SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

Revisiting the Porter Hypothesis: A Nonparametric Analysis on the impact of Pollution Abatement Technologies on firms 'performances

by

Davide Golinelli

06/2022

SEEDS Working Paper 06/2022
July 2022
By Davide Golinelli

Revisiting the Porter Hypothesis:

A Nonparametric Analysis on the impact of Pollution Abatement Technologies on firms 'performances

Davide Golinelli

Università degli Studi di Ferrara

Abstract

Nonparametric regression models are designed to relax the Gauss-Markov assumptions needed to obtain an unbiased and consistent estimator from the traditional parametric regression. The rationale is to let the function be defined by the data locally, without imposing a linear relationship or higher orders polynomials to fit possible non-linearity, at global level. This paper has the aim to investigate the Porter Hypothesis with the use of nonparametric analysis using kernel regression, in particular the local constant estimator developed by Nadaraya and Watson (1955; 1956) and the linear extension proposed by Stone (1977) and Cleveland (1979). The use of kernel to deal with discrete variables is extremely useful to study the effect of the introduction of pollution abatement technologies, used as a proxy assessing for policy stringency, over the value added and hence to test the effect of regulations on firm's performances: in doing so, starting from the estimator designed by Aitchinson and Aitkens (1976), the extension proposed by Li and Racine (2007) is used. The nonparametric analysis provides a model with a better goodness-of-fit, furthermore the value of the bandwidth referred to the introduction of pollution abatement technology obtained through Kullback-Leibler cross-validation, underlines heterogeneity between groups, and suggesting the positive effect of the introduction of environmental regulation on the performance of firms, leading to the so called Porter hypothesis.

Introduction

In economics, models are used to inquire and study the effect of the relationship between agents and goods in influencing markets mechanisms and social processes; based on available data and techniques, the aim is to explain and/or predict the effects and the probability of occurrence of these events, to combine thereafter the information in an effective and useful way. Due to the high temperatures related to global greenhouse gas emissions, acidification and other forms of pollution that arose in recent years, the necessity to change to new patterns of production and consumption is strongly required, enhancing both government, firms and individuals to act in a more eco-friendly way.

To promote relevant changes proper regulations are needed, aimed at favoring the transition from non-green to green technologies, furthermore different policy measures are likely to affect innovation in different ways (Johnstone et al. 2010).

The relationship between policies and innovation must be studied to understand how changing the way the system is seen, perceived and evaluated by policy-makers could lead to changes in the development and adoption of technologies (Michael Porter, 1991).

Policies are fundamental to change the market structure, in particular to avoid that path dependency and lock-ins that hamper innovation (and in particular environmental innovation). The main types (of policies) are Market-based Regulations and Command-and-Control Instruments. Evidences suggest that market-based instruments (environmental related tax, subsidies or tradeable permits) induce more innovation than command and control regulations (Bergquist et al, 2013; Jaffe et al, 2002).

The focus of this paper is on food processing industry firms characterized by the following research question: would firms' performances increase (or at least not decrease) while augmenting pollution abatement costs?

In the case taken in exam, according to the Porter Hypothesis (PH), increasing pollution abatement costs should result in a win-win situation, thanks to well-designed regulations that have fostered firms to introduce such sort of innovations that lead to profitable results while lowering the environmental impact.

However, it could be difficult for all firms to properly adapt because of different situations arising in the decision processes.

The work is organized as follows: the starting point of the work is defining the main literature behind the Porter Hypothesis, then the econometric method is introduced describing kernel as smoothers used in a regression framework with both continuous and discrete variaables, furthermore the data are presented and the estimates derived. In the end the results are briefly described and discussed followed by the conclusion. The estimations are provided using the statistical software RStudio. In the Appendix it's possible to find the descriptive statistics.

Literature Review

After twenty years the Porter hypothesis (PH) is still a huge theme of discussion.

"Strict environmental regulations do not inevitably hinder competitive advantage against rivals; indeed, they often enhance it." *(Porter, 1991)*

The first definition by Porter (1991) and Porter and Van der Linde (1995) argue the classic assumption about regulations that, forcing the introduction of new standards or taxes, would increase the costs of production, leading the a reallocation of labour and capital to control pollution, and resulting in a loss of efficiency and profitability. In contrast, according to their view, properly defined regulation may favor innovation. Increasing the cost of production, as a consequence of an innovative behaviour (enhanced by proper regulation and aimed at reducing waste and pollution), would result in an increase in the efficiency of the production (cost-efficient), leading to a more profitable and eco-friendly situation.

"Pollution is often a waste of resources and a reduction in pollution may lead to an improvement in the productivity with which resources are used".

Palmer, Oates and Portney (1995) criticized the theory by Porter, assuming that it would not be compatible with the profit-maximization theory of firms. Porter (1991) suggested that properly designed policies should favor the selection of the best "low hanging fruit" that often firms tend to miss.

The PH critiques, led by Jaffe and Palmer(1997), gave birth to three versions of the hypothesis: the weak, strong and narrow Porter hypothesis that are summarized as follows:

WEAK P.H.	*Well-designed environmental regulation may spur innovation*
STRONG P.H.	*In many cases (often) innovation more than offset regulatory costs*
NARROW P.H.	*Flexible regulatory policies give firms greater incentives to innovate*

Table 1: The three versions of Porter hypothesis

These three versions framed the approach proposed by Porter (1991) and Van der Linde (1995) distinguishing in different moments: in the first step the properly designed regulation should foster innovation, as a consequence of this introduction firms should reduce the waste, the pollution and then perform well in the production; moreover, the presence of a flexible regulatory policy should give the firms the incentive to innovate and adapt to the situation.

However, regulation could increase uncertainty related to future investments.

One of the main problems is linked to the decision-making process taken by firm's managers: plenty of literature arguing about managers being risk-averse or rationally bounded (Kennedy, 1994; Aghion, Dewatripont and Rey, 1997; Gable and Sinclair-Desgagné, 1998); furthermore Ambec and Barla (2006) reported about manager present-biased preferences leading to the avoidance of costly investments that gave reward in the long term perspective preferring short-term results.

This aversion could be understood going a little deep in the history of the term and its perception.

Innovation derived from "Novation", first appeared in the thirteen century, meaning "*renewing an obligation by changing a contract for a new debtor"*. The term was not used so much until the twentieth century, presenting two main sides from which there was a strong opposition: political and religious parties. Innovation was identified as a deviation from the common political or religious affairs. Anthropologists in the eighteenth and nineteenth century defined innovation as culture change: innovation was recognized as the change between different cultures and in time related to agriculture, trade, social and political organizations leading to a huge amount of literature focusing on diffusion, borrowing, migration, invasions (Godin, 2008).

Schumpeter (1912:66) defined five different types of innovation:

- Product
- Process
- Market
- Sources
- Organization

Environmental innovations should include product, process and organizational innovation (OECD, 1997b), moreover, being different from other innovations, they decrease the external environmental costs of production or product , leading to what Rennings (1998) called a “double externality effect”.

Environmental regulations are extremely important to lead to the adoption of new products, processes or organizational schemes.

Horbach (2008) define technological opportunities as a third factor, and emphasize how expected government regulations are fundamental in pushing firms to act in a more sustainable way, recalling the predictability as the characteristics of a policy to induce innovation (Horbach, 2012); so, uncertainty about policies could significantly diminish the investments (Pindyck, 2007; Dixit and Pindyck, 1994).

The induced innovation theory, in fact, suggests that the stringency of policies enhance innovation and the reduction of pollution emissions, because firms had an incentive to abate taxation costs in order to reduce prices of the factors (Hicks, 1932). Hicks’s theory is correlated to stringency: greater the possible effect to diminish a certain type of behavior, due to the increasing opportunity costs in the use of factors, greater the incentive to invest in new production/consumption patterns in order to act in a relatively less expensive way.

Stringency and predictability are two main aspects of policies: stringency can be measured as the changes that a regulation induced to a firm, while predictability refers to innovation in reducing the uncertainty and the risks that managers more than often perceive. More specific studies found out that air pollution regulation, in US electric utilities in 1970, significantly increased the age of capital; Gray and Shadbegian (1998) reported that more stringent air and water regulations had a significant impact on paper mills’ technological choice in the US.

Regulations that force firms to produce green products, where the beginners did not have any disadvantages, should help them achieve a Pareto-improving equilibrium. Furthermore, given the public good nature of knowledge, Mohr (2002) reported that when the R&D's investments of a firm are easily captured by its competitors, it will underinvest in such activity (e.g. green technologies).

