_0, y \leq Y_i, i = 1, \ldots, m\}, \tag{7}$$

analogously to (2) we can define:

$$\tilde{\vartheta}_m(x_0, y_0) = \sup\{\vartheta > 0 | (x_0, \vartheta y\ ) \in P_m(x_0)\}. \tag{8}$$

Then a robust output oriented efficiency measure can be defined as:

$$\vartheta_m(x_0, y_0) = \mathbb{E}(\tilde{\vartheta}_m(x_0, y_0) | X \leq x_0). \tag{9}$$

Finally, the unconditional $\vartheta_m(x_0, y_0)$ and the time-dependent conditional $\vartheta_{t,m}(x_0, y_0|w_0)$ Order-m (robust) output oriented efficiency measures can be represented as:

$$\hat{\vartheta}_m(x_0, y_0) = \hat{\vartheta}(x_0, y_0) - \int_0^{\hat{\vartheta}(x_0,y_0)} \left(1 - \hat{\Phi}_{Y|X}(uy_0|X \leq x_0)\right)^m du, \tag{10}$$

$$\hat{\vartheta}_{t,m}(x_0, y_0|w_0) = \hat{\vartheta}_t(x_0, y_0|w_0) - \int_0^{\hat{\vartheta}_t(x_0,y_0|w_0)} \left(1 - \hat{\Phi}^t_{X,Y|W}(uy_0|X \leq x_0, W = w_0, T = t_0)\right)^m du. \tag{11}$$

Following Bădin et al. (2012) we create the following ratio:

$$\hat{Q} = \hat{\vartheta}_{t,m}(x_0, y_0|w_0)/\hat{\vartheta}_m(x_0, y_0). \tag{12}$$

Then in second-stage analysis we apply a local linear nonparametric regression[1] in order to examine the effect of time and competitive conditions on the sectors' efficiency measures. Specifically we apply:

$$\hat{Q} = f(w_i, t_i) + \epsilon_i. \tag{13}$$

By examining the ratios $(\hat{Q})$ as a function of $w$ and $t$ we will be able to determine the effect on sectors production performance levels. An increasing shape for $\hat{Q}$ as a function of $w$ and $t$ would indicate a favorable effect of competitive conditions and time on sectors' production efficiency levels. However, a decreasing shape will signify a negative effect.

## 3. Results and discussion

As suggested by Cazals et al. (2002) partial frontiers (i.e. Order-m) are less sensitive to outliers. This is evident in the following figure, where we plot confidence intervals for some contours for selected sample years. A first look indicates an inverted U shaped curvature between HHI and productive efficiency.

[1] As in Jeong et al. (2010) we have applied Epanechnikov kernels and Least Squares Cross-Validation (LSCV) criterion (Li and Racine 2007; Hayfield and Racine, 2008) for bandwidth selection.

**<Insert Figure 1 about here>**

In the output oriented case a sector has an efficiency score below unity (i.e. super-efficient sector) if performs better compared to the randomly drawn $m$ (in our case m=50) sectors with input levels $\leq x$. If a sector lies on the frontier then is said to be efficient with an Order-m values equal to unity, while values greater than one indicate inefficiency.

Figure 2 portrays the probability density function of the output oriented Order-m frontiers over selective sample years. The results suggest that the U.S. sectors performed better between 1958 up to 1990. However it is evident that after 2000 the results indicate a deterioration of sectors' performance levels.

**<Insert Figure 2 about here>**

Figure 3 presents the effect of time and competition on sectors' production efficiency levels. The results signify a nonlinear relationship between competition and production efficiency (see upper panel). Specifically, for higher competitive conditions the effect is positive up to a certain threshold point after which the effect becomes negative (*inverted 'U'-shape*) with technological catch-up.[2] It is also worth emphasizing that the peak of the curvature lies approximately around the median of the distribution (4,000) so that industries are well spread across the U-shape (Aghion et al, 2005).

**<Insert Figure 3 about here>**

The above findings are in alignment with previous studies (Hashmi 2013; Aghion et al, 2005; Mukoyama, 2003) suggesting that there is a nonlinear relationship between competition and innovation. On the downward part of the curve there is the argument that innovation dissipates

[2] Highly competitive conditions are indicated by values less than $100$. Competitive conditions are indicated by values greater than $1500$; Moderate competitive conditions are indicated by values between $1500$ and $2500$ and low competition is indicated for values greater than $2500$ (DOJ, 2010).

with competition since more competition reduces profits. The latter fully justifies the "*Schumpeterian*" hypothesis indicating that monopoly power can better serve innovation and stimulate productive efficiency in a sector. However, when competition is in the intermediate range (upward part), the level of innovation is high. Combining these results, we get an inverted-U shaped curve between PMC and productive efficiency.

