



Universidade de Brasília

Instituto de Ciências Humanas

Departamento de Economia

Mestrado em Economia do Setor Público

Luiz Antônio Zenóbio da Costa

Orientadora : Prof^ª. Dr.^a. Maria da Conceição Sampaio de Sousa

DPF

DETECÇÃO ACELERADA DE PENETRAS
POR INSPEÇÃO DE FRONTEIRAS NÃO PARAMÉTRICAS

BRASÍLIA

2003



Universidade de Brasília

Instituto de Ciências Humanas

Departamento de Economia

Mestrado em Economia do Setor Público

Luiz Antônio Zenóbio da Costa

DPF

DETECÇÃO ACELERADA DE PENETRAS
POR INSPEÇÃO DE FRONTEIRAS NÃO PARAMÉTRICAS

Dissertação apresentada ao Departamento de Economia do Instituto de Ciências Humanas da Universidade de Brasília para obtenção do grau de mestre em Economia.

Orientadora : Prof^ª. Dr^ª. Maria da Conceição Sampaio de Sousa.

BRASÍLIA

2003

Costa, Luiz Antônio Zenóbio da.

DPF: detecção acelerada de penetras por inspeção de fronteiras não paramétricas / Luiz Antônio Zenóbio da Costa. – 2003.

102 p.

Orientadora: Doutora Maria da Conceição Sampaio de Sousa.

Dissertação (mestrado) – Universidade de Brasília, Instituto de Ciências Humanas, 2003.

1. Economia I. Título.

SUMÁRIO

LISTA DAS TABELAS.....	iii
LISTA DOS GRÁFICOS.....	iii
LISTA DAS FIGURAS.....	iv
LISTA DAS DEFINIÇÕES.....	v
LISTA DAS NOTAÇÕES.....	v
LISTA DOS EXEMPLOS.....	vi
LISTA DOS TEOREMAS.....	vi
RESUMO.....	vii
ABSTRACT.....	viii
1. INTRODUÇÃO.....	1
2. MÉTODOS NÃO PARAMÉTRICOS DE ESTIMATIVA DE EFICIÊNCIA.....	5
2.1. Motivação.....	5
2.2. Definições de termos utilizados.....	5
2.3. Método AED-C.....	8
2.4. Método AED-V.....	10
2.5. Método AED-N.....	10
2.6. Método FDH (<i>Free Disposal Hull</i>).....	11
2.7. Principais problemas dos métodos AED.....	12
2.7.1. Erro de especificação.....	13
2.7.2. Sensibilidade em relação a variações nos dados.....	15
2.7.3. Erro de amostragem.....	16
2.7.4. Eficiência não referendada.....	17
2.7.5. Erros nos dados e existência de penetras.....	18
3. DETECÇÃO DE PENETRAS.....	20
3.1. Método de inspeção de observações influentes.....	20
3.2. Método da super eficiência.....	20
3.3. Método da hiper eficiência.....	22
3.4. Método MRE (<i>Robust Efficiency Measurement - REM</i>).....	22
4. DETECÇÃO ACELERADA DE PENETRAS POR INSPEÇÃO DE FRONTEIRAS NÃO PARAMÉTRICAS - DPF.....	24
4.1. Introdução.....	24
4.2. Simplificações aceleradoras.....	25
4.3. Determinação e ordenação das influências.....	26
4.4. Descrição do algoritmo.....	26
4.5. Tempos de execução do método em testes realizados.....	30
5. DETECÇÃO DE PENETRAS E ESTIMAÇÃO DA EFICIÊNCIA TÉCNICA DAS MUNICIPALIDADES BRASILEIRAS.....	35
5.1. Indicadores de insumos e produtos para as municipalidades brasileiras.....	36
5.2. Análise preliminar dos dados dos municípios brasileiros: aplicando o método AED diretamente.....	38
5.2.1. Subamostra contendo 377 municípios da Bahia.....	38
5.2.2. Amostra contendo 4.796 municípios brasileiros.....	42
5.3. Ilustração empírica do método DPF.....	45
5.3.1. Subamostra contendo 377 municípios da Bahia.....	45

5.3.2. Municípios brasileiros	54
6. CONCLUSÃO	66
7. APÊNDICE – TEORIA UTILIZADA	68
7.1. Notações e definições básicas	68
7.2. Medida de eficiência no uso dos insumos	81
7.3. Estimativa de eficiência no uso dos insumos a partir de amostra	81
7.4. Teoremas usados	85
BIBLIOGRAFIA	100

LISTA DAS TABELAS

Tabela 1:	Quantidades e durações na etapa de definição das camadas	31
Tabela 2 :	Quantidades e durações do cálculo MRE (etapa 1).....	32
Tabela 3 :	Quantidades e durações do cálculo MRE (etapa 2).....	32
Tabela 4 :	Quantidades e durações do cálculo MRE (etapa 3).....	32
Tabela 5 :	Durações do cálculo final das eficiências	33
Tabela 6 :	Indicadores de insumos e produtos e serviços municipais correspondentes	37
Tabela 7 :	Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos	38
Tabela 8 :	Comparação da produtividade de Sapeaçu com as dos demais municípios da Bahia – amostra original - um insumo e nove produtos	40
Tabela 9 :	Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original – quatro insumos e nove produtos.....	40
Tabela 10 :	Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos	42
Tabela 11 :	Municípios brasileiros com maiores eficiências – amostra original - um insumo e nove produtos	43
Tabela 12 :	Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - quatro insumos e nove produtos	44
Tabela 13 :	Municípios mais influentes – Bahia com um insumo e nove produtos	46
Tabela 14 :	Teste F para identificar quantidade de penetras – Bahia com um insumo e nove produtos	46
Tabela 15 :	Municípios mais influentes – Bahia com quatro insumos e nove produtos	51
Tabela 16 :	Teste F para identificar quantidade de penetras – Bahia com quatro insumos e nove produtos	51
Tabela 17 :	Quantidades de municípios nas camadas selecionadas – Brasil com um insumo e nove produtos	55
Tabela 18 :	Municípios mais influentes – Brasil com um insumo e nove produtos.....	56
Tabela 19 :	Teste F para identificar quantidade de penetras – Brasil com um insumo e nove produtos.....	56
Tabela 20 :	Quantidades de municípios nas camadas selecionadas – Brasil com quatro insumos e nove produtos	60
Tabela 21 :	Municípios mais influentes – Brasil com quatro insumos e nove produtos	61
Tabela 22 :	Teste F para identificar quantidade de penetras – Brasil com quatro insumos e nove produtos	62

LISTA DOS GRÁFICOS

Gráfico 1 :	Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos	39
Gráfico 2 :	Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original – quatro insumos e nove produtos.....	41
Gráfico 3 :	Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos	42
Gráfico 4 :	Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - quatro insumos e nove produtos	44
Gráfico 5 :	Evolução da soma dos quadrados das ineficiências – Bahia com um insumo e nove produtos.....	47
Gráfico 6 :	Evolução das eficiências de municípios selecionados – Bahia com um insumo e nove produtos ..	48
Gráfico 7 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – municípios da Bahia – um insumo e nove produtos.....	49
Gráfico 8 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 15 municípios mais influentes – municípios da Bahia – um insumo e nove produtos	49
Gráfico 9 :	Distribuição das eficiências retirando 5 penetras – municípios da Bahia – um insumo e nove produtos	50
Gráfico 10 :	Evolução da soma dos quadrados das ineficiências – Bahia com quatro insumos e nove produtos	52
Gráfico 11 :	Evolução das eficiências de municípios selecionados – Bahia com quatro insumos e nove produtos	52

Gráfico 12 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Bahia com quatro insumos e nove produtos	53
Gráfico 13 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 20 municípios mais influentes - Bahia com quatro insumos e nove produtos ..	53
Gráfico 14 :	Distribuição das eficiências dos municípios retirando 11 penetras - Bahia com quatro insumos e nove produtos.....	54
Gráfico 15 :	Evolução da soma dos quadrados das ineficiências – Brasil com um insumo e nove produtos.....	57
Gráfico 16 :	Evolução das eficiências de municípios selecionados – Brasil com um insumo e nove produtos ..	57
Gráfico 17 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Brasil com um insumo e nove produtos	58
Gráfico 18 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 23 municípios mais influentes - Brasil com um insumo e nove produtos.....	58
Gráfico 19 :	Distribuição das eficiências retirando 17 penetras - Brasil com um insumo e nove produtos	59
Gráfico 20 :	Evolução da soma dos quadrados das ineficiências – Brasil com quatro insumos e nove produtos	63
Gráfico 21 :	Evolução das eficiências de municípios selecionados – Brasil com quatro insumos e nove produtos	63
Gráfico 22 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Brasil com quatro insumos e nove produtos	64
Gráfico 23 :	Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 30 municípios mais influentes – Brasil com quatro insumos e nove produtos ..	64
Gráfico 24 :	Distribuição das eficiências retirando 23 penetras– Brasil com quatro insumos e nove produtos ..	65

LISTA DAS FIGURAS

Figura 1 :	CPG de tecnologia convexa	6
Figura 2 :	CPG de tecnologia convexa e com rendimentos de escala constantes	6
Figura 3 :	CPG de tecnologia convexa e com rendimentos de escala não crescentes	7
Figura 4 :	CPG de tecnologia com livre disponibilidade de insumos e produtos	8
Figura 5 :	CPG de tecnologia usada no método AED-C	9
Figura 6 :	CPG de tecnologia usada no método AED-V	10
Figura 7 :	CPG de tecnologia usada no método AED-N	11
Figura 8 :	CPG de tecnologia usada no método FDH	12
Figura 9 :	Erro de especificação	14
Figura 10 :	Sensibilidade em relação a variações nos dados	15
Figura 11 :	Erro de amostragem	17
Figura 12 :	Erro nos dados	19
Figura 13 :	Exemplo de mascaramento de penetras	21
Figura 14 :	Conjunto de produção com tecnologia convexa	73
Figura 15:	CPG(A, c, C).....	75
Figura 16:	CPG(A, c, N)	76
Figura 17:	CPG(A, c, V)	77
Figura 18:	CPG(A, DIn).....	77
Figura 19:	CPG(A, DPr).....	78
Figura 20:	CPG(A, AED-FDH).....	78
Figura 21:	CPG(A, AED-C).....	79
Figura 22:	CPG(A, AED-N).....	80
Figura 23:	CPG(A, AED-V).....	80

LISTA DAS DEFINIÇÕES

Definição 1 :	Seqüência de elementos de um conjunto C	68
Definição 2 :	Insumos : In	69
Definição 3 :	Produtos : Pr	69
Definição 4 :	Universo de possibilidades de produção : \mathbf{U}	69
Definição 5 :	Conjunto de produção de uma dada tecnologia : $P(T)$	69
Definição 6 :	Correspondência de insumos : $In(y T)$	70
Definição 7 :	Correspondência de produtos : $Pr(x T)$	70
Definição 8 :	Complemento de um conjunto de produção : $P^c(T)$	70
Definição 9 :	Bola de raio δ e centro em $(\mathbf{x}^0, \mathbf{y}^0)$: $B(\mathbf{x}^0, \mathbf{y}^0, \delta)$	70
Definição 10 :	Interior de um conjunto de produção $P(T)$: $Int(P(T))$	70
Definição 11 :	Fronteira de um conjunto de produção $P(T)$: $Front(P(T))$	70
Definição 12 :	$Front(P(T)) \subset P(T)$	71
Definição 13 :	Conjunto amostra de produção de uma dada tecnologia : A	71
Definição 14 :	Matriz de insumos de uma amostra : X	71
Definição 15 :	Matriz de produtos de uma amostra : Y	71
Definição 16 :	Vetor unitário da dimensão k : \mathbf{e}^k	71
Definição 17 :	Diagonal principal do K -cubo unitário : \mathbf{e}	72
Definição 18 :	Combinação linear convexa de insumos de uma amostra : Xz	72
Definição 19 :	Combinação linear convexa de produtos de uma amostra : Yz	72
Definição 20 :	Tecnologia convexa	72
Definição 21 :	Tecnologia com rendimentos de escala constantes : C	74
Definição 22 :	Tecnologia com rendimentos de escala não crescentes : N	74
Definição 23 :	Tecnologia com rendimentos de escala variáveis : V	74
Definição 24 :	Tecnologia com livre disponibilidade de insumos : DIn	74
Definição 25 :	Tecnologia com livre disponibilidade de produtos : DPr	74
Definição 26 :	$CPG(A, a, b, c, \dots)$	74
Definição 27 :	Fronteira de eficiência de insumos (no sentido de Farrell)	81
Definição 28 :	Função grau de eficiência de insumos : $E_i(x, y T)$	81
Definição 29 :	Função grau de eficiência de insumos : $E_i(x, y T)$	81
Definição 30 :	Estimativa de eficiência - caso geral	81
Definição 31 :	Estimativa de eficiência - caso AED-C.....	82
Definição 32 :	Estimativa de eficiência - caso AED-N.....	82
Definição 33 :	Estimativa de eficiência - caso AED-V.....	83
Definição 34 :	Estimativa de eficiência - caso AED-FDH.....	84
Definição 35 :	Dominância de produção.....	85

LISTA DAS NOTAÇÕES

Notação 1 :	Vetores em \mathfrak{R}^N e matrizes-coluna $N \times 1$ em \mathfrak{R}	68
Notação 2 :	Seqüências de números reais ou de vetores em \mathfrak{R}^N	68
Notação 3 :	Coordenadas de vetores em \mathfrak{R}^N	68
Notação 4 :	Matrizes $M \times N$ de números reais	69
Notação 5 :	Elementos de matrizes	69
Notação 6 :	Desigualdade entre vetores : $\mathbf{x} > \mathbf{y}$, $\mathbf{x} \geq \mathbf{y}$, $\mathbf{x} < \mathbf{y}$, $\mathbf{x} \leq \mathbf{y}$	69

LISTA DOS EXEMPLOS

Exemplo 1 :	\mathbf{x}^k (k-ésimo elemento de uma seqüência de vetores).....	68
Exemplo 2 :	y^k (k-ésimo elemento de uma seqüência de números reais)	68
Exemplo 3 :	x_n (n-ésima coordenada do vetor \mathbf{x}).....	68
Exemplo 4 :	x_n^k (n-ésima coordenada do k-ésimo vetor \mathbf{x}^k).....	68
Exemplo 5 :	X, Y (matrizes).....	69
Exemplo 6 :	x_{mn} (elemento da linha m e coluna n).....	69
Exemplo 7 :	Conjunto de produção com tecnologia convexa.....	72
Exemplo 8 :	$CPG(A, c, C)$	75
Exemplo 9 :	$CPG(A, c, N)$	76
Exemplo 10 :	$CPG(A, c, V)$	76
Exemplo 11 :	$CPG(A, DIn)$	77
Exemplo 12 :	$CPG(A, DPr)$	78
Exemplo 13 :	$CPG(A, DIn, DPr)$ ou $CPG(A, AED-FDH)$	78
Exemplo 14 :	$CPG(A, c, C, DIn, DPr)$ ou $CPG(A, AED-N)$	79
Exemplo 15 :	$CPG(A, c, N, DIn, DPr)$ ou $CPG(A, AED-N)$	79
Exemplo 16 :	$CPG(A, c, V, DIn, DPr)$ ou $CPG(A, AED-V)$	80

LISTA DOS TEOREMAS

Teorema 1 :	$P(T) \cup P^c(T) = \mathbf{U}$	85
Teorema 2 :	$P(T) \cap P^c(T) = \Phi$	85
Teorema 3 :	$(P^c(T))^c = P(T)$	85
Teorema 4 :	$Front(P(T)) = Front(P^c(T))$	85
Teorema 5 :	$Front(P(T)) \cap Int(P(T)) = \Phi$	86
Teorema 6 :	$Int(P(T)) \subset P(T)$	86
Teorema 7 :	$Front(P(T)) \cup Int(P(T)) \cup Int(P^c(T)) = \mathbf{U}$	87
Teorema 8 :	$Int(P(T)) = P(T) - Front(P(T))$	87
Teorema 9 :	$\{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} \neq \Phi$	88
Teorema 10 :	$E_i(\mathbf{x}, \mathbf{y} \mid T) \in [0, 1]$	88
Teorema 11 :	$(\mathbf{x}, \mathbf{y}) \in P(S)$ e $P(S) \subset P(T) \Rightarrow E_i(\mathbf{x}, \mathbf{y} \mid S) \geq E_i(\mathbf{x}, \mathbf{y} \mid T)$	89
Teorema 12 :	$\mathbf{x} \in Front_i(\mathbf{y} \mid T) \Leftrightarrow E_i(\mathbf{x}, \mathbf{y} \mid T) = 1$	89
Teorema 13 :	$E_i(\mathbf{x}, \mathbf{y} \mid T) = 1 \Rightarrow (\mathbf{x}, \mathbf{y}) \in Front(P(T))$	90
Teorema 14 :	$\mathbf{x} \in Front_i(\mathbf{y} \mid T) \Rightarrow (\mathbf{x}, \mathbf{y}) \in Front(P(T))$	90
Teorema 15 :	$\theta^* = E_i(\mathbf{x}, \mathbf{y} \mid T) \Rightarrow (\theta^*\mathbf{x}, \mathbf{y}) \in Front(P(T))$	91
Teorema 16 :	$\theta^* = E_i(\mathbf{x}, \mathbf{y} \mid T) \Rightarrow \theta^* \in \{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$	91
Teorema 17 :	$CPG(A, AED-C), \theta \in \mathfrak{R}_+ \subset P(T)$	93
Teorema 18 :	$E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, AED-C) \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid T)$	93
Teorema 19 :	Se $B \subset A \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k \mid B, AED-C) \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, AED-C)$	93
Teorema 20 :	$(\mathbf{x}^k, \mathbf{y}^k) \in Front(A) \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, AED-C) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid Front(A), AED-C)$	94
Teorema 21 :	$(\mathbf{x}^k, \mathbf{y}^k) \in Int(A) \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, AED-C) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^k, \mathbf{y}^k)\}, AED-C)$	94
Teorema 22 :	$E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, AED-C) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid Front(A), AED-C)$	97

RESUMO

Apresentação do método Detecção Acelerada de Penetras por Inspeção de Fronteiras não Paramétricas – DPF desenvolvido para detecção de erros e discrepâncias nos dados de amostras contendo grandes quantidades de Unidades Decisórias – UDs (*Decision Making Units* - DMUs) objetivando o cálculo da eficiência no uso de insumos por um dos métodos de Análise de Envoltória de Dados – AED (*Data Envelopment Analysis* - DEA). O conjunto a ser pesquisado é reduzido às unidades localizadas nas camadas mais externas da amostra. O método Medição Robusta de Eficiência – MRE (*Robust Efficiency Measurement* – REM) é utilizado para classificar as UDs que mais influenciam o cálculo das eficiências das demais unidades. O método foi aplicado a quatro amostras, sendo duas com 377 UDs e duas com 4.796 UDs. Os testes comprovam a viabilidade do método em todos os casos, sendo que o processamento da amostra mais complexa foi feito em 26 horas, incluindo o tempo de processamento e de atividades manuais. Os resultados permitiram validar também o critério para estimar a quantidade de UDs com dados discrepantes, baseado na hipótese de que as razões das somas dos quadrados das ineficiências seguem uma distribuição F.

Palavras-chave

Análise de Envoltória de Dados - AED

Medição Robusta de Eficiência - MRE

Medida de eficiência não paramétrica

Detecção de erros e discrepâncias nos dados

ABSTRACT

Presentation of Accelerated Outlier Detection by Inspection of Nonparametric Frontiers (Detecção Acelerada de Penetras por Inspeção de Fronteiras não Paramétricas – DPF) method for outlier detection in samples containing large quantities of Decision Making Units - DMUs in order to calculate input efficiency using one of Data Envelopment Analysis – DEA methods. The set of DMUs to be inspected is reduced to those lying in the sample's outer layers. Robust Efficiency Measurement – REM is used to sort the most influential DMUs when calculating input efficiency. Four samples were analyzed by this method, two with 377 DMUs and two with 4.796 DMUs. The method proved to be viable in all tested samples and the worst case was processed in 26 hours, including computation and manual activities time. The criterium for estimating outlier quantity, based on the assumption that the ratio of the sums of the squares of the inefficiencies follow an F distribution, was also validated.

Key words

Data Envelopment Analysis - DEA

Robust Efficiency Measurement – REM

Nonparametric efficiency measurement

Outlier detection

1. INTRODUÇÃO

A necessidade de se medir a eficiência relativa de firmas, instituições, ou qualquer tipo de unidade de interesse, surge em diversas situações práticas. Consultores que desejam otimizar a operação de uma firma se interessam em identificar quais são as melhores práticas existentes no mercado. Instituições governamentais de controle, como o Tribunal de Contas da União, têm a incumbência de analisar o correto e eficiente uso dos recursos colocados à disposição das instituições pela sociedade. Agências fiscalizadoras ou reguladoras, cada vez mais presentes no mundo atual, precisam desenvolver parâmetros capazes de medir a eficiência das empresas reguladas, não só durante a fase normal de operação, mas principalmente nos processos de revisão de preços ou nos processos licitatórios de concessão de novos empreendimentos. Em geral, dispõem-se de dados, mas pouco conhecimento sobre os detalhes dos processos produtivos e das funções de produção que caracterizam as operações das empresas ou instituições. A abordagem mais utilizada é a análise individual, envolvendo o envio de equipes de auditoria que passam diversos dias nas unidades analisadas, envolvendo enormes custos. No caso das instituições de controle, devido à enorme quantidade de locais a serem auditados, a auditoria é feita anualmente em pequenas amostras, determinadas por critérios estatísticos ou em função de denúncias e outros tipos de indícios detectados. Portanto, a possibilidade de se analisar, de forma sistemática e automática, esses grandes volumes de dados disponíveis, detectando *a priori* os casos mais prováveis a receberem auditorias individualizadas, pode representar uma enorme economia de tempo e dinheiro, bem como um aumento na eficácia e abrangência do controle.

Uma forma de analisar a eficiência produtiva das unidades decisórias (UDs ou *Decision Making Units* - DMUs) é supor que elas otimizam seus resultados segundo limitações impostas pela tecnologia. Nessa abordagem, as variações observadas em torno da fronteira de produção são creditadas a flutuações estocásticas. Ampliando o raciocínio e admitindo que nem todas as UD's são eficientes, podemos estimar a

eficiência de uma UD, comparando seus resultados com os valores da fronteira de produção da tecnologia. As fronteiras de produção podem ser estimadas a partir de dados de uma amostra usando métodos paramétricos ou não paramétricos.

As técnicas paramétricas partem de especificações funcionais, cujos parâmetros são estimados usando informações da tecnologia ou ajustando a função aos dados da amostra. Essa abordagem é muito dependente da correta especificação do tipo de função a ser utilizada na estimação da fronteira de produção. Quando a análise envolve diversos tipos de insumos e produtos, a especificação da função depende de se estabelecer pesos para que se possa agregar os diversos dados em uma fórmula de cálculo da eficiência. Devido a esta característica, esses parâmetros são muito difíceis de serem estimados e dependem muito de aspectos subjetivos. Em geral, conhecemos apenas parte do conjunto produtivo e não temos informações sobre a forma específica assumida pela fronteira desse conjunto. Nesse caso, os resultados obtidos são afetados tanto pelos erros contidos nos dados como pelo erro de especificação da função de produção.

As abordagens não paramétricas, ao invés de se basearem em uma função especificada *a priori*, consideram apenas que o conjunto de produção deva satisfazer determinadas propriedades tais como: (1) livre disponibilidade de insumos e produtos; (2) convexidade (envolve a noção de UDs virtuais que podem ser construídas a partir de combinações lineares convexas das UDs da amostra); e (3) rendimentos de escala (que podem ser constantes, não crescentes ou variáveis). Dentre os métodos não paramétricos, a Análise de Envoltória de Dados - AED (*Data Envelopment Analysis* – DEA, Charnes et al. (1978)) é um dos mais estudados. Originalmente tratando de tecnologias com retornos constantes de escala, a AED foi expandida para tratar de tecnologias com ganhos não decrescentes ou variáveis de escala. As diferentes variantes dessa metodologia permitem calcular a eficiência relativa das UDs.

Porém, em razão de seu caráter determinístico, o uso adequado da AED requer uma descrição acurada do conjunto produtivo já que os escores computados de eficiência dependem, crucialmente, da qualidade da informação utilizada. Como na prática, as amostras são incompletas e contaminadas por diferentes tipos de erros –

erros de mensuração, de agregação, omissão de variáveis relevantes – isto constitui uma das mais sérias restrições ao uso dessa metodologia. Faz-se, pois, necessário desenvolver técnicas que permitam corrigir esses problemas.

Nesse contexto, diversos métodos foram propostos para detectá-los, como os métodos de super e hiper eficiências. Dentre eles, a Medição Robusta de Eficiência – MRE (*Robust Efficiency Measurement* – REM), apresentada por Kuosmanen e Post (1999), oferece uma abordagem interessante para estimar a eficiência a partir de amostras de produção contendo erros de informação. Contudo, o algoritmo, da forma como foi apresentado, pode ser impraticável em amostras muito grandes, em função do altíssimo volume de cálculo exigido para sua implementação. Outra limitação para sua aplicação em instituições governamentais é que se concentra apenas em estimar a eficiência das UDs, sem descrever critério para identificação e eliminação de UDs com erros.

Ademais, os trabalhos analisados, tanto sobre MRE, quanto sobre a super e a hiper eficiências, exemplificam seus cálculos com exemplos contendo até 100 UDs. As instituições governamentais constantemente se deparam com situações em que é necessário comparar centenas ou milhares de UDs.

O objetivo do presente trabalho é apresentar uma metodologia que possa detectar e eliminar observações discrepantes e obter estimativas robustas para a eficiência de UDs em grandes amostras. O método DPF, Detecção Acelerada de Penetras por Inspeção de Fronteiras não Paramétricas, desenvolvido neste trabalho, parte dos resultados introduzidos pela MRE e faz algumas adaptações, usando resultados de teoremas demonstrados no apêndice, para acelerar os cálculos em amostras com grandes quantidades de UDs. A partir dos teoremas demonstrados, o método trabalha com subamostras contendo apenas as UDs contidas nas fronteiras mais externas da amostra original. O algoritmo desenvolvido foi testado em dois conjuntos de dados, um contendo 377 UDs e outro contendo 4.796 UDs. Para cada conjunto, as eficiências foram calculadas usando 9 produtos e 1 insumo e 9 produtos e 4 insumos.

O trabalho é desenvolvido da seguinte forma : no capítulo 2, são descritos os principais métodos não paramétricos e seus principais problemas; no capítulo 3 são apresentados os principais métodos utilizados para a detecção de erros; no capítulo 4 é apresentado o método DPF, juntamente com seu embasamento teórico; no capítulo 5 são apresentados os resultados para os casos analisados; no capítulo 6, são apresentadas as conclusões e as sugestões para estudos posteriores; o capítulo 7 contém um apêndice com as notações, definições e teoremas utilizados neste trabalho.

2. MÉTODOS NÃO PARAMÉTRICOS DE ESTIMATIVA DE EFICIÊNCIA

2.1. Motivação

Por sua simplicidade e pouca dependência em conhecimentos detalhados da tecnologia, os métodos não paramétricos, em especial os de AED, têm-se mostrado muito úteis na avaliação da eficiência de instituições e empresas. No entanto, estes métodos são muito sensíveis aos erros contidos nos dados da amostra sendo analisada. Apesar de pouco estudado no início do desenvolvimento da teoria de AED, nos anos recentes a gravidade deste problema foi reconhecida e diversos autores apresentaram artigos envolvendo estudos sobre a influência destes erros, bem como a detecção e retiradas dos dados problemáticos. Os métodos não paramétricos partem dos dados de insumos e produtos de uma amostra de UD's e de algumas hipóteses sobre as propriedades da tecnologia utilizada para estimarem a fronteira do conjunto de produção real. A seguir, são definidas as principais propriedades utilizadas nas definições desses métodos.

2.2. Definições de termos utilizados

Tecnologia convexa: Sejam T uma tecnologia e $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\} \subset P(T)$ um conjunto amostra de produção. Dizemos que T é uma tecnologia convexa se $(Xz, Yz) \in P(T)$, sempre que $z \in \mathfrak{R}_+^K$ e que $e^t z = 1$. Nessa definição, Xz e Yz são combinações lineares convexas dos insumos e produtos da amostra A , a saber:

$$Xz = \sum_{i=1}^K z_i x^i = \left(\sum_{i=1}^K z_i x_1^i, \sum_{i=1}^K z_i x_2^i, \dots, \sum_{i=1}^K z_i x_N^i \right) \in \mathfrak{R}_+^N$$

$$z \in \mathfrak{R}_+^K, \quad e^t z = 1$$

$$Yz = \sum_{i=1}^K z_i y^i = \left(\sum_{i=1}^K z_i y_1^i, \sum_{i=1}^K z_i y_2^i, \dots, \sum_{i=1}^K z_i y_M^i \right) \in \mathfrak{R}_+^M$$

$$z \in \mathfrak{R}_+^K, \quad e^t z = 1$$

A figura a seguir ilustra o conjunto de produção gerado (CPG) a partir de uma amostra contendo 3 UDs, que usam uma tecnologia convexa, 1 insumo e 1 produto.

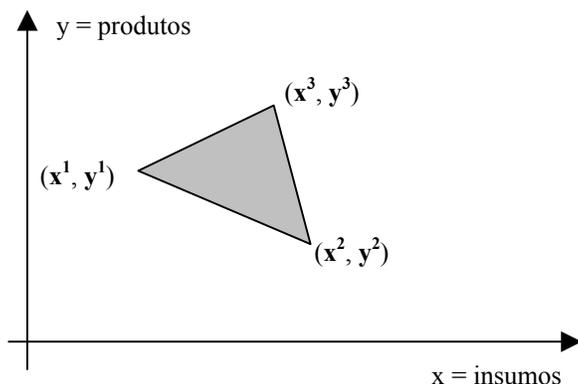


Figura 1 : CPG de tecnologia convexa

Tecnologia com rendimentos de escala constantes: Sejam T uma tecnologia, $P(T)$ seu conjunto de produção e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala constantes se $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$, sempre que $\alpha \in \mathbb{R}_+$. Ou seja, se (\mathbf{x}, \mathbf{y}) é factível, então qualquer dos seus múltiplos ou submúltiplos positivos também será factível. A figura a seguir ilustra o conjunto de produção gerado (CPG) a partir de uma amostra contendo 3 UDs, que usam uma tecnologia convexa, com rendimentos de escala constantes, 1 insumo e 1 produto.

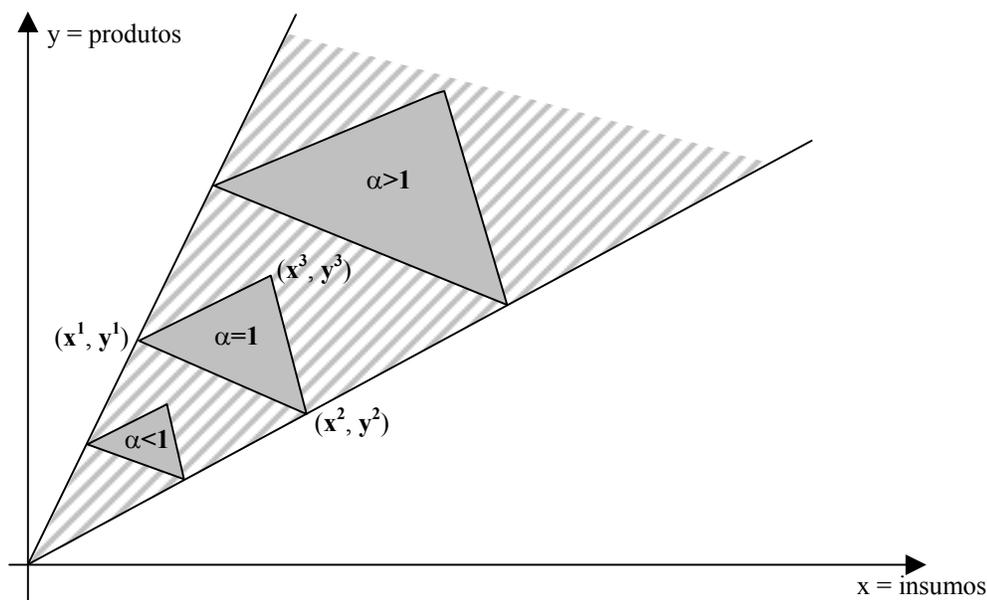


Figura 2 : CPG de tecnologia convexa e com rendimentos de escala constantes

Tecnologia com rendimentos de escala não crescentes: Sejam T uma tecnologia, $P(T)$ seu conjunto de produção e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala não crescentes se $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$, sempre que $0 < \alpha \leq 1$. Ou seja, se (\mathbf{x}, \mathbf{y}) é factível, então qualquer dos seus submúltiplos (neste caso, não valem mais os múltiplos) positivos também será factível. A figura a seguir ilustra o conjunto de produção gerado (CPG) a partir de uma amostra contendo 3 UD's, que usam uma tecnologia convexa, com rendimentos de escala não crescentes, 1 insumo e 1 produto.

