Accepted Manuscript

Cost Constrained Industry Inefficiency

Antonio Peyrache

PII: S0377-2217(15)00600-1
DOI: 10.1016/j.ejor.2015.06.067
Reference: EOR 13074

To appear in: *European Journal of Operational Research*

Received date: 8 May 2014
Revised date: 26 June 2015
Accepted date: 29 June 2015

Please cite this article as: Antonio Peyrache, Cost Constrained Industry Inefficiency, *European Journal of Operational Research* (2015), doi: 10.1016/j.ejor.2015.06.067

Highlights

- Industry inefficiency is defined for cost constrained production environments;
- Industry prices are defined and the optimal industry configuration is determined;
- The industry inefficiency indicator is decomposed into sources components;
- An empirical application to Ontario energy data is presented;
- The empirical application shows how to use the methodology in order to estimate the effect of changes in the regulation regime.

Cost Constrained Industry Inefficiency

Antonio Peyrache*

July 8, 2015

Abstract

In this paper a definition of industry inefficiency in cost constrained production environments is introduced. This definition uses the indirect directional distance function and quantifies the inefficiency of the industry in terms of the overall output loss, given the industry cost budget. The industry inefficiency indicator is then decomposed into sources components: reallocation inefficiency arising from sub-optimal configuration of the industry; firm inefficiency arising from a failure to select optimal input quantities (given the prevalent inputs prices); firm inefficiency due to lack of best practices. The method is illustrated using data on Ontario electricity distributors. These data show that lack of best practices is only a minor component of the overall inefficiency of the industry (less than 10%), with reallocation inefficiency accounting for more than 75% of the overall inefficiency of the system. An analysis based on counter-factual input prices is conducted in order to illustrate how the model can be used to estimate the effects of a change in the regulation regime.

Keywords: Data Envelopment Analysis (DEA), Industry Inefficiency, Indirect Directional Distance Function, Productivity, optimal configuration.

*Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, The University of Queensland, Australia (a.peyrache@uq.edu.au).

Introduction.

In this paper a definition of industry inefficiency in cost constrained production environments is introduced. The idea behind cost constrained production is that firms or decision making units (DMU) are allocated a certain cost and are supposed to produce as much output as possible with this given cost (this implies optimizing when selecting input quantities). It is assumed that data on the inputs, the prices of the inputs and the outputs produced by each firm in a given industry are available. With such data, it is possible to determine the overall cost budget of the industry by looking at the inputs used and the input prices faced by each firm. The problem that the central planner (or a market) now faces is how to allocate this overall budget across the different production units (given a certain number of constraints) in order to maximize the overall output. This problem corresponds to the implicit determination of the optimal structure of the industry via the determination of the optimal number of firms that should populate the industry and the optimal allocation of resources across these firms. Once this optimization problem is solved, one is able to quantify inefficiency in terms of the directional distance function (DDF), where inefficiency is measured in terms of the overall output loss due to different sources: *i)* the inefficiency of the firms actually operating in the industry (lack of best practice) and *ii)* the inefficiency arising from a sub-optimal configuration of the industry.

To the best of our knowledge the first paper to address this issue explicitly in a linear programming framework is due to Ray and Hu (1997). This contribution introduced the basic model in a primal context, where only input and output quantities are observed. Later on Lozano and Villa (2004) introduced the centralized resource allocation model which uses the same idea (and it is in fact a special case of Ray and Hu (1997) where the number of firms is fixed to the observed one). Following these attempts a literature developed to accommodate alternative empirical settings (see Aparicio et al (2013), Aparicio and Pastor (2012), Asmild et al (2009), Asmild et al (2012), Fang (2013), Fang and Zhang (2008), Gimenez-Garcia et al (2007), Lotfi et al (2010), Lozano and Villa (2004, 2005), Lozano et al (2004, 2009, 2011), Mar-Molinero et al (2012), Ray (2007), Ray and Mukherjee (1998)). Ray et al (2008) extended the industry efficiency model using a cost function approach with input prices varying across locations.

The method introduced in this paper is illustrated using data on Ontario electricity distributors. The data show that lack of best practices at the firm level is only a minor component of the overall inefficiency of the industry (less than 10%). The bulk of the industry inefficiency is accounted for by severe deviations from the optimal configuration. In this empirical part a counterfactual analysis is also provided.

The rest of the paper is organized as follows. Section 1 presents the definition of firm technology and inefficiency. Section 2 extends these notions to the industry level. Section 3 is dedicated to the empirical illustration. Finally section 4 concludes.

1 Firm Technology and Inefficiency

Consider an industry where $\mathbf{x} \in R_+^N$ inputs produce $\mathbf{y} \in R_+^M$ outputs. It is assumed that data on inputs and outputs are available for a number K of firms $(k = 1, ..., K)$. The data can be collected into two matrices: an input matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]'$ and an output matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K]'$, where the observations relative to each firm are collected on the rows of these matrices. The dataset is the collection of these matrices:

$$(\mathbf{X}, \mathbf{Y}) \tag{1}$$

The data generated firm technology set (production set, or production possibilities set) is given by the picece-wise linear envelop of the observations available (Banker et al., 1984, 2005):

$$\Psi = \left\{ (\mathbf{x}, \mathbf{y}) : \lambda \mathbf{X} \leq \mathbf{x}' , \ \lambda \mathbf{Y} \geq \mathbf{y}' , \ \sum_k \lambda_k = 1 , \ \lambda \geq \mathbf{0} \right\} \tag{2}$$

