

APLICAÇÃO DO MÉTODO BOOTSTRAP NA ESTIMAÇÃO DE FRONTEIRAS NÃO-PARAMÉTRICAS: O CASO DOS FRUTICULTORES DO VALE DO SÃO FRANCISCO

Emanoel de Souza Barros

Doutor em Ciências Econômicas pela Université Paris I – Pantheon Sorbonne, Paris, France.
Pesquisador e professor da Universidade Federal de Pernambuco - UFPE.

Alcides Jeronimo de Almeida Tenorio Junior

Mestre em Economia Aplicada pela Universidade Federal de Pernambuco - UFPE.
Professor convidado da Universidade de Pernambuco - UPE.

Resumo:

A agricultura irrigada é a grande responsável pelo marcante desenvolvimento ocorrido no Vale do São Francisco. Ela possibilitou a captação de novas tecnologias, criando um pólo de produção diversificada de frutas destinadas ao mercado europeu e norte-americano. Tudo isto devido aos investimentos do Governo Federal nos anos setenta, o qual criou toda a infraestrutura da irrigação, possibilitando a conjuntura atual da região. Assim, supondo a atual tecnologia utilizada na irrigação, combinada com a aplicação de insumos modernos, pode-se esperar um nível de produção próximo ou sobre a fronteira de produção. Para isto, este artigo visa analisar a eficiência dos colonos do pólo Petrolina/Juazeiro através do método DEA-V e submeter tais resultados à abordagem bootstrap proposta por Silverman [1986], o *smoothed bootstrap*, a fim de encontrar intervalos de confiança para testar a confiabilidade dos estimadores DEA-V e, em seguida, testar a convexidade do conjunto de produção. A análise dos resultados mostra que os fatores Insumo e Capital são os grandes responsáveis pela ineficiência dos colonos. Todavia, a simulação bootstrap mostrou que o método convencional tende a superestimar os níveis de eficiência. A hipótese de convexidade é violada, abrindo margem a múltiplos equilíbrios.

Palavras-chave: Irrigação, eficiência, produtividade, renda, DEA-V, bootstrap.

Abstract:

Irrigated agriculture is largely responsible for the expressive development occurred in the Valley of the São Francisco. New technologies were introduced, creating a pole of diversified production of fruit destined to the European and North American market. This is occurred because of investments of the Federal Government in the seventies, which created all the infrastructure of the irrigation, allowing the current conjuncture of the region. Thus, assuming the current technology used in irrigation, combined with the application of modern inputs, it can expect a level of production on or near the frontier of production. For this, this paper aims to analyse the efficiency of farmers from pole Petrolina / Juazeiro with the DEA-V method and submitting such results to the approach bootstrap proposed by Silverman [1986], *smoothed bootstrap*, in order to find intervals of confidence to test the trustworthiness of estimators DEA-V and the convexity of the set of production. Results show that inputs and capital factors are the responsible for inefficiency of farmers. However, the bootstrap simulation showed that the conventional method tends to overestimate the efficiency levels. The hypothesis of convexity is violated, opening room for multiple balances.

Key-words: Irrigation, efficiency, productivity, rentability, DEA-V, bootstrap.

1. Introdução

No final dos anos sessenta, o Governo Brasileiro criou o Plano Plurianual de Irrigação. Este plano tinha como objetivo implementar pesquisas no sentido de criar perímetros de irrigação no semi-árido nordestino, de modo a equilibrar as diferenças estruturais entre o interior e a costa. Estas implementações deveriam levar em conta a existência de reservatórios naturais de águas perenes e subterrâneas, possibilitando a formulação de estratégias de desenvolvimento sócio-econômico do Vale do São Francisco a partir da agricultura irrigada.

Estes investimentos foram realizados ao longo daquela década e da subsequente. Inicialmente com a implementação da infra-estrutura de irrigação, de auto-estradas, aeroportos e todos os aportes necessários para atrair a atenção de indústrias processadoras de matérias-primas, e empresas agrícolas do centro-sul do Brasil bem como do exterior. Assim, foram gerados efeitos diretos e indiretos sobre o emprego e a renda (novos postos de trabalho, formação de mão-de-obra especializada, etc), freando o êxodo rural em direção às grandes cidades do litoral e gerando um desenvolvimento durável para a região.

A área mais desenvolvida do Vale é o pólo Petrolina/Juazeiro, com uma superfície de 32.295 km² e 565.355 habitantes, localizado no submédio São Francisco, justamente na região conhecida como “polígono da seca”. Ele agrupa oito cidades: Juazeiro, Sobradinho, Casa Nova e Curaçá, nos estados da Bahia, e Petrolina, Lagoa Grande, Santa Maria da Boa Vista e Orocó, no estado de Pernambuco. E sua produção é diversificada, se concentrando sobretudo no campo da fruticultura irrigada, através da produção de variedades como manga, uva, goiaba, acerola e coco, destinadas em boa parte ao mercado europeu e norte-americano.

Entre os principais organismos que apoiaram a atração de capitais para a irrigação podem ser citados: a Secretaria para o Desenvolvimento do Nordeste (SUDENE); O Fundo de Investimentos do Nordeste (FINOR), responsável pela implantação das agro-indústrias; o Banco do Nordeste, através do programa de crédito subsidiado; e, o Banco Nacional de Desenvolvimento (BNDES).

Os primeiros projetos de irrigação foram introduzidos no pólo Petrolina/Juazeiro ao longo dos anos sessenta, quando a SUDENE encarregou-se de dar o passo inicial com a implantação dos projetos Bebedouro e Mandacaru, em 1968. Na década seguinte, foram introduzidos mais quatro projetos públicos: Maniçoba e Curaçá, em 1982, e, Tourão e Nilo Coelho, em 1984, que juntos totalizavam, na época, 46.279 hectares irrigados.

A orientação da produção, antes e após a implantação destes projetos, era completamente antagônica. Antes, a produção se restringia a pequenas unidades produtivas baseadas no cultivo do algodão com utilização de técnicas rudimentares. A partir daí a iniciativa do Setor Público, quando da criação desses projetos, foi considerada pioneira, pois viabilizou a utilização intensiva da terra em regiões de baixa densidade demográfica (típicas do semi-árido), permitindo o estabelecimento de empresas agrícolas e agro-industriais, dinamizando assim a toda a região.