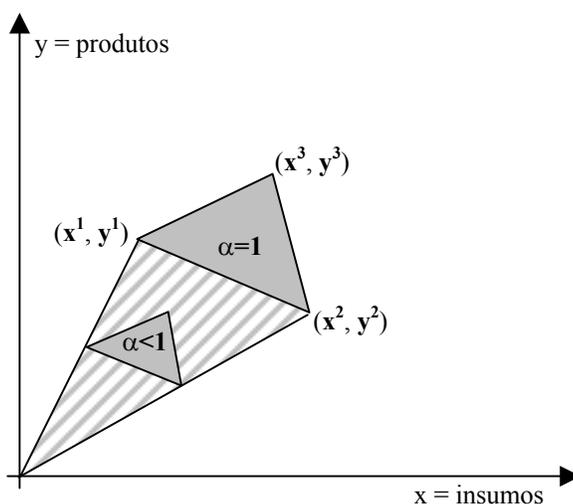


Figura 3 : CPG de tecnologia convexa e com rendimentos de escala não crescentes

Tecnologia com rendimentos de escala variáveis: Sejam T uma tecnologia, $P(T)$ seu conjunto de produção e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala variáveis se podemos garantir que $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$ apenas quando $\alpha = 1$. Ou seja, não podemos garantir a viabilidade nem dos múltiplos nem dos submúltiplos de (\mathbf{x}, \mathbf{y}) . A figura que ilustra este caso é a mesma da tecnologia convexa.

Tecnologia com livre disponibilidade de insumos: Sejam T uma tecnologia, $P(T)$ seu conjunto de produção e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com livre disponibilidade de insumos se $(\mathbf{z}, \mathbf{y}) \in P(T)$, sempre que $\mathbf{z} \geq \mathbf{x}$. Ou seja, aumentar os insumos é sempre factível.

Tecnologia com livre disponibilidade de produtos: Sejam T uma tecnologia, $P(T)$ seu conjunto de produção e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com livre disponibilidade de produtos se $(\mathbf{x}, \mathbf{z}) \in P(T)$, sempre que $\mathbf{z} \leq \mathbf{y}$. Ou seja, diminuir os produtos é sempre factível.

A figura a seguir ilustra o conjunto de produção gerado (CPG) a partir de uma amostra contendo 3 UD's, que usam uma tecnologia com livre disponibilidade de insumos e produtos, 1 insumo e 1 produto.

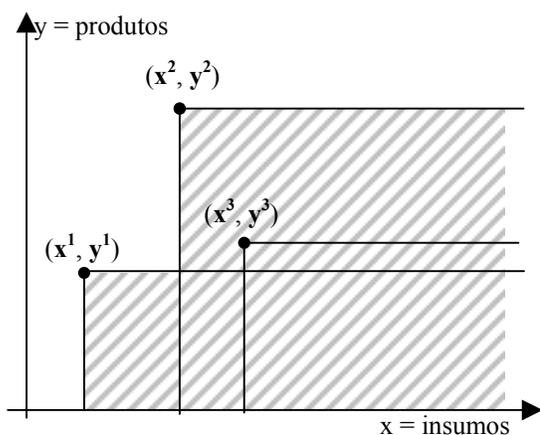


Figura 4 : CPG de tecnologia com livre disponibilidade de insumos e produtos

2.3. Método AED-C

A Análise de Envoltória de Dados – AED foi inicialmente descrita por Farrell (1957) e operacionalizada por Charnes, Cooper e Rhodes (1978). São métodos não paramétricos, que independem da especificação de uma função de produção. Todos supõem tecnologia convexa, com livre disponibilidade de insumos e produtos. O método original, usualmente referido por CCR, pressupõe tecnologia com rendimentos de escala constantes, razão pela qual o denominamos de método AED-C. Partindo dos dados relativos a um conjunto de UD's, o conjunto de produção gerado a partir da amostra e das propriedades da tecnologia pode ser descrito como :

$$\text{CPG}(A, \text{AED-C}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

Na expressão acima, X e Y são matrizes cujas colunas são os vetores dos insumos (X) e produtos (Y) das UD's contidas na amostra.

A figura a seguir, combina as definições de tecnologia convexa, de tecnologia com rendimentos constantes de escala e de tecnologia com livre disponibilidade de insumos e produtos, para ilustrar o conjunto de produção gerado com as hipóteses do método AED-C, usando 1 insumo e 1 produto.

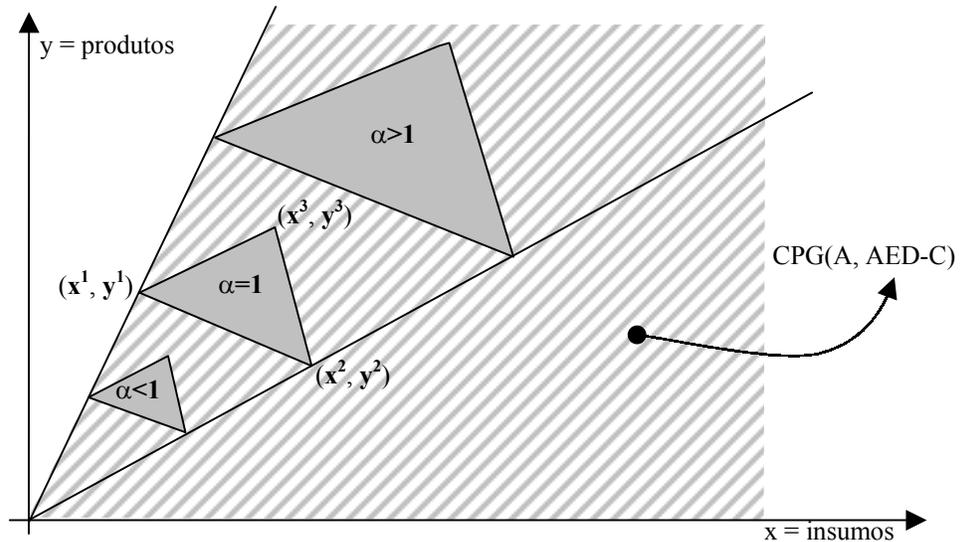


Figura 5 : CPG de tecnologia usada no método AED-C

A eficiência no uso de insumos de uma UD, cujos dados são (x, y) , relativamente ao conjunto de produção gerado pelas observações da amostra, é a menor proporção dos insumos x da UD capaz de gerar os produtos y . Esta definição pode ser apresentada da forma a seguir :

$$\theta^{\text{AED-C}}(x, y) = \min \{ \theta \mid \theta x \geq Xz, y \leq Yz, \theta \in \mathcal{R}_+, x \in \mathcal{R}_+^N, y \in \mathcal{R}_+^M, z \in \mathcal{R}_+^K \}.$$

Para obtermos a eficiência de uma observação (x^k, y^k) , precisamos resolver o seguinte problema de minimização, que pode ser resolvido utilizando-se o método simplex :

$$\theta^{\text{DEA-C}} = \min \theta \quad \text{s.a.} \quad \begin{cases} \theta x_n^k \geq \sum_{j=1}^K z_j x_n^j & n = 1, 2, \dots, N \\ y_m^k \leq \sum_{j=1}^K z_j y_m^j & m = 1, 2, \dots, M \\ z_j \geq 0 & j = 1, 2, \dots, K \\ \theta \geq 0 \end{cases}$$

2.4. Método AED-V

Posteriormente, Banker, Charness e Cooper (1984) propuseram um modelo, usualmente denominado de BCC, que adota suposições similares ao modelo anterior, mudando apenas para tecnologia com rendimentos de escala variáveis, razão pela qual o denominamos de método AED-V. Matematicamente, o que muda é o acréscimo de uma nova restrição relativa aos pesos \mathbf{z} :

$$\mathbf{e}^t \mathbf{z} = z_1 + z_2 + \dots + z_k = 1$$

A figura a seguir, combina as definições de tecnologia convexa, de tecnologia com rendimentos variáveis de escala e de tecnologia com livre disponibilidade de insumos e produtos, para ilustrar o conjunto de produção gerado com as hipóteses do método AED-V, usando 1 insumo e 1 produto.

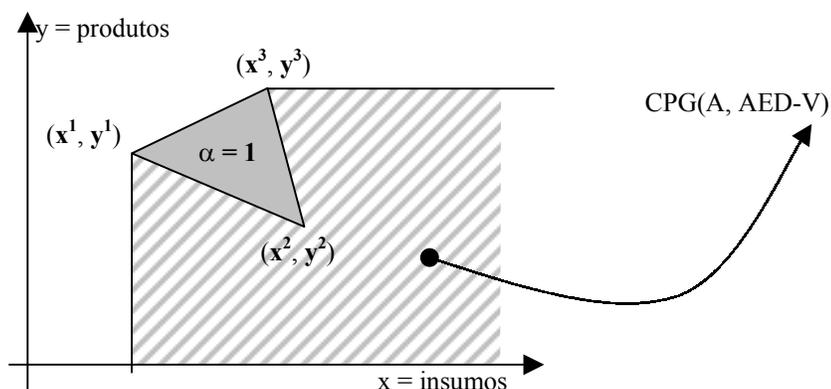


Figura 6 : CPG de tecnologia usada no método AED-V

2.5. Método AED-N

O método AED-N é análogo, pressupondo tecnologia com rendimentos de escala não crescentes. Matematicamente, o que muda é a substituição da restrição relativa aos pesos \mathbf{z} para :

$$\mathbf{e}^t \mathbf{z} = z_1 + z_2 + \dots + z_k \leq 1$$

A figura a seguir, combina as definições de tecnologia convexa, de tecnologia com rendimentos não crescentes de escala e de tecnologia com livre disponibilidade de

insumos e produtos, para ilustrar o conjunto de produção gerado com as hipóteses do método AED-N, usando 1 insumo e 1 produto.

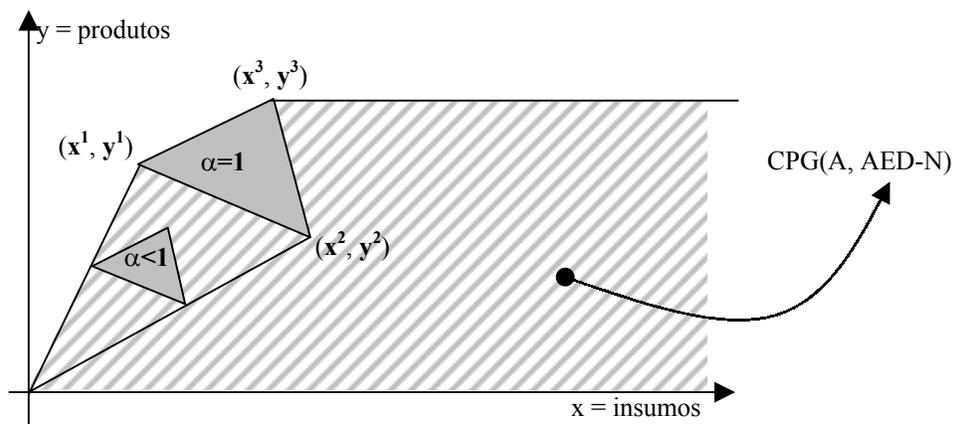


Figura 7 : CPG de tecnologia usada no método AED-N

2.6. Método FDH (*Free Disposal Hull*)

O método FDH (Tulkens, 1990) desconsidera as hipóteses de convexidade e rendimentos de escala, mantendo apenas a de livre disponibilidade de insumos e produtos. Comparativamente aos métodos AED anteriormente apresentados, o que muda é a substituição da restrição relativa aos pesos \mathbf{z} para :

$$\mathbf{e}^t \mathbf{z} = z_1 + z_2 + \dots + z_k = 1, \mathbf{z} \in \{0, 1\}^K.$$

Neste método, a determinação da eficiência de uma UD se dá comparando-a com uma e apenas uma UD dentre as demais UD's da amostra. Uma das consequências desta restrição na comparação entre UD's é a maior ocorrência de eficiência não referendada (eficiência por *default*), caracterizada por não haver uma UD diferente que satisfaça as condições do método, que equivalem às condições de dominância entre UD's. Neste caso a UD é considerada eficiente apenas porque não pode ser comparada com nenhuma outra da amostra. No entanto, devido a esta característica em que o cálculo da eficiência é direto, sem necessidade de se criar o conceito de UD virtual, resultante de uma combinação linear convexa de outras UD's, o método FDH pode ser muito útil na identificação de UD's ineficientes. Supondo inexistência de erros nos dados, a comprovação da ineficiência se dá de forma clara e inquestionável. Esta

característica pode ser muito útil para agências governamentais de controle, como o caso do Tribunal de Contas da União, por dar mais confiança de que as UD's consideradas ineficientes o são de fato, inclusive com a apresentação das UD's dominantes, que servem de comparação e confirmação.

Pelas hipóteses do método FDH, a figura que ilustra o conjunto de produção gerado é a mesma que foi apresentada no caso de tecnologia com livre disponibilidade de insumos e produtos.

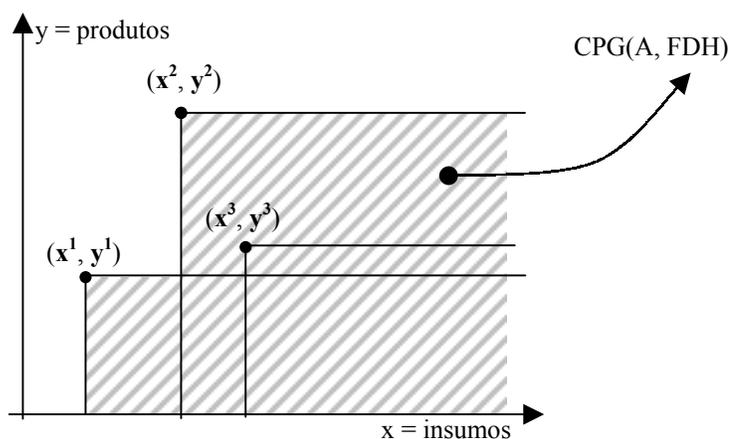


Figura 8 : CPG de tecnologia usada no método FDH

Neste exemplo gráfico, podemos apresentar alguns conceitos utilizados ao longo deste trabalho. A UD 3 é dominada pela UD 2, pois produz menos consumindo mais. Nesta comparação, podemos afirmar que a UD 2 é mais eficiente do que a UD 3. Considerando apenas as 3 UD's, o método FDH irá concluir que as UD's 1 e 3 são eficientes (eficiências iguais a 1) e que a UD 2 é ineficiente (eficiência menor que 1). No caso da UD 1, diferentemente do caso da UD 2, a eficiência calculada é 1 como consequência da inexistência de outras UD's que possam ser comparadas com ela. Designaremos esta situação por eficiência não referendada (também designada na literatura por eficiência por *default*).

2.7. Principais problemas dos métodos AED

A grande vantagem dos métodos AED é a sua característica não paramétrica, que permite a estimativa da eficiência a partir de suposições mínimas a respeito da

tecnologia de produção. Os métodos partem do princípio de que o conjunto gerado a partir da amostra de dados e das características da tecnologia é suficientemente próximo do real conjunto de produção. De maneira simplificada, se os dados possuem uma distribuição com probabilidade não nula sobre todo o conjunto de produção, a eficiência estimada pode ter a precisão que desejarmos, desde que tenhamos uma quantidade suficiente de dados na amostra e que esses dados sejam efetivamente representativos do conjunto de produção, ou seja, não contenham erros ou UDs que se utilizam de diferentes tecnologias.

Originalmente, esses métodos foram desenvolvidos e utilizados segundo uma orientação voltada para aplicações, com pouca atenção às suposições subjacentes dos modelos. No entanto, mais recentemente, tem-se reconhecido a importância de se estudar os problemas que podem advir da não observância das suposições desses tipos de método. Apesar de não muito explicitadas nos estudos iniciais, essas suposições são as seguintes:

- Propriedades da tecnologia (convexidade, rendimento de escala, etc.);
- Amostra representativa do conjunto de produção.

Os problemas aparecem, portanto, quando essas suposições não são atendidas. A seguir apresentamos uma breve descrição dos principais problemas.

2.7.1. Erro de especificação

Apesar de não depender da especificação de uma função de produção, os métodos AED partem da suposição de que a tecnologia possui determinadas propriedades. Se a tecnologia não segue as propriedades escolhidas pelo método utilizado, o conjunto de produção gerado será diferente do conjunto real, a fronteira estimada será incorreta e as eficiências calculadas poderão ser inconsistentes.

A maneira mais direta de se minimizar este tipo de erro é o perfeito conhecimento da tecnologia utilizada. Quando isto não for de todo possível, a alternativa é contar com uma amostra suficientemente grande para se obter uma melhor aproximação do conjunto de produção, mesmo com premissas erradas sobre as

propriedades da tecnologia. No caso limite, em que se está inseguro sobre as características da tecnologia, pode-se restringir o cálculo da eficiência às UD's com produções próximas às da UD analisada, ou dividir a amostra em subamostras.

A figura a seguir, ilustra este tipo de problema e a “solução” mais adequada, em função dos dados disponíveis. Neste exemplo, como a tecnologia não é convexa, a fronteira estimada pelo método AED-V faz com que a eficiência calculada seja muito diferente da eficiência real. No entanto, se temos uma amostra suficientemente representativa do conjunto de produção, a fronteira estimada ao se usar apenas as UD's com produções mais próximas da UD analisada pode ser bastante próxima da fronteira real. A eficiência estimada será, então, uma boa aproximação da eficiência real.

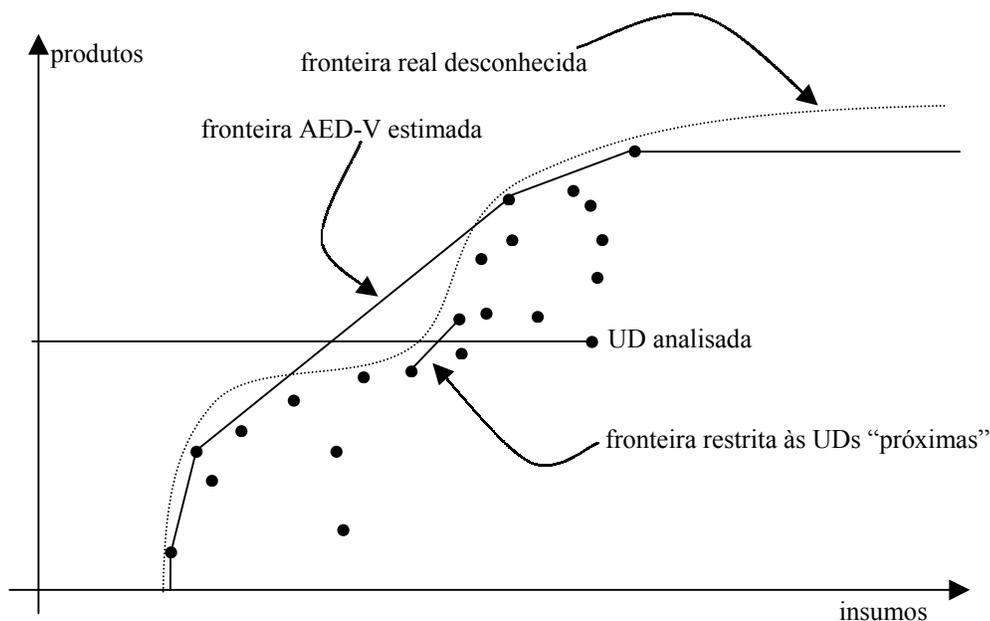


Figura 9 : Erro de especificação

Apesar da aparente simplicidade do exemplo gráfico, minimizar este tipo de problema em grandes amostras e com vários insumos e produtos pode ser praticamente inviável. Este tipo de problema é pouco tratado pelos estudiosos do assunto, de forma que soluções sistemáticas não estão disponíveis.

2.7.2. Sensibilidade em relação a variações nos dados

Os métodos AED utilizam técnicas determinísticas para calcular as eficiências em relação a fronteiras obtidas a partir dos dados disponíveis. Em alguns casos, pequenas distorções existentes nos dados, ocasionadas por erro de amostragem ou por erro nos dados, podem gerar grandes diferenças nos resultados. Este problema, denominado de sensibilidade do algoritmo às variações nos dados, pode invalidar os resultados.

A figura a seguir ilustra esta situação com um exemplo do método AED-C usando 1 insumo e 1 produto. Como pode-se observar, a pequena diferença em uma observação provoca grande deslocamento da fronteira estimada. Como a eficiência é obtida pela relação entre o consumo da UD da fronteira, e o consumo da UD analisada, o valor resultante ficará bastante distorcido.

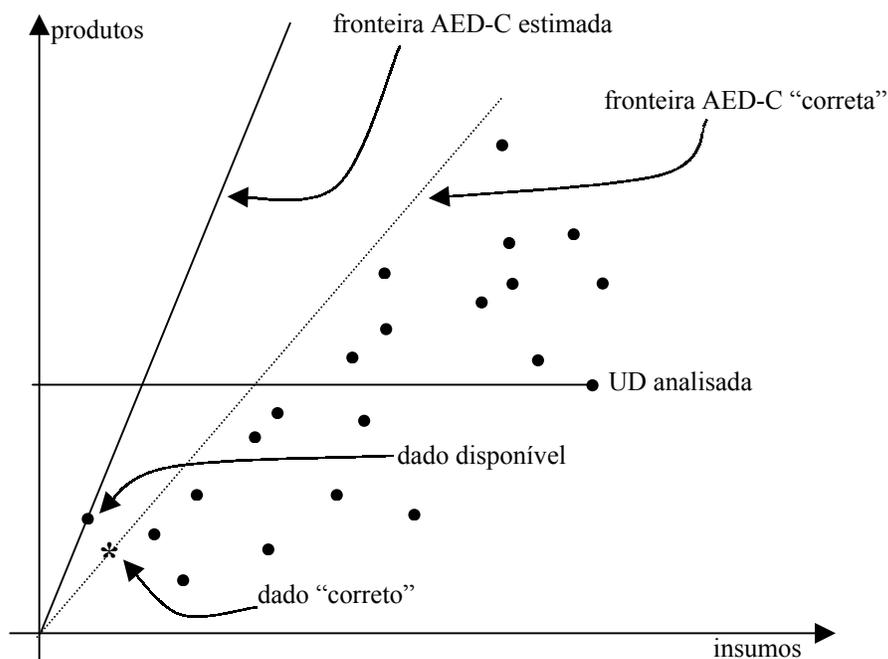


Figura 10 : Sensibilidade em relação a variações nos dados

2.7.3. Erro de amostragem

Um dos problemas que podem invalidar a suposição de que a amostra seja representativa do conjunto de produção é o do erro na amostragem das UD's utilizadas nos cálculos AED. Quando as amostras possuem grandes quantidades de dados, este problema tende a ser minimizado. Atualmente, apesar de muitas áreas de estudo disporem de grandes massas de dados, ainda existem casos em que o estudioso precisa analisar pequenas amostras, principalmente devido à pouca homogeneidade nas unidades disponíveis, devido a características sigilosas dos dados ou mesmo devido à necessidade de se usar tipos de insumos e produtos em números relativamente altos em relação às unidades de referência disponíveis.

Se a distribuição dos dados for conhecida, existem métodos para corrigir as distorções geradas por pequenas amostras e para estimar intervalos de confiança. Duas abordagens tratam destes problemas: (1) análise assintótica e (2) técnicas de *bootstrap*. Os resultados atualmente existentes a partir da primeira abordagem estão restritos aos casos envolvendo um insumo e um produto (p.ex.: Gijbels *et al.* (1999)). Outro problema advindo dessas técnicas se origina da necessidade de se estimar parâmetros de distribuição desconhecidos, perdendo, com isto, a vantagem dos métodos AED de serem não paramétricos. Alternativamente, Efron (1979) e Efron e Gong (1983) introduziram a técnica *bootstrap* para melhorar as amostras disponíveis. Esta técnica envolve a simulação repetida do processo de geração dos dados e a aplicação do estimador original a cada amostra simulada de forma a gerar estimadores cujas distribuições se aproximem das distribuições do estimador original. Esta técnica tem sido muito utilizada para analisar a sensibilidade dos estimadores empíricos às variações amostrais em casos em que a distribuição é difícil ou impossível de se obter analiticamente.

A figura a seguir ilustra uma situação em que uma amostra inadequada resulta em uma estimativa de fronteira bastante diferente da fronteira “ótima”, resultando em estimativas de eficiência distorcidas.

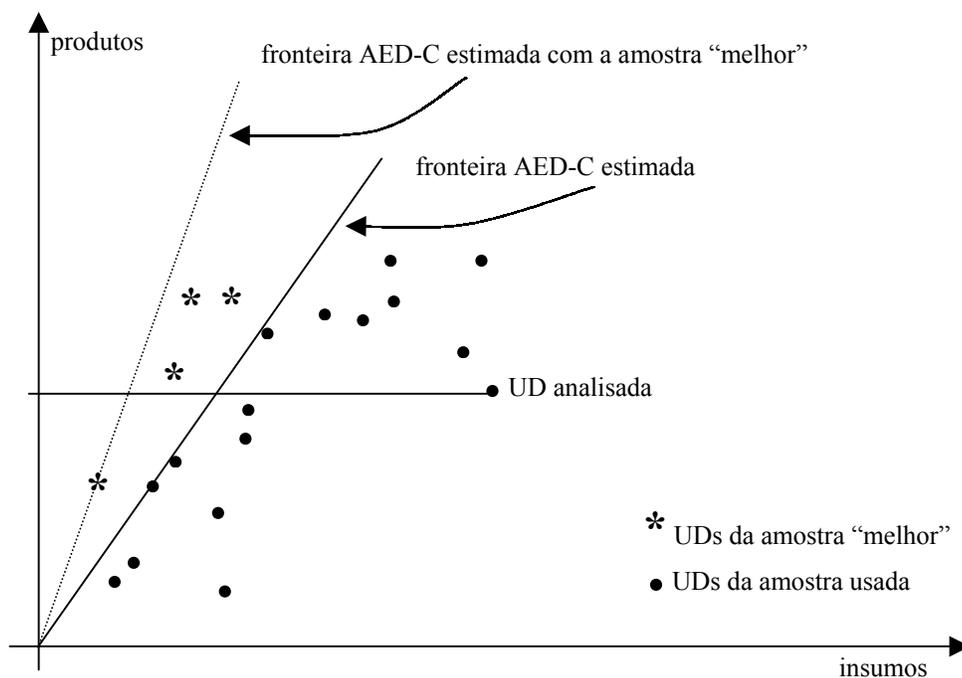


Figura 11 : Erro de amostragem

2.7.4. Eficiência não referendada

A eficiência não referendada (por *default*) ocorre quando uma UD não é dominada por nenhuma outra UD (no método FDH) ou por nenhuma combinação de UDs (nos demais métodos). Esta situação é tanto mais comum quanto maior for o número de insumos e produtos analisados ou quanto menor for a amostra. Neste caso, o método pode por na fronteira uma UD bastante ineficiente no uso de diversos insumos importantes, mas eficiente em apenas um insumo. O método FDH, dentre os métodos não paramétrico, é o que mais está sujeito a este tipo de problema. Em casos extremos, podemos ter amostras em que o método estima todas as eficiências iguais a 1, perdendo, assim, toda a capacidade de discriminação. A eficiência não referendada pode ser considerada como uma consequência do erro de amostra. Dependendo do número de insumos e produtos a solução deste problema pode ser bastante difícil em função do grande volume de dados necessários para se obter uma dada precisão.

2.7.5. Erros nos dados e existência de penetras

Um outro problema, que pode invalidar a suposição de que a amostra seja representativa do conjunto de produção, surge com a existência de erros nos dados ou de UD's penetras. Nem todo tipo de erro nos dados apresenta problema grave para os métodos AED. Além do mais, UD's sem erros nos dados também podem representar problema, desde que suas características de produção não sejam comparáveis com as demais UD's, causando distorção no cálculo de suas eficiências. Como já foi possível observar, o problema ocorre quando a UD, com erro ou não comparável, altera a fronteira estimada para o conjunto de produção. Pelo fato de ficarem fora da fronteira do conjunto de produção real, podemos designar estas UD's problemáticas como UD's penetras (são indesejáveis e entram na festa sem fazer parte do grupo de convidados). Em inglês elas são designadas por UD's *outliers* (em contrapartida das UD's *inliers*). Nos demais casos, os erros afetam apenas o cálculo da eficiência da UD com erro. Esta situação pode ocorrer por medições erradas, por erro de transcrição, por critérios diferentes de medição (parcelas que são consideradas por algumas UD's e são abatidas em outras, etc.), ou mesmo pela existência de UD's que fazem uso de tecnologias não comparáveis.

Como já foi apresentado anteriormente, os métodos AED podem ser muito sensíveis a variações nos dados que possam alterar a fronteira estimada do conjunto de produção. As conseqüências são análogas às apresentadas no caso de erro de amostragem. Para entender a analogia, basta designarmos as UD's da amostra aumentada (representadas com asterísticos na Figura 11, da página 17 ilustrativa do erro de amostragem) de UD's com erros. A figura a seguir, idêntica à do exemplo de erro de amostragem, ilustra esta situação. As UD's, ilustradas com asterísticos, se encontram em posições diferentes das suas posições reais, devido aos erros em seus dados. Como conseqüência, a fronteira estimada fica deslocada em relação à fronteira ideal, distorcendo o cálculo da eficiência das demais UD's.

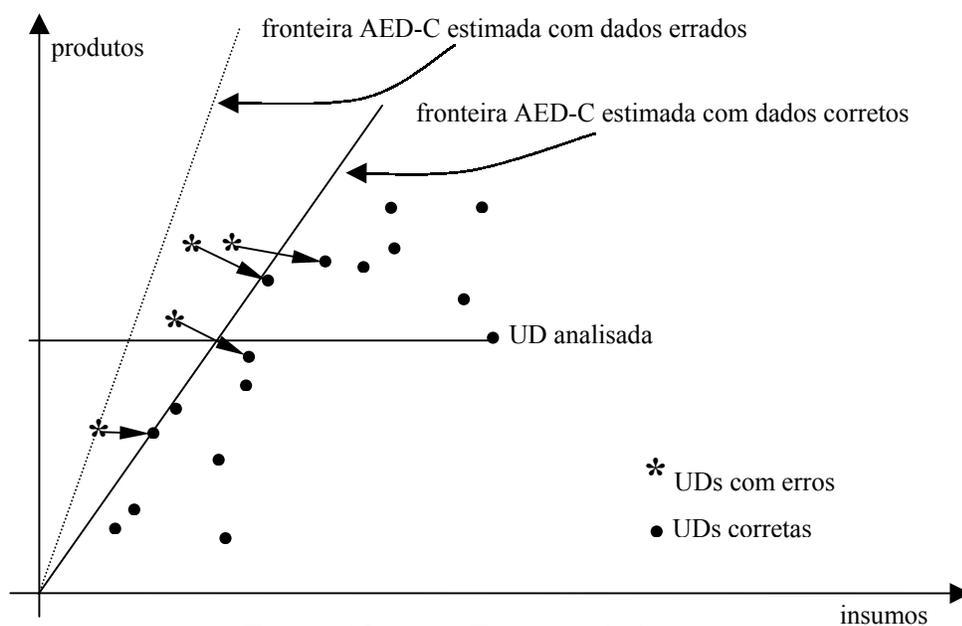


Figura 12 : Erro nos dados

A diferença em relação ao erro de amostragem está na forma de resolver cada problema. No caso anteriormente analisado, o problema pode ser resolvido aumentando-se a quantidade de elementos da amostra e/ou melhoria nos critérios de seleção, para torná-la mais representativa. Esta solução, ou mesmo os métodos anteriormente descritos de análise assintótica ou de *bootstrap*, não resolvem o problema de erro nos dados, podendo até mesmo potencializá-lo. Na realidade, enquanto não se eliminar os erros que afetem a estimativa da fronteira do conjunto de produção, os métodos AED gerarão resultados inconsistentes.

Várias abordagens diferentes foram propostas para tratar este problema na metodologia AED. detecção de penetras, análise de sensibilidade, *chance constrained programming*, e regressão não paramétrica.

Neste trabalho, os principais métodos utilizados na detecção de penetras serão analisados, suas principais características serão apresentadas, com destaque para suas principais limitações. Em seguida, apresentaremos o método DPF (Detecção Acelerada de Penetras por Inspeção de Fronteiras não Paramétricas) desenvolvido e proposto no presente trabalho, que objetiva minimizar as limitações dos demais métodos, principalmente ao se lidar com amostras contendo grandes quantidades de UD.

3. DETECÇÃO DE PENETRAS

3.1. Método de inspeção de observações influentes

Em um dos primeiros trabalhos sobre detecção de penetras, Wilson (1995) propõe analisar a influência causado pela retirada de uma UD da amostra antes de se calcular a eficiência das UDs remanescentes. Se esta influência for grande, podemos supor que poucas outras UDs respaldam a relação produto/insumo da UDs retirada. Conseqüentemente, esta UD é uma penetra em potencial e recebe uma alta prioridade para uma análise posterior. As principais limitações deste método são a sua dependência na habilidade do analista e a inviabilidade de tratar grandes volumes de dados, devido à sua característica artesanal.

3.2. Método da super eficiência

Em todos os métodos AED, e na própria definição de eficiência, a UD avaliada deve obrigatoriamente pertencer ao conjunto de produção. No método de cálculo da super eficiência, as UDs eficientes, ou seja, com eficiências iguais à unidade, são comparadas com as demais UDs excluindo-se a UD comparada. Desta forma, o valor obtido pode ser superior à unidade, uma vez que a UD comparada pode não estar no conjunto de produção gerado pelas demais UDs. Banker e Gifford (1988) e Banker, Das e Datar (1989) usaram este método para a identificação e eliminação de erros nos dados.

