



Universidade de Brasília - UnB
Faculdade de Economia, Administração,
Contabilidade e Ciência da Informação e
Documentação – FACE
Departamento de Economia

“Eficiência Bancária: Uma abordagem não paramétrica
aplicada ao Banco do Brasil”

João Carlos Félix Souza

Brasília – DF
Dezembro de 2006

Agradecimentos

Às professoras Maria Eduarda Tannuri Piantó e Maria Conceição Sampaio, pelas valiosas sugestões e orientações, auxiliando no aprimoramento deste trabalho.

Aos professores do Departamento de Economia pelos ensinamentos e contribuições durante o decorrer do curso.

Aos funcionários deste Departamento pelo apoio e colaboração.

Aos colegas que sempre me auxiliaram nos estudos durante o curso e nas opiniões para o desenvolvimento do trabalho.

À minha família: Rosângela, João Gabriel, Carolina, Maria do Carmo e Spike pelo carinho e compreensão.

Ao BANCO DO BRASIL S.A. pelo patrocínio e oportunidade. Empresa que com esta atitude contribui para o engrandecimento da pesquisa no Brasil.

“Eficiência Bancária: Uma abordagem não paramétrica aplicada ao Banco do Brasil”

João Carlos Félix Souza

Tese de Doutorado apresentada em 15/12/2006 ao Departamento de Economia da UNIVERSIDADE DE BRASÍLIA, como parte dos requisitos para obtenção do grau de Doutor em Economia.

Aprovada por

**Maria Eduarda Tannuri-Piantó
Presidente da banca - Orientadora**

**Maria da Conceição Sampaio de Sousa
Membro da banca - Coorientadora**

**Walter Belluzzo Jr.
Membro da banca**

**Carlos Eduardo Gasparini
Membro da banca**

**Joanílio Teixeira
Membro da banca**

RESUMO

Este trabalho baseia-se na comparação empírica da eficiência bancária, calculados por modelos não paramétricos, utilizando a distribuição dessa variável e pesquisando seus *outliers*. Não necessariamente os *outliers* são decorrentes de problemas de mensuração, podem ser analisados como mudança de atuação estratégica, o que efetivamente ocorre na aplicação no Banco do Brasil. Os resultados do cálculo de eficiência são submetidos ao modelo de Regressão Quantílica como variável endógena. As variáveis dependentes são formadas por indicadores do potencial de mercado do município, bem como alguns indicadores internos da Instituição. A análise desses resultados confirma a importância e impacto de certas variáveis de mercado ou *dummies* de localização na medição do grau de eficiência relativa. Dois modelos são testados na técnica DEA (*Data Envelopment Analysis*) optando-se por um modelo mais apropriado para a aplicação em questão.

Palavras chave: DEA, eficiência, *outliers*, *leverage*

ABSTRACT

This project is based on the empirical comparison of banking efficiency calculated by nonparametric models, using the distribution of this variable e researching its outliers. The outliers do not necessarily result from problems of measuring, but they can be analyzed as strategic change, something which actually occurs when it applies to Banco do Brasil. Efficiency calculation results are submitted to the Quantile Regression models as an endogenous variable. The dependent variables are formed by inside indicators of the institution and municipalities. The analysis of these results confirms the importance and the impact of certain market variables or localization dummies on the measuring of relative efficiency rates. Two models are tested in the DEA (Data Envelopment Analysis) technique, another more appropriate model being chosen to apply in the case.

Key words : DEA, efficiency, outliers, leverage

SUMÁRIO

Resumo	vi
Abstract.....	vii
INTRODUÇÃO.....	x
CAPÍTULO 1 – Eficiência Bancária.....	1
CAPÍTULO 2 – Mensurações Não Paramétricas.....	10
2.1 – DEA e FDH – Características e Limitações.....	12
2.1.1 - Modelo dos Multiplicadores - Retorno Constante de Escala – CCR ou CRS ..	14
2.1.2 - Modelo do Envelope - Retorno Variável de Escala – BCC ou VRS	16
2.2 – O problema dos <i>Outliers</i> : Diferentes abordagens	20
2.3 – A abordagem “ <i>Jackstrap</i> ”	22
CAPÍTULO 3 – Índices de eficiência não paramétricos - Estudo Empírico e aplicação de DEA ao Banco do Brasil	26
3.1 – Dados e Estatísticas Descritivas	27
3.2 - Resultados dos valores de θ :	30
3.3 - Identificação dos <i>outliers</i> e análise das distribuições para os modelos CRS e VRS: .	37
3.4 - Avaliação das diferenças dos modelos - Aplicação de testes não paramétricos:	47
3.5 – <i>Clusterização</i> ou Agrupamento das Agências.....	51
3.5.1 - Agrupamento pelo critério de Segmentação do Banco:	51
3.5.2 - Clusterização pelo critério de Mercado:	62
CAPÍTULO 4 – Análise do impacto do mercado no grau de eficiência relativa: Uma aplicação de Regressão Quantílica	67
4.1 - Modelo de Regressão Quantílica.....	68
4.2 - Teoria Assintótica para Regressão Quantílica	69
4.3 - Dados e Resultados	70
4.3.1 - Regressão Quantílica – CRS com <i>outliers</i> - PF	73
4.3.2 - Regressão Quantílica – CRS sem <i>outliers</i> - PF.....	77
4.3.3 - Regressão Quantílica - VRS com <i>outliers</i> - PF.....	81
4.3.4 - Regressão Quantílica - VRS sem <i>outliers</i> - PF	85
CAPÍTULO 5 – Conclusões e Sugestões	92
Apêndice 1 - Testes não paramétricos.....	97
Apêndice 2 - Tabela de <i>outliers</i> Pessoa Física	100
Apêndice 3 - Testes não paramétricos dos <i>clusters</i> e segmentos	103
Apêndice 4 - Gráficos de eficiência x variáveis independentes	106
Apêndice 5 - Tabelas de <i>output</i> das variáveis independentes da regressão quantílica.....	110
Apêndice 6 – <i>Output</i> ANOVA e Coeficientes do modelo MQO	122
Apêndice 7 – Modelo DEA	126
Referências Bibliográficas.....	128
Figuras	
Figura 1.1: Fluxo de Eficiência	9
Figura 2.1: Fronteira de Eficiência.....	16
Figura 2.2: LG Fronteira de Produção.....	17
Figura 2.3: Fronteira VRS	20
Figura 3.1: Histograma Geral – CRS 5 <i>outputs</i>	38
Figura 3.2: Distribuição de θ – CRS 5 <i>outputs</i>	39

Figura 3.3: Distribuição de θ – CRS 4 <i>outputs</i>	41
Figura 3.4: Distribuição de θ – VRS	44
Figura 4.0: Distribuição de $\rho_d(u)$	69
Figuras 4.1: PF com 4 <i>outputs</i> e <i>outliers</i> - CRS: β do τ quantílico X θ	75
Figuras 4.1: continuação.....	76
Figuras 4.2: PF com 4 <i>outputs</i> sem <i>outliers</i> - CRS: β do τ quantílico X θ	79
Figuras 4.2: continuação.....	80
Figuras 4.3: PF com 4 <i>outputs</i> e <i>outliers</i> - VRS: β do τ quantílico X θ	83
Figuras 4.3: continuação.....	84
Figuras 4.4: PF com 4 <i>outputs</i> sem <i>outliers</i> - VRS: β do τ quantílico X θ	87
Figuras 4.4: continuação.....	88
Tabelas	
Tabela 3.1: Variáveis de <i>Inputs</i> e <i>Outputs</i> das 3663 agências bancárias	28
Tabela 3.2: Estatística Descritiva	29
Tabela 3.3: Correlação entre as variáveis.....	30
Tabela 3.4: Eficiência e <i>Leverage</i> - modelo CRS - 30 agências	32
Tabela 3.5: Dominância por FDH	34
Tabela 3.6: Eficiência e <i>Leverage</i> - modelo VRS - 30 agências	36
Tabela 3.7: Análise Descritiva de θ – 5 <i>outputs</i> - CRS	38
Tabela 3.8: Análise Descritiva de θ – 4 <i>outputs</i> - CRS	40
Tabela 3.9: Análise Descritiva de θ – 5 <i>outputs</i> – CRS e VRS.....	43
Tabela 3.10: Geral: DEA e <i>Heaviside step</i> – quantidade de clientes	46
Tabela 3.11: Geral: DEA e <i>Heaviside step</i> – classes de renda de clientes	47
Tabela 3.12: Análise descritiva dos valores de $\delta^j = \theta_5^j - \theta_4^j$	48
Tabela 3.13: Análise Descritiva de θ – segmento BB	52
Tabela 3.14: Eficiência e <i>Leverage</i> – segmento Governo	54
Tabela 3.15: Eficiência e <i>Leverage</i> – segmento PJ	55
Tabela 3.16: Análise Descritiva de θ – 5 <i>outputs</i> – segmento PF	56
Tabela 3.17: Eficiência e <i>Leverage</i> – segmento PF.....	57
Tabela 3.18: PF: DEA e <i>Heaviside step</i> – quantidade de clientes.....	59
Tabela 3.19: PF: DEA e <i>Heaviside step</i> – classes de renda de clientes	60
Tabela 3.20: PF: DEA e <i>Heaviside step</i> – agências concorrentes	61
Tabela 3.21: Comparação grupos <i>Clusters</i>	64
Tabela 4.1: Eficiência PF – 4 <i>outputs</i> – CRS com <i>outliers</i>	73
Tabela 4.2: Eficiência PF – 4 <i>outputs</i> – CRS sem <i>outliers</i>	77
Tabela 4.3: Eficiência PF – 4 <i>outputs</i> – VRS com <i>outliers</i>	81
Tabela 4.4: Eficiência PF – 4 <i>outputs</i> – VRS sem <i>outliers</i>	86

INTRODUÇÃO

Muitos pesquisadores, em livros ou artigos especializados, discutem o papel das empresas. Para determinadas pessoas, o objetivo principal e único das empresas é o lucro. Todavia as organizações passam por uma lógica de transformar fatores de produção utilizados no seu processo em produtos/serviços que determinados clientes desejam receber. O resultado financeiro não deve ser visto somente como a única meta, pois, se isto fosse verdadeiro, como seriam analisadas as empresas sem fins lucrativos e que a cada dia são de vital importância no cotidiano das famílias?

Embora não sendo um conceito econômico para a Teoria Sistêmica a empresa pode ser descrita como um conjunto de três elementos interdependentes: Entrada, Processamento e Saída. O que a caracteriza como sistema aberto e dinâmico, sem necessariamente se limitar ao lucro. Atualmente, as empresas começam a agir com responsabilidade social. Este fato é facilmente confirmado através dos balanços sociais corporativos que passaram a ser divulgados por diversas organizações. Portanto, existem diversas formas de se medir o desempenho através de índices, quocientes e outras formas quantitativas e que não visam exclusivamente o resultado financeiro. A teoria define muitas expressões relacionadas ao desempenho, tais como: eficiência, eficácia e produtividade. Pode-se conceituar produtividade como a relação entre entradas e saídas (insumos e produtos). Já a eficiência difere do conceito de produtividade em função do fato de expressar uma relação ótima entre recursos produzidos e insumos utilizados. A eficácia está relacionada ao cumprimento dos objetivos traçados (Silva e Azevedo, 2004).

A conceituação de modelo é essencial em qualquer estudo quantitativo que se possa realizar no campo da economia e finanças. Ela expressa uma visão simplificada da realidade. Através de funções matemáticas e distribuições probabilísticas este trabalho tenta explicar o comportamento de sistemas de produção, particularmente com uma aplicação no Banco do Brasil. Berger e Humphrey (1997) pesquisaram modelos paramétricos e não paramétricos para a medição de eficiência. Um dos problemas das técnicas paramétricas é a necessidade de se conhecer a função de produção, bem como a distribuição da estatística dos desvios (resíduos) entre os valores esperados e observados. Por outro lado a técnica

não paramétrica não necessita de suposição no que diz respeito ao formato da fronteira de produção.

A grande popularidade das medidas técnicas de eficiência não paramétricas, tais como DEA (Data Envelopment Analysis), vem de sua flexibilidade operacional. A exceção se faz pela propriedade de suposição de convexidade, no caso DEA original, e a exigência de correta definição dos *outputs* e *inputs*.

O conceito básico trata de fronteira de produção ou função de fronteira de produção que por definição é a máxima quantidade de *outputs* (ou produtos) que pode ser obtida dado uma série de *inputs* (insumos) ou recursos utilizados. Conforme definição de Samuelson e Nordhaus (2004): “A fronteira de possibilidade de produção mostra as quantidades máximas de produção que podem ser obtidas por uma economia ou firma, dado o seu conhecimento tecnológico e a quantidade de insumos disponíveis.”

Neste contexto são definidas DMU - *Decision Making Unit* - como uma firma, um departamento, uma divisão administrativa e, no caso da aplicação empírica deste trabalho, uma filial ou agência bancária cuja eficiência esta sendo avaliada. O conjunto de DMU's adotado para análise DEA deve ter em comum a utilização dos mesmos *inputs* e *outputs*, bem como serem homogêneos e ter autonomia na tomada de decisão. As unidades de medidas das variáveis iguais devem ser as mesmas, no entanto podem ser diferentes entre as demais.

Von Neumann empregou modelos de programação matemática para construir ou representar uma tecnologia na sua análise de atividades e modelos de crescimento, bem como Leontief fez uso de análise de atividades e programação linear, utilizando modelos lineares de “*input-output*” dedicados à construção do equilíbrio geral interno a um sistema produtivo (Lins, 2000). Diferentemente a modelagem DEA está mais apropriada para o desenvolvimento dos modelos de programação de produção microeconômicos, portanto mais adequada a esse trabalho. Os modelos de programação matemática provêm uma maneira elegante de, simultaneamente, construir a fronteira para uma dada tecnologia a partir do conjunto de observações e calcular a distância da fronteira a cada uma das observações individuais.

Para Seiford e Thrall (1990) os enfoques e interesses em DEA são diversificados: os estatísticos consideram esta técnica como exercício em análise exploratória de dados; os

econometristas como uma técnica que estima uma função de produção empírica; os matemáticos como uma metodologia para determinar soluções não dominadas em problema multicritério¹.

Importantes contribuições surgiram, no sentido de capacitar o modelo DEA a lidar com situações reais, com a existência de variáveis exógenas e categóricas e com a incorporação do conhecimento de especialistas e usuários. Com a interação da economia e estatística surgiram métodos que incluem DEA, ou baseados em DEA, para determinar eficiências: método para avaliação de desempenho, para seleção de unidades eficientes, para determinação de preços, para avaliação de risco, para classificação de centros universitários e de pesquisa e outras aplicações (Deprins et al., 1984; Souza et al., 1997; Wheelock e Wilson, 2003; Rosa e Mazzon, 2003; Silva e Azevedo 2004, só para citar alguns trabalhos e autores). Embora a aplicação seja relativamente recente, tem tido um rápido desenvolvimento. Assim, atualmente, ela conta com uma ampla base teórica e variedade de aplicações práticas também em economia e no sistema financeiro em eficiência bancária (Banker, Charnes e Cooper, 1984; Berger e Humphrey, 1991 e 1997; Golany e Storbeck, 1999; Kantor e Maital, 1999; Thanassoulis, 1999 e Zenios et al., 1999).

Banker Charnes e Cooper (1984) destacam-se entre os estudiosos mais proeminentes na literatura a respeito de mensuração da eficiência de instituições financeiras. A eles se seguiram diversos outros autores e trabalhos sobre o assunto, focando em eficiência bancária. Posteriormente, Berger e Humphrey (1997) consolidaram e avaliaram diversos estudos sobre eficiência técnica para instituições financeiras.

Para modelos de produção determinísticos mostra-se que os estimadores DEA são de máxima verossimilhança, não-paramétricos, consistentes e geram ineficiência técnicas que convergem em distribuição. Essa distribuição assintótica conjunta permite testes de hipóteses estatísticas. Esses resultados foram originalmente obtidos por Banker (1993) e, posteriormente, com outros autores foram refinados e generalizados. No Brasil Souza (2002) fez extenso trabalho sobre isso. Os modelos de produção com base na análise envoltória de dados - DEA – tem sido objeto de uso intenso na literatura econométrica. No entanto, essa abordagem tem sido criticada por sua natureza essencialmente determinística,

¹ Algumas aplicações interessantes em Análise Multicritério podem ser vistas em Belton e Vickers (1993) e Barba-Romero e Pomerol (1997).

mesmo porque os modelos de produção que conduzem as medidas de eficiência DEA não levam em conta a influência de erros de medida e ruídos na especificação da fronteira, consideram-se que os desvios da fronteira resultam de ineficiências técnicas.

Souza (2002) demonstrou que os estimadores DEA de funções de produção têm a propriedade de máxima verossimilhança, são consistentes e convergem em distribuição sob hipóteses que não excluem a possibilidade de heterocedasticidade². O comportamento dos resíduos corrobora no sentido que os componentes estocásticos no modelo de produção são resultantes de ineficiências técnicas. Os resultados são assintóticos e permitem utilização de modelagens flexíveis.

Normalmente DEA apresenta nos resultados, principalmente em situações onde as unidades de decisão são bem comportadas e as variáveis de *input* e *output* são equilibradas, variáveis sem grandes dispersões. No entanto, se alguma ou algumas unidades têm um ótimo desempenho há alteração no resultado das demais unidades indicando que as mesmas têm baixa eficiência. A distribuição de frequência da eficiência torna-se altamente assimétrica e com escala não linear. Muito se tem trabalhado para dirimir este efeito. Entretanto, quase sempre isto depende de uma inspeção visual dos dados o que é virtualmente impossível para grandes bases ou conjunto de dados (Sousa e Stosic, 2005). A essência da metodologia de Sousa e Stosic (2005) é reduzir estocasticamente o impacto das DMU's mais influentes quanto aos escores finais de eficiência usando o conceito de *leverage* ou poder de influência da DMU (Cribari e Zarkos, 2004), que consiste em apurar o efeito que uma determinada DMU produz no escore final de eficiência. Este efeito é medido removendo esta DMU do conjunto de dados e recalculando a eficiência das demais.

Posteriormente, para analisar variáveis determinantes dos *scores* de eficiência, propõe-se utilização de métodos econométricos, tais como modelos de Mínimos Quadrados Ordinários (MQO), Tobit, Logit ou, mais recentemente, utilização de Regressão Quantílica, a fim de estabelecer o comportamento da medida de eficiência como variável resposta submetida a variáveis exógenas definidas por indicadores de potencial de mercado ou de gestão interna.

² O que reforça a aplicação, mais adiante, da Regressão Quantílica para modelar e ajustar a variável endógena de eficiência com demais variáveis de mercado.

A técnica de regressão quantílica foi introduzida por Koenker e Basset (1978) e pode ser entendida como a extensão dos quantis ordinários em um modelo de locação para uma classe mais geral de modelos lineares em que os quantis condicionais têm forma linear. Eles definem o τ -ésimo quantil de regressão como a solução de uma função objetivo num modelo de programação linear.

A regressão quantílica apresenta características que a tornam extremamente interessante, pois os modelos podem ser usados para caracterizar toda a distribuição condicional de uma variável resposta dado um conjunto de regressores (Buchinsky, 1997). O modelo permite estimar os parâmetros na forma de programação linear; a função objetivo é a soma ponderada de desvios absolutos, portanto, fornece uma medida de locação robusta de modo que os valores estimados não são sensíveis às observações extremas na variável dependente; quando não existe distribuição normal nos erros os estimadores da regressão quantílica podem ser mais eficientes que os estimadores de mínimos quadrados. Por fim, soluções diferentes para quantis distintos podem ser interpretadas como diferenças na resposta da variável dependente às mudanças nos regressores em vários pontos da distribuição condicional da variável dependente. Neste particular existem diversos exemplos de pesquisas empíricas que a utilização da técnica de regressão quantílica é ajustada.

Diante disso, os objetivos desta pesquisa são:

- Aplicação do DEA em um grande banco nacional de tal forma a avaliar melhor modelo de eficiência relativa que permita comparar as agências bancárias.
- Escolha e aplicação do melhor modelo DEA (utilizando como opção Retorno Constante de Escala – CRS ou Retorno Variável de Escala – VRS) para o caso em questão.
- Identificar e analisar possíveis diferenças entre grupos de agências para que se possa estabelecer *ranking*.
- Identificar o impacto de determinados *outputs* na composição do modelo.
- Utilizar distribuições para detectar *outliers* sugerindo modelos teóricos para seleção e possível mudança de atuação estratégica da instituição estudada,

tanto na distribuição de segmentos de unidades como em unidades individuais de destaque.