O desenvolvimento do pólo Petrolina/Juazeiro foi acompanhado pela expansão marcante dos setores industrial e de serviços. Os investimentos feitos pelo Governo Federal em infra-estrutura ligadas à captação, adução e distribuição de água possibilitaram o cultivo de culturas permanentes, consolidando a expansão de vários setores produtivos. Estudos mostram que, somente em 1998, foram gerados na região, 86.668 empregos, seguido de um PIB de US\$ 609 milhões, todos os dois concernentes aos impactos gerados de forma direta e indireta (Barros, Costa e Sampaio [2004a, 2004b] e Barros et al. [2004c]).

Atualmente, o pólo Petrolina/ Juazeiro se concentra em três atividades econômicas principais: arboricultura e oleaginosos irrigados, com potencial para diversificação de frutas

exportáveis a grande escala; criação de ovelhas e gados; e agro-indústrias especializadas no processamento de frutos e comercialização de produtos ligados à criação de gados e ovelhas. Entre suas principais vantagens competitivas podem ser citadas: a disponibilidade de terra e água de boa qualidade; mão-de-obra abundante; infra-estrutura de irrigação em expansão; proximidade dos mercados europeus e americanos; e, ciclo produtivo precoce, com altos níveis de produtividade.

Esta nova realidade é fortemente dirigida no sentido de se obter uma produção eficiente e competitiva, capaz de uma boa penetração comercial no mercado externo, em virtude da qualidade dos seus produtos. Diante do novo contexto mundial (globalização), a associação entre colonos (fazendeiros) e empresas agrícolas se tornou indispensável, pois os dois contribuem simultaneamente ao crescimento da renda total. Vale lembrar que as relações comerciais estabelecidas entre as agro-indústrias, as empresas produtoras de insumos e os colonos, ultrapassam a fronteira Petrolina/Juazeiro, aumentando os efeitos a montante e a jusante sobre a cadeia produtiva do Vale do São Francisco.

Diante dessa realidade, pode-se supor que a alta tecnologia utilizada na irrigação, combinada com o uso de insumos modernos, possibilita um nível de produção próximo ou sobre a fronteira (de produção).

Neste artigo em particular, há o interesse de estudar a eficiência dos colonos do Vale do São Francisco. Para isto, será estimada uma fronteira baseada no método *Data Envelopment Analysis (DEA)*, proposta por Barnker, Charnes e Cooper [1984], com rendimento variável de escala (DEA-V), para o conjunto dos colonos inseridos nos cinco principais projetos de irrigação do pólo Petrolina/Juazeiro (Bebedouro, Maniçoba, Curaçá, Tourão e Nilo Coelho).

Após estas estimações, usar-se-á a abordagem bootstrap proposta por Silverman [1986] para testar a confiabilidade e as propriedades dos estimadores do modelo DEA-V. O objetivo é encontrar uma aproximação da verdadeira lei estatística de modo a verificar se a distribuição assintótica é uma boa imagem de sua verdadeira distribuição, a qual na realidade desconhecemos. Pode-se, neste caso, melhorar a qualidade dos resultados estatísticos caso a distribuição assintótica não funcione perfeitamente.

Este artigo está dividido em três partes : a segunda sessão, referente à metodologia de estimação dos níveis de eficiência pelo método DEA-V e pelo bootstrap; a terceira sessão, onde estão delineados os resultados encontrados; e, por fim, as conclusões.

2. Metodologia

A análise de fronteiras de produção tem como marco inicial o trabalho de Farrell [1957] quando utilizou técnicas de programação linear para encontrar a eficiência econômica. Seu trabalho era simples e se baseava na medida de uma isoquanta eficaz de modo a estabelecer uma combinação de insumos e tecnologia capaz de gerar uma fronteira máxima de produção. Nenhuma forma funcional foi utilizada neste trabalho.

Os modelos não paramétricos foram desenvolvidos diretamente a partir da análise de Farrell [1957]. Seu trabalho foi considerado não paramétrico, no sentido de ter sido construído a partir de uma série de inequações lineares para obtenção de um conjunto de produção submetido à regra “*free disposal*” das taxas de insumo-produto observado. Färe e Lovell [1985] mostraram que a regra de Farrell é equivalente ao cálculo das funções distância-insumo de Shepard [1970]. Charnes, Cooper e Rhones [1978] desenvolveram uma técnica chamada Análise Envoltória de Dados (DEA), onde é possível construir fronteiras de produção cujas limitações impostas são de uma tecnologia convexa (perfeita divisibilidade) e livre disponibilidade de insumos e produtos.

A principal motivação para se utilizar o método DEA está no fato de que, para avaliar a eficiência das firmas, não é necessário supor uma forma funcional para a tecnologia. Além do mais, este método é bastante útil quando se pretende analisar múltiplos insumos e múltiplos produtos, e a variável “preço” não pode ser quantificada, como nos serviços prestados por universidades, hospitais, dentre outros.

A suposição fundamental do modelo DEA está no fato que é possível construir uma fronteira com segmentos lineares, uma “fronteira de melhor prática”, utilizando as firmas reais nos seus pontos extremos e firmas “virtuais” ou “compostas”, criadas a partir de combinações convexas das firmas reais. As firmas eficientes situam-se sobre fronteira enquanto que as não eficientes em baixo dela.

O objetivo é encontrar o melhor produto virtual com relação a cada produto real. Se o produto virtual é melhor que o produto real, utilizando a mesma quantidade de insumos que a firma real, esta firma (real) é considerada ineficiente. As firmas reais e virtuais são chamadas *Decision Making Units (DMU)*. Elas recebem este nome porque, de forma geral, o método DEA não leva em conta os preços dos insumos e produtos.

O modelo DEA com rendimento variável de escala (DEA-V) foi introduzido por Banker, Charnes e Cooper [1984] como uma extensão do modelo DEA-C, de Charnes, Cooper e Rhodes [1978]. A principal novidade foi a incorporação da hipótese de rendimento variável de escala à análise. Para isto, eles introduziram a idéia de que a soma dos pesos de cada firma no processo de produção, $\sum \lambda_j$, é igual à unidade, representada pela notação matricial $e^T \lambda = 1$.

Supondo que exista n DMUs representadas por $j \in J = \{j \mid j = 1, 2, \dots, n\}$, dado um vetor de insumos $X_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T > 0$ e um vetor de produtos $Y_j = (y_{1j}, y_{2j}, \dots, y_{sj})^T > 0$. Sob a hipótese de rendimento variável de escala, existem diferenças no tocante ao tipo de distância a ser usada para estimar os níveis de eficiência: distância produto ou distância insumo. No caso deste artigo, adotaremos a distância produto. Assim, o problema de maximização é dado por:

$$\begin{aligned}
& \underset{\theta, \lambda}{\text{Max}} \theta & (1) \\
& \text{s. a. } X\lambda \leq X_0 \\
& \theta Y_0 - Y\lambda \leq 0 \\
& q \text{ é livre } \quad e^T \lambda = 1
\end{aligned}$$

A eficiência será dada pela inversa da solução do problema de maximização, $1/\theta$.