Ambec and Barla (2002) focusing on information asymmetries find that their presence could be consistent in supporting the PH, this because managers could use some private information opportunistically exaggerating costs of new technologies and being able to extract some rent from those even if the government imposes an environmental regulation.

Domazlicky and Weber (2004) and Weiss (2015), to study technical change as changes in efficiency, used data on toxic releases together with real value added in a non-parametric analysis, concluding that regulation has a negative impact on total factor productivity.

In many cases, research leads to the conclusion that environmental regulation will have a negative impact on firm's performances, however, according to Johnstone and Hasic (2010), different environmental-related taxes could have different attributes: a CO^2 tax would be more flexible and targeted while a value-added tax wouldn't.

According to Rennings and Rammer (2010) competitiveness is not harmed by regulation (according to the PH) and the contribution of environmental policies, in fostering green technologies, should induce innovation to more profitable situations compared to the return on other voluntary innovations.

Gabel and Sinclair-Desgagné (2001; 1999) provided a theoretical analysis of the PH suggesting that environmental regulation should be considered as "*...an industrial policy instrument aimed at increasing the competitiveness of firms, the underlying rationale for this statement being that well-designed environmental regulation could force firms to seek innovation that would turn to be both privately and socially profitable* (Sinclair-Desgagné, 1999)". Then, the increases in production costs should lead to observable effect on a firm's profitability, a situation in which cheap innovation is logically most likely to occur when firms are not near the efficient frontier (Sinclair-Desgagné, 2001). These considerations are important in detecting differences within firms which are related under stringent regulations and investigating which type of government policies are most supportive to avoid organizational failures.

Albrecht (1998b) has focused his analysis on specific industry products such as refrigerators, air conditioning, freezers, air conditioning equipment, fire extinguishers, foams, aerosol etc…, affected by Montreal Protocol (used as product-specific test of the PH). The analysis underlined that countries with more active policies and relatively high pollution-substitution costs could improve their competitiveness and enhance performances.

The Econometric Method

Kernel smoother and Bandwidth selection

In general, econometric analysis started trying to describe the data using parametric estimators, predetermining assumptions on distributions and functions under study. This approach required precise constraints that must be fulfilled to obtain a consistent and unbiased estimator necessary for inference, assumptions that most of the times are not fulfilled. On the other side, non-parametric methods started without any specific assumptions. This method is extremely flexible, being designed to provide a local estimator, rather than global, and providing better goodness-of-fit.

In a nonparametric framework, parameters aren't defined as prior, since their nature and their number are unknown, a subset of infinite dimensional vector spaces determines the parameter: particularly, kernel methods are extremely useful to deal with this situation (Mahmoud, 2019).

The main earlier contributors to the idea of *local polynomial (LP) fitting* were the Italian meteorologist Schiaparelli (1866), American mathematic De Forest (1873) and Danish actuary Gram (1879), then followed by Macauley (1931) (Loader, 1999). The *local regression* method was created starting from the concept of linear regression to be applied to different part of the distribution (intervals) considering every point in the respective interval. Henderson (1924) initialized the work, while Katkovnik (1985), Cleveland and Devlin (1988) significantly developed the related theory (Loader, 1999). In this framework, kernels represented a great tool to deal with the process of smoothing in a nonparametric context, furthermore since it presents some limitations, mainly related to poor approximation and the requirement of restrictive conditions, improvement has been developed based on the same concept (Silverman and Jones, 1989). *Splines*, being pointwise continuous polynomial functions, are a possible example. Moreover, *Kernels* are employed in various types of algorithms (e.g., Support vector machine, supervised learning, gradient descent, etc...) that are being used for different kind of analysis.

Hastie and Loader (1993), in the abstract that introduce a paper on kernel regression, defined kernel smoother as:

"[…] an intuitive estimate of a regression function or conditional expectation; at each point x_o the estimate of $E(Y \mid x_o)$ is a weighted mean of the sample Y_i, with observations close to x_o receiving the largest weights."

(*Hastie and Loader, 1993*)

Henderson (1916) studied the local polynomial regression and the problems related to the selection of the adequate smoothing parameter. Spencer (1904) used the "*15 points rule*" defining a more fitted curve along the distribution with respect to the linear counterpart performing a moving average, however, constraints related to the estimation of the first and last seven points in the distribution should require some extrapolation due to the presence of a low amount of observation in the "*tails*". A general extension of this method is the *k-points moving average*, in which the number of points could vary (Loader 1999). A formalization of this method, the Nearest-Neighbour (k-NN) smoother, arrived in 1951 by Fix and Hodges, in a technical report on discriminant analysis, in which also kernel weighting function were presented (Silverman and Jones, 1989).

Fitting the model to the data lead to a *Bias-Variance trade-off*; Henderson and Shepard (1919), to deal with this balance, defined the y_i observation as

$$y_i = \mu_j + e_j \tag{3.1}$$

where the former (μ_j) is the true value and the latter (e_j) the error or distance from that value. The *graduation rule* was then defined as

$$\hat{y}_j = \sum l_k y_{j+k} = \sum l_k \mu_{j+k} + \sum l_k e_{j+k} \tag{3.2}$$

in which the first graduation component (*systematic component*) should be close to μ_j, while reducing the error as much as possible the associated error term:

$$\sum l_k \mu_{j+k} \approx \mu_j; \tag{3.3}$$

$$\sum l_k e_{j+k} \approx 0, \tag{3.4}$$

Then, the errors were supposed to be uncorrelated and to have all the same variance (σ^2), or the same probable error $\sigma^2 \sum l_k$: this includes the variance reducing factor $\sum l_k$ that measure the reduction in probable error for the graduation rule.

Shorter the graduation rule, slower the systematic error.

The selection of the appropriate length of a graduation rule, or more commonly *bandwidth*, represents a trade-off between systematic and random error.

Introduced in 1952 as smoothing technique, *Kernel methods* define a flexible continuous probability distribution rather than a discontinuous one with a possible extension of the application also to regression estimation (Nadaraya, 1965; Watson, 1964).

The first contribution to kernel weight function is given by Fix and Hodges in 1951, defining the so-called *naïve estimator* (see in figure below) in an unpublished report; then, some year later, Rosenblatt (1956) published the first paper of probability density estimation.

$$\hat{f}(z) = m^{-1} \sum_{i=1}^{m} K_m\left(X_i - z\right), \tag{3.5}$$

Where the kernel K is the uniform probability density (Fix and Hodges, 1951). The main idea was to substitute the frequency approach, giving discontinuous results, allowing the distribution to be smoothed using a *moving average* (Rosenblatt,1956; Parzen, 1962). Rosenblatt (1956) suggested a kernel method to estimate univariate probability density distribution for continuous variables; Parzen (1962) and van Ryzin (1969) were the first to study a general class of consistent and asymptotically normal estimators $\left[P_n(x)\right]$ as a kernel weighted average.

The *Rosenblatt-Parzen density estimator* is defined through the normalization of the weights:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3.6}$$

Kernel density estimation could be also called “*Parzen - Rosenblatt window method*”, used to estimate density function at a specific point ***x*** using neighbouring observations.

The weights are designed from a continuous function $K(\cdot)$ rather than a discontinuous one, with the aim to produce a smooth estimator. The degree of smoothness is defined by the *bandwidth*.

To introduce the kernel estimator in a regression framework were individually Watson (1964) and Nadaraya (1965) with the definition of a local average that was evaluated starting by the *Rosenblatt-Parzen estimator*.

$$\widehat{f}_h(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} \tag{3.7}$$

Being ***K*** a kernel function (weight function) with bandwidth ***h***. The above is known as the Nadaraya-Watson (NW) estimator.

The kernel function is evaluated at a certain design point, that is, then, divided by the sum of the weights, in this way the total will be 1. The “*moving average*” defines the curvature through the slide of a “*window function*” weighting each point proportionally to the height of each kernel above each x coordinate (Watson, 1964; Nadaraya, 1965).

The N-W estimator is a simple weighted average of observations that usually reports boundary bias due to the impossibility to define half of the weights (Racine, 1998).

An alternative to the local constant estimator, such as NW, is the local polynomial estimator. The most common is the *local linear* (LL), introduced by Stone (1977) and Cleveland (1979), and then studied by Fan (1992), Fan and Gijbels (1992; 1996), that performs a local OLS rather than a local average (Li and Racine, 2007). This estimator is better when dealing with linear data and perform better at boundaries.

$$y_i = \alpha + \beta'\left(X_i - x\right) + \varepsilon_i \tag{3.8}$$

LL could be seen as a generalization of the linear OLS estimator, since as $h \longrightarrow \infty$ all the observations receive equal weights, leading to the full-sample OLS estimator $\widehat{g}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$ (Hansen, 2013).