Um dos problemas desse método é que as restrições existentes no cálculo da AED-V e AED-N, relativas aos pesos das combinações lineares, faz com que o problema de minimização seja eventualmente inviável. Esta situação pode ocorrer quando a UD que está sendo comparada produza um dado produto em quantidade superior a todas as demais UDs da amostra. Neste caso, esta UD produzirá mais que qualquer combinação linear convexa das demais UDs. Pelas restrições dos métodos AED-V e AED-N, a tecnologia não garante a viabilidade de produções múltiplas das

existentes na amostra. Como conseqüência, o conjunto objetivo a ser minimizado é vazio, ficando sem sentido a definição de seu mínimo. Além do mais, para a identificação de penetras, o algoritmo precisa levar em conta a possibilidade de que essas UD's possam ser mascarados pela presença de outras UD's penetras. O valor da super eficiência de uma UD penetra pode resultar em um valor muito próximo de 1, descaracterizando-a como potencial penetra, caso existam outras UD's penetras que estejam em uma fronteira muito próxima da UD analisada. Esse mascaramento pode ocorrer em diversas camadas.

A figura a seguir apresenta um exemplo simples, com 1 insumo e 1 produto e usando o caso AED-C. A UD penetra 1, que está na fronteira, teria uma super-eficiência pouco maior do que a unidade, em virtude da presença da UD penetra 2. Neste caso, o método da super-eficiência, conforme descrito na literatura, não iria classificá-la como penetra. No entanto, como podemos observar, as duas UD's são penetras e estão prejudicando o cálculo correto das eficiências das demais UD's.

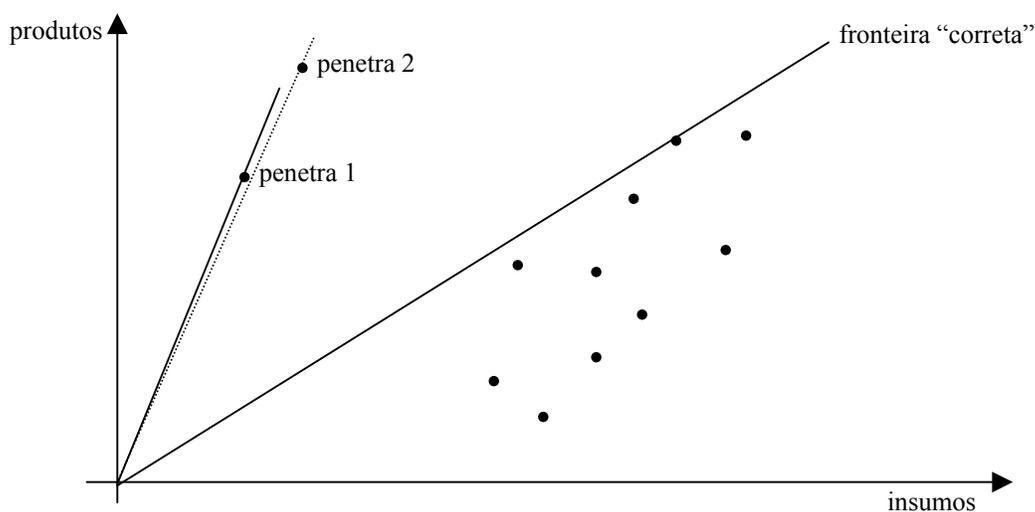


Figura 13 : Exemplo de mascaramento de penetras

Ao avaliar a adequação do método da super eficiência para a identificação de penetras, Banker e Chang (2000) não analisam a possibilidade de ocorrer mascaramento de penetras, usam amostras geradas artificialmente para testar os resultados, em geral com poucas UD's, e não concluem sobre critérios gerais para estimar a quantidade de UD's penetras a serem removidas da amostra.

3.3. Método da hiper eficiência

O método da hiper eficiência segue raciocínio análogo ao do método da super eficiência. As UD's eficientes, ou seja, com eficiências iguais à unidade, são comparadas com as demais UD's excluindo-se a UD analisada. A diferença é que, neste método, todas as demais UD's eficientes da amostra são também excluídas da comparação. Dessa forma, o valor obtido certamente será superior à unidade, uma vez que a UD não estará no conjunto de produção gerado pelas demais UD's.

Valem para este método as observações sobre o método da super eficiência, no que se refere a problemas inviáveis, ao mascaramento de penetras e à não estimação da quantidade de UD's penetras a serem removidas.

3.4. Método MRE (*Robust Efficiency Measurement - REM*)

Este método, apresentado por Kuosmanen e Post (1999), usa os conceitos do método AED-V (ou BCC) para desenvolver um algoritmo robusto de estimativa de eficiência na presença de penetras. Neste artigo, os autores não enfocam o problema da identificação desses penetras. O método parte da premissa de que existe uma dada quantidade q de UD's penetras na amostra sendo analisada e que essas UD's são as que geram as maiores distorções nos cálculos da eficiência das demais UD's. Como não sabemos quais são essas UD's, podemos calcular a eficiência de uma dada UD retirando todas as possíveis combinações de q UD's. Como uma das combinações será composta das UD's penetras, podemos supor que o maior dos valores obtidos será o valor mais provável da eficiência da UD sendo analisada. Para garantir a existência de solução, a UD sendo analisada não é retirada da amostra.

Para definirmos a eficiência MRE, introduziremos algumas notações. Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^k, y^k)\}$ um conjunto com os dados de produção de uma amostra de UD's com k elementos. Seja $K_A = \{1, 2, \dots, k\}$ o conjunto de índices das UD's da amostra. Seja (x^j, y^j) a UD a ser avaliada e $K_{A|j} = K_A - \{j\}$ o conjunto dos índices das demais UD's da amostra. Seja $R \subset K_{A|j}$ os índices de UD's a serem retiradas e $\text{card}(R)$ a quantidade de elementos de R . Sejam X^R e Y^R as matrizes cujas colunas são os

insumos e produtos das observações contidas na amostra sem as UD's cujos índices estão no conjunto R . Para uma dada quantidade r de UD's a serem retiradas, definimos:

$$\theta^{\text{MRE},r}(\mathbf{x}^j, \mathbf{y}^j) = \max_R \min_z \{ \theta | \theta \mathbf{x}^j \geq X^R \mathbf{z}, \mathbf{y}^j \leq Y^R \mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, R \subset K_A | j, \text{card}(R) \leq r, \\ \theta \in \mathfrak{R}_+, \mathbf{x}^j \in \mathfrak{R}_+^N, \mathbf{y}^j \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^k \}.$$

O artigo demonstra que, se retirarmos uma quantidade r de UD's superior à quantidade q de penetras, na hipótese de que as observações sejam consideradas como variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade positiva sobre todo o conjunto de produção, o estimador MRE é um estimador consistente para a real eficiência de uma UD pertencente ao conjunto de produção.

No trabalho analisado, é proposto um critério para estimar a quantidade de UD's a serem retiradas da amostra. O artigo parte da demonstração feita por Banker (1993) em que, se as ineficiências (ou seja, $(1 - \theta)$) forem consideradas como variáveis aleatórias independentes e identicamente distribuídas de uma distribuição semi-normal, a razão das somas dos quadrados das ineficiências sobre todas as UD's para dois modelos, sendo um subconjunto do outro, segue uma distribuição F , supondo os dois modelos consistentes. Denominando a eficiência da j -ésima UD calculada retirando-se p ou r UD's de θ_p^j ou θ_r^j , a expressão da razão entre as somas dos quadrados das ineficiências será :

$$S_p = \sum_{j=p+1}^k (1 - \theta_p^j)^2 \quad S_r = \sum_{j=r+1}^k (1 - \theta_r^j)^2 \quad F_{r,p} = \frac{S_r}{S_p} \sim f_{k-q, k-q, 1-\alpha}$$

Como q não é conhecido, supondo $r \geq q$, $f_{k-r, k-r, 1-\alpha}$, pode ser usado como um limite superior para $f_{k-q, k-q, 1-\alpha}$. O valor a ser usado para r será o menor dentre os que $F_{r,p}$ seja inferior ao limite para um dado grau de confiança. No entanto, permanece em debate a correção das suposições utilizadas para a demonstração feita no artigo. O presente trabalho, apesar de não entrar no aspecto formal de demonstrar a validade dessas suposições, apresenta quatro casos no capítulo 5 em que os dados parecem respaldá-las.

4. DETECÇÃO ACELERADA DE PENETRAS POR INSPEÇÃO DE FRONTEIRAS NÃO PARAMÉTRICAS - DPF

4.1. Introdução

Nenhum dos métodos AED, isoladamente, possibilita a identificação de penetras. No entanto, a maioria dos métodos determinísticos, desenvolvidos com este objetivo, utiliza um desses métodos, ou pequenas variações, como parte de seus algoritmos. O método AED-C é o mais afetado pela existência de penetras. Este aparente problema pode servir para melhor identificar as UDs contendo erros. Por este motivo, escolhemos este método como base para o cálculo das eficiências feito pelo método DPF apresentado neste trabalho. No entanto, os resultados do método DPF independem de qual método AED seja escolhido. A estimativa das influências das UDs no cálculo das eficiências foi baseada, em parte, no método MRE apresentado na seção 3.4.

Como vimos na seção 3.4, o método MRE objetiva estimar as eficiências de UDs em amostras contendo penetras, sem explicitar como identificar esses penetras. Além do mais, o método MRE, aplicado diretamente como apresentado em seu artigo, pode ser impraticável para amostras contendo grandes quantidades de UDs, em função do altíssimo volume de cálculo. O presente trabalho procurou superar estas limitações, de forma a criar um procedimento computacionalmente viável para identificar penetras em amostras contendo grandes quantidades de UDs e calcular as eficiências das demais UDs, caracterizando-se como método robusto de estimativa de eficiência. Como teste de sua viabilidade, amostras contendo 377 UDs e 4.796 UDs foram analisadas e os resultados são apresentados no próximo capítulo.

O método MRE depende do cálculo de todas as UDs da amostra em cada combinação de UDs retiradas. Os casos apresentados no artigo de Kuosmanen e Post (1999) continham 100 UDs. No entanto, em casos com quantidades maiores, o tempo de processamento pode subir exponencialmente.

4.2. Simplificações aceleradoras

A primeira simplificação é baseada no Teorema 22, demonstrado no apêndice à página 97, que nos permite executar os cálculos de eficiência usando como base apenas as UD's da fronteira. No método DPF, na fase 0, calculamos as eficiências das UD's e determinamos as que estão na fronteira. Na fase 1, retiramos uma das UD's da fronteira determinada na fase 0 e calculamos as eficiências das demais UD's. O procedimento da fase 1 é repetido para todas as UD's da fronteira determinada na fase 0. Na fase 2, o procedimento é executado para cada combinação da UD retirada no cálculo da fase 1 com uma das UD's da fronteira resultante do cálculo efetuado sem essa UD. Nas fases subsequentes, repetimos o cálculo das eficiências acrescentando, a cada conjunto de UD's retiradas na fase anterior, uma das UD's da fronteira resultante da retirada desse conjunto. Esse procedimento é repetido até a quantidade R desejada.

A redução nos cálculos necessários à aplicação do método MRE, usando esta simplificação, poderia ser suficiente para muitos casos. No entanto, quando as amostras são muito grandes e envolvendo diversos insumos, as fronteiras podem conter um número demasiado grande de UD's. Nestes casos, a redução do escopo dos cálculos às UD's da fronteira, pode não ser suficiente para que o tempo de processamento caia a níveis viáveis.

A simplificação seguinte utiliza o mesmo Teorema 22, demonstrado no apêndice à página 97, para diminuir a quantidade de UD's utilizadas na fase da identificação das UD's penetras. Nesta fase, basta usarmos uma subamostra contendo as primeiras camadas de UD's da amostra original. Definimos a camada 0 de uma amostra como o conjunto de UD's da fronteira original. A camada 1 é o conjunto de UD's da fronteira resultante ao usarmos a amostra original sem as UD's da camada 0. A camada n é o conjunto de UD's da fronteira resultante ao usarmos a amostra original sem as UD's das $n-1$ camadas anteriores. Usando esse procedimento, a amostra de 4.796 UD's pode ser reduzida a pouco mais de 500 UD's. Para o sucesso deste procedimento, precisamos selecionar uma quantidade de camadas suficientemente grande para garantir que nenhuma das UD's identificadas como penetra esteja na última das camadas selecionadas. Este cuidado é importante para que não fique de fora

nenhuma UD penetra mascarada pelas camadas superiores de UD's penetras e para garantir a existência de camadas inferiores que sirvam de comparação para a correta estimativa da influência das UD's candidatas.

4.3. Determinação e ordenação das influências

Usando a subamostra anteriormente descrita, o método DPF identifica as UD's penetras a cada grupo de fases calculadas. A identificação é feita da seguinte forma : para cada UD, guardamos o valor da eficiência calculada em cada fase, ou seja, o valor máximo da eficiência da UD retirando r UD's; na fase seguinte, verificamos qual combinação de $(r + 1)$ UD's retiradas resultou no maior acréscimo da eficiência em relação à fase anterior (na realidade guardamos a razão $\text{eficiência}_{\text{fase}_r} / \text{eficiência}_{\text{fase}_{r+1}}$, que é sempre um número no máximo igual a 1); as combinações de UD's são ordenadas segundo os maiores acréscimos nas eficiências (ou as menores razões); as UD's são retiradas da amostra segundo a ordem decrescente de influência; as somas dos quadrados das ineficiências são calculadas a cada retirada de UD; o procedimento é repetido para um número p suficientemente grande de UD's, supostamente superior ao de penetras; o valor de r é estimado usando o critério apresentado pelo método MRE, ou seja, o menor r tal que :

$$\frac{\sum_{j=r+1}^k (1-\theta_r^j)^2}{\sum_{j=p+1}^k (1-\theta_p^j)^2} < f_{k-r,k-r,1-\alpha}$$

k = número de UD's na amostra

$p \gg q$ (onde q é o número de penetras na amostra)

r = número estimado de penetras pelo método

4.4. Descrição do algoritmo

A seguir, resumimos o procedimento definido :

- Obtenção da subamostra das camadas mais externas da amostra :

- Partindo da amostra original, calculamos as eficiências e identificamos as UDs eficientes (com eficiência igual a 1), que denominaremos de camada 0;
- Retirando as UDs da camada 0, repetimos o procedimento anterior e identificamos as novas UDs eficientes, que denominaremos de camada 1;
- O procedimento é repetido para uma quantidade de camadas que não inviabilize os cálculos, mas permita identificar todas as UDs pentras;
- Como regra simplificada, podemos avançar até que totalizemos uma quantidade em torno de 500 UDs selecionadas, para que a duração do processamento não fique muito longo;
- Este conjunto precisará ser aumentado, caso a quantidade estimada de pentras resulte na identificação de uma UD pentra que esteja na última camada selecionada. Este cuidado é importante, pois, neste caso, podem restar outras UDs pentras na amostra, além da própria identificação da referida UD não poder ser respaldada por outras UDs de camadas inferiores;
- Cálculo das influências das UDs
 - Usando a subamostra selecionada, utilizamos o método MRE para calcular as eficiências das UDs;
 - Inicialmente, calcula-se a eficiência de uma UD j pelo método AED, que denominaremos $REM(0, j)$;
 - Para cada UD j da amostra, são calculadas as eficiências retirando uma UD da fronteira (camada 0) de cada vez. A maior das eficiências da UD j assim calculadas será denominada de $REM(1, j)$. Guardamos então a eficiência $REM(1, j)$, a UD da fronteira cuja retirada resultou no cálculo da $REM(1, j)$ e a razão $REM(0, j) / REM(1, j)$;

- A cada retirada de uma UD da fronteira, calcula-se a nova fronteira resultante e guarda-se todas as combinações da UD retirada com as UDs da nova fronteira;
- O conjunto formado por todas as combinações, montadas na fase em que são retiradas uma UD de cada vez, será utilizado na etapa seguinte, onde são retiradas duas UDs de cada vez;
- O procedimento de cálculo da eficiência da UD j é repetido, retirando cada par de UDs do conjunto formado na etapa anterior. O maior valor resultante da eficiência da UD j será denominado de $REM(2, j)$. Novamente, guardamos a eficiência $REM(2, j)$, as UDs da combinação cuja retirada resultou no cálculo da $REM(2, j)$ e a razão $REM(1, j) / REM(2, j)$
- A cada retirada de uma combinação de UDs, calcula-se a nova fronteira resultante e guarda-se todas as combinações formadas pelas UDs da combinação original com as UDs da nova fronteira;
- O conjunto formado por todas as combinações, montadas na fase em que são retiradas duas UDs de cada vez, será utilizado na etapa seguinte, onde são retiradas três UDs de cada vez;
- O procedimento é repetido até que se tenha retirado as combinações de p UDs de cada vez. O valor de p não é conhecido *a priori* e deve ser estimado, de forma a ser bem maior que a quantidade r de UDs estimadas pelo teste F (digamos, pelo menos 30% acima de p);
- Caso a quantidade p resulte muito próxima de r , deve-se repetir o procedimento para um valor maior;
- Ordenação das UDs mais influentes
 - Ao longo do processamento foram guardadas todas as combinações de UDs cujas retiradas resultaram nos cálculos das eficiências $REM(i, j)$, juntamente com as razões anteriormente descritas ($REM(i-1, j) / REM(i, j)$);

- Estes dados são agrupados segundo as combinações de UD's retiradas e as menores razões de cada combinação são determinadas;
- As combinações de UD's são então ordenadas, de forma crescente, segundo as razões determinadas conforme anteriormente descrito;
- As UD's de cada combinação recebem o mesmo posto. Como uma UD pode aparecer em mais de uma combinação, seu posto será aquele da combinação de menor razão (menor posto);
- Retirada das UD's mais influentes :
 - A partir do conjunto original, contendo todas as UD's, retiramos a primeira UD da ordenação feita na fase anterior (menor posto);
 - Calculamos as eficiências das UD's restantes e somamos os quadrados das ineficiências : $S_1 = \sum_{j=2}^k (1 - \theta_1^j)^2$;
 - O procedimento é repetido retirando uma nova UD a cada vez até a UD de ordem p . Como o valor de p não é conhecido, usa-se um valor suficientemente grande, de forma que o valor de r , estimado no passo seguinte, obedeça a condição $r \ll p$. Caso isso não venha a ocorrer, repete-se o procedimento com um valor de p maior;
- Estimativa da quantidade de penetras :
 - A quantidade de penetras estimada será o menor valor r , tal que :

$$\frac{\sum_{j=r+1}^k (1 - \theta_r^j)^2}{\sum_{j=p+1}^k (1 - \theta_p^j)^2} < f_{k-r, k-r, 1-\alpha}$$
- Resultado final :
 - As eficiências estimadas serão as obtidas retirando as r UD's mais influentes.

Caso, durante o procedimento de cálculo das influências das UDs, verifique-se que a quantidade de combinações a serem processadas na fase seguinte seja muito grande para viabilizar seu cálculo, pode-se adotar o seguinte procedimento :

- Para-se o procedimento de cálculo das influências e passa-se para a etapa de ordenação e retirada das UDs mais influentes;
- Neste caso, após a retirada dessas UDs mais influentes, pode-se repetir o processo, de forma a fazer uma nova classificação de UDs influentes;
- As novas combinações de UDs assim obtidas devem ser ordenadas em conjunto com as combinações anteriores, de forma a se criar uma ordenação única;
- Feita esta ordenação única, pode-se passar para as etapas seguintes de retirada das UDs, estimativa de quantidade de penetras e cálculo final das eficiências.

O risco desta simplificação é a existência de mascaramento múltiplo, onde temos um conjunto de UDs penetras que se situam muito próximas de um trecho da fronteira, de forma que a retirada de qualquer subconjunto que não contenha todas estas penetras resulte em pouca variação na fronteira, de forma a gerar pouca variação nas eficiências das demais UDs. Caso esta simplificação tenha que ser usada, a distribuição das eficiências na amostra final deverá ser analisada, para verificar se o resultado é consistente.

4.5. Tempos de execução do método em testes realizados

Como exemplo, relacionamos a seguir os tempos necessários para processar as diversas etapas relativas à maior e mais complexa das amostras analisadas, contendo 4.796 municípios brasileiros com 9 produtos e 4 insumos. Neste processamento, foi utilizado um microcomputador Pentium III de 650 Mhz. Por comodidade e por facilidade, usamos o software Access 97 para tratarmos os dados e a linguagem interpretada visual basic, nativa a este software. Se usarmos linguagens compiladas e mais eficientes, como C++, certamente podemos obter tempos bem inferiores.

Os tempos de processamento da etapa em que são determinadas as diversas camadas a serem utilizadas nos demais cálculos são relacionados na tabela apresentada a seguir :

Tabela 1: Quantidades e durações na etapa de definição das camadas

Camada	Quantidades na fronteira	Duração
0	41	00:06:00
1	87	00:21:03
2	136	00:16:29
3	176	00:20:48
4	228	00:26:41
Totais	668	01:31:01

O programa para calcular as eficiências, usando o método AED-C, foi desenvolvido de forma a otimizar o processamento usando apenas as UD's da fronteira. Como consequência, as durações tendem a ser maiores quanto mais UD's estiverem na fronteira. Em contrapartida, a quantidade de UD's sendo calculadas diminui a cada nova camada, pois retiramos as UD's das fronteiras anteriormente calculadas. A combinação destes efeitos nem sempre resultou em tempos crescentes no processamento destas camadas.

Selecionadas as 668 UD's contidas nas 5 primeiras camadas, passamos à etapa em que usamos o algoritmo MRE para calcularmos as influências de todos os possíveis subconjuntos de UD's. Nesta etapa, calculamos as eficiências das UD's, retirando uma UD de cada vez, duas UD's de cada vez, e assim por diante. À medida que aumentamos o número de UD's a serem retiradas de cada vez, a quantidade de combinações possíveis aumenta muito, tornando o tempo de processamento impraticável. Devido à inviabilidade do processamento para combinações com mais de duas UD's, realizamos três etapas de ordenação e retirada de UD's. O quadro a seguir

apresenta as durações de cada fase, com as quantidades de combinações e de UDs cujas eficiências foram calculadas.

Tabela 2 : Quantidades e durações do cálculo MRE (etapa 1)

Qt de municípios retirados por vez	Qt de combinações calculadas	Qt de escores de eficiência calculados	Duração
0	1	668	00:01:33
1	41	3.366	00:27:54
2	852	97.608	10:53:40

Como podemos observar, foram 852 combinações de 2 UDs, resultando no cálculo da eficiência de 97.608 UDs. Foram quase 11 horas de cálculo. Como a seqüência do método MRE seria impraticável, retiramos as 9 UDs mais influentes, cujas razões ficaram abaixo de 0,8, e repetimos o processo para as UDs restantes. Esta nova etapa foi necessária para garantirmos que não ficaríamos mascarados. Os tempos e as quantidades resultantes foram os seguintes:

Tabela 3 : Quantidades e durações do cálculo MRE (etapa 2)

Qt de municípios retirados por vez	Qt de combinações calculadas	Qt de escores de eficiência calculados	Duração
0	1	659	00:03:49
1	70	2.694	01:28:05

Seguindo o raciocínio já apresentado, foram retiradas novas 11 UDs e feito novo processamento MRE. Os tempos e as quantidades resultantes foram os seguintes:

Tabela 4 : Quantidades e durações do cálculo MRE (etapa 3)

Qt de municípios retirados por vez	Qt de combinações calculadas	Qt de escores de eficiência calculados	Duração
0	1	648	00:05:22
1	79	2.559	02:04:36

Após este cálculo, as UDs foram ordenadas em conjunto, segundo suas influências. Finalmente, retornamos ao conjunto inicial de 4.796 UDs e calculamos 31 conjuntos de eficiências, um para a amostra original e os demais 30 retirando, de forma cumulativa, uma das 30 UDs mais influentes de cada vez. As durações nesta etapa foram as seguintes :

Tabela 5 : Durações do cálculo final das eficiências

Qt de municípios retirados	Duração
0	0:05:56
1	0:07:49
2	0:07:53
3	0:07:50
4	0:07:22
5	0:07:14
6	0:07:25
7	0:07:23
8	0:08:31
9	0:09:05
10	0:11:18
11	0:13:19
12	0:13:16
13	0:13:16
14	0:12:15
15	0:12:11
16	0:12:01
17	0:11:33
18	0:11:23
19	0:12:58
20	0:12:31
21	0:15:11
22	0:15:01
23	0:16:05
24	0:15:55
25	0:16:06
26	0:15:23
27	0:16:37
28	0:15:55
29	0:15:55
30	0:16:25

Os resultados são apresentados no próximo capítulo. O processo todo levou 26 horas, incluindo o tempo necessário para anotar dados, digitar as UDs selecionadas e outros procedimentos manuais necessários. Apesar do grande volume de dados, o método se mostrou computacionalmente viável.

As quantidades selecionadas em cada etapa podem ser alteradas, sem grandes influências no resultado final. Para verificar a validade destas simplificações, refizemos os cálculos, usando apenas os 400 municípios das 4 camadas mais externas e processamos os dados até a retirada de todos os conjuntos de 3 municípios por vez. Com estes critérios, o processamento levou mais de 100 horas e foram calculadas mais de um milhão de eficiências. Comparando os 23 municípios mais influentes selecionados desta forma com os do procedimento anterior, houve apenas a inversão do 24° e 26° da ordenação anterior que ficaram entre os 23 municípios mais influentes em contrapartida pela não seleção do 22° e 23°. Os resultados finais em ambos os casos, para todos os efeitos práticos, foram idênticos.

5. DETECÇÃO DE PENETRAS E ESTIMAÇÃO DA EFICIÊNCIA TÉCNICA DAS MUNICIPALIDADES BRASILEIRAS

Em trabalhos anteriores, diferentes autores (Sampaio de Sousa e Ramos de Sousa (1999a, 1999b), Vanden Eeckaut, Tulkens e Jamar (1991)), utilizando as técnicas não paramétricas analisadas no capítulo 2, mensuraram o nível de eficiência técnica para os governos locais. Esses trabalhos preencheram uma lacuna importante no que diz respeito à avaliação do desempenho dessas cidades, pois permitiu elaborar um instrumental, solidamente apoiado na teoria econômica e suficientemente flexível, para ser operacionalizado e usado como forma de avaliação sistemática e permanente por seus governos.

Entretanto, é importante enfatizar a natureza exploratória desses trabalhos. Isto porque os níveis calculados de ineficiência podem resultar, não somente da incompetência administrativa e gerencial ou da inexistência de incentivos adequados ao bom funcionamento do município, mas podem, também, decorrer da fragilidade da base de dados usada ou da heterogeneidade das comunidades analisadas. De fato, em razão da natureza determinística dos modelos não paramétricos, os escores computados de eficiência dependem, crucialmente, da qualidade da informação utilizada. Assim, a existência de municipalidades super eficientes, erros de mensuração, variáveis omitidas e outros tipos de discrepâncias estatísticas podem comprometer seriamente a credibilidade dos indicadores obtidos. É preciso, pois, garantir a qualidade da base de dados sobre a qual o modelo se assenta, assegurar-se de que as variáveis relevantes sejam corretamente identificadas, antes que o diagnóstico de ineficiência seja atribuído a uma determinada municipalidade.

Neste contexto, a aplicação do método DPF (Detecção Acelerada de Penetras por Inspeção de Fronteiras não Paramétricas) permite identificar as inconsistências porventura existentes na base de dados, podendo, assim, contribuir, de maneira significativa, para a estimação de indicadores de eficiência mais robustos. Note-se, porém, que não é objetivo do presente trabalho analisar os resultados para tirar

conclusões sobre a realidade dos municípios. Os dados e os resultados das eficiências, obtidos antes e após a retirada dos penstras pela aplicação do método DPF, foram usados apenas para analisar e validar as premissas deste método.

5.1. Indicadores de insumos e produtos para as municipalidades brasileiras

A implementação da metodologia DPF utilizou a base de dados municipais desenvolvida por Sampaio de Sousa e Stosic (2003), que contém informações variadas sobre 4.796 municipalidades brasileiras. Esse conjunto de dados é particularmente apropriado para testar o método DPF, por ser uma amostra suficientemente grande e bastante heterogênea, favorecendo o surgimento de penstras (erros de mensuração, ou municipalidades atípicas).

Nesta base de dados utilizada, os custos agregados totais, computados como o valor da despesa corrente,¹ serviram como um dos indicadores de insumo. A taxa de mortalidade infantil, estudada como consequência das ações governamentais, deveria ser computada como produto. No entanto, como o objetivo é a sua minimização, seus valores foram considerados como insumos, para compatibilizar com os critérios dos modelos usados. A análise da disponibilidade dos serviços municipais foi focada no tripé : habitação, educação e saúde.

As informações sobre a despesa corrente foram obtidas da Secretaria do Tesouro Nacional (STN). Os indicadores de produto 1 a 4, da Tabela 6, foram obtidos no Censo de 2002. Dados sobre matrículas nas escolas municipais, no primeiro e segundo graus, referem-se ao ano de 1991 e foram obtidos junto à Coordenadoria Geral do SEEC/INEP, do Ministério da Educação e dos Desportos (MEC). Para mais detalhes sobre a base de dados, ver Sampaio de Sousa e Stosic (2003).

A descrição dos insumos e produtos utilizados, as fontes dos dados e as equivalências dessas variáveis com os serviços municipais encontram-se relacionados na Tabela 6.

¹As despesas com investimento foram eliminadas pelo fato de serem mais erráticas, o que dificultaria a comparação entre os municípios.

Tabela 6 : Indicadores de insumos e produtos e serviços municipais correspondentes

Indicadores de insumos e produtos	Fonte	Serviços Municipais para os quais os indicadores servem como proxies
a) Indicadores de insumos		
1. Despesa corrente (em R\$ 1.000.000)	STN (1)	Custo total agregado
2. Número de docentes	Censo Escolar/MEC	Mão de obra empregada
3. Taxa de mortalidade infantil (1998)	IBGE (2)	Serviços públicos de saúde
4. Hospitais e unidades de atendimento médico por 1000 habitantes	IBGE	Serviços públicos de saúde
b)- Indicadores de produtos		
1. População total em 2000 (em milhares)	IBGE	
2. Pessoas alfabetizadas	MEC (3)	Serviços educacionais
3. Matrículas por escola (total) (em milhares) (4)	Censo Escolar/MEC	Serviços educacionais
4. Frequência dos alunos por escola (em milhares)	Censo Escolar/MEC	Serviços educacionais
5. Alunos aprovados por escola (em milhares) (4)	Censo Escolar/MEC	Serviços educacionais
6. Alunos na série correta por escola (em milhares) (4)	Censo Escolar/MEC	Serviços educacionais
7. Domicílios com acesso à rede geral de água (em milhares)	IBGE	Condições habitacionais
8. Domicílios com acesso à rede geral de saneamento (em milhares)	IBGE	Condições habitacionais
9. Coleta de lixo por empresa pública ou terceirizada (em milhares)	IBGE	Condições habitacionais

Fonte : Sampaio de Sousa e Stosic (2003), pp. 8.

Obs. : 1: STN – Secretaria do Tesouro Nacional; 2: IBGE – Instituto Brasileiro de Estatística, Censo Demográfico de 2000; 3: MEC - Ministério da Educação; 4: Todos os dados se referem às escolas municipais de primeiro e segundo graus.

5.2. Análise preliminar dos dados dos municípios brasileiros: aplicando o método AED diretamente

Inicialmente, calculamos as eficiências dos municípios brasileiros utilizando o método AED-C (ou CCR), antes da eliminação dos penétris pelo método DPF. A base de dados já havia sido cuidadosamente inspecionada por procedimentos manuais e os erros e discrepâncias detectados já haviam sido corrigidos. Os cálculos foram feitos para as mesmas amostras que foram posteriormente utilizadas para testar o método DPF. Uma subamostra, contendo apenas 377 municípios da Bahia, e a amostra completa, contendo 4.796 municípios brasileiros, foram analisadas, inicialmente com um insumo (a despesa agregada) e, em seguida, com quatro insumos, conforme explicitado na Tabela 6.

5.2.1. Subamostra contendo 377 municípios da Bahia

UM INSUMO E NOVE PRODUTOS

Neste conjunto, apenas o município Sapeaçu apresentou eficiência igual a 1 e os demais apresentam eficiências inferiores a 0,07. Trata-se, claramente, de uma distribuição altamente assimétrica e que ilustra bem os limites da aplicação direta do método DEA nesse tipo de conjunto de dados. Dada a discrepância entre estes resultados, temos uma situação em que, no mínimo, Sapeaçu tem informações não comparáveis com os demais municípios. A tabela e o gráfico apresentados a seguir, mostram o resultados obtidos.