Entretanto, todas essas aproximações geram fronteiras de produção supondo distribuições conhecidas somente assintoticamente. A partir dos anos noventa, a Econometria tomou consciência que o custo de uma má especificação de um modelo está intimamente ligado ao risco crescente de conclusões errôneas. O desenvolvimento de técnicas de estimação menos restritivas e a procura da boa especificação de um modelo fizeram com que ela se dedicasse à descoberta de um melhor teste. A preocupação principal, percebida nas estimativas das décadas precedentes, voltou-se para a inferência estatística.

Foi neste contexto que diversos métodos de simulação, como Monte Carlo e Bootstrap, vieram à tona. O objetivo destes métodos é encontrar uma resposta à necessidade de testar a confiabilidade das estatísticas largamente utilizadas e estimar testes de hipóteses e intervalos de confiança mais próximos da verdadeira distribuição estatística, que é desconhecida.

Beran [1986] e Hall e Titterington [1989] foram os primeiros a propor uma simulação baseada no Bootstrap. Derivado do método de Monte Carlo, ele se diferencia deste primeiro pelo fato do termo de erro ser “desenhado” de amostras construídas por simulação a partir de uma amostra original. Eles são tiragens aleatórias da amostra original, e não existe um modelo (a priori) que suponha um formato para a distribuição. O Bootstrap tem diversos tipos de aplicação, especialmente nas situações onde os modelos tradicionais têm certas dificuldades em obter resultados satisfatórios. Ele pode estabelecer, sob certas condições, testes de hipóteses e intervalos de confiança mais confiáveis do que os dos modelos tradicionais (Mínimos Quadrados Ordinários, Máxima de Verossimilhança, Mínimos Quadrados Ponderados, etc.).

No caso da abordagem não paramétrica, as fronteiras são construídas a partir de um DGP, considerado pivô assintótico de um DGP desconhecido compatível com todas as definições, hipóteses e propriedades dos conjuntos de produção. Isto nos permite estabelecer uma fronteira de produção compatível com a teoria da produção, mesmo na ausência de uma forma funcional. Jensen [2000], Simar [1992], Horrace e Richards [2004], Hall, Hardle e Simar [1995], Simar e Wilson [2000a, 2000b] estimaram fronteiras de produção baseadas nesses pressupostos.

Simar e Wilson [1998] propuseram uma estratégia bootstrap para analisar a sensibilidade das medidas de eficiência ao longo das variações de amostras bootstrap. Eles também estimaram intervalos de confiança de modo a corrigir o viés inerente ao procedimento DEA. Todavia, o método utilizado é condicionado à regra de homogeneidade restrita das distribuições de eficiência entre as firmas. Outros métodos, como o *free disposal hull* (FDH), são discutidos por estes mesmos autores, em um trabalho publicado em 2000 (Simar et Wilson [2000b]), como alternativas menos restritivas para este impasse. Este método (FDH) é derivado do método DEA e se baseia na ausência de regras que imponham a convexidade da fronteira de produção¹. Eles definiram cinco hipóteses para caracterizar o

¹ Ver Deprins et al. [1984] e Kneip et al. [1998] para maiores detalhes.

DGP do modelo DEA. Estas hipóteses servem de complementos àquelas propostas por Shepard's [1970] e dão consistência ao formato do conjunto de produção sob simulação. Paralelamente, estas hipóteses dão consistência à formulação de uma densidade gaussiana de Kernel, a partir da qual serão estimados os intervalos de confiança bootstrap.

De modo geral, o método bootstrap utilizado por Simar e Wilson [1998] tem como objetivo construir amostras bootstrap (x_i^*, y_i^*) de insumos e produtos a partir de uma função de densidade $\hat{f}(x, y)$ sujeita a uma certa tecnologia. A função $\hat{f}(x, y)$ representa o estimador consistente da densidade sub-jacente $f(x, y)$ do conjunto de produção dos pontos disponíveis desta tecnologia. Isto equivale a estimar a densidade $f(y, h, d)$ e construir as amostras bootstrap (y_i^*, h_i^*, d_i^*) , através de coordenadas polares propostas por Kneip et al. [1998], onde h_i^* representa o número de observações de cada amostra bootstrap e d_i^* o nível de eficiência estimado, por simulação, para cada unidade de produção.

Silverman [1986] utiliza os estimadores de $f(y, h, d)$ a partir de um conjunto $\{(y_i, h_i, d_i), i = 1, \mathbf{K}, n\}$ no qual \hat{d}_i é um estimador DEA consistente de d_i . Através de alguns cálculos, eles encontraram o estimador da densidade gaussiana de Kernel, dado por:

$$\tilde{f}_h(z) = \frac{1}{2nh^{(p+q)}} \sum_{i=1}^n \left[K_1\left(\frac{z - z_i}{h}\right) + K_2\left(\frac{z - z_{Ri}}{h}\right) \right] \quad (2)$$

Onde $z = (y, h, d)$, K_l é uma função densidade de média zero para $l = 1, 2$, n o número de observações e h o parâmetro de amortização.

A partir deste estimador, pode-se calcular o estimador consistente \tilde{f}_h de f , contido em A e descrito pela equação $(y_i, h_i, d_i) \in A = R_+^q \times [0, p/2] \times [1, \infty)$

$$\tilde{f}_h(z) = \begin{cases} 2\tilde{f}_h(z) & \text{si } z \in A \\ 0 & \text{autres cas} \end{cases} \quad (3)$$

A consistência de $\tilde{f}_h(z)$ requer que $h \rightarrow 0$ a medida que $n \rightarrow \infty$, mas não tão rápido. Assim, tem-se necessidade de um h que tenha a propriedade seguinte: $h = O(n^{-1/(p+q+4)})$.