- Detectar efeitos do ambiente (mercado ou variáveis internas de gestão) nas medidas de eficiência, estabelecendo modelo adequado para esse fim e identificando variáveis vinculadas a estimação de parâmetros, possivelmente voláteis condicionalmente, que influenciam a eficiência das unidades em estudo.

Além da introdução, esse trabalho se completa com mais cinco capítulos. O capítulo um trata dos conceitos de eficiência bancária e faz uma revisão bibliográfica sobre o que atualmente está sendo feito sobre este tema no Brasil e no exterior. O capítulo dois apresenta o modelo DEA identificando as equações de restrição e função objetivo definidas para cálculo dos “pesos” que geram estimativas de eficiência relativa. Este capítulo mostra, também, as técnicas de utilização das distribuições linear, exponencial, inversa e *heaviside step*, para identificar os *outliers* ou as melhores práticas nas DMU’s de maior impacto. O capítulo três analisa o resultado de uma aplicação no Banco do Brasil, com uma amostra inicial de 3.663 agências com informações do ano de 2004, avaliando os valores de eficiência das agências, bem como o motivo de algumas delas serem responsáveis pelo impacto e posicionamento da eficiência relativa. Neste capítulo esses resultados são avaliados considerando grupos mais homogêneos de comparação destas agências. Primeiramente utilizando grupos de atuação estratégica para o Banco, posteriormente redefinindo os grupos por uma metodologia de *clusterização* utilizando variáveis de mercado. O capítulo quatro analisa os valores de eficiência como variável dependente de uma Regressão Quantílica considerando algumas variáveis exógenas e *dummies* que possivelmente influenciam a mudança e classificação dessa eficiência técnica relativa. Por fim, o quinto e último capítulo conclui e faz sugestões sobre os resultados.

CAPÍTULO 1 – EFICIÊNCIA BANCÁRIA

A indústria bancária brasileira, acompanhando a indústria financeira internacional, vem sofrendo alterações substanciais de regulação e estruturais com mudanças tecnológicas importantes nas últimas décadas (com impacto direto na melhoria do atendimento e dos serviços).

A alta inflação, além de toda perversidade de suas conseqüências, foi fundamental para o aperfeiçoamento das técnicas de atendimento, variedade e facilidade de produtos e serviços na indústria bancária nacional. No entanto, contribuiu, para um comodismo dos bancos pela facilidade de ganhos em *spread* bancário³.

Por outro lado, a nova conjuntura econômica verificada após a implementação do Plano Real e a baixa inflação provocou uma seleção natural, permanecendo no mercado instituições mais rentáveis e mais eficientes. Um grande número de conhecidas e tradicionais Instituições Financeiras, públicas ou privadas, fecharam suas portas ou foram incorporadas a outras instituições nacionais ou internacionais após a estabilização econômica (Banco Econômico, Banco Nacional, Bamerindus, Banco Real e diversos outros Bancos Estaduais como o Banerj, Banespa, Mercosul e outros de menor expressão).

Aquelas que sobreviveram foram forçosamente obrigadas a se reciclar e buscar maior índice de eficiência. Com a alteração na forma operacional, os incentivos de reorganização societária e os programas de reestruturação e fortalecimento do mercado financeiro brasileiro, evidenciaram um aumento de competição entre as instituições financeiras, inclusive com maior participação do capital estrangeiro nesse mercado e um acirramento na busca pela eficiência das atividades e maximização dos lucros. O resultado foi maior automação, investimento em alta tecnologia e pesquisa e, por outro lado, fechamento de agências, redução de postos de trabalho, redução de cargos, revisão de taxas, juros e serviços bem como outras medidas de igual impacto. As conseqüências dos fatos acima descritos estão hoje espelhadas nos balanços semestrais divulgados nos principais meios de comunicação em que as maiores instituições desse setor apresentam seguidamente recordes de rentabilidade e produtividade. Evidentemente que a taxa de juros, exageradamente alta no Brasil, tem contribuído e com uma grande parcela para esses

³ Diferença do que é arrecadado com o que é aplicado num Banco.

resultados, porém, não se pode desprezar os esforços dessas instituições na busca da eficiência.

No entanto, a indústria bancária quando comparada às demais indústrias, apresenta características peculiares que a distinguem na avaliação de *performance*, eficiência e resultados, bem como são mais específicas a regulamentação e registros contábeis. Os bancos, essencialmente aqueles com carteira comercial, são vulneráveis a corridas bancárias que, ainda que tenham como alvo apenas uma instituição financeira, acarretam risco sistêmico devido a considerável probabilidade de contaminação do sistema financeiro (efeito manada), tornando os bancos extremamente susceptíveis a alterações na economia, e a economia susceptível ao desempenho dos bancos. Conseqüentemente, a atividade bancária constitui alvo de regulação prudencial e monitoramento intenso (Freixas e Rochet, 1999 e Krause, Portella e Tabak, 2006).

O acompanhamento da *performance* das instituições financeiras configura-se, portanto, como constante preocupação não apenas para depositantes e investidores, mas também para analistas de mercado, pesquisadores, gestores administrativos e financeiros, órgãos reguladores e instituições governamentais, sendo bastante divulgada e consagrada na literatura, a motivação para os estudos sobre a mensuração da eficiência bancária e a exploração de seus modelos (Krause, Portella e Tabak, 2003).

Para o conceito de eficiência duas definições podem ser consideradas: (1) Eficiência Total ou de Pareto a qual assinala que uma unidade de tomada de decisão é eficiente, se e só se, é impossível dar um melhor uso aos seus recursos e modificar os resultados sem piorá-los; (2) Eficiência Relativa indica que uma unidade de tomada de decisão é classificada como 100% eficiente, se e somente se, não hajam provas de que outra unidade de tomada de decisão possa melhorar seu desempenho e superar o nível de desempenho daquela que foi qualificada como 100% eficiente (para um dado nível de recursos). Metodologicamente utiliza-se o conceito de eficiência alocativa, que consiste em incluir uma componente capaz de refletir a habilidade em selecionar *inputs* e *outputs* adequados considerando os preços (Farrel, 1957). Charnes et al. (1985) preocupados com a precisão destas medidas de preço introduziram o conceito de eficiência técnica. Desta maneira, considerou-se que a unidade 100% eficiente é aquela que tem o maior e melhor desempenho dentro da amostra utilizada. Este raciocínio indica que deve existir uma

combinação técnica e alocativa ótima o que se define como eficiência econômica. É importante mencionar que este trabalho foi conduzido por esta última definição, visto que a primeira definição é utópica e de difícil mensuração.

Muitos trabalhos foram produzidos sobre o assunto. Os mais conhecidos autores que exploraram o tema foram Berger e Humphrey (1991) que pesquisaram exaustivamente a matéria e mostraram empiricamente que diferenças entre eficiências dos bancos norte-americanos constituem fator importante em termos de vantagem competitiva de forma mais significativa do que economias de escala ou de escopo. Isto é, a habilidade gerencial de reduzir custos operacionais, para um determinado nível de produção, é mais relevante do que a escala e escopo das atividades. Estes mesmos autores (Berger e Humphrey, em 1997) avaliaram 130 estudos que aplicaram análise de eficiência técnica para instituições financeiras em 21 países (nenhum caso considerando bancos brasileiros), 18 casos avaliando agências bancárias de um mesmo banco. Segundo os autores, as informações obtidas no estudo de eficiência são importantes para subsidiar políticas governamentais (efeitos de regulamentação e desregulamentação, fusões e estrutura de mercado), pesquisa científica a respeito da descrição de eficiência de uma indústria, *ranking* de firmas ou subsidiárias, avaliação das diferentes técnicas e aperfeiçoamento do processo de gerenciamento e gestão identificando as “melhores” e ou “piores práticas”.

As técnicas de mensuração de eficiência são classificadas por não paramétricas e paramétricas. Berger e Humphrey (1997) subdividiram estes dois grupos em: *Fronteiras Não Paramétricas* que utilizam as técnicas (1) DEA (Data Envelopment Analysis) que será a metodologia utilizada neste trabalho e (2) FDH (Free Disposal Hull) que corresponde a um caso especial do modelo DEA e as *Fronteiras Paramétricas* com três principais técnicas: (1) A Stochastic Frontier Approach ou Econometric Frontier Approach (SFA), (2) Distribution-Free Approach (DFA) e finalmente (3) Thick Frontier Approach (TFA). A técnica utilizada neste segundo grupo consiste de se especificar uma forma funcional para a fronteira eficiente, atribuindo-se uma distribuição de probabilidade para as eficiências, geralmente utilizando-se a distribuição *half-normal*⁴, dentre várias outras distribuições também comuns às técnicas⁵.

⁴ Normal truncada a 50%.

⁵ Ver maior detalhes em Berger e Humphrey (1997) páginas 177 a 180.

Não há consenso na literatura sobre a melhor técnica para a mensuração da eficiência em instituições financeiras. Entretanto, como mostrado no trabalho de Berger e Humphrey (1997), é perceptível a preferência dada à aplicabilidade dos modelos DEA, considerando as diversas aplicações deste modelo nos estudos publicados até o momento. Dentre os 130 casos estudados por eles, 69 foram abordados por meio de técnicas não paramétricas, 62 dos quais com uso da metodologia DEA. Os autores concluem que os diversos métodos de mensuração não necessariamente levam à resultados consistentes e sugerem algumas alternativas de aperfeiçoamento em cada caso. Neste contexto, apresentam numa tabela, a relação dos estudos e pesquisas efetuadas considerando o país, o método, os autores, a eficiência média estimada e o tipo de grupo estudado (agências ou bancos)⁶.

Krause, Portella e Tabak (2003) apresentam, também para Bancos, uma “Síntese dos Estudos Empíricos de Eficiência Analisados” onde consolidam (nos moldes de Berger e Humphrey, 1997) diversos estudos na utilização de DEA em função de produção, Fronteira Estocástica para função custo e outras técnicas, especificando a eficiência média e indicando os *inputs* e *outputs*. Esta tabela se destaca pela comparação de diversas aplicações empíricas no Brasil⁷ e apresentação com sugestões de variáveis de insumo e produto.

Quanto a aplicação do modelo DEA em Bancos, Yeh (1996) ressalta que a seleção apropriada das variáveis de *input* e *output* constitui a mais importante tarefa no uso da metodologia e esclarece que, para justificar suas escolhas, pesquisadores e estudiosos geralmente fundamentam-se na adoção de uma das três seguintes abordagens:

- a) Bancos como prestadores de serviços: são definidos, em geral, como *outputs* o número de clientes, número de transações por unidade de tempo, o número de empréstimos e como *input* aluguel, custos de capital e de operação, condições de *marketing*, despesas de pessoal.
- b) Regra categórica: se produzir receita é *output*, se requerer desembolso é *input*. Então as receitas financeiras e não financeiras são definidas como produtos. As

⁶ Ver Berger e Humphrey (1997), páginas 181 a 184.

⁷ Ver Krause, Portella e Tabak (2003), páginas 3 e 4.

despesas de captação, administrativas e de depreciação geradas na atividade bancária são insumos.

- c) Bancos como intermediários financeiros: os *inputs* são o total de depósitos e diversos custos. Os *outputs* são o total de empréstimos, receitas financeiras e não financeiras.

Wheelock e Wilson (2003), examinam o desempenho da indústria bancária comercial dos USA utilizando estimadores não paramétricos DEA e FDH com resultados robustos em relação aos *outliers*. Segundo os autores, os resultados empíricos (no período de 1993 a 2002) revelam um substancial crescimento na produtividade destes bancos, com grande contribuição do progresso tecnológico, principalmente aqueles de maiores ativos.

No Brasil estudos desse tipo foram feitos por Souza, Tabak e Staub (2006), analisando a eficiência do sistema financeiro nacional (agregado por Bancos) utilizando modelo paramétrico e não paramétrico. Os autores contribuem para a literatura específica sugerindo uma coleção de modelos estatísticos que podem ser usados em aplicações DEA, utilizando estimadores de máxima verossimilhança no contexto das distribuições normal truncada, exponencial e modelos Tobit em geral, segundo eles, modelos tão bem estimados quanto as análises de covariância não paramétrica visto em Banker (1993), Souza (2001) e Souza (2005). Outra de suas conclusões empíricas interessantes em outro artigo de Souza e Staub (2006), em relação ao modelo e que acompanha e corrobora com alguns autores, é a utilização da eficiência relativa DEA como variável dependente em modelos paramétricos (Sampaio, Cribari e Stosic, 2005; Cribari e Zarkos, 2004; Cook, Kress e Seiford, 1996 entre outros).

Na aplicação em Bancos, comparando agências de uma mesma instituição, Thanassoulis (1999) descreve, num modelo hipotético de quatro agências, os passos necessários à utilização do modelo DEA para comparação de eficiência entre elas. A medida de eficiência é analisada sobre dois aspectos: liquidez e eficiência de mercado. O que muda nestas duas abordagens é a utilização da variável “tamanho ou potencial de mercado” como *input*. Ao mesmo tempo o autor sugere e descreve a utilização de *clusters*, tornando as DMU's, ou unidades de comparação, mais homogêneas.

Golany e Storbeck (1999) apresentam um estudo de eficiência, usando um modelo DEA, de agências bancárias de um grande Banco do USA, num total de 182 agências, utilizando seis períodos consecutivos (segundo trimestre de 1992 ao terceiro trimestre de 1993), calculando o índice de eficiência para cada período e, desta maneira, acompanhando numa série temporal o comportamento dos índices. Outra contribuição dos autores diz respeito às definições de *input* e *output* em aplicações desta natureza. O modelo contempla 5 *inputs* e 4 *outputs* subdivididos em 7 grupos de *clusters*. Golany e Storbeck (1999) apresentam a definição de *inputs* não-discrecionários tais como representativo do mercado de venda de produtos financeiros, tamanho e status das agências e atividade competitiva que são fatores que influenciam as atividades das agências, mas não dependem da decisão de sua administração. Os demais *inputs* (considerados discrecionários) definem três categorias de funcionários (medidas por horas trabalhadas). Por outro lado, no *output* consideraram volume de depósitos, volume de empréstimos, uma medida de fidelidade (número de transações por cliente) e o resultado da pesquisa de satisfação de clientes por agência.

Os sete subgrupos, onde foram aplicados os modelos DEA, dividiram-se pelo volume de depósitos (4 quartis) e empréstimos agregados (5 quintis), quatro cidades centro, localização geográfica (7 tipos de localização), número de dias que as agências ficam abertas (5, 6 e 7)⁸, espaço para tratamento privativo ao cliente, e número de máquinas ATM⁹.

Os autores concluem o artigo com a necessidade de monitoramento e acompanhamento para otimização dos insumos da indústria financeira, altamente competitiva, sobretudo com o avanço tecnológico. Acrescentando que a ferramenta é de vital importância no processo decisório ou de suporte a decisão para os gerentes de cada unidade.

Ainda sobre a definição das variáveis Rosa e Mazzon (2003), apresentam um quadro (Quadro 1 a seguir) no qual identificam as variedades de *inputs* e *outputs* utilizados em trabalhos para Bancos.

⁸ Nos USA a definição de horário e dias de abertura de agências é determinada por lei estadual, o que pode diferenciar em cada estado.

⁹ Automathic Technic Machine – máquinas de automação bancária.

QUADRO 1 – *Inputs* e *outputs* utilizados para medir a eficiência operacional em bancos – utilizando-se modelos de uma única fase

Autores	<i>Inputs</i>	<i>Outputs</i>
Sherman e Gold (1985)	- trabalho em horas - espaço ocupado - custo de suprimentos	- Quantidades de cada um dos 17 tipos de transações, agregadas em 4 grupos Segundo a dificuldade de realização
Vassiloglou e Giokas (1990)	- Trabalho expresso em horas-homem - Custo de suprimentos - Espaço ocupado pela agência - Quantidade de terminais de Computadores	- Quantidades de transações distribuídas em quatro classes, segundo a dificuldade de realização
Schaffnit, Rosen e Paradi (1997)	- Quantidade de pessoal alocado em diversos tipos de função (caixas, atendimento, gerente, suporte, supervisão, análise de crédito, etc.)	- Quantidade de transações e vendas de produtos - Quantidade de contas para Manutenção
Soteriou e Stavrinides (1997)	- Quantidade de horas trabalhadas por escriturários e gerentes - Quantidade de horas de computadores - Espaço ocupado - Quantidade de contas de pessoa física, poupança, pessoa jurídica e empréstimos	- Índice de percepção de Qualidade em Serviços baseado em SERVQUAL
Golany e Storbeck (1999)	- Quantidade de horas semanais trabalhadas por caixas e não-caixas - Facilidades na agência - Espaço ocupado - Índice de emprego na região	- Volumes de empréstimos - Volumes de depósitos - Retenção de clientes - Satisfação de clientes
Zenios et al. (1999)	- Recursos utilizados em horas (gerentes, escriturários e terminais de computadores) - Espaço ocupado - Microambiente expresso por quantidade de contas correntes, poupança, câmbio e pessoa jurídica, e propostas de empréstimos	- Trabalho total produzido pela agência (em horas)

Outro exemplo interessante e clássico na indústria financeira é apresentado por Zenios et. al. (1999) quando trabalha no desenvolvimento de modelo que estuda a eficiência de agências bancárias do Banco do Chipre utilizando DEA. Três princípios motivaram a utilização da metodologia:

- *Benchmarks* na eficiência relativa das agências;
- Melhoramento nas estratégias de gerenciamento das agências e
- Isolar e apurar o efeito do ambiente na eficiência de cada unidade (agência).

Os resultados apurados indicaram recursos subutilizados, impactando a eficiência, principalmente no período de baixa temporada turística que poderiam ser redistribuídos

para outras unidades. Após estes resultados o Banco estabeleceu políticas e estratégias de aproveitamento dos recursos criando uma unidade de suporte operacional para melhoria do desempenho.

O artigo se destaca pela utilização e minuciosa avaliação do microambiente de inserção de cada DMU. Zenios et. al. (1999) analisam a normalidade da distribuição dos estimadores de eficiência e testa a hipótese não paramétrica das agências urbanas e turísticas (teste de Mann-Withney)¹⁰, que segundo Souza (2002) é o procedimento padrão para a comparação de dois grupos de DMU's. Para mais de dois grupos a estatística apropriada é a de Kruskal-Wallis¹¹. Esses testes são robustos à presença de heterocedasticidade e de observações atípicas (*outliers*) nos dados (Souza, 2002).

A utilização da metodologia DEA foi a que se apresentou como mais indicada às comparações que aqui serão desenvolvidas. Isto é, comparações entre agências de uma mesma instituição, visto que comparações de instituições diferentes podem esbarrar em padrões diferentes de atuação e gestão, principalmente quando se consideram instituições públicas e privadas, não obstante deve-se considerar a dificuldade de informações disponíveis nas demais instituições financeiras cercadas de todo o sigilo inerente às suas atividades, particularmente para efeito comparativo.

Um questionamento que pode ser feito, principalmente por não pesquisadores e gestores de instituições envolvidas e não acadêmicas, é sobre o motivo e a praticidade da utilização de metodologia complexa para cálculo de eficiência quando se trabalha com variáveis sobre mesma medida e que facilmente, em seu conjunto, definem a receita, despesa e, conseqüentemente, o resultado das unidades e suas instituições. Num modelo simples a eficiência poderia ser calculada numa fração: resultado sobre despesas ou resultado sobre número de horas trabalhadas em cada unidade. Embora atrativa pela sua simplicidade essa metodologia não proporciona cálculos comparativos com valores relativos entre as unidades, o modelo DEA estabelece comparação entre as unidades. Ademais os pesos e a importância relativa de cada variável de *output* ou *input* não seriam considerados a não ser arbitrariamente. Os “pesos” no modelo DEA são calculados por

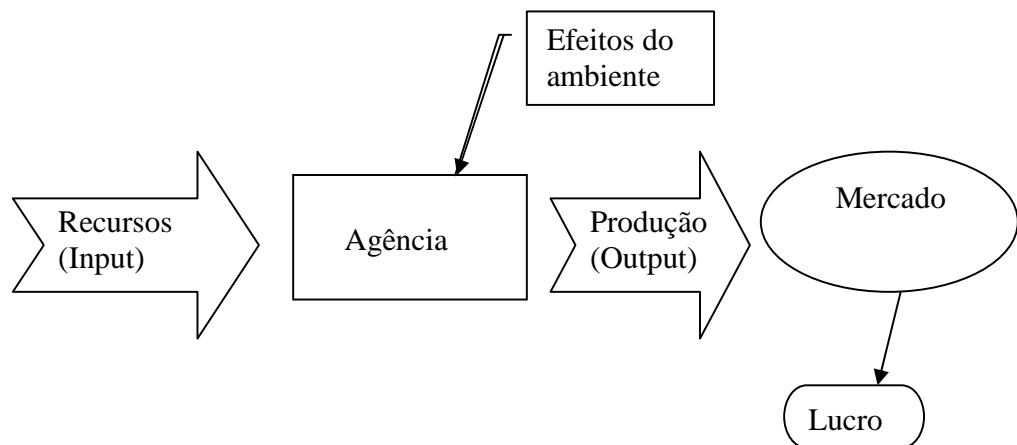
¹⁰ Zenios et. al. (1999) utiliza o teste de Shapiro-Wilks para testar normalidade. No resultado, rejeitando a normalidade da distribuição, utiliza o teste não paramétrico de Mann-Whitney com $p < 0,001$ de diferença entre as duas distribuições. Ver detalhes sobre o teste não paramétrico em Siegel (1956), páginas 131 a 144.