This production set is built under the assumption that the technology satisfies convexity and free disposability of inputs and outputs. The constraint on the intensity vector $\sum \lambda_k = 1$ means that the technology allows for variable returns to scale (VRS). An equivalent representation of the technology is given by the output sets, which are the collection of all producible outputs given a certain input vector $P(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \Psi\}$ and can be represented in functional form via the directional output distance function (DDF) (see Chambers et al., 1996, 1998):

$$TE = D(\mathbf{x}, \mathbf{y}; \mathbf{g}_y) = \sup_{\beta} \{\beta : (\mathbf{y} + \beta \mathbf{g}_y) \in P(\mathbf{x})\} \tag{3}$$

The DDF provides a measure of technical efficiency (TE) or, more precisely, of technical inefficiency because it represents the total loss in output due to inefficient use of the inputs available to the firm (with respect to the benchmark represented by technology (2)). This measure of inefficiency is expressed in terms of the *numeraire* $\mathbf{g}_y$ which is therefore assumed to be fixed across firms and time periods (comparisons of technical inefficiency using different *numeraires* would be equivalent to comparing apples with oranges). The function (3) can also be interpreted as a shortage function, inasmuch it is a measure of total output loss with respect to a potential output benchmark.

It is now interesting to consider an alternative representation of the technology (2) which assumes availability of information on input prices in the form of a row vector $\mathbf{w} \in \mathbb{R}_+^N$. In this case, one may think of the production possibilities as the collection of all the possible output vectors which are feasible when the cost budget is set at level C. This gives rise to the indirect output sets $IP(\mathbf{w}/C) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \Psi , \ \mathbf{wx} \leq C\}$ and their functional representation via the indirect directional output distance function (IDDF):

$$CE = ID(\mathbf{w}/C, \mathbf{y}; \mathbf{g}_y) = \sup_{\beta} \{\beta : (\mathbf{y} + \beta \mathbf{g}_y) \in IP(\mathbf{w}/C)\} \tag{4}$$

This alternative measure of inefficiency involves the use of input prices and it measures how much output production could be expanded, given that the overall cost the firm is facing is given. Since the constraint here is the overall cost rather than a specific input vector, this type of inefficiency has also been called cost constrained inefficiency (see Färe and Grosskopf, 1994; Grosskopf et al., 1997). It should be emphasized that, though the cost is constraining production, the inefficiency measure is defined on the output side for a given *numeraire* $\mathbf{g}_y$. This means that the quantities in equations (3) and (4) share the same underlying *numeraire* and they can be compared. Invoking the duality theorem proved in Färe and Primont (2006), it holds that $ID(\mathbf{w}/C, \mathbf{y}; \mathbf{g}_y) \geq D(\mathbf{x}, \mathbf{y}; \mathbf{g}_y)$, where the difference between the two indicators is an allocative efficiency component, interpreted as the loss in output due to the choice of a non-optimal input mix:

$$AE = ID(\mathbf{w}/C, \mathbf{y}; \mathbf{g}_y) - D(\mathbf{x}, \mathbf{y}; \mathbf{g}_y) \tag{5}$$

The allocative inefficiency definition embedded in equation (5) determines the quantity of output which is lost because the firm fails to choose the optimal input mix given the prevailing input prices. Given the duality between the direct and the indirect DDF, it is thus possible to decompose the firm level cost constrained inefficiency (CE) into the two components defined in equations (5) and (3):

$$CE = TE + AE \tag{6}$$

The left hand side of this equation is a measure of the overall loss in output for a firm operating at the specified cost level. The first component on the right hand side attributes part of this inefficiency to a less than optimal use of the given input vector; while the second component is a measure of loss in output attributable to the firm failing to choose an optimal input vector.

2 Industry Technology and Inefficiency

The previous section introduced the main representation of the firm technology set and the firm inefficiency. The purpose of this section is to extend these notions to the industry level. For the purposes of this study the industry is defined in terms of inputs and outputs homogeneity, so that all the firms operating in the industry uses the same set of inputs to produce the same set of outputs. It is assumed that any number of firms can operate in the industry (entry and exit of firms is allowed) and all the firms in the industry (already operating in it or potentially entrant) face the same technology set Ψ defined in equation (2); in other words all the firms in the industry face the same production trade-offs. Under these assumptions the industry technology set is defined as (see Peyrache, 2013):

$$\Psi_I = \cup_{S=1}^{+\infty} \left(\sum_{s=1}^{S} \Psi \right) \tag{7}$$

The summation in parentheses is a special case of the aggregation discussed in Li and Ng (1995) and Zelenyuk (2006)[1]. It should be noted that, though the firm technology set is convex, the industry technology set may show some non-convexity. This non-convexity at the industry level arises because of the indivisibility of the firm: only an integer number of firms can operate in the industry (of course this non-convexity becomes less important as the number of firms grows large). The set defined in (7) can be also written as follow:

$$\Psi_I = \left\{ (\mathbf{x}, \mathbf{y}) : \lambda \mathbf{X} \leq \mathbf{x}' \, , \; \lambda \mathbf{Y} \geq \mathbf{y}' \, , \; \sum_k \lambda_k = S \, , \; \lambda \geq \mathbf{0} \, , \; S \in \mathbb{N} \right\} \tag{8}$$

