INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0921

INNOVATION AND EXPORT ACTIVITIES IN THE GERMAN MECHANICAL ENGINEERING SECTOR: AN APPLICATION OF TESTING RESTRICTIONS IN PRODUCTION ANALYSIS

SCHUBERT, T. and L. SIMAR

Innovation and Export Activities in the German Mechanical Engineering Sector: An Application of Testing Restrictions in Production Analysis

Torben Schubert
Fraunhofer Institute for Systems and Innovation Research
and Chair of Innovation Economics, Technical University of Berlin
torben.schubert@isi.fraunhofer.de

Léopold Simar*
Institute of Statistics
Université Catholique de Louvain-la-Neuve, Belgium
and Toulouse School of Economics
Université de Toulouse, France
leopold.simar@uclouvain.be

January 19, 2010

In memory of Professor Hariolf Grupp, Chairman of the Institute for Economic Policy Research (IWW), Karlsruhe University, who died after a tragic accident on January 20^{th}, 2009. Although we may not hope to close the void which he left behind he and his work will live on in ours.

*L. Simar acknowledges support from the "Interuniversity Attraction Pole", Phase VI (No. P6/03) of the Belgian Science Policy, from the Helga & Wolfgang Gaul Stiftung, Fakultät für Wirtschaftswissenschaften, Universität Karlsruhe and from the Chair of Excellency "Pierre de Fermat", Région Midi-Pyrénées, France.

Abstract

Since Solow (1956) the economic literature has widely accepted innovation and technological progress as the central drivers of long-term economic growth. From the microeconomic perspective, this has led to the idea that the growth effects on the macroeconomic level should be reflected in greater competitiveness of the firms. Although innovation effort does not always translate into greater competitiveness, it is recognized that innovation, is in an appropriate sense, unique and differs from other inputs like labor or capital. Nonetheless, often this uniqueness is left unspecified. We analyze two arguments rendering innovation special, the first related to partly non-discretionary innovation input levels and the second to the induced increase in the firm's competitiveness on the global market. Methodologically the analysis is based on restriction tests in non-parametric frontier models, where we use and extend tests proposed by Simar and Wilson (2001, 2009). The empirical data is taken from the German Community Innovation Survey 2007 (CIS 2007), where we focus on mechanical engineering firms. Our results are consistent with the explanation of the firms' inability to freely choose the level of innovation inputs. However, we do not find significant evidence that increased innovation activities correspond to an increase in the ability to serve the global market.

Keywords: data envelopment analysis; bootstrap; subsampling; nonparametric efficiency estimation, technical efficiency; production; innovation; exports; CIS; mechanical engineering; Germany; discretionary

JEL Classification: C14, C40, C60, D20, L60, O30

1 Introduction

Innovation has become an important topic in both scientific and public discussion, because it is believed to have laid the basis for the wealth of the Western economies.

This almost unequivocal acceptance of the benefits of innovation probably finds its origins in Solow (1956), who demonstrated that economic growth cannot be sustained by an increase in the supply of regular input factors such as labor or physical capital but only by technological progress. The endogenous growth literature (e.g. Romer 1990, Aghion and Howitt 1992, Grossman and Helpman 1993), finally linked technological progress (in Solow's model treated as exogenous) to firms' decisions on their innovation activities.[1]

But innovation is not only analyzed in its role as a driver of macroeconomic technological progress. On the firm level it is also analyzed as a driver of competitive advantage. If this were true, innovation input should be fundamentally different from other input factors such as capital or labor in some way.

To justify this hypothesis the postulated uniqueness of innovation inputs must be given a clear economic interpretation, where, in this paper, we intend to test two of them. We think that firms are not completely free in choosing the level of innovation input, because the induced organizational changes may result in inertia a firm is unable to easily adapt to. That means that innovation may cause organizational costs on top of the direct. This would lead firms to invest relatively little in innovation. However, firms that are more able to adapt their internal structure can gain a competitive edge by innovating (rather than spending additionally on regular inputs), because the marginal product of innovation input will be higher than that of regular inputs. Another explanation, already known from the literature, implies that innovation raises competitiveness and therefore allows firms enter to global markets enabling them to profit from increasing returns to scale (e.g. Beise-Zee and Rammer 2006).

Knowledge about which of these channels is more important (if any at all) is crucial, because it affects the strategies firms can take to benefit from innovation. If the first explanation were right, firms could gain a competitive edge by implementing an organizational structure that better allows them to cope with technological progress. This calls for an active and efficient innovation and change management.[2] The second approach states that firms

[1] The theoretical results have given new impetus to the research efforts in innovation economics. Special interest was lately paid to the empirical measurement of innovation on the national level (Grupp 1997, Schubert and Grupp 2009, Grupp and Schubert 2009), on the firm level, well described in the OSLO Manual - a handbook that proposes procedures to construct innovation surveys (OECD, 2005), and even at the level of technological artifacts (Grupp and Maital, 1998).

[2] See Penrose (1959) for the role of management capabilities in the innovation process.

should take advantage of innovation by increasing production capacities. It might also imply that firms should seek structures which allow them to serve a global market on a large scale.

The aim of the paper is to test which of the two explanations is consistent with empirical data. We do so using restriction tests in frontier models proposed by Simar and Wilson (2001, 2009).[3] We use data from the German Community Innovation Survey 2007 (CIS 2007) and focus on the mechanical engineering sector as a reference case, because firms here are both export-oriented (export ratio of 75.8% in 2008) and have relatively high innovation expenditures. (Roughly 5% of total turnover were spent on innovation in 2008.) This makes this sector a very attractive candidate for our questions.

The paper is organized as follows. Section 2 gives theoretical and literature background on the two proposed mechanisms by which firms which benefit from innovation. In Section 3 to 4 we present the methodology of testing restrictions based on subsampling used in this paper. We limit the presentation of the test to the case of testing aggregation of inputs (or of outputs), but the procedure can easily be adapted, following the scheme of Simar and Wilson (2001) for testing the relevance of some inputs (or of some outputs). Additionally we present some variations of this methodology, showing that it is also possible to define bootstrap procedures which recover the distribution of the test statistic when the null is true. Following that, in Section 5 we will apply the methods discussed previously to data on mechanical engineering firms in Germany using data from the German Community Innovation Survey 2007 (CIS 2007). Section 6 concludes.

2 The Role of Innovation Input

In the following we will further develop the role of innovation input from a theoretical perspective. Based on these consideration we will derive the research hypotheses, where the first two are based on a theoretical model presented in the next subsection. The third, dealing with the relationship of innovation and export activities, is based on a review of the literature.

2.1 Partly Non-discretionary Innovation Inputs

Following contributions of endogenous growth theory (e.g. Romer 1986), the importance of innovation input is often highlighted by incorporating it as a separate factor in the production

[3]A restriction test analyses whether inputs or outputs can be added up and used as a compound input or output instead of treating each of them separately. For example we could ask whether it is necessary to include innovation expenditures as a separate input or whether it is possible to add up all expenditures (irrespective if related to innovation, physical capital, or labor) and treat this as an aggregate input.

function:[4]

$$Y = f(C, L, A), \tag{1}$$

where Y is a scalar output, C is physical capital, L is labor input, A is innovation input, and f is the production function. Suppose that it fullfils the usual regularity conditions, i.e. positive first partial derivatives, negative second partial derivatives, and non-negative cross derivatives.

However, enthusiasm about innovation would not be justified, if it were just another input. In the following we propose that its uniqueness results from constraints on choosing the level of innovation input. This could be due to the fact that innovation is not a routine task and requires considerable resources on the side of the firm. A firm might need specialized human capital or management capabilities, which it might not have or be unable to buy. It might also need a flexible structure to cope with the organizational inertia following innovations. This implies that firms might be kept away from choosing a higher level of innovation, because their organizational constraints hinder them in doing so.

To illustrate the implications of this within a theoretical framework, consider a profit-maximizing firm. Let us further assume that this firm is small, so its choices will not affect market outcomes. Let p be the price of one unit of output, r be capital costs, w be wages, and $c_i(A)$ be costs per unit of innovation input, which (might) depend on firm i. Additionally, they are allowed to depend on the current level of innovation input through a component that reflects organizational costs induced by innovation. Profit maximization thus implies solving the following problem with respect to the levels of each input:

$$\begin{aligned} \max \pi_i &= pf(C, L, A) - rC - wL - c_i(A)A \\ \text{subject to:} &\quad C, L \geq 0 \end{aligned} \tag{2}$$

In order to keep things simple we will assume for marginal innovation to be determined by $c_i(A) = c_1 + c_{2i}(A)$, where c_1 denotes constant direct marginal costs (e.g. for buying it) and $c_{2i}(A)$ captures the costs that have to be incurred for "incorporating" that marginal unit into the production process. Thus, the latter reflects organizational costs. We assume that $c_{2i}(A)$ is zero for all A smaller than some firm-specific threshold $\overline{A_i}$ and prohibitively large above. Therefore, firms will never use more innovation input than $\overline{A_i}$. For simplicity we

[4] In fact, there are numerous ways of modeling innovation. An common alternative is to represent it as a cost reducing activity. However, defining it as an input is more in line with our empirical frontier approach.

normalize all direct input costs to unity.[5] Furthermore, taking into account the structure of the innovation costs, the maximization problem can be rewritten as:

$$\begin{aligned}\max \pi_i &= pf(C,L,A) - rC - wL - c_1 A \\ &= f(C,L,A) - C - L - A \\ &\text{subject to:} \quad C, L \geq 0 \quad \text{and} \quad A \in [0, \overline{A_i}]\end{aligned} \tag{3}$$

From now on we will assume that the innovation constrained is binding. Under this assumption, firm i will make a constrained choice, which is charakterized by following properties:

$$\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A} > \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial C} = \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial L} = 1, \tag{4}$$

where optimal capital and labor choices of firm i may depend on its innovation constraint. To see why innovation can grant a firm a competitive advantage in this framework, suppose the firm i manages to move up its constraint by a marginal unit without having to incur costs. Using (4) the marginal initial change in profits is given by:[6]

$$\begin{aligned}\frac{\mathrm{d}\pi(C_i^*, L_i^*, \overline{A_i})}{\mathrm{d}A} &= \left(\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial C} - 1\right)\frac{\mathrm{d}C_i^*}{\mathrm{d}A} \\ &+ \left(\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial L} - 1\right)\frac{\mathrm{d}L_i^*}{\mathrm{d}A} + \left(\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A} - 1\right) \\ &= \left(\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A} - 1\right) > 0\end{aligned} \tag{5}$$

Since (5) is true for all $\overline{A_i} < A^*$, the absolute increase in profits associated with a discrete change in the upper bound is also positive. Thus the firm can increase profits by spending more on innovation. Of course this hinges on the assumption that moving upward the constraint is costless, which need not be true, but at least it highlights the mechanism.