¹¹ Ver em Siegel (1956), páginas 209 a 218.

comparações entre as unidades utilizando modelos de programação linear. Outro obstáculo à aplicação da simplicidade seria a dificuldade de se comparar variáveis em escalas diferentes, mesmo com mesma unidade de medida.

Além desta situação e conforme o diagrama da figura 1.1 abaixo o conceito de eficiência a ser explorado, em instituição definida como Banco Comercial, considera agências que utilizam recursos para gerar produção de acordo com o ambiente (interno e externo) com que elas operam. Essa produção agindo no mercado gera o lucro.

Figura 1.1 : Fluxo de Eficiência



É nesse contexto com a definição de produtividade relativa por unidade, que serão avaliadas agências de uma determinada instituição, estabelecendo as relações dos “Recursos” até a “Produção”, considerando o “Mercado” como fator vital de influência e, conseqüentemente, mesmo não discutido, trabalhado e comparado empiricamente, o “Lucro”. No próximo capítulo será apresentada a teoria sobre a metodologia DEA utilizada e que será base da aplicação para os estudos empíricos do capítulo 3.

CAPÍTULO 2 – MENSURAÇÕES NÃO PARAMÉTRICAS

Os modernos estudos sobre eficiência, aplicando mensurações não paramétricas, acompanham a história da Análise Envoltória de Dados – DEA (Data Envelopment Analysis) – que surgiu, formalmente, em 1978 a partir do trabalho de Charnes, Cooper e Rhodes, como consequência da tese para obtenção de grau de Ph.D de Edward Rhodes e sob a orientação de W.W. Cooper. O problema abordado tinha como objetivo medir eficiência de unidades tomadoras de decisão (DMU - *Decision Making Unit*) na presença de múltiplos *inputs* (fatores de produção) e múltiplos *outputs* (produtos). Particularmente era desenvolver um método para comparar a eficiência de escolas públicas (DMU's) com os seguintes *outputs*:

- Scores aritméticos;
- Melhoria de auto-estima medida em testes psicológicos e
- Habilidade psicomotora.

Quanto aos *inputs* consideravam-se:

- Número de professores-hora e
- Tempo gasto pela mãe em aprendizagem com o filho.

O objetivo da tese foi desenvolver um modelo para estimar a eficiência técnica sem recorrer ao arbítrio de pesos para as variáveis de *input* e *output*, e, principalmente, sem converter todas as variáveis em valores econômicos comparáveis.

Na realidade, a abordagem analítica rigorosa, aplicada à medida de eficiência na produção, teve origem anterior com o trabalho de Debreu (1951). Para esses autores a eficiência técnica para um vetor *input-output* só ocorre quando:

- I. Nenhum dos *outputs* pode ser aumentado sem que algum outro *output* seja reduzido ou algum *input* necessite ser aumentado.
- II. Nenhum dos *inputs* pode ser reduzido sem que algum outro *input* seja aumentado ou algum *output* seja reduzido.

A formulação de problemas de medidas de eficiência como “problemas de programação linear” foi concebida inicialmente por Boles, Bressler, Seitz e Sitorus em 1966 para o caso linear por partes, conforme Färe et al. (1994). Entretanto, foi com o empenho de Charnes, Cooper e Rhodes (1978) que os modelos DEA ganharam maior penetração, a partir do modelo original de rendimentos constantes de escala CRS ou CCR (sigla para Charnes, Cooper e Rhodes) e do modelo de rendimentos variáveis de escala VRS ou BCC (Banker, Charnes e Cooper, 1984).

Posteriormente, Charnes et al. (1985) reforçaram o fato da necessidade de tratar esta definição como um conceito relativo onde a eficiência de 100% é atingida por uma unidade quando comparações com outras unidades relevantes não provêm evidência de ineficiência no uso de qualquer *input* ou *output*. Este conceito nos permite diferenciar entre estados de produção eficientes e ineficientes. No entanto, não permite medir o grau de ineficiência de um vetor ou identificar um vetor ou uma combinação de vetores eficientes com os quais se podem comparar um vetor ineficiente.

Anteriormente Debreu (1951), preocupado com isso, abordou o problema introduzindo uma medida radial de eficiência técnica conhecida como coeficiente de utilização de recursos. Esta medida radial pode buscar a máxima redução equiproporcional de todos os *inputs* ou a máxima expansão equiproporcional de todos os *outputs*. A vantagem do uso deste coeficiente é que ele independe da unidade de medida de cada variável. Porém um vetor “*input-output*” eficiente na medida radial de Debreu pode não ser eficiente na definição de Pareto-Koopmans.

Como citado no capítulo anterior, Farrel (1957) estendeu o trabalho de Koopmans de forma a incluir uma componente capaz de refletir a habilidade dos produtores em selecionar o vetor “*input-output*” eficiente considerando os respectivos preços. Esta componente foi denominada eficiência alocativa. Entretanto, a dificuldade em medir estes preços de forma mais acurada, segundo Charnes et al. (1985), foi um dos motivos que levou os trabalhos em DEA enfatizarem medidas de eficiência técnica.

2.1 – DEA E FDH – CARACTERÍSTICAS E LIMITAÇÕES

Segundo as situações apresentadas anteriormente algumas características importantes podem ser destacadas para o método DEA conforme Esttelita Lins e Meza, (2000):

- Difere dos métodos baseados em avaliação puramente econômica, que necessitam converter todos os *inputs* e *outputs* em unidades monetárias;
- Os índices de eficiência são baseados em dados reais;
- Generaliza o método de Farrel (1957), construindo um único *output* virtual e um único *input* virtual;
- É uma alternativa e um complemento aos métodos da análise da tendência central e análise custo benefício;
- Considera a possibilidade de que *outliers* não representem apenas desvios em relação ao comportamento "médio", mas possíveis *benchmarks* (Zenios et al., 1999) a serem estudados pelas demais DMU's (*Decision Making Unit*). A DMU é definida como a alternativa cuja eficiência está sendo avaliada, o conjunto de DMU's adotado em uma análise DEA deve ter em comum a mesma utilização de *inputs* e *outputs*, ser homogêneo e ter autonomia na tomada de decisão.
- A abordagem DEA, ao contrário das abordagens paramétricas tradicionais, otimiza cada observação individual com o objetivo de determinar uma fronteira por partes que compreende o conjunto de DMU's, Pareto-eficiente.
- Além destas características, pode-se ainda destacar que o método:
 - a) caracteriza cada DMU por uma única medida de eficiência relativa,
 - b) realiza projeções de melhoria de eficiência baseada nas melhores práticas,

c) provê informações complementares sobre rendimento de escala (crescente e decrescente).

Estas características conferem ao método uma potencialidade de resgatar a natureza essencialmente aplicada. Particularmente, os modelos DEA fornecem um indicador quantitativo e comparativo, o que são as características desejadas para indicadores em modelos financeiros utilizados em Bancos (Golany and Storbeck, 1999; Thanassoulis, 1999; Zenios et al., 1999 e Wheelock e Wilson, 2003 entre outros). Entretanto, a despeito de suas vantagens o modelo apresenta limitações ao número de observações (número grande de DMU's) principalmente na detecção dos desvios ou *outliers*, que podem ser oriundos de unidades “modelo” ou ocasionadas por possíveis distorções e erros na base de dados, o que será discutido mais adiante.

Representando uma alternativa técnica para aproximações não paramétricas a técnica DEA conta com uma hipótese restritiva na estrutura de produção. Esta suposição tem sido proposta por Deprins, Simar e Tulkens (1984). Eles postularam que a fronteira de produção é simplesmente o limite do Free Disposal Hull (FDH) do conjunto de dados. É mantida importante disponibilidade de *inputs* e *outputs* com retornos variáveis de escala, porém a hipótese de convexidade não é requerida. Neste método – FDH – a fronteira é obtida pela comparação de *inputs* e *outputs*. Consistente com os retornos variáveis de escala os elementos de intensidade de um vetor λ são restritos a soma da unidade. A restrição de integralidade nas variáveis implica que as combinações lineares das diferentes observações são excluídas e a convexidade não é imposta. Em termos da técnica DEA adiciona-se a restrição:

$$\sum_{j=1}^K \lambda_{kj} = 1 \quad \lambda_{kj} \in (0,1) \quad \text{onde} \quad k = 1, 2, \dots, K$$

Um dos conceitos e limitações associadas à utilização do FDH, quando é necessária uma comparação dos dados, é que a metodologia tende a gerar grande número de observações eficientes, limitando seu poder discriminatório o que faz com que as estimativas empíricas aqui desenvolvidas concentrem-se principalmente nos modelos CCR

e BCC. Entretanto, será utilizada a técnica de FDH para calcular e melhor estabelecer as relações de dominância na fronteira de produção.

2.1.1 - Modelo dos Multiplicadores - Retorno Constante de Escala – CCR ou CRS

O modelo original CCR também conhecido como CRS (Constant Returns to Scale) trabalha com retornos constantes de escala (Charnes et al., 1996), na ótica dos multiplicadores e do *input*¹², que é o modelo utilizado na orientação desse trabalho.

Considera-se que cada DMU k é uma unidade de produção que utiliza n *inputs* x_{ik} , $i=1, \dots, n$, para produzir m *outputs* y_{jk} , $j=1, \dots, m$. Este modelo maximiza o quociente entre a combinação linear dos *outputs* e a combinação linear dos *inputs*, com a restrição de que para qualquer DMU este quociente não pode ser maior que 1. Um índice de eficiência é condicionalmente definido como a combinação linear dos *outputs* dividida pela combinação linear dos *inputs* de determinada DMU $_k$ variando de 0 a 1. Em termos matemáticos, tem-se o seguinte problema de programação fracionária (Charnes et al., 1996):

$$\max \theta_k = \frac{\sum_{j=1}^m u_{jk} \cdot y_{jk}}{\sum_{i=1}^n v_{ik} \cdot x_{ik}}$$

Com as restrições de

$$\frac{\sum_{j=1}^m u_{jk} \cdot y_{jk}}{\sum_{i=1}^n v_{ik} \cdot x_{ik}} \leq 1$$

Para $\forall k$; u_{jk} e v_{ik} não negativos.

¹² A técnica DEA sugere dois modelos: orientado para o *Input* ou orientado para *Output*. No primeiro caso deseja-se minimizar a utilização dos recursos tal que o nível dos *outputs* ou produtos se mantenha o mesmo. Isto é, pergunta-se quanto posso reduzir meus *inputs* sem afetar as quantidades de *outputs* produzidos? No caso de orientação *output* o objetivo é maximizar os produtos obtidos sem alterar o nível atual dos *inputs*. Maiores detalhes na página 33 de Estellita Lins e Meza (2000).

Nesta expressão u e v são os vetores de pesos ou multiplicadores dados aos *outputs* e *inputs*, respectivamente. Em outra situação estes pesos seriam arbitrados por especialistas. A proposta do modelo evita decisões arbitrárias na definição de importância das variáveis.

Portanto, para cada DMU a ser analisada, formula-se um problema de otimização com o objetivo de determinar os valores que esta DMU atribuiria aos multiplicadores u e v de modo a aparecer com a maior eficiência possível. Este processo é repetido K vezes conforme o número de DMU's.

Este modelo é linearizado pelas suas restrições, transformando-se em um Problema de Programação Linear. Como o problema tem infinitas soluções fixa-se um valor constante para o denominador da função objetivo. Charnes, Cooper e Rhodes (1978) apresenta o seguinte modelo:

$$\max \theta_k = \sum_{j=1}^m u_j y_{jk}$$

tal que
$$\sum_{i=1}^n v_i x_{ik} = 1$$

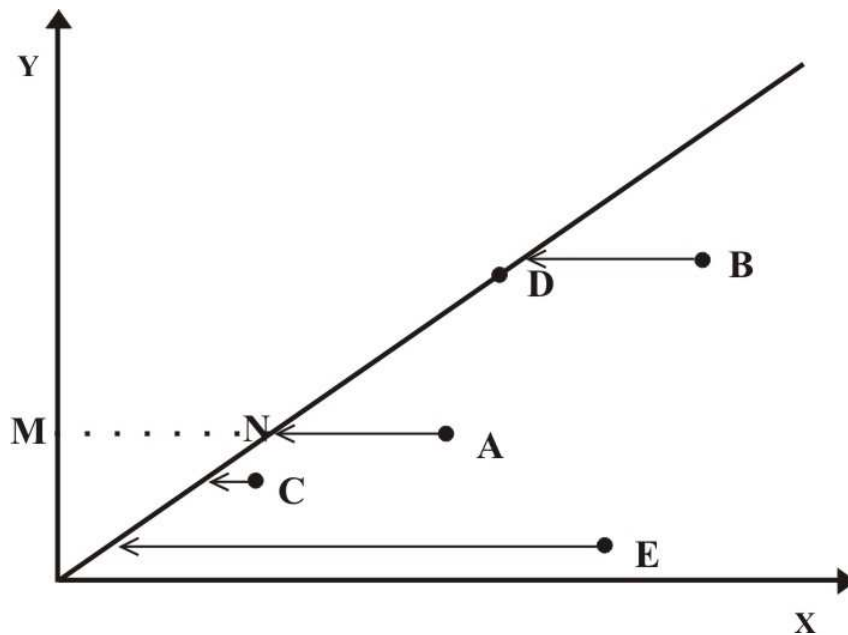
$$\sum_{j=1}^m u_j y_{jk} - \sum_{i=1}^n v_i x_{ik} \leq 0 \quad \text{com } k = 1, \dots, K$$

e
$$u_j, v_i \geq 0, \quad \forall x, y$$

Este problema de programação linear assume rendimentos constantes de escala (CRS) a partir de multiplicadores aplicados aos *inputs* e *outputs* (v_i e u_j respectivamente).

Um exemplo que se pode visualizar graficamente a posição de eficiência num modelo CRS de multiplicadores foi apresentado por Estellita Lins e Meza (2000) com duas variáveis: um *input* (x) e um *output* (y) e cinco DMU's, onde a fronteira em forma de semi-reta passa pela origem e pela DMU eficiente D conforme a figura 2.1. As demais DMU's, que não se encontram na fronteira, são ineficientes. As setas indicam a projeção (ou posição) de cada DMU ineficiente na fronteira, indicando que na redução do *input* tal DMU alcançaria sua eficiência. No exemplo a eficiência da DMU A é dada pela distância do eixo Y até o ponto ótimo N dividida pela distância do eixo Y até o ponto A.

Figura 2.1 : Exemplo de fronteira de eficiência



2.1.2 - Modelo do Envelope - Retorno Variável de Escala – BCC ou VRS

Algumas definições e propriedades são necessárias para representar este modelo:

Definição: Uma tecnologia de produção transformando *inputs* $X = (x_1 \ x_2 \ \dots \ x_n) \in \mathcal{R}_+^n$ em *outputs* $Y = (y_1 \ y_2 \ \dots \ y_m) \in \mathcal{R}_+^m$ que pode ser representado por um conjunto de possibilidades de produção como:

$$T : \{ (x, y) \in \mathcal{R}_+^{n+m} \text{ tal que } y \geq 0 \text{ pode ser produzido por } x \geq 0 \}$$

O conjunto T é definido a partir das observações de K DMU's onde (x_k, y_k) para $k = 1, \dots, K$ assumindo as seguintes propriedades desejáveis para T, conforme Banker, Charnes e Cooper (1984):

1ª. Propriedade: Convexidade

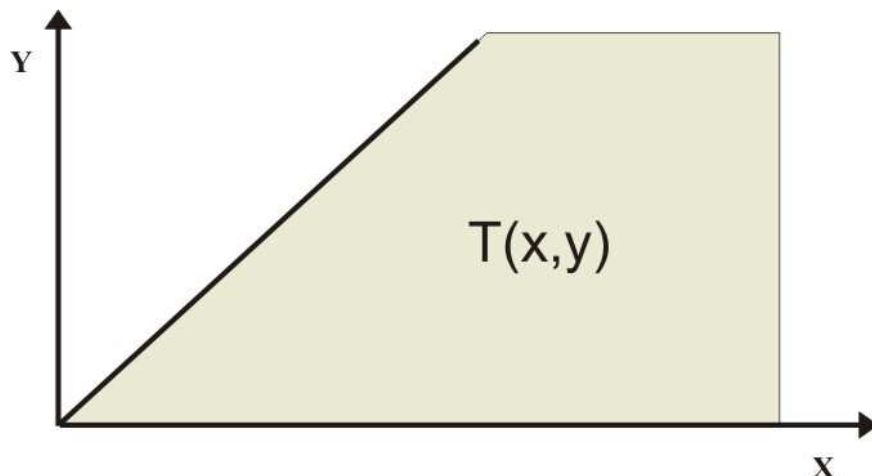
Se $(x_k, y_k) \in T$ e $\lambda_k, k = 1, 2, \dots, K$ são escalares não negativos tais que

$\sum \lambda_k = 1$, então :

$$(\sum \lambda_k x_k, \sum \lambda_k y_k) \in T.$$

O Lugar Geométrico (LG) dos pontos de operação viáveis é dado pela combinação linear convexa definida acima e representada pelo polígono hachurado na figura 2.2

Figura 2.2 : Lugar Geométrico da Fronteira de Produção



2ª. Propriedade: Ineficiência

a) se $(x, y) \in T$ e $\bar{x} \geq x \Rightarrow (\bar{x}, y) \in T$

b) se $(x, y) \in T$ e $\bar{y} \leq y \Rightarrow (x, \bar{y}) \in T$

O LG dos pontos que satisfazem a primeira e segunda propriedade é a área a direita e abaixo de qualquer ponto do polígono hachurado indicado na figura 2.2.

3ª. Propriedade: Raio ilimitado (válido para o modelo que admite rendimentos constantes de escala – CRS). Se $(x, y) \in T \Rightarrow (c x, c y) \in T, \forall c > 0$ onde c é uma constante qualquer.

Esta propriedade faz com que qualquer ponto pertencente ao segmento de reta que passa pela origem e por um ponto viável (conforme as propriedades anteriores) também seja viável.

Da 1ª. e 3ª. propriedades temos que: $\forall (x,y) = (c \sum \lambda_k x_k, c \sum \lambda_k y_k)$, com $c > 0$,

$\lambda_k \geq 0$ e $\sum \lambda_k = 1$ está em T , ou seja $(\sum \mu_k x_k, \sum \mu_k y_k)$ com $\mu_k \geq 0$ está em T , com a propriedade 2:

$$(x, y) \in T \text{ se e somente se } x \geq \sum \mu_k x_k \text{ e } y \leq \sum \mu_k y_k \text{ para } \mu_k \geq 0 \quad \forall k \quad (2.1)$$

Isto é, para um novo vetor de *inputs* e *outputs* (x, y) , só será um vetor pertencente ao conjunto de possibilidades de produção se existir uma combinação linear de *inputs* e *outputs* de DMU observadas tal que este novo vetor consuma *inputs* em nível maior ou igual, e produza *outputs* em nível menor ou igual do que os da combinação linear.