When coming into a confrontation, since both estimators presents the same asymptotic variance, the literature “prefers” the LL estimator, but the choice could be misleading and caution is recommended (Hansen, 2013).

To measure the accuracy and to assess the properties of the kernel the pointwise mean squared error (MSE) is used.

$$\begin{aligned} MSE(\widehat{f}(x)) &= E(\widehat{f}(x) - f(x))^2 \\ &= (E\widehat{f}(x) - f(x))^2 + E(\widehat{f}(x) - E\widehat{f}(x))^2 \\ &= Bias^2(\widehat{f}(x)) + Var(\widehat{f}(x)) \end{aligned} \tag{3.9}$$

The main aspect that must be considered, once defined the MSE, is the bandwidth definition, implying the trade-off between bias and variance (Henderson and Shepard, 1919).

Using a Taylor series expansion[1], Pagan and Ullah (1999; also, Li and Racine, 2007a) showed that it is possible to obtain the approximate variance and bias:

$$var\widehat{f}(x) \approx \frac{f(x)}{nh}\int K^2 z\, dz \qquad\qquad bias\widehat{f}(x) \approx \frac{h^2}{2} f''(x)\, k_2 \tag{3.10}$$

Once defined, the following relationship between bias, variance and bandwidth is straightforward:

BIAS / VARIANCE TRADEOFF		***Bias***	***Variance***
Bandwidth “h”	***Increase***	⬆	⬇
	Decrease	⬇	⬆

Table 2: Relation between the Bandwidth and the Bias/Variance trade-off

[1] The Taylor series expansion is a powerful tool utilizing Taylor series, being determined through an infinite sum of terms expressed as function’s derivatives, to obtain the proper estimate of a function.

What arise is that the trade-off bias/variance is very relevant in the bandwidth selection process, constituting a crucial aspect of nonparametric analysis (Racine, 2008).

The four general approaches to choose the best bandwidth are:

1) *The Reference Rules-of-Thumb*
2) *Plug-in methods*
3) *Cross-Validation methods*
4) *Bootstrap methods*

The Reference Rules-of-Thumb defined the bandwidth starting from the choice of a specific family of distributions and then the sample standard deviation $\hat{\sigma}$ is used. The *Plug-in method*, started in mid-1980s, was introduced by Park and Marron (1990) in trying to provide another selection criterion due to the limits of the least square cross validation, related to undersmoothing and variability, and then proposed also by Sheater and Jones (1991) and Ruppert, Sheater and Wand (1995).

The two approaches that follows, *Cross Validation* (CV) and *Bootstrap method* (BM), are key techniques in what is being defined as Data Mining (DM) and Machine Learning (ML) (DMML), where DM refers to the research of patterns to find the unknown structure of the data through statistical analysis, while ML to the use of these data-driven structures to do inference. These approaches were designed to deal with unknown distributions: their objective is to properly define the span, or bandwidth, being guided by the data itself, investigating through them based on predetermined statistical rules (Clarke et al., 2009). Cross-Validation methods are usually the preferred choice.

The main role of kernel function is to impact the smoothness and differentiability on the resulting estimates, responding to the issue of density estimation next to many other tools, such as histogram, scatter plot, orthogonal series density, nearest neighbour, and projection pursuit density estimation (Ogbeide and Osemwenkhae, 2019).

The selection of the appropriate kernel function is not a relevant issue, while bandwidth selection is a relevant topic (Silverman, 1986). The selection criterion through which the bandwidth is determined could vary with respect to the kind of data to deal with; large sample size would lead to the use of higher order kernel functions that Li and Racine (2007) showed to provide a better convergence rate to determine the true density function (Elamin, 2013).

In this analysis, local constant and local linear estimator are proposed, deriving the bandwidth through the use of Kullback-Leibler cross-validation; to deal with the green capital, the methods proposed by Li and Racine (2004) is performed.

Aitchinson and Aitkens (A-A), in 1976, defined a kernel estimator that could be used to deal with discrete observations. Also, the contribute of van Ryzin and Wang (1978), Hall (1981), Hall and Wand (1988), and Simonoff (1983, 1996) was fundamental.

A-A (1976) were the first to introduce the discrete kernel estimation, while Simonoff (1996) and Tutz (2000) were fundamental in the development; then Li and Racine (2004) extended discrete kernel for categorical data and finite discrete distribution (Kokonendji and Kiesse, 2014).

The A-A function to smooth discrete unordered variables was defined as follow

$$l\left(X_i, x, \lambda\right) = \begin{cases} 1-\lambda & X_i = x \\ \dfrac{\lambda}{c-1} & X_i \neq x \end{cases} \tag{3.11}$$

With λ being the smoothing parameter, or bandwidth, that is bounded in between 0 and $\frac{c-1}{c}$.

Starting from the A-A (1976) kernel estimator developed to determine the joint distribution involving both continuous and discrete data, Li and Racine (2004) proposed an extension useful to determine models that can represent interactions between variables that the determination of different sub cells might inevitably lose. This new estimator is better than the previous counterpart, giving significant results against the once obtained through the frequency estimator.

They started with considering a nonparametric regression model:

$$Y_i = g\left(X_i\right) + u_i \tag{3.12}$$

Being $g(\cdot)$ an unknown function, they performed a non-parametric kernel estimator for both discrete ($x_i^{\,d}$) and continuous variables ($x_i^{\,c}$) using a data-driven bandwidth selection. This type of smoothing over discrete variable significantly increase the out-of-sample predictions with respect to the classical splitting method. (Li and Racine, 2007)

After defining $\mathscr{D} = \prod_{t=1}^{k} \{0, 1, \ldots, c_t - 1\}$ as the range for $X_i^{\,d}$, for $X_i^{\,d}, x^d \in \mathscr{D}$, the *unordered kernel function* defined by Li and Racine was:

$$l\left(X_{t,i}^{\,d}, x_i^{\,d}, \lambda\right) = \begin{cases} 1 & X_{t,i}^{\,d} = x_i^{\,d} \\ \lambda & X_{t,i}^{\,d} \neq x_i^{\,d} \end{cases} \tag{3.13}$$

With $\lambda \in [0, 1]$. It behaves similarly to A-A function when $\lambda = 0$, taking the value 1 with ($X_i = x$), and the value 0 otherwise, as an indicator function, while giving uniform weights when $\lambda = 1$.

They introduced the function W(·) that defines the kernel function associated to continuous variables $x_i^{\,c}$ and ***h*** being the smoothing parameters for the continuous variables. The kernel estimator resulted is

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h,ix} \tag{3.14}$$

Considering the case when the endogenous variable Y_i is continuous, the estimator $g(x) = E\left(Y_i | X_i = x\right)$ is given by

$$\hat{g}(x) = \frac{\int y\hat{v}(y,x)\,dy}{\hat{f}(x)} = \frac{n^{-1}\sum_{i=1}^{n} YK_{h,ix}}{\hat{f}(x)} \tag{3.15}$$

Assigning to the bandwidth the value 0 would lead to the classic frequency estimator and hence to the possible cases in which there could sub-cells without any, or enough, observations to perform an adequate estimation; on the other way, smoothing the discrete variable led to a more efficient situation. The above estimator could be viewed as an extension of the general discrete kernel functions proposed by Ahmad and Cerrito (1994), who underlined the relevance in selecting the smoothing parameters rather than in the choice of the discrete kernel function.

As bandwidth selection criterion the least square cross validation (LSCV) is proposed by Li and Racine and they demonstrated to be quite efficient (Li and Racine, 2001; 2003).

Data description

Data are referred to French food processing industries, provided by different surveys such as the ANTIPOL and the *Enquête Annuelle d' Entreprise* (EAE), that have been merged, resulting in an unbalanced panel of 6145 firms between 1993 and 2007.

The table presenting the descriptive statistics is reported below, while all the graphs are presented in the Appendix at the end.

In average, green capital represents the 2% of the total capital invested, however, it is defined as a dummy having value of 1 if firms introduced pollution abatement technologies and 0 otherwise, useful to classify firms between green and non-green.

Looking at the distributions (reported in the Appendix) it is possible to see a high concentration of firms on lower value, giving distribution more skewed to the right, suggesting a mean value greater than the median value. Logarithmic transformation provides normal distributions. Labour shows a decreasing trend over years suggesting a diminishing number of employees per firm. Green capital and traditional capital, on the contrary, underlined an increase in the same period, with capital being quite stable across firms while reporting some differences in pollution abatement costs (green capital), however the median value showed a stable growth suggesting that most of the firms tend to invest in green capital over time. It is considered as an external variable because it is not known by the producers and gives a reflection of both effects of policies in changing the structure of firms (policy stringency) and the technological push that it gives to them.