Further note from (4) that a firm still equates the marginal products of the discretionary capital and labor expenditures. This means that investing 1 Euro in capital increases the output by the same amount as an investment of 1 Euro in labor inputs, if firms have made

[5]The normalization is without loss of generality, because we can always rescale the units of the input factors appropriately. However, normalization allows easier interpretation, because the terms C, L, and A actually denoting the physical units can now be viewed as expenditures for the respective factor. We adopt this interpretation from now on.

[6]Note that this is a special application of the envelope theorem.

optimal decisions. In other words, at optimal levels of capital and labor input, capital and labor cannot be distinguished from one another. So the inputs can be added up and treated as a compound output. More formally, we will prove the following:

Proposition: Given the concave production function defined in (1), there exists a monotonous and convex function h_1 whose inverse satisfies $f(C_i^*, L_i^*, A) = h_1^{-1}(C_i^* + L_i^*, A)$ for any value A.

We will see in Section 3.2 that it is exactly this property that links implications from the economic model of this section to the restriction tests in production analysis. To prove the proposition, consider the isoquants in $L \times C$ space, which can be calculated from the relationship $\mathrm{d}f(C, L, A)|_{\mathrm{d}A=0} = 0$. We obtain:

$$\frac{\mathrm{d}C}{\mathrm{d}L} = -\frac{\frac{\partial f(C,L,A)}{\partial L}}{\frac{\partial f(C,L,A)}{\partial C}}, \tag{6}$$

which implies a function $C = C(L)$. Further, because the marginal products of capital and labor will be equated by profit-maximizing firms, the ratio in (6) equal -1 for optimally chosen capital and labor expenditures.

Now, consider the tangent of the function $C(L)$ at the optimum. It is a straight line of the form $h_1(Y, A) - L$, where $h_1(Y, A)$ simply is the intercept at the C-axis.[7] We should note here that $h(\cdot)$ is convex, if f is concave, because with any repeated fixed increase in the output level the isoquants will be more and more distant from each other. Further, by construction, $C(L)$ and the tangent are identical in the optimum, i.e.:

$$C_i^* = h_1(Y, A) - L_i^* \tag{7}$$

Rearranging and taking the inverse gives $Y \equiv f(C_i^*, L_i^*, A) = h_1^{-1}(C_i^* + L_i^*, A)$, where h_1^{-1} is concave, because h_1 is convex. This completes the proof of the proposition.

Of course, because A is taken as generic, we can also write $f(C_i^*, L_i^*, \overline{A_i}) = h_1^{-1}(C_i^* + L_i^*, \overline{A_i})$. In terms of testing restrictions, we will therefore expect capital and labor inputs to be aggregable, meaning that they can be added up and treated as a compound input. In fact, the procedures presented in the next sections can test exactly this property. So, we state this as the explicit hypothesis:

[7] h_1 depends on Y , because Y determines uniquely which isoquant it will be tangent to. Writing A explicitly is only to remind us of its presence. Otherwise it is of minor interest.

H1: Labor and capital expenditures can be added up and treated as a compound input.

To avoid confusion, we stress clearly that **H1** only holds at profit-maximizing (i.e. efficient) choices of capital and labor input. It does not follow for arbitrary ratios of capital and labor input.[8] We expect this relationship to be reflected by the efficient frontier estimated from the empirical data.

In any case, it is not possible to aggregate innovation input and capital and/or labor, because the marginal product of innovation is too high. To derive that formally, set the total differential of the production function to zero. Using (4) we obtain:

$$\begin{aligned} \mathrm{d}f(C_i^*, L_i^*, \overline{A_i}) &= \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial C}\mathrm{d}C + \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial L}\mathrm{d}L + \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A}\mathrm{d}A \\ &= \mathrm{d}(C+L) + \frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A}\mathrm{d}A \stackrel{!}{=} 0, \end{aligned} \tag{8}$$

which by (4) implies that $\frac{\mathrm{d}(C+L)}{\mathrm{d}A} = -\frac{\partial f(C_i^*, L_i^*, \overline{A_i})}{\partial A} < -1$. Thus we cannot find a function h_2 such that $\overline{A_i} = h_2(Y) - (C_i^* + L_i^*)$, which is the basis for representing f by some function h_2^{-1} depending on the compound input only. We conclude this subsection with the second hypothesis:

H2: Innovation expenditures and compound non-innovation related capital and labor expenditures cannot be aggregated.

In the empirical Section 5 we will explicitly test these hypotheses and check whether the data from German mechanical engineering is consistent with the implications of this model. Before we present the results we move on by presenting our third hypothesis in the next section.

2.2 The Relationship between Innovation and Export Activities

As seen in the previous subsection, not being able to choose the innovation expenditures as desired can make this input unique.

[8]In fact, the aggregability of capital and labor input would be quite surprising to hold everywhere, because it would imply that 1 Euro spent on labor could be replaced by 1 Euro spent on physical capital, irrespective of the current capital-to-labor ratio. In consequence, **H1** would have the curious implication that a firm could produce with capital only (not having any employees!). This is not sensible for obvious reasons.

A second view considers innovation to be a special input, because it may raise competitiveness, which allow firms to enter global markets and thus to realize economies of scale.[9]

This argument is reflected in the commonly made claim that innovation activities and exports are positively related. Very explicitly, Beise-Zee and Rammer (2006), state in one of their fundamental hypotheses that "innovations are a major determinant for the export performance of firms". Their argument is underlined by theoretical considerations where relative firm efficiency determines market share (Hay and Liu, 1997) and, in consequence, also the possibility to survive in a highly competitive world market. Another strand from the literature - the resource-based view of firms - stresses the importance of intangible resources (e.g. as technological capacity) for competitiveness and export behavior (Rodriguez and Rodriguez, 2005, Braunerhjelm, 1996).

Indeed there is much empirical evidence, emphasizing the positive link between innovation and export activities. Sterlacchini (1999) states that higher innovation activities exert a significantly positive effect on share of exports in sales for Italian manufacturing. Tomiura (2007) demonstrates that there is a positive association between R&D intensity and exports. The same result is shown by Pla-Barber and Alegre (2007) for French firms in the biotech sector. Alvarez (2007) analyzes Chilean manufacturing plants and claims that "technological innovation is positively associated with exporting". A similar result is given by Yang et al. (2004) for the case of Taiwanese SMEs (2004). Lachenmaier and Wossmann (2006) find such results also for Germany.

However, there are also more agnostic voices in the literature. For instance, a more differentiated perspective is taken by Bleaney and Wakelin (2002), who state that innovation activities may only be one of the paths to increase export performance. For non-innovators they find that low-capital intensity and lower unit labor costs are more important drivers. In an earlier paper Wakelin (1998) even claims that "non innovative firms are found to be more likely to export than innovative firms of the same size." For the case of the UK, Harris and Li (2009) emphasize that although R&D activities may reduce entry barriers to export markets, given a market entry has already taken place, R&D no longer affects

[9] Although there is already a broad literature in this topic (which we will review in this section), we regard it as worthwhile to deal with this question once again. The main reason is that all of the mentioned empirical contributions rely on regression-based approaches. Abstracting from the problem of strong function assumptions, we regard it as an even more severe drawback that the estimated relationship not is one that reflects behavior of roughly efficient firms (i.e. firms that are on or close to the frontier). It rather reflects an "average" relationship between exports and innovation, which is estimated based primarily on observations corresponding inefficient firms. However, from an economic point of view, this may be quite misleading, because this relationship need not carry over to the efficient frontier. Even worse, it is not even guaranteed that it is not the very source of inefficiency. The restrictions test we use instead do not have this problem, because they do not rely on expected relationships but on relationships implied by the shape of the frontier.

exporting behavior. Similarly Becchetti and Santoro (2001) find evidence for the claim that technological innovation reduces the chances of creating sale structures abroad.

Despite however the more agnostic voices most economists tend to assume a positive relationship between innovation and export activities, which we summarize in our last hypothesis.

H3: More innovative firms have higher export ratios.

Before we present the empirical results concerning the hypotheses, in the next section we will link the previously presented theoretical framework to the statistical model.