Tabela 7 : Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos

Eficiência	Municípios
<0,07	376
1,00	1
TOTAL	377

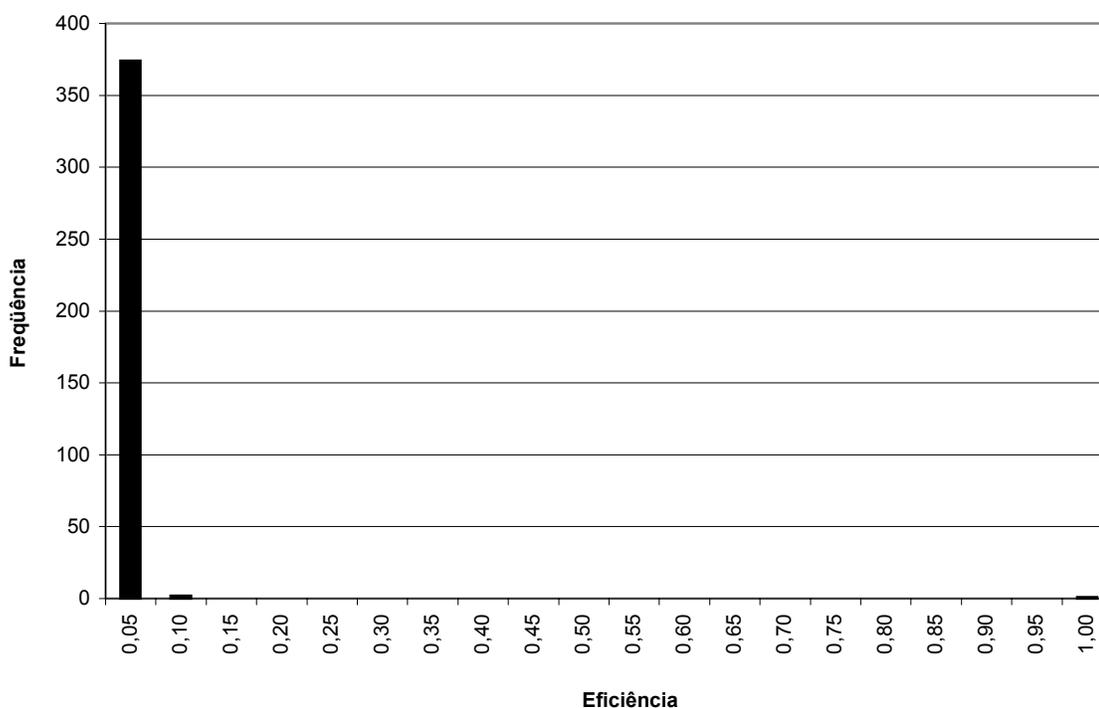


Gráfico 1 : Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos

De fato, os dados obtidos em relação a Sapeaçu devem estar errados, apresentando relações produto/insumo bastante improváveis de se observar na prática. A Tabela 8 apresenta as proporções entre as produções relativas aos nove produtos analisados e a despesa corrente de Sapeaçu, em comparação com os demais municípios da Bahia. Sapeaçu apresenta proporções de 15 a 55 vezes maiores do que as maiores proporções encontradas entre os demais municípios.

No entanto, a simples eliminação dessa municipalidade não resolve todos os problemas dessa amostra, pois logo surge o município de Salinas da Margarida, cuja inclusão distorce, sensivelmente, os índices de eficiência calculados. Eliminando essa última cidade, Serrinha agora se qualifica como penetra e assim por diante. Nesse caso, a simples inspeção visual para identificar os erros nos dados se mostrou insuficiente em virtude da quantidade de municípios sendo analisados.

Tabela 8 : Comparação da produtividade de Sapeaçú com as dos demais municípios da Bahia – amostra original - um insumo e nove produtos

Produto	Sapeaçú (A)	Máximos dos demais municípios (B)	Razão A/B
1. População total em 2000	378,204	7,901	48
2. Pessoas alfabetizadas	190,275	3,982	48
3. Matrículas por escola	2,414	0,122	20
4. Frequência dos alunos por escola	2,138	0,099	22
5. Alunos aprovados por escola	1,563	0,081	19
6. Alunos na série correta por escola	0,897	0,061	15
7. Domicílios com acesso à rede geral de água	32,946	1,266	26
8. Domicílios com acesso à rede geral de saneamento	72,284	1,313	55
9. Coleta de lixo por empresa pública ou terceirizada	36,073	1,156	31

A = produção / despesa agregada (relativo a Sapeaçú)

B = produção / despesa agregada (maior valor dentre os demais municípios)

A/B = razão entre os valores das colunas A e B

QUATRO INSUMOS E NOVE PRODUTOS

Em seqüência, os mesmos municípios foram analisados com os quatro insumos disponíveis na base de dados. A tabela e o gráfico, apresentados a seguir, mostram o resultados obtidos.

Tabela 9 : Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original – quatro insumos e nove produtos

Eficiência	Municípios
0,20-<0,40	5
0,40-<0,60	146
0,60-<0,80	150
0,80-<1,00	40
1,00	36
TOTAL	377

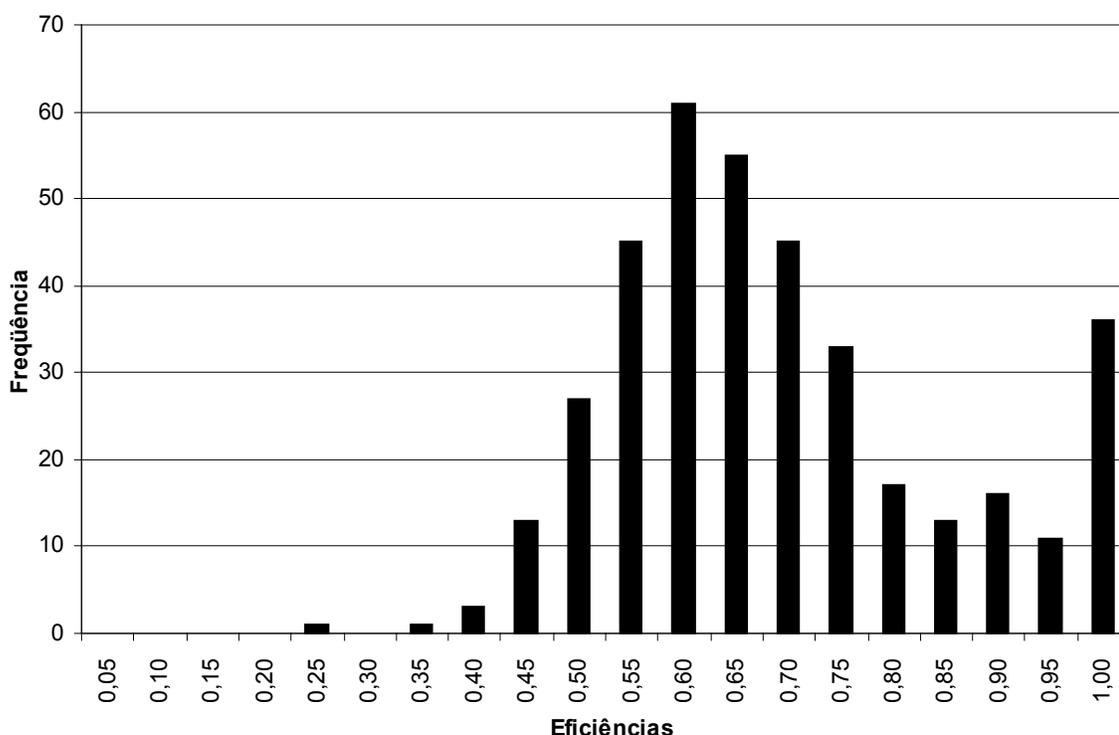


Gráfico 2 : Distribuição das eficiências das municipalidades baianas, computadas utilizando-se o método DEA-C (ou CCR) - amostra original – quatro insumos e nove produtos

Comparando com o Gráfico 1 do caso com um insumo, podemos notar que a inclusão dos outros três insumos mascarou os erros existentes na massa de dados, tornando a distribuição menos assimétrica. Como a eficiência é calculada levando em consideração todos os insumos e produtos, erros para menos nas quantidades de um único insumo nem sempre resultam em deslocamento da fronteira. Apesar de não ficar evidente, visualmente falando, existem erros nos dados pois, como a base usada é a mesma do exemplo anterior, apenas acrescentando-se 3 insumos, sabemos que existem erros, pelo menos no caso de Sapeaçu. Neste caso, fica mais evidente a necessidade de se utilizar um processo analítico e sistematizado para a identificação de penetras, pois a inspeção manual ou visual foi insuficiente para identificarmos esses erros.

5.2.2. Amostra contendo 4.796 municípios brasileiros

UM INSUMO E NOVE PRODUTOS

Neste conjunto, apenas quatro municípios apresentam eficiência igual a 1, seis acima de 0,10 e os demais 4.786 apresentam eficiências inferiores a 0,10. Analogamente ao exemplo da Bahia, trata-se de uma distribuição altamente assimétrica, caracterizando uma distorção gerada por problemas nos dados e que ilustra bem os limites da aplicação direta do método DEA neste conjunto de dados. A tabela e o gráfico, apresentados a seguir, mostram os resultados obtidos.

Tabela 10 : Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos

Eficiência	Municípios
<0,10	4.786
0,10-<1,00	6
1,00	4
TOTAL	4.796

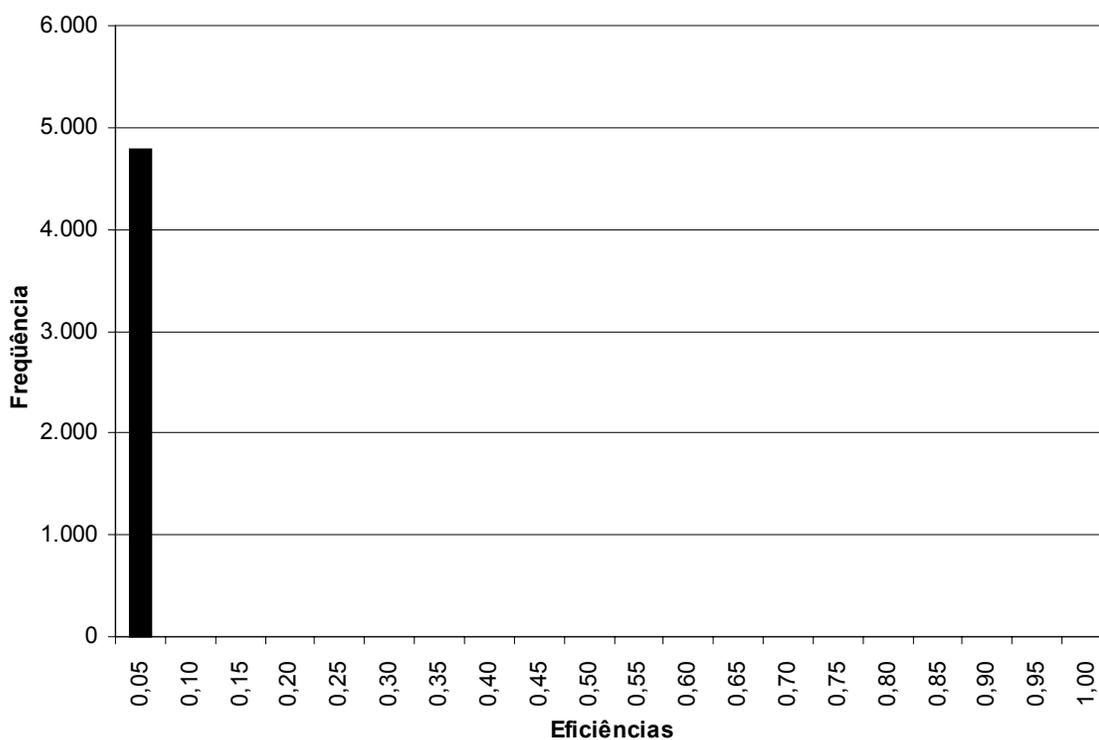


Gráfico 3 : Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - um insumo e nove produtos

De forma semelhante ao caso restrito aos municípios da Bahia, podemos inferir sobre a existência de erros nos dados, devido à grande concentração de eficiências nas faixas inferiores da distribuição. Apesar desta evidência, os erros não se apresentam de forma muito fácil e imediata. Isto pode ser visto na Tabela 11, que mostra os municípios mais eficientes na amostra não tratada, onde o município de Sapeaçu se apresenta mascarado por outros municípios discrepantes.

Tabela 11 : Municípios brasileiros com maiores eficiências – amostra original - um insumo e nove produtos

Município	Eficiência
Santo Antônio do Sudoeste	1
Pontal do Araguaia	1
Salto Veloso	1
Barra do Garças	1
São Francisco de Itabapoana	0,9658653
Trajano de Moraes	0,3237392
Sapeaçu	0,1189448

Em uma inspeção isolada dos dados acima, poderíamos concluir que Sapeaçu não contém erros em seus dados. Os resultados do método DPF, apresentados na seção 5.3, indicam que, na realidade, Sapeaçu se encontra apenas na oitava posição dentre os municípios mais influentes. No entanto, como corretamente o método estimou, este município se encontra entre os municípios escolhidos para serem retirados da massa de dados.

QUATRO INSUMOS E NOVE PRODUTOS

Em seqüência, os mesmos municípios foram analisados com os quatro insumos disponíveis na base de dados. A tabela e o gráfico apresentados a seguir, mostram o resultados obtidos.

Tabela 12 : Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - quatro insumos e nove produtos

Eficiência	Municípios
<0,20	1.285
0,20-<0,40	2.661
0,40-<0,60	546
0,60-<0,80	186
0,80-<1,00	77
1,00	41
TOTAL	4.796

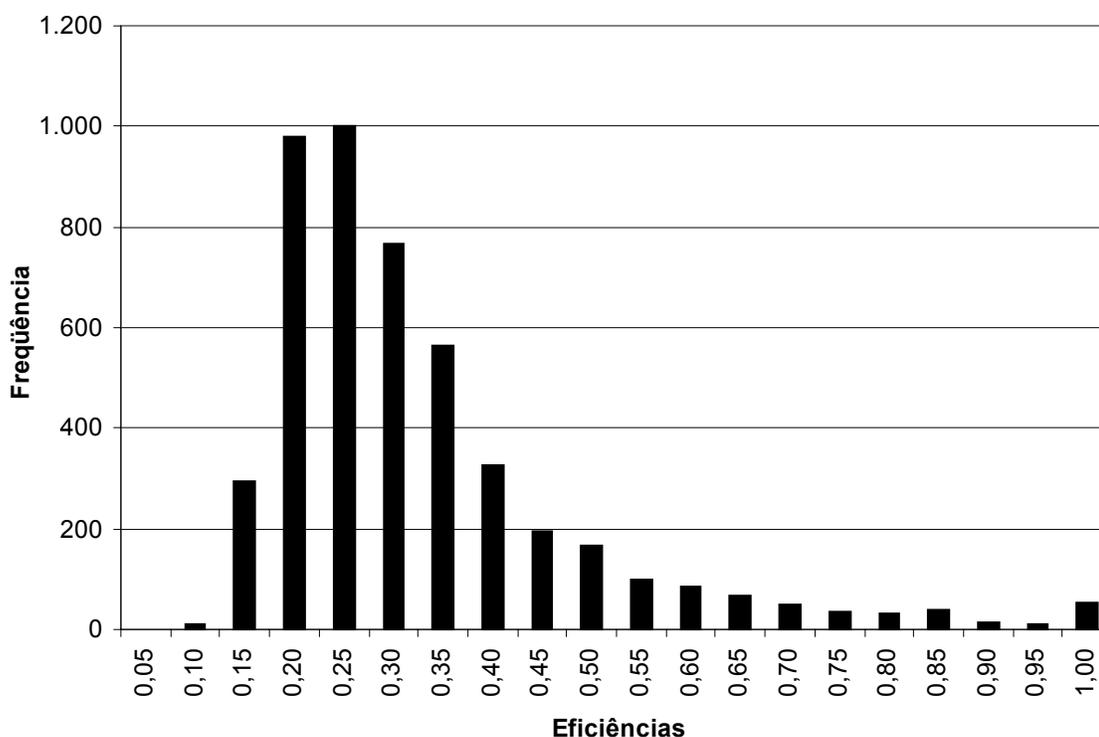


Gráfico 4 : Distribuição das eficiências das municipalidades brasileiras, computadas utilizando-se o método DEA-C (ou CCR) - amostra original - quatro insumos e nove produtos

Analogamente à subamostra da Bahia, a inclusão dos outros insumos mascarou os erros existentes na massa de dados, tornando a distribuição menos assimétrica do que a do caso com apenas um insumo, como se pode comparar com o Gráfico 3. Apesar de não ser visualmente tão evidente quanto o caso com um insumo, o formato da distribuição é indicativo da presença de erros nos dados. Devido à

quantidade de municípios, a utilização de método automatizado de inspeção de erros é essencial para o caso em questão.

Concluída a análise preliminar dos dados, com a constatação da presença de erros nos dados, passaremos a apresentar os resultados obtidos pelo método DPF.

5.3. Ilustração empírica do método DPF

Para facilitar o processo evolutivo do desenvolvimento do método DPF, os testes iniciais foram feitos em uma subamostra contendo apenas 377 municípios da Bahia. Após desenvolvidos e testados os algoritmos, foi usada a amostra total contendo 4.796 municípios brasileiros. Em ambas as amostras, os cálculos foram feitos para um (1) insumo (a despesa agregada) e para quatro insumos, conforme explicitado na Tabela 6. Ao passarmos da amostra contendo 377 municípios da Bahia para a amostra maior com 4.796 municípios brasileiros, tornou-se necessário o desenvolvimento da técnica de seleção da subamostra contendo os municípios das camadas mais externas como forma de viabilizar o processamento dos dados.

A seguir, apresentamos os resultados dos casos analisados, utilizando o método DPF.

5.3.1. Subamostra contendo 377 municípios da Bahia

UM INSUMO E NOVE PRODUTOS

Inicialmente serão apresentados os resultados dos cálculos da subamostra relativa a 377 municípios da Bahia usando apenas um insumo. Pelo reduzido tamanho desta subamostra, pulou-se a etapa de seleção prévia das camadas mais externas. As eficiências foram calculadas pelo método DEA-C (ou CCR). A influências foram calculadas usando o algoritmo descrito na seção 4.4.

A Tabela 13 apresenta as 15 municipalidades mais influentes, ordenadas decrescentemente segundo a influência que causam no cálculo da eficiência das demais cidades.

A Tabela 14 apresenta os valores utilizados na estimativa da quantidade de penetras. Pelo teste F, com confiança de 95%, foram estimados cinco (5) penetras.

Tabela 13 : Municípios mais influentes – Bahia com um insumo e nove produtos

Seq	Município
1	Sapeaçu
2	Salinas da Margarida
3	Serrinha
4	Saubara
5	Santa Cruz da Vitória
6	Teodoro Sampaio
7	Jandaíra
8	Cravolândia
9	Feira de Santana
10	Santo Antônio de Jesus
11	Cardeal da Silva
12	Cachoeira
13	Muritiba
14	Tapiramutá
15	Guanambi

Tabela 14 : Teste F para identificar quantidade de penetras – Bahia com um insumo e nove produtos

Estatísticas retirando um município influente por vez				
Qt municípios	(1-Teta) ²	Razão	Teste F	
			alfa n-k	0,05 f
0	363,5256889	8,113132	377	1,184881
1	76,78786828	1,713744	376	1,185149
2	68,11846943	1,520262	375	1,185418
3	53,97672916	1,204648	374	1,185688
4	53,28632435	1,189239	373	1,185959
5	52,50276512	1,171752	372	1,186232
6	50,62140378	1,129764	371	1,186505
7	50,01789121	1,116295	370	1,186780
8	49,85831125	1,112733	369	1,187056
9	49,85830251	1,112733	368	1,187333
10	49,36721846	1,101773	367	1,187611
11	49,21831358	1,098450	366	1,187891
12	49,21594694	1,098397	365	1,188171
13	47,52966158	1,060763	364	1,188454
14	46,37288224	1,034946	363	1,188737
15	44,80707187	1,000000	362	1,189021

Os gráficos apresentados a seguir servem para corroborar a adequação do método, mostrando a melhoria de diversos indicadores sobre a massa dos dados após a retirada dos penetas identificados pelo método DPF.

O Gráfico 5 mostra a evolução da soma dos quadrados das ineficiências $(1-\theta)^2$. Como podemos verificar, a partir da retirada dos cinco municípios mais influentes, essa soma tende a se estabilizar, caracterizando que a retirada dos demais municípios classificados pouco influencia na determinação das eficiências.

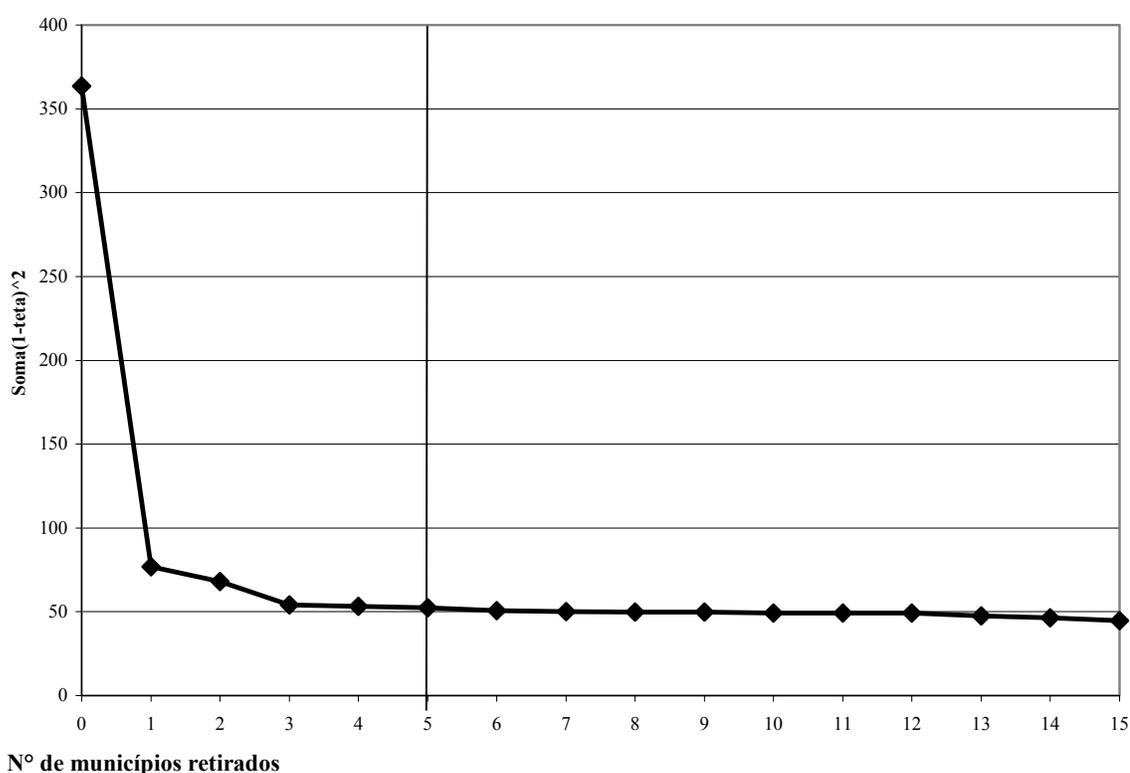


Gráfico 5 : Evolução da soma dos quadrados das ineficiências – Bahia com um insumo e nove produtos

Para a elaboração do

Gráfico 6, foram selecionadas aleatoriamente 10 municipalidades. Cada curva representa a evolução das eficiências de cada entidade, à medida que são retirados os municípios mais influentes. Como podemos notar, após a retirada do quinto penetra, as eficiências e as posições relativas tendem a se estabilizar, mostrando que a retirada dos demais municípios afeta pouco os valores calculados.

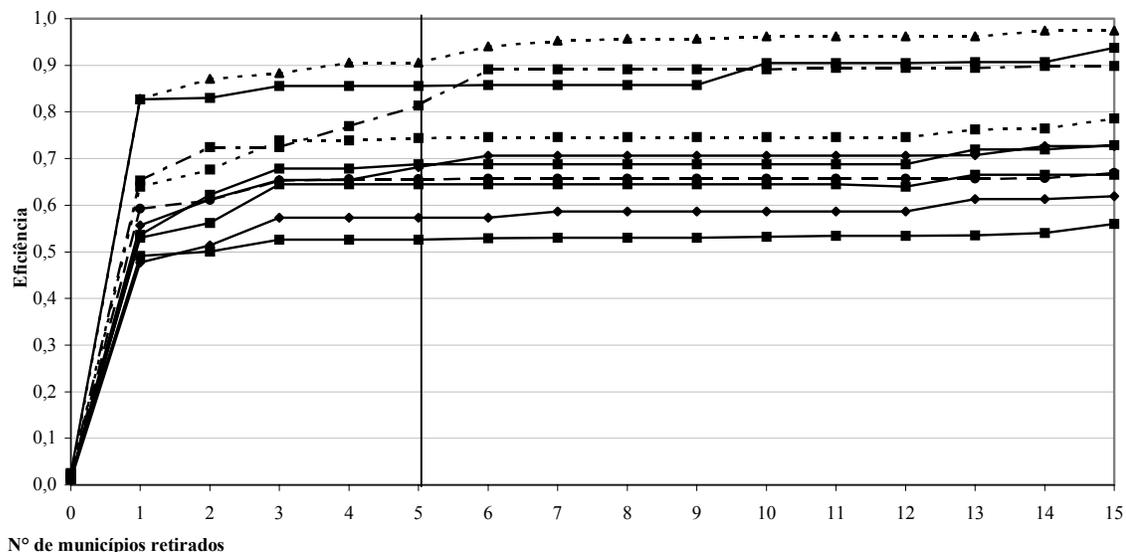


Gráfico 6 : Evolução das eficiências de municípios selecionados – Bahia com um insumo e nove produtos

Para montar o gráfico a seguir, os municípios foram classificados de forma decrescente em relação às eficiências calculadas sucessivamente quando são retirados da amostra os municípios mais influentes. Podemos então montar uma tabela com 377 linhas e 16 colunas. A coluna 0 correspondendo à posição que cada município recebe na classificação da amostra original e as colunas de 1 a 15 correspondendo às posições que cada município recebe na classificação após a retirada dos 15 primeiros municípios mais influentes. O gráfico apresenta a correlação entre os valores de duas colunas adjacentes. Ou seja, o primeiro ponto da curva é obtido pela correlação entre a ordem de cada município na amostra original (retirando 0 municípios) e a ordem que este município recebe na ordenação das eficiências calculadas após a retirada do município mais influente. O segundo ponto é obtido pela correlação entre as ordens após a retirada do município mais influente e as ordens após a retirada dos dois municípios mais influentes, e assim por diante. Novamente, podemos verificar a correção da quantidade estimada de penetras, pois as posições relativas dos municípios pouco variam, após a retirada dos cinco municípios mais influentes.

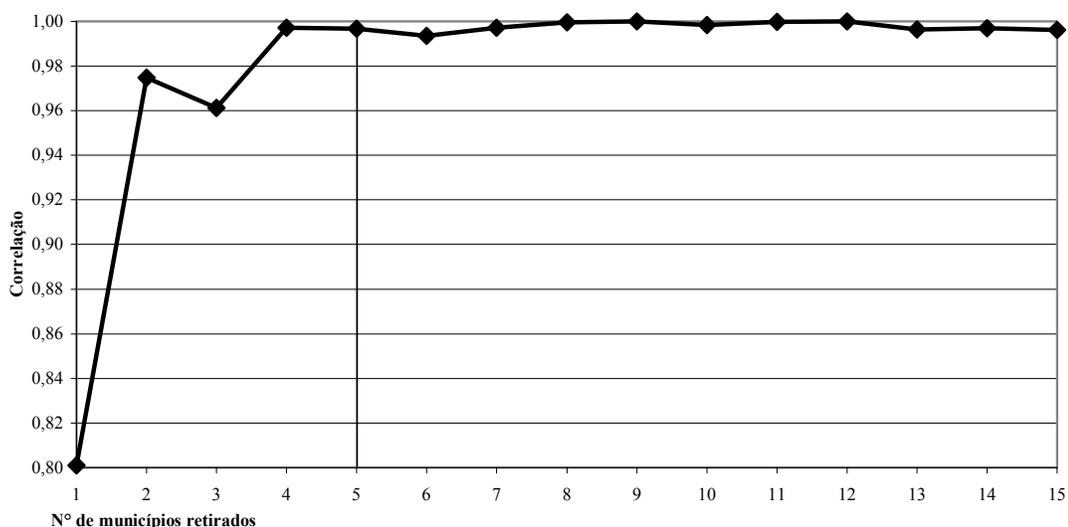


Gráfico 7 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades sequenciais de municípios retirados da amostra – municípios da Bahia – um insumo e nove produtos

Procedimento análogo foi usado para montar o

Gráfico 8. A diferença é que esse gráfico apresenta a correlação entre as ordens dos municípios comparando cada caso com aquele em que são retirados os 15 municípios mais influentes. Novamente, podemos verificar a correção da quantidade estimada de penetras.

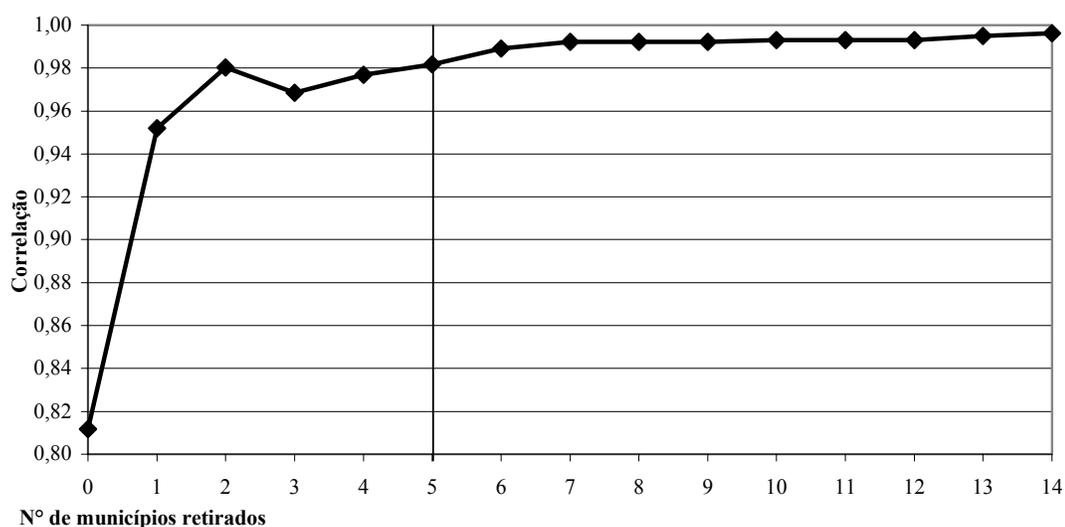


Gráfico 8 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 15 municípios mais influentes – municípios da Bahia – um insumo e nove produtos

O Gráfico 9 apresenta a distribuição das eficiências após retirarmos os 5 penetras estimados pelo método DPF.

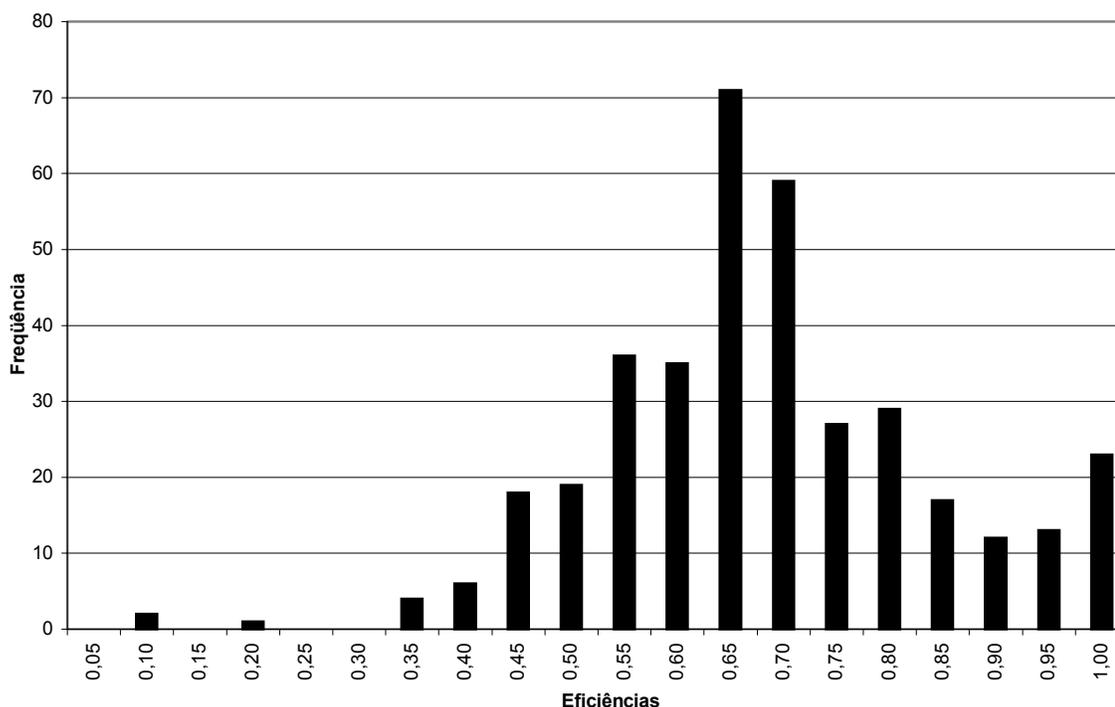


Gráfico 9 : Distribuição das eficiências retirando 5 penetras – municípios da Bahia – um insumo e nove produtos

Como podemos observar, o resultado final se apresenta muito melhor distribuído em comparação com o resultado obtido pela massa de dados original, conforme apresentado no Gráfico 1. A pequena subida da quantidade de municípios na parte final da distribuição pode ser considerada normal pelo fato do algoritmo limitar a eficiência em 1. Em uma distribuição não limitada, parte desses municípios estaria distribuídas em faixas superiores a 1.

QUATRO INSUMOS E NOVE PRODUTOS

Para testar a robustez do nosso método, a análise foi repetida para os mesmos municípios da Bahia, considerando agora os 4 insumos relacionados na Tabela 6.