Contrary to definition (2), the intensity vector is now constrained to sum up to the the integer number S which represents the number of firms operating (actually or potentially) in the industry. Interestingly enough, this set collapses to the firm technology set (2) when $S = 1$. The industry technology returns all the possible input and output combinations which are feasible at the industry level and it is an enlargement of the firm production set. Since the industry as a whole is operating within a given cost budget, it is interesting to describe the industry technology set in terms of the associated indirect output sets which represent all the output combinations the industry is able to produce given the overall cost budget available. It would be tempting to define the industry indirect output sets as $IP_I (\mathbf{w}/C) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \Psi_I \, , \; \mathbf{wx} \leq C\}$ (note that this definition uses the industry technology (7) rather than the firm technology (2) as a benchmark). Unfortunately this definition would ignore the structure of prices that the industry is facing. Observed industry inputs, outputs and cost can be defined quite easily as the sum of inputs, outputs and costs across the firms operating in the industry. In fact, the total output of the industry will be equal to the sum of all the firm level outputs ($\mathbf{Q} = \sum_k \mathbf{y}_k$) and, similarly, the industry overall cost budget will be the sum of the firm level costs ($C_I = \sum_k C_k$). On the contrary (unless the law of one price applies in the inputs market) there will be a heterogeneous variety of different input price vectors which segments the inputs market and make the price vector faced by each firm different. In this general case, it is not clear how to define an industry input price vector and therefore it is not clear which input price vector to use in the definition of the industry indirect DDF. Suppose that the inputs market is divided into different segments and in each segment a particular input price vector prevails (this embeds the law of one price as a special case). An example of a segmented inputs market is when input prices vary across different geographical locations within a nation or a macro region (this for example is the case considered in the Multi-Location Minimum Cost model of Ray et al, 2008). Each firm, by deciding in which segment of the market operate production, implicitly decides what vector of input prices to face. The number of segments of the market can be smaller, equal or larger than the number of firms actually operating in

[1]The reader should note that if the production set is non-convex, then $\sum_{s=1}^{S} \Psi \neq S\Psi$. The proof of this is in Li and Ng (1995). In the case of a convex production set the equality holds.

the industry. In fact, the first two cases corresponds to a situation in which every firm is facing a different input price vector or some of the firms share the same input price vector. The last case corresponds to a situation in which there are more price vectors than firms. Though it may seem an unlikely situation, this case corresponds to a counter-factual analysis in which one considers not only the price vectors that the firms are currently facing in the industry, but also an additional number of price vectors that will drive the counter-factual analysis. So, for example, one may be interested in assessing the potential efficiency gains obtainable by the industry via a shift of production towards locations where the industry is not currently operating. This counter-factual analysis may be important when assessing if to open a new firm (an entrant) in a segment of the inputs market where no other firms are operating yet. All in all, what is important from a mathematical point of view is to distinguish and keep separate theoretically the number of segments of the market from the number of firms which actually operate in the industry. As noted by Ray et al (2008) the location idea imposes an integer constraint to guarantee that the number of firms operating in each location is an integer number. In other words, the vector of prices at a specific location can be activated only if a firm is physically there; and the number of firms at each location must be an integer number (which is the reason for the integer constraint). Though the location idea is a very attractive way of thinking of input price variation, it is not the only possibility. Price variation will arise, for example, every time that there is price discrimination, limited information on prices (or costly search) or information is organized in networks (therefore belonging to a specific non-physical network is what determines prices). Contrary to the location idea, in this case input price variation is not linked to the presence of a firm in a specific location. This means that, in this second case, at the industry level it is possible to choose any combination of the observed input prices (irrespective of the number of firms). In order to formalize these ideas, consider now a partition of the input vector into two components of dimension $N_1 + N_2 = N$: $\mathbf{x}' = \begin{pmatrix} \boldsymbol{\upsilon}' & \boldsymbol{\mu}' \end{pmatrix}$. Associated to this partition of the input vector we consider two input price matrices $\mathbf{\Omega}$ and $\mathbf{\Pi}$ of dimension $J \times N_1$ and $P \times N_2$ respectively. The first component will respond to the logic behind the location idea: in order to activate such a vector of prices, a firm needs to operate in that segment of the market (J is the total number of segments). The second component will not respond to such a logic: it is possible to access a specific vector of prices in the second component without necessarily activate any production in that specific segment of the market (for example by becoming part of a network; P is the total number of alternative prices). The indirect output set for the industry will include different constraints for these different components:

$$IP_I\left(\mathbf{\Omega}, \mathbf{\Pi}, C\right) = \left\{ \begin{array}{c} \sum_j \mathbf{y}_j : \quad \left(\boldsymbol{\upsilon}_{\mathrm{j}}, \boldsymbol{\mu}_j, \mathbf{y}_j\right) \in \Psi_I \,, \\ \sum_j \left(\boldsymbol{\omega}_j \boldsymbol{\upsilon}_{\mathrm{j}} + \sum_p \boldsymbol{\pi}_p \boldsymbol{\mu}_p^j\right) \leq C \,,\ \sum_p \boldsymbol{\mu}_p^j = \boldsymbol{\mu}_j \end{array} \right\}$$

This discussion implies the following definition of industry inefficiency (IE), which takes both the number of firms and the allocation of cost across different segments as variables of choice:

$$ID_I(\mathbf{y}, \mathbf{\Omega}, \mathbf{\Pi}, C; \mathbf{g}_y) = \sup_{\beta}\{\beta : (\mathbf{y} + \beta\mathbf{g}_y) \in IP_I(\mathbf{\Omega}, \mathbf{\Pi}, C)\} \tag{9}$$