Até agora foi definido o conjunto de viabilidade de produção os quais modelam o conjunto de restrições do problema de programação linear (PPL). Para a função objetivo temos a seguinte definição:

Definição: A função distância de Shepard para um conjunto de *outputs* $L(y)$ é $g(x, y)$ onde

$$g(x, y) = \frac{1}{\theta(x, y)} \quad \text{e} \quad h(x, y) = \min\{\theta / \theta x \in L(y), \theta \geq 0\}$$

Portanto, na análise de determinada DMU, θ é o menor valor (e sempre menor do que 1) tal que, multiplicado pelo vetor de insumo da citada DMU, resulta em redução equiproporcional destes insumos. O que se pretende é a máxima redução equiproporcional, ou seja, o mínimo de θ , para que se possa garantir que a DMU, operando com esta nova combinação de insumos, ainda pertença ao conjunto de possibilidades de produção T . Para isto, precisa-se exigir que o vetor de *inputs* e *outputs* $h(x, y)$ satisfaça às restrições estabelecidas em (2.1). A expressão do PPL será achar:

O valor mínimo de θ

tal que $\theta x \geq \sum \mu_k x_k$

$$y \leq \sum \mu_k y_k \quad \text{onde} \quad \mu_k \geq 0 \quad \forall k \quad \text{e} \quad \theta \geq 0 \quad (2.2)$$

Onde θ será interpretado como o indicador de eficiência da DMU analisada, baseado na possibilidade de redução de insumos para obter a eficiência máxima. Esta é a ótica do input com rendimento constante de escala quando o conjunto de pontos, tais que $\theta=1$, é definido como a fronteira de eficiência, embora não seja sempre Pareto-eficiente o

que é definido e garantido no modelo não-arquimediano (Charnes, Cooper, Lewin and Seiford, 1996 2ª. Edição)¹³.

Na eliminação da terceira propriedade, de raio ilimitado (rendimento constante de escala), o modelo DEA torna-se BCC \equiv VRS (Variable Returns to Scale), passando a considerar a possibilidade de rendimentos crescentes ou decrescentes de escala na fronteira eficiente. Os coeficientes μ_i são substituídos pelos λ_k , que definem uma combinação linear convexa, conforme figura 2.3. O PPL e a fronteira que envolve os pontos viáveis são definidos como:

O mínimo de θ

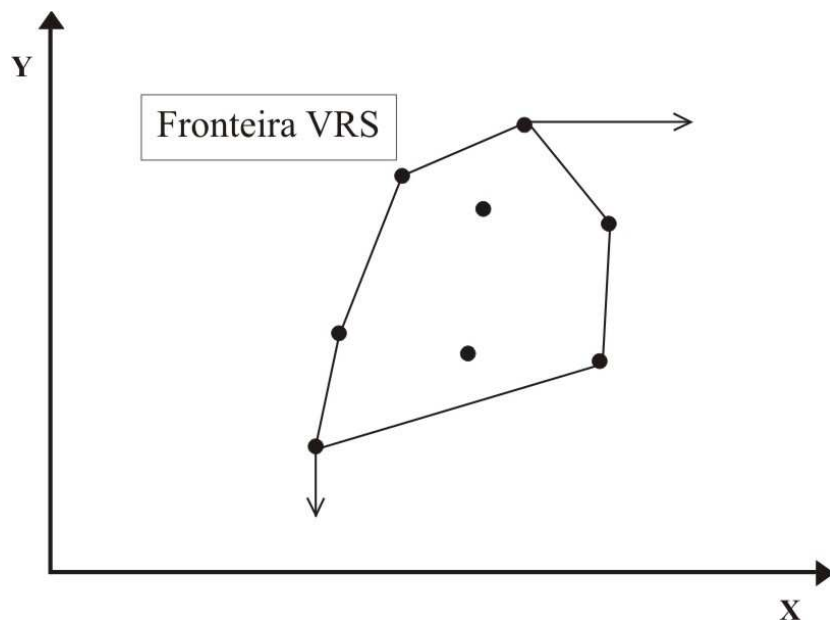
tal que $\theta x \geq \sum \lambda_k x_k$

$$y \leq \sum \lambda_k y_k \quad \text{onde} \quad \lambda_k \geq 0 \quad \forall k \quad \text{e} \quad \theta \geq 0$$

$$\sum \lambda_k x_k = 1 \quad (2.3)$$

¹³ Maiores detalhes veja em Charnes, Cooper, Lewin and Seiford (1996, 2ª. edição - KAP). Data Envelopment Analysis, Theory Methodology and Applications, 3º capítulo página 76 .

Figura 2.3 – Fronteira VRS



2.2 – O PROBLEMA DOS *OUTLIERS*: DIFERENTES ABORDAGENS

A metodologia DEA é extremamente atrativa visto que não requer conhecimentos *a priori* da relação funcional entre as variáveis de *input* e *output*. Por outro lado, como visto na seção anterior, a metodologia é baseada no conceito de fronteira de produção, onde um simples erro no conjunto de dados (ou um desempenho excepcional de uma unidade de produção) pode causar sérios compromissos na análise.

Devido a sensibilidade dos modelos DEA na presença de *outliers* um dos problemas encontrados na técnica em questão diz respeito à limitação na utilização da metodologia para grandes bases de dados, portanto com número considerável de DMU's, exigindo alguns procedimentos adicionais para tornar robustos os estimadores de eficiência. A não homogeneidade dos dados, bem como a complexidade e repetição dos cálculos no modelo de otimização é responsável pela dificuldade natural na sua aplicação.

Vários estudos foram propostos na literatura especializada para tentar amenizar e solucionar esse tipo de situação. Wilson (1993 e 1995) introduziu métodos descritivos para detectar a influência destas observações nas medidas não paramétricas, Seaver e Triantis (1992 e 1995) propuseram um modelo denominado de *Fuzzy Clustering Strategies* e *High*

Breakdown Procedures para identificar os *outliers* e os pontos ou unidades de maior impacto (*leverage*).

Técnicas mais recentes foram desenvolvidas por Cazals, Florens e Simar (2002) e Simar (2003) utilizando medidas de fronteira de *ordem-m*, bem como os modelos de fronteira de eficiência robusta de Kuosmanem e Post (1999) e Cherchye, Kuosmanem e Post (2000). A aproximação de *ordem-m* é baseada no conceito de expectativa da função de insumo mínimo ou função de produção máxima que são técnicas utilizadas em aplicações de estimadores FDH. As medidas de eficiência robusta – Robust Efficiency Measurements (REM) – são modelos de decomposição do DEA original, ambas são inferências estatísticas com características de natureza não paramétrica.

No entanto, todas estas propostas são orientadas por modelos que basicamente se utilizam de inspeção visual dos dados, o que torna impraticável com grandes bases de dados. Sampaio de Sousa e Stosic (2005), apresentaram medidas de eficiência através do modelo DEA testando robustez para grandes amostras ou grandes populações utilizando solução computacionalmente intensiva. O modelo está baseado na combinação de técnicas amostrais de *Bootstrap* e *Jackknife*¹⁴, principalmente para grandes amostras. Inicialmente um algoritmo aplicando “*Jackknife*” é usado para extrair *leverage* (impacto de uma DMU). O procedimento consiste em calcular todas as medidas de eficiência para cada DMU, posteriormente repete-se o procedimento retirando uma DMU. Nesta operação procura-se saber a influência dessa unidade para os índices de eficiência das demais. Na sequência, utiliza-se o modelo estocástico “*Bootstrap*” para analisar as informações geradas pela distribuição dos *leverage*. A associação destas duas técnicas de “reamostragem” foi batizada de “*Jackstrap*”.

É demonstrado pelos autores que as aproximações são robustas na presença de *outliers* e erros no conjunto de dados, o que torna o modelo adequado para implementação em grandes bases de dados. Este modelo é testado em outro artigo dos mesmos autores e no mesmo ano com 4.796 municípios brasileiros. A intenção é comparar o grau de eficiência dos municípios quanto aos recursos e políticas públicas utilizadas testando a

¹⁴ Modelos de geração de amostras em que os dados originais são repetidamente criados em subamostras para geração de modelos de estimação, viabilizados através de estimação empírica ao invés de paramétrica. Maiores detalhes podem ser vistos em Hair, Anderson and Tatham (2005), *Análise Multivariada de Dados*. Editora Bookman, 5ª. edição, páginas 466 a 470.

adequabilidade do modelo, para isso foram consideradas quatro variáveis de *input* e nove de *output*. Os *inputs* são definidos pela despesa corrente, número de professores, taxa de mortalidade infantil, hospitais e serviços de saúde. O *output* é apresentado por variáveis e indicadores educacionais, população e condições de saúde e infraestrutura (Sampaio de Sousa e Stosic, 2005). As análises indicaram medidas de eficiência e ineficiência associadas ao porte dos municípios, bem como outros resultados interessantes de causalidade.

2.3 – A ABORDAGEM “JACKSTRAP”

Inicialmente é utilizado o procedimento *Jackknife* a fim de extrair o *leverage*, isto é, o impacto que a remoção de uma observação ou DMU causa para a medida de comparação do desempenho no conjunto dos dados restantes. A idéia subjacente é que os *outliers* ocasionam grande impacto na eficiência média e por isso deveriam ocorrer com baixa probabilidade em relação aos demais. Em seqüência aplica-se uma “reamostragem” estocástica utilizando o *Bootstrap* no cálculo dos *leverages* para cada DMU, de tal maneira que todas as unidades sejam submetidas a, aproximadamente, um número n de repetições, como será visto adiante. Este procedimento é necessário para reduzir e tornar viável o tempo de processamento.

Para formalização considere K Unidades de Tomada de Decisão (Decision Making Units – DMU), onde a k -ésima DMU ($k = 1, 2, 3, \dots, K$) usa N não negativos *inputs* $x^k = (x_{k1}, x_{k2}, \dots, x_{kN}) \in \mathcal{R}_+^N$ para produzir M não negativos *outputs*. Isto é, $y^k = (y_{k1}, y_{k2}, \dots, y_{kM}) \in \mathcal{R}_+^M$.

Então, conforme (2.2) ou (2.3) (CCR – 1978 ou BBC - 1984) procura-se avaliar o efeito dos erros ou *outliers* sobre todas as unidades e o motivo de sua classificação de eficiência.

O *leverage* de uma DMU pode ser entendido como a quantidade que mede o impacto de se remover uma DMU do conjunto total de dados, informando um escore de eficiência com relação a todas as outras DMU’s. Ele pode ser entendido como o desvio padrão das medidas de eficiência antes e depois de removida a informação. A forma sugerida é a técnica de “reamostragem” que segue “passo a passo” (Sampaio de Sousa e Stosic, 2005):

1. Calculam-se as medidas de eficiência para cada DMU utilizando DEA para os dados originais, obtendo $\{\theta_k \mid k = 1, 2, \dots, K\}$.
2. Uma por uma as DMU's são sucessivamente removidas e, portanto, a cada processamento o conjunto de eficiência é recalculado onde $\{\theta_{kj}^* \mid k = 1, 2, \dots, K\}$ e $j=1, 2, \dots, K$ representa a DMU removida.

O *leverage* da j -ésima DMU pode ser definido como o desvio-padrão :

$$l_j = \sqrt{\frac{\sum_{k=1; k \neq j}^K (\theta_{kj}^* - \theta_k)^2}{K-1}}$$

Como o processo torna-se computacionalmente intensivo, mesmo porque para estes procedimentos ter-se-ia que solucionar $K(K-1)$ PPL (Programas de Programação Linear) que, para grandes conjuntos de dados, tornaria o procedimento inviável necessitando muitas horas de processamento. Assim, Sampaio de Sousa e Stosic (2005) propuseram a combinação da técnica de *Bootstrap* associada a técnica *Jackknife* definida acima. Em seguida, o processo, detalhadamente, consiste dos seguintes passos:

1. Selecionar aleatoriamente um conjunto de L DMU's (usa-se, normalmente, 10% de K) e desenvolve o procedimento anterior de tal forma a se obter subconjuntos de valores \tilde{l}_{jl} , onde o índice l representa L diferentes valores do conjunto $\{1, 2, \dots, K\}$.
2. Repetindo o passo 1 acima B vezes acumula-se o subconjunto de informações \tilde{l}_{jb} para todas as DMU's, de tal forma que B seja grande suficiente para que cada DMU tenha sido selecionada, aproximadamente, $n_j \approx BL/K$ vezes.

3. Para cada DMU calcula-se a média $\tilde{l}_j = \frac{\sum_{b=1}^{n_j} \tilde{l}_{jb}}{n_j}$ e a média global de todos os

$$\text{leverage} \quad \tilde{l} = \frac{\sum_{j=1}^K \tilde{l}_j}{K}$$

As informações geradas pelo *leverage* podem ser usadas para detectar, eliminar ou dar tratamento diferenciado para os erros, *outliers* ou valores de destacada eficiência no conjunto de dados. Para detectar ou eliminar DMU's com valores altos no *leverage* pode-se testar algumas funções distribuição de probabilidade. As funções propostas são as funções: Linear, Inversa, Exponencial e a função *Heaviside step*. A distribuição Normal foi testada, porém com resultados insatisfatórios. A seguir serão apresentadas estas funções com algum detalhe:

1. A função probabilidade linear é dada por:
$$P(\tilde{l}_k) = \frac{l_{\max} - \tilde{l}_k}{l_{\max} - l_{\min}}$$

onde l_{\max} e l_{\min} representam os valores máximos e mínimos, respectivamente, do conjunto de *leverage* $\{ \tilde{l}_k : k = 1, \dots, K \}$. A distribuição indica que a probabilidade de seleccionar uma DMU com l_{\min} é um, no entanto essa probabilidade torna-se zero caso l_{\max} .

2. Outra distribuição apropriada é a distribuição de probabilidade inversa representada por

$$P(\tilde{l}_k) = \frac{l_0 (l_{\max} - \tilde{l}_k)}{\tilde{l}_k (l_{\max} - l_0)}$$

onde l_0 é um valor pequeno na fronteira de \tilde{l}_k (valores abaixo deste valor são seleccionados com probabilidade 1). Neste trabalho utilizou-se um valor pequeno para l_0 , de tal forma que $l_0 = 0,0001$. Esta escolha traz certa arbitrariedade à distribuição, de tal forma que o valor escolhido influencia na seleção das DMU's retirando ou mantendo algumas unidades conforme se diminui ou se aumenta o valor de l_0 , respectivamente.

3. A distribuição exponencial é dada por:

$$P(\tilde{l}_k) = \frac{e^{-\tilde{l}_k} - e^{-l_{\max}}}{e^{-l_{\min}} - e^{-l_{\max}}}$$

4. Finalmente tem-se a função de distribuição *Heaviside step*

$$P(\tilde{l}_k) = \begin{cases} 1 & \text{se } \tilde{l}_k < \tilde{l} \log K \\ 0 & \text{se } \tilde{l}_k \geq \tilde{l} \log K \end{cases}$$

onde K representa o tamanho da amostra de DMU's utilizada, a fronteira é escolhida como o produto $\tilde{l} \log k$. A utilização desta distribuição torna os resultados mais confiáveis pela maior robustez e menor arbitrariedade para os valores.

Mesmo que estes *outliers* não representem erros, a utilização das distribuições consegue apartar os de maior impacto no conjunto das observações, indicando aqueles que mais se destacam. Convencionalmente, como a distribuição de θ é comportada e, razoavelmente, bem ajustada para uma distribuição simétrica, aplicou-se $(\log_{10} N) \times \tilde{l}$ ¹⁵ para o ponto de corte na função de distribuição *Heaviside step*. Para todas as demais funções (linear, exponencial e inversa), aplicado o cálculo de probabilidade na distribuição, eliminam-se os valores da amostra que apresentam probabilidade abaixo de 95%¹⁶ e, por isso, considerados *outliers*.

No capítulo seguinte serão definidos os dados que compõem informações e variáveis de uma grande e importante instituição financeira do Brasil, sobre a qual será feita aplicação do modelo e análise de seus resultados.

¹⁵ O valor N corresponde ao número total K de DMU's consideradas na aplicação do capítulo 3. No caso do estudo empírico serão consideradas 3663 agências.

¹⁶ Por esta escolha arbitrária da probabilidade que a opção preferencial será do modelo *Heaviside step*.

CAPÍTULO 3 – ÍNDICES DE EFICIÊNCIA NÃO PARAMÉTRICOS - ESTUDO EMPÍRICO E APLICAÇÃO DE DEA AO BANCO DO BRASIL

Neste capítulo será feita uma análise empírica numa Instituição Financeira do Brasil. A amostra deste estudo é composta de 3.663 agências presentes em todos os estados brasileiros e em diversos municípios em funcionamento durante todo o ano de 2004 do Banco do Brasil S.A. O Banco em questão é um dos maiores bancos comerciais com R\$ 282 bilhões de ativos, R\$ 2,5 bilhões de intermediação financeira, R\$ 118,34 bilhões de carteira de crédito e com 36% de retorno sobre o P.L. em 2006. É uma empresa de economia mista, com atuação nacional e internacional, possui uma enorme carteira de clientes (24,1 milhões de correntistas), bem como um grande número de agências atuando no país e no exterior. Oferece uma diversidade muito grande de produtos e serviços customizados para cada segmento de clientes.

De uma forma geral segmenta, estrategicamente, seus clientes em:

- Pessoa física - conforme o nível de relacionamento (chamado de NR) atende basicamente três segmentos – NR1, NR2 e NR3 (com 3.552 observações).
- Pessoa jurídica – conforme o tamanho da empresa divide-se em *Corporate* e *Empresarial* (com 73 observações).
- Governo – Divide-se em três segmentos conforme a atuação de cada cliente: executivo, legislativo e judiciário (38 observações).
- Internacional – Com diversas agências e escritórios de representação em muitos países do mundo¹⁷.

Como membro do Sistema Financeiro Nacional e regido pelas normas do Banco Central do Brasil todas as agências estudadas tem contabilidade própria e oferecem um conjunto padrão de produtos e serviços diferenciados apenas nas particularidades e especificidades das características de cada cliente, isto é: produtos de captação (conta corrente, poupança e outras), aplicação (empréstimos, cheques especiais, cdc, linhas de crédito em geral), captação em fundos (de investimento, renda fixa, renda variável e ações) e taxas e serviços (inclusive cartões).

¹⁷ Essas agências e escritórios, pela sua diversidade e particularidade característica de cada país sede, não serão contempladas no estudo.

Os custos para manter uma agência em funcionamento são contabilizados na conta “Despesas Gerais” que contempla todas as despesas, inclusive de funcionários.

Portanto, dependendo da base de clientes, cada agência é organizada para servir diferentes tipos de negócios adequados aos segmentos acima identificados. Tanto o processo de produção quanto o ambiente no qual a agência está inserida afetam sua eficiência (conforme observado na figura 1.1 do capítulo 1)

Alguns municípios comportam diversas agências, e muitas delas apresentam particularidades no que tange a tipos de clientes e de produtos de maior destaque segundo suas próprias características. No entanto, conforme a metodologia, as variáveis que compõem os *inputs* e *outputs* são iguais para todas as unidades de decisão (agências). Existe mudança, evidentemente, nos seus valores e na dimensão e escala destes valores.

3.1 – DADOS E ESTATÍSTICAS DESCRITIVAS

Nessa metodologia a escolha de variáveis é uma fase extremamente delicada. Neste caso, foi considerado o fato de que as agências possuem uma cultura de trabalho influenciada pela estratégia da empresa. Este conjunto de variáveis corresponde ao modelo vigente de avaliação das unidades. Para composição das variáveis foram selecionados resultados de subitens orçamentários pelo período de um ano (2004). Esses resultados foram agregados e somados de acordo com o item a que pertencia, mês a mês, dentro do período estudado. Posteriormente, foi realizado o cálculo da média dos 12 meses para cada variável, chegando a um único valor para cada agência. Apesar da técnica DEA não exigir que os *inputs* e *outputs* sejam transformados em uma única unidade de medida (Silva e Azevedo, 2004), os dados foram transformados numa escala de 0 a 1 para preservar a privacidade das informações das agências. Inicialmente se utilizou sete variáveis: duas como *inputs* e cinco como *outputs*. Na tabela 3.1 a seguir são apresentados seus conceitos e definições:

Tabela 3.1 – Variáveis de *Inputs* e *Outputs* das 3663 agências bancárias¹⁸

Variáveis	Definição	Medidas de Escala
INPUT		
Despesas Administrativas	Todas as despesas necessárias para o funcionamento de uma agência, inclusive despesas de funcionários.	Valores entre 0 e 1. *
Homens/hora	Total de utilização de mão de obra por hora de trabalho em uma semana	Valores entre 0 e 1. *
OUTPUT		
Aplicação	Volume médio de aplicações da agência na sua praça, ponderado pela qualidade do crédito (contratos cumpridos)	Valores entre 0 e 1. *
Captação	Volume médio captado pela agência entre conta corrente, poupança, RDB, CDB.	Valores entre 0 e 1. *
Fundos	Volume médio captado pela agência em Fundos de renda fixa e variável.	Valores entre 0 e 1. *
Tarifas e Serviços	Volume médio de pagamentos de tarifas e serviços prestados pela agência	Valores entre 0 e 1. *
Média de quantidade de contas	Quantidade de contas de clientes pessoa física, pessoa jurídica e governo	Valores entre 0 e 1. **

* Valores entre 0 e 1, onde o valor 1 é o maior desta variável entre as 3663 agências. Isto é:

$$\text{valor}_{DMU} = \text{variável}_{DMU} / \text{variável}_{MAX}$$

** Valores entre 0 e 1 de três variáveis (quantidade de contas pessoa física, jurídica e clientes do governo). Após encontrar os valores entre 0 e 1 para cada uma das três variáveis aplicou-se a média aritmética entre elas¹⁹. Todos os valores foram apurados pela média do ano de 2004.