VALUE ADDED	Measurement of output, deflated by its annual industry price index; ***Value Added*** in the food processing is low in France; according to Agreste (2017) it accounted for only 18% of turnover of the sector in 2013. It assess for firms' performances.

LABOUR	Measured by the number of employees expressed in annual full-time equivalent. Factor of production.
CAPITAL	Measured as the amount of fixed assets, deflated by the annual price index for capital goods. Factor of Production.
GREEN CAPITAL	Pollution Abatement Technologies are collected annually in a survey conducted by the French Ministry of Agriculture called *Enquête Annuelle sur les Dépenses pour Protéger l' Environment* (ANTIPOL), since the early 1990s. (Huiban et al. 2018). It is a proxy used to assess for policy stringency.

Table 3: Data Description

Moreover, to investigate the effect of green capital on value added the sample is split in two subgroups, once composed of firms that invested in pollution abatement technologies and the other of firms who

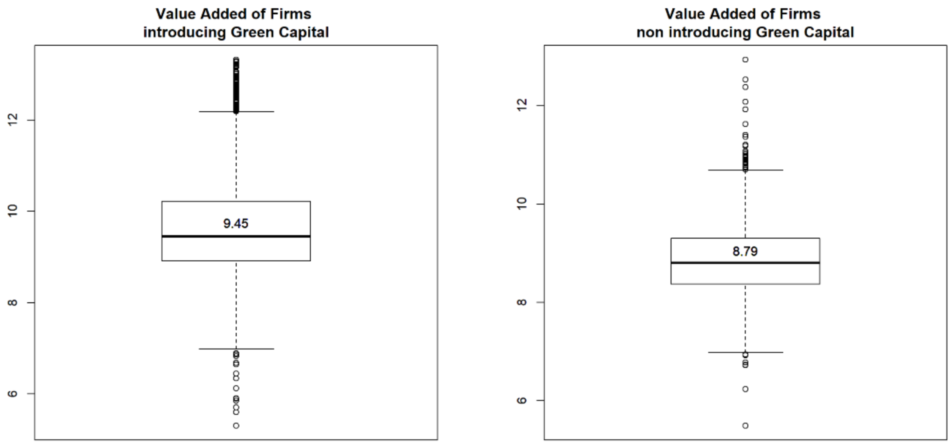


Figure 1: Boxplot of firms investing in green capital compared to firms non investing in it

didn't. It is possible to see the difference in the boxplots presented below, which underlined the fact that the median of value added of "green" firms is greater with respect to other firms not introducing green capital. Moreover, the difference in the 75th percentile remarked the fact that there are more "green" firms performing better, in terms of value added, while "nongreen" firms account for few firms at the same level. Moreover, as shown in figure 1, green capital has seen a progressive in the number of firms introducing it.

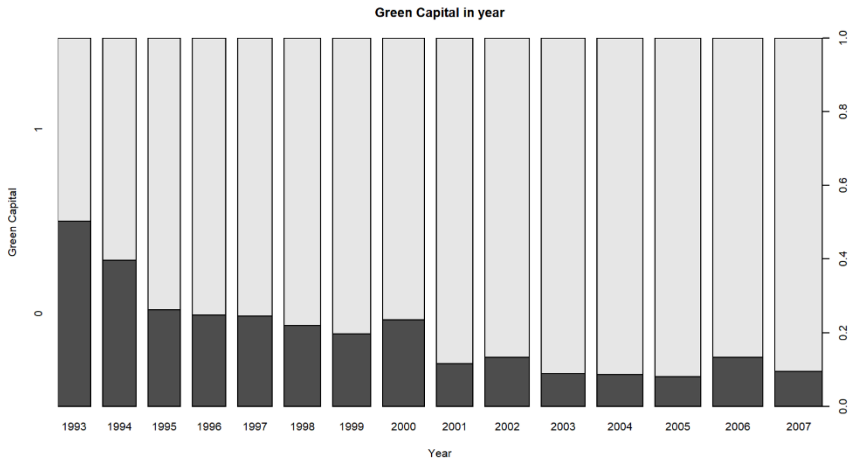


***Figure 2*: Value Added of firms investing in green capital compared to firms non investing in it over time**

Parametric and Non-Parametric Regression

For simplification the data are treated as pooled, without considering the time series dimension and the individual effects.

The parametric specification for regression analysis is the following

$$Y = f\,(X, \beta) + \varepsilon \tag{4.1}$$

The Cobb-Douglas production function is presented both in simple framework as:

$$\ln(VA) = \beta_0 + \widehat{\beta}_1 \cdot \ln(K) + \varepsilon \tag{4.2}$$

$$\ln(VA) = \beta_0 + \widehat{\beta}_1 \cdot \ln(K) + \widehat{\beta}_2 \cdot \ln(K)^2 + \varepsilon \tag{4.3}$$

And in the multiple frameworks as:

$$VA = \beta_0 + \beta_1 \cdot K + \beta_2 \cdot L + \beta_3 \cdot GK \tag{4.4}$$

$$\ln(VA) = \beta_0 + \beta_1 \cdot \ln(K) + \beta_2 \cdot \ln(L) + \beta_3 \cdot GK + \varepsilon \tag{4.5}$$

Where *GK* (Green Capital) is a dummy variable that takes value 1 for firms investing in green capital or value 0 otherwise.

The Nonparametric specification is

$$y = g(x) + u$$
$$g(x) = E(Y|X) \tag{4.6}$$

Allowing for flexibility in defining the model and being determined using kernel weighting functions.

The results obtained from the parametric estimators could be extended into a nonparametric framework, that is presented starting from a simple model, allowing only one independent variable, for example total capital, to deal with the dependent one, value added. The results of simple regression, presented in table 4, show that with respect to the parametric counterpart the nonparametric kernel estimator provide more reliable results having a greater R^2. Then, the analysis is extended to a multiple framework accounting for capital, labour and green capital, the latter defined as discrete, so assuming value of 0 for firms that did not introduce pollution abatement technologies and 1 in the opposite case.

Results and Discussion

The simple models proposed using capital, in a parametric and nonparametric regression, highlighted that the R^2 for the nonparametric counterparts is greater; moreover, since the values of the range defined by the bandwidth let closer point between each other being weighted more, having a low value for the bandwidth suggests heterogeneity between observations, while a high value homogeneity. The results obtained are reported in table 4 refers to the models in which the variable capital enter in level, while table 5 reports the results using the variables in logarithmic specification; the latter shows that quadratic and cubic estimators perform better than linear, and they are quite like kernel regression in the explained variability, in fact the R^2 for both local constant, local linear and polynomials is around 70%, with nonparametric estimator giving a bit more explained variability. The trade-off between variance and bias depends to the value of the bandwidth, in the case of total capital being quite low and suggesting that the effective sample size, being quite low using both the specifications, provides the adequate balance between bias and variance and the best fit to the function.

VARIABLES	***LINEAR***	***QUADRATIC***	***CUBIC***	***LC***	***LL***
Intercept	*9.032* *(0.0108)*	*8.76* *(0.0103)*	*8.58* *(0.0121)*	-	-

K	*1.152E-5* *(1.487E-7)*	*2.25E-5* *(2.71E-7)*	*3.31* *(4.59E-7)*	-	-
K²	-	*-3.17E-11* *(6.9E-13)*	*-1.82E-10* *(2.84E-12)*	-	-
K³	-	-	*1.093E-16* *(3.96E-18)*	-	-
Bandwidth: K	-	-	-	*20.17*	*24.28*
R²	*0.49*	*0.63*	*0.576*	*0.93*	*0.94*
Standard error is reported in brackets					

Table 4: Results of parametric regression –Total Capital

VARIABLES	*LINEAR*	*QUADRATIC*	*CUBIC*	*LC*	*LL*
Intercept	*2.36* *(0.06)*	*12.53* *(0.41)*	*24.02* *(2.4)*	-	-
	(0.006)	*(0.08)*	*(0.727)*		
Log(K)²	-	*0.087* *(0.004)*	*0.455* *(0.072)*	-	-
Log(K)³	-	-	*0.011* *(0.002)*	-	-
Bandwidth Log(K)	-	-	-	*0.0922*	*0.0925*
R²	***0.66***	***0.694***	***0.695***	***0.70***	***0.70***
Standard error is reported in brackets					

Then, the multiple regression model is proposed, both using parametric (linear) and nonparametric (local constant and local linear) approaches; the results for the parametric approach are showed in table 6, highlighting that green capital having low significance, while the R^2 of 0.78 suggests that most of the variability in the regression is explained by the model. Moreover, in table 7 are showed the results obtained from kernel regression using both local constant and local linear estimator.