This new definition represents, for a given input price structure and a given industry cost budget, the overall loss in output when the benchmark is the industry technology rather than the firm technology. In other words, this definition is comparing the observed output produced to a benchmark dictated by the industry technology. If the observed total output and total cost budget of the industry are inserted in equation (9), the industry inefficiency indicator is obtained:

$$IE = ID_I(\mathbf{Q}, \mathbf{\Omega}, \mathbf{\Pi}, C_I; \mathbf{g}_y) \tag{10}$$

The solution to this problem provides: *i)* the optimal number of firms that should populate the industry; *ii)* the optimal allocation of the total industry cost across the different segments of the inputs market; *iii)* the overall inefficiency of the industry. The inefficiency of the industry is, once again, measured in terms of the numeraire $\mathbf{g}_y$. The definition of industry inefficiency just provided involves two very different types of inefficiency: on one hand, the inefficiency arising from a non-optimal allocation of cost across different firms (which includes a non-optimal number of firms as a special case); on the other hand, a failure to optimally allocate the overall cost budget across the different segments of the inputs market. The price effect vanishes if the matrices of prices $(\mathbf{\Omega}, \mathbf{\Pi})$ contains only one element, i.e. in the case in which the law of one price applies.

2.1 Industry inefficiency decomposition.

Given the previous definitions, it is now possible to seek for a decomposition of the industry inefficiency indicator defined in (10) into its sources components. The first step in this direction is to consider the impact of the inefficiencies arising from the firms already operating in the industry. These were defined in the previous section in equations (3) and (4) and can now be aggregated by a simple sum across firms (the reader should note how simple aggregation of directional distance functions is). This is possible because an additive notion of inefficiency has been used and the *numeraire* $\mathbf{g}_y$ is common to all firms. The industry technical inefficiency is therefore defined as (see Briec et al, 2003; Färe et al, 2008; Färe and Primont, 2003; Färe and Zelenyuk, 2003):

$$ITE = \sum_{k=1}^{K} D(\mathbf{x}_k, \mathbf{y}_k; \mathbf{g}_y) \tag{11}$$

Since this is only one part of the total inefficiency of the firm, one should also aggregate the cost constrained inefficiency into an industry indicator:

$$ICE = \sum_{k=1}^{K} ID(\mathbf{w}_k / C_k, \mathbf{y}_k; \mathbf{g}_y) \tag{12}$$

A decomposition of the firm inefficiency was given in the previous section and it still holds at the industry level, giving rise to an industry allocative inefficiency indicator:

$$\begin{aligned} IAE &= ICE - ITE = \\ &= \sum_{k=1}^{K} \left[ID\left(\mathbf{w}_k/C_k, \mathbf{y}_k; \mathbf{g}_y\right) - D\left(\mathbf{x}_k, \mathbf{y}_k; \mathbf{g}_y\right) \right] \end{aligned} \tag{13}$$

The difference between the two indicators is a measure of the impact on industry output loss of non optimal input choices at the firm level. Therefore overall industry inefficiency can be decomposed as follow:

$$IE = ITE + IAE + IRE \tag{14}$$

In this decomposition: the ITE component accounts for the technical inefficiency of the firms actually operating in the industry; the IAE component accounts for the overall effect of the failure to optimize the choice of input vectors by the firms actually operating in the industry; the IRE component corresponds to the effect of the potential reallocation of cost across firms and potential gains obtainable via a reallocation of the industry cost budget across the different segments of the inputs market. A more direct interpretation of this result can be obtained by expressing all these quantities in percentage terms:

$$\%ITE + \%IAE + \%IRE = 1 \tag{15}$$

where $\%ITE = \frac{ITE}{IE}$, $\%IAE = \frac{IAE}{IE}$, $\%IRE = \frac{IRE}{IE}$.

2.2 Counter-factual Analysis.

It is interesting to consider the case in which the researcher is interested in investigating a situation in which input prices change from the observed segmentation $(\mathbf{\Omega}, \mathbf{\Pi})$, to a new one say $(\mathbf{\Omega}_C, \mathbf{\Pi}_C)$ (where the subscript stays for "counter-factual"). The question now becomes how the industry responds to such a change. Assuming that the total cost budget of the industry stays unchanged to the same level, the new optimization problem for the industry is:

$$IE_C = ID_I\left(\mathbf{Q}, \mathbf{\Omega}_C, \mathbf{\Pi}_C, C_I; \mathbf{g}_y\right)$$

It should be noted that it may be the case that $IE_C < IE$ and in fact it could also be negative if counter-factual prices are set to such a level as to make overall potential production lower than the observed one. This counter-factual analysis may inform on changes associated with the elimination of subsidies for some of the production factors in the industry. An example of this counter-factual analysis is given in the empirical section.