Uma vez obtido os valores $\hat{d}_b^*(x, y), b = 1, \mathbf{K}, n$, pode-se estimar o viés bootstrap em relação ao estimador original $\hat{d}(x, y)$ através de:

$$Biais_{est}[\hat{d}(x, y)] = B^{-1} \sum_{b=1}^B \hat{d}_b^*(x, y) - \hat{d}(x, y) \quad (4)$$

O estimador corrigido deste viés será então

$$\hat{\hat{d}}(x, y) = \hat{d}(x, y) - bias_{B}[\hat{d}(x, y)] = 2\hat{d}(x, y) - B^{-1} \sum_{b=1}^B \hat{d}_b^*(x, y) \quad (5)$$

A variância dos valores bootstrap $\hat{d}_b^*(x, y)$ nos dá o estimador da variância de $\hat{d}(x, y)$ ou (\hat{S}^2) . Então, a média estimada dos erros quadráticos de $\hat{d}(x, y)$ será $4\hat{S}^2$ se $B \rightarrow \infty$, e $\left[\hat{S}^2 + (bias_{est}[\hat{d}(x, y)])^2\right]$ para $\hat{d}(x, y)$. Neste caso, a variância \hat{S}^2 será :

$$\hat{S}^2 < \frac{1}{3}(bias_B[\hat{d}(x, y)])^2 \quad (6)$$

Os valores bootstrap $\hat{d}_b^*(x, y)$ podem ser utilizados para construir os intervalos de confiança com relação aos verdadeiros valores de $d(x, y)$. Todavia, a distribuição de $(\hat{d}(x, y) - d(x, y))$ não é conhecida. Neste caso, os intervalos de confiança são calculados a partir das amostras bootstrap. Portanto, tem-se um intervalo de confiança com alta probabilidade de que seja verdadeiro:

$$Pr ob\left(-b_a^* \leq \hat{d}^*(x, y) - \hat{d}(x, y) \leq -a_a^* | \mathfrak{S}\right) = 1 - a \quad (7)$$

Mecanicamente, isto equivale a organizar os valores $(\hat{d}^*(x, y) - \hat{d}(x, y))$, $b = 1, \mathbf{K}, B$ em ordem crescente por valores algébricos, excluindo $(a/2 \times 100)$ – percent dos últimos elementos desta série. Os valores $-b_a^*$ e $-a_a^*$ serão iguais aos últimos pontos do conjunto preservado após a exclusão dos elementos citados acima.

2.1. A base de dados

Os dados mais recentes relativos à conjuntura dos colonos no Vale do São Francisco foram obtidos através do questionário *Investimento Público e Privado em Agricultura Irrigada e seus Determinantes sobre o Emprego e a Renda*, realizado entre março e setembro de 1998, sob demanda da Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba (CODEVASF) e da Universidade Federal de Pernambuco (UFPE). Ele descreve dados, em cortes transversais, para o ano agrícola de 1997 para o quadro de colonos, empresas agrícolas e empresas agro-industriais de todo o Vale do São Francisco.

Todavia, diante da complexidade desta base de dados e para extrair somente os dados que serão analisados neste artigo (colonos), selecionou-se apenas os dados referentes aos colonos produtores de fruticultura irrigada. São eles os produtores que possuem maior valor agregado em termos de produção e que foram submetidos diretamente à tecnologia da irrigação.

A estimação das fronteiras de produção será efetuada com a agregação de todos os produtos por colono, contemplando um total de 149 colonos especializados na produção de uva, manga, goiaba, acerola, banana e coco. A variável dependente será o “Valor da Produção”, considerada uma *proxy* da quantidade produzida.

As variáveis explicativas são:

- “Área irrigada”, em hectare, que descreve o espaço onde o colono cultivou sua produção;
- “Despesas com Insumos”, em dólar (US\$), que descreve o total das despesas com sementes, mudas, adubos, herbicidas, combustíveis, água e energia;
- “O Capital”, em dólar (US\$), que descreve o inventário dos prédios e equipamentos presentes na propriedade do colono;
- “A mão-de-obra”, em horas/ano, que corresponde à mão-de-obra familiar ou contratada por cada colono no processo de produção, comercialização e transporte das culturas irrigadas.

3. Resultados

3.1. Análise descritiva dos dados

A estatística descritiva dos dados (Tabela 1) mostra que, embora haja disparidade entre os valores mínimos e máximos, a maior parte dos colonos utilizam, em média, fatores de produção em intervalos equilibrados. As variáveis “valor da produção”, “área irrigada”, “insumos”, “capital” e “mão-de-obra” possuem índices médios de US\$ 10.030,43; 4,43 h; US\$ 2.579,61; e US\$ 8.812,55, respectivamente. Todavia, seus desvios-padrões são relativamente grandes, sinônimo de variâncias também bastante expressivas. Isto explica os valores mínimos e máximos observados.

Tabela 1 : Estatística descritiva dos dados

<i>Variáveis</i>	<i>Unidade de medida</i>	<i>de Média</i>	<i>Desvio-padrão</i>	<i>Mínimo</i>	<i>Máximo</i>
Valor da produção	Dólar (US\$)	10.030,43	16.596,48	150,00	120.000,00
Área irrigada	Hectare	4,43	2,32	0,50	16,00
Insumos	Dólar (US\$)	2.579,61	2.274,69	238,00	12.955,00
Capital	Dólar (US\$)	8.812,55	10.506,28	400,00	76.600,00
Mão-de-obra	Horas/ano	295,00	209,66	12,00	1.210,00

Fonte : CODEVASF.

É necessário salientar aqui que alguns colonos são especialistas na produção de dois ou três tipos de culturas, enquanto que outros se concentram na produção de uma única cultura, como por exemplo a uva, uma cultura de alto valor agregado. Isto pode gerar diferenças entre colonos no tocante ao nível de despesas com insumos, capital e mão-de-obra, que são proporcionais às quantidades de culturas que eles produzem.

A variável “área irrigada” tem uma média de 4,43 ha por colono. Este índice mostra a eficiência do plano da CODEVASF em distribuir lotes de 5,00 ha por proprietário. O objetivo desta estratégia é promover uma distribuição igualitária de terras. Assim, teoricamente, cada colono dispõe de um espaço razoável para produzir suas respectivas culturas utilizando a tecnologia disponível.

A variável mão-de-obra apresenta grande disparidade entre seus valores mínimo e máximo. Isto se deve basicamente ao tipo de cultura produzida pelo colono. Culturas como acerola e banana são caracterizadas por produções de grande escala e com várias colheitas ao longo do ano, e assim, a mão-de-obra se concentra basicamente na colheita. Os controles fitossanitários são menos frequentes do que os exigidos para culturas como uva, goiaba e manga. Deste modo, um colono produtor de uva é obrigado a contratar mão-de-obra especializada e sua permanência no espaço cultivado é, às vezes, o dobro ou o triplo do número de horas exigidas para os produtores de banana.