Table 5: Results of simple regression using polynomials (linear, quadratic, and cubic)

Local constant estimation presents lower bandwidth results with respect to the local linear and greater R^2, particularly using the extension proposed by Li and Racine.

Since the bandwidth is very small for all the variables, especially green capital, both using local constant or local linear estimator, this would suggest that variables are both relevant. Moreover, as discrete values are closed to each other the amount of weight that would receive will be higher since, following the equation 3.11 equal values are weighted receiving the complement of 1: lower the value of the bandwidth, higher the complement (*1-h*) rewarding these observations. On the other side, data far from each other are going to be weighted according to the value of the bandwidth itself when the categories are only two, as in this case (A-A's approach used for unordered variables). The method proposed by Li and Racine (2004), equation 3.13, is somewhat more direct in going to reward similar observations with the classic indicator equal to 1, and the value of the bandwidth otherwise; it is straightforward to observe that as the bandwidth approaches zero, the result is going to be very close to the classical indicator function, while on the other way, being equal to one will led to uniform weights.

In general, as $h \rightarrow 0$, or $\lambda \rightarrow 0$, heterogeneity is present recognizing observations as very different from each other; on the other side, as $h \rightarrow \infty$, or $\lambda \rightarrow 1$, values are considered as homogeneous, underlining no differences between observations, receiving uniform weights (Hall et al. 2004). Cross validation methods are useful since they are perfect in removing the irrelevant variables (Li and Racine, 2007).

Variables	***Coefficient***	***Std. Error***	***p-value*** **>** **|*t*|**
Intercept	*1.67*	*0.05*	*0.12*
Log(Capital)	*0.428*	*0.007*	*2.2e-16****
Log(Labour)	*0.633*	*0.0109*	*2.2e-16****
Green Capital (dummy)	*0.03*	*0.0165*	*0.059.*
R^2 = 0.78			
Significance: 0 (***), 0.001 (**), 0.01 (*), 0.1 (.), 1 ()			

.

Variables	*Multiple Model: Local Constant*	*Multiple Model: Local Linear*
Log(Capital)	*0.122*	*0.143*
Log(Labour)	*0.079*	*0.1*
Green Capital (dummy variable)	*0.078*	*0.049*
R^2	***0.83***	***0.828***

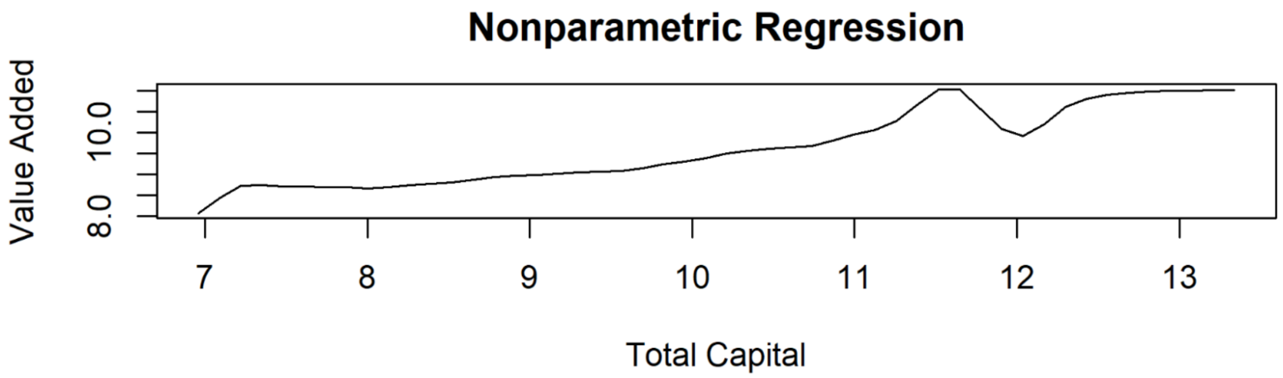
Table 7: Multiple regression comparison between local constant and local linear specifications

Table 6: Results of multiple linear regression

It is possible to see that the values of the bandwidth increase, especially for local linear, indicating that, for the continuous variables, capital and labour, a linear fitting is appropriate.

The bandwidth reports that a linear relation fit well with the continuous variables Capital and Labour since the estimator collapses to the classic OLS. A reminder that nonparametric regression is a generalization of the OLS. While GK, the discrete studied to assess the effect on value added due to the introduction of pollution abatement technologies, being closed to 0.5 for local constant is partially smoothed, but until the value shift to 1, the variable is significant and is not smoothed out. Local linear, on the other side, give a little more relevance with a value of 0.18.

Figure 3 shows how, letting the other variables fixed at the median point, capital, and labour, the two continuous, seems to be defined linearly, labour in particular, so the Cobb-Douglas specification is useful and could provide an adequate linear fit (variables are in log).



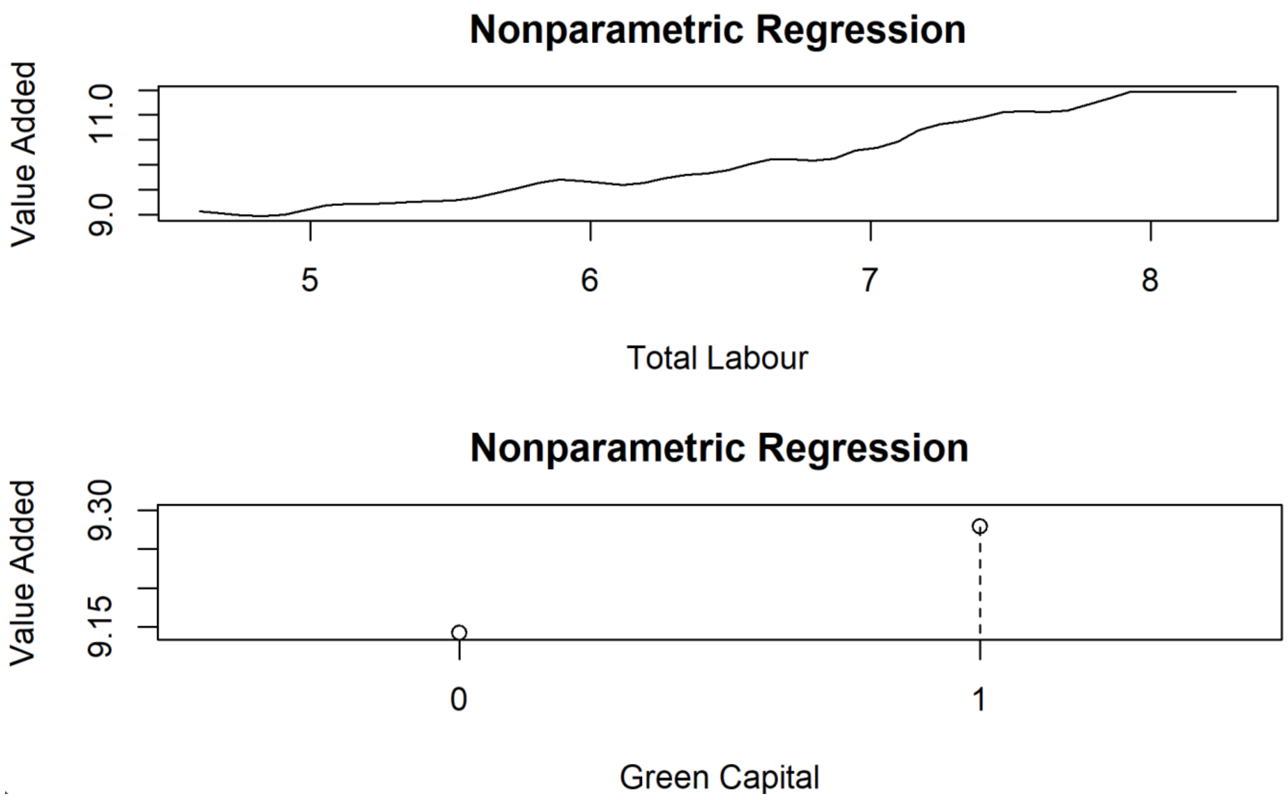


Figure 3: Plot of the model described on table 13 from Local Linear Estimator

The plot obtained using the ***np*** package in RStudio, presented in the figure above, describe the behaviour of the variables under study, capital, and labour, without giving information about the function at every point, but focusing only in a specific one, the median value.

Conclusions

Economic models, such as PF, are extremely useful to control the level of production, analyse potential problems, predict possible results providing some tools to deal with uncertainty.