3 Empirical Illustration

In this section the methodology will be applied to data on 92 electricity distributors in Ontario, Canada for the year 2011. These data provide information on 3 outputs and 3 inputs. The outputs are: electricity distributed, other products and services and transmission losses. The inputs are labor (in full time equivalent), capital services and other materials services (intermediate inputs). It is worth mentioning the inclusion of transmission losses on the output side. This variable highly correlates with the level of electricity distributed and the transmission losses in the data are of the order of between 1% to 10% of the total electricity distributed. Prices for inputs were obtained from different sources. Monthly wage rates and producer and consumer price indexes were obtained from Statistics Canada. Bond yields on capital (which is a measure of the cost of capital) were obtained from the Bank of Canada. The remaining data were extracted from the Yearbooks of Electricity Distributors published by the Ontario Energy Board (OEB). All price and quantity measures combine to yield the total revenue, total expenses and earnings before interest and taxes reported in the Ontario Energy Board (OEB) yearbooks. Figure A1 reports boxplots of the variables used standardized by the sample mean. It is quite clear from this picture that the Ontario electricity distribution system show quite a high degree of heterogeneity on the size of firms, with a few big firms and a large number of smaller firms. Since the main output produced in the industry is the electricity distributed, this is reflected in the choice of the directional vector: $\mathbf{g}_y = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$. This means that the projection involves an expansion of the quantity of electricity distributed towards the frontier, while keeping the other outputs constant. This also means that the unit of measurement for the inefficiency indicators discussed in this section is the same as the unit of measurement of the electricity distributed. It is unfortunate that data on fixed assets (like plants and machinery) is not available. If such information was available, it was possible to specialize our models to a short term perspective in which reallocation happens only with the variable factors and the fixed assets are kept constant. Due to this data limitation, the results presented below should be interpreted as long run potential gains. In fact, in the long run fixed assets can be reallocated as well.

Table A1 reports the results for the industry inefficiency decomposition (14). The total inefficiency of the system is equal to 1.93E+11 KWh of loss in electricity for the given overall cost budget of the industry. The meaning of this figure is that with the observed overall budget of the industry, if all the inefficiencies were to be eliminated, that very cost would be able to dispatch almost two times the quantity of energy that it is currently distributing. In fact by taking the ratio of actual to potential production gives an inefficiency score of around 0.4, which means the industry is operating well below its potential. In this table it is quite clear that 79.3% of this overall inefficiency is given by a reallocation effect, which implies a transfer of cost across the different firms in the industry in order to optimally exploit scale and scope economies and bring the number of firms operating in the industry to the optimal. The remaining 20.7% of the industry inefficiency is due to the inefficiency of the firms actually operating in the industry. According to equation (6) there are two components that induce this type of inefficiency. The technical inefficiency component (ITE) represents lack of best practices, while the allocative inefficiency component (IAE) represents a failure to choose optimal input quantities given the input prices. It is clear from these figures that the IAE component is the most important component in the overall firm inefficiency. Summing up, the main component in the inefficiency of the industry is a reallocation effect and the second most important component is a failure to optimize with respect to the prevalent input prices.

A more interesting result is the one reported in table A2. Here a counter-factual analysis is conducted based on the following idea. It is known that there is a structure of subsidies and regulations in the Ontario electricity distribution sector which is influencing the choices of the agents operating in it. Without the ambition of doing any justice to the complexity of the system, here two anecdotal facts are used to run a counter-factual analysis. The first has to do with regulation on the cost of capital, with regulation imposing some type of cap on the return on equity. It is well known that this type of regulation provides an incentive for manager to expand the capital stock above the optimal level in order to maintain the correct proportion of return on equity. In other words this type of regulation implies, de facto, a reduction in the price of capital. The other anecdotal stylized fact is that worker unions in the energy sector were able to bargain higher salaries with respect to the average of other sectors, due to the special structure of subsidies of this sector. This implicitly result in a higher than average price of labor in the electricity distribution sector. Based on this anecdotal evidence the counter-factual analysis assumes a scenario in which wages decline by 10% and the cost of capital increases. Table A2 reports the results for such an exercise. The table reports input prices and the associated optimal input quantities. This optimal input quantities are determined as the best response to the prevailing input prices and are the quantities that should be chosen by the average firm in order to make the industry operate at its potential. The table reports how these optimal quantities vary when the input prices are varied from the observed ones to the counter-factual ones (and the cost is kept fixed at the observed one). The capital input is the one that shows the highest percentage reduction (6.5%) followed by labor (6.1%) and intermediate materials (5.5%). This reduction in optimal input quantities induce a reduction in the overall potential output of the industry of 6.7%. It is also interesting to note that the optimal cost shares of the various inputs change dramatically with the capital share increasing from 14.6% to 22.1% and the share of the other two inputs declining. It should be noted that the reduction in the output of the industry is a reduction in the potential output. In fact, since the observed output is well below this optimal, the reduction induced by the counter-factual could be offset by an increase in the inudstry efficiency. Therefore a central planner that would want to relax regulation in the sector could do so, according to this counter-factual analysis, by providing at the same time stronger incentives towards efficiency, so that the reduction in the subsidies would not affect the overall output of the industry.

4 Conclusion

In this paper a definition of industry inefficiency in cost constrained production environments is introduced. This definition makes use of the Indirect Directional Distance Function (IDDF) and quantifies the inefficiency of the industry in terms of overall output loss for a given overall industry cost budget. The methodology proposed in this paper identifies the optimal configuration of the industry and shows how to decompose it into sources components: inefficiencies arising from sub-optimal configuration of the industry (reallocation inefficiency); inefficiency arising from a failure to optimize input quantities given the prevalent inputs prices; inefficiencies due to lack of best practices at the firm level. The method is illustrated using data on Ontario electricity distributors. The data shows that lack of best practices is only a minor component of the overall inefficiency of the sector (less than 10%), with reallocation inefficiency accounting for more than 75% of the overall inefficiency of the system. A counter-factual analysis based on counter-factual inputs prices is conducted in order to illustrate how the model can be used to estimate the effects of a change in the regulation regime.