3 The Statistical Model

The theoretical model described in Section 2.1 was based on a production function that was, except for a few quite general assumptions on the first and second partial derivatives, left unspecified. This very general representation was chosen, because in pracical applications the production function f will generally be unknown. Thus, in order to test the hypotheses derived from the theoretical model, the production function needs to be estimated. In what follows we present a nonparametric statistical model and estimation techniques that reflect the general formulation of the theoretical model. Adding to the presented economic model the statistical model will allow firms to be inefficient, i.e. they can fall short of the production function defined in Equation (1).

3.1 Basic notations

Formally, if $x \in \mathbb{R}_+^p$ is a vector of inputs and $y \in \mathbb{R}_+^q$ is a vector of output, the production set (or the technically attainable set) can be defined as

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}. \tag{9}$$

To be coherent with the economic model described above, we assume that Ψ is convex. The Farrell-Debreu efficiency score (Debreu 1951, Farrell 1957) for a point (x_0, y_0) is determined by the distance from (x_0, y_0) to the efficient frontier of the attainable set, defined as

$$\Psi^\partial = \left\{(x, y) \in \Psi \mid \left(\gamma x, \gamma^{-1} y\right) \notin \Psi \text{ for any } \gamma < 1\right\}. \tag{10}$$

The boundary Ψ^∂ of Ψ constitutes the **technology**. In principal, the production function given in Equation (1) is just another way of expressing this technology in mathematical terms. The input-oriented Farrell-Debreu efficiency measure $\theta(x_0, y_0)$ is "radial", the efficiency of a point (x_0, y_0) is defined in terms of how much all input quantities can be proportionately reduced so that the output levels can still be produced, i.e. are not beyond the boundary of Ψ; formally,

$$\theta(x_0, y_0) = \inf\{\theta \geq 0 \mid (\theta x_0, y_0) \in \Psi\}. \tag{11}$$

Similarly, the output-oriented Farrell-Debreu efficiency measure for the point (x_0, y_0) is defined by

$$\lambda(x_0, y_0) = \sup\{\lambda \geq 0 \mid (x_0, \lambda y_0) \in \Psi\}. \tag{12}$$

By construction, $\theta(x_0, y_0) \leq 1$ is the proportionate reduction of inputs this unit should perform to achieve (input) efficiency. If $\theta(x_0, y_0) = 1$, the unit is on the efficient frontier Ψ^∂ of Ψ. Similarly, $\lambda(x_0, y_0)$ gives the maximum, proportionate, feasible expansion of y_0, holding input quantities x_0 fixed, with $\lambda(x_0, y_0) = 1$ if $(x_0, y_0) \in \Psi^\partial$.

Since Ψ (and hence Ψ^∂) is unknown, it must be estimated from a sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ of data on firms' input and output quantities. Nonparametric envelopment estimators have been very popular in the efficiency literature because they rely on very few assumptions on the DGP. A DGP $\mathbb{P}$ is thus characterized by the definition of Ψ (which depends of course on the choice of the inputs X and of the outputs Y) and the probability law that generates $(X, Y) \in \Psi$ (e.g. its density $f(x, y)$).

Under the assumption of convexity, we obtain the DEA estimator of Ψ as the convex free disposal hull of $\mathcal{X}_n$ (Farrell 1957, Charnes et al. 1978):

$$\begin{aligned}\widehat{\Psi}_{DEA} = \quad & \{(x,y) \in \mathbb{R}_+^{p+q} | y \leq \sum_{i=1}^n \gamma_i y_i; x \geq \sum_{i=1}^n \gamma_i x_i \text{ for } (\gamma_1, \ldots, \gamma_n), \\ & \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0, i = 1, \ldots, n\}.\end{aligned} \tag{13}$$

Substituting $\widehat{\Psi}_{DEA}$ for Ψ in (11) yields the DEA estimator of $\theta(x_0, y_0)$ given by

$$\begin{aligned}\widehat{\theta}(x_0, y_0) = \min_{\theta, \gamma_1, \ldots, \gamma_n} & \Big\{\theta > 0 \mid y_0 \leq \sum_{i=1}^n \gamma_i Y_i,\ \theta x_0 \geq \sum_{i=1}^n \gamma_i X_i, \\ & \sum_{i=1}^n \gamma_i = 1,\ \gamma_i \geq 0 \ \forall\, i = 1,\ \ldots,\ n\Big\}.\end{aligned} \tag{14}$$

Similarly, the DEA estimator of $\lambda(x_0, y_0)$ is given by

$$\begin{aligned}\widehat{\lambda}(x_0, y_0) = \max_{\lambda, \gamma_1, \ldots, \gamma_n} & \Big\{\lambda > 0 \mid \lambda y_0 \leq \sum_{i=1}^n \gamma_i Y_i,\ x_0 \geq \sum_{i=1}^n \gamma_i X_i, \\ & \sum_{i=1}^n \gamma_i = 1,\ \gamma_i \geq 0 \ \forall\, i = 1,\ \ldots,\ n\Big\}.\end{aligned} \tag{15}$$

The properties of these estimators are well known today (for a recent survey see Simar and Wilson, 2008 and the references therein). To summarize, under technical regularity conditions (smoothness of the frontier surface and of the density of (X, Y),...), these estimators have nice asymptotic properties: consistency and regular limiting distribution. Although the limiting distribution is too complicated to allow conventional statistical inference in practice, the results are important for proving the consistency of appropriate bootstrap techniques (for details see Kneip, Simar and Wilson, 2008).

3.2 Testing restrictions

Suppose that we want to test the hypothesis that the technology is such that some of the inputs can be aggregated (the case where outputs could be aggregated follows the same lines and is summarized in Appendix A). This corresponds to the case where the technologically efficient isoquants in the space of the inputs to be aggregated are parallel hyperplanes (if only 2 inputs are considered, the isoquants are parallel lines with negative slope equal to -1 in the plane $x^1 \times x^2$).

We can now formalize the approach described in Simar and Wilson (2001). Let $x = (x^1, x^2)$ where $x^2 \in \mathbb{R}^r_+$ is the vector of the inputs we are considering to aggregate. We denote $x^+ = i_r' x^2 \in \mathbb{R}_+$ the resulting aggregated input.[10] So, we would like to solve the following test problem

$$\begin{aligned} H_0 &: x^2 \text{ can be aggregated in } x^+, \\ H_1 &: x^2 \text{ cannot be aggregated in } x^+. \end{aligned}$$

H_0 implies the peculiar form of the technology described above, where only the sum of the elements of x^2 is relevant (and the input space can be reduced to $(p - r + 1)$ dimensions) and H_1 specifies that at least some component of x^2 are really individually relevant (so we need to keep the p-dimensional input x). Now, for any given point $(x, y) = (x^1, x^2, y) \in \Psi$

[10] As in Simar and Wilson (2001), we make here the implicit assumptions that the units of the aggregated inputs are the same. If this is not the case, it is some appropriate linear combinations of the inputs that should be considered. The procedure explained here could then be easily adapted to this case: we would simply replace $x^+ = i_r' x^2$ by $x^a = a' x^2$ where $a \in \mathbb{R}^r$ is given.

we define

$$\begin{aligned}\lambda(x,y) &= \sup\{\lambda | (x^1, x^2, \lambda y) \in \Psi\} \\ &= \sup\{\lambda | \lambda \in A_{(x,y)}\}\end{aligned} \tag{16}$$

$$\begin{aligned}\lambda_0(x,y) &= \sup\{\lambda | (x^1, u, \lambda y) \in \Psi, \text{ with } u \in \mathbb{R}^r_+, \\ &\qquad (x^1, u, y) \in \Psi \text{ and } i_r' u = x^+\} \\ &= \sup\{\lambda | \lambda \in A_{0,(x,y)}\}.\end{aligned} \tag{17}$$

Clearly $\lambda \in A_{(x,y)} \Rightarrow \lambda \in A_{0,(x,y)}$, so that $A_{(x,y)} \subseteq A_{0,(x,y)}$. Therefore, since we search for a supremum for all $(x,y) \in \Psi$, we have in general $1 \leq \lambda(x,y) \leq \lambda_0(x,y)$. In addition, we have the following basic inequalities

$$\begin{aligned}&\text{if } H_0 \text{ is true }, \ 1 \leq \lambda(x,y) = \lambda_0(x,y), \text{ for all } (x,y) \in \Psi \\ &\text{if } H_1 \text{ is true }, \ 1 \leq \lambda(x,y) < \lambda_0(x,y), \text{ for some } (x,y) \in \Psi.\end{aligned} \tag{18}$$