The use of nonparametric models provide a different tools to deal with such a lack of knowledge. The use of both parametric and nonparametric models, as many other methods, is a powerful tool rather than a mutually exclusive choice that would be inefficient and lacking possible information. Nonparametric methods are an extremely useful extension of parametric ones, and primarily to provide more flexibility once the underlying distribution isn’t known without the imposition of restrictive model requiring a set of assumptions, such as parametric ones. Kernel regression extends the parametric regression at local level, providing an adequate management of the information available without any loss of information as it usually happens when dealing with an indicator function that led to the classic frequency estimator. A problem that is going to lead to a worst situation when introducing categorical variables.

The smoothness done by kernel functions provide a more adequate fit to the function, and the use of data-driven method is useful to study the behaviour of the distribution accounting for the relevant variables, the once that define the amount of local information that must be provided through the window width or bandwidth. The information contained in the windows defined by the bandwidth are the once that are used by the local estimator to obtain the local average or the local regression: having small values will lead to undersmoothing, while large values provoke over smoothing. The method proposed by Li and Racine to deal with both continuous and discrete variables is an extremely useful tool to deal with *the curse of dimensionality*, a problem usually arising while using the frequency estimator when dealing with a mix of these variables, primarily due to the presence of categorical that in defining more cells as the number of categories increase could led to a situation in which data are sparse. The extension proposed, did not separate observations but smoothed them in, or out, depending on the relevance of the variable. In applying kernel functions, the value of the bandwidth is extremely relevant, acting as a litmus paper to recognize homogeneity or heterogeneity among observations.

In the case under study, the bandwidth evaluated through cross-validation method, recognized as the most valuable method to obtain an optimal value that will assure an adequate trade-off between variance and bias, reported values underlining the relevance of green capital as shown in figure 3 (firms that introduced pollution abatement technologies reports higher value added).

The relevance of the bandwidth in smoothing out relevant and irrelevant variables is straightforward, allowing to deal with all the information in an efficient way and lowering the amount of data lost.

Nonparametric regression is extremely useful to deal with data that presented complex distribution that could be described by non-linear functions; next to parametric methods, and in particular polynomials of degree higher than one, to provide some curvature, the use of kernel functions to create weights that smooth the value of the observations will provide a more adequate fit. When the function under study is linear, the bandwidth is going to collapse to infinity saying that the linear fit is appropriate, on the other side, a cross-validation method, such as leave-one-out, will provide the optimal bandwidth that reduce the IMSE, and in doing so, recognizing the variable as relevant, when the bandwidth tends to zero, and irrelevant when tends to infinity; moreover, it is able to smooth out discrete variables, being considered as uniformly weighted, and so assessing for homogeneity rather than to heterogeneity among observations in different groups.

What seems to arose is that capital and labour described a quite linear relation, with capital increasing in the range 7-13 with a small decline around the values 11-12, while labour increase in the range 5-8; both variables are quite flat due to boundary bias (lack of information to obtain reliable estimates).

Green capital seems to confirm the strong version of the Porter hypothesis, since it is possible to record a slightly positive difference in the performance of firms adopting pollution abatement technology due to the introduction of new policies.

It is straightforward to observe that in figure 3, the plot could also led to misinterpretation, since the values are plotted maintaining fixed the other variables at the median point, the relation between the values under study could be different (and most of the time it is) in the other points of the distribution, and the median value could be not relevant especially when the studies are performed to small or large values, or when the median did not represent a reliable information about the sample assessed. [2]

The plot referred to green capital suggests that the introduction of pollution abatement technologies led, in the mean/median value, to higher performances by these firms, recalling the strong Porter hypothesis and suggesting that the introduction of environmental regulations could enhance an innovative behaviour targeted to the adoption, and the adaptation of new technologies and/or new processes to obtain better results in the production process, or at least not worst.

[2] It could be useful a multivariate function's plot able to describe the underlined behaviour across all the distribution.

REFERENCES

- Agreste (2017), *Graphagri France 2017*. Ministère de agriculture et de l'Alimentation, Paris;
- Ahmad, I. A., and Cerrito P. (1994), *Nonparametric estimation of joint discrete-continuous probability densities with application*S, Journal of Statistical Planning and Inference 41: 349-364.
- Aitchison J, Aitken C.G.G., (1976). *Multivariate Binary Discrimination by the Kernel Method,* Biometrika, 63(3), 413–420.
- Alavi, M., & Leidner, D. E. (2001). *Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues*. MIS Quarterly, 25(1), 107–136.
- Albrecht J.A.E., (1998b). *Environmental Costs and Competitiveness. A Product-Specific Test of the Porter Hypothesis,* University of Ghent, Nr.98/50.
- Ambec, S., and Barla, P., (2002). *A theoretical foundation of the Porter hypothesis,* Economics Letters, 75, issue 3, p. 355-360.

- Bergquist, Ann-Kristin, et al., (2013). *Command-and-control revisited: Environmental compliance and technological change in Swedish industry 1970–1990*. Ecological Economics 85: 6-19.
- Bierens, H.J. (1987), *Kernel Estimators of Regression Functions*, Advances in Econometrics: Fifth World Congress, Cambridge University Press, Cambridge, 99-144.
- Blundell and Duncan, (1998), *Kernel Regression in Empirical Microeconomics*, Journal of Human Resources, Winter, 1998, Vol. 33, No. 1, pp. 62-87.
- Cleveland, W.S., (1979), *Robust locally weighted regression and smoothing scatterplots*. J. Amer. Statist. Assoc. 74, 829–836.
- Cleveland, W.S., Delvin, S.J., (1988). *Locally weighted regression: an approach to regression analysis by local fitting*. J. Amer. Statist. Assoc. 83, 579–610.
- Cobb, C.W. and Douglas, P.H., (1928). *A Theory of Production*. American Economic Review, 18, 139-165Czekaj, Henningsen, 2013.
- De Forest, E. L. (1877). *On adjustment formulas*. The Analyst 4, 79–86.
- Domazlicky, B. and Weber, W., (2004), *Does Environmental Protection Lead to Slower Productivity Growth in the Chemical Industry?*, Environmental & Resource Economics, 28, issue 3, p. 301-324
- Elamin, O.A. (2013). *Nonparametric kernel estimation methods for discrete conditional functions in econometrics*.
- Fan, J., (1992). *Design-adaptive nonparametric regression*, J. Amer. Statist. Assoc. 87, 998–1004.
- Fan, J. and Gijbels I., (1992). *Variable bandwidth and local linear regression smoothers*, Annals of Statistics 20, 2006-2036.
- Fan, J. and Gijbels, I., (1996). *Local Polynomial modelling and its applications*, London: Chapman and Hall.
- Fan, Y., (1998), *Goodness-of-fit tests based on kernel density estimators with fixed soothing parameters*, Econometric Theory, 14(5), 604-621..
- Griliches, Z., and Mairesse, J., (1995). *Production Functions: The Search for Identification*, NBER Working Paper No. w5067.
- Godin B., *Innovation the history of a category*, Project on the Innovation History of Innovation, Working Paper No.1, 2008;

- Hall, P., and Titterington, D.M., (1987). *On smoothing sparse multinomial data*, Australian J. Statist. 29(1), 19–37.
- Hall, P., Wolff R. C. L., and Yao Q., (1999): "*Methods for Estimating a Conditional Distribution Function*," Journal of the American Statistical Association, 94, 154–163.
- Hall, P., Racine, J.S., (2004). *Cross-Validation and the estimation of conditional probability densities*, Journal of the American Statistical Association 99 (468), 1015-1026.
- Hansen B. E., (2000). *Econometrics*, Univ. of Wisconsin, Dept. of Econ.
- Hastie, T. J. and Loader C. R. (1993). *Local regression: Automatic kernel carpentry (with discussion)*. Statistical Science 8, 120–143.
- Heady, E. O. & Dillon, J. L. (1961). *Agricultural production functions*. Ames, Iowa: Iowa State University Press.
- Henderson, R. (1916). *Note on graduation by adjusted average*. Transactions of the Actuarial Society of America 17, 43–48.
- Henderson, R. (1924a). *A new method of graduation*. Transactions of the Actuarial Society of America 25, 29–40.
- Henderson, R. (1924b). *Some points in the general theory of graduation*. In Proceedings of the International Mathematical Congress, pp. 815–820.
- Henderson, R. and H. N. Sheppard (1919). *Graduation of Mortality and Other Tables*. New York: Actuarial Society of America.
- Hendry, D.F, (2020). *A short history of Macro-econometric modelling*, Climate Econometrics, Nuffield College, Univ. of Oxford, UK.
- Hicks, J. R., (1937), *Mr Keynes and the Classics: a suggested interpretation*, Econometrica, volume 5, p. 147 – 59.
- Horbach, J., (2008). *Determinants of environmental innovation—New evidence from German panel data sources*. Research policy 37, 163–173.
- Horbach, J., Rammer, C., Rennings, K., (2012). *Determinants of eco-innovations by type of environmental impact — The role of regulatory push/pull, technology push and market pull*, Ecological Economics, Volume 78, pp: 112-122,