References

[1] Aparicio J., Pastor J.T. (2012), "Directional distance function and Rate-of-Return Regulation", Advances in Decision Sciences, doi:10.1155/2012/731497

[2] Aparicio J., Pastor J.T., Ray S.C. (2013), "An overall measure of technical inefficiency at the firm and at the industry level: the 'lost profit on outlay' ", European Journal of Operational Research, 226, 154-162.

[3] Asmild, M., Paradi, J.C., Pastor, J.T. (2009), "Centralized Resource Allocation BCC Models", *Omega. The International Journal of Management Science* 37(1), 40-49.

[4] Asmild, M., Paradi, J.C., Pastor, J.T. (2012), "DEA based models for reallocations of police personnel", *OR Spectrum,* 34(4), 921-941.

[5] Banker R.D., Chang H., Natarajan R. (2005), "Productivity change, technical progress and relative efficiency change in the public accounting industry", *Management Science*, 51, 291-304.

[6] Banker R.D., Charnes A., Cooper W.W. (1984), "Some models for estimating technical and scale inefficiencies in data envelopment analysis", *Management Science*, 30, 1078-1092.

[7] Briec W., Dervaux B., Leleu H. (2003), "Aggregation of directional distance functions and industrial efficiency", *Journal of Economics*, 79, 237-261.

[8] Chambers R. G., Chung Y., Färe R. (1996), "Benefit and distance functions", *Journal of Economic Theory*, 70, 407-419.

[9] Chambers R.G., Chung Y., Färe R. (1998), "Profit, directional distance functions and Nerlovian efficiency", *Journal of Optimization Theory and Applications*, 98, 351-364.

[10] Fang, L. (2013), "A generalized DEA model for centralized resource allocation", *European Journal of Operational Research* 228(2), 405-412.

[11] Fang, L., Zhang, C.Q. (2008), "Resource Allocation Based on the DEA Model", *Journal of the Operational Research Society*, 59 1136-1141.

[12] Färe R., Grosskopf S. (1994), "Cost and revenue constrained production", Springer-Verlag.

[13] Färe R., Grosskopf S., Zelenyuk V. (2008), "Aggregation of Nerlovian profit indicator", *Applied Economics Letters*, 15, 845-847.

[14] Färe R., Primont D. (2003), "Luenberger productivity indicators: aggregation across firms", *Journal of Productivity Analysis*, 20, 425-435.

[15] Färe R., Primont D. (2006), "Directional duality theory", *Economic Theory*, 29, 239-247.

[16] Färe R., Zelenyuk V. (2003), "On aggregate Farrell efficiencies", *European Journal of Operational Research*, 146, 615-620.

[17] Giménez-García, V.M., Martínez-Parra, J.L., Buffa, F.P. (2007), "Improving Resource Utilization in Multi-Unit Networked Organizations: The Case of a Spanish Restaurant Chain", *Tourism Management* 28(1), 262-270.

[18] Grosskopf S., Hayes K.J., Taylor L.L., Weber W.L. (1997), "Budget-constrained frontier measures of fiscal equality and efficiency in schooling", *The Review of Economics and Statistics*, 79, 116-124.

[19] Li S.K., Ng Y.C. (1995), "Measuring the productive efficiency of a group of firms", *International Advances in Economic Research*, 1, 377-390.

[20] Lotfi, F.H., Noora, A.A., Johanshahloo, G.R., Gerami, J. Mozaffari, M.R. (2010), "Centralized resource allocation for enhanced Russell models", *Journal of Computational and Applied Mathematics*, 235, 1-10.

[21] Lozano, S., Villa, G. (2004), "Centralized Resource Allocation Using Data Envelopment Analysis", *Journal of Productivity Analysis* 22(1-2), 143-61.

[22] Lozano, S., Villa, G. (2005), "Centralized DEA Models with the Possibility of Downsizing", *Journal of the Operational Research Society* 56, 357-364.

[23] Lozano, S., Villa, G., Adenso-Díaz, B. (2004), "Centralized Target Setting for Regional Recycling Operations Using DEA", *Omega. The International Journal of Management Science* 32(2), 101-110.

[24] Lozano, S., Villa, G., Braennlund, R. (2009), "Centralized Reallocation of Emission Permits using DEA", *European Journal of Operational Research* 193(3), 752-760.

[25] Lozano, S., Villa, G., Canca, D. (2011), "Application of Centralized DEA Approach to Capital Budgeting in Spanish Ports", *Computers & Industrial Engineering,* 60(3), 455-465.

[26] Maindiratta A. (1990), "Largest size-efficient scale and size efficiencies of decision-making units in data envelopment analysis", *Journal of Econometrics*, 46, 57-72.

[27] Mar-Molinero, C., Prior, D., Segovia, M.M., Portillo, F. (2012), "On Centralized Resource Utilization and its Reallocation by using DEA", *Annals of Operations Research*, 221, 273-283, DOI: 10.1007/s10479-012-1083-8.

[28] Peyrache A. (2013), "Industry structural inefficiency and potential gains from mergers and break-ups: a Comprehensive approach", *European Journal of Operational Research*, 230, 422-430.