Note that from a sample $\mathcal{X}_n$, these two quantities are estimated by

$$\begin{aligned}\widehat{\lambda}(x,y) = \max\Big\{\lambda \ \ | \ \ \lambda y \leq \sum_{i=1}^n \gamma_i Y_i, \ x^1 \geq \sum_{i=1}^n \gamma_i X_i^1, \ x^2 \geq \sum_{i=1}^n \gamma_i X_i^2 \\ \sum_{i=1}^n \gamma_i = 1, \ \gamma_i \geq 0 \ \forall \ i = 1, \ \ldots, \ n\Big\}\end{aligned} \tag{19}$$

and

$$\begin{aligned}\widehat{\lambda}_0(x,y) = \max\Big\{\lambda \ \ | \ \ \lambda y \leq \sum_{i=1}^n \gamma_i Y_i, \ x^1 \geq \sum_{i=1}^n \gamma_i X_i^1, \ x^+ \geq \sum_{i=1}^n \gamma_i X_i^+, \\ \sum_{i=1}^n \gamma_i = 1, \ \gamma_i \geq 0 \ \forall \ i = 1, \ \ldots, \ n\Big\},\end{aligned} \tag{20}$$

where $X_i^+ = i_r' X_i^2$ for $i = 1, \ldots, n$. For the same reasons as above, for all $(x,y) \in \widehat{\Psi}_{DEA}$,

$$1 \leq \widehat{\lambda}(x,y) \leq \widehat{\lambda}_0(x,y). \tag{21}$$

Denote by $\mathbb{P}_0$ the restricted DGPs where the null hypothesis is true and $\mathbb{P}_1$ its complement, so we have $\mathbb{P}_0 \cap \mathbb{P}_1 = \emptyset$ and $\mathbb{P} = \mathbb{P}_0 \cup \mathbb{P}_1$. In fact we want to test $H_0 \ : \ P \in \mathbb{P}_0$ versus $H_1 \ : \ P \in \mathbb{P}_1$. Consider now a particular model $P \in \mathbb{P}$ and the model characteristic $t(P)$ defined as

$$t(P) = E\left(\frac{\lambda_0(X,Y)}{\lambda(X,Y)} - 1\right). \tag{22}$$

Due to the basic inequalities (18) discussed above, we have $t(P) \geq 0$ for all $P \in \mathbb{P}$, but $t(P) = 0$ if $P \in \mathbb{P}_0$ and $t(P) > 0$ if $P \in \mathbb{P}_1$.

A consistent estimator of $t(P)$ is easy to derive. Consider the sample empirical mean in place of the expectation and replace the unknown λs by their DEA estimates. We obtain:

$$t_n(\mathcal{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\widehat{\lambda}_0(X_i, Y_i)}{\widehat{\lambda}(X_i, Y_i)} - 1 \right). \tag{23}$$

We know by construction, see (21), that $t_n(\mathcal{X}_n) \geq 0$. But since it is a consistent estimator of $t(P)$, small values of $t_n(\mathcal{X}_n)$ would not give evidence that H_0 is false, but large values of $t_n(\mathcal{X}_n)$ would lead us to reject the null. The question is "how large is large enough?". So we should either compute the p-value of H_0 given by $\text{Prob}\big(t_n(\mathcal{X}_n) \geq t_{n,\text{obs}} | H_0\big)$, where $t_{n,\text{obs}}$ is the observed value of $t_n(\mathcal{X}_n)$, or find a critical value $t_n^c(1-\alpha)$ such that $\text{Prob}\big(t_n(\mathcal{X}_n) \geq t_n^c(1-\alpha) | H_0\big) = \alpha$, to obtain a test of size α. In both cases we need the sampling distribution of the test statistics $t_n(\mathcal{X}_n)$ under the null hypothesis, or at least, some reasonable approximation, which the bootstrap can provide.

There are several ways of implementing bootstrap methods in frontier models. However, not all of them will work well in our setup. The bootstrap proposed in Simar and Wilson (1998), and adapted in Simar and Wilson (2001) to our testing problem, is known as the homogeneous bootstrap and usually provides a reasonable approximation of the sampling distribution (certified by numerous Monte Carlo experiments). However it relies on a very restrictive assumption that, for the output orientation, the distribution of the inefficiencies $\lambda(X, Y)$ is homogeneous (constant) whatever the level of the inputs X and the value of the output mix (the ray of Y) are. This assumption simplifies the implementation of the bootstrap and is similar to the assumption of homoscedasticity in regression analysis. If input orientation is used, we need the same assumption for the distribution of $\theta(X, Y)$, which has to be the same for all values of the output level Y, and for all value of the input mix (the ray of X).

Even if in some specific problems this homogeneity assumption may be a reasonable approximation, it is certainly restrictive. In addition, when analyses in both output and input orientations are required, the assumptions cannot hold simultaneously in both directions.[11]

The double smoothed bootstrap of Kneip et al.(2008), allowing heterogeneity, cannot be used in our setting, because it is dependent on the point of interest where $\widehat{\lambda}(x_0, y_0)$ is

[11] As a simple probability exercise, the reader will easily verify that if $Y = X \exp(-V)$ where $X \sim \text{Unif}(0, 1)$ and $V \sim \text{Expo}(\eta)$ independent of X, we have indeed homogeneity in the output direction but not in the input direction. As another example, if (X, Y) are uniformly distributed on $\Psi = \{(x, y) | 0 \leq y \leq x \leq 1\}$, the homogeneity condition is wrong in both directions.

evaluated, but we need a bootstrap valid for all the data points considered simultaneously. [12] So we do not have any other choice than using the subsampling approach.

4 Subsampling

The bootstraps obtained by drawing m observations out of the n with and without replacement have the same asymptotic behavior, with $m \to \infty$ and $m/n \to 0$ when $n \to \infty$. Kneip et al. (2008) have proved their consistency for estimating individual efficiency scores. The selection of the appropriate subsamples size m in practice is important, because in finite samples the accuracy of the obtained approximation depends on this choice.

Simar and Wilson (2009) have investigated this issue and show in intensive Monte Carlo experiments that data-driven procedures coming from the literature (e.g. Politis et al., 2001 and Bickel and Sakov, 2008) can be adapted to the setup of frontier estimation using DEA estimators. The procedure exploits the duality between confidence intervals and testing, and allows us to build consistent tests with desired asymptotic size, without imposing the null when generating the subsamples. We will see however that in our testing problem, we will also be able to generate the subsamples under the null.

4.1 The Main Idea

The population parameter of interest is $t(P)$, as defined in (22). It discriminates between H_0 where $t(P) = 0$ when $P \in \mathbb{P}_0$ and H_1 where $t(P) > 0$ when $P \in \mathbb{P}_1$. Its estimate is our test statistics $t_n(\mathcal{X}_n)$ given in (23), where by construction $t_n(\mathcal{X}_n) \geq 0$.

Suppose first we would know the (asymptotic) sampling distribution of $\tau_n(t_n(\mathcal{X}_n) - t(P))$, where τ_n is the rate of convergence of $t_n(\mathcal{X}_n)$ to $t(P)$,

$$G_n(z) = \text{Prob}_P(\tau_n(t_n(\mathcal{X}_n) - t(P)) \leq z), \tag{24}$$

then we would obtain $t_n^c(1-\alpha)$, the upper $(1-\alpha)$ quantile of this distribution, and a one-sided confidence interval of level $(1-\alpha)$ for $t(P)$, would be given by

$$CI = [t_n(\mathcal{X}_n) - (1/\tau_n)t_n^c(1-\alpha), \infty). \tag{25}$$

In this case, we would reject the null $H_0 : t(P) = 0$ if $t_n(\mathcal{X}_n) - (1/\tau_n)t_n^c(1-\alpha) > 0$ (because CI would not not contain 0). So the rejection region of H_0 would be $\tau_n t_n(\mathcal{X}_n) \geq t_n^c(1-\alpha)$.

Of course, we do not know the sampling distribution of $\tau_n(t_n(\mathcal{X}_n) - t(P))$ (and its quantiles) but we can approximate it by an appropriate bootstrap or subsampling distribution.

[12] Remember that $t_n(\mathcal{X}_n)$ depends on $\widehat{\lambda}(X_1, Y_1), \ldots, \widehat{\lambda}(X_n, Y_n)))$.

Since under the null, $t(P) = 0$, a suitable approximation will be given by the following algorithm suggested by Politis et al. (2001). We construct the $N_m = \binom{n}{m}$ subsets $\mathcal{X}^*_{m,b}$, $b = 1, \ldots, N_m$, of size m from $\mathcal{X}_n$, the sampling distribution of $\tau_n t_n(\mathcal{X}_n)$ is approximated by

$$\widehat{G}_{m,n}(z) = \frac{1}{N_m} \sum_{b=1}^{N_m} \mathbb{I}\Big(\tau_m t_{m,b}(\mathcal{X}^*_{m,b}) \leq z)\Big), \tag{26}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $t_{m,b}(\mathcal{X}^*_{m,b})$ is the version of the test statistics applied to the sample $\mathcal{X}^*_{m,b} = \{X^*_{1,b}, \ldots, X^*_{m,b}\}$.[13] The critical value is given by the $(1-\alpha)$ quantile of $\widehat{G}_{m,n}(z)$ given by

$$\hat{t}^c_m(1-\alpha) = \inf\{z | \widehat{G}_{m,n}(z) \geq 1-\alpha\}. \tag{27}$$

The nominal level α test rejects H_0 if and only if $\tau_n t_n(\mathcal{X}_n) \geq \hat{t}^c_m(1-\alpha)$.

Under some regularity conditions on the sampling distribution of $t_n(\mathcal{X}_n)$ under the null and because $t_n(\mathcal{X}_n)$ is a consistent estimator of $t(P)$ for all $P \in \mathbb{P}$, the test is asymptotically of size α and consistent (the probability of rejecting H_0 tends to 1 as $n \to \infty$ for all $P \in \mathbb{P}_1$) as far as $m \to \infty$ such that $m/n \to 0$ as $n \to \infty$ with $\liminf_n n/m > 1$ (see Theorem 3.1 in Politis et al., 2001). This results relies on the condition that under the null, $P \in \mathbb{P}_0$, the scaled statistics $\tau_n t_n(\mathcal{X}_n)$ has a regular asymptotic distribution and that for all $P \in \mathbb{P}$, $t_n(\mathcal{X}_n) \xrightarrow{P} t(P)$. In Appendix B we show how this is achieved by adapting the arguments developed in Simar and Wilson (2009) and by choosing $\tau_n = \sqrt{n}\, n^{2/(p+q+1)}$.