A seguir, são apresentados os gráficos e tabelas resultantes desta análise. Os comentários e explicações são os mesmos já feitos para o exemplo anterior.

Tabela 15 : Municípios mais influentes – Bahia com quatro insumos e nove produtos

Seq	Município
1	Dias d'Ávila
2	Salvador
3	Feira de Santana
4	Salinas da Margarida
5	Sapeaçu
6	Saubara
7	Santa Cruz da Vitória
8	Nova Viçosa
9	Serrinha
10	Cravolândia
11	Madre de Deus
12	Maetinga
13	Camaçari
14	Guanambi
15	Alagoinhas
16	Irecê
17	Ilhéus
18	Teodoro Sampaio
19	Santo Antônio de Jesus
20	Terra Nova

Tabela 16 : Teste F para identificar quantidade de penetras – Bahia com quatro insumos e nove produtos

Estatísticas retirando um município influente por vez				
Qt municípios	Soma de $(1-Teta)^2$	Razão	Teste F	
			alfa n-k	0,05 f
0	51,876430	1,685135	377	1,184881
1	51,377550	1,668930	376	1,185149
2	50,301980	1,633991	375	1,185418
3	47,418914	1,540339	374	1,185688
4	46,551304	1,512156	373	1,185959
5	42,971669	1,395876	372	1,186232
6	40,673069	1,321209	371	1,186505
7	39,297825	1,276536	370	1,186780
8	39,062124	1,268880	369	1,187056
9	37,052768	1,203609	368	1,187333
10	36,852282	1,197096	367	1,187611
11	36,364395	1,181248	366	1,187891
12	35,762045	1,161681	365	1,188171
13	35,745696	1,161150	364	1,188454
14	35,350489	1,148312	363	1,188737
15	34,697584	1,127104	362	1,189021
16	34,325161	1,115006	361	1,189307
17	33,924597	1,101994	360	1,189594
18	33,445903	1,086445	359	1,189882
19	31,281406	1,016134	358	1,190172
20	30,784732	1,000000	357	1,190463

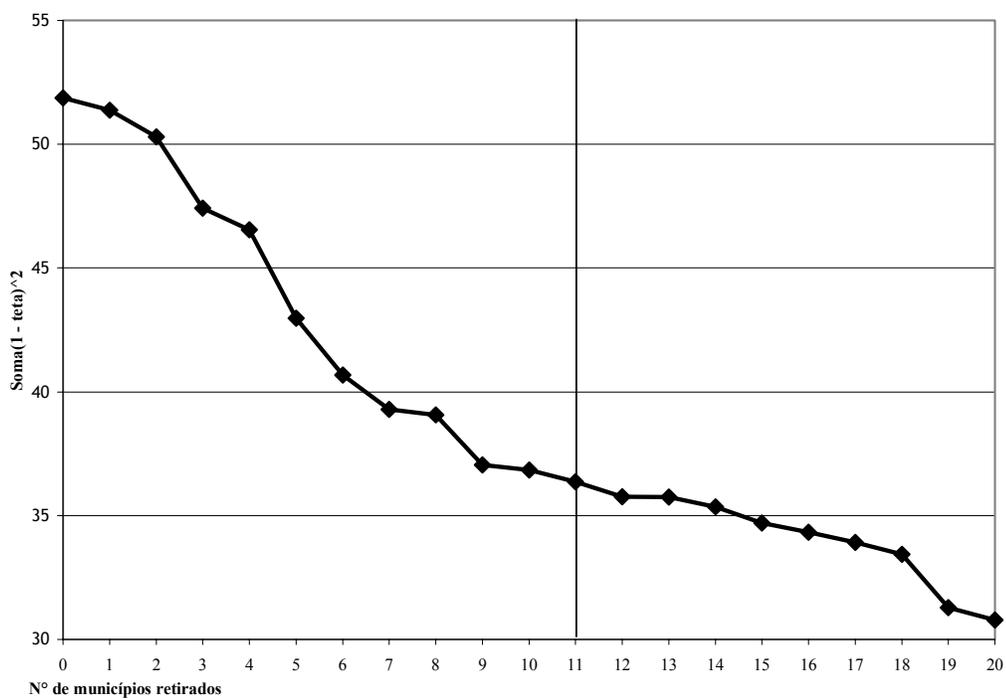


Gráfico 10 : Evolução da soma dos quadrados das ineficiências – Bahia com quatro insumos e nove produtos

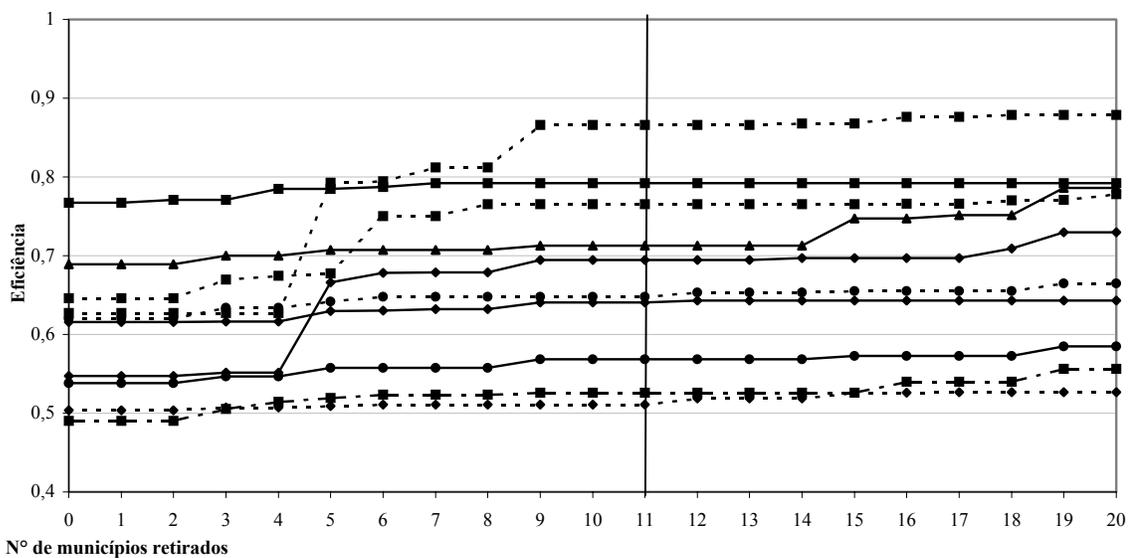


Gráfico 11 : Evolução das eficiências de municípios selecionados – Bahia com quatro insumos e nove produtos

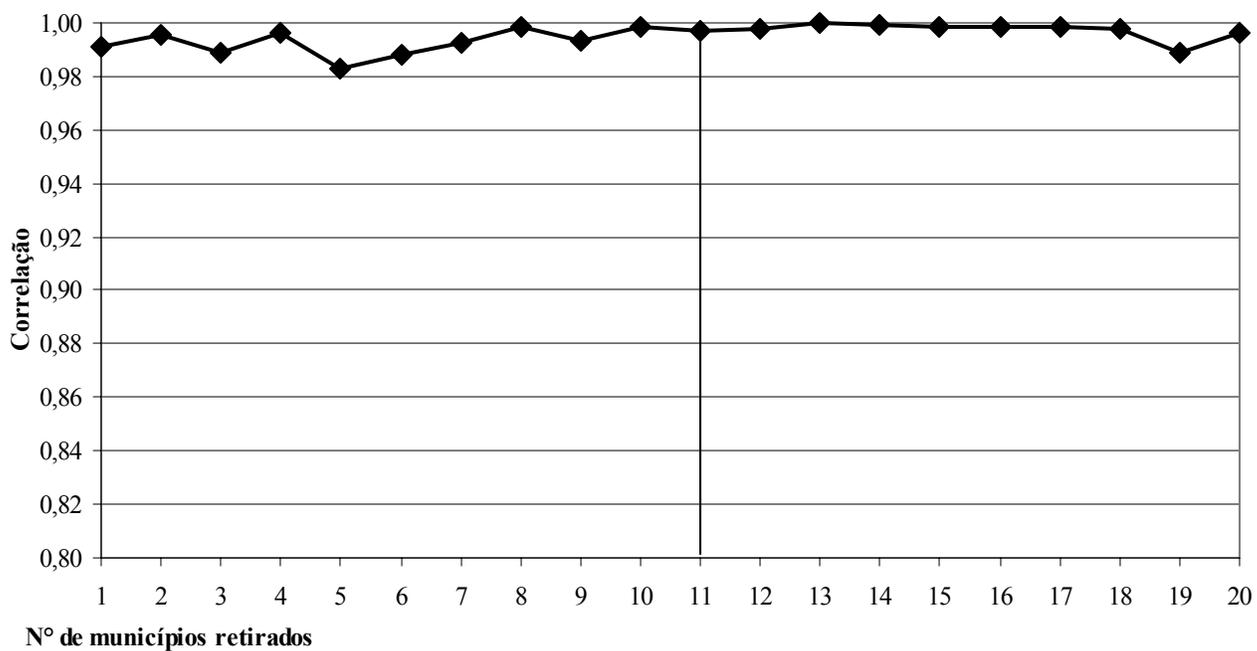


Gráfico 12 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Bahia com quatro insumos e nove produtos

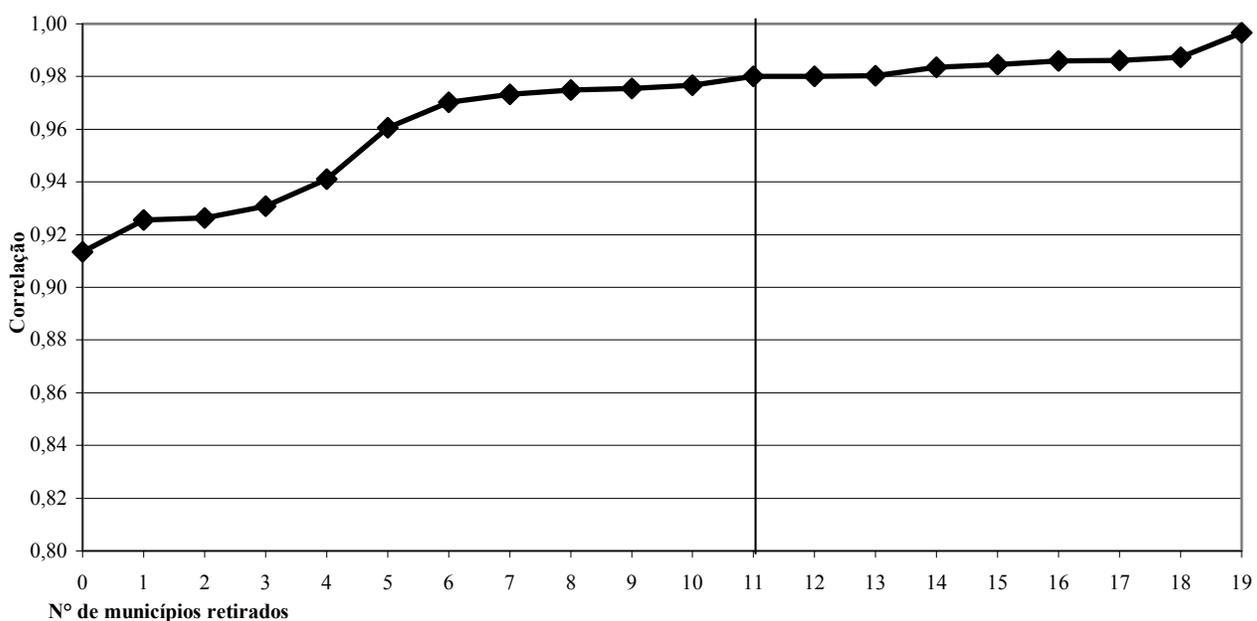


Gráfico 13 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 20 municípios mais influentes - Bahia com quatro insumos e nove produtos

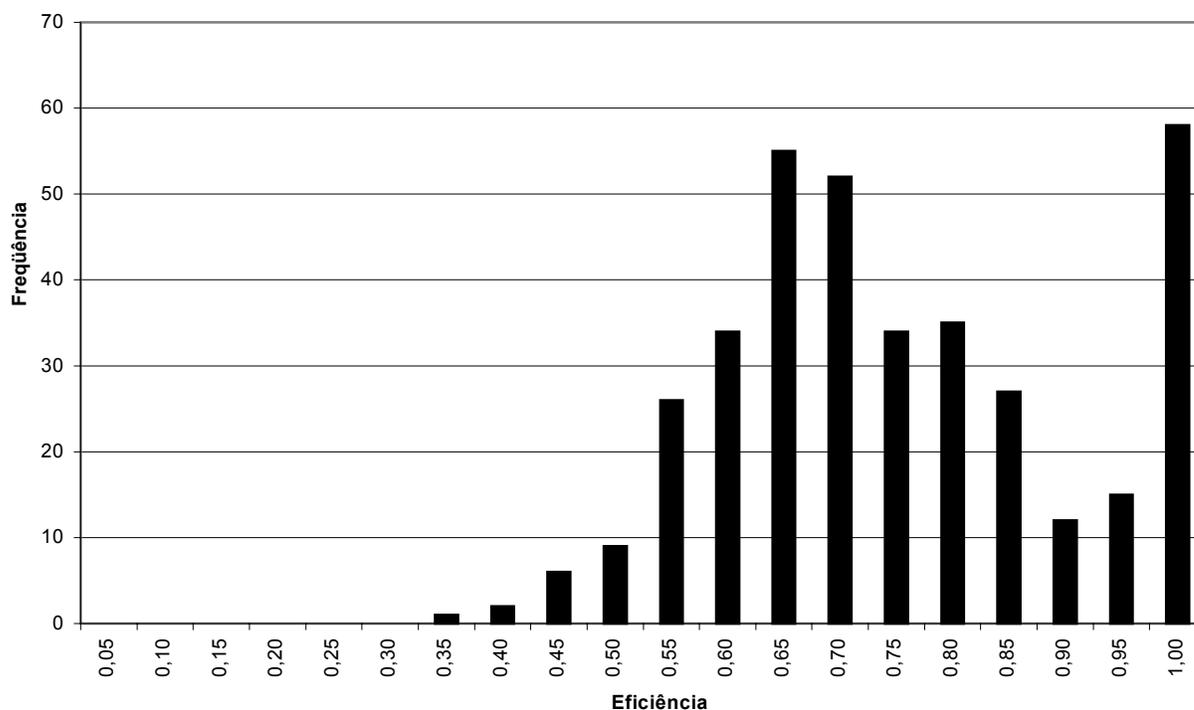


Gráfico 14 : Distribuição das eficiências dos municípios retirando 11 penetas - Bahia com quatro insumos e nove produtos

Como podemos observar, os gráficos e tabelas corroboram a adequação do método e da estimativa de quantidade de penetas a serem retirados. As observações são análogas às apresentadas no caso da amostra com apenas 1 insumo.

5.3.2. Municípios brasileiros

UM INSUMO E NOVE PRODUTOS

Para verificar a viabilidade computacional do método DPF, os cálculos citados na seção anterior foram repetidos para a totalidade dos municípios brasileiros. Novamente, tratamos o caso com apenas um insumo e, posteriormente, o caso com quatro insumos. Como já mencionado, a inspeção manual é ineficaz e impraticável, principalmente neste caso contendo 4.796 municípios.

Devido ao tamanho da amostra, foi necessário fazer a seleção prévia das camadas mais externas. Neste caso, com um insumo, selecionamos uma subamostra contendo 719 municípios das 17 camadas mais externas. A quantidade foi escolhida

em função da capacidade computacional e suficientemente “profunda” para garantir que todos os penetas pudessem ser detectadas, mesmo os que estivessem mascarados por outros penetas. As quantidades de municípios por camada, onde a camada 0 é a fronteira original, são relacionadas na Tabela 17. Esta tabela relaciona também as quantidades de penetas selecionados que se encontravam em cada camada.

Tabela 17 : Quantidades de municípios nas camadas selecionadas – Brasil com um insumo e nove produtos

Camada	Quantidades de municípios na camada	Quantidades de penetas na camada
0	4	4
1	3	3
2	2	2
3	1	1
4	2	2
5	11	4
6	16	1
7	25	0
8	26	0
9	51	0
10	47	0
11	66	0
12	66	0
13	86	0
14	100	0
15	109	0
16	104	0
Total	719	17

As tabelas e gráficos, apresentados em seqüência, seguem o mesmo raciocínio do exemplo anterior, com dados restritos ao estado da Bahia. Neste caso, o método DPF estimou 17 penetas, conforme apresentado na Tabela 19. As demais observações são as mesmas já apresentadas e, portanto, não serão repetidas.

Tabela 18 : Municípios mais influentes – Brasil com um insumo e nove produtos

Seq	Municípios mais influentes	Camada
1	Salto Veloso	0
2	Barra do Garças	0
3	Santo Antônio do Sudoeste	0
4	Pontal do Araguaia	0
5	São Francisco de Itabapoana	1
6	São José do Mantimento	1
7	Trajano de Moraes	1
8	Sapeaçu	2
9	Santo André	2
10	Congonhas do Norte	3
11	Nazareno	4
12	Quixadá	4
13	Divisa Alegre	5
14	Oratórios	6
15	Propriá	5
16	Orizânia	5
17	Caiçara do Norte	5
18	Nova Olinda do Maranhão	5
19	São José de Ribamar	5
20	Brasília de Minas	6
21	Alvorada	6
22	Palhoça	6
23	São Gonçalo	6

Tabela 19 : Teste F para identificar quantidade de penetras – Brasil com um insumo e nove produtos

Estatísticas retirando um município influente por vez				
Qt municípios	(1-Teta)^2	Razão	Teste F	
			alfa	0,05
			n-k	f
0	4.777,404129	3,014031	4796	1,048654
1	4.774,412714	3,012144	4795	1,048659
2	4.773,571613	3,011613	4794	1,048664
3	4.772,031220	3,010641	4793	1,048669
4	4.691,351069	2,959741	4792	1,048675
5	4.677,492724	2,950998	4791	1,048680
6	4.659,359041	2,939557	4790	1,048685
7	4.552,487079	2,872132	4789	1,048690
8	4.443,556681	2,803409	4788	1,048696
9	4.209,608643	2,655813	4787	1,048701
10	2.707,536081	1,708166	4786	1,048706
11	2.432,531645	1,534667	4785	1,048711
12	1.945,143356	1,227177	4784	1,048716
13	1.934,929135	1,220733	4783	1,048722
14	1.928,778778	1,216853	4782	1,048727
15	1.753,539123	1,106296	4781	1,048732
16	1.750,288866	1,104245	4780	1,048737
17	1.658,852049	1,046558	4779	1,048742
18	1.635,669408	1,031932	4778	1,048748
19	1.632,314169	1,029816	4777	1,048753
20	1.616,159907	1,019624	4776	1,048758
21	1.613,097792	1,017692	4775	1,048763
22	1.607,568392	1,014204	4774	1,048769
23	1.585,054741	1,000000	4773	1,048774

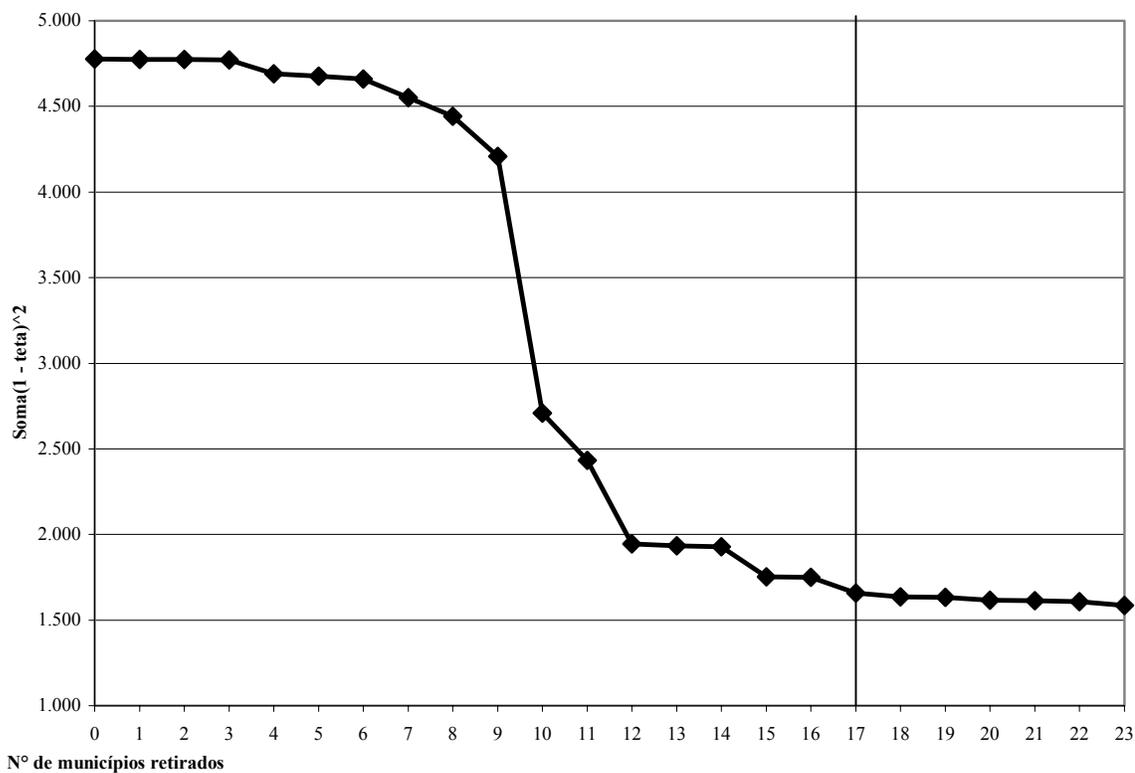


Gráfico 15 : Evolução da soma dos quadrados das ineficiências – Brasil com um insumo e nove produtos

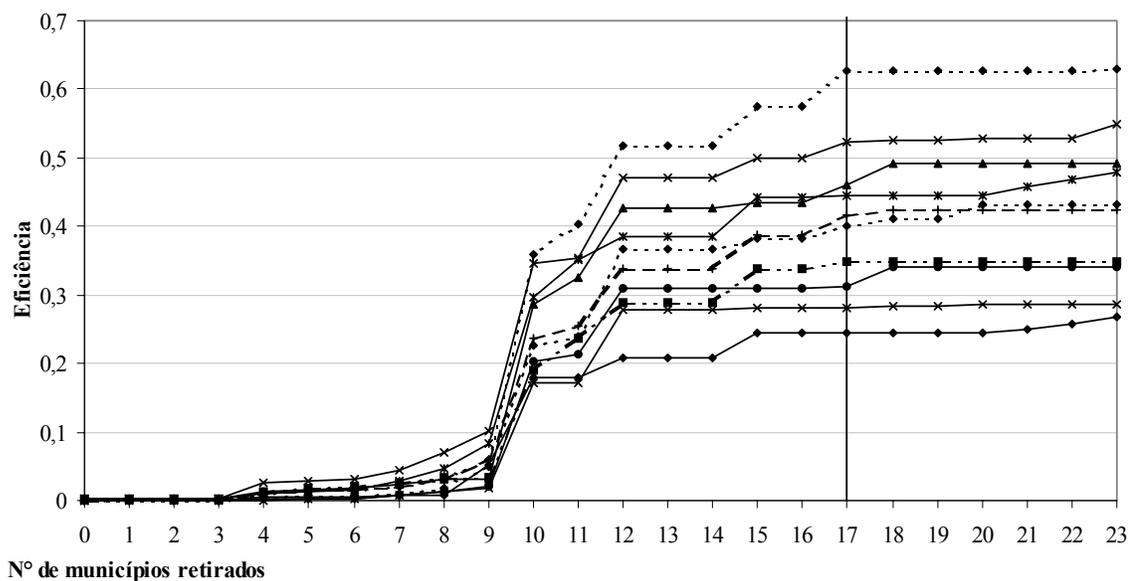


Gráfico 16 : Evolução das eficiências de municípios selecionados – Brasil com um insumo e nove produtos

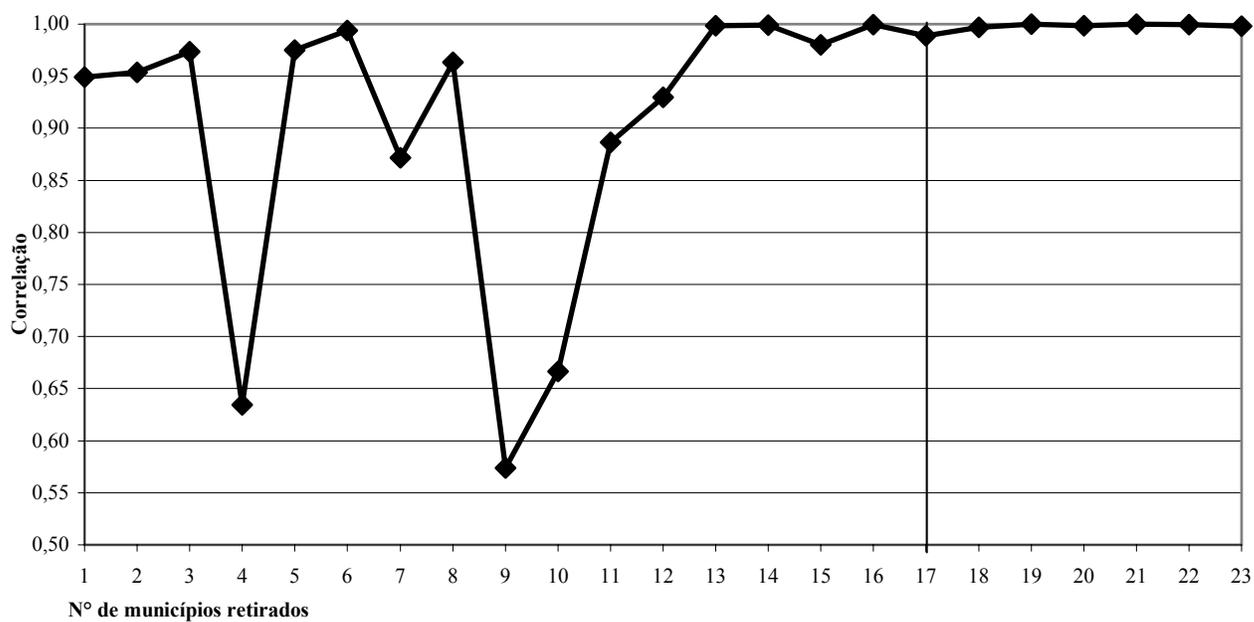


Gráfico 17 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Brasil com um insumo e nove produtos

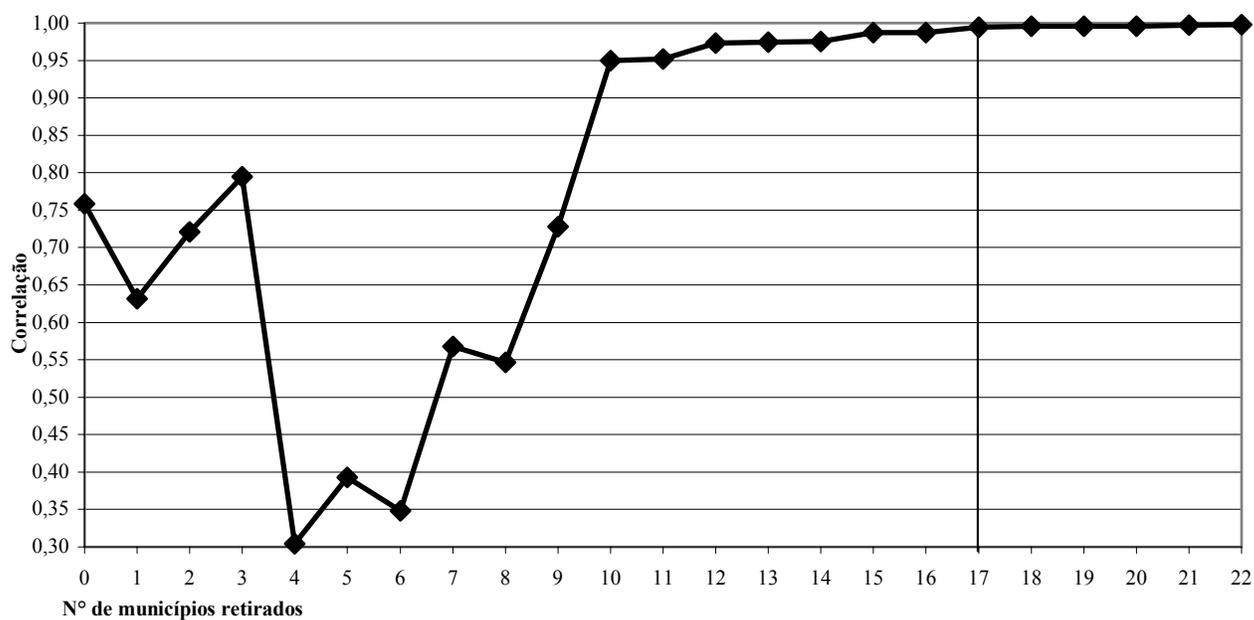


Gráfico 18 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 23 municípios mais influentes - Brasil com um insumo e nove produtos

O gráfico a seguir apresenta a distribuição das eficiências após retirarmos os 17 penetras estimados pelo método DPF.

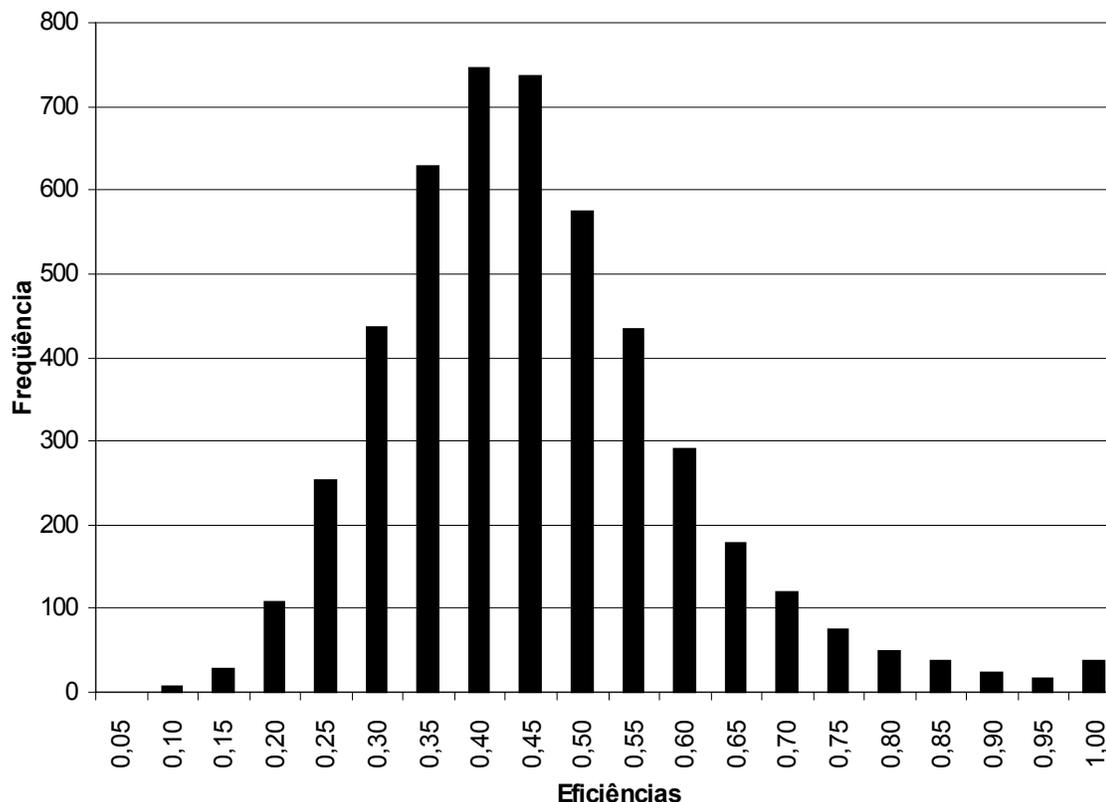


Gráfico 19 : Distribuição das eficiências retirando 17 penetras - Brasil com um insumo e nove produtos

O Gráfico 19 permite verificar a enorme melhora que obtivemos nos resultados, após o tratamento de retirada de penetras, comparativamente com os resultados obtidos com os dados originais, conforme apresentado no Gráfico 3. A pequena subida das quantidades na parte final da distribuição pode ser considerada normal pelo fato do algoritmo limitar a eficiência em 1. Em uma distribuição não limitada, parte desses municípios estariam distribuídos em faixas superiores a 1. A subida nesse caso não foi tão grande quanto nos casos anteriores em virtude da moda desta distribuição ser bem menor, ficando mais distante da cauda da distribuição.

QUATRO INSUMOS E NOVE PRODUTOS

No caso com quatro insumos, selecionamos uma subamostra contendo 668 municípios das 5 camadas mais externas. A quantidade foi escolhida em função da capacidade computacional e suficientemente “profunda” para garantir que todos os penstras pudessem ser detectadas, mesmo os que estivessem mascarados por outros penstras. A Tabela 20 relaciona as quantidades de municípios por camada, onde a camada 0 é a fronteira original. Esta tabela relaciona também as quantidades de penstras selecionados que se encontravam em cada camada.