¹⁸ Foi utilizada a Regra Categórica conforme Yeh (1996) comentada no capítulo 1 página 4.

¹⁹ Como quase todas as agências possuem clientes pessoa física (pf), pessoa jurídica (pj) e governo a utilização do *output* com essa transformação atenua o problema da tipificação do cliente.

As variáveis de *input* e *output* foram exaustivamente discutidas e convenientemente escolhidas com a intenção de apurar o desempenho financeiro das unidades do BB. Portanto, a utilização da Regra Categórica foi a que mais se ajustou aos objetivos inicialmente definidos para a empresa que, para a finalidade em questão, define o número de variáveis adequadamente. Segundo Simar e Wilson (2007), com aplicação em dois estágios para grandes amostras, o número de variáveis finais deve ser suficientemente limitado. Na mesma linha de pesquisa, utilizando dois estágios em grandes amostras geradas aleatoriamente por processo de Monte Carlo, Souza e Staub (2006) analisam a inferência das medidas de eficiência DEA e concluem pela ineficácia na utilização de grande número de variáveis representativas de *inputs* ou *outputs*.

Cabe destacar que foram testados outros critérios de transformação das medidas de escala, além do critério aplicado (variáveis transformadas no intervalo de 0 e 1). Aplicou-se, também, logaritmo nas variáveis originais. Esse procedimento não logrou sucesso pela dificuldade de transformação aos dados originais, bem como resultados de eficiência com pequena diferença entre as DMU's. Isto é, os valores apurados de θ (índice de eficiência) entre as unidades eram inconvenientemente próximos, de tal sorte que não se pode apurar ou diferenciar graus de eficiência relativa e significativa entre as unidades de decisão. Já com a transformação utilizada os resultados foram idênticos às variáveis originais.

Algumas estatísticas foram geradas para melhor entendimento e conhecimento das distribuições de cada variável. A Tabela 3.2 apresenta as estatísticas descritivas de cada variável:

Tabela 3.2: Estatística Descritiva

	N	Mínimo	Máximo	Média	Desvio Padrão
Despesas	3663	.020466	1.000000	.169136	.130074
Homens/hora	3663	.018767	1.000000	.148448	.121048
Aplicação	3663	.000000	1.000000	.008320	.038947
Captação	3663	.000016	1.000000	.006040	.027245
Fundos	3663	.000000	1.000000	.001709	.021091
Tarifas	3663	.000616	1.000000	.039720	.053247
Média contas	3663	.002376	.398502	.069387	.054601
Valid N (listwise)	3663				

Pela Tabela 3.2 pode-se observar que a variável Despesas é a que apresentou maior média relativa, sendo a variável captação de Fundos a de menor média. As maiores amplitudes são verificadas nas variáveis Aplicação e Fundos. Todas as variáveis apresentaram alta variabilidade. Pelas características das distribuições percebe-se a dimensão das amplitudes das variáveis, indicando unidades altamente heterogêneas.

Tabela 3.3: Correlação entre as variáveis

	Despesas	Homens/h	Aplicação	Captação	Fundos	Tarifas	Contas
Despesas	1	.978(**)	.245(**)	.294(**)	.103(**)	.686(**)	.862(**)
Homens/hora		1	.227(**)	.274(**)	.102(**)	.683(**)	.876(**)
Aplicação			1	.263(**)	.155(**)	.534(**)	.046(**)
Captação				1	.304(**)	.370(**)	.104(**)
Fundos					1	.456(**)	.015
Tarifas						1	.481(**)
Média contas							1

** Correlação significativa no nível de 0.01.

Na Tabela 3.3 é apresentada análise de correlação das variáveis que compõem a base de dados do modelo DEA. Convém destacar que altas correlações existem entre as variáveis de *input* e destas com as variáveis de *output* Tarifas e Contas, o que pode ser explicado pela necessidade de administrar um número maior de contas (onde se cobram as tarifas) e, conseqüentemente, há uma maior movimentação de recursos. Isto, provavelmente, pode contribuir para a importância das variáveis na modelagem. É o que será testado posteriormente colocando e retirando as variáveis de *output* (principalmente a variável “média do número de contas de clientes”) nos modelos.

A alta correlação entre as variáveis de *input* (Despesas Administrativas e quantidade de Homens/hora trabalhada) confirma o peso e a importância do custo de pessoal em instituições financeiras onde predomina a atividade de serviço. Isto não altera o desempenho do modelo visto que se trata de cálculo de medidas não paramétricas.

3.2. RESULTADOS DOS VALORES DE θ :

Nesta seção serão discutidos os resultados da aplicação do modelo DEA de eficiência comentados no capítulo 2 e os resultados das aplicações dos procedimentos do

modelo *Jackstrap*, utilizando as informações e dados apresentados²⁰. No Apêndice 7 estão representados exemplos da função de Programação Linear do DEA, bem como um modelo de parametrização em arquivo texto para os cálculos e estimativas apresentados no capítulo anterior e aplicados neste capítulo.

O valor θ representativo da eficiência comparativa, tais como os *leverages* são calculados para todas as 3.663 agências usando os dois modelos CCR (ou CRS) com 5 *outputs* e CRS com 4 *outputs* do qual é retirada a variável de *output* representativa do número de clientes. Devido ao tamanho da amostra e muitas DMU's terem inexpressivos impactos (*leverages*) na eficiência geral, serão mostrados os resultados das 30 agências (Tabela 3.4) que provocam maior impacto no cálculo da eficiência comparativa (por ordem dos maiores valores dos *leverages* no modelo CRS – 5 *outputs*).

Observando a Tabela 3.4, a primeira coluna identifica o nome da agência, a segunda o indicador de eficiência θ , a coluna seguinte (*N Hits*) o número de vezes que a remoção daquela DMU produz impacto diferente de zero (variância diferente de zero) na amostra de *Total Hits*, que corresponde ao total de vezes que a agência é escolhida no procedimento *Jackstrap*. O *leverage* aparece na quinta coluna. Este valor já está ponderado pela razão *N Hits / Total Hits*. Todas as agências têm a mesma chance de serem escolhidas. A coluna *Rank* do modelo CRS esta representada na sexta coluna, as demais colunas (da 7^a. a 11^a.) obedecem a mesma ordem do CRS com 5 *outputs* para o modelo CRS com 4 *outputs*.

²⁰ O *software* utilizado para o cálculo das eficiências foi desenvolvido em linguagem “C” pelo Prof. Dr. Borko D. Stosic da Universidade Federal Rural de Pernambuco com auxílio da Profa. Dra. Maria da Conceição Sampaio de Sousa da UnB. Maiores detalhes poderão ser vistos no Apêndice 7.

Tabela 3.4

Eficiência e Leverage para 30 agências selecionadas: CRS - 5 outputs e CRS - 4 outputs

Agência	CRS – 5 outputs					CRS - 4 outputs				
	Eficiência	N Hits	Total Hits	Lev	Rank	Eficiência	N Hits	Total Hits	Lev	Rank
S.PUBLICO BRASILIA	1.0000	80	94	0.1666	1	0.2862	53	94	0.0125	35
VERA CRUZ	1.0000	86	106	0.1652	2	0.2630	15	106	0.0026	60
SANTA CRUZ DO SUL	0.7257	86	112	0.0809	3	0.2333	10	112	0.0021	64
EMPRES.CAMPINAS	0.6785	84	110	0.0582	4	0.1874	0	110	0.0000	148
GOVERNO BRASILIA	1.0000	93	103	0.0512	5	1.0000	93	103	0.0818	1
MARES	1.0000	83	89	0.0460	6	0.1635	83	89	0.0000	130
CORPORATE RJ	1.0000	95	104	0.0445	7	1.0000	95	104	0.0749	2
CORPORATE AV.PLTA.	1.0000	95	111	0.0350	8	1.0000	92	111	0.0614	4
CORPORATE DF	0.8480	53	92	0.0339	9	0.7753	0	92	0.0556	5
SINIMBU	0.5981	94	98	0.0314	10	0.1439	1	98	0.0000	211
GOVERNO SAO PAULO	1.0000	68	104	0.0293	11	1.0000	0	104	0.0662	3
ITAIM BIBI	1.0000	90	99	0.0278	12	1.0000	90	99	0.0272	15
VICENTE MACHADO	0.6110	98	114	0.0275	13	0.0827	97	114	0.0000	1392
CORPORATE PETROLEO	0.9379	57	80	0.0271	14	0.9162	0	80	0.0338	11
CORPORATE RS	1.0000	89	96	0.0247	15	1.0000	89	96	0.0399	7
GOVERNO R.DE JANEIRO	0.6555	110	126	0.0235	16	0.6359	109	126	0.0386	9
CORPORATE MG	1.0000	56	84	0.0231	17	1.0000	0	84	0.0389	8
PONTAL	0.5718	65	86	0.0230	18	0.1022	65	86	0.0000	534
SHOP.LITORAL NORTE	0.6290	89	99	0.0205	19	0.1099	83	99	0.0000	420
CORPORATE SAO JOAO	0.8619	85	116	0.0204	20	0.8568	89	116	0.0405	6
CORPORATE ABC	0.9954	59	105	0.0192	21	0.9954	0	105	0.0283	14
CORPORATE BA	0.5492	48	73	0.0188	22	0.5393	48	73	0.0345	10
P.JUDIC.SAO PAULO	0.9740	73	96	0.0184	23	0.9740	77	96	0.0189	23
CAAPORA	0.5842	52	100	0.0180	24	0.0801	0	100	0.0000	1588
P.JUDIC.BRASILIA	1.0000	60	101	0.0167	25	1.0000	0	101	0.0182	24
VENANCIO AIRES	0.5853	91	101	0.0164	26	0.1401	83	101	0.0000	224
EUSEBIO	0.5900	30	88	0.0156	27	0.0957	0	88	0.0000	711
S.PUBLICO SAO LUIS	0.6710	80	105	0.0154	28	0.6190	84	105	0.0326	12
CORPORATE S.J.CAMPOS	0.6046	52	89	0.0141	29	0.6018	51	89	0.0256	17
GOVERNO CURITIBA	0.5522	63	115	0.0139	30	0.5477	0	115	0.0196	21

Aquela unidade que apresentar grande quantidade em “N Hits” relativamente ao “Total Hits” e, ao mesmo tempo (como é de se esperar), maiores valores na medida de impacto ou *leverage* representam um potencial *outliers*.

Identificam-se, nos resultados da Tabela 3.4, que das 30 agências consideradas na linha de maior Rank do CRS (com 5 outputs), 19 delas correspondem ao conjunto de agências atendidas, segundo a estratégia do Banco, pelo segmento de pessoa jurídica (a característica é que iniciam com o nome *Empresarial* ou *Corporate*) ou governo (chamada

também de *pública, Setor Público* ou *Governo*). Destaca-se que estes dois segmentos, normalmente, trabalham com pequena quantidade de funcionários altamente especializados no atendimento personalizado dos clientes, administrando e atendendo enormes empréstimos e contas de captação como contas corrente e fundos de investimento. Estranheza poderia ser causada para as agências consideradas de varejo (ou do segmento *Pessoa Física*) e que trabalham quase que exclusivamente com clientes pessoa física, micro e pequenas empresas e prefeituras de pequenas cidades, e estão entre as de maior impacto. Nesta situação, segundo a Tabela 3.4, temos Vera Cruz (RS), Santa Cruz do Sul (RS), Sinimbu (RS), Venâncio Aires (RS), Vicente Machado – Ponta Grossa (PR), Mares – Salvador (BA), Shopping Litoral Norte – Lauro Freitas (BA), Caapora (PB), Eusébio (CE), Itaim Bibi (SP) e Pontal (SP) que são classificadas no segmento de varejo e, como foi dito acima, atendem, na sua grande maioria, micro, pequenas empresas e clientes pessoa física.

Aquelas que se destacam no modelo de eficiência por trabalharem com grande quantidade de clientes e grandes aplicações proporcionalmente ao seu tamanho foram: Vera Cruz (RS), Santa Cruz do Sul (RS), Sinimbu (RS), Venâncio Aires (RS) e Vicente Machado (PR) porém, em contra partida, utilizaram recursos de *input* limitados. As cidades do interior do Rio Grande do Sul obedecem esta característica por concentrarem, tradicionalmente, seus negócios com este Banco. A agência de Itaim Bibi (SP) chama atenção pela grande quantidade e volume de operações de captação. As agências Eusébio (CE), Pontal (SP), Sinimbu (RS), Vicente Machado (PR) e Caapora (PB) estão, no modelo CRS, entre as 30 primeiras de maior impacto devido a influência da variável de *output* que mede o número de contas de clientes de cada unidade. Neste caso, embora com pouca eficiência, algumas agências passam a dominar, na fronteira de produção, um número considerável de outras agências o que as posiciona num elevado *Rank* pelos valores do *leverage*. Calculando a matriz de dominância, pelo modelo FDH, tem-se para algumas destas agências (quatro de maior poder dominante) os resultados apresentados na Tabela 3.5 abaixo:

Tabela 3.5 – Dominância por FDH

Agências dominantes	# de Agências dominadas
Vera Cruz (RS)	360
Sinimbu (RS)	206
Pontal (SP)	196
Eusébio (CE)	62

Como exemplo, algumas agências dominadas pela agência Vera Cruz (RS) são: Codó, Crateús, Quixadá (CE), Pedreiras, Bragança (SP), Goiana (GO) e Januária (MG). Estas agências são muito semelhantes à Vera Cruz, exceto por alguma variável de *input* ou *output* com valores maiores ou menores respectivamente, o que as faz estarem próximas as suas dominadas num particular limite imaginário da fronteira de produção. Este comportamento é replicado para as demais agências dominantes e suas dominadas, considerando que uma unidade dominante pode perfeitamente assumir, em determinado contexto de relacionamento, simultaneamente e com outra unidade, o papel de dominada.

As outras agências destacam-se com impacto ou *leverage* nas demais DMU's pelo fato de utilizarem poucos recursos (*inputs*) transformando-os em valores de *output* (principalmente nas aplicações, empréstimos e ganhos com tarifas) acima da média, mas não apresentaram características de dominância na fronteira de produção.

Outra análise pode ser feita considerando apenas quatro variáveis de *output* (representadas nas últimas cinco colunas da Tabela 3.4), isto é, retirando a variável “média da quantidade de contas de clientes” (CRS com 4 *outputs*). Esta tentativa se justifica pela hipótese, defendida por alguns especialistas, de que o número de contas não necessariamente provoca ou contribui com o conceito de eficiência, uma vez que a quantidade normalmente provoca pouca qualidade na seleção do cliente. Se considerarmos as agências do segmento de varejo, onde a base de seus clientes é de micro e pequenas empresas, bem como clientes pessoa física, esta hipótese torna-se discutível. De qualquer forma, este novo resultado para as 30 agências de maior impacto na eficiência, inclui apenas a agência de Itaim Bibi (SP) na categoria de agência de varejo. O resultado não é surpresa, visto que ela está localizada em local privilegiado na cidade de São Paulo e atende

clientes especiais que elevam o índice de captação²¹. As demais são agências do segmento governo e pessoa jurídica. Comparando as duas colunas de *Rank* (CRS – 5 *outputs* e CRS – 4 *outputs*) observa-se que nenhuma agência do segmento pessoa física, exceto Itaim Bibi – SP, está posicionada entre as 30 primeiras. Estes resultados atestam a importância da variável “média da quantidade de contas de clientes” para definir a estratégia das agências do segmento de varejo e que trata basicamente de clientes pessoa física.

Destaque pode se dar à mudança de *Rank* na Tabela 3.4. Confirma-se o fato de que existem alterações significativas de posicionamento da importância para medida de impacto quando da redução de um *output*. Alguns exemplos podem ser notados: 1) a agência do segmento Governo *Setor Público de Brasília*, 1ª. no *ranking* de eficiência com $\theta = 1$ e $l = 0,1666$ (*leverage*), muda seu posicionamento para 35ª. com $\theta = 0,2862$ e $l = 0,0125$; 2) outro caso interessante trata-se da agência do segmento Pessoa Física *Vicente Machado (PR)* na 13ª. posição com $\theta = 0,6110$ e $l = 0,0275$, quando retirado o *output* modifica-se a posição para 1392ª. com $\theta = 0,0827$ e $l = 0$. Portanto, existe uma mudança visualmente significativa no posicionamento após a retirada de um *output*. Mudanças semelhantes ocorrem, principalmente, com agências do segmento Pessoa Física o que reforça o fato de que a variável retirada (média da quantidade de contas de clientes) tem impacto com contribuição positiva para agências de menor expressão e de pouco investimento em insumos.

As hipóteses de impacto na eficiência pela variável “média da quantidade de contas de clientes” foram analisadas, com maior profundidade, em comparações dos segmentos do Banco, bem como quando da utilização da Regressão Quantílica para explicar as medidas de eficiência.

Sob as mesmas condições foram observados os resultados da aplicação do modelo DEA- BCC (ou VRS – Retorno Variável de Escala). Eles estão representados na Tabela 3.6 e, posteriormente, serão consolidados e comparados com o modelo CCR (Retorno Constante de Escala).

²¹ A agência é responsável por depósitos judiciais, ou seja, alvarás judiciais não executados e que, normalmente, são depositados em conta de poupança específica.