- **Practical choice of the subsample siz** m

The subsampling procedures are consistent for any subsample size m, such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$, but choosing m in practice may have considerable influence on the test decision in finite samples. The crucial question then is how to choose the "optimal" subsample size. Let us consider the trade-off that is governed by the choice of m. When doing the bootstrap (subsampling with replacement), if m is too large, we are not far from the naive bootstrap and we know that inference for individual inefficiency measures will be inconsistent. When subsampling without replacement, values of m too close to n will result in values of t_m too close to t_n with too small variation. However, in both cases, if we choose m to be too small, then we discard too much information. In consequence, an intermediate value for the size of the subsamples should be optimal. A first rule to find that optimal value for m when using DEA estimators was proposed by Simar and Wilson (2008) who

[13] In practice, of course, we do not compute all these subsets, but we would just take a random selection of B such subsamples, where B should not be too small.

recommend to choose a subsample size such that resulting distribution of the inefficiency measures, or some of its features, are stable with respect to the variation of m (this is in fact the rule proposed in a more general setup by Politis et al., 2001 and by Bickel and Sakov, 2008). The Monte-Carlo experiments done by Simar and Wilson (2009) in a similar testing problem to ours (testing returns to scale or testing convexity of Ψ) indicates that this data-driven procedure for choosing m does not only give tests with the appropriate size, but also tests with desirable power properties (rejecting H_0 with high probabilities when H_0 is false).

In the application below, we will first select a grid of values for $m \in [m_{min}, m_{max}]$, where $m_{min} = [n^{\kappa_1}]$ and $m_{max} = [n^{\kappa_2}]$ with $0 < \kappa_1 < \kappa_2 < 1$, then we evaluate for the given sample $\mathcal{X}_n$ the quantity of interest, i.e. $\hat{t}^c_m(1-\alpha)$ for each m. The visual inspection of the plot of $\hat{t}^c_m(1-\alpha)$ against m indicates where the results are stable as a function of m. This gives the order of magnitude for a reasonable choice of m. This simple plot will be mostly sufficient in our application, but the data-driven procedure (described below) allows to fix this choice in a more objective way.

4.2 Subsampling under the null

The subsampling above is a bootstrap where the draws are done without replacement. It is clear that the asymptotic theory, and so the procedure described above, remain valid if the draws are done with replacement. Simar and Wilson (2009) found in their Monte-Carlo experiments that subsampling without replacement performed better than subsampling with replacement. Interestingly, in our setup here, we could also draw random samples $\mathcal{X}^*_{m,b}$, $b = 1, \ldots, B$ (with or without replacement) by imposing the null hypothesis. By doing so, we can hope to improve the performance of the test both for the achieved level and for the power.

In order to implement this, consider the random label of the observation that has been drawn. Call this i^*. We can define the random subsample as follows: $(X,Y)^*_{m,b} = (X^1_{i^*}, X^{2,*}, Y_{i^*})$ where $X^{2,*}$ is generated component by component as follows (we denote by $X^{+,*}$ the sum of the elements of $X^2_{i^*}$):

$$
\begin{aligned}
X^{(1),2,*} &\sim \text{Uniform}(0, X^{+,*}) \\
X^{(j),2,*} &\sim \text{Uniform}(0, X^{+,*} - \sum_{k=1}^{j-1} X^{(k),2,*}), \text{ for } j = 2, \ldots, r.
\end{aligned}
\tag{28}
$$

In this case, by the subsampling, we can immediately estimate the p-value of H_0.

$$\begin{aligned} p\text{-value} &= \text{Prob}\big(t_n(\mathcal{X}_n) \geq t_{n,\text{obs}})|H_0\big) \\ &\approx \text{Prob}\big(t_{m,b}(\mathcal{X}_m^*) \geq (n/m)^{2/(p+q+1)} t_{n,\text{obs}}|H_0\big) \\ &\approx \frac{\#\big(t_{m,b}(\mathcal{X}_{m,b}^*) \geq (n/m)^{2/(p+q+1)} t_{n,\text{obs}}\big)}{B}. \end{aligned} \tag{29}$$

In our application, it appeared that this approach gave very similar results to the general subsampling algorithm described in Section 4.1, at least in terms of the decision relative to H_0.

5 Innovation in Mechanical Engineering

In the following, we will test **H1-H3** using the methods described in Sections 3 and 4.[14] However, before we do so, we will shortly present the data source.

• The setup and the dataset

The data used in this paper comes from the German CIS 2007. The Community Innovation Survey is a large European survey conducted every two years in all member states of the EU as well as some outside the EU. It is harmonized across the participating countries. Concerning the data structure, CIS is a moving cross-section, comprising a yearly sample size of about 5,500 firms (participation rate of about 20%) from all major sectors including, besides manufacturing firms, also provision of services. The main focus of the survey is on the innovative activities, including product and process innovations and lately also marketing and organizational innovations, but contains many questions which are deemed to be relevant for the entrepreneurial innovation process, such as market conditions, public subsidies, hampering factors, and firm-specific features.

In order to base our analyses on a homogeneous population, we focus on mechanical engineering firms (NACE-Code 29). A further reason lies in the sector's orientation both towards exports and innovation.[15] In total the implied sample contains information on 408 firms, where 215 firms have provided answers for *all* variables used in this paper. Table 1, presenting summary statistics, is based on these 215 firms in order to keep certain defining relations (e.g. mean of domestic turnover plus mean of exports must equal mean of total

[14] All the computations have been done by based on own code provided by the FEAR package of Wilson (2008).

[15] Despite the fact that we limit our analysis to mechanical engineering, there is still great heterogeneity. For example, the smallest firm only has 5 employees, while the largest has as many as 34,000.

turnover). It should be noted, however, that the subsequently described tests are based on subsets of these variables. Thus the test samples can be larger (sizes given below).

Table 1: Summary statistics for key variables in Mln. euros (firms from mechanical engineering)

Variable	mean	std.
Exp. for equipment and materials	88.61	712.01
Exp. for personnel	19.05	70.67
Exp. for innovation	4.66	26.24
Exp. for equipment and personnel (excl. innovation expenses)	103.00	733.90
Turnover	81.03	339.65
Domestic turnover	33.49	143.99
Exports	47.54	211.70

Referring to the discussion of Section 2.1 materials and equipment shall reflect expenditures for physical capital C while personnel expenditures of course reflect expenditures for labor L.

- **The methodology**

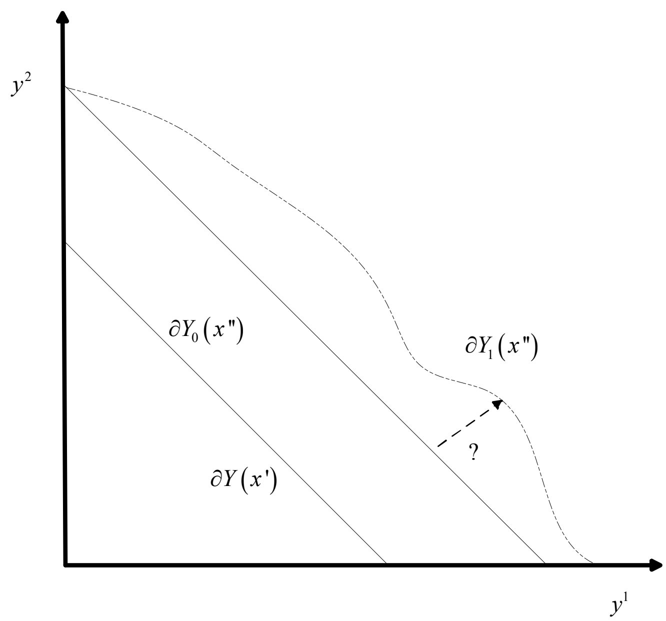
Building on the previously described restriction tests we define a sequence of tests, which we run in a step-wise setup. We will shortly outline how this allows us to test **H1**-**H3**. This is summarized also in Table 2.

*Testing **H1**:* To test **H1** we define a preliminary model of the production process (Model 1). Inspired by the discussion in Section 2 we assume that a firm's inputs are given by expenses for materials and expenditures for personnel. The outputs are taken to be domestic turnover and exports, where we first test whether we can aggregate the outputs. Since this test (see Table 3) does not reject, we can simplify our model and add up domestic turnover and exports. Based on this simplification we define Model 2 and test **H1** explicitly by considering the null-hypothesis that material and labor expenditures can be aggregated.

*Testing **H2**:* Since we do not find evidence against **H1**, we feel comfortable to aggregate capital and labor cost and based on this define Model 3. However, to test **H2** we split up total expenditures (equaling the sum of material plus labor expenditures) into innovation-related and non-innovation-related expenditures. If **H2** was true, we should reject the null-hypothesis that both can be aggregated. In fact, we find that the null can be rejected.

*Testing **H3**:* **H3** postulates a relationship between innovation and export activities. Because of the lack of parameters in frontier models it does not seem obvious on first sight how

Figure 1: Isoquants in output-space (one input and two outputs)



restriction tests can be used to test for an association between the two variables (something that would be obvious using a regression).

However, consider Model 4, where we use innovation and non-innovation expenditures as inputs and domestic turnover and exports as outputs. We test the null that the outputs can be aggregated. As will turn out the null of aggregability of outputs is equivalent to the null-hypotheses that there is no relation between innovation and export activities.