Detectou-se também a presença de heterocedasticidade causada pela variável “despesas com insumos”. Para detectá-la, utilizou-se o método gráfico e os testes de Park, Glejser e White. Todavia, como esta variável é indispensável à estimação dos índices de eficiência DEA, ela foi mantida no modelo. Vale lembrar que não existe um método satisfatório para amortizar o efeito da heterocedasticidade quando se utiliza uma abordagem não paramétrica.

3.2. Resultados das estimações DEA-V

A análise dos cinco primeiros níveis de eficiência (Tabela 2) mostra que os fatores que causam a ineficiência são o capital e a mão-de-obra, respectivamente. Os colonos possuem capital superior ao necessário para produzir as culturas irrigadas. No caso particular do quarto colono, a diferença entre o valor do capital utilizado e o valor ideal (entre parêntese) chega ao patamar de US\$12.000,00. De fato, este fator, ligado à utilização de mão-de-obra em quantidade de horas excessivas, são os responsáveis por sua quarta colocação.

O número de horas de trabalho por ano é o segundo fator responsável pela redução da eficiência. Há duas explicações possíveis para este empasse : ou os colonos trabalharam mais que deveriam, ou utilizaram um excedente de mão-de-obra que os levaram a despesas excedentes. Em alguns casos, este nível está correlacionado ao processo de transporte e comercialização do produto, que representa, na maior parte dos casos, despesas extras para os colonos, incluindo a contratação de novos empregados.

O primeiro colono da Tabela 2 é o mais eficiente num sentido relativo. Embora ele deva reduzir um pouco de cada fator de produção, estas reduções não são de grandes magnitudes. Os níveis de eficiência próximos de zero estão diretamente ligados à grandes diferenças entre os valores/quantidades utilizados e os ideais, estimados pelo modelo.

Estes resultados também mostram que os colonos devem gerenciar bem a aquisição de novas máquinas e equipamentos, bem como o número de horas por trabalhadores, se quiserem aumentar seus níveis de eficiência. O primeiro e o quinto colocados da Tabela 2 apresentam rendimento decrescente de escala, enquanto os demais são caracterizados por rendimentos crescentes de escala. Vale lembrar que estão ainda em processo de expansão de suas capacidades produtivas.

Tabela 2: Os cinco primeiros níveis de eficiência do modelo DEA-V e seus valores/quantidades utilizadas e ideais (entre parêntese)

Rank	Nível	VP (US\$)	EI (ha)	IN (US\$)	K (US\$)	MO (horas/ano)	Rendimento de escala
1°	0,9791	75.000,00 (76.593,66)	6,00 (4,10)	7.274,00 (7.274,00)	7.956,00 (7.235,48)	936 (611)	Decrescente
2°	0,9291	45.000,00 (56.285,94)	3,00 (3,00)	6.865,00 (6.865,00)	16.624,00 (13.339,42)	1148 (948)	Crescente
3°	0,8238	19.160,00 (20.264,00)	14,00 (6,90)	7.730,00 (5.709,26)	3.000,00 (3.000,00)	624 (610)	Crescente
4°	0,7716	22.670,00 (27.388,29)	6,00 (6,00)	1.707,00 (1.707,00)	16.270,00 (4.752,44)	1.160 (914)	Crescente
5°	0,7058	50.900,00 (72.108,57)	9,00 (4,02)	13.272,00 (6.198,05)	5.170,00 (5.170,00)	1.184 (319)	Decrescente

Fonte : CODEVASF.

Observação : VP, EI, IN, K e MO representam o valor da produção e os fatores área irrigada, despesas com insumos, capital e mão-de-obra, respectivamente.

Segundo a estimacão do modelo DEA-V, 52,35% dos colonos estão situados na parte da fronteira onde se observa rendimento crescente de escala, enquanto que 28,19% estão situados na região de rendimento decrescente de escala e somente 19,46% apresentam rendimento constante de escala. Tais colonos possuem níveis de eficiência situados no intervalo 0,2 a 0,6. Isto mostra que, no conjunto das estimacões DEA-V, a maior parte deles está em fase de expansão de suas capacidades produtivas e aqueles que não são ainda “eficientes” podem se tornar, se souberem gerenciar seus recursos.

3.3. Resultados do método de simulação bootstrap

Para estimar os intervalos de confiança bootstrap, utilizou-se todas as observações de colonos com exceção das que obtiveram nível de eficiência igual à unidade, a fim de evitar problemas com *outliers*. Em seguida, gerou-se 2.000 replicações bootstrap a partir da amostra original (149 colonos). Estes resultados se encontram na Tabela 3.

A primeira coluna mostra a distância produto $\hat{d}(x, y)$ estimada (ou nível de eficiência); a segunda, o nível de eficiência estimado pela simulação bootstrap; a terceira o viés estimado pela equação (4), a diferença entre o nível de eficiência estimado e simulado; a quarta coluna descreve o desvio-padrão estimado pela equação (6). A quinta e sexta colunas apresentam os valores mínimo e máximo do intervalo de confiança dado por (7) ao nível de significância de 5%.

Após as 2.000 replicações bootstrap, os níveis de eficiência são idênticos aos estimados na Tabela 2. A surpresa vem do fato de que os níveis de eficiência simulados são inferiores aos índices estimados pelo método convencional. Somente o segundo colono teve seu índice de eficiência aumentado. Isto poderia indicar uma superestimação dos níveis de eficiência pelo método DEA-V tradicional. Os vieses são positivos com exceção do segundo colono, pois seu nível de eficiência simulado foi superior ao estimado pelo modelo clássico. Os desvios-padrão são relativamente pequenos, sinônimo de variâncias centradas na média. Todavia, as variâncias dos cinco colonos são 0,0012; 0,000013; 0,00015; 0,0051; e, 0,000072, respectivamente. Ou seja, diferentes uma das outras.

Devido a estas variâncias, os intervalos de confiança possuem larguras diferentes. Vale também ressaltar que os dados são heterocédasticos e este “pequeno detalhe” pode influenciar a construção da fronteira de produção e, por conseqüência, os níveis de eficiência e os intervalos de confiança bootstrap (que são mais largos neste caso).

Tabela 3 : Os cinco primeiros níveis de eficiência do modelo DEA-V e seus resultados com relação aos níveis de eficiência bootstrap, o viés estimado, o desvio-padrão e o intervalo de confiança (para um nível de significância de 5% e 2.000 replicações).