- Huiban, J.P., Mastromarco, C., Musolesi, A., Simioni, M., (2018). *Reconciling the Porter Hypothesis with the traditional paradigm about environmental regulation: a nonparametric approach*, Journal of Productivity Analysis, Springer Science, pp. 50:85-100.
- Jaffé, W., (1980). *Walras' Economics as Others See it*, Journal of Economic Literature, American Economic Association, Vol. 18, No.2, pp. 528-549.
- Johnstone N., Haščič I., Kalamova M., (2010). *Environmental Policy Design Characteristics and Technological Innovation: Evidence from Patent Data*, OECD Environment Working Papers No.16;
- Katkovnik, V. Y. (1985). *Nonparametric Identification and Smoothing of Data: Local Approximation Method*. Moscow: Nauka.
- Li, Q., Racine, J., (2001). *Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data*, Journal of Econometrics, Vol. 119, pp. 99-13.
- Li, Q., Racine, J., (2004). *Cross-validated local linear nonparametric regression*. Statist. Sinica 14, 485–512.
- Li, Q., Racine, J., (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Livanis et al, (2009). *A nonparametric kernel representation of the agricultural production function*: *implications for economic measures of technology*, Presented at the 83rd Annual Conference of the Agricultural Economics Society.
- Loader, C., (1999). *Local Regression and Likelihood*, Statistics and Computing, Springer, New York.
- Lucas, R. E., (1976). *Econometric policy evaluation: a critique, in K. Brunner and A. Meltzer (eds), Stabilization of the Domestic and International Economy*, Amsterdam :North-Holland , pp. 7–29 .
- Mahmoud H.F.F., (2019). *Parametric versus Semi and Nonparametric Regression Mo*dels,Cornell University.
- Marshak, J. and Andrews, W. H. (1944). *Random Simultaneous Equations and the Theory of Production*, Econometrica, 12 (3, 4), 143-2.
- Mundlak, Y., (1961). *Empirical Production Function Free of Management Bias*, Journal of Farm Economics, Vol. 43, pp. 44-56.

- Mundlak, Y., and Hoch I., (1965). *Consequences of Alternative Specifications of Cobb-Douglas Production Function*, Econometrica.
- Mussa, H.Y., (2013). *The Aitchison and Aitken Kernel Function Revisited*, Journal of Mathematics Research Vol.5 No.1.
- Nadaraya, E. A., (1964). *"On Estimating Regression*," Theory of Probability and Its Applications, 9, 141–142.
- Ogbeide, M., and Osemwenkhae, J., (2019). *Nonparametric Statistics, Adaptive data kernel density estimation and data mining experiment*, Athens Journal of Science, Vol.6, pp. 287-308.
- Pagan, A., and Ullah, A., (1999). *Nonparametric Econometrics*, Cambridge Books, Cambridge Univ. Press.
- Racine J., (1998). *Bias Corrected Kernel Regression*, North American Summer Meetings of the Econometric Society, Caltech.
- Rajagopalan B. and Lall U., (1995). *A Kernel estimator for discrete distribution*, Journal of Nonparametric Statistics 4:409-42.
- Rennings, K. (1998). *Towards a theory and policy of eco-innovation-neoclassical and (Co-) evolutionary perspectives*. ZEW Discussion Papers, No. 98–24.
- Rennings, K., and Rammer, C., (2011). *The Impact of Regulation-Driven Environmental Innovation on Innovation Success and Firm Performance*, Industry and Innovation, 18:3, 255-283.
- Rosenblatt, M., (1956). *Remarks on some nonparametric estimates of a density function*. Annals of Mathematical Statistics, 27, 832-837.
- Rudemo, M., (1982). *Empirical choice of histograms and kernel density estimators*. Scandinavian Journal of Statistics, 9, 65-78.
- Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). *An eective bandwidth selector for local least squares regression*. Journal of the American Statistical Association 90, 1257.
- Schiaparelli, G. V., (1866). *Sul modo di ricavare la vera espressione delle leggi della natura dalle curve empiricae*, Effemeridi Astronomiche di Milano per l'Arno 857, 3–56.

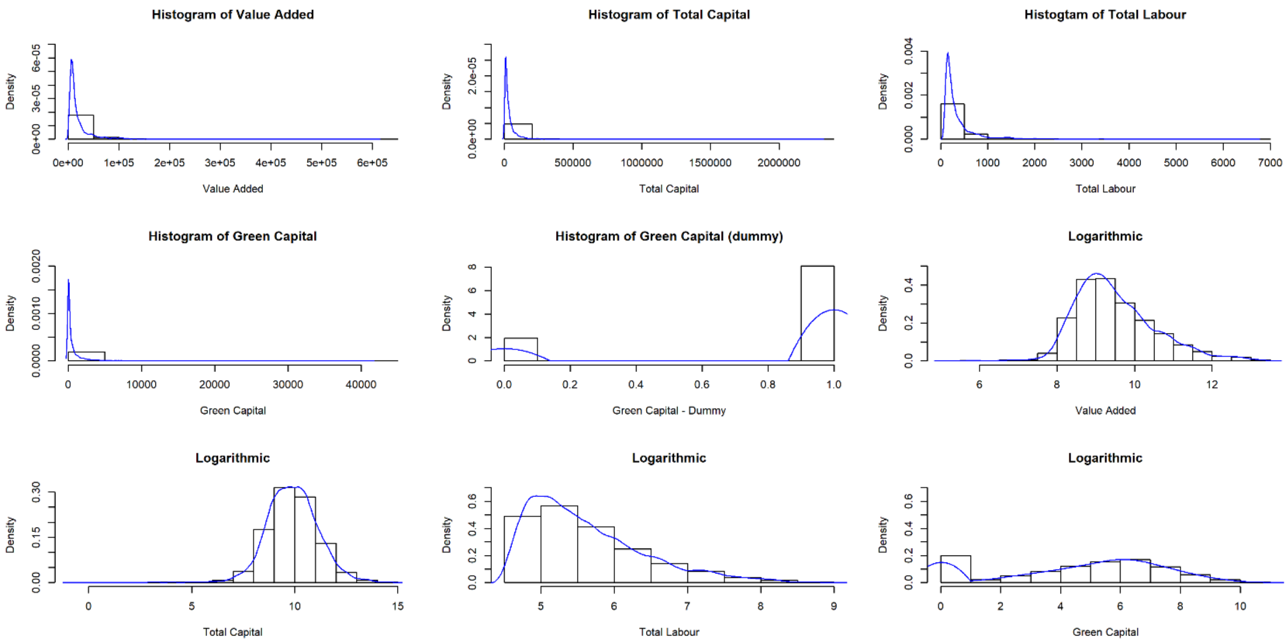
- Sheather, S. J., & Jones, M. C. (1991). *A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation*, Journal of the Royal Statistical Society. Series B (Methodological), 53(3), 683–690.
-
- Schumpeter, J.A., (1954). *History of Economic Analysis*, edited from manuscript by Elizabeth Boody Schumpeter.
- Shultz, H., (1928). *Statistical Laws of Demand and Supply with Special Application to Sugar*, Chicago: University of Chicago Press, I928.
- Silverman B. W. and Jones M. C., *E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951)*, International Statistical Review, Vol.57, No.3, 1989, International Statistical Institute.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- Sinclair-Desgagné, B., (1999). *Remarks on Environmental Regulation, Firm Behavior and Innovation*, CIRANO Working Papers 99s-20.
- Spencer, J. (1904). *On the graduation of rates of sickness and mortality*. Journal of the Institute of Actuaries 38, 334–343.
- Stock J.H., (1989). *Nonparametric Policy Analysis*, Journal of the American Statistical Association, 84, 567-75.
- Stone, R. H., (1954). *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom 1920–1938*, Cambridge: Cambridge University Press.
- Stone, M., (1974). *Cross-Validatory Choice and Assessment of Statistical Predictions*, Journal of the Royal Statistical Society, Ser. B, 39, 44-47.
- Stone, C. J., (1977). "*Consistent Nonparametric Regression*," The Annals of Statistics, 5, 595–645;
- Tutz, G., (1995). *Smoothing for categorical data: discrete kernel regression and local likelihood approaches*, in: H.H. Bock, W. Polasek (Eds.), Data Analysis and Information Systems, Springer, Berlin, pp. 261–271.
- Wand, M. P., and Jones M. C., (1995). *Kernel Smoothing*, London: Chapman and Hall.