To see this, we have to reconsider the geometry of the isoquants in output space. Under H_0 in Model 4, i.e. aggregability of domestic turnover and exports, the isoquants are a straight line with slope of -1. [16] Any change of inputs therefore corresponds just to a parallel shift of this curve, just like the shift from $\partial Y(x')$ to $\partial Y_0(x'')$, where in Figure 1 we assumed $x'' > x'$. With a more than one-dimensional input this clearly implies that increasing the inputs ($x'' > x'$ should be understood component-wise with a strict inequality for at least one element) does not have a relatively stronger impact on export than it does on domestic turnover, because, under the null, they can be aggregated. In other words, aggregability of domestic turnover and exports for all levels of input implies that innovation expenditures do

[16] In contrast to the inputs in Section 2 this relationship should not be subject to some optimizing behavior, because structure of turnover (i.e. whether domestic or not) is primarily decided on the customer.

not have a promoting effect for exports. On the contrary, if they had, then we would either expect the slope or the curvature of the isoquant to change, both being departures from H_0. In Figure 1 this would correspond to a shift from $\partial Y(x')$ to e.g. $\partial Y_1(x'')$.

Table 2: Summary of the tests

Model/Test	Inputs	Outputs	H_0
1	Expenses for equipment and materials / expenses for personnel	Domestic turnover / exports	The outputs can be aggregated
2	Expenses for equipment and materials / expenses for personnel	Turnover	The inputs can be aggregated
3	Total expenses excluding innovation expenses / innovation expenses	Turnover	The inputs can be aggregated
4	Total expenses excluding innovation expenses / innovation expenses	Domestic turnover / exports	The outputs can be aggregated

• **The results**

This section presents the estimation results. For each of the tests we provide three versions. The first is based on the homogeneous bootstrap (presented for comparison only) The remaining two are based on subsampling, one imposing the null and one not imposing the null.

As already said, a first issue is the selection of the subsample size in the case of the latter two tests. The results are given in the following two tables, where the vertical line indicates the optimal sample size and the horizontal line gives the original value of the test statistic.

More specifically, Figure 2 illustrates for each of the four tests described in Table 2, how the 95%-quantile of $t_{m,b}(\mathcal{X}^*_{m,b})$ varies as a function of m. Likewise, Figure 3 displays the 95%-quantiles for the same four tests, obtained by sampling under the null.

The empirical selection rule suggested by Politis et al. (2001) determines the value of m associated to the smallest volatility of the result of interest. The volatility index can be measured, e.g., by the standard deviation of the results on a running window going from

$m-2$ to $m+2$, m being the running value where the volatility is evaluated.[17] For the four tests depicted in Figure 2, these values of m are 198 ($n = 232$), 206 ($n = 232$), 181 ($n = 235$) and 183 ($n = 215$) respectively. Concerning the results of Figure 3, the automatic selection rule gives values of 200, 204, 180, and 151. These values seem to correspond quite well with regions we would identify as being stable by visual inspection.

As far as the decision of rejection or non-rejection H_0 is concerned, in all four tests, the decision is clear by looking to both pictures and we observe, in particular when subsampling under the Null, that the decision is not very sensitive to the choice of m. In Figure 2 and 3, since the observed values of $t_n(\mathcal{X}_n)$ are 0.1772, 0.1295, 8.3189, and 0.3036, we see that for the Tests 1,2 and 4 we do not reject H_0 but for Test 3, we reject H_0 for all values of m.

The results for all of the tests are summarized in Table 3.

Table 3: Test Results (lowest possible Rejection Levels).

Test	homogeneous Bootstrap (Sampling under H_0)	heterogeneous Bootstrap (Subsampling)	heterogeneous Bootstrap (Subsampling under H_0)	Decision
1	0.33	0.07 ($m = 198$)	0.37 ($m = 200$)	Do not reject H_0
2	0.06	0.12 ($m = 206$)	0.90 ($m = 204$)	Do not reject H_0
3	0.00	0.00 ($m = 181$)	0.00 ($m = 180$)	Reject H_0
4	0.80	0.13 ($m = 183$)	0.60 ($m = 151$)	Do not reject H_0

In the case of Test 1 all the 3 bootstraps do not reject the Null at the usual level of 5%. The same is true for Test 2, but the homogeneous bootstrap is not far from rejection, but since it is based on very restrictive assumptions, it is better not to reject the Null. We see that the pictures and the decisions are even much clearer when using the subsampling under the Null (which is more appropriate when available), since the probabilities in the table are real p-values. The preliminary Model 1 suggests that there is no considerable evidence that domestic turnover and exports a separate output dimension. Furthermore, In line with **H1** we also do not find significant evidence against the null-hypothesis that capital and labor inputs can be aggregated.

Additionally, looking at the results of Test 3, it becomes clear that regular expenses substantially differ from expenses for innovation. The hypothesis that both kinds of inputs might be aggregated is rejected by all three tests at levels being computationally zero. Thus

[17] In order to reduce the computational burden, Figures 2 and 3 are based on calculations at every second m. Thus the window size corresponds to choosing the immediate neighbors for calculating the volatility index.

innovation expenses reflect something different than regular expenses. This corroborates **H2**. Taking together to decisions on **H1** and **H2**, we can say that the data is consistent with the implications of a model which treats innovation as not completely discretionary input.

Turning to the claim that benefits from innovation are channeled through greater export orientation, the non-rejection of Test 4 does not corroborate this view. There is no apparent association between innovation and export activities. Thus, we cannot confirm **H3**.

Note again, that in the four Tests, the bootstrap by subsampling under the Null seems to provide, as expected, much more clear results.

Figure 2: 95%-quantiles of $\{t_{m,b}(\mathcal{X}^*_{m,b})|b = 1, \ldots, 2000\}$ for heterogeneous bootstrap based on subsampling without replacement, as a function of m.