Tabela 20 : Quantidades de municípios nas camadas selecionadas – Brasil com quatro insumos e nove produtos

Camada	Quantidades de municípios na camada	Quantidades de penstras na camada
0	41	19
1	87	4
2	136	0
3	176	0
4	228	0
Total	668	23

Como pode-se observar, pelos dados apresentados na Tabela 20, as quantidades de municípios nas camadas mais externas, levando-se em conta os quatro insumos, é bastante superior às quantidades apresentadas no caso com apenas um insumo (Tabela 17).

A Tabela 21 apresenta os 30 municípios mais influentes, ordenados decrescentemente segundo a influência que causam no cálculo da eficiência dos demais municípios. Usando o teste F, estimou-se a quantidade de 23 penstras, conforme apresentado na Tabela 22. Dos 30 municípios, 9 se encontravam na camada 1 e os demais na camada 0. Dos 23 penstras selecionados, 4 se encontravam na camada 1 e os demais na camada 0.

Tabela 21 : Municípios mais influentes – Brasil com quatro insumos e nove produtos

Seq	Município	Camada
1	Santo André	0
2	Barra do Garças	0
3	Salto Veloso	0
4	Taubaté	0
5	Alto Alegre	0
6	Nazareno	0
7	Oratórios	0
8	Francisco Morato	0
9	São Gonçalo	0
10	Salvador	0
11	São Paulo	0
12	Guarulhos	0
13	Iperó	0
14	Itanhaém	0
15	Orizânia	1
16	Pontal do Araguaia	0
17	Rio de Janeiro	1
18	Nova Iguaçu	1
19	Panorama	0
20	Divisa Alegre	0
21	São Francisco de Itabapoana	1
22	Novo Horizonte do Sul	0
23	Santo Antônio do Sudoeste	0
24	Piracicaba	0
25	Votuporanga	0
26	Mauá	1
27	Porto Alegre	1
28	Trajano de Moraes	1
29	Quixadá	1
30	Nossa Senhora do Socorro	1

Tabela 22 : Teste F para identificar quantidade de penetras – Brasil com quatro insumos e nove produtos

Estatísticas retirando um município influente por vez				
Qt municípios	Soma de (1-Teta)^2	Razão	Teste F	
			alfa	0,05
			n-k	f
0	2.458,195645	1,662674	4.796	1,048654
1	1.829,335527	1,237326	4.795	1,048659
2	1.697,677263	1,148275	4.794	1,048664
3	1.697,562308	1,148197	4.793	1,048669
4	1.696,527658	1,147497	4.792	1,048675
5	1.676,038842	1,133639	4.791	1,048680
6	1.640,571531	1,109650	4.790	1,048685
7	1.640,564848	1,109645	4.789	1,048690
8	1.614,994289	1,092350	4.788	1,048696
9	1.607,580753	1,087335	4.787	1,048701
10	1.602,894666	1,084166	4.786	1,048706
11	1.602,047795	1,083593	4.785	1,048711
12	1.601,851663	1,083460	4.784	1,048716
13	1.599,512586	1,081878	4.783	1,048722
14	1.598,972987	1,081513	4.782	1,048727
15	1.598,673259	1,081310	4.781	1,048732
16	1.591,864856	1,076705	4.780	1,048737
17	1.591,827179	1,076680	4.779	1,048742
18	1.590,246324	1,075611	4.778	1,048748
19	1.585,333618	1,072288	4.777	1,048753
20	1.572,002607	1,063271	4.776	1,048758
21	1.567,163813	1,059998	4.775	1,048763
22	1.563,060249	1,057223	4.774	1,048769
23	1.538,070663	1,040320	4.773	1,048774
24	1.537,825268	1,040154	4.772	1,048779
25	1.533,553786	1,037265	4.771	1,048784
26	1.533,086597	1,036949	4.770	1,048789
27	1.532,952623	1,036858	4.769	1,048795
28	1.496,128594	1,011951	4.768	1,048800
29	1.489,605844	1,007539	4.767	1,048805
30	1.478,459034	1,000000	4.766	1,048810

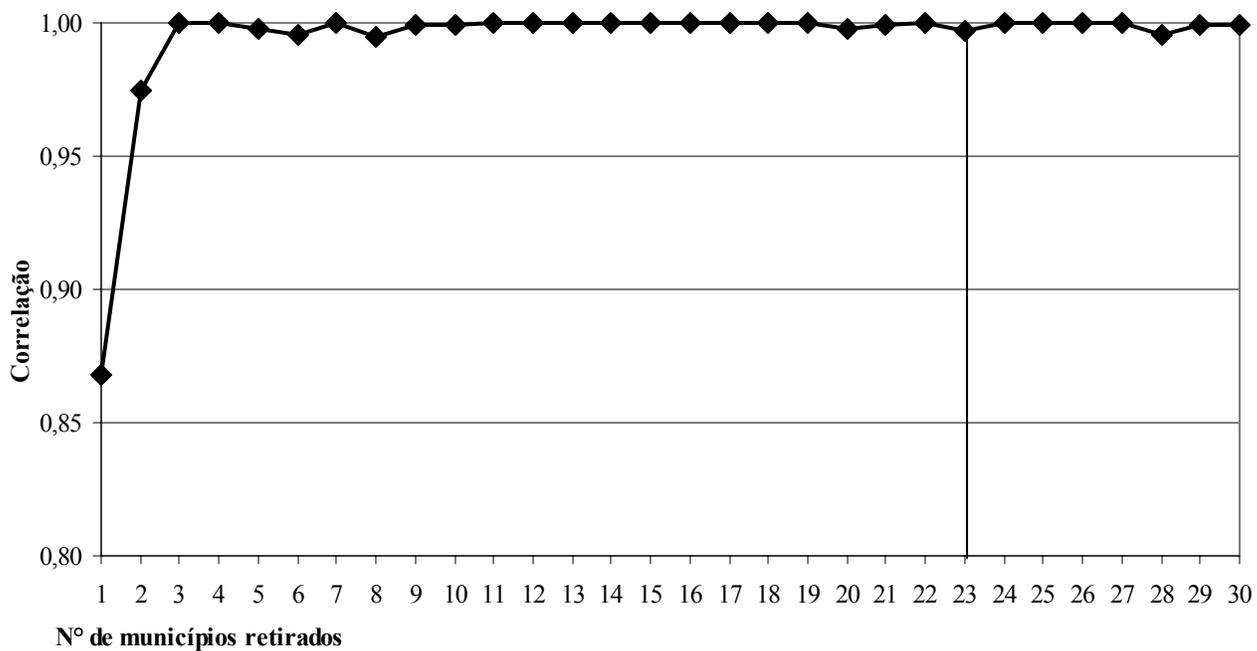


Gráfico 22 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando quantidades seqüenciais de municípios retirados da amostra – Brasil com quatro insumos e nove produtos

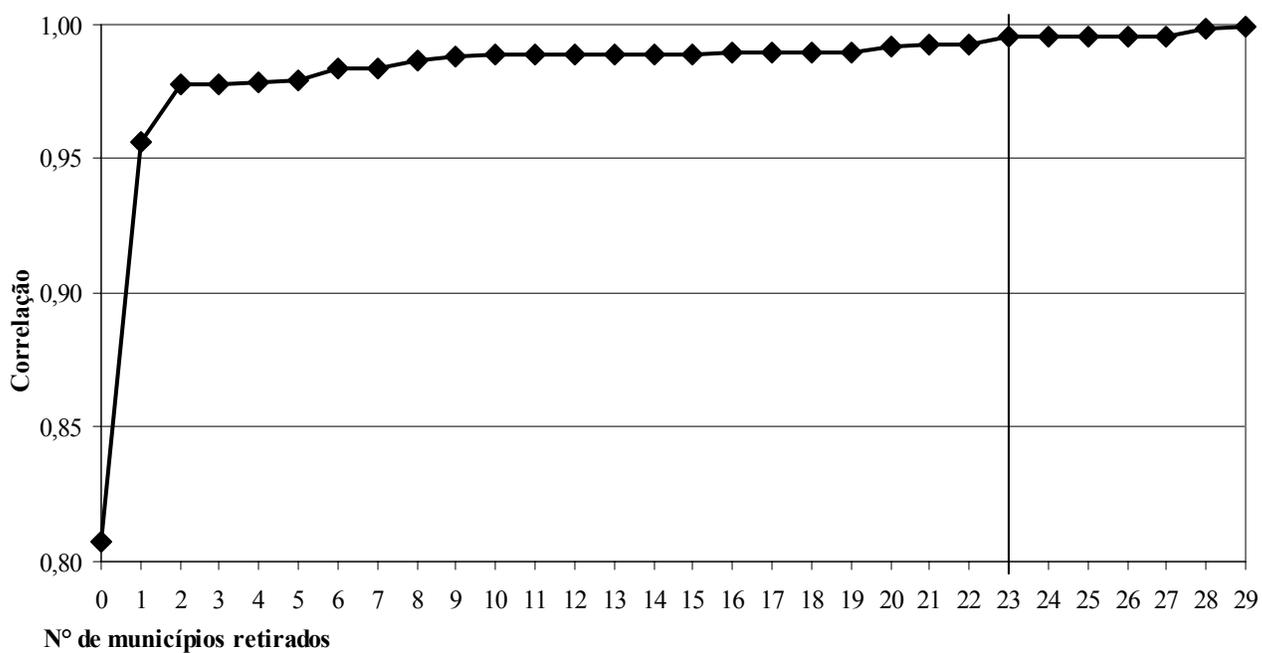


Gráfico 23 : Correlação das posições dos municípios na ordenação decrescente de eficiências, comparando-se com a retirada dos 30 municípios mais influentes – Brasil com quatro insumos e nove produtos

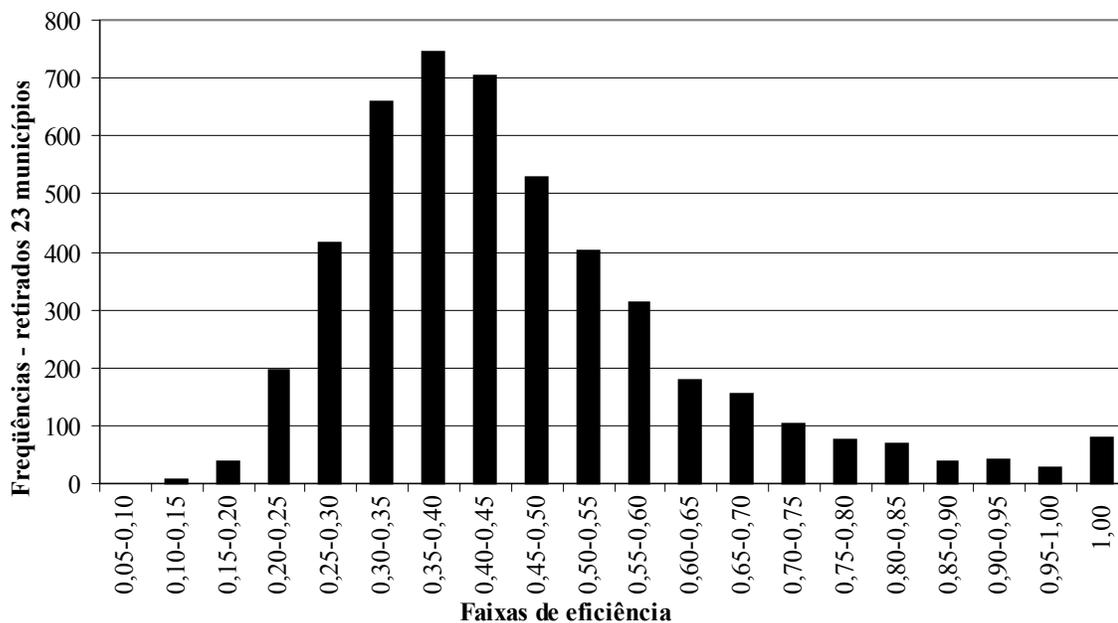


Gráfico 24 : Distribuição das eficiências retirando 23 penetas– Brasil com quatro insumos e nove produtos

A distribuição, após a retirada dos penetas identificados, está muito mais bem equilibrada do que a distribuição original, apresentada no Gráfico 4. A pequena subida das frequências na parte final da distribuição pode ser considerada normal pelo fato do algoritmo limitar a eficiência em 1. Em uma distribuição não limitada, parte desses municípios estariam distribuídos em faixas superiores a 1.

6. CONCLUSÃO

Este trabalho apresentou um novo método de detecção de penstras para ser utilizado em métodos não paramétricos de determinação de fronteiras de eficiência denominado Detecção Acelerada de Penstras por Inspeção de Fronteiras não Paramétricas, DPF. Essa metodologia, inspirada na abordagem MRE (Mensuração Robusta de Eficiência) desenvolvida por Cherchye e Kuosmanen, ao identificar erros e observações atípicas, permite construir um conjunto de dados capaz de gerar estimativas robustas dos índices de eficiência produtiva.

Para ilustrar o funcionamento desse método, ele foi aplicado em diferentes bases de dados contendo informações, que incluíram *proxies* para nove produtos e até quatro insumos, referentes a até 4796 municipalidades brasileiras. Compararam-se, então, os índices de eficiência calculados com base na versão original da abordagem AED (Análise Envoltória de Dados), em sua variante que apresenta retornos constantes de escala, com aqueles obtidos por meio do uso da metodologia DPF, utilizando a mesma variante do método AED.

Os resultados obtidos mostraram que os indicadores calculados pelo nosso método são mais plausíveis e bem mais robustos do aqueles computados por meio da aplicação direta da AED. Os resultados empíricos corroboraram, também, o teste de estimação da quantidade de penstras. Assim, a adequação do método DPF ficou demonstrada, tanto na detecção e eliminação de penstras quanto na viabilidade computacional em casos de grandes amostras. O trabalho completo na amostra contendo 4.796 municipalidades pôde ser realizado em 26 horas de processamento, usando um microcomputador de média velocidade (Pentium III de 650 Mhz). Este ponto é importante visto que, até agora, poucos são os métodos que funcionam comprovadamente bem em grandes amostras.

As possibilidades de aplicações do método proposto são amplas e incluem as diversas situações econômicas em que a construção de fronteiras não paramétricas de eficiência produtiva é utilizada. Ademais, essa abordagem é particularmente eficaz

para revelar a existência de penetras em grandes amostras, o que a torna apropriada para o uso em questões que envolvem auditoria e controle, em uma grande massa de dados.

Com a retirada dos penetras, os valores calculados com a amostra resultante são estimativas robustas para as reais eficiências relativas, permitindo melhor embasamento para as conclusões a serem retiradas dos resultados. As unidades mais eficientes podem servir de exemplo às demais unidades e as unidades menos eficientes poderão ser inspecionadas para identificação das causas das ineficiências (dolosas ou não), bem como para receber treinamento e sugestões de melhoria..

Uma extensão natural desse trabalho é ampliar a análise para incluir os resultados referentes às demais variantes da metodologia AED, incluindo aquelas em que a tecnologia é não convexa (FDH – Free Disposal Hull) , bem como a modificação do algoritmo para inserir a detecção de *inliers*, isto é, aquelas observações atípicas ou que contêm erros, porém, que não estão na fronteira do conjunto produtivo.

7. APÊNDICE – TEORIA UTILIZADA

Neste anexo, são apresentadas as notações, as definições básicas e os teoremas que são utilizados na apresentação dos diversos modelos AED para conjuntos de produção e nos métodos de estimação de eficiência do uso de insumos em unidades de produção.

7.1. Notações e definições básicas

Notação 1 : Vetores em \mathfrak{R}^N e matrizes-coluna $N \times 1$ em \mathfrak{R}

Os vetores em \mathfrak{R}^N e as correspondentes matrizes-coluna $N \times 1$ em \mathfrak{R} são designados por letras minúsculas em negrito. Em função da associação biunívoca entre esses espaços, seus elementos correspondentes são referidos, de forma indistinta, pelas mesmas letras.

Definição 1 : Seqüência de elementos de um conjunto C

Uma seqüência de elementos de C é uma função cujo domínio é um subconjunto dos números naturais e cuja imagem é um subconjunto de C .

$$S : D(S) \subset \mathbf{N} \rightarrow C$$

Notação 2 : Seqüências de números reais ou de vetores em \mathfrak{R}^N

Os elementos de seqüências são designados por índices sobrescritos.

Exemplo 1 : \mathbf{x}^k (k-ésimo elemento de uma seqüência de vetores).

Exemplo 2 : y^k (k-ésimo elemento de uma seqüência de números reais)

Notação 3 : Coordenadas de vetores em \mathfrak{R}^N

As coordenadas de vetores são designadas por índices subscritos.

Exemplo 3 : x_n (n-ésima coordenada do vetor \mathbf{x}).

Exemplo 4 : x_n^k (n-ésima coordenada do k-ésimo vetor \mathbf{x}^k).

Notação 4 : Matrizes $M \times N$ de números reais

As matrizes são designadas por letras maiúsculas.

Exemplo 5 : X, Y (matrizes)

Notação 5 : Elementos de matrizes

Os elementos de matrizes são designados por letras minúsculas com primeiro subscrito indicando a linha e o segundo subscrito indicando a coluna.

Exemplo 6 : x_{mn} (elemento da linha m e coluna n)

Notação 6 : Desigualdade entre vetores : $\mathbf{x} > \mathbf{y}$, $\mathbf{x} \geq \mathbf{y}$, $\mathbf{x} < \mathbf{y}$, $\mathbf{x} \leq \mathbf{y}$

$\mathbf{x} \geq \mathbf{y}$ equivale a : $x_i \geq y_i$, $i = 1, 2, \dots$

$\mathbf{x} \leq \mathbf{y}$ equivale a : $x_i \leq y_i$, $i = 1, 2, \dots$

$\mathbf{x} > \mathbf{y}$ equivale a : $x_i \geq y_i$, $i = 1, 2, \dots$ e $x_j > y_j$ para algum j

$\mathbf{x} < \mathbf{y}$ equivale a : $x_i \leq y_i$, $i = 1, 2, \dots$ e $x_j < y_j$ para algum j

Definição 2 : Insumos : \mathbf{In}

Conjunto de todos os possíveis insumos que podem ser utilizados em uma dada tecnologia.

$$\mathbf{In} = \mathfrak{R}_+^N = \{(x_1, x_2, \dots, x_N) \mid x_i \in \mathfrak{R}_+, i = 1, 2, \dots, N\}$$

Definição 3 : Produtos : \mathbf{Pr}

Conjunto de todos os possíveis produtos que podem ser gerados por uma dada tecnologia.

$$\mathbf{Pr} = \mathfrak{R}_+^M = \{(y_1, y_2, \dots, y_M) \mid y_i \in \mathfrak{R}_+, i = 1, 2, \dots, M\}$$

Definição 4 : Universo de possibilidades de produção : \mathbf{U}

$$\mathbf{U} = \mathbf{In} \times \mathbf{Pr}$$

Definição 5 : Conjunto de produção de uma dada tecnologia : $\mathbf{P}(T)$

Conjunto de todos os possíveis pares de insumos e os correspondentes produtos que podem ser gerados por uma dada tecnologia.

$P(T) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in \mathfrak{R}^N_+, \mathbf{y} \in \mathfrak{R}^M_+ \text{ e } \mathbf{y} \text{ pode ser produzido a partir de } \mathbf{x}, \text{ usando a tecnologia } T\}$

$$P(T) \subset \mathbf{U} = \text{In} \times \text{Pr} = \mathfrak{R}^N_+ \times \mathfrak{R}^M_+$$

Definição 6 : Correspondência de insumos : $\text{In}(\mathbf{y} \mid T)$

Seja T uma tecnologia, $P(T)$ seu conjunto de produção e $\mathbf{y} \in \text{Pr} = \mathfrak{R}^M_+$.

A seção $\text{In}(\mathbf{y} \mid T) \subset \text{In} = \mathfrak{R}^N_+$, definida como :

$$\text{In}(\mathbf{y} \mid T) = \{\mathbf{x} \mid (\mathbf{x}, \mathbf{y}) \in P(T)\}$$

É denominada correspondência de insumos.

Definição 7 : Correspondência de produtos : $\text{Pr}(\mathbf{x} \mid T)$

Seja T uma tecnologia, $P(T)$ seu conjunto de produção e $\mathbf{x} \in \text{In} = \mathfrak{R}^N_+$.

A seção $\text{Pr}(\mathbf{x} \mid T) \subset \text{Pr} = \mathfrak{R}^M_+$, definida como :

$$\text{Pr}(\mathbf{x} \mid T) = \{\mathbf{y} \mid (\mathbf{x}, \mathbf{y}) \in P(T)\}$$

É denominada correspondência de produtos. Logo :

$$P(T) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in \text{Pr} = \mathfrak{R}^M_+ \text{ e } \mathbf{x} \in \text{In}(\mathbf{y} \mid T)\}$$

$$P(T) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in \text{In} = \mathfrak{R}^N_+ \text{ e } \mathbf{y} \in \text{Pr}(\mathbf{x} \mid T)\}$$

Definição 8 : Complemento de um conjunto de produção : $P^c(T)$

$$P^c(T) = \{(\mathbf{x}, \mathbf{y}) \mid (\mathbf{x}, \mathbf{y}) \notin P(T)\}$$

Definição 9 : Bola de raio δ e centro em $(\mathbf{x}^0, \mathbf{y}^0)$: $B(\mathbf{x}^0, \mathbf{y}^0, \delta)$

$$B(\mathbf{x}^0, \mathbf{y}^0, \delta) = \{(\mathbf{x}, \mathbf{y}) \mid \text{dist}((\mathbf{x}, \mathbf{y}), (\mathbf{x}^0, \mathbf{y}^0)) < \delta\}$$

Definição 10 : Interior de um conjunto de produção $P(T)$: $\text{Int}(P(T))$

$$\text{Int}(P(T)) = \{(\mathbf{x}, \mathbf{y}) \mid \exists \delta > 0 \text{ tal que } B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)\}$$

Definição 11 : Fronteira de um conjunto de produção $P(T)$: $\text{Front}(P(T))$

$$\text{Front}(P(T)) = \{(\mathbf{x}, \mathbf{y}) \mid \forall \delta > 0, B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) \neq \Phi \text{ e } B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi\}$$

Definição 12 : $\text{Front}(P(T)) \subset P(T)$

Matematicamente, essa condição não é obrigatória. No entanto, como estamos modelando situações reais, seria estranho se existisse $(\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$ tal que $(\mathbf{x}, \mathbf{y}) \notin P(T)$. Portanto, convencionaremos que $\text{Front}(P(T)) \subset P(T)$.

Definição 13 : Conjunto amostra de produção de uma dada tecnologia : A

Conjunto formado de pares de insumos e produtos observados.

$$A = \{(\mathbf{x}^k, \mathbf{y}^k) \mid \mathbf{x}^k \in \mathfrak{R}_+^N, \mathbf{y}^k \in \mathfrak{R}_+^M, k = 1, 2, \dots, K \text{ em que } \mathbf{y}^k \text{ foi produzido a partir de } \mathbf{x}^k\}$$

$$K_A = \{1, 2, \dots, K\}$$

$$A \subset P(T) \subset \mathbf{U} = \text{In} \times \text{Pr} = \mathfrak{R}_+^N \times \mathfrak{R}_+^M$$

Definição 14 : Matriz de insumos de uma amostra : X

Matriz cujas colunas são os vetores de insumos observados.

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^K \\ x_2^1 & x_2^2 & \dots & x_2^K \\ \dots & \dots & \dots & \dots \\ x_N^1 & x_N^2 & \dots & x_N^K \end{bmatrix} \quad X \text{ é uma matriz } N \text{ por } K$$

Definição 15 : Matriz de produtos de uma amostra : Y

Matriz cujas colunas são os vetores de produtos observados.

$$Y = \begin{bmatrix} y_1^1 & y_1^2 & \dots & y_1^K \\ y_2^1 & y_2^2 & \dots & y_2^K \\ \dots & \dots & \dots & \dots \\ y_M^1 & y_M^2 & \dots & y_M^K \end{bmatrix} \quad Y \text{ é uma matriz } M \text{ por } K$$

Definição 16 : Vetor unitário da dimensão k : \mathbf{e}^k

Vetor com coordenadas nulas, exceto a k -ésima que é igual à unidade.

$$\mathbf{e}^k = (0, \dots, 1, \dots, 0)$$

Definição 17 : Diagonal principal do K-cubo unitário : \mathbf{e}

Vetor com K coordenadas iguais à unidade : $\mathbf{e} = (1, 1, \dots, 1)$

Definição 18 : Combinação linear convexa de insumos de uma amostra : $X\mathbf{z}$

Vetor resultante de uma combinação linear convexa dos vetores de insumos observados.

$$X\mathbf{z} = \sum_{i=1}^K z_i \mathbf{x}^i = \left(\sum_{i=1}^K z_i x_{11}^i, \sum_{i=1}^K z_i x_{12}^i, \dots, \sum_{i=1}^K z_i x_{1N}^i \right) \in \mathfrak{R}_+^N$$

$$\mathbf{z} \in \mathfrak{R}_+^K, \quad \mathbf{e}'\mathbf{z} = 1$$

Definição 19 : Combinação linear convexa de produtos de uma amostra : $Y\mathbf{z}$

Vetor resultante de uma combinação linear convexa dos vetores de produtos observados.

$$Y\mathbf{z} = \sum_{i=1}^K z_i \mathbf{y}^i = \left(\sum_{i=1}^K z_i y_{11}^i, \sum_{i=1}^K z_i y_{12}^i, \dots, \sum_{i=1}^K z_i y_{1M}^i \right) \in \mathfrak{R}_+^M$$

$$\mathbf{z} \in \mathfrak{R}_+^K, \quad \mathbf{e}'\mathbf{z} = 1$$

Definição 20 : Tecnologia convexa

Sejam T uma tecnologia e $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\} \subset P(T)$ um conjunto amostra de produção. Dizemos que T é uma tecnologia convexa se $(X\mathbf{z}, Y\mathbf{z}) \in P(T)$, sempre que $\mathbf{z} \in \mathfrak{R}_+^K$ e que $\mathbf{e}'\mathbf{z} = 1$.

Exemplo 7 : Conjunto de produção com tecnologia convexa

Em uma tecnologia convexa, qualquer combinação linear de seus elementos, cujos pesos sejam não negativos e cuja soma seja igual a 1 (combinação linear convexa), também será um elemento do conjunto de produção. Para melhor entender este conceito, mostramos graficamente uma amostra contendo dois elementos representados pelos pontos A e B. Por simplicidade, usamos uma caso com 1 insumo e 1 produto.

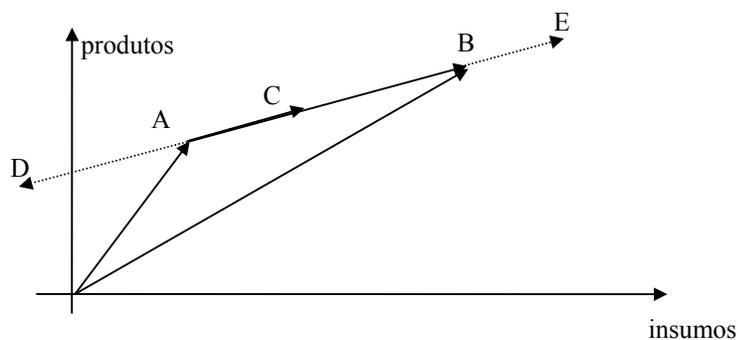


Figura 14 : Conjunto de produção com tecnologia convexa

Os pontos sobre a reta que passa por A e B podem ser escritos, usando notação vetorial, como :

$$P = A + \alpha AB \quad \text{onde } \alpha \in \mathfrak{R}$$

No caso do ponto C, situado no segmento entre A e B, temos que :

$$C = A + AC$$

Mas, como AC tem a mesma direção de AB, mesmo sentido e módulo menor, esse vetor pode ser escrito como :

$$AC = \alpha AB \quad \text{onde } 0 \leq \alpha \leq 1$$

Logo :

$$C = A + AC = A + \alpha AB = A + \alpha(B - A) = (1 - \alpha)A + \alpha B$$

Denominando $z_1 = (1 - \alpha)$ e $z_2 = \alpha$, podemos observar que :

$$z_1 + z_2 = 1, \quad 0 \leq z_1 \leq 1, \quad 0 \leq z_2 \leq 1$$

Portanto, as combinações lineares convexas de A e B são os pontos localizados no segmento entre A e B.

No caso do ponto D, situado fora do segmento entre A e B (no exemplo mostrado na figura, à esquerda de A), podemos verificar que $\alpha < 0$. Nesse caso $z_1 > 1$ e $z_2 < 0$.

No caso do ponto E, situado fora do segmento entre A e B (no exemplo mostrado na figura, à direita de B), podemos verificar que $\alpha > 1$. Nesse caso $z_1 < 0$ e $z_1 > 1$.

Generalizando para mais pontos, pode-se verificar que os pontos associados a combinações convexas situam-se no interior e na fronteira do menor poliedro convexo que contém todos os pontos da combinação.

Definição 21 : Tecnologia com rendimentos de escala constantes : C

Sejam T uma tecnologia e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala constantes se $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$, sempre que $\alpha \in \mathfrak{R}_+$.

Definição 22 : Tecnologia com rendimentos de escala não crescentes : N

Sejam T uma tecnologia e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala não crescentes se $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$, sempre que $0 \leq \alpha \leq 1$.

Definição 23 : Tecnologia com rendimentos de escala variáveis : V

Sejam T uma tecnologia e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com rendimentos de escala variáveis se podemos garantir que $(\alpha\mathbf{x}, \alpha\mathbf{y}) \in P(T)$ apenas quando $\alpha = 1$.

Definição 24 : Tecnologia com livre disponibilidade de insumos : DIIn

Sejam T uma tecnologia e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com livre disponibilidade de insumos se $(\mathbf{z}, \mathbf{y}) \in P(T)$, sempre que $\mathbf{z} \geq \mathbf{x}$.

Definição 25 : Tecnologia com livre disponibilidade de produtos : DPr

Sejam T uma tecnologia e $(\mathbf{x}, \mathbf{y}) \in P(T)$. Dizemos que T é uma tecnologia com livre disponibilidade de produtos se $(\mathbf{x}, \mathbf{z}) \in P(T)$, sempre que $\mathbf{z} \leq \mathbf{y}$.

Definição 26 : CPG(A, a, b, c, ...)

Conjunto de produção gerado a partir de um conjunto amostra A e das características da tecnologia a, b, c, ...

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia que obedece a propriedades a, b, c, ...

Seja o conjunto $\{(x, y) \mid (x, y) \text{ pode ser obtido a partir dos elementos de } A, \text{ aplicando as propriedades } a, b, c, \dots\}$.

Denominamos esse conjunto de $\text{CPG}(A, a, b, c, \dots)$.

Exemplo 8 : $\text{CPG}(A, c, C)$

CPG de tecnologia convexa (c) e com rendimentos de escala constantes (C).

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c) e com rendimentos de escala constantes (C).

Combinando as definições de tecnologia convexa e com rendimentos de escala constantes, o conjunto de produção gerado a partir de A pode ser definido como :

$$\text{CPG}(A, c, C) = \{(x, y) \mid x = \alpha Xw, y = \alpha Yw, e^t w = 1, \alpha \in \mathfrak{R}_+, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, w \in \mathfrak{R}_+^K\}.$$

O conjunto de produção acima pode ser reescrito como :

$$\text{CPG}(A, c, C) = \{(x, y) \mid x = Xz, y = Yz, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, z \in \mathfrak{R}_+^K\}.$$

Para isso, basta fazermos $\alpha w = z$.

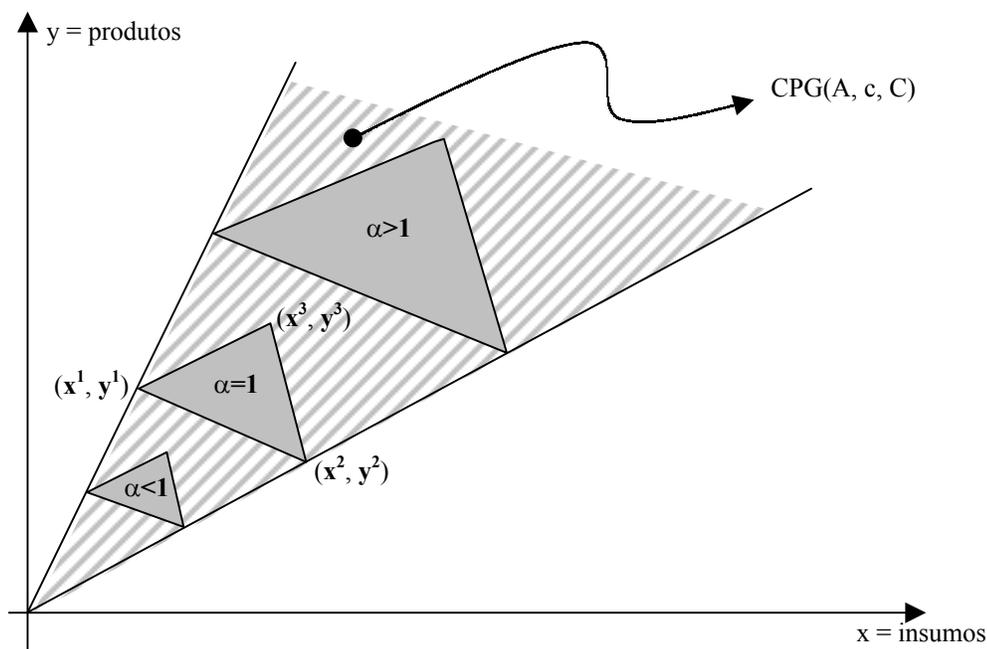


Figura 15: $\text{CPG}(A, c, C)$

Exemplo 9 : $CPG(A, c, N)$

CPG de tecnologia convexa (c) e com rendimentos de escala não crescentes (N).