Tabela 3.6

Eficiência e *Leverage* para 30 agências selecionadas : VRS - 5 *outputs* e VRS - 4 *outputs*

Agência	VRS – 5 <i>outputs</i>					VRS - 4 <i>outputs</i>				
	Eficiência	N Hits	Total Hits	Lev	Rank	Eficiência	N Hits	Total Hits	Lev	Rank
S.PUBLICO BRASILIA	1.0000	80	94	0.1433	1	0.3373	53	94	0.0146	30
VERA CRUZ	1.0000	86	106	0.1334	2	0.4089	17	106	0.0031	78
SANTA CRUZ DO SUL	1.0000	96	112	0.0816	3	0.2840	13	112	0.0019	99
EMPRES.CAMPINAS	0.7804	92	110	0.0680	4	0.2262	0	110	0.0000	2263
GOVERNO BRASILIA	1.0000	93	103	0.0508	5	1.0000	93	103	0.0709	1
CORPORATE RJ	1.0000	83	89	0.0449	6	1.0000	83	89	0.0596	2
CORPORATE AV.PLTA.	1.0000	95	104	0.0359	7	1.0000	95	104	0.0510	4
CORPORATE DF	0.8673	96	111	0.0344	8	0.7976	92	111	0.0451	5
AMERICA DOURADA	1.0000	85	92	0.0331	9	1.0000	62	92	0.0013	115
MARES	1.0000	94	98	0.0328	10	0.9240	71	98	0.0005	148
VENANCIO AIRES	0.7275	76	104	0.0326	11	0.1934	1	104	0.0000	2632
GOVERNO SÃO PAULO	1.0000	90	99	0.0288	12	1.0000	82	99	0.0525	3
CORPORATE PETROLEO	0.9462	98	114	0.0283	13	0.9304	97	114	0.0321	9
VICENTE MACHADO	0.6181	58	80	0.0267	14	0.1875	0	80	0.0000	2693
ITAIM BIBI	1.0000	90	96	0.0267	15	1.0000	89	96	0.0261	13
CORPORATE RS	1.0000	110	126	0.0264	16	1.0000	109	126	0.0363	6
SINIMBU	0.7192	65	84	0.0262	17	0.3884	3	84	0.0001	199
GOVERNO R.DE JANEIRO	0.6654	66	86	0.0248	18	0.6383	65	86	0.0308	11
CORPORATE MG	1.0000	89	99	0.0239	19	1.0000	86	99	0.0338	8
CORPORATE SAO JOAO	1.0000	88	116	0.0232	20	1.0000	84	116	0.0349	7
CAAPORA	0.8493	85	105	0.0215	21	0.5039	0	105	0.0000	542
CORPORATE BA	0.5935	49	73	0.0192	22	0.5935	50	73	0.0311	10
CORPORATE ABC	0.9978	80	96	0.0191	23	0.9978	83	96	0.0259	14
CAMARA CASCUDO	0.5991	57	100	0.0189	24	0.1775	0	100	0.0000	2800
PONTAL	0.6523	69	101	0.0181	25	0.2842	0	101	0.0000	1620
P.JUDIC.SAO PAULO	0.9964	91	101	0.0178	26	0.9964	85	101	0.0174	23
VILA MILITAR	0.5800	50	88	0.0177	27	0.1408	0	88	0.0000	3281
P.JUDIC.BRASILIA	1.0000	84	105	0.0162	28	1.0000	85	105	0.0173	24
CORPORATE PR	0.5118	52	89	0.0158	29	0.4950	51	89	0.0226	16
EUSEBIO	0.6470	70	115	0.0151	30	0.2500	0	115	0.0000	1991

No caso do modelo VRS o número sobe de 11 para 13 agências de atendimento do varejo no total de 30 agências de maior *leverage*. O modelo privilegia, colocando nas 30 primeiras posições, ordenada na tabela pelo indicador de impacto *leverage*, a agência *Corporate* Paraná e outras três agências do segmento Pessoa Física: América Dourada (BA), Câmara Cascudo - Natal (RN), Vila Militar – Rio de Janeiro (RJ). Desta forma, substitui, na classificação do modelo CRS da Tabela 3.4, as agências *Corporate* São José

dos Campos (SP), Governo Curitiba (PR), Setor Público São Luis (MA) e Shopping Litoral Norte – Lauro Freitas (BA). Como retorno variável de escala (VRS), o modelo passa a restringir a propriedade de raio ilimitado e considera a possibilidade de rendimentos crescentes e decrescentes de escala na fronteira eficiente, resultando num maior número de unidades eficientes (capítulo 2), portanto alterando as posições de algumas DMU's e reposicionando com índice maior de eficiência muitas agências do segmento varejo. Não considerando a variável de *output* “média da quantidade de contas de clientes” (com apenas 4 variáveis de *output*) a alteração torna-se menos significativa e, para as 30 agências de maior *leverage* comparativamente com o CRS, existe pequena mudança e apenas nos segmentos Governo e Pessoa Jurídica, entrando na relação as agências Setor Público Brasília (DF) de 1^a. para 30^a. colocada e *Corporate* (PR) de 29^a. para 16^a. colocada.

Na seção seguinte serão feitas análises dos resultados, utilizando as distribuições identificadas no capítulo 2 como técnica para identificação e, se for o caso, eliminação de *outliers* nos dados.

3.3. IDENTIFICAÇÃO DOS *OUTLIERS* E ANÁLISE DAS DISTRIBUIÇÕES PARA OS MODELOS CRS E VRS:

Na Figura 3.1(histograma geral com a frequência de θ e curva normal com todos os valores da amostra) é mostrada, através do histograma do indicador de eficiência, a influência que os *outliers* causam em todos os estimadores DEA. Vale ressaltar que o conjunto de dados está inalterado e com 3.663 unidades (agências). Percebe-se pelos resultados que a distribuição tende a ser normal e simétrica, apenas é claro o destaque na cauda direita nas condições em que onze unidades apresentam padrão de eficiência com $\theta=1$, como esperado na presença de poucos casos de *outliers*.

No conjunto das figuras 3.2 (histograma das curvas das distribuições com a presença da curva normal) são apresentados os resultados das distribuições que selecionam e eliminam as DMU's consideradas *outliers*.

A Tabela 3.7 retrata as alterações de θ quando da aplicação das distribuições e eliminação dos *outliers*.

Figura 3.1– Histograma Geral – CRS 5 outputs

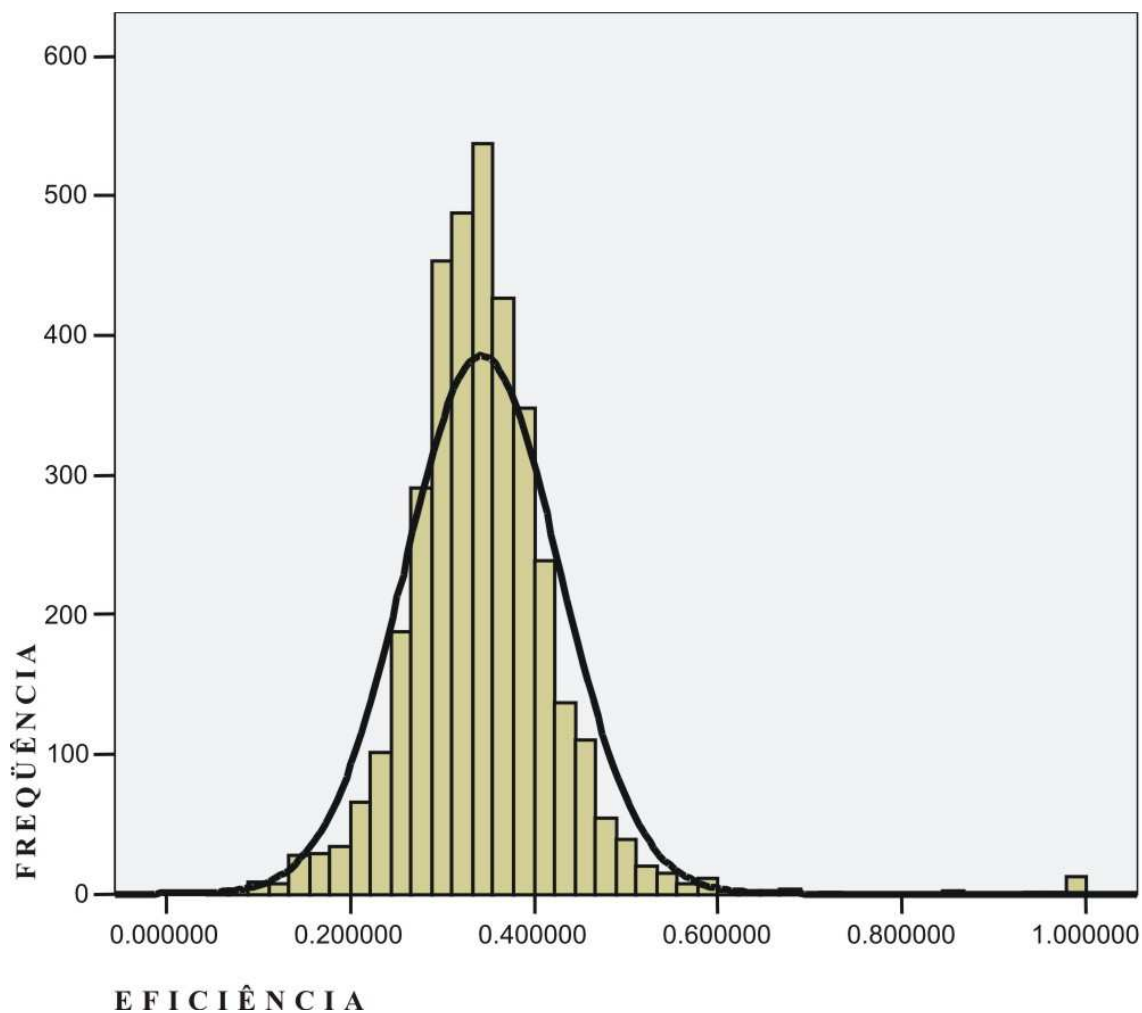


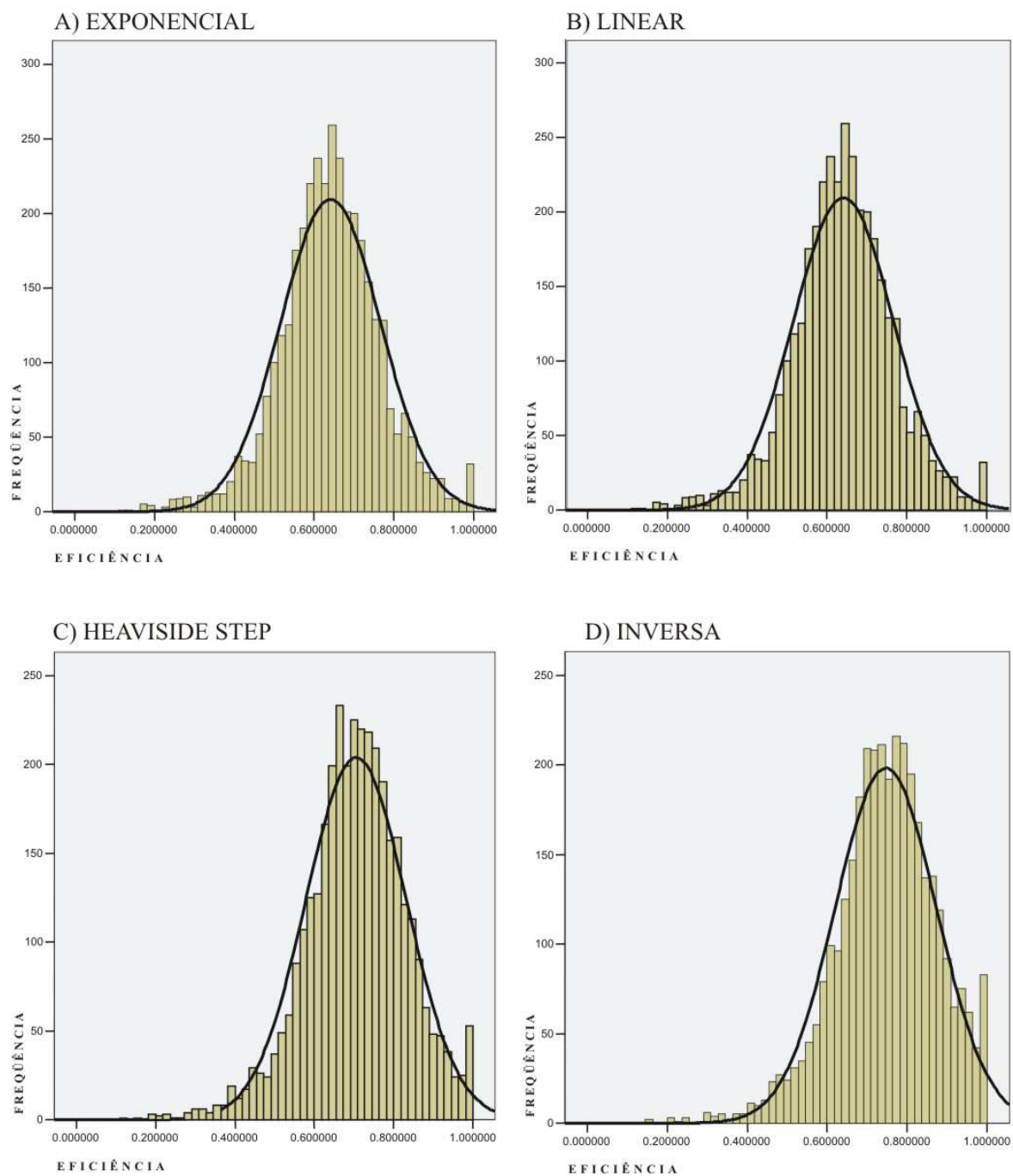
Tabela 3.7 – Análise descritiva dos valores de θ para 5 variáveis de output - CRS

Distribuição	Média	Desv-padrão	CV(%)	Assimetria	Amostra
Exponencial*	0,64123528	0,12530664	19,54	-0,157	3616(47)
Heaviside step	0,70550839	0,126867402	17,98	-0,381	3564(99)
Inversa	0,74670240	0,126749175	16,97	-0,530	3463(200)
Geral	0,34325273	0,084273606	24,55	1,95	3663(0)

Os valores em parênteses correspondem ao número de DMU's removidas da amostra.

* As DMU's selecionadas com a distribuição Linear foram idênticas as da distribuição Exponencial.

Figura 3.2– Distribuição de θ - CRS 5 outputs



Nota-se que a eliminação de um maior número de observações de considerável impacto (maior valor de *leverage*) provoca, na distribuição de θ , mais homogeneidade relativa com diminuição do coeficiente de variação (CV), embora carregue maior peso na sua cauda esquerda invertendo sua assimetria e concentrando seus resultados à direita e acima da média (coeficiente de assimetria negativo). Destaca-se que o motivo da diminuição do coeficiente de assimetria deve-se a eliminação de DMU's de alto impacto e conseqüente aproximação de uma considerável quantidade de valores de $\theta = 1$. Numa linguagem mais objetiva, aumentou-se o número de agências com maior grau de eficiência relativa quando se retirou os casos que apresentaram condições particulares e muitas vezes excepcionais (*outliers*). Trata-se de agências de segmentos de clientes especiais ou que trabalham com grandes aplicações e empréstimos provocando distorção nos demais segmentos. Como se destacou anteriormente, utilizando como exemplo a distribuição *Heaviside step*, a maioria (55 das 99 agências) trabalha com clientes do segmento pessoa jurídica de grandes empresas e contas do governo.

Tabela 3.8 – Análise descritiva dos valores de θ para 4 variáveis de *output* - CRS

Distribuição	Média	Desv-padrão	CV(%)	Assimetria	Amostra
Exponencial*	0,2652	0,1107	41,72	2,985	3612(51)
Heaviside step	0,3107	0,1172	37,74	2,395	3590(73)
Inversa	0,3942	0,1238	31,42	1,563	3551(112)
Geral	0,0887	0,0714	80,50	8,404	3663(0)

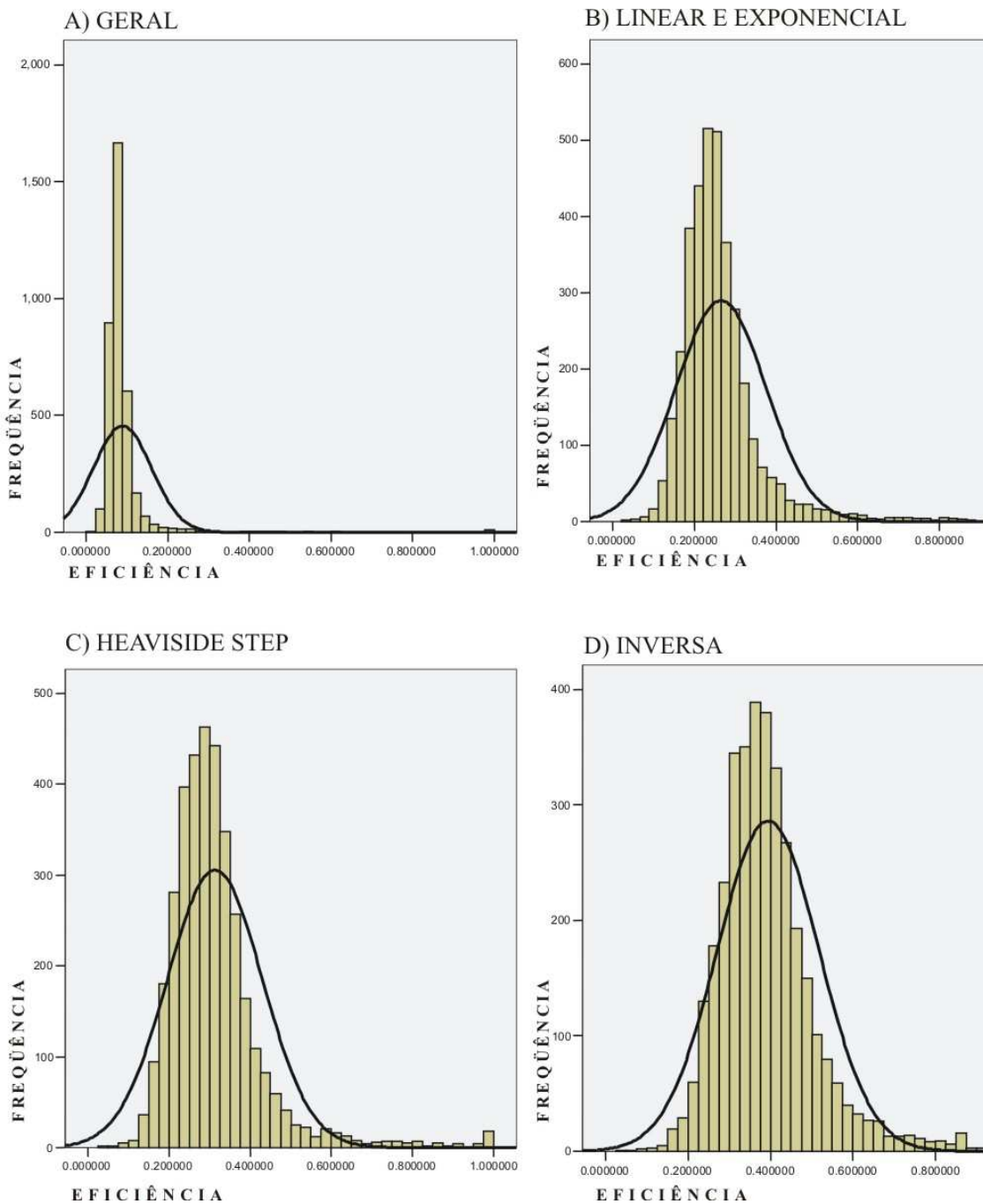
Os valores em parênteses correspondem ao número de DMU's removidas da amostra.

* As DMU's selecionadas com a distribuição Linear foram idênticas as da distribuição Exponencial.

Novamente na intenção de analisar melhor os resultados e o comportamento comparativo das unidades, foi retirada uma variável de *output* (média da quantidade de contas de clientes), cujos resultados estão representados na Tabela 3.8 acima. Observando a Figura 3.3(cálculo da eficiência θ com 4 outputs modelo CRS) pode-se perceber um maior distanciamento da média em relação às DMU's eficientes (este resultado deve ser observado também na comparação com a Tabela 3.7), provocando crescimento no grau de assimetria, bem como da variabilidade relativa. Um simples teste de hipóteses de diferença

entre duas médias para a variável eficiência, nestas condições, identifica uma diferença significativa entre as duas populações, de tal forma que a hipótese nula (H_0 : as duas populações são iguais) é rejeitada com qualquer nível de significância. Lembrando que este teste não avalia o ordenamento dos valores da variável eficiência ou *leverage*, teste necessário para apurar se houve mudança significativa no ranking de θ ou l , conforme sugere a Tabela 3.4, e que será confirmado na próxima seção.

Figura 3.3– Distribuição de θ - CRS 4 outputs



Portanto, principalmente nos segmentos de varejo, que corresponde aproximadamente a 97% das DMU's, a “média da quantidade de contas de clientes” é um fator decisivo e importante para o ganho de escala. A remoção desta variável de *output* provoca maior diferença e menor importância no impacto destas unidades na comparação diminuindo a média geral, porém aumentando a dispersão relativa entre os diversos graus de eficiência.

Esta abordagem reduz, sobremaneira, os *leverages* das DMU's, indicando que a variável de *output* comentada é de grande importância na medição relativa do impacto para eficiência. Ao retirá-la, muitas unidades perdem em eficiência relativa.

Conforme a Tabela 3.8 e observando as distribuições que identificam e removem os *outliers*, nota-se que houve uma diminuição da quantidade removida em quase todas as distribuições comparativamente a Tabela 3.7. Analisando especificamente a distribuição *Heaviside step* das 73 removidas apenas 13 foram identificadas como do segmento de varejo, onde a variável “média da quantidade de contas de clientes” é destaque. As agências que mais perdem e que, portanto, “beneficiaram-se” com essa variável são, respectivamente, no modelo CRS: Pontal (SP), Caapora (PB), Mares e Shopping Litoral Norte - Lauro Freitas (BA), Eusébio (CE), Vera Cruz e Santa Cruz do Sul (RS) só para citar algumas. Nota-se que todas foram identificadas na seção 3.2 (tabela 3.4) entre as 30 agências de alto impacto e influência relativa (*leverage*) pertencentes ao segmento de varejo.