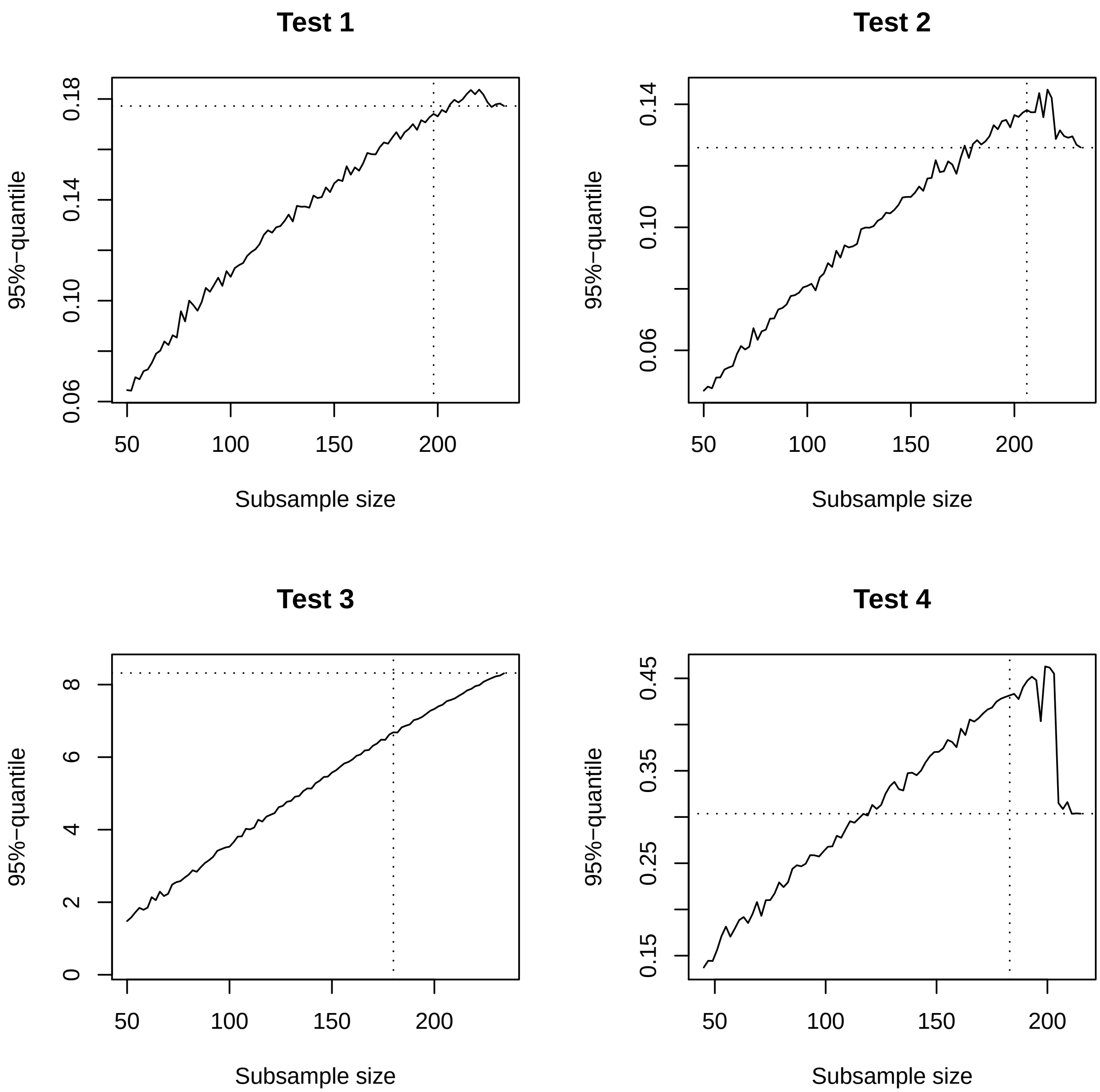
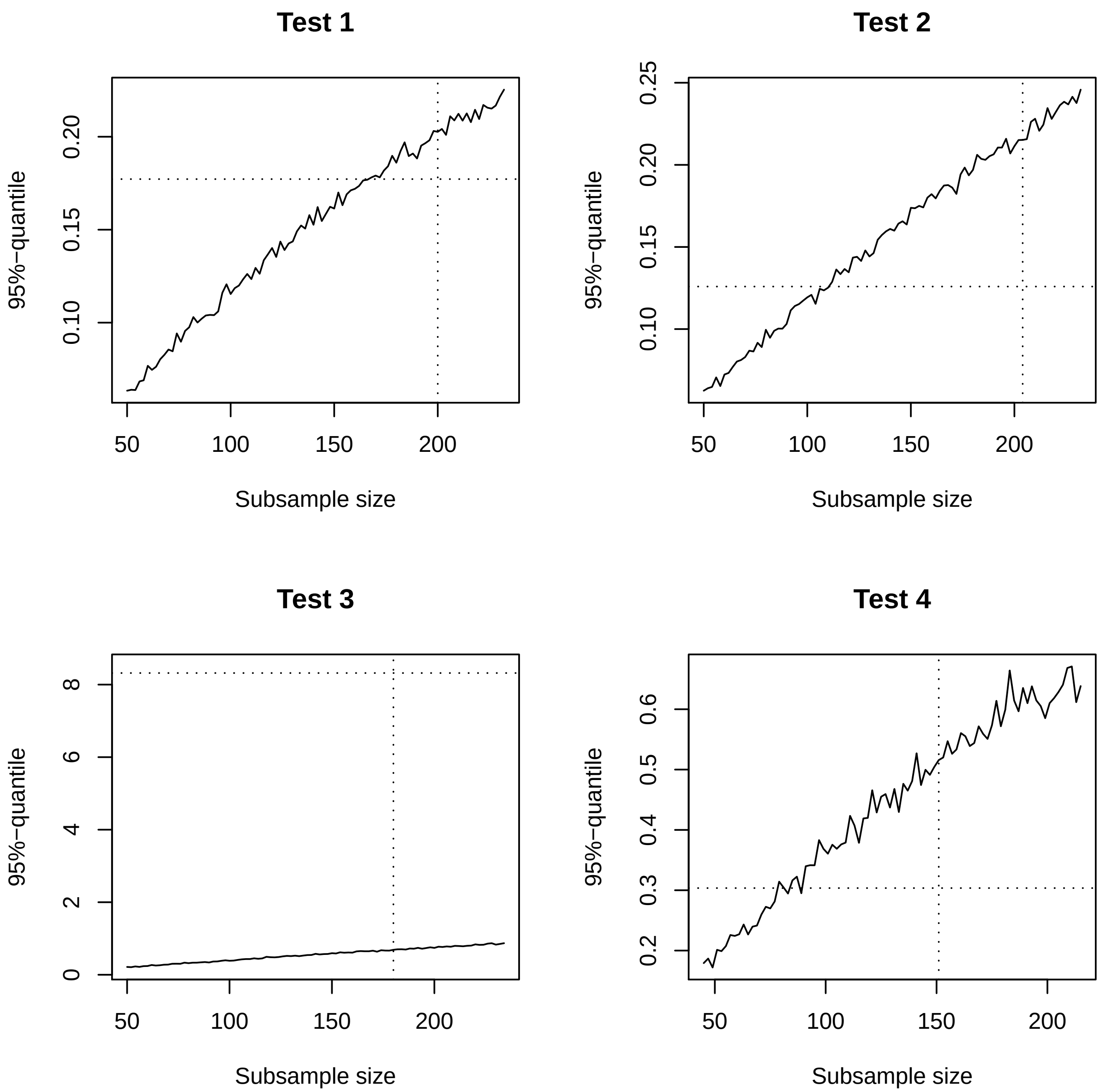


Figure 3: 95%-quantiles of $\{t_{m,b}(\mathcal{X}^*_{m,b})|b = 1, \dots, 2000\}$ for heterogeneous bootstrap based on subsampling without replacement under H_0, as function of m.



6 Conclusion

In this paper we have revisited the question of testing restrictions in non-parametric frontier models, where we extended existing methologies and applied it to the analysis of the role of innovation activities in mechanical engineering. Economics often argues that firms can gain a competitive edge through innovating, which is why special interest is paid to it.

We investigated two different channels through which this might work. According to the first view, the level of innovation input is not completely discretionary, hindering firms in equating marginal products of all inputs (necessary for profit maximization). Firms with lower limitations could benefit from innovating, because they are closer to equating marginal products of innovation-related and regular input. This theory highlights the importance of an efficient innovation management and structures that are flexible enough to cope with the organizational change induced by innovation. The empirical evidence found in this paper is indeed consistent with this view. We also analyzed an alternative explanation common to the literature. According to this, innovation activities allow firms to enter the global market. Resulting from the increased market demand, they can benefit from economies of scale and thereby increase the competitiveness. However, we do not find empirical evidence for this explanation.

References

[1] Aghion, P. and P. Howitt (1992), A Model of Growth through Creative Destruction, *Econometrica*, 60, 323 - 351.

[2] Alvarez, R. (2007), Explaining Export Success: Firm Characteristics and Spillover Effects, *World Development*, 35, 377-393.

[3] Becchetti, L. and M.I.M. Santoro (2001), The Determinants of Small and Medium-sized Firm Internationalization and its relationship with Productive Efficiency, *Weltwirtschaftliches Archiv - Review of World Economics*, 137,297-319.

[4] Beise-Zee, R. and C. Rammer(2006), Local user-producer interaction Innovation and Export Performance of Firms, *Small Business Economics*, 27, 207-222.

[5] Bickel, P.J. and A. Sakov (2008), On the choice of m in the m Out of n bootstrap and Confidence Bound for Extrema,*Statistica Sinica*, 18, 967-985.

[6] Bleaney, M. and K. Wakelin (2002), Efficiency, Innovation and Exports, *Oxford Bulletin of Economics and Statistics*, 64, 3+.

[7] Braunerhjelm, P. (1996), The Relation between Firm-specific Intangibles and Exports, *Economics Letters*, 53, 213-219.

[8] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2 (6), 429-444.

[9] Debreu, G. (1951), The coefficient of resource utilization,*Econometrica*, 19:3, 273-292.

[10] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, A(120), 253-281.

[11] Grossman, G. and E. Helpman (1993), Innovation and Growth in the Global Economy, Cambridge, MA, The MIT Press.

[12] Grupp, H. (1997), Messung und Erklärung des Technischen Wandels, Springer, Berlin.

[13] Grupp, H. and S. Maital (1998), Interpreting the sources of market value in a capital goods market: R&D management in industrial sensors, *R&D Management*, 68, 65-77.

[14] Grupp, H. and Schubert, T. (2009), Review and new evidence on composite innovation indicators for evaluating national performance, *Research Policy*, forthcoming.

[15] Hay, D. A. and G.S. Liu (1997), The Efficiency of Firms: What Difference does Competition Make? *Economic Journal*, 107, 597-617.

[16] Harris, H. and Q.C. Li(2009), Exporting, R&D, and Absorptive Capacity in UK establishments, *Oxford Economic Papers - New Series*, 61, 74-103.

[17] Kneip, A, L. Simar and P.W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory*, 24, 1663–1697.

[18] Lachenmaier, S. and L. Wossmann(2006), Does Innovation Cause exports? Evidence from Exogenous Innovation Impulses and Obstacles Using German Micro Data, *Oxford Economic Papers - New Series*, 58, 327-350.

[19] OECD (2005), OSLO Manual: Guidelines for Collecting and Interpreting Innovation Data, OECD and Eurostat, 3rd edition

[20] Pla-Barber, J. and J. Alegre (2007), Analysing the Link between Export Intensity, Innovation and Firm Size in a Science-based Industry, *International Business Review*, 16, 275-293.

[21] Penrose, E.T. (1959), The Theory of the Growth of the Firm, Basil Blackwell, Berlin.

[22] Politis, D.N., J.P. Romano and M. Wolf (2001), On the asymptotic theory of subsampling, *Statistica Sinica*, 11, 1105–1124.

[23] Rodriguez, J.L. and R.M.G Rodriguez, R. M. G. (2005), Technology and Export Behaviour, A Resource-based View Approach, *International Business Review*, 14, 539-557.

[24] Romer, R.M. (1986): Increasing Returns and Long-Run Growth, *Journal of Political Economy*, 94, 1002-1037.

[25] Romer, P.M. (1990), Endogenous Technological Change, *Journal of Political Economy*, 98, 72 - 102.

[26] Schubert, T. and H. Grupp (2009), Tests and Confidence Intervals for a Class of Scientometric, Technological and Economic Specialisation Ratios, *Applied Economics*, forthcoming

[27] Serfling, R.S. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New-York.

[28] Simar, L. and P. Wilson (1998), Sensitivity of efficiency scores : How to bootstrap in Nonparametric frontier models, *Management Sciences*, 44, 1, 49–61.

[29] Simar L. and P. Wilson (2001), Testing Restrictions in Nonparametric Efficiency Models, *Communications in Statistics, simulation and computation*, 30 (1), 159–184.

[30] Simar, L. and P.W. Wilson (2008), Statistical Inference in Nonparametric Frontier Models: recent developments and perspectives, in *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A. Knox Lovell and Shelton Schmidt, editors, Oxford University Press.