## 4. Robustness checks

In this section, we re-formulate our basic model by dividing the three inputs (capital, labour and energy) over output (Q) in order to isolate from the whole sample the capital, labour and energy intensive sectors of the US manufacturing industry.

As it is evident from Figure 3 there is a clear inverted "*U-shape*" between time and competition on sectors' production efficiency levels in all of the three sub-samples supporting the validity of our previous findings (see lower panel). In the capital intensive sectors, we observe that the peak of the inverted U is larger compared to labour and energy intensive sectors.

This can be attributed to the fact that these sectors are usually comprised by technologically equal firms (neck-and-neck industries) where the number of new entrants and subsequently the level of potential competition is low due to the existence of high barriers to entry. As a result, lower levels of competition lead to greater productive efficiency since firms in these industries innovate more when competition is low and the opposite (Hashmi, 2013).

## 5. Conclusions

This paper tries to fill this gap in the literature by applying an innovative application of Operational Research to examine the effect of PMC on sectors' economic efficiency levels. By using a novel time-dependent robust conditional DEA estimator we evaluate the dynamic effects alongside with the effects of competition on the sectors' technological change and technological catch-up levels.

The empirical analysis reveals that competition affects sectors' efficiency level in a non-linear relationship. Our findings indicate that when competition is too high or too low, the level of efficiency is low. On the contrary, when competition is in the moderate range, the level of productive efficiency is high. Moreover, we argue that sectors' lower competition levels enhance their efficiency levels (technological catch-up), supporting the validity of the inverted U shaped hypothesis. Our results remain robust when we include only capital, labour and energy intensive sectors. We notice though that the peak of the inverted U is larger to the capital intensive sectors.

## References

Aghion, P, Bloom, N, Blundell, R, Griffith, R. & Howitt, P. (2005). Competition and Innovation: An Inverted-U Relationship. *Quarterly Journal of Economics*, 120(2): 701–728.

Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *European Journal of Operational Research*, *201*(2), 633-640.

Bădin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, *223*(3), 818-833.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, *106*(1), 1-25.

Correa, J. A., and Ornaghi C (2014). Competition and Innovation: Evidence from US patent and productivity data. *The Journal of Industrial Economics,* 62(2): 258-285.

Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24(1), 93-121.

Daraio, C., & Simar, L. (2014). Directional distances and their robust versions: computational and testing issues. *European Journal of Operational Research*, 237(1), 358-369.

Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the '*separability'* condition in non-parametric, two-stage models of production. *Econometrics Journal*, 21(2), 170-191.

DOJ (2010). Horizontal Merger Guidelines, U.S. Department of Justice and the Federal Trade Commission, Can be downloaded from: https://www.justice.gov/sites/default/files/atr/legacy/2010/08/19/hmg-2010.pdf

Hashmi A.R., & Biesebroeck, J.V. (2016). The Relationship between Market Structure and Innovation in Industry Equilibrium: A Case Study of the Global Automobile Industry. *Review of Economics and Statistics* 98 (1): 192-208.

Hashmi, A. R., (2013). Competition and Innovation: The Inverted-U Relationship Revisited. *Review of Economics and Statistics*, 95(5): 1653–1668.

Hayfield, T., & Racine, J. (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software*, 27(5), 1 - 32. doi:http://dx.doi.org/10.18637/jss.v027.i05

Huang, CW., Foo Nin Ho, Yung-ho Chiu (2014). Measurement of tourist hotels′ productive efficiency, occupancy, and catering service effectiveness using a modified two-stage DEA model in Taiwan, *Omega*, 48: 49-59,

Jeong, S. O., Park, B. U., & Simar, L. (2010). Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, *173*(1), 105-122.

Li, Q., and Racine, J.S. (2007). Nonparametric Econometrics: Theory and Practice. Princeton University Press.

López, G.P., Diego Prior, José L. Zafra-Gómez (2018). Temporal scale efficiency in DEA panel data estimations. An application to the solid waste disposal service in Spain, *Omega*, 76: 18-27.