Há de se considerar que, mesmo não influenciando diretamente na rentabilidade, a indicação de busca, cadastramento e participação de um número maior de clientes, é fundamental para iniciar uma venda de produtos, o que contribui favoravelmente para essa variável compor medidas de eficiência produtiva. Por outro lado, com a diferença significativa entre as medidas de eficiência, associada com a avaliação do tipo de agências que perdem eficiência quando se retira essa variável de *output* (no caso analisado grande parte são de agências localizadas em pequenos municípios), pode-se inferir quanto à contribuição socialmente positiva da instituição financeira nestas localidades (externalidade positiva), mesmo porque grande parte destas agências trabalha com resultados, senão negativos, muito pouco compensatórios, conforme foi comentado anteriormente.

A Tabela 3.9 faz uma comparação agregada dos resultados das duas distribuições CRS e VRS para 5 *outputs* antes e depois da aplicação da distribuição *Heaviside step*. Essa distribuição foi usada como instrumento de seleção de *outliers* das distribuições originais, conforme exposto no capítulo 2. O motivo desta escolha é pela facilidade de seleção das unidades, bem como pelo critério mostrado de robustez dos seus resultados. Como dito anteriormente as demais distribuições apresentam critérios arbitrários na seleção de *outliers*²².

Tabela 3.9 – Análise descritiva dos valores de θ para 5 *outputs* – CRS e VRS

Distribuição	Média	Desv-padrão	CV(%)	Assimetria	Amostra
Geral (CRS)	0,3433	0,0843	24,55	1,95	3663(0)
Geral (VRS)	0,4525	0,1334	29,49	1,567	3663(0)
Heav. (CRS)	0,7055	0,1269	17,98	-0,381	3564(99)
Heav. (VRS)	0,7541	0,1212	16,07	-0,273	3507(156)

Os valores em parênteses correspondem ao número de DMU's removidas da amostra.

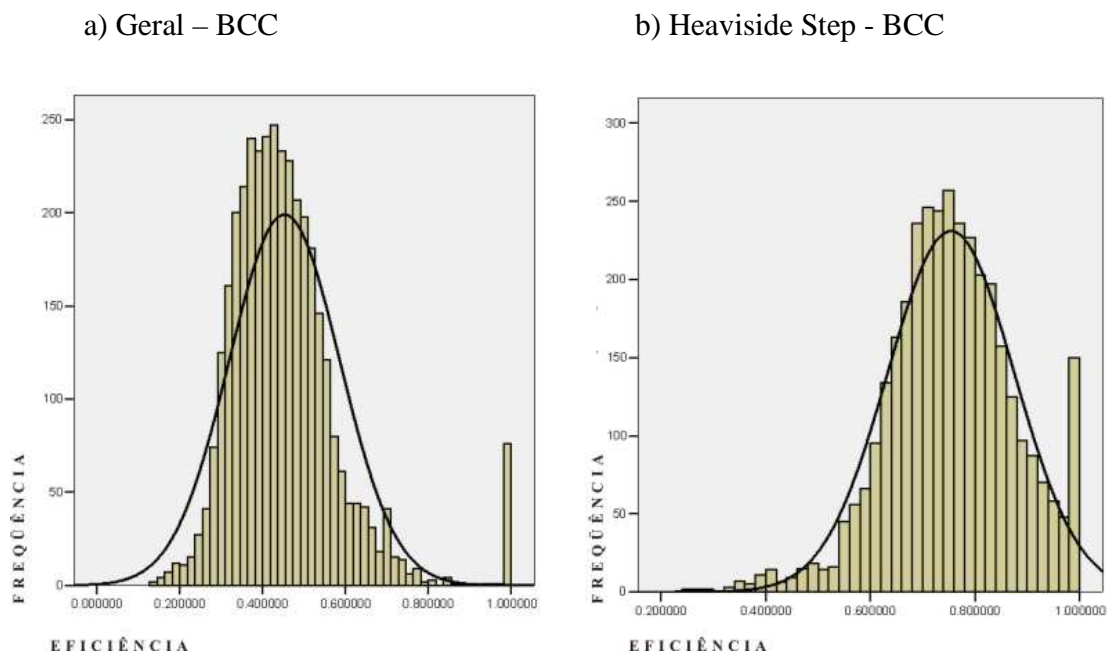
A distribuição com retorno variável de escala (VRS) apresenta uma média maior, bem como menor coeficiente de assimetria. No entanto, cresce ligeiramente seu grau de dispersão relativa de variação no modelo Geral. Para um simples teste de hipóteses existe uma variação significativa entre as amostras, o que quer dizer uma diferença significativa (a qualquer nível de significância) na medida de eficiência utilizando o VRS comparado à medida de Retorno Constante de Escala (CRS). Como se trata de medidas relativas é de se esperar que uma “correção” na medida de variação de escala aproxime da eficiência algumas unidades, ocasionando aumento no índice θ com a aproximação de um número maior de unidades à fronteira de produção eficiente. A diminuição, embora pequena, do índice de assimetria mostra uma aproximação melhor da distribuição do grau de eficiência VRS à curva normal conforme se visualiza pelos gráficos da Figura 3.4 comparativamente

²² A escolha de 0,95 nas distribuições reduz (e.g. 0,90) ou aumenta (e.g. 0,98) o número de *outliers* da amostra.

a Figura 3.2, embora tenha aumentado o número de agências eficientes “deformando” a característica gráfica de normalidade da distribuição.

Através dos cálculos para encontrar os *leverages* e procurando utilizar a distribuição *Heaviside step* para solucionar o problema dos *outliers*, chegam-se aos resultados apresentados na Tabela 3.9. Isto é, a retirada de 156 observações de *leverage* onde 105 são do segmento varejo e apenas 51 agências são dos segmentos Governo e Pessoa Jurídica. Isto ratifica as observações feitas anteriormente sobre o retorno variável de escala, confirmando o modelo CRS como mais adequado para discriminar os segmentos do Banco. A Figura 3.4 abaixo mostra o gráfico das duas distribuições e suas aproximações à curva normal. Novamente, com a retirada das 156 observações e analisando o gráfico e a tabela, a distribuição ajusta-se melhor à normal muito embora considerando as mesmas condições e restrições identificadas anteriormente a respeito do número de unidades eficientes.

Figura 3.4: Distribuição de θ - modelo VRS



Depois das constatações acima procura-se, através da Tabela 3.10, categorizar as agências segundo a quantidade de contas de clientes comparando o modelo DEA normal com o modelo usando a distribuição *Heaviside step*. Os resultados são extremamente interessantes conjugados às análises anteriores. O comportamento do índice de eficiência θ para os modelos DEA, principalmente para CRS – 4 *outputs* e VRS, são caracterizados por médias de valores de θ iniciais grandes (nas classes de até 7.000 clientes) decrescendo posteriormente. As agências das classes com poucas contas estão “contaminadas” com a presença entre elas das agências de segmentos especiais como Governo e Pessoa Jurídica que, como já comentado, utilizam em sua gestão modelo particularizado e diferenciado das demais unidades, personalizado aos clientes. Estas agências com altas médias de eficiência tornam a variância maior e alteram, significativamente, as medidas de variabilidade e assimetria. Ao ser aplicada uma técnica de seleção e eliminação de *outliers*, como a distribuição *Heaviside step*, a situação se inverte indicando, nos três casos, que agências com maior número de clientes apresentam maior grau de eficiência. Os indicadores do grau de eficiência passam a ser crescentes. Portanto, nestas unidades a influência da quantidade de clientes se faz notória e importante, indicando que essa variável é fundamental para o processo de gestão, contribuindo tanto para o ganho de escala, como que, se mal dimensionada, para perda de eficiência. Deve-se destacar que a perda de eficiência (quando retirada a variável “média do número de contas de clientes”) não necessariamente deve ser atribuída à quantidade “exagerada” de clientes, mas também, pela possível má gestão da relação *input/output* das demais variáveis.

A Tabela 3.11 compara os três modelos com classes de renda dos clientes, baseada na renda média dos clientes das agências. Neste caso o resultado observado do DEA para o modelo *Heaviside step* apenas altera a dimensão, aumentando as médias de eficiência após a eliminação dos *outliers*. No entanto, sem modificar a ordem crescente ou decrescente do comportamento dos modelos quanto ao grau de eficiência. Isto é, os modelos CRS e VRS, ambos com 5 *outputs*, diminuem o grau de eficiência relativa para agências de maior renda média que, diferentemente da situação anterior (Tabela 3.10), a tendência é acompanhada pelos modelos *Heaviside step* (que corrigem os *outliers*). Entretanto, quando retirado um *output* (média da quantidade de contas de clientes – modelo de 4 *outputs*), inverte-se a

ordem de crescimento e o grau de eficiência é correlacionado diretamente com a renda média.

Tabela 3.10 – Estatística Descritiva da eficiência θ para o DEA e a abordagem *Heaviside step* para os modelos CRS com 5 e 4 *outputs* e VRS com 5 *outputs* – *Comparação com quantidade de clientes*

Classes de Clientes	# Observações		Média		Desvio-padrão		Assimetria		Mínimo	
	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav
CRS – 5										
0-3000	869	812	0.3053	0.5964	0.1198	0.1399	2.9313	0.0958	0.0615	0.1273
3001-5000	937	931	0.3497	0.7112	0.0600	0.1024	0.2132	0.0891	0.1491	0.3556
5001-7000	597	589	0.3629	0.7513	0.0646	0.1000	0.4323	0.0176	0.1597	0.4019
7001-10000	520	510	0.3655	0.7654	0.0645	0.1044	0.4498	0.1140	0.2067	0.5170
10001-15000	396	388	0.3481	0.7391	0.0725	0.1034	2.4875	0.3224	0.1728	0.4260
15001-20000	175	173	0.3375	0.7341	0.0503	0.0891	1.0609	0.5566	0.2257	0.5721
>20000	169	161	0.3598	0.7541	0.0842	0.0815	3.6271	0.3725	0.2266	0.5561
CRS – 4										
0-3000	869	806	0.1026	0.2760	0.1343	0.1548	4.5958	2.5775	0.0081	0.0357
3001-5000	937	933	0.0794	0.2938	0.0287	0.0985	3.8119	2.3615	0.0302	0.1228
5001-7000	597	596	0.0858	0.3278	0.0276	0.1053	2.7726	2.4000	0.0429	0.1598
7001-10000	520	519	0.0863	0.3393	0.0257	0.1070	3.4114	3.0178	0.0477	0.1706
10001-15000	396	395	0.0881	0.3436	0.0498	0.0912	15.7175	3.7829	0.0550	0.2128
15001-20000	175	175	0.0880	0.3553	0.0248	0.1027	3.7042	3.3009	0.0396	0.1633
>20000	169	166	0.0890	0.3491	0.0250	0.0567	4.6523	2.8009	0.0583	0.2310
VRS – 5										
0-3000	869	783	0.5392	0.7097	0.1936	0.1518	0.7407	-0.2045	0.1301	0.2449
3001-5000	937	927	0.4728	0.7390	0.0879	0.1045	0.2425	-0.0403	0.1981	0.3865
5001-7000	597	590	0.4369	0.7627	0.0797	0.1016	0.4788	0.0262	0.1857	0.4097
7001-10000	520	509	0.4121	0.7731	0.0746	0.1049	0.4491	0.1361	0.2238	0.5322
10001-15000	396	387	0.3704	0.7738	0.0768	0.1066	2.1606	0.1432	0.1836	0.4199
15001-20000	175	169	0.3461	0.8052	0.0518	0.0907	1.2602	0.2598	0.2338	0.6142
>20000	169	142	0.3762	0.8792	0.1202	0.0888	3.6260	-0.3891	0.2266	0.6212

Este comportamento indica o quanto a renda é importante quando se considera, para o grau de eficiência, apenas variáveis financeiras. Conclui-se com isso que, provavelmente neste caso, a expansão da base de clientes não leva, necessariamente, maior eficiência financeira, principalmente nos segmentos Governo e Pessoa Jurídica. Posteriormente estas conclusões serão melhor fundamentadas no momento em que os segmentos do Banco serão analisados separadamente.

Tabela 3.11 – Estatística Descritiva da eficiência θ para o DEA e a abordagem Heavyside para os modelos CRS com 5 e 4 *outputs* e VRS com 5 *outputs* – comparação com Classes de Renda²³

Classes de Renda	# Observações		Eficiência Média de θ		Desvio-padrão de θ		Assimetria de θ		Mínimo de θ	
	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav
CRS – 5										
0-1000	2.269	2197	0,3618	0,7184	0,0851	0,1249	2,2793	-0,3464	0,0957	0,1870
1000-1500	952	948	0,3198	0,6924	0,0558	0,1107	0,3675	-0,6051	0,0779	0,1522
1500-2000	252	248	0,3028	0,6710	0,0802	0,1249	3,0177	-0,3164	0,0615	0,1273
2000-3000	111	104	0,2982	0,6695	0,1083	0,1918	1,4541	0,0127	0,0945	0,2210
>3000	66	61	0,2747	0,6483	0,1137	0,2018	0,5649	0,0532	0,0674	0,2304
CRS – 4										
0-1000	2.269	2.224	0,0813	0,2750	0,0753	0,0814	8,9790	2,6030	0,0081	0,0357
100-1500	952	950	0,0889	0,3466	0,0290	0,0981	7,9694	1,8883	0,0129	0,0442
1500-2000	252	251	0,1059	0,4056	0,0674	0,1477	9,5846	1,5939	0,0213	0,0707
2000-3000	111	102	0,1408	0,4802	0,1075	0,2220	3,6725	0,8906	0,0339	0,1250
>3000	66	58	0,1431	0,4944	0,0958	0,2109	1,9417	0,9418	0,0394	0,2107
VRS – 5										
0-1000	2.269	2.148	0,4939	0,7717	0,1340	0,1137	1,6258	-0,1888	0,1691	0,2522
100-1500	952	940	0,3909	0,7325	0,0858	0,1136	2,3593	-0,2267	0,1626	0,2692
1500-2000	252	248	0,3747	0,7131	0,1076	0,1236	1,5747	-0,2962	0,1643	0,2449
2000-3000	111	103	0,3784	0,7171	0,1412	0,1836	1,8149	-0,0274	0,1460	0,2846
>3000	66	62	0,3366	0,6930	0,1231	0,1975	0,6545	-0,0403	0,1301	0,2870

3.4 AVALIAÇÃO DAS DIFERENÇAS DOS MODELOS - APLICAÇÃO DE TESTES NÃO PARAMÉTRICOS²⁴:

Após esses resultados procurou-se estimar a distribuição da diferença da eficiência calculada com 5 variáveis de *output* (θ_5) e com 4 variáveis de *output* (θ_4) e testar estatisticamente essa diferença. Portanto, a estatística $\delta_i^j = \theta_{i5}^j - \theta_{i4}^j$ onde $i = 1, 2, \dots, 3663$ e j corresponde a variável de *output* retirada do modelo ($j =$ aplicação, capitação, fundos, tarifas ou média do número de contas de clientes), é quase sempre positiva. Essas distribuições auxiliam: a) na constatação da diferença da utilização e escolha entre os modelos CRS e VRS; b) na confirmação de que, com a retirada de um *output* no modelo DEA, provoca diminuição nas medidas do grau de eficiência, c) na importância de cada

²³ A média do grau de eficiência gerada por essa e outras tabelas semelhantes não consideram o cálculo computacional da eficiência por classes (clientes, renda e agências) e sim a média das eficiências daquela classe gerada por todas as agências (do modelo geral ou dos segmentos Governo, Pessoa Jurídica e Pessoa Física) que serão vistos posteriormente.

²⁴ Para a aplicação destes testes não paramétricos foi usado o *software* SPSS v.13.

output para o grau de eficiência relativa de cada unidade ou na média das unidades e d) na avaliação prática e administrativa da logística da instituição. Isto é, diferenças significativamente grandes, indicam poucos recursos de *input* para o atendimento de grande quantidade de *outputs*. Essa informação auxilia, tanto no remanejamento, quanto na contratação de novos funcionários ou de outros recursos produtivos contabilizados nas Despesas Administrativas (máquinas e equipamentos para transações bancárias). A Tabela 3.12 apresenta estatísticas descritivas de δ para o total da amostra de agências:

Tabela 3.12. – Análise descritiva dos valores de $\delta^j = \theta_3^j - \theta_4^j$

Output retirado (j)	Média	Desvio-padrão	CV(%)	Amostra
Clientes	0,254539	0,0781	30,68	3663 (CRS)
Clientes	0,156111	0,0781	50,05	3663 (VRS)
Tarifas	0,026350	0,0334	128,97	3663 (CRS)
Aplicação	0,003874	0,0269	693,57	3663 (CRS)
Captação	0,001396	0,0258	1.845,19	3663 (CRS)
Fundos	0,000056	0,0030	5.320,64	3663 (CRS)

Conforme a tabela acima, a diferença média de perda é de $\bar{\delta}^{clientes} = 0,254539$ e $\bar{\delta}^{clientes} = 0,156111$ na eficiência para os dois modelos (CRS e VRS, respectivamente) ao se retirar o *output* que identifica a quantidade de clientes por unidade. Para esta comparação existe uma diferença significativa entre as duas amostras, tanto para o teste paramétrico de *Student* como os testes não paramétricos dos sinais ou de *Wilcoxon* na diferença de duas amostras relacionadas. Portanto, confirmando a hipótese de que existe diferença significativa “favorável” ao modelo CRS onde $\bar{\delta}^{clientes} (CRS) > \bar{\delta}^{clientes} (VRS)$ permanecendo a influência do modelo VRS de considerar valores maiores de eficiência relativa, mesmo retirando a variável “média da quantidade de contas dos clientes”.

As diferenças significativas encontradas em teste paramétricos nos cálculos de θ dos quatro modelos analisados nas seções anteriores (CRS e VRS com 5 variáveis de *output*, CRS e VRS com 4 variáveis de *output*) não identificam, como já foi explanado, uma avaliação sobre a ótica da posição de cada DMU e sua possível modificação em face da mudança metodológica. Isto é, o que se pretende apurar são as diferenças ordinais, se existentes, decorrentes das mudanças de um modelo para outro. Isso só pode ser feito com a aplicação de testes não paramétricos. Como se trata de testes de 2 amostras relacionadas, optou-se pela prova de Sinais e a prova de Wilcoxon (mais poderosa)²⁵ para testar as diferenças de posição. Antes disso, necessitou-se alterar os valores de eficiência por valores que representavam a posição no *ranking* do cálculo de eficiência para cada DMU, de tal forma que resultou em valores ordinais de 1 a 3652 (para 11 DMU's eficientes no CRS) e de 1 a 3589 (para 74 DMU's eficientes no VRS). Feita esta mudança testamos as hipóteses de que as posições não se alteravam significativamente na mudança de metodologia de CRS para VRS (hipótese nula ou H_0). A hipótese alternativa H_1 (de que havia alteração significativa) foi aceita com alto nível de significância para ambos os testes (Apêndice 1).