[31] Simar, L. and P.W. Wilson (2009), Inference by Subsampling in Nonparametric Frontier Models, (manuscript).

[32] Solow, R.M (1956), A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics*, 70, 6594.

[33] Sterlacchini, A. (1999), Do Innovative Activities Matter to Small Firms in non-R&D-intensive Industries? An Application to Export Performance, *Research Policy*, 28, 819-832.

[34] Tomiura, E.(2007), Effects of R&D and Networking on the export Decision of Japanese Firms, *Research Policy*, 36, 758-767.

[35] Wakelin, K.(1998): Innovation and Export Behaviour at the Firm Level, *Research Policy*, 26, 829-841.

[36] Wilson, P.W. (2008), FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R, *Socio-Economic Planning Sciences* 42, 247–254.

[37] Yang, C.H., Chen, J.R. and W.B. Chuang (2004), Technology and export decision, *Small Business Economics*, 22, 349-364.

A Appendix: Aggregation of outputs

We only give a sketch of how to proceed for testing if some outputs can be aggregated. We again follow the ideas of Simar and Wilson (2001). Let $y = (y^1, y^2)$, where $y^2 \in \mathbb{R}^r_+$ is the vector of the outputs we consider to aggregate. We denote as $y^+ = i_r' y^2 \in \mathbb{R}_+$ the resulting aggregated output. So we would like to solve the following test problem

$$\begin{aligned} H_0 &: y^2 \text{ can be aggregated in } y^+, \\ H_1 &: y^2 \text{ cannot be aggregated in } y^+. \end{aligned}$$

Now for any given point $(x, y) = (x, y^1, y^2) \in \Psi$ we define

$$\theta(x, y) = \sup\{\theta | (\theta x, y^1, y^2) \in \Psi\} \tag{A.1}$$

$$\theta_0(x, y) = \sup\{\theta | (\theta x, y^1, v) \in \Psi, \text{ with } v \in \mathbb{R}^r_+, (x, y^1, v) \in \Psi \text{ and } i_r' v = y^+\} \tag{A.2}$$

Clearly, we have in general $\theta_0(x, y) \leq \theta(x, y) \leq 1$. In addition, we have the following basic inequalities

$$\begin{aligned} &\text{if } H_0 \text{ is true , } \theta_0(x, y) = \theta(x, y) \leq 1, \text{ for all } (x, y) \in \Psi \\ &\text{if } H_1 \text{ is true , } \theta_0(x, y) < \theta(x, y) \leq 1, \text{ for some } (x, y) \in \Psi. \end{aligned} \tag{A.3}$$

Note that from a sample $\mathcal{X}_n$, these two quantities are estimated by

$$\widehat{\theta}(x, y) = \max\Big\{\theta \ \mid \ \theta x \geq \sum_{i=1}^n \gamma_i X_i, \ y^1 \leq \sum_{i=1}^n \gamma_i Y_i^1, \ y^2 \leq \sum_{i=1}^n \gamma_i Y_i^2 \ \sum_{i=1}^n \gamma_i = 1, \ \gamma_i \geq 0 \ \forall \ i = 1, \ \ldots, \ n\Big\}, \tag{A.4}$$

$$\widehat{\theta}_0(x, y) = \max\Big\{\theta \ \mid \ \theta x \geq \sum_{i=1}^n \gamma_i X_i, \ y^1 \leq \sum_{i=1}^n \gamma_i Y_i^1, \ y^+ \leq \sum_{i=1}^n \gamma_i Y_i^+, \ \sum_{i=1}^n \gamma_i = 1, \ \gamma_i \geq 0 \ \forall \ i = 1, \ \ldots, \ n\Big\}, \tag{A.5}$$

whereas for y^+, $Y_i^+ = i_r' Y_i^2$ for $i = 1, \ldots, n$. For the same reasons as above, for all $(x, y) \in \widehat{\Psi}_{DEA}$, we have the basic inequality for the estimators

$$\widehat{\theta}_0(x, y) \leq \widehat{\theta}(x, y) \leq 1. \tag{A.6}$$

Denote by $\mathbb{P}_0$ the restricted DGPs where the null hypothesis is true and $\mathbb{P}_1$ its complement, so we have $\mathbb{P}_0 \cap \mathbb{P}_1 = \emptyset$ and $\mathbb{P} = \mathbb{P}_0 \cup \mathbb{P}_1$. In fact we want to test $H_0 : P \in \mathbb{P}_0$ versus $H_1 : P \in \mathbb{P}_1$. Consider now a particular model $P \in \mathbb{P}$ and the model characteristic $t(P)$ defined as

$$t(P) = E\left(\frac{\theta(X,Y)}{\theta_0(X,Y)} - 1\right). \tag{A.7}$$

Due to the basic inequalities (A.3) discussed above, we have $t(P) \geq 0$ for all $P \in \mathbb{P}$, but $t(P) = 0$ if $P \in \mathbb{P}_0$ and $t(P) > 0$ if $P \in \mathbb{P}_1$.

A consistent estimator of $t(P)$ is

$$t_n(\mathcal{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\theta}(X_i,Y_i)}{\widehat{\theta}_0(X_i,Y_i)} - 1\right). \tag{A.8}$$

We know by construction, see (A.6), that $t_n(\mathcal{X}_n) \geq 0$, and we will reject H_0 if $t_n(\mathcal{X}_n)$ is too large. The subsampling algorithms can be developed along the same lines as described above for the aggregation of inputs case.

B Appendix: Some Asymptotics

DEA estimators, as defined e.g. by (15), suffer from the *curse of dimensionality* shared by most of the nonparametric approaches, which means that to achieve the same accuracy, when the dimension $(p+q)$ of the input-output space increases, we need much more data. Kneip et al. (2008) show that for every point $(x_0, y_0) \in \Psi$

$$n^{2/(p+q+1)}\big(\widehat{\lambda}(x_0,y_0) - \lambda(x_0,y_0)\big) \xrightarrow{\mathcal{L}} Q(\cdot;\eta), \tag{B.1}$$

where $Q(\cdot;\eta)$ is a regular distribution function defined on $\mathbb{R}_-$ depending on a vector of unknown parameters η (these parameters depends on the DGP, like the density of (X,Y) near the frontier, the slope and curvature of the frontier, etc).[18] The curse of dimensionality is reflected by the fact that the rate of convergence $n^{2/(p+q+1)}$ is far below the usual parametric rate $n^{1/2}$ when p or q increases with $p+q>3$.

To derive the asymptotic distribution of our test statistics, we follow the same arguments as in Simar and Wilson (2009). Let $Z = (X,Y)$ denote a generic observation and define

$$T(Z) = \frac{\lambda_0(Z)}{\lambda(Z)} - 1 \text{ and } \widehat{T}(Z;\mathcal{X}_n) = \frac{\widehat{\lambda}_0(Z)}{\widehat{\lambda}(Z)} - 1, \tag{B.2}$$

[18]Note that no closed form of this limiting distribution is available, but its existence is crucial for proving the consistency of bootstrap approximations.

where $\mathcal{X}_n = \{Z_1, \ldots, Z_n\}$. So that

$$t(P) = E(T(Z)) \;\text{ and }\; t_n(\mathcal{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \widehat{T}(Z_i; \mathcal{X}_n). \tag{B.3}$$

We know that for all $P \in \mathbb{P}$, $T(Z) \overset{a.s.}{\geq} 0$ and $\widehat{T}(Z; \mathcal{X}_n) \overset{a.s.}{\geq} 0$ but if $P \in \mathbb{P}_0$, $T(Z) \overset{a.s.}{=} 0$ and still $\widehat{T}(Z; \mathcal{X}_n) \overset{a.s.}{\geq} 0$. We know also from (B.1) that for all $P \in \mathbb{P}$ and for all $z = (x, y) \in \Psi$,

$$n^{2/(p+q+1)}\big(\widehat{T}(z; \mathcal{X}_n) - T(z)\big) \xrightarrow{\mathcal{L}} G(\cdot; z), \tag{B.4}$$

where $G(.; z)$ is a nondegenerate distribution whose characteristics depends on z. By marginalizing on Z we have

$$n^{2/(p+q+1)}\big(\widehat{T}(Z; \mathcal{X}_n) - T(Z)\big) \xrightarrow{\mathcal{L}} Q(\cdot), \tag{B.5}$$

where $Q(\cdot) = \int_z G(\cdot \mid z) f_Z(z) dz$ is, under regularity conditions, a nondegenerate distribution with finite mean μ_0 and finite variance $\sigma_0^2 > 0$.

Now, by using central limit theorem for triangular arrays (see e.g. Serfling, 1980, section 1.9.3), it is easy to show that if $P \in \mathbb{P}_0$,

$$n^{2/(p+q+1)}\sqrt{n}\,\big(t_n(\mathcal{X}_n) - \mu_0/n^{2/(p+q+1)}\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_0^2). \tag{B.6}$$

So the rate of convergence for obtaining a regular distribution of $t_n(\mathcal{X}_n)$ under the null is $\tau_n = n^{2/(p+q+1)}\sqrt{n}$. Note that $\mu_0/n^{2/(p+q+1)}$ acts as a bias term that can be neglected. It can also be proven that for all $P \in \mathbb{P}$, $t_n(\mathcal{X}_n)$ converges in probability to $t(P)$. For technical details and proofs, see Simar and Wilson (2009). So we can apply Theorem 3.1 of Politis et al. (2001) for using subsampling.