- Wang M.C., and van Ryzin J., (1981). *A class of smooth estimator for discrete distribution*, Biometrika, Vol. 68, pp. 301-309.
- Watson, G. S. (1964): "*Smooth Regression Analysis*," Sankhya, 26, 359–372.
- Weiss, J. F. (2015). *Essays on Externalities, Regulation, Institutions, and Firm Performance*. Jönköping International Business School, Jönköping University, Dissertation No, 102.
- Wong, W. H., (1983). *On the Consistency of Cross-Validation in Kernel Nonparametric Regression*, The Annals of Statistics, 11(4), 1136–1141.
- Zamagni, S., (1994). *Economia e Filosofia*, Dipartimento di Scienze Economiche, Bologna.

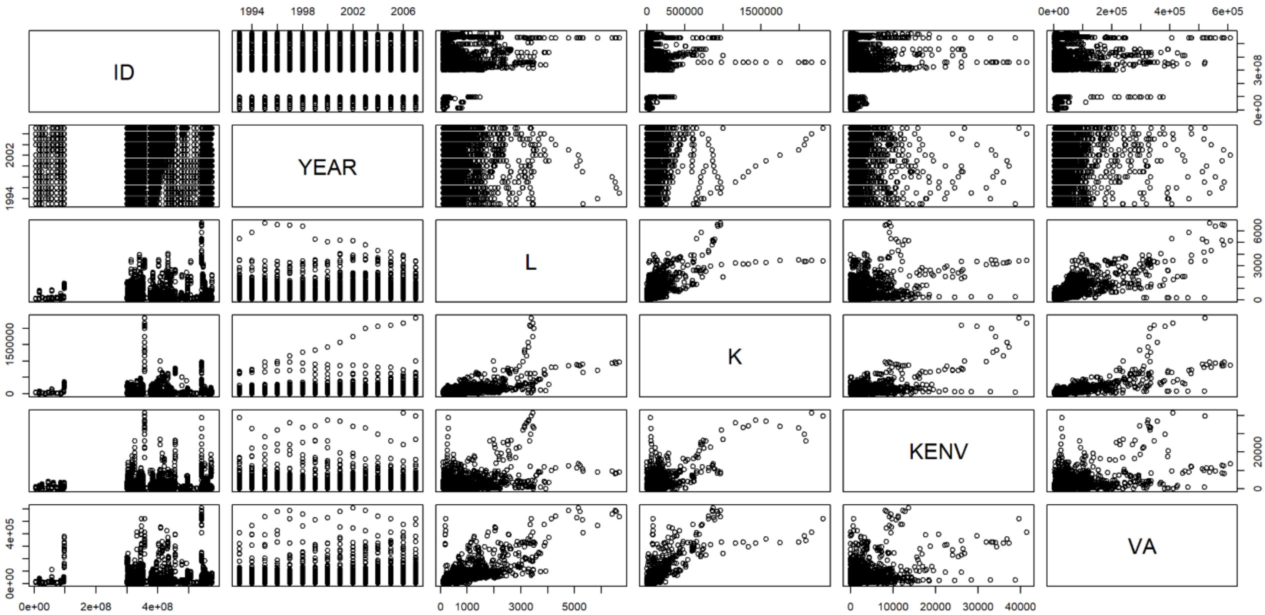
APPENDIX

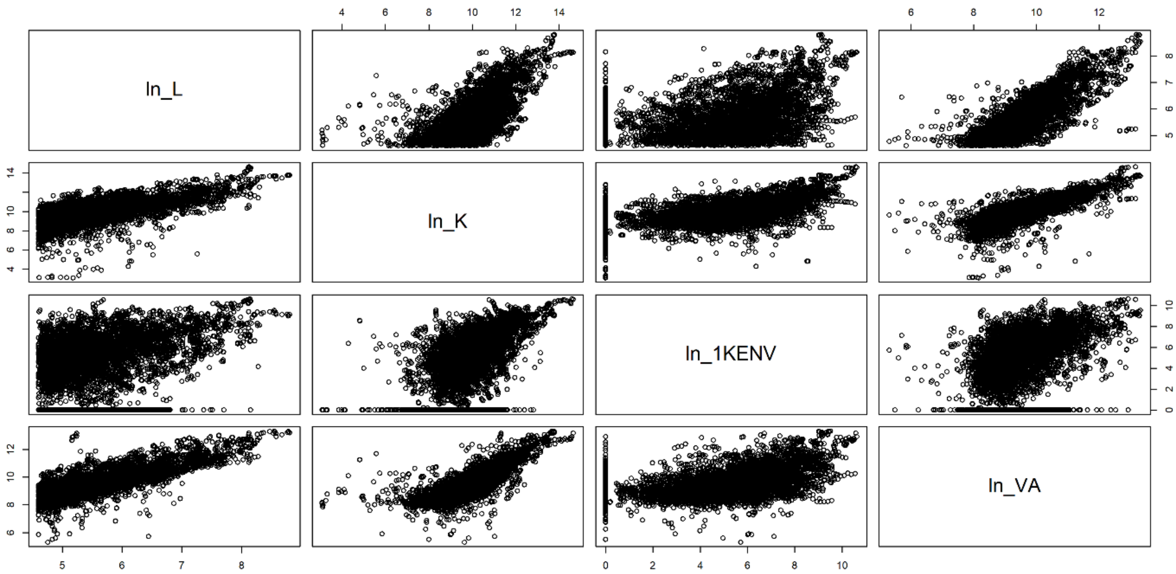
Descriptive Statistics

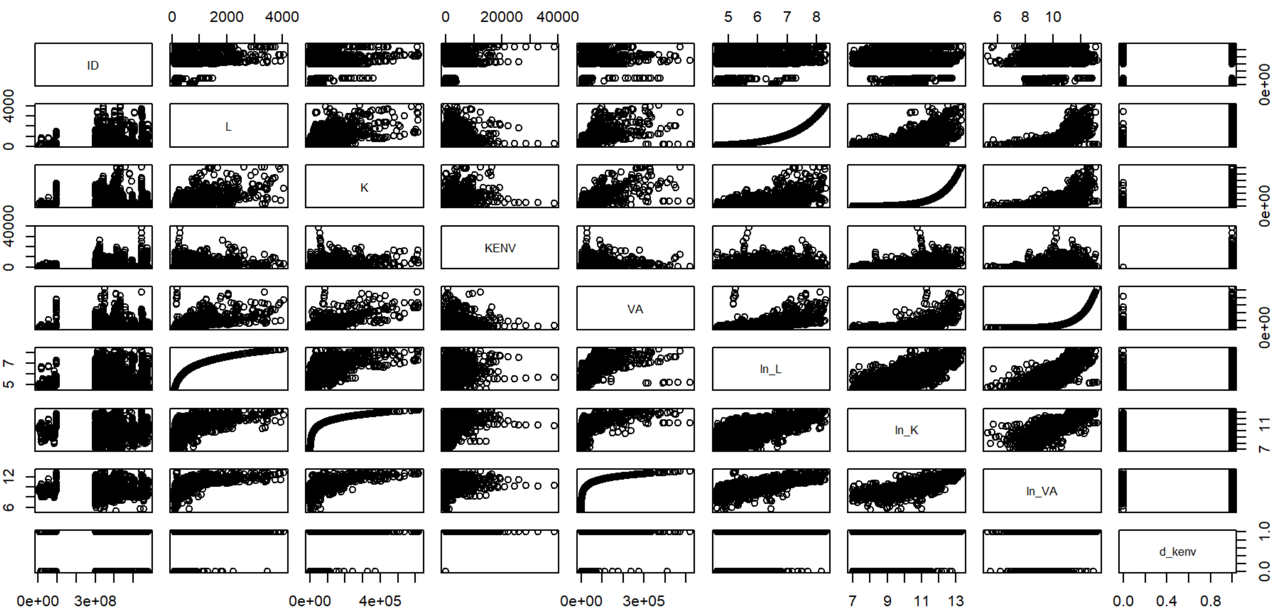
Histograms



Correlation Analysis






ID
L
K
KENV
VA
ln_L
ln_K
ln_VA
d_kenv
0 2000 4000
0 20000 40000
5 6 7 8
6 8 10
0e+00 3e+08
0e+00 4e+05
0e+00 3e+05
7 9 11 13
0.0 0.4 0.8
0 4000
0 40000
5 7
6 12
0e+00
7 11
0.0 1.0