Rank	$\hat{d}(x, y)$	$\hat{d}^*(x, y)$	$Biais_{est}[\hat{d}(x, y)]$	Desvio-padrão	Valor mínimo	Valor Máximo
1º	0,9791	0,9248	0,0543	0,0314	0,9026	0,9666
2º	0,9291	0,9351	-0,0060	0,0035	0,9194	0,9471
3º	0,8238	0,8031	0,0207	0,0120	0,7833	0,8216
4º	0,7716	0,6521	0,1191	0,0688	0,6501	0,6917
5º	0,7058	0,6918	0,0140	0,0081	0,6914	0,7142

Fonte : CODEVASF.

Os níveis de eficiência bootstrap e seus respectivos intervalos de confiança são úteis quando se está interessado em investigar as distâncias produto corretas que permitam estabelecer uma fronteira de produção mais próxima do que seria sua verdadeira representação. Todavia, existe uma condição que pode inverter toda a balança de um lado ou de outro. Trata-se da hipótese de convexidade do conjunto de produção.

De fato, um conjunto de produção convexo implica diretamente numa função de produção côncava. Se esta hipótese não pode ser verificada, sua função de produção correspondente não será mais côncava. Diante de uma função de produção não-côncava, ter-se-ia múltiplos equilíbrios e a balança tenderia a um valor de produção não-linear.

Como a simulação bootstrap é feita a partir do modelo DEA, poder-se-ia propor uma outra abordagem, baseada sobre a validade da hipótese de convexidade do conjunto de produção. Todavia, como esta simulação é efetuada no contexto de uma abordagem não

paramétrica, não é possível calcular a elasticidade de produção sobre esta fronteira. Contudo, ter-se-ia um meio de verificar a validade da hipótese de convexidade e, assim, a existência de múltiplos equilíbrios ou não.

Para isto, utilizar-se-á um estimador de fronteira tecnológica não repousando sobre a hipótese de convexidade do conjunto de produção. Trata-se do estimador proposto por Deprins et al. [1984], o FDH (*Free Disposal Hull*). Desta forma, pode-se testar uma especificação convexa contra uma especificação não convexa do conjunto de produção.

Park, Simar e Weiner [2000] mostraram que, sob algumas condições, o estimador FDH converge em direção à verdadeira fronteira. Esta velocidade é geralmente inferior à do estimador DEA, pois se impõe um número menor de estruturas ao modelo (ausência da hipótese de convexidade). Eles derivaram também uma distribuição assintótica da distância FDH em relação à fronteira. O estimador, corretamente normalizado pelo tamanho da amostra, tende a uma distribuição *Weibull* e depende de parâmetros aleatórios inobservados. Isto permite construir intervalos de confiança assintóticos para os colonos e testar se um determinado colono é realmente eficiente. Segundo Adjemian [2002], os parâmetros não observados não são tão fáceis de estimar, e a aproximação assintótica pode se revelar insuficiente. Portanto, neste artigo, iremos utilizar o estimador FDH somente para estabelecer um teste de hipótese sobre a convexidade do conjunto de produção.

Por construção, o conjunto de produção FDH está contido no conjunto de produção DEA-NIRS (rendimento de escala não crescente), que por sua vez pertence ao conjunto de produção DEA-V; assim, a taxa d_i^{DEA-V} / d_i^{FDH} será necessariamente inferior ou igual a 1 para todo $i = 1, \mathbf{K} N$.

Os níveis de eficiência média dos modelos DEA-V e FDH são 0,4350 e 0,6534, respectivamente. Por construção, as distâncias em relação à fronteira FDH são mais importantes que as estimadas pela fronteira DEA-V. Assim, observam-se 25 colonos eficientes no modelo DEA-V contra 56 do modelo FDH. Para estimar estas densidades, supõe-se um parâmetro de amortização ótimo dado por $h = 1,06 \cdot std(d(x, y)) \cdot N^{-1/5}$, onde $std(d(x, y))$ representa o desvio-padrão da variância (6) e N o número de observações de colonos com exceção dos que obtiveram nível de eficiência igual a 1. Este procedimento é o de *cross-validation* (ver Silverman [1986, páginas 51-52]). Os valores estimados deste parâmetro são $\hat{h}_{DEA-V} = 0,1384$ e $\hat{h}_{FDH} = 0,1511$.

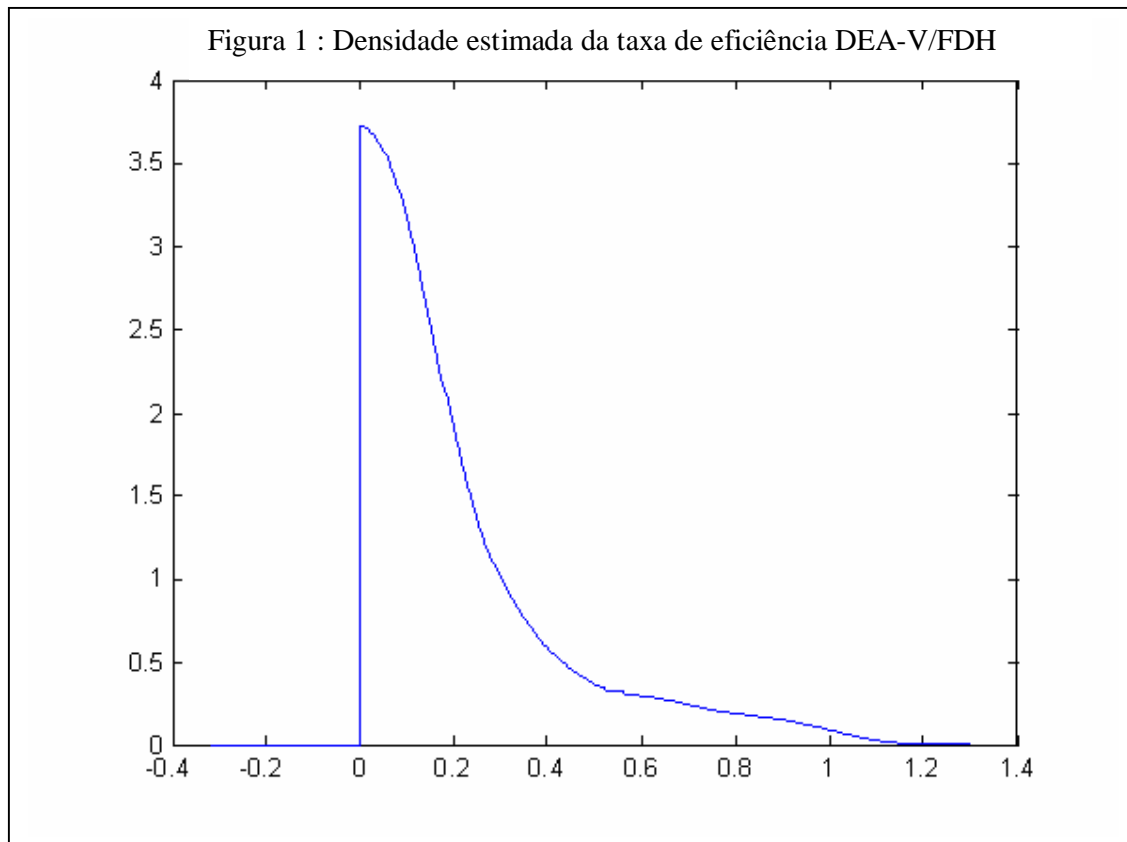
Agora, pode-se estabelecer um teste para a convexidade do conjunto de produção. Para isto, será utilizada a taxa d_i^{DEA-V} / d_i^{FDH} para todo $i = 1, \mathbf{K} N$, que no caso da estimação deste estudo, possui o valor médio de $\bar{d} = 0,3335$. Para verificar a natureza da convexidade ou não do conjunto de produção, deve-se testar se a taxa média é significativamente diferente da unidade. Neste intuito, será considerada a estatística seguinte para testar a convexidade :