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c) e com rendimentos de escala não crescentes (N).

Combinando as definições de tecnologia convexa e com rendimentos de escala não crescentes, o conjunto de produção gerado a partir de A pode ser definido como :

$$CPG(A, c, N) = \{(x, y) \mid x = \alpha Xw, y = \alpha Yw, e^t w = 1, 0 \leq \alpha \leq 1, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, w \in \mathfrak{R}_+^K\}.$$

O conjunto de produção acima pode ser reescrito como :

$$CPG(A, c, N) = \{(x, y) \mid x = Xz, y = Yz, e^t z \leq 1, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, z \in \mathfrak{R}_+^K\}.$$

Para isso, basta fazermos $\alpha w = z$.

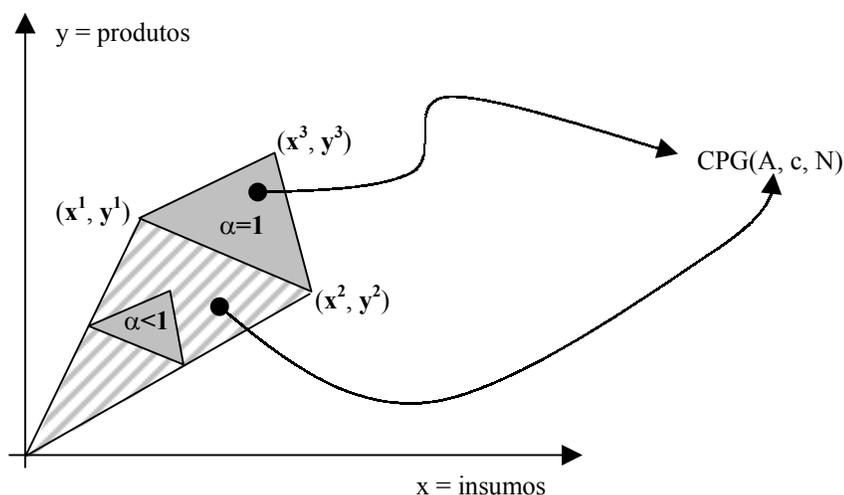


Figura 16: $CPG(A, c, N)$

Exemplo 10 : $CPG(A, c, V)$

CPG de tecnologia convexa (c) e com rendimentos de escala variáveis (V).

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c) e com rendimentos de escala variáveis (V).

Combinando as definições de tecnologia convexa e com rendimentos de escala variáveis, o conjunto de produção gerado a partir de A pode ser definido como :

$$\text{CPG}(A, c, V) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} = \alpha X\mathbf{w}, \mathbf{y} = \alpha Y\mathbf{w}, \alpha = 1, \mathbf{e}^t \mathbf{w} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{w} \in \mathfrak{R}_+^K\}.$$

O conjunto de produção acima pode ser reescrito como :

$$\text{CPG}(A, c, V) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} = X\mathbf{z}, \mathbf{y} = Y\mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

Para isso, basta fazermos $\alpha \mathbf{w} = \mathbf{z}$.

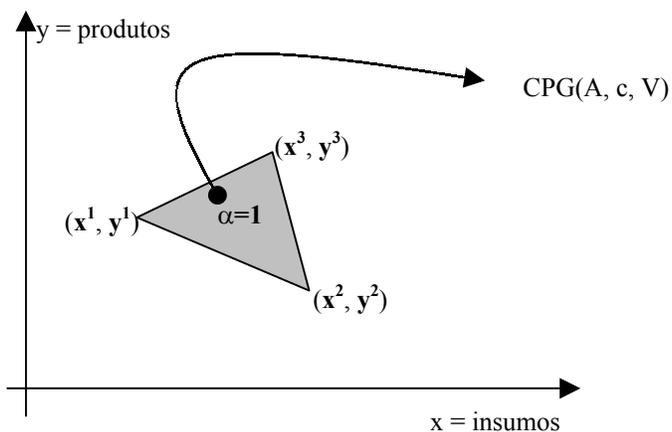


Figura 17: CPG(A, c, V)

Exemplo 11 : CPG(A, DIn)

CPG de tecnologia com livre disponibilidade de insumos (DIn).

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia com livre disponibilidade de insumos (DIn).

Então $\text{CPG}(A, \text{DIn}) = \{(\mathbf{x}, \mathbf{y}) \mid \exists k = 1, 2, \dots, K \text{ tal que } \mathbf{x} \geq X\mathbf{e}^k \text{ e } \mathbf{y} = Y\mathbf{e}^k\}$.

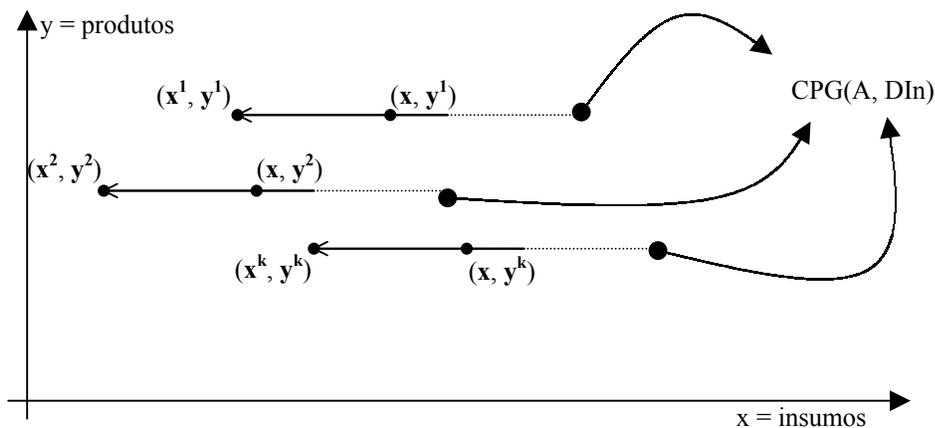


Figura 18: CPG(A, DIn)

Exemplo 12 : $\text{CPG}(A, \text{DPr})$

CPG de tecnologia com livre disponibilidade de produtos (DPr).

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia com livre disponibilidade de produtos (DPr).

Então $\text{CPG}(A, \text{DPr}) = \{(\mathbf{x}, \mathbf{y}) \mid \exists k = 1, 2, \dots, K \text{ tal que } \mathbf{x} = X\mathbf{e}^k \text{ e } \mathbf{y} \leq Y\mathbf{e}^k\}$.

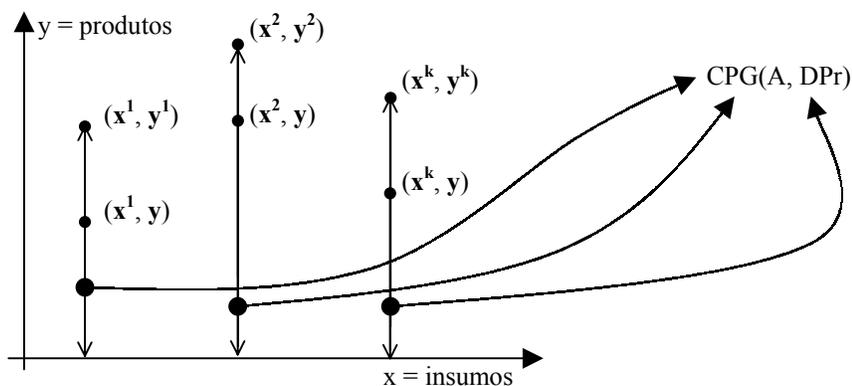


Figura 19: $\text{CPG}(A, \text{DPr})$

Exemplo 13 : $\text{CPG}(A, \text{DIn}, \text{DPr})$ ou $\text{CPG}(A, \text{AED-FDH})$

CPG de tecnologia com livre disponibilidade de insumos (DIn) e produtos (DPr). Este conjunto de propriedades será denominado de AED-FDH.

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia com livre disponibilidade de insumos (DIn) e produtos (DPr).

Então $\text{CPG}(A, \text{AED-FDH}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \{0,1\}^K\}$.

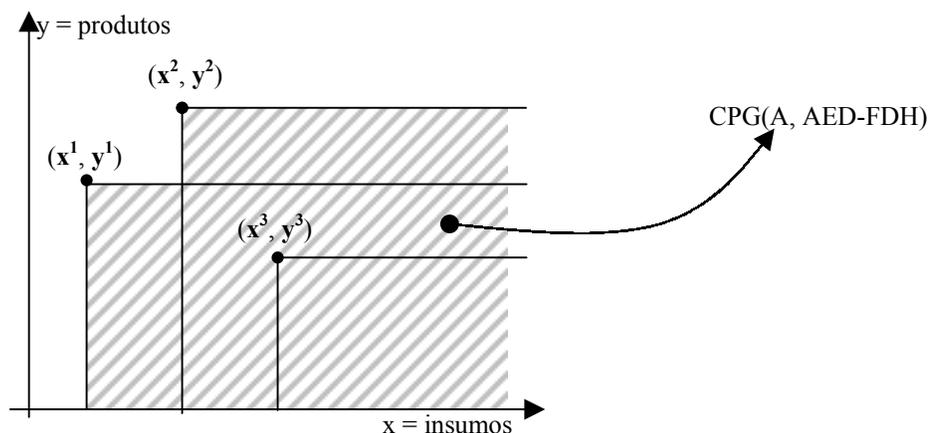


Figura 20: $\text{CPG}(A, \text{AED-FDH})$

Exemplo 14 : CPG(A, c, C, DIn, DPr) ou CPG(A, AED-C)

CPG de tecnologia convexa (c), com rendimentos de escala constantes (C) e livre disponibilidade de insumos (DIn) e produtos (DPr). Este conjunto de propriedades será denominado de AED-C.

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala constantes (C) e livre disponibilidade de insumos (DIn) e produtos (DPr).

Então $CPG(A, AED-C) = \{(x, y) \mid x \geq Xz, y \leq Yz, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, z \in \mathfrak{R}_+^K\}$.

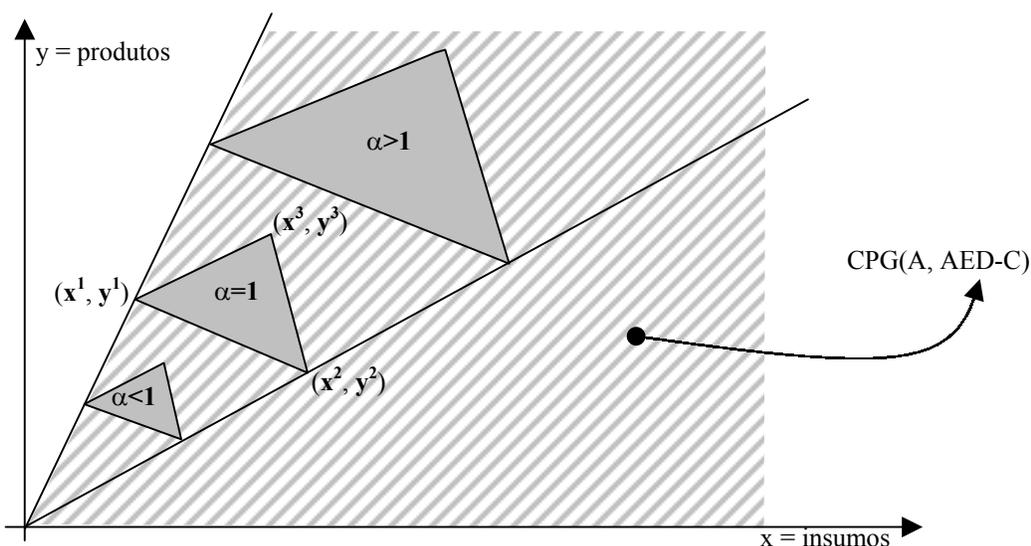


Figura 21: CPG(A, AED-C)

Exemplo 15 : CPG(A, c, N, DIn, DPr) ou CPG(A, AED-N)

CPG de tecnologia convexa (c), com rendimentos de escala não crescentes (N) e livre disponibilidade de insumos (DIn) e produtos (DPr). Este conjunto de propriedades será denominado de AED-N.

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala não crescentes (N) e livre disponibilidade de insumos (DIn) e produtos (DPr).

Então $CPG(A, AED-N) = \{(x, y) | x \geq Xz, y \leq Yz, e^t z = 1, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, z \in \mathfrak{R}_+^K\}$.

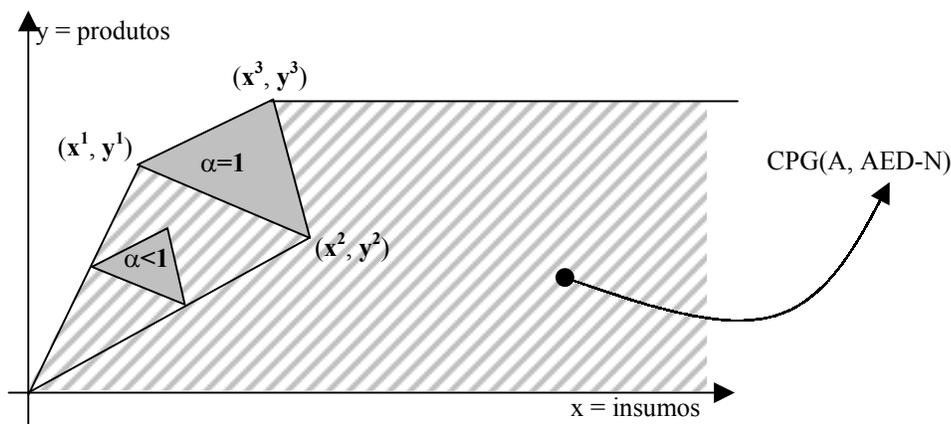


Figura 22: CPG(A, AED-N)

Exemplo 16 : CPG(A, c, V, DIn, DPr) ou CPG(A, AED-V)

CPG de tecnologia convexa (c), com rendimentos de escala variáveis (V) e livre disponibilidade de insumos (DIn) e produtos (DPr). Este conjunto de propriedades será denominado de AED-V.

Seja $A = \{(x^1, y^1), (x^2, y^2), \dots, (x^K, y^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala variáveis (V) e livre disponibilidade de insumos (DIn) e produtos (DPr).

Então $CPG(A, AED-V) = \{(x, y) | x \geq Xz, y \leq Yz, e^t z = 1, x \in \mathfrak{R}_+^N, y \in \mathfrak{R}_+^M, z \in \mathfrak{R}_+^K\}$.

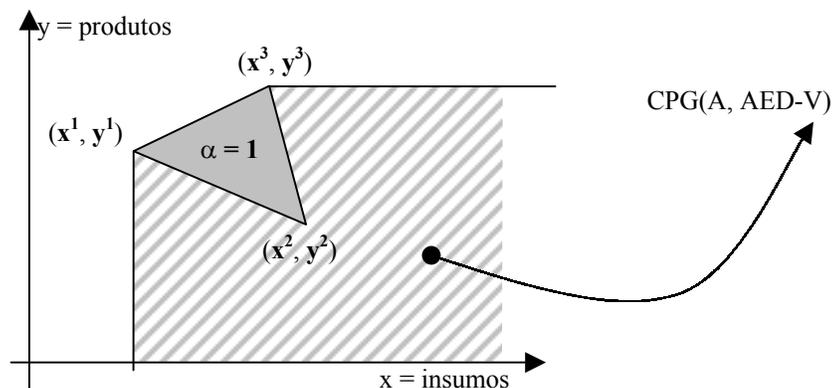


Figura 23: CPG(A, AED-V)

7.2. Medida de eficiência no uso dos insumos

Analizamos apenas as medidas de eficiência no uso dos insumos. As medidas de eficiência na geração dos produtos seguem raciocínio análogo.

Definição 27 : Fronteira de eficiência de insumos (no sentido de Farrell)

$$\text{Front}_f(\mathbf{y} | T) = \{\mathbf{x} | \mathbf{x} \in \text{In}(\mathbf{y} | T) \text{ e } \theta\mathbf{x} \notin \text{In}(\mathbf{y} | T) \forall \theta \in [0, 1)\}$$

Definição 28 : Função grau de eficiência de insumos : $E_i(\mathbf{x}, \mathbf{y} | T)$

$$E_i(\mathbf{x}, \mathbf{y} | T) = \text{mín} \{\theta | \theta\mathbf{x} \in \text{In}(\mathbf{y} | T)\}.$$

Como, da definição de $\text{In}(\mathbf{y} | T)$, temos que $\theta\mathbf{x} \in \text{In}(\mathbf{y} | T) \Leftrightarrow (\theta\mathbf{x}, \mathbf{y}) \in P(T)$, podemos usar a definição apresentada a seguir.

Definição 29 : Função grau de eficiência de insumos : $E_i(\mathbf{x}, \mathbf{y} | T)$

Seja T uma tecnologia e (\mathbf{x}, \mathbf{y}) uma produção observada de T .

Seja $(\theta\mathbf{x}, \mathbf{y})$ uma produção factível em T , onde $\theta \in \mathfrak{R}_+$.

Seja o conjunto $R = \{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.

Definimos $E_i(\mathbf{x}, \mathbf{y} | T) = \min R = \min \{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.

7.3. Estimativa de eficiência no uso dos insumos a partir de amostra

Definição 30 : Estimativa de eficiência - caso geral

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia que obedece a propriedades a, b, c, \dots

Seja $\text{CPG}(A, a, b, c, \dots)$ o conjunto de produção gerado a partir de A e das características a, b, c, \dots da tecnologia.

Seja o conjunto $R^k = \{\theta | (\theta\mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, a, b, c, \dots), \theta \in \mathfrak{R}_+\}$.

Definimos $E_i(\mathbf{x}^k, \mathbf{y}^k | A, a, b, c, \dots) = \min R^k = \min \{\theta | (\theta\mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, a, b, c, \dots), \theta \in \mathfrak{R}_+\}$.

Apesar de, em geral, só usarmos a definição para a estimativa dos elementos de A, essa definição e as seguintes podem ser estendidas para qualquer ponto $(\mathbf{x}, \mathbf{y}) \in \text{CPG}(A, a, b, c, \dots)$.

Definição 31 : Estimativa de eficiência - caso AED-C

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala constantes (C) e livre disponibilidade de insumos (DIn) e produtos (DPr). Neste caso, temos:

$$\text{CPG}(A, \text{AED-C}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$\begin{aligned} R(\mathbf{x}, \mathbf{y}) &= \{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in \text{CPG}(A, \text{AED-C}), \theta \in \mathfrak{R}_+\} = \{\theta \mid \theta\mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \\ &\mathbf{Yz}, \theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}. \end{aligned}$$

$$\begin{aligned} E_i(\mathbf{x}, \mathbf{y} \mid A, \text{AED-C}) &= \theta^{\text{AED-C}} = \min R(\mathbf{x}, \mathbf{y}) = \min \{\theta \mid \theta\mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \\ &\theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}. \end{aligned}$$

Para obtermos a eficiência de uma observação $(\mathbf{x}^k, \mathbf{y}^k)$, precisamos resolver o seguinte problema de minimização :

$$E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = \theta^{\text{AED-C}} =$$

$$\min_{\theta, \mathbf{z}} \quad s.a. \quad \begin{cases} \theta x_n^k \geq \sum_{j=1}^K z_j x_n^j & n = 1, 2, \dots, N \\ y_m^k \leq \sum_{j=1}^K z_j y_m^j & m = 1, 2, \dots, M \\ z_j \geq 0 & j = 1, 2, \dots, K \\ \theta \geq 0 \end{cases}$$

Este sistema de inequações pode ser resolvido utilizando-se o método simplex.

Definição 32 : Estimativa de eficiência - caso AED-N

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala não crescentes (N) e livre disponibilidade de insumos (DIn) e produtos (DPr). Neste caso, temos :

$$\text{CPG}(A, \text{AED-N}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \mathbf{e}^t \mathbf{z} \leq 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$R(\mathbf{x}, \mathbf{y}) = \{\theta \mid (\theta \mathbf{x}, \mathbf{y}) \in \text{CPG}(A, \text{AED-N}), \theta \in \mathfrak{R}_+\} = \{\theta \mid \theta \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} \leq 1, \theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$E_i(\mathbf{x}, \mathbf{y} \mid A, \text{AED-N}) = \theta^{\text{AED-N}} = \min R(\mathbf{x}, \mathbf{y}) = \min \{\theta \mid \theta \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} \leq 1, \theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

Para obtermos a eficiência de uma observação $(\mathbf{x}^k, \mathbf{y}^k)$, precisamos resolver o seguinte problema de minimização :

$$E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-N}) = \theta^{\text{AED-N}} = \min_{\theta, \mathbf{z}} \quad s.a. \quad \begin{cases} \theta x_n^k \geq \sum_{j=1}^K z_j x_n^j & n = 1, 2, \dots, N \\ y_m^k \leq \sum_{j=1}^K z_j y_m^j & m = 1, 2, \dots, M \\ \sum_{j=1}^K z_j \leq 1 \\ z_j \geq 0 & j = 1, 2, \dots, K \\ \theta \geq 0 \end{cases}$$

Este sistema de inequações pode ser resolvido utilizando-se o método simplex.

Definição 33 : Estimativa de eficiência - caso AED-V

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia convexa(c), com rendimentos de escala variáveis (V) e livre disponibilidade de insumos (DIn) e produtos (DPr). Neste caso, temos :

$$\text{CPG}(A, \text{AED-V}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$R(\mathbf{x}, \mathbf{y}) = \{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, \text{AED-V}), \theta \in \mathfrak{R}_+\} = \{\theta \mid \theta \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, \theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$E_i(\mathbf{x}, \mathbf{y} \mid A, \text{AED-V}) = \theta^{\text{AED-V}} = \min R(\mathbf{x}, \mathbf{y}) = \min \{\theta \mid \theta \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{e}^t \mathbf{z} = 1, \theta \in \mathfrak{R}_+, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

Para obtermos a eficiência de uma observação $(\mathbf{x}^k, \mathbf{y}^k)$, precisamos resolver o seguinte problema de minimização :

$$E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-V}) = \theta^{\text{AED-V}} =$$

$$\min_{\theta, z} \quad s.a. \quad \begin{cases} \theta x_n^k \geq \sum_{j=1}^K z_j x_n^j & n = 1, 2, \dots, N \\ y_m^k \leq \sum_{j=1}^K z_j y_m^j & m = 1, 2, \dots, M \\ \sum_{j=1}^K z_j = 1 \\ z_j \geq 0 & j = 1, 2, \dots, K \\ \theta \geq 0 \end{cases}$$

Este sistema de inequações pode ser resolvido utilizando-se o método simplex.

Definição 34 : Estimativa de eficiência - caso AED-FDH

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia com livre disponibilidade de insumos (DIn) e produtos (DPr). Neste caso, temos :

$$\text{CPG}(A, \text{AED-FDH}) = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \mathbf{e}^t \mathbf{z} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \{0, 1\}^K\}.$$

$$\begin{aligned} \text{R}(\mathbf{x}, \mathbf{y}) &= \{\theta | (\theta \mathbf{x}, \mathbf{y}) \in \text{CPG}(A, \text{AED-FDH}), \theta \in \mathfrak{R}_+\} = \{\theta | \theta \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \mathbf{e}^t \mathbf{z} \\ &= 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \{0, 1\}^K\}. \end{aligned}$$

$$\begin{aligned} E_i(\mathbf{x}, \mathbf{y} | A, \text{AED-FDH}) &= \theta^{\text{AED-FDH}} = \min \text{R}(\mathbf{x}, \mathbf{y}) = \min \{\theta | \theta \mathbf{x} \geq \mathbf{Xz}, \mathbf{y} \leq \mathbf{Yz}, \\ &\mathbf{e}^t \mathbf{z} = 1, \mathbf{x} \in \mathfrak{R}_+^N, \mathbf{y} \in \mathfrak{R}_+^M, \mathbf{z} \in \{0, 1\}^K\}. \end{aligned}$$

Para obtermos a eficiência de uma observação $(\mathbf{x}^k, \mathbf{y}^k)$, podemos usar o seguinte algoritmo :

- Para cada $j = 1, 2, \dots, K$, compare \mathbf{x}^k com \mathbf{x}^j e \mathbf{y}^k com \mathbf{y}^j .
- Se $\mathbf{x}^k \geq \mathbf{x}^j$ e $\mathbf{y}^k \leq \mathbf{y}^j$, dizemos que $(\mathbf{x}^k, \mathbf{y}^k)$ é dominado por $(\mathbf{x}^j, \mathbf{y}^j)$, pois $(\mathbf{x}^k, \mathbf{y}^k)$ usa pelo menos os insumos usados por $(\mathbf{x}^j, \mathbf{y}^j)$ e gera no máximo os produtos gerados por $(\mathbf{x}^j, \mathbf{y}^j)$.
- Nesse caso, para cada $x_i^k \neq 0$, calculamos $\theta^{kj}_i = x_i^j / x_i^k$.
- Tomemos $\theta^{kj} = \max \theta^{kj}_i, i = 1, 2, \dots, N$.

- Podemos, então, garantir que θ^{kj} é o menor valor tal que $\theta^{kj} \mathbf{x}^k \geq \mathbf{x}^j$.
- Tomemos $\theta^k = \min \theta^{kj}$, $j = 1, 2, \dots, K$
- Pelo algoritmo, $\theta^k = \theta^{\text{AED-FDH}}$.

Definição 35 : Dominância de produção.

Dizemos que $(\mathbf{x}^d, \mathbf{y}^d)$ é dominado por $(\mathbf{x}^D, \mathbf{y}^D)$ se $\mathbf{x}^d \geq \mathbf{x}^D$ e $\mathbf{y}^d \leq \mathbf{y}^D$.

7.4. Teoremas usados

Teorema 1 : $P(T) \cup P^c(T) = \mathbf{U}$

Demonstração :

1. Axioma : $\forall (\mathbf{x}, \mathbf{y}) \in \mathbf{U}$, temos que $(\mathbf{x}, \mathbf{y}) \in P(T)$ ou $(\mathbf{x}, \mathbf{y}) \notin P(T)$
2. De 1 e da definição de $P^c(T)$, $(\mathbf{x}, \mathbf{y}) \in P(T)$ ou $(\mathbf{x}, \mathbf{y}) \in P^c(T)$
3. De 2, $(\mathbf{x}, \mathbf{y}) \in P(T) \cup P^c(T)$
4. De 1 e 3, $\mathbf{U} \subset P(T) \cup P^c(T)$
5. Mas, da definição de universo : $P(T) \cup P^c(T) \subset \mathbf{U}$
6. De 4 e 5 : $P(T) \cup P^c(T) = \mathbf{U}$

CQD

Teorema 2 : $P(T) \cap P^c(T) = \Phi$

Demonstração :

1. Segue direto da definição de $P^c(T)$

Teorema 3 : $(P^c(T))^c = P(T)$

Demonstração :

1. Segue direto da definição de $P^c(T)$

Teorema 4 : $\text{Front}(P(T)) = \text{Front}(P^c(T))$

Demonstração :

1. Usando a definição :

$$\text{Front}(P(T)) = \{(\mathbf{x}, \mathbf{y}) | \forall \delta > 0, B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) \neq \Phi \text{ e } B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi\}$$

$$\text{Front}(P^c(T)) = \{(\mathbf{x}, \mathbf{y}) | \forall \delta > 0, B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi \text{ e } B(\mathbf{x}, \mathbf{y}, \delta) \cap (P^c(T))^c \neq \Phi\}$$

2. Mas, $(P^c(T))^c = P(T)$
3. De 1 e 2, $\text{Front}(P(T)) = \text{Front}(P^c(T))$

CQD

Teorema 5 : $\text{Front}(P(T)) \cap \text{Int}(P(T)) = \Phi$

Demonstração :

1. Seja $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$
2. Da definição, $\exists \delta > 0$ tal que $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
3. De 2 e do fato, já demonstrado, de que $P(T) \cap P^c(T) = \Phi$, temos que $\exists \delta > 0$ tal que $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) = \Phi$
4. Logo, a negativa de 3 é falsa, ou seja, é falso que : $\forall \delta > 0 B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi$
5. De 4, concluímos que $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$
6. De 1 e 5, $\text{Front}(P(T)) \cap \text{Int}(P(T)) = \Phi$

CQD

Teorema 6 : $\text{Int}(P(T)) \subset P(T)$

Demonstração :

1. Seja $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$
2. Da definição, $\exists \delta > 0$ tal que $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
3. Mas $(\mathbf{x}, \mathbf{y}) \in B(\mathbf{x}, \mathbf{y}, \delta)$, logo, de 2, concluímos que $(\mathbf{x}, \mathbf{y}) \in P(T)$
4. De 1 e 3, $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T)) \Rightarrow (\mathbf{x}, \mathbf{y}) \in P(T)$

5. De 4, $\text{Int}(P(T)) \subset P(T)$

CQD

Teorema 7 : $\text{Front}(P(T)) \cup \text{Int}(P(T)) \cup \text{Int}(P^c(T)) = \mathbf{U}$

Demonstração :

1. Da definição $(\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$ se e só se $\forall \delta > 0, B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) \neq \Phi$ e $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi$
2. Logo $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$ se e só se $\exists \delta > 0$ tal que $B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) = \Phi$ ou $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) = \Phi$
3. Mas, $B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) = \Phi$ equívale a $B(\mathbf{x}, \mathbf{y}, \delta) \subset P^c(T)$ e $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) = \Phi$ equívale a $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
4. Logo $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$ se e só se $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$ ou $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P^c(T))$
5. Ou seja $(\text{Front}(P(T)))^c = \text{Int}(P(T)) \cup \text{Int}(P^c(T))$
6. Como, para qualquer conjunto, $A \cup A^c = \mathbf{U}$, temos :

$$\text{Front}(P(T)) \cup (\text{Front}(P(T)))^c = \text{Front}(P(T)) \cup \text{Int}(P(T)) \cup \text{Int}(P^c(T)) = \mathbf{U}$$

CQD

Teorema 8 : $\text{Int}(P(T)) = P(T) - \text{Front}(P(T))$

Demonstração :

1. Seja $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$
2. Já demonstramos que $\text{Int}(P(T)) \subset P(T)$, logo $(\mathbf{x}, \mathbf{y}) \in P(T)$
3. Já demonstramos que $\text{Front}(P(T)) \cap \text{Int}(P(T)) = \Phi$, logo $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$
4. De 2 e 3, $(\mathbf{x}, \mathbf{y}) \in P(T) - \text{Front}(P(T))$
5. De 1 e 4, $\text{Int}(P(T)) \subset P(T) - \text{Front}(P(T))$
6. Seja $(\mathbf{x}, \mathbf{y}) \in P(T) - \text{Front}(P(T))$

7. De 6, $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$, logo, da definição, é falso que $\forall \delta > 0, B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) \neq \Phi$ e $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) \neq \Phi$, portanto, $\exists \delta > 0$ tal que ou $B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) = \Phi$ ou $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) = \Phi$
8. Mas, de 6, $(\mathbf{x}, \mathbf{y}) \in P(T)$ e, da definição de bola, $(\mathbf{x}, \mathbf{y}) \in B(\mathbf{x}, \mathbf{y}, \delta)$, logo $B(\mathbf{x}, \mathbf{y}, \delta) \cap P(T) \neq \Phi$
9. De 7 e 8, podemos concluir que : $B(\mathbf{x}, \mathbf{y}, \delta) \cap P^c(T) = \Phi$
10. De 9, $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
11. Logo, $\exists \delta > 0$ tal que $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
12. De 11 e da definição de $\text{Int}(P(T))$, temos que $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$
13. De 6 e 12, $P(T) - \text{Front}(P(T)) \subset \text{Int}(P(T))$
14. De 5 e 13, $\text{Int}(P(T)) = P(T) - \text{Front}(P(T))$

CQD

Teorema 9 : $\{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} \neq \Phi$

Demonstração :

1. Da definição, $(\mathbf{x}, \mathbf{y}) \in P(T)$
2. Logo, fazendo $\theta = 1$, temos que $(\theta\mathbf{x}, \mathbf{y}) \in P(T)$
3. Logo, $1 \in \{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
4. Logo, $\{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} \neq \Phi$

CQD

Teorema 10 : $E_i(\mathbf{x}, \mathbf{y} \mid T) \in [0, 1]$

Demonstração :

1. Da definição, $E_i(\mathbf{x}, \mathbf{y} \mid T) = \min R = \min \{\theta \mid (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.
2. Já vimos que o conjunto R é não vazio, pois $\theta = 1 \in R$.
3. Logo $E_i(\mathbf{x}, \mathbf{y} \mid T) = \min R \leq 1$.

4. Mas, da definição de R , $\forall \theta \in \mathbb{R}$, $\theta \geq 0$.
5. Logo $E_i(\mathbf{x}, \mathbf{y} | T) = \min R \geq 0$.
6. De (3) e (5), temos que : $0 \leq E_i(\mathbf{x}, \mathbf{y} | T) = \min R \leq 1$

CQD

Teorema 11 : $(\mathbf{x}, \mathbf{y}) \in P(S)$ e $P(S) \subset P(T) \Rightarrow E_i(\mathbf{x}, \mathbf{y} | S) \geq E_i(\mathbf{x}, \mathbf{y} | T)$.

Demonstração :

1. Como $P(S) \subset P(T)$, temos que $\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(S), \theta \in \mathfrak{R}_+\} \subset \{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
2. Logo, $\min\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(S), \theta \in \mathfrak{R}_+\} \geq \min\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
3. Da definição, $E_i(\mathbf{x}, \mathbf{y} | S) \geq E_i(\mathbf{x}, \mathbf{y} | T)$.