Mastromarco, C., & Simar, L. (2015). Effect of FDI and time on catching up: New insights from a conditional nonparametric frontier analysis. *Journal of Applied Econometrics*, *30*(5), 826-847.

Mukoyama, T, (2003). Innovation, Imitation and Growth with Cumulative Technology. *Journal of Monetary Economics* 50: 361–380

Polemis, M., & Stengos, T. (2015). Does market structure affect labour productivity and wages? Evidence from a smooth coefficient semiparametric panel model. *Economics Letters*. 137(C): 182-186.

Simar, L., & Wilson, P. W. (2011). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis*, *36*(2), 205.

Tran, T.H., Mao, Y., Nathanail, P., Siebers, P.O., and Darren R.(2018). Integrating slacks-based measure of efficiency and super-efficiency in data envelopment analysis, *Omega*, https://doi.org/10.1016/j.omega.2018.06.008.

## Tables and Figures

**Table 1:** Summary statistics

| **Variables** | **Observations** | **Mean** | **Standard deviation** | **Min** | **Max** |
|---|---|---|---|---|---|
| Output | 24,024 | 6,286 | 17,650 | 3.041 | 898,019 |
| Labour | 24,024 | 34.77 | 45.09 | 0.200 | 559.9 |
| Capital | 24,024 | 2,760 | 6,405 | 4.100 | 120,110 |
| Energy | 24,024 | 128.0 | 394.6 | 0.149 | 8,052 |
| HHI | 24,024 | 935.4 | 233.9 | 12.00 | 9,406 |