$$R = N^{-1} \sum_{i=1}^N \frac{d_i^{DEA-V}}{d_i^{FDH}} \quad (8)$$

É necessário somente simular a distribuição desta estatística sob a hipótese nula de convexidade ($R = 1$) contra a de não convexidade ($R < 1$) e situar a estatística estimada $\hat{R} = 0,3335$ com relação a esta distribuição.

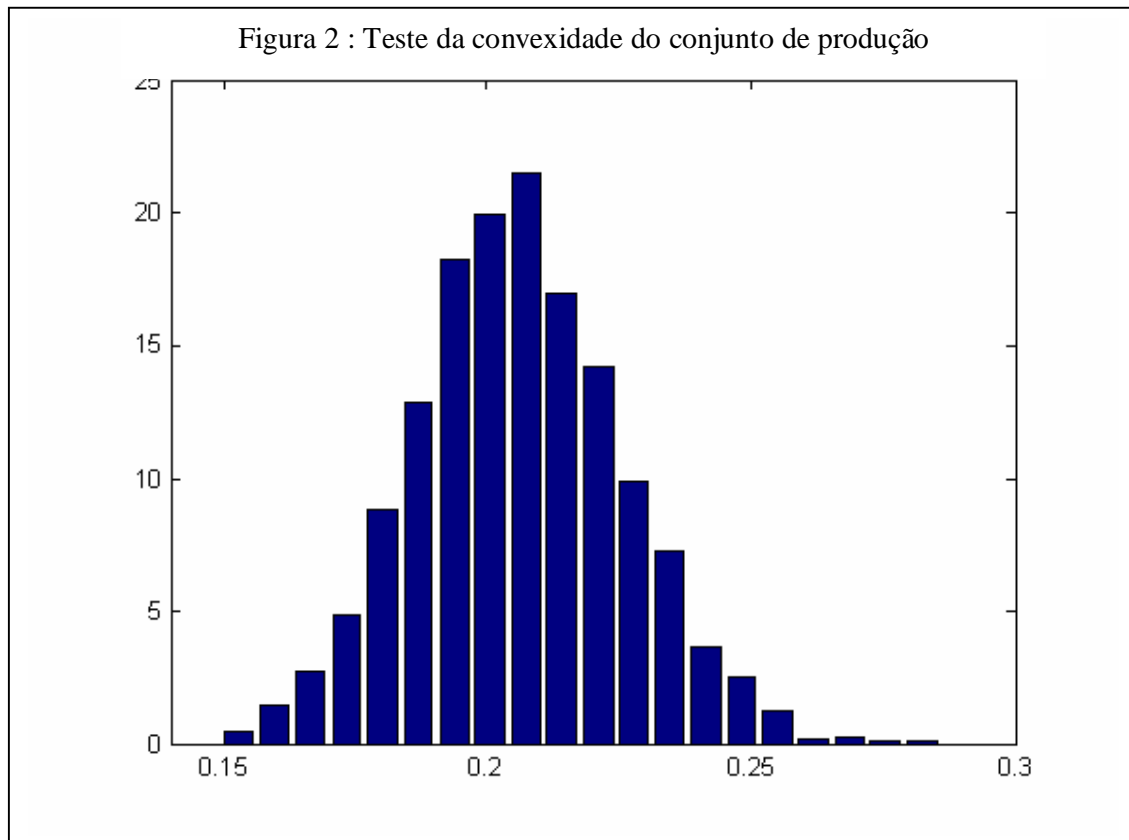
Para isto, utiliza-se o procedimento proposto por Simar et Wilson [1998] e Silverman [1986, páginas 142-147]. Trata-se de estimar a densidade da eficiência com ajuda do estimador proposto por Silverman [1986] e formar amostras bootstrap sob a hipótese nula estimando a eficiência a partir desta distribuição.

A Figura 1 mostra a densidade de kernel para a taxa d_i^{DEA-V} / d_i^{FDH} . Após a exclusão das observações com nível de eficiência igual a 1, observa-se a que a densidade se torna monótona decrescente. Ou seja, a massa de probabilidade se concentra à esquerda, diminuindo à medida que o nível de eficiência aumenta.



Source : CODEVASF.

Este procedimento foi aplicado com 2.000 replicações bootstrap. A Figura 2 dá a estimação da distribuição da estatística de teste sob hipótese nula de convexidade. Normalmente, o estimador FDH é mais viesado que o estimador DEA sob a hipótese nula de uma fronteira DEA-V com distância finita. Isto se deve ao fato que o estimador FDH não explora informações sobre a convexidade do conjunto de produção. O estimador DEA converge mais rápido em direção á fronteira que o estimador FDH. Segundo Kneip et al. [1998], o estimador DEA converge em direção à verdadeira fronteira à uma velocidade dada por $N^{-\frac{2}{q+m+1}}$ para q produtos e m insumos. O estimador FDH converge à $N^{-\frac{1}{q+m}}$. Assim, a taxa d_i^{DEA-V} / d_i^{FDH} convergirá em direção à fronteira à uma taxa dada por $N^{-\frac{q+m-1}{(q+m)(q+m+1)}}$. Desta forma, a variância da distribuição desta taxa sob hipótese nula tenderá à zero à medida que a dimensão da amostra tende ao infinito, para todo $(q, m) \geq (1,1)$. No caso em particular deste artigo, para $q = 1$ e $m = 4$, tem-se a velocidade $\tilde{R} \sim N^{-\frac{2}{15}}$.