CQD

Teorema 12 : $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Leftrightarrow E_i(\mathbf{x}, \mathbf{y} | T) = 1$

Demonstração :

1. Da definição, $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Leftrightarrow \mathbf{x} \in \text{In}(\mathbf{y} | T)$ e $\theta\mathbf{x} \notin \text{In}(\mathbf{y} | T) \forall \theta \in [0, 1)$
2. Isso equivale a : $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Leftrightarrow (\mathbf{x}, \mathbf{y}) \in P(T)$ e $(\theta\mathbf{x}, \mathbf{y}) \notin P(T) \forall \theta \in [0, 1)$
3. Seja o conjunto $\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
4. Logo, de 1 e 2, temos que : $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Leftrightarrow \{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} \subset [1, \infty)$
5. Mas, $1 \in \{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.
6. Logo, $\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} \subset [1, \infty) \Leftrightarrow \min\{\theta | (\theta\mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\} = 1$
7. De 4 e 6, concluímos que : $\text{Front}_f(\mathbf{y} | T) \Leftrightarrow E_i(\mathbf{x}, \mathbf{y} | T) = 1$

CQD

Teorema 13 : $E_i(\mathbf{x}, \mathbf{y} | T) = 1 \Rightarrow (\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

Demonstração :

1. Suponhamos que $E_i(\mathbf{x}, \mathbf{y} | T) = 1$
2. Suponhamos, por absurdo, que $(\mathbf{x}, \mathbf{y}) \notin \text{Front}(P(T))$.
3. De 2 e como $(\mathbf{x}, \mathbf{y}) \in P(T)$, então concluímos que $(\mathbf{x}, \mathbf{y}) \in \text{Int}(P(T))$
4. De 3, $\exists \delta > 0$, tal que $B(\mathbf{x}, \mathbf{y}, \delta) \subset P(T)$
5. Seja $\theta^0 = \max\left\{\frac{1}{2}, 1 - \frac{\delta}{2\|\mathbf{x}\|}\right\}$
6. De 5, $1/2 \leq \theta^0 < 1$
7. De 5, $0 < \frac{1}{2} \leq 1 - \theta^0 \leq \frac{\delta}{2\|\mathbf{x}\|} < \frac{\delta}{\|\mathbf{x}\|}$
8. Mas, $\text{dist}((\theta^0 \mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y})) = \|(1 - \theta^0)\mathbf{x}\| = (1 - \theta^0)\|\mathbf{x}\|$
9. De 7 e 8, temos que $\text{dist}((\theta^0 \mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y})) < \delta$
10. De 9, $(\theta^0 \mathbf{x}, \mathbf{y}) \in B(\mathbf{x}, \mathbf{y}, \delta)$
11. De 4 e 10, $(\theta^0 \mathbf{x}, \mathbf{y}) \in P(T)$
12. De 11, $\theta^0 \in \{\theta | (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
13. De 12 e da definição de $E_i(\mathbf{x}, \mathbf{y} | T)$, temos que $E_i(\mathbf{x}, \mathbf{y} | T) \leq \theta^0 < 1$
14. De 1 e 13, chegamos a um absurdo, provando que é falsa a hipótese 2, ou seja, é verdadeiro que $(\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$.
15. De 1 e 14, provamos que : $E_i(\mathbf{x}, \mathbf{y} | T) = 1 \Rightarrow (\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

CQD

Teorema 14 : $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Rightarrow (\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

Demonstração :

1. Seja $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T)$
2. Então, já demonstramos que $E_i(\mathbf{x}, \mathbf{y} | T) = 1$

3. Então, já demonstramos que $(\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$
4. Logo, de 1 e 3, $\mathbf{x} \in \text{Front}_f(\mathbf{y} | T) \Rightarrow (\mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

CQD

Observação :

O contrário não é necessariamente verdadeiro. Basta construir o contra-exemplo $(\mathbf{x}, \mathbf{0})$, onde exista \mathbf{y} tal que $(\mathbf{x}, \mathbf{y}) \in P(T)$ e supondo a tecnologia com livre disponibilidade de insumos. Como podemos verificar, $(\mathbf{x}, \mathbf{0}) \in \text{Front}(P(T))$, $(\mathbf{x}, \mathbf{0}) \in P(T)$ e $(0.5 \mathbf{x}, \mathbf{0}) \in P(T)$. Portanto $\mathbf{x} \notin \text{Front}_f(\mathbf{y} | T)$.

Teorema 15 : $\theta^* = E_i(\mathbf{x}, \mathbf{y} | T) \Rightarrow (\theta^* \mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

Demonstração :

1. Seja $\mathbf{z} = \theta^* \mathbf{x}$, então $\theta \mathbf{z} = \theta \theta^* \mathbf{x}$
2. Suponhamos, por absurdo, $E_i(\mathbf{z}, \mathbf{y} | T) < 1$, logo $\exists \theta < 1$ tal que $(\theta \mathbf{z}, \mathbf{y}) \in P(T)$
3. Mas, se $\theta < 1$, então $\theta \theta^* < \theta^*$
4. Definindo $\theta^0 = \theta \theta^*$, concluimos que $\theta^0 < \theta^*$ e que $(\theta^0 \mathbf{x}, \mathbf{y}) \in P(T)$
5. Mas, da definição de $E_i(\mathbf{x}, \mathbf{y} | T)$, $\theta^* = \min\{\theta | (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$
6. De 4 e 5, chegamos a um absurdo, ou seja $E_i(\mathbf{z}, \mathbf{y} | T) = 1$, ou seja, $E_i(\theta^* \mathbf{x}, \mathbf{y} | T) = 1$
7. Logo, como já demonstramos, $(\theta^* \mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$

CQD

Teorema 16 : $\theta^* = E_i(\mathbf{x}, \mathbf{y} | T) \Rightarrow \theta^* \in \{\theta | (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$

Demonstração :

1. Como já demonstramos, $(\theta^* \mathbf{x}, \mathbf{y}) \in \text{Front}(P(T))$.
2. De 1 e como, por convenção, $\text{Front}(P(T)) \subset P(T)$, temos que $(\theta^* \mathbf{x}, \mathbf{y}) \in P(T)$
3. De 2, temos que $\theta^* \in \{\theta | (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.

CQD

Como anteriormente demonstrado, vimos que o conjunto $\{\theta \mid (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$ é não vazio e limitado por 0 e 1. Vimos também que θ^* pertence a esse conjunto. Pela teoria dos conjuntos, todo subconjunto não vazio e limitado tem ínfimo e, se o ínfimo pertence ao subconjunto, ele é denominado mínimo. Logo, podemos garantir que é correto definirmos $E_i(\mathbf{x}, \mathbf{y} \mid T) = \min \{\theta \mid (\theta \mathbf{x}, \mathbf{y}) \in P(T), \theta \in \mathfrak{R}_+\}$.

Os teoremas seguintes são relativos a conjuntos amostra de produção. Faremos a demonstração básica para o caso AED-C e, em seguida, demonstraremos que eles valem também para os casos AED-V e AED-N.

Seja $A = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$ um conjunto amostra de produção de uma tecnologia convexa (c), com rendimentos de escala constantes (C) e livre disponibilidade de insumos (DIn) e produtos (DPr). Neste caso, temos :

$$\text{CPG}(A, \text{AED-C}) = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\mathbf{z}, \mathbf{y} \leq Y\mathbf{z}, \mathbf{z} \in \mathfrak{R}_+^K\}.$$

$$R^k = \{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, \text{AED-C}), \theta \in \mathfrak{R}_+\} = \{\theta \mid \theta \mathbf{x}^k \geq X\mathbf{z}, \mathbf{y}^k \leq Y\mathbf{z}, \mathbf{z} \in \mathfrak{R}_+^K, \theta \in \mathfrak{R}_+\}.$$

$$E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = \theta^{\text{AED-V}} = \min R^k = \min \{\theta \mid \theta \mathbf{x}^k \geq X\mathbf{z}, \mathbf{y}^k \leq Y\mathbf{z}, \mathbf{z} \in \mathfrak{R}_+^K, \theta \in \mathfrak{R}_+\}.$$

Sejam $K_A = \{1, \dots, K\}$, $K_{\text{Front}} = \{1, \dots, k_{\text{Front}}\}$ e $K_{\text{Int}} = \{k_{\text{Front}} + 1, \dots, K\}$, onde K_{Front} contenha os elementos de A tais que $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = 1$ e K_{Int} contenha os elementos de A tais que $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) < 1$. Da definição, temos que $K_{\text{Front}} \cup K_{\text{Int}} = K_A$ e $K_{\text{Front}} \cap K_{\text{Int}} = \Phi$.

$$\text{Definimos } \text{Front}(A) = \{(\mathbf{x}^k, \mathbf{y}^k) \mid (\mathbf{x}^k, \mathbf{y}^k) \in A \text{ e } k \in K_{\text{Front}}\}.$$

$$\text{Definimos } \text{Int}(A) = \{(\mathbf{x}^k, \mathbf{y}^k) \mid (\mathbf{x}^k, \mathbf{y}^k) \in A \text{ e } k \in K_{\text{Int}}\}.$$

Das definições, temos que $\text{Front}(A) \subset A$, $\text{Int}(A) \subset A$, $\text{Front}(A) \cup \text{Int}(A) = A$, $\text{Front}(A) \cap \text{Int}(A) = \Phi$, $\text{Front}(A) = A - \text{Int}(A)$, $\text{Int}(A) = A - \text{Front}(A)$.

Sejam θ^k e \mathbf{z}^k tais que $\theta^k = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$ e $\theta^k \mathbf{x}^k = X\mathbf{z}^k + \mathbf{s}^{k-}$ e $\mathbf{y}^k = Y\mathbf{z}^k - \mathbf{s}^{k+}$. Onde \mathbf{s}^{k-} e \mathbf{s}^{k+} são os vetores de excessos e faltas de $(\theta^k \mathbf{x}^k, \mathbf{y}^k)$ em relação à tecnologia e $s_i^{k-} \geq 0$ e $s_i^{k+} \geq 0$, $i = 1, 2, \dots, K$.

Teorema 17 : $\text{CPG}(A, \text{AED-C}), \theta \in \mathfrak{R}_+\} \subset P(T)$

Demonstração :

1. Segue da definição de $\text{CPG}(A, \text{AED-C})$ e do fato de que a tecnologia T possui as propriedades c, C, DIn, DPr .

CQD

Este Teorema vale também para os casos AED-V e AED-N , pois a demonstração independeu da condição de z .

Teorema 18 : $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid T)$

Demonstração :

1. Segue do Teorema anterior e das definições.

CQD

Este Teorema vale também para os casos AED-V e AED-N , pois a demonstração independeu da condição de z .

Podemos concluir que a estimativa do cálculo da eficiência de uma UD usando um conjunto amostra de produção é um limite superior ao valor da eficiência real. Na realidade, à medida que a amostra aumenta e se torna mais representativa de $P(T)$, o valor da estimativa converge assintoticamente para o valor real, supondo que os pontos da amostra possam ser considerados variáveis aleatórias identicamente e independentemente distribuídas com uma densidade estritamente positiva sobre todo o conjunto de produção (Banker 1993, Korostlev et al. 1995a, 1995b).

Teorema 19 : Se $B \subset A \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k \mid B, \text{AED-C}) \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$.

Demonstração :

1. Como $B \subset A$, temos que $\text{CPG}(B, \text{AED-C}) \subset \text{CPG}(A, \text{AED-C})$
2. Logo, $\{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(B, \text{AED-C}), \theta \in \mathfrak{R}_+\} \subset \{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, \text{AED-C}), \theta \in \mathfrak{R}_+\}$
3. Logo, $\text{mín} \{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(B, \text{AED-C}), \theta \in \mathfrak{R}_+\} \geq \text{mín} \{\theta \mid (\theta \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A, \text{AED-C}), \theta \in \mathfrak{R}_+\}$

4. Da definição, $E_i(\mathbf{x}^k, \mathbf{y}^k | B, \text{AED-C}) \geq E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C})$

CQD

Este Teorema vale também para os casos AED-V e AED-N, pois a demonstração independeu da condição de \mathbf{z} .

Teorema 20 : $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Front}(A) \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C})$.

Demonstração :

1. Seja $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Front}(A)$
2. De 1 e da definição de $\text{Front}(A)$, $E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = 1$
3. Mas, $\text{Front}(A) \subset A$. Logo, como já foi demonstrado, $E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C}) \geq E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = 1$
4. Mas, como já demonstramos, $E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C}) \in [0, 1]$
5. De 3 e 4, concluímos que $E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C}) = 1$
6. De 2 e 5, $E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C})$
7. De 1 e 6, $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Front}(A) \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k | \text{Front}(A), \text{AED-C})$.

CQD

Este Teorema vale também para os casos AED-V e AED-N, pois a demonstração independeu da condição de \mathbf{z} .

Teorema 21 : $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Int}(A) \Rightarrow E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k | A - \{(\mathbf{x}^k, \mathbf{y}^k)\}, \text{AED-C})$.

Demonstração :

1. Seja $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Int}(A)$
2. De 1, $E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C}) < 1$
3. Sejam θ^k e \mathbf{z}^k tais que $\theta^k = E_i(\mathbf{x}^k, \mathbf{y}^k | A, \text{AED-C})$ e $\theta^k \mathbf{x}^k = \mathbf{Xz}^k + \mathbf{s}^{k-}$ e $\mathbf{y}^k = \mathbf{Yz}^k - \mathbf{s}^{k+}$.

4. $\theta^k \mathbf{x}^k = \mathbf{Xz}^k + \mathbf{s}^{k-} = z^k_1 \mathbf{x}^1 + \dots + z^k_k \mathbf{x}^k + \dots + z^k_K \mathbf{x}^K + \mathbf{s}^{k-}$
5. Da definição, temos que $z^k_k \geq 0$
6. Caso $z^k_k = 0$, a demonstração é imediata, pois as parcelas em \mathbf{x}^k e \mathbf{y}^k são anuladas. Portanto, basta analisarmos a condição $z^k_k > 0$
7. De 4, temos : $(\theta^k - z^k_k) \mathbf{x}^k = z^k_1 \mathbf{x}^1 + \dots + z^k_{k-1} \mathbf{x}^{k-1} + z^k_{k+1} \mathbf{x}^{k+1} + \dots + z^k_K \mathbf{x}^K + \mathbf{s}^{k-}$
8. Das definições, todos os termos da equação acima são não negativos, logo :
 $(\theta^k - z^k_k) \geq 0$, ou seja, $\theta^k \geq z^k_k > 0$
9. Suponhamos, por absurdo, que $z^k_k = \theta^k$
10. De 9, $\theta^k - z^k_k = 0$
11. Logo, por serem não negativos, todos os valores do lado direito da equação teriam que ser nulos, ou seja $z^k_i = 0$ para todo $i \neq k$.
12. De 3 e 11, $\mathbf{y}^k = \mathbf{Yz}^k - \mathbf{s}^{k+} = z^k_k \mathbf{y}^k - \mathbf{s}^{k+}$
13. $(1 - z^k_k) \mathbf{y}^k = -\mathbf{s}^{k+}$
14. De 2, 3 e 9, temos que $z^k_k = \theta^k < 1$, logo, $1 - z^k_k > 0$
15. Portanto, o lado esquerdo da equação só possui parcelas não negativas e o lado direito apenas parcelas não positivas. A única possibilidade seria se todas as parcelas fossem nulas. Como $\mathbf{y}^k \neq \mathbf{0}$, chegamos a um absurdo.
16. De 9 e 15, concluímos que $z^k_k \neq \theta^k$
17. De 2, 3, 6, 8 e 16, temos que $1 > \theta^k > z^k_k > 0$
18. Nesse caso, $(1 - z^k_k) > 0$ e $(\theta^k - z^k_k) > 0$
19. Usando as equações definidas em 3, passando para a esquerda as parcelas relativas a $(\mathbf{x}^k, \mathbf{y}^k)$, temos :

$$(\theta^k - z^k_k) \mathbf{x}^k = z^k_1 \mathbf{x}^1 + \dots + z^k_{k-1} \mathbf{x}^{k-1} + z^k_{k+1} \mathbf{x}^{k+1} + \dots + z^k_K \mathbf{x}^K + \mathbf{s}^{k-}$$

$$(1 - z^k_k) \mathbf{y}^k = z^k_1 \mathbf{y}^1 + \dots + z^k_{k-1} \mathbf{y}^{k-1} + z^k_{k+1} \mathbf{y}^{k+1} + \dots + z^k_K \mathbf{y}^K - \mathbf{s}^{k+}$$

20. Dividindo essas equações por $(1 - z^k_k)$, temos :

$$\frac{\theta^k - z_k^k}{1 - z_k^k} \mathbf{x}^k = \frac{z_1^k}{1 - z_k^k} \mathbf{x}^1 + \dots + \frac{z_{k-1}^k}{1 - z_k^k} \mathbf{x}^{k-1} + \frac{z_{k+1}^k}{1 - z_k^k} \mathbf{x}^{k+1} + \dots + \frac{z_K^k}{1 - z_k^k} \mathbf{x}^K + s^{k-}$$

$$\mathbf{y}^k = \frac{z_1^k}{1 - z_k^k} \mathbf{y}^1 + \dots + \frac{z_{k-1}^k}{1 - z_k^k} \mathbf{y}^{k-1} + \frac{z_{k+1}^k}{1 - z_k^k} \mathbf{y}^{k+1} + \dots + \frac{z_K^k}{1 - z_k^k} \mathbf{y}^K - s^{k+}$$

Sejam $\mathbf{z}^0 = \left(\frac{z_1^k}{1 - z_k^k}, \dots, \frac{z_{k-1}^k}{1 - z_k^k}, 0, \frac{z_{k+1}^k}{1 - z_k^k}, \dots, \frac{z_K^k}{1 - z_k^k} \right)$ e $\theta^0 = \frac{\theta^k - z_k^k}{1 - z_k^k}$

21. De 2 e 3, $\theta^k < 1$, logo $(1 - \theta^k) > 0$ e de 17, $z_k^k > 0$

22. De 21, $(1 - \theta^k)z_k^k > 0 \therefore z_k^k - \theta^k z_k^k > 0$

23. Somando θ^k a ambos os lados da última inequação, temos :

$$z_k^k + \theta^k - \theta^k z_k^k > \theta^k$$

$$\theta^k - \theta^k z_k^k > \theta^k - z_k^k$$

$$\theta^k (1 - z_k^k) > \theta^k - z_k^k$$

$$\theta^k > \frac{\theta^k - z_k^k}{1 - z_k^k} = \theta^0$$

24. De 20 e 23, concluímos que existem \mathbf{z}^0 e θ^0 tais que $\theta^0 \mathbf{x}^k \geq \mathbf{Xz}^0$, $\mathbf{y}^k \leq \mathbf{Yz}^0$ e $\theta^0 < \theta^k$

25. Pela definição de θ^k , chegamos a um absurdo, pois $\theta^k = \min \{ \theta \mid \theta \mathbf{x}^k \geq \mathbf{Xz}, \mathbf{y}^k \leq \mathbf{Yz}, \mathbf{z} \in \mathfrak{R}_+^K, \theta \in \mathfrak{R}_+ \}$

26. De 25, podemos concluir que a hipótese de $z_k^k > 0$ não é possível, restando apenas a solução $z_k^k = 0$. Mas isso significa que podemos calcular $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$ sem considerar o ponto $(\mathbf{x}^k, \mathbf{y}^k)$, ou seja, $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^k, \mathbf{y}^k)\}, \text{AED-C})$

CQD

$$\text{Mas } \mathbf{e}^t \mathbf{z}^0 = \frac{\sum_{i \neq k} z_i^k}{1 - z_k^k} = \frac{\mathbf{e}^t \mathbf{z}^k - z_k^k}{1 - z_k^k}$$

Pela equação acima, podemos concluir que o Teorema vale também para o caso AED-V, pois $\mathbf{e}^t \mathbf{z}^k = 1$ implica que $\mathbf{e}^t \mathbf{z}^0 = 1$.

Pela equação acima, podemos concluir que o Teorema vale também para o caso AED-N, pois $e^t z^k \leq 1$ implica que $e^t z^0 \leq 1$.

Teorema 22 : $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid \text{Front}(A), \text{AED-C})$.

Demonstração :

1. A condição acima já foi demonstrada no caso de $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Front}(A)$
2. De 1, podemos analisar apenas o caso em que $(\mathbf{x}^k, \mathbf{y}^k) \in \text{Int}(A)$
3. Seja $(\mathbf{x}^k, \mathbf{y}^k) \in A$, $\theta^k = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$ e \mathbf{z}^k tais que :

$$\theta^k \mathbf{x}^k = \mathbf{Xz}^k + s^{k-} = z_1^k \mathbf{x}^1 + \dots + z_K^k \mathbf{x}^K + s^{k-}$$

$$\mathbf{y}^k = \mathbf{Yz}^k - s^{k+} = z_1^k \mathbf{y}^1 + \dots + z_K^k \mathbf{y}^K - s^{k+}$$

4. De 2 e da definição de θ^k em 3, temos que $\theta^k < 1$
5. Suponhamos que exista $j \in K_{\text{Int}}$ tal que $z_j^k \neq 0$
6. De 5, existem $\theta^j < 1$ e \mathbf{z}^j tais que :

$$\theta^j \mathbf{x}^j = \mathbf{Xz}^j + s^{j-} = z_1^j \mathbf{x}^1 + \dots + z_K^j \mathbf{x}^K + s^{j-}$$

$$\mathbf{y}^j = \mathbf{Yz}^j - s^{j+} = z_1^j \mathbf{y}^1 + \dots + z_K^j \mathbf{y}^K - s^{j+}$$

7. Como $(\mathbf{x}^j, \mathbf{y}^j) \in \text{Int}(A)$, já demonstramos que $z_j^j = 0$, logo :

$$\theta^j \mathbf{x}^j = z_1^j \mathbf{x}^1 + \dots + z_{j-1}^j \mathbf{x}^{j-1} + z_{j+1}^j \mathbf{x}^{j+1} + \dots + z_K^j \mathbf{x}^K + s^{j-}$$

$$\mathbf{y}^j = z_1^j \mathbf{y}^1 + \dots + z_{j-1}^j \mathbf{y}^{j-1} + z_{j+1}^j \mathbf{y}^{j+1} + \dots + z_K^j \mathbf{y}^K - s^{j+}$$

8. Substituindo as equações 7 nas equações 3, temos :

$$\begin{aligned} \theta^k \mathbf{x}^k &= \left(z_1^k + z_1^j \frac{z_j^k}{\theta^j} \right) \mathbf{x}^1 + \dots + \left(z_{j-1}^k + z_{j-1}^j \frac{z_j^k}{\theta^j} \right) \mathbf{x}^{j-1} + \\ &+ \left(z_{j+1}^k + z_{j+1}^j \frac{z_j^k}{\theta^j} \right) \mathbf{x}^{j+1} + \dots + \left(z_K^k + z_K^j \frac{z_j^k}{\theta^j} \right) \mathbf{x}^K + s^{k-} + \frac{z_1^k}{\theta^j} s^{j-} \end{aligned}$$

$$\begin{aligned} \mathbf{y}^k &= \left(z_1^k + z_1^j \frac{z_j^k}{\theta^j} \right) \mathbf{y}^1 + \dots + \left(z_{j-1}^k + z_{j-1}^j \frac{z_j^k}{\theta^j} \right) \mathbf{y}^{j-1} + \\ &+ \left(z_{j+1}^k + z_{j+1}^j \frac{z_j^k}{\theta^j} \right) \mathbf{y}^{j+1} + \dots + \left(z_K^k + z_K^j \frac{z_j^k}{\theta^j} \right) \mathbf{y}^K - s^{k+} - z_1^k s^{j+} \end{aligned}$$

9. Mas, $\theta^j < 1$, logo :

$$z_i^k + z_i^j \frac{z_j^k}{\theta^j} > z_i^k + z_i^j z_j^k$$

$$\text{Chamando } w_i = z_i^k + z_i^j z_j^k \quad \text{e} \quad w_i + \delta_i = z_i^k + z_i^j \frac{z_j^k}{\theta^j}$$

Temos que : $w_i \geq 0$ e $\delta_i \geq 0$

10. Substituindo 9 em 8, temos :

$$\begin{aligned} \theta^k \mathbf{x}^k &= (w_1 + \delta_1) \mathbf{x}^1 + \dots + (w_{j-1} + \delta_{j-1}) \mathbf{x}^{j-1} + (w_{j+1} + \delta_{j+1}) \mathbf{x}^{j+1} \\ &+ \dots + (w_K + \delta_K) \mathbf{x}^K + s^{k-} + \frac{z_1^k}{\theta^j} s^{j-} \end{aligned}$$

$$\begin{aligned} \theta^k \mathbf{x}^k &= w_1 \mathbf{x}^1 + \dots + w_{j-1} \mathbf{x}^{j-1} + w_{j+1} \mathbf{x}^{j+1} + \dots + w_K \mathbf{x}^K + \\ &+ \delta_1 \mathbf{x}^1 + \dots + \delta_{j-1} \mathbf{x}^{j-1} + \delta_{j+1} \mathbf{x}^{j+1} + \dots + \delta_K \mathbf{x}^K + s^{k-} + \frac{z_1^k}{\theta^j} s^{j-} \end{aligned}$$

$$\mathbf{y}^k = w_1 \mathbf{y}^1 + \dots + w_{j-1} \mathbf{y}^{j-1} + w_{j+1} \mathbf{y}^{j+1} + \dots + w_K \mathbf{y}^K - s^{k+} - z_1^k s^{j+}$$

11. Chamando $\mathbf{w} = (w_1, \dots, w_{j-1}, 0, w_{j+1}, \dots, w_K)$, concluímos que :

$$\theta^k \mathbf{x}^k \geq X \mathbf{w}$$

$$\mathbf{y}^k \leq Y \mathbf{w}$$

$$\mathbf{w} \geq 0$$

12. Como $w_j = 0$, concluímos que $(\theta^k \mathbf{x}^k, \mathbf{y}^k) \in \text{CPG}(A - \{(\mathbf{x}^j, \mathbf{y}^j)\}, \text{AED-C})$.

Logo, da definição de $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^j, \mathbf{y}^j)\}, \text{AED-C})$, temos que $\theta^k \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^j, \mathbf{y}^j)\}, \text{AED-C})$.

13. Mas, $A - \{(\mathbf{x}^j, \mathbf{y}^j)\} \subset A$. Logo, como já demonstramos, $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^j, \mathbf{y}^j)\}, \text{AED-C}) \geq E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$.

14. Mas, conforme definimos, $\theta^k = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C})$. Portanto, de 12 e 13, concluímos que $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \{(\mathbf{x}^j, \mathbf{y}^j)\}, \text{AED-C})$.

15. Repetindo as operações de 5 a 14 para todos os pontos de $\text{Int}(A)$, podemos concluir que $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid A - \text{Int}(A), \text{AED-C})$

16. Mas, já vimos que $\text{Front}(A) = A - \text{Int}(A)$. Logo : $E_i(\mathbf{x}^k, \mathbf{y}^k \mid A, \text{AED-C}) = E_i(\mathbf{x}^k, \mathbf{y}^k \mid \text{Front}(A), \text{AED-C})$

CQD

$$\text{Mas } \mathbf{e}^t \mathbf{w} = \sum_{i \neq j} z_i^k + z_j^k \sum_{i \neq j} z_i^j = \mathbf{e}^t \mathbf{z}^k - z_j^k + z_j^k (\mathbf{e}^t \mathbf{z}^j - z_j^j)$$

$$\text{Mas } z_j^j = 0, \text{ assim } \mathbf{e}^t \mathbf{w} = \mathbf{e}^t \mathbf{z}^k + z_j^k (\mathbf{e}^t \mathbf{z}^j - 1)$$

Pela equação acima, podemos concluir que o Teorema vale também para o caso AED-V, pois $\mathbf{e}^t \mathbf{z}^k = 1$ e $\mathbf{e}^t \mathbf{z}^j = 1$ implicam que $\mathbf{e}^t \mathbf{w} = 1$.

Pela equação acima, podemos concluir que o Teorema vale também para o caso AED-N, pois $\mathbf{e}^t \mathbf{z}^k \leq 1$ e $\mathbf{e}^t \mathbf{z}^j \leq 1$ implicam que $\mathbf{e}^t \mathbf{w} \leq 1$.

Este Teorema é importante porquê, nos casos em que estivermos trabalhando com amostras contendo grandes quantidades de UD's, ele garante que podemos usar apenas as UD's da fronteira para efetuarmos os cálculos das estimativas de eficiência das demais UD's.

BIBLIOGRAFIA

ADLER, N.; FRIEDMAN, L.; SINUANY-STERN, Z. Review of ranking methods in the data envelopment analysis context. **European Journal of Operational Research**, 140, p. 249-265, 2002.

BANKER, R.D. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. **Management Science**, 39(10), p.1265-1273, 1993.

BANKER, R.D.; CHANG, H. Evaluating the super-efficiency procedure in data envelopment analysis for outlier identification and for ranking efficient units. **School of Management, J043**. The University of Texas at Dallas. Richardson, Texas, 75083-0688, 19 p., nov. 2000.

BANKER, R.D.; CHARNES, A.; COOPER, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. **Management Science**, 30(9), p.1078-1092, 1984.

BANKER, R.D.; DAS, S.; DATAR, S.M. Analysis of cost variances for management control in hospitals. **Research in governmental and nonprofit accounting**, 5, p. 269-291, 1989.

BANKER, R.D.; GIFFORD, J.L. **A relative efficiency model for the evaluation of public health nurse productivity**. 1988.

CHARNESS, A.; COOPER, W.W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**, 2, p.429-444, 1978.

CHERCHYE, L.; POST, T. **Methodological advances in AED: A survey and an application for the Dutch electricity sector**.

COOPER, W.W.; LAWRENCE, M.S.; TONE, K. **Data envelopment analysis: A comprehensive text with models, applications, references and AED-solver software**. Norwell, Massachusetts: Kluwer Academic Publishers, 318 p., 2002.

DEPRINS, D.; SIMAR, L.; TULKENS, H. Measuring labor efficiency in post offices. **The performance of public enterprises: Concepts and measurement**, Amsterdam, p.243-267, 1984.

EFFRON, B. Bootstrap methods: Another look at the jackknife, **Annals of Statistics**, 7, p. 1-26, 1979.

EFFRON, B.; GONG, G. A leisurely look at the bootstrap, the jackknife, and cross validation. **American Statistician**, 37, p. 36-48, 1983.

FARRELL, M.J. The measurement of productive efficiency. **Journal of Royal Statistical Society**, Series A 120, p.253-290, 1957,.

GIJBELS, I. E. *et al.* On estimation of monotone and concave frontier. **Journal of the American Statistical Association**, 94, p. 220-228, 1999.

KITTELSEN, S.A.C. Monte Carlo simulations of AED efficiency measures and hypothesis test. In: 5TH EUROPEAN WORKSHOP ON EFFICIENCY AND PRODUCTIVITY MEASUREMENT, 1997, Copenhagen.

KOROSTLEV, A.; SIMAR, L.; TSYBAKOV, A.B. Efficient estimation of monotone boundaries. **The Annals of Statistics**, 23, p.476-489, 1995.

KOROSTLEV, A.; SIMAR, L.; TSYBAKOV, A.B. On estimation of monotone and convex boundaries. **Pub. Inst. Stat. Univ. Paris**, XXXIX, 1, p.3-18, 1995.

KUOSMANEN, T.; POST, T. Robust efficiency measurement. **Rotterdam Institute for Business Economic Studies (RIBES)**, Rotterdam, Report 9911, 11 p., mar. 1999.

KUOSMANEN, T.; POST, T.; SCHOLTES, S. **Testing for production efficiency with errors-in-variables**: With an application to the Dutch electricity sector.

PASTOR, J.T.; RUIZ, J.L.; SIRVENT, I. A statistical test for nested radial AED models. Working Paper, Departamento de Estadística e Investigación Operativa. Universidad de Alicante, Alicante, España, 03071, 1995.

SAMPAIO DE SOUSA, M.C.; RAMOS DE SOUZA, F. Eficiência Técnica e Retornos de Escala na Produção de Serviços Públicos Municipais. **Revista Brasileira de Economia**, 53, p. 433-461, 1999.

SAMPAIO DE SOUSA, M.C.; RAMOS DE SOUZA, F. Measuring Public Spending Efficiency in Brazilian Municipalities: A Non Parametric Approach. **Data Envelopment Analysis in the Service Sector**. Wiesbaden: Westermann, G., Ed., 1999. p. 237-267.

SAMPAIO DE SOUSA, M.C.; STOŠIĆ, B. **Technical Efficiency of the Brazilian Municipalities**: Correcting Non-Parametric Frontier Measurements For Outliers. mimeo, 2003.

TULKENS, H. The Measurement of Productive Efficiency by FDH Frontiers. Document de Travail, CORE, Université Catholique de Louvain , Louvain-la-Neuve, 1990

VANDEN EECKAUT; TULKENS, H; JAMAR, M.A. A Study of Cost-Efficiency and Returns of Scale for 235 Municipalities in Belgium. Core Discussion Paper n° 9158, CORE, Université Catholique de Louvain, 1991.

WILSON, P.W. Detecting influential observations in Data Envelopment Analysis, *Journal of productivity analysis*, 6, p. 27-46, 1995.