[29] Ray S.C. (2007), "Are some Indian banks too large? An examination of size efficiency in Indian banking", *Journal of Productivity Analysis*, 27, 41-56

[30] Ray S.C., Hu X. (1997), "On the technically efficient organization of an industry: a study of U.S. airlines", *Journal of Productivity Analysis*, 8, 5-18.

[31] Ray S.C., Mukherjee K. (1998), "Quantity, quality and efficiency for a partially super-additive cost function: Connecticut public schools revisited", *Journal of Productivity Analysis*, 10, 47-62.

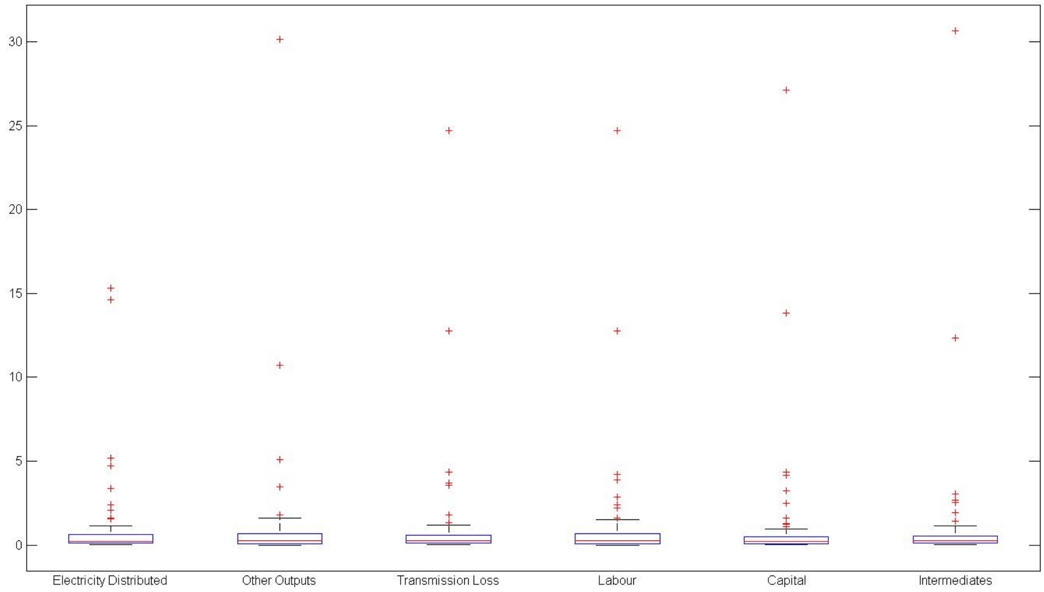
[32] Ray S.C., Chen L., Mukherjee K. (2008), "Input price variation across locations and generalized measure of cost efficiency", *International Journal of Production Economics*, 116, 208-218.

[33] Zelenyuk V. (2006), "Aggregation of Malmquist productivity indexes", *European Journal of Operational Research*, 174, 1076-1086.

Industry Inefficiency Decomposition.

Year	IE	IRE	ICE	IAE	ITE
Levels	1.93E+11	1.53E+11	4.00E+10	2.80E+10	1.20E+10
Percentage		79.3%	20.7%	14.5%	6.2%

A Tables and Figures



B APPENDIX: Optimization programs

The purpose of this appendix is to provide the optimization programs associated with the definitions used in the text in a more explicit way. The firm direct DDF (3) is obtained solving the following linear program:

$$\begin{aligned} D\left(\mathbf{x},\mathbf{y};\mathbf{g}_y\right) = sup_{\beta,\lambda} \quad & \beta \\ s.t.: \quad & \boldsymbol{\lambda}\mathbf{X} \leq \mathbf{x}' \\ & \boldsymbol{\lambda}\mathbf{Y} \geq \mathbf{y}' + \beta\mathbf{g}_y' \\ & \sum_k \lambda_k = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{16}$$

The firm indirect DDF defined in (4) corresponds to the solution of a linear program, which can be explicitly written as follow:

$$\begin{aligned} ID\left(\mathbf{w}/C,\mathbf{y};\mathbf{g}_y\right) = sup_{\beta,\lambda,\mathbf{x}} \quad & \beta \\ s.t.: \quad & \boldsymbol{\lambda}\mathbf{X} \leq \mathbf{x}' \\ & \boldsymbol{\lambda}\mathbf{Y} \geq \mathbf{y}' + \beta\mathbf{g}_y' \\ & \sum_k \lambda_k = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0} \\ & \mathbf{wx} \leq C \end{aligned} \tag{17}$$

It should be noted that while $\mathbf{x}$ is given in equation (16), it is a decision variable in equation (17). Consider now the partition of the input vector and input price vector into two components: $\mathbf{x} = \begin{pmatrix} \boldsymbol{\upsilon} & \boldsymbol{\mu} \end{pmatrix}$, $\mathbf{W} = \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{\Pi} \end{pmatrix}$. The industry inefficiency defined in (10) can be computed using the following mixed

Industry Inefficiency Counterfactual.