**Figure 1:** Regression plots of the effect of HHI

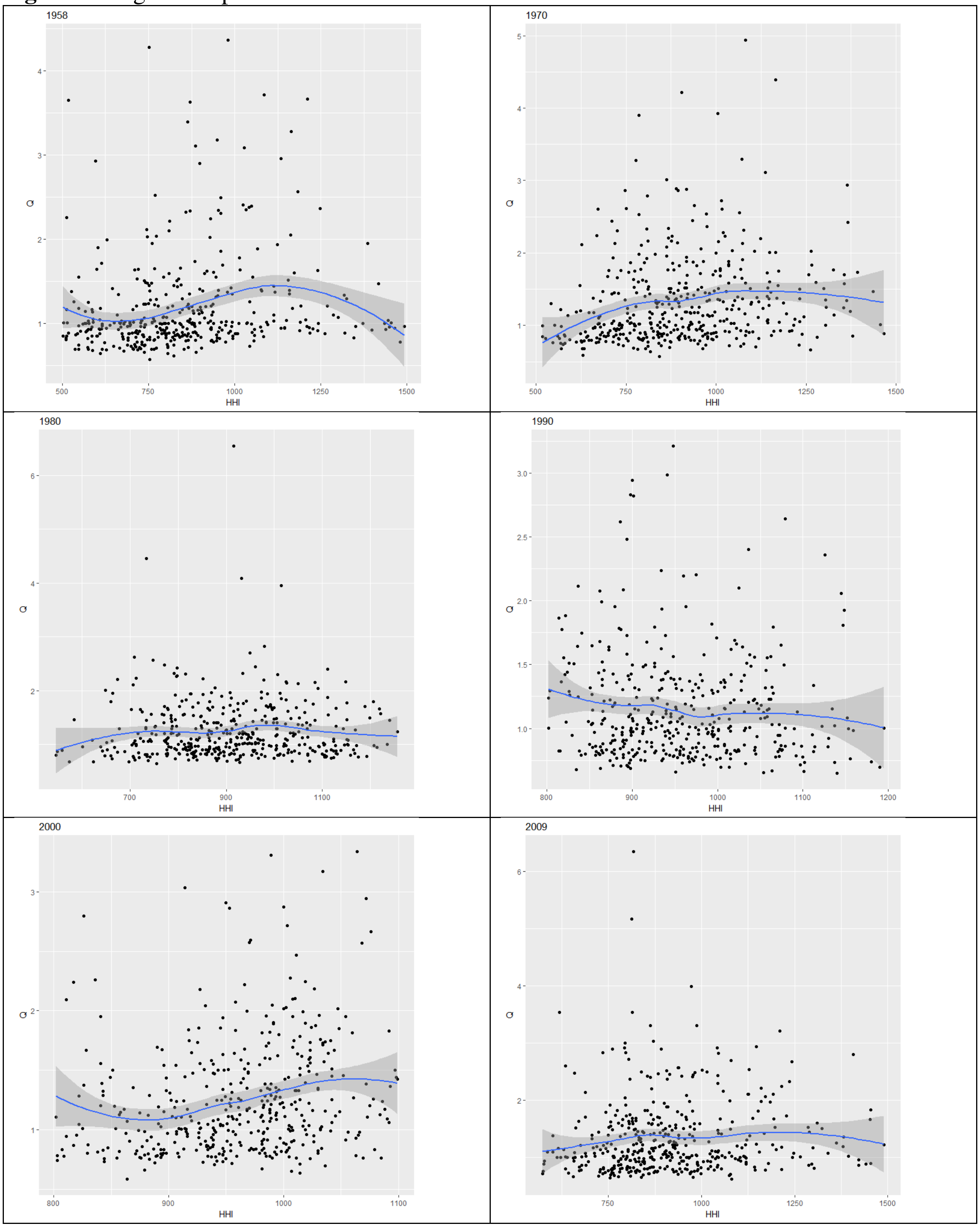
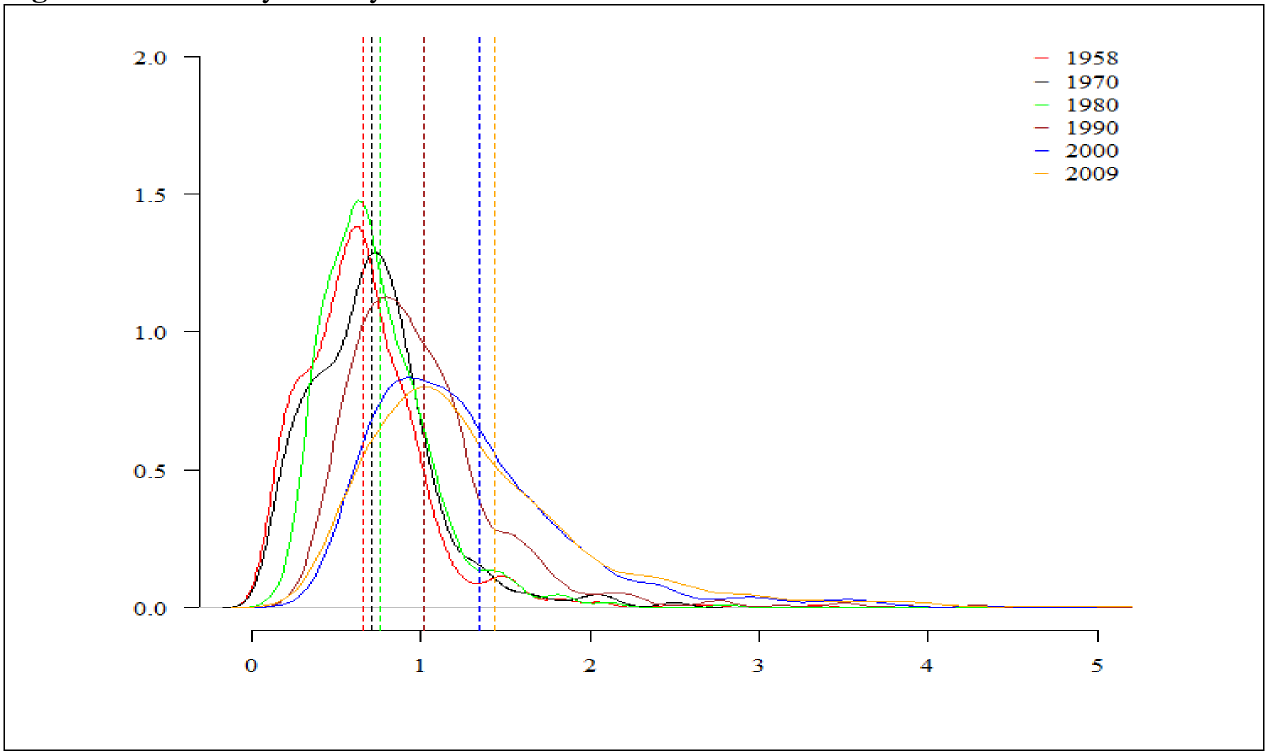


**Figure 2**: Probability Density Functions.




**Note:** Vertical dotted lines indicate per period average efficiency estimates

**Figure 3**: The effect on sectors' production efficiency levels