Source : CODEVASF.

Como se pode observar na Figura 2, a distribuição é relativamente concentrada devido aos fatores de convergência citados acima. O estimador deste artigo ($\hat{R} = 0,3335$) não pertence a esta distribuição e se situa à sua direita. A hipótese de uma fronteira côncava é rejeitada não importa quais índices de eficiência sejam utilizados na simulação. Se fosse efetuado o mesmo teste supondo um modelo DEA-C, os resultados tenderiam a ser idênticos aos encontrados pelo método DEA-V para as combinações convexas dos fatores de produção observados. Vale lembrar, no entanto, que é a convexidade que está sendo rejeitada, não a hipótese de rendimento de escala.

Pode-se observar também que os níveis de eficiência, após as 2.000 replicações, se concentram no intervalo de eficiência [0-0,3]. Ou seja, eles são reduzidos sob simulação. Normalmente, espera-se melhorar tais resultados com a ajuda do bootstrap. Todavia, como o conjunto de produção não é convexo, é possível que existam múltiplos equilíbrios, o que pode ocasionar alguns problemas quando os dados são submetidos a uma técnica de simulação como o bootstrap. Além do mais, os dados são heterocédásticos, o que pode explicar este resultado contraditório.

4. Conclusões

Diante dos resultados encontrados pelo método DEA-V e pelas simulações com o bootstrap, pode-se afirmar que os colonos do pólo Petrolina/Juazeiro assimilaram a tecnologia da irrigação, embora tenham apresentado certa dificuldade em gerenciar suas despesas com capital e mão-de-obra. O pólo Petrolina/Juazeiro é, sem nenhuma margem de dúvida, uma referência em termo de desenvolvimento para a região do semi-árido nordestino. A disponibilidade de avançados sistemas de irrigação ligados à utilização de insumos modernos pode conduzir o conjunto destes colonos a produções eficientes.

Todavia, mesmo que ele esteja bem posicionado em termos de competição internacional, deve-se melhorar a capacidade de gerar mais renda e amplificar sua capacidade de concorrência ao nível internacional. Ganhos são possíveis através da implementação de toda a infra-estrutura necessária, de modo a antecipar a evolução da demanda e adaptar as técnicas utilizadas à regulamentação, melhorando a formação de capital humano.

Já o método de simulação bootstrap apontou problemas ligados a heterocedasticidade e à não convexidade do conjunto de produção. Tais problemas foram acentuados e vistos no seu verdadeiro ângulo quando os dados foram submetidos à simulação. Isto mostra a importância de se comparar os métodos de estimação de fronteiras de produção convencionais aos resultados encontrados pelo bootstrap. Em alguns casos, os resultados encontrados pelos métodos tradicionais estão viesados e os resultados tendem a divergir dos que seriam encontrados se conhecêssemos a verdadeira fronteira de produção. O método bootstrap vem justamente dar o suporte necessário à identificação de tais problemas e filtrar as informações verdadeiramente pertinentes nos resultados.

Bibliografia

- Adjemian, S. (2002) « Divergence des nations et progrès technique ». *Thèse de doctorat à l'Université d'Evry-Val-d'Essonne*.
- Barros, E. S., E. Costa et Y. Sampaio (2004a) « Analyse de eficiência através da estimação de fronteiras paramétricas Cobb-Douglas et Translog : o caso das empresas agrícolas do polo Petrolina/Juazeiro. » *Revista Econômica do Nordeste*, 35, 1, 7-19.
- Barros, E. S., E. Costa et Y. Sampaio (2004b) « Analise da eficiência das empresas agrícolas do polo Petrolina/Juazeiro utilizando a fronteira paramétrica Translog. » *Revista de Economia e Sociologia Rural, Rio de Janeiro*, 42, 04, 597-614.
- Beran, R. (1986) Discussion of "Jackknife bootstrap and others resampling methods in regression analysis" by C. F. J. Wu. *Annals of Statistics*, 14, 1295-1298.
- Charnes, A. W. W. Cooper, et E. Rhodes. (1978) "Measuring Efficiency of Decision Making Units". *European Journal of Operational Research*, 2, 429-444.
- Deprins, D., Simar, L. & Tulkens, H. (1984) Measuring labor inefficiency in post offices. In: M.
- Färe, R., Grosskopf, S. and Lovell, C. K. (1985) *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff Publishing.
- Farrell, M. J. (1957) "The Measurement of Economic Efficiency". *Journal of the Royal Statistical Society*, 120, Series A Pt. III, 252-281.
- Hall, P. et D. M. Titterton. (1989) "The effect of simulation order on level of accuracy and power of Monte Carlo tests." *J. R. Statistic Society B*, 51, 459-467.
- Hall, P. Hardle et L. Simar (1995) "Iterated bootstrap with applications to frontier models." *Journal of productivity analysis*, 6, 63-76.
- Horrace, W. C. et S. O. Richards (2004) « Bootstrapping efficiency probabilities in parametric stochastic frontier models. » *Syracuse University, NY, (in press)*.
- Jensen, U. (2000) "Is it efficiency to analysis efficiency ranking?" *Empirical Economics*, 25, 189-208.
- Kneip, A., Park, B. & Simar, L. (1998) A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory*, 14, pp. 783± 793.
- Park, B. L. Simar et C. Weiner (2000) « The FDH estimator for productivity efficiency scores : asymptotic properties. » *Econometric Theory*, 16, 855-877.
- Sherpard, R. W. (1970) "Theory of Cost and Production Functions". *Princeton: Princeton University Press*.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis* (London, Chapman and Hall).
- Simar, L. (1992) "Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping." *Journal of productivity analysis*, 6, 63-76.
- Simar, L. & Wilson, P. (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science*, 44, pp. 49± 61.
- Simar, L. & Wilson, P. (1999) Estimating and bootstrapping Malmquist indices, *European Journal of Operational Research*, 115, pp. 459± 471.
- Simar, L.; Wilson, P.W. (2000a) "A general Methodology for Bootstrapping in Nonparametric Frontier Models". *J. Appl. Statist.*, 27, 779–802.
- Simar, L.; Wilson, P.W. (2000b) "Statistical Inference in Nonparametric Frontier Models: The State of the Art." *J. Productivity Anal*, 13, 49–78.