	Labour	Capital	Intermediates
Input prices (Observed)	38,500	3.09	178
Input prices (Counterfactual)	34,700	5.00	178
Input Quantities (Optimal for Observed)	1.64E+04	1.06E+08	7.16E+06
Input Quantities (Optimal for Counterfactual)	1.54E+04	9.86E+07	6.76E+06
Percentage reduction in Input quantities	6.1%	6.5%	5.5%
Input Cost shares (Optimal for Observed)	28.3%	14.6%	57.1%
Input Cost shares (Optimal for Counter-factual)	23.9%	22.1%	53.9%

Total Industry Output	Inefficiency (Observed)	Inefficiency (Counterfactual)	Potential Output reduction (Percentage)
1.21E+11	1.93E+11	1.7221E+11	6.7%

integer linear program:

$$
\begin{aligned}
IE = sup_{\boldsymbol{\lambda},\boldsymbol{S},\boldsymbol{\nu},\boldsymbol{\mu}} \quad & \beta \\
s.t.: \quad & \sum_k \lambda_{\mathrm{kj}} \boldsymbol{\upsilon}_k \leq \boldsymbol{\upsilon}_{\mathrm{j}}\,,\ j = 1,\ldots,J \\
& \sum_k \lambda_{\mathrm{kj}} \boldsymbol{\mu}_k \leq \boldsymbol{\mu}_j\,,\ j = 1,\ldots,J \\
& \sum_k \lambda_{\mathrm{kj}} \mathbf{y}_k = \mathbf{y}_j\,,\ j = 1,\ldots,J \\
& \sum_j \mathbf{y}_j \geq \mathbf{Q} + \beta \mathbf{g}_y \\
& \sum_k \lambda_{kj} = S_j \\
& S_j \in \mathbb{N} \\
& \lambda_{\mathrm{kj}} \geq 0\,,\ \forall k, j \\
& \sum_j \left(\boldsymbol{\omega}_j \boldsymbol{\upsilon}_{\mathrm{j}} + \sum_p \boldsymbol{\pi}_p \boldsymbol{\mu}_p^j \right) \leq C \\
& \sum_p \boldsymbol{\mu}_p^j = \boldsymbol{\mu}_j
\end{aligned}
\tag{18}
$$

We note that there are two special cases for which this program can be greatly simplified. The first one is if we consider the MLMC program of Ray et al (2008). In this case the only component in the input partition is $\boldsymbol{\upsilon}$. The program becomes the following mixed integer linear program:

$$
\begin{aligned}
IE = sup_{\boldsymbol{\lambda},\boldsymbol{S},\boldsymbol{\nu},\boldsymbol{\mu}} \quad & \beta \\
s.t.: \quad & \sum_k \lambda_{\mathrm{kj}} \boldsymbol{\upsilon}_k \leq \boldsymbol{\upsilon}_{\mathrm{j}}\,,\ j = 1,\ldots,J \\
& \sum_k \lambda_{\mathrm{kj}} \mathbf{y}_k = \mathbf{y}_j\,,\ j = 1,\ldots,J \\
& \sum_j \mathbf{y}_j \geq \mathbf{Q} + \beta \mathbf{g}_y \\
& \sum_k \lambda_{kj} = S_j \\
& S_j \in \mathbb{N} \\
& \lambda_{\mathrm{kj}} \geq 0\,,\ \forall k, j \\
& \sum_j \boldsymbol{\omega}_j \boldsymbol{\upsilon}_{\mathrm{j}} \leq C
\end{aligned}
\tag{19}
$$

Another simplification can be obtained by assuming that the price vector can be picked independently of the location. In this case we obtain a linear program with only 1 integer variable. In other words location does not matter anymore in this second model. This is a mixed integer linear program (that can be easily solved using its NLP relaxation):

$$
\begin{aligned}
IE = sup_{\boldsymbol{\lambda},\boldsymbol{S},\boldsymbol{\nu},\boldsymbol{\mu}} \quad & \beta \\
s.t.: \quad & \sum_k \lambda_{\mathrm{kj}} \boldsymbol{\mu}_k \le \boldsymbol{\mu}_j \,,\ j = 1, \dots, J \\
& \sum_k \lambda_{\mathrm{kj}} \mathbf{y}_k = \mathbf{y}_j \,,\ j = 1, \dots, J \\
& \sum_j \mathbf{y}_j \ge \mathbf{Q} + \beta \mathbf{g}_y \\
& \sum_k \lambda_{kj} = S_j \\
& \lambda_{\mathrm{kj}} \ge 0 \,,\ \forall k, j \\
& \sum_j \sum_p \boldsymbol{\pi}_p \boldsymbol{\mu}_p^j \le C
\end{aligned}
\tag{20}
$$

That definitions (7) and (8) are equivalent can be shown by considering the sum of S identical firm technology sets:

$$
\begin{aligned}
\Psi + \dots + \Psi &= \left\{ \begin{array}{c} (\mathbf{x}, \mathbf{y}): \quad \lambda_1 \mathbf{X} + \dots + \lambda_S \mathbf{X} \le \mathbf{x}' \\ \lambda_1 \mathbf{Y} + \dots + \lambda_S \mathbf{Y} \ge \mathbf{y}' \\ \sum_s \sum_k \lambda_{sk} = S \,,\ \lambda_{sk} \ge 0 \,,\ \forall s, k \end{array} \right\} = \\
&= \left\{ \begin{array}{c} (\mathbf{x}, \mathbf{y}): \quad (\lambda_1 + \dots + \lambda_S)\, \mathbf{X} \le \mathbf{x}' \\ (\lambda_1 + \dots + \lambda_S)\, \mathbf{Y} \ge \mathbf{y}' \\ \sum_s \sum_k \lambda_{sk} = S \,,\ \lambda_{sk} \ge 0 \,,\ \forall s, k \end{array} \right\}
\end{aligned}
$$