A mesma análise, com variáveis ordenadas, foi feita para testar a diferença entre os modelos com 5 variáveis e 4 variáveis de *output* (retirando-se a variável “média da quantidade de contas de clientes”) tanto para os modelos CRS, como para VRS²⁶. Neste caso os resultados mostraram, utilizando a prova de Wilcoxon, que não se pode rejeitar H_0 ao nível de significância de 0,432 e 0,208, respectivamente. Isto mostra a não existência de diferença significativa na ordem de eficiência entre os modelos com ou sem a variável “média do número de contas de clientes”. Entretanto, analisando pela prova dos Sinais, os resultados não foram tão conclusivos, visto que para determinado nível de significância não se pode afirmar com segurança para *Rejeição* ou *Aceitação* de H_0 conforme Apêndice 1, muito embora esta prova não seja tão robusta quanto a anterior. Esta diferença significativa é atribuída, particularmente, a influência do *output* “média do número de contas de clientes” com as duas variáveis de *input* (alta correlação conforme tabela 3.3). Para testar

²⁵ Maiores detalhes sobre a prova de Wilcoxon e dos Sinais pode ser visto em Siegel (1956), versão nacional e da mesma editora em 1975 página 84. Souza (2002) e Zenios et.al (1999) utilizam e propõem os testes Mann-Withney (caso de duas amostras independentes) e Kruskal-Wallis para k amostras independentes. No entanto, as amostras não são relacionadas e o objetivo deles é comparar o grau de eficiência relativa entre amostras e não seu ordenamento.

essa possibilidade foram feitos os cálculos do grau de eficiência relativa do modelo com 4 *outputs* retirando as demais variáveis de *output*. Analisando detalhadamente a variável “captação em Fundos” (por ser aquela que apresentou a menor correlação com os *inputs* e de menor impacto e alteração ao modelo original de 5 *outputs*), chegaram-se a resultados para a média do grau de eficiência de $\theta = 0,343196873$ e desvio-padrão de 0,083956675, resultados bem próximos do modelo geral CRS de 5 *outputs* (veja tabela 3.10 e 3.13), o que reforça a hipótese de que a mudança ou alteração de *inputs* é importante, causando alterações significativas, apenas nos casos de *inputs* correlacionados com os *outputs* que compõem o modelo. Destaca-se que retirada essa variável de *output* (“captação em fundos”) apenas 9 agências tiveram alteração, sendo que 4 delas eram *Corporate* (de 1 a 3 casas decimais de diferença do modelo com 5 *outputs*), 1 do Setor Público (4 casas decimais de diferença) e 4 de Varejo (4 a 5 casas decimais de diferença). As demais variáveis de *output* ocasionam impacto à medida de eficiência proporcionalmente ao grau de correlação entre os *inputs* (comparação entre as tabelas 3.3 e 3.13) o que determina uma ordem de importância dessas variáveis para a composição do modelo e do grau de eficiência estimado para cada DMU (segundo a ordem da Tabela 3.12). Com esta afirmativa e constatação caberiam estudos mais detalhados a respeito da influência de cada variável (*output* ou *input*) sobre cada unidade, desta forma contribuindo para melhor adequação e decisão dos gestores dessas agências. Pelos resultados percebe-se a importância do *output* “média da quantidade de contas de clientes”, exatamente por isso essa variável está sendo, no decorrer deste trabalho, exaustivamente estudada.

A seguir serão analisadas a adequabilidade dos grupos de segmentação para estratégia da empresa considerando os critérios do Banco (que em algumas unidades dos segmentos governo e pessoa jurídica fazem parte dos *outliers* selecionados pela distribuição *Heaviside step*), bem como pela aplicação da Análise Multivariada de *Cluster* que forma grupos homogêneos utilizando algumas variáveis de impacto no mercado de atuação destas DMU's.

É parte dos objetivos desse estudo analisar, teoricamente e comparativamente, as vantagens e desvantagens empíricas das metodologias CRS e VRS. No entanto, como

²⁶ Para os demais *outputs* foi utilizado apenas o modelo CRS como padrão, visto que a intenção era avaliar o impacto de cada variável de *output* ao modelo geral com 5 *outputs*.

foram identificados, no modelo geral CRS, os *outliers* parte integrante dos segmentos estratégicos da instituição avaliada (aproximadamente 2/3 dos segmentos governo e pessoa jurídica fazem parte dos *outliers*) e que não se confirmaram tão bem na utilização do modelo VRS, adotar-se-á o CRS como modelo padrão nas comparações que se seguem.

3.5 – CLUSTERIZAÇÃO OU AGRUPAMENTO DAS AGÊNCIAS

Pretende-se nesta seção analisar, pela metodologia DEA, tanto os resultados das medidas de eficiência relativa, como também os impactos ocasionados por determinadas DMU's *outliers* no conjunto total das demais DMU's. A abordagem explora duas situações:

- I. A primeira investiga os resultados a luz da segmentação utilizada pela empresa, que separa três tipos de clientes, os quais são atendidos por agências específicas, quais sejam: Governo, Pessoa Jurídica e Pessoa Física. O processo de segmentação é totalmente empírico e relacionado com o perfil do cliente.
- II. A segunda separa 4 (quatro) grupos por características do mercado sob uma visão de potencial bancário, considerando, também, 3 variáveis: (1) número de agências da praça; (2) operações de crédito da praça; (3) depósitos bancários na praça. Esta segmentação utiliza o procedimento multivariado de Análise de Cluster²⁷.

A seguir serão mostrados detalhes das metodologias aplicadas e os resultados gerados por estas duas abordagens.

3.5.1. AGRUPAMENTO PELO CRITÉRIO DE SEGMENTAÇÃO DO BANCO:

Um aspecto de necessidade prática na manipulação e gestão de um conjunto de agências para um Banco é o agrupamento ou segmentação das filiais em alguma característica específica, onde se busca homogeneização das variáveis utilizadas. Esta homogeneização é conveniente na medida em que padroniza certos indicadores de gestão, facilitando, principalmente em grandes corporações, tomadas de decisão no que se refere à dotação de mão-de-obra, investimento em *lay out*, *marketing* de relacionamento, medidas

²⁷ Técnica de Análise Multivariada para agrupamento de observações semelhantes, veja Hair et. al (2005). O modelo será melhor explicado na seção 3.5.2.

de variação de risco, previsão de rentabilidade, eficiência e muitos outros aspectos importantes para análise e decisão do gestor.

Na seção anterior chamou-se atenção para as dependências ou DMU's que são administradas de forma diferenciada por características particularizadas de seus clientes (Pessoa Física, Pessoa Jurídica e Governo, divididos pelo Banco pelas suas respectivas Diretorias). Pelo tamanho total da população em estudo até o momento da apuração e levantamento dos dados em 2004 o segmento de Pessoa Física continha 3.552 agências, Pessoa Jurídica 73 e Governo 38. Para análises destes segmentos repetiram-se os cálculos de avaliação da medida de eficiência apurada segundo a metodologia DEA para cada agrupamento. Mesmo porque, existe uma diferença significativa no modelo de gestão para cada um destes segmentos e que foram, estatisticamente, detectadas nas análises anteriores. Alguns resultados descritivos de eficiência são apresentados na Tabela 3.13 abaixo:

Tabela 3.13 – Análise descritiva dos valores de θ – comparação dos segmentos BB

Distribuição	Média	Desv-padrão	CV(%)	Assimetria	Amostra
Governo	0,5582	0,2726	48,84	0,486	38
Pessoa Jurídica	0,3929	0,2572	65,62	1,417	73
Pessoa Física	0,3821	0,0872	22,82	1,542	3552
Geral	0,3433	0,0843	24,55	1,95	3663

As médias dos resultados de eficiência relativa aumentaram para cada agrupamento de clientes identificando uma distribuição com características de maior normalidade dos grupos, caracterizado, também, pelo indicador de assimetria. A maior média de eficiência foi alcançada pelas agências administradas pela Diretoria de Governo (0,558), seguida pela Diretoria de Pessoa Jurídica (0,392) e Pessoa Física (0,382), respectivamente.

O segmento Pessoa Jurídica destaca-se por uma maior heterogeneidade decorrente das diferenças significativas do porte e escala das empresas classificadas como “Empresarial” e “Corporate”. O aumento da dispersão no segmento Governo é explicado pelas diferenças existentes entre agências que tratam com o “poder executivo” e “poder judiciário” e que apresentam diferenças expressivas em suas variáveis.

Confirmou-se, portanto, o que fora comentado anteriormente e que pode ser observado nas Tabelas 3.4 e 3.6 quanto ao grande impacto ou *leverage* ocasionado na população estudada quando não considerados os agrupamentos. Isto é, grande parte desse impacto ocasionado na apuração de eficiência pelo modelo DEA é proveniente das agências com características de gestão voltadas para clientes Governo e Pessoa Jurídica.

3.5.1.1. Segmento Governo:

A Tabela 3.14 retrata 30 agências do segmento Governo (ordem decrescente da coluna *Lev – leverage*), numa nova abordagem, destacando a diminuição significativa do número de *leverages*, bem como de seus valores absolutos com o número total de clientes atendidos pela unidade. Desta maneira, este segmento é confirmado como de maior impacto no estilo de gestão e eficiência. Principalmente as DMU's “Setor Público de Brasília” e “Governo de Brasília” mostram, mesmo na classificação interna do seu próprio segmento, destaque nos resultados do *leverage*. Nota-se que no caso da agência Setor Público Brasília a classificação como primeira no *rank* se deve ao número e quantidade de clientes com contas na sua base. Isto é visivelmente perceptível na última coluna e primeira linha com um número total de 2.716 clientes e, quando da retirada desta variável como *output*, seu posicionamento fica alterado (o indicador de impacto *Leverage* mudou de 0,0865 para 0,0008). Trata-se de unidade com grande número de contas porém pequeno volume de serviços. Sua ineficiência relativa esta caracterizada pelo pequeno volume de serviços associado por um grande número de clientes. Essa quantidade de clientes proporciona um número elevado de *inputs* sem a devida contrapartida de *outputs*. A instituição estuda um modelo para unificar as unidades “Setor Público de Brasília” e “Governo de Brasília”, de tal maneira a aproveitar e distribuir melhor os insumos e recursos dessas unidades, o que pode proporcionar maiores resultados em *outputs* na nova unidade. Por isso um acompanhamento do comportamento temporal se faz necessário.

No entanto, para as demais agências Governo não existem diferenças significativas quando da alteração de *outputs*, confirmando que o ganho de escala não é fator tão importante para a eficiência relativa do segmento, porém, a quantidade de clientes deve ser reavaliada e estudada caso a caso.

Tabela 3.14

Relação de Eficiência e *Leverage* para o segmento Governo

Agência	CRS - 5 outputs				CRS - 4 outputs				Clientes
	Efic	N Hits	Tot Hits	Lev	Efic	N Hits	Tot Hits	Lev	
S.PUBLICO BRASILIA	1,0000	385	998	0,0865	0,3060	61	998	0,0008	2716
GOVERNO BRASILIA	1,0000	345	984	0,0333	1,0000	345	984	0,0508	114
GOVERNO SÃO PAULO	1,0000	407	991	0,0233	1,0000	406	991	0,0201	331
GOVERNO R.DE JANEIRO	0,8492	212	962	0,0154	0,8492	196	962	0,0241	415
P.JUDIC.BRASILIA	1,0000	364	1016	0,0150	1,0000	364	1016	0,0167	45
GOVERNO SALVADOR	1,0000	361	1027	0,0107	1,0000	361	1027	0,0127	358
P.JUDIC.SAO PAULO	1,0000	148	1019	0,0098	1,0000	131	1019	0,0110	576
S.PUBLICO C.GRANDE	0,8316	165	973	0,0072	0,7860	50	973	0,0020	380
S.PUBLICO CUIABA	1,0000	335	1020	0,0059	1,0000	376	1020	0,0101	275
S.PUBLICO MANAUS	0,4607	66	962	0,0018	0,4067	0	962	0,0000	329
P.JUDIC.ESTADUAL	0,7278	37	981	0,0015	0,7278	37	981	0,0016	25
GOVERNO CURITIBA	0,5913	33	1041	0,0009	0,5654	28	1041	0,0008	144
S.PUBLICO TERESINA	0,6870	50	1028	0,0006	0,6461	2	1028	0,0001	229
S.PUBLICO SAO LUIS	0,7851	71	995	0,0005	0,7563	56	995	0,0002	246
S.PUBLICO FORTALEZA	0,6584	47	988	0,0003	0,6460	47	988	0,0003	226
GOVERNO RECIFE	0,6495	26	1005	0,0003	0,6355	49	1005	0,0003	208
GOVERNO P.ALEGRE	0,4147	39	979	0,0001	0,3901	40	979	0,0006	174
S.PUBLICO PALMAS	0,4382	4	1003	0,0000	0,4119	4	1003	0,0000	228
S.PUBLICO MACAPA	0,4328	5	1008	0,0000	0,3997	0	1008	0,0000	126
S.PUBLICO MACEIO	0,3424	5	1014	0,0000	0,3014	0	1014	0,0000	206
GOVERNO B.HORIZONTE	0,4910	9	1044	0,0000	0,4648	10	1044	0,0001	186
S.PUBLICO RIO BRANCO	0,3688	2	982	0,0000	0,3278	0	982	0,0000	204
S.PUBLICO GOIANIA	0,5170	0	1037	0,0000	0,5024	0	1037	0,0000	219
S.PUBLICO J.PESSOA	0,4910	0	983	0,0000	0,4880	0	983	0,0000	171
S.PUBLICO BELEM	0,3722	0	1004	0,0000	0,3389	0	1004	0,0000	387
P.JUDIC.R.DE JANEIRO	0,4389	0	1019	0,0000	0,4320	0	1019	0,0000	247
S.PUBLICO P.VELHO	0,4679	0	1006	0,0000	0,4634	0	1006	0,0000	172
P.JUDIC.RECIFE	0,2101	0	1040	0,0000	0,2052	0	1040	0,0000	123
P.JUDIC.SALVADOR	0,3130	0	1010	0,0000	0,3090	0	1010	0,0000	92
S.PUBLICO FLORIANOP	0,2456	0	966	0,0000	0,2211	0	966	0,0000	142

3.5.1.2. Segmento Pessoa Jurídica:

Resultado semelhante é apresentado na Tabela 3.15 que seleciona as 30 maiores agências de impacto ou *Leverage* do segmento Pessoa Jurídica. As agências Empresarial Campinas, Empresarial São José dos Campos e Empresarial Sul com 5.079, 3.096 e 1.584 clientes, respectivamente, apresentaram resultados diferentes comparando os dois modelos com 5 e 4 *outputs*. Além do grau de eficiência relativa ter diminuído significativamente os valores de *N Hits* (que indica o número de vezes que a unidade interferiu no modelo quando retirada da amostra) anulando-os. Analisando principalmente a agência Empresarial Campinas, pelos resultados, pode-se observar que a unidade está com um número elevado

de clientes que supostamente, analisando o outro modelo com 4 *outputs*, pode estar prejudicando sua *performance* financeira, trabalhando acima de sua capacidade de atendimento. Entretanto, outras análises devem ser feitas para melhor conclusão sobre esta afirmativa, visto que esta *performance* medíocre pode estar associada a insuficiente geração de outros *outputs*. As outras unidades desse segmento pouco ou nada se modificaram. Por outro lado, as agências *Corporate* (RJ) e *Corporate* (DF) com 1295 e 1074 clientes, respectivamente, mostraram grande capacidade de atendimento, muito acima da média das demais agências, permanecendo, mesmo retirando uma das variáveis de output, com eficiência máxima. Desta maneira, estas duas agências são grandes candidatas a posição de *benchmark* como exemplo de atuação e utilização das melhores práticas.

Tabela 3.15

Relação de Eficiência e *Leverage* para o segmento Pessoa Jurídica

Agência	CRS - 5 outputs				CRS - 4 outputs				Clientes
	Efic	N Hits	Tot Hits	Lev	Efic	N Hits	Tot Hits	Lev	
EMPRES.CAMPINAS	1,0000	364	991	0,0441	0,2809	0	991	0,0000	5079
CORPORATE RJ	1,0000	391	1005	0,0205	1,0000	391	1005	0,0247	1295
CORPORATE DF	1,0000	379	955	0,0162	1,0000	370	955	0,0129	1074
CORPORATE AV.PLTA.	1,0000	364	990	0,0158	1,0000	364	990	0,0208	589
EMPRES.S.J.CAMPOS	0,5195	151	1034	0,0133	0,1322	0	1034	0,0000	3096
CORPORATE PETROLEO	1,0000	365	1004	0,0126	1,0000	366	1004	0,0140	528
EMPRES.SUL	0,3985	51	1037	0,0048	0,1717	0	1037	0,0000	1584
CORPORATE MG	1,0000	378	1046	0,0038	1,0000	377	1046	0,0036	727
CORPORATE RS	1,0000	383	1022	0,0019	1,0000	372	1022	0,0017	646
CORPORATE RIO SUL	0,4988	32	1034	0,0013	0,4807	33	1034	0,0017	547
CORPORATE ABC	0,9954	93	1007	0,0006	0,9954	153	1007	0,0008	362
CORPORATE SAO JOAO	0,8671	19	988	0,0005	0,8621	19	988	0,0005	872
EMPRES.NORTE PR	0,4112	6	967	0,0004	0,2984	0	967	0,0000	697
EMPRES.BOA VIAGEM	0,4693	9	1018	0,0002	0,4234	8	1018	0,0002	524
CORPORATE PR	0,6206	14	1020	0,0001	0,6045	18	1020	0,0002	516
EMPRES.BRASILIA	0,3439	1	1021	0,0000	0,2986	1	1021	0,0000	559
CORPORATE MORUMBI	0,5994	3	1024	0,0000	0,5714	3	1024	0,0000	537
CORPORATE SC	0,5443	1	1082	0,0000	0,5031	0	1082	0,0000	645
EMPRES.VILA RICA	0,3743	1	979	0,0000	0,2895	0	979	0,0000	739
CORPORATE S.J.CAMPOS	0,6108	3	1009	0,0000	0,6108	4	1009	0,0000	304
CORPORATE BA	0,5651	1	957	0,0000	0,5462	1	957	0,0000	263
EMPRES.FORTALEZA	0,4474	0	1015	0,0000	0,4064	0	1015	0,0000	449
EMPRES.SEN.DANTAS	0,4402	0	1040	0,0000	0,3977	0	1040	0,0000	715
EMPRES.MANAUS	0,2240	0	1000	0,0000	0,1410	0	1000	0,0000	473
EMPRES.SOROCABA	0,1986	0	980	0,0000	0,1397	0	980	0,0000	457
EMPRES.SANTO AMARO	0,1921	0	1002	0,0000	0,1389	0	1002	0,0000	444
EMPRES.CAMPO GRANDE	0,2316	0	1007	0,0000	0,1546	0	1007	0,0000	415
CORPORATE CAMPINAS	0,5297	0	1000	0,0000	0,5297	0	1000	0,0000	317
EMPRES.JUNDIAI	0,1870	0	943	0,0000	0,1399	0	943	0,0000	381

3.5.1.3. Segmento Pessoa Física:

O único segmento que se pode avaliar ou teria sentido apurar a *leverage* das DMU's é o segmento Pessoa Física, devido à quantidade de unidades. Após os resultados repetiu-se o procedimento de eliminar algumas agências, o que foi feito utilizando a distribuição *Heaviside step function*. A Tabela 3.16 mostra os resultados consolidados para o modelo CRS – 5 *outputs*.

Tabela 3.16 – Análise descritiva dos valores de θ para o segmento Pessoa Física CRS – 5 *outputs*

Distribuição	Média	Desv-padrão	CV(%)	Assimetria	Amostra
Pessoa Física	0,38209951	0,087188813	22,82	1,542	3552
Pessoa Física Heaviside step	0,7181149	0,135429597	18,86	- 0,283	3429(123)
Pessoa Física Outliers	0,625806	0,149913	23,96	0,171	123
Geral	0,3432527	0,084273606	24,55	1,95	3663

Os valores em parênteses correspondem ao número de DMU's removidas da amostra. Repetiu-se a última linha das tabelas 3.7 e 3.10 para melhor entendimento e comparação.

Conforme se pode observar aumentou, consideravelmente, a média de eficiência da distribuição, caracterizando forte influência dos *outliers* na amostra. Com a retirada dessas unidades o resultado apresentou maior homogeneidade, bem como inverteu o grau de assimetria no sentido de concentração da maior eficiência. Da mesma forma, observando a Tabela 3.17 nota-se que o impacto (valor dos maiores *Lev*) diminui em relação ao modelo com todas as unidades. Conclui-se que o impacto e o número de *outliers* no modelo Geral são muito maiores, portanto, atestando a homogeneidade dos segmentos. Isso confirma, como comentado anteriormente, a influência representada pelos *leverages* dos demais segmentos (Pessoa Jurídica e governo).

Tabela 3.17

Relação de Eficiência e *Leverage* para o segmento Pessoa Física

Agência	CRS - 5 outputs					CRS - 4 outputs				
	Efic	N Hits	Tot Hits	Lev	Rank	Efic	N Hits	Tot Hits	Lev	Rank
VERA CRUZ	1,0000	98	122	0,1304	1	1,0000	97	122	0,0105	37
PLANALTO	1,0000	89	100	0,0556	2	1,0000	89	100	0,0930	3
VICENTE MACHADO	0,6848	80	110	0,0502	3	0,2298	0	110	0,0000	2930
SANTA CRUZ DO SUL	0,9219	78	99	0,0486	4	1,0000	74	99	0,0064	50
ENTRE RIOS	0,9520	97	102	0,0485	5	1,0000	97	102	0,1135	1
ACREUNA	1,0000	85	91	0,0447	6	1,0000	85	91	0,0939	2
SINIMBU	0,6053	85	102	0,0420	7	0,6268	34	102	0,0010	98
PETROBRAS	0,6848	78	86	0,0406	8	0,6341	77	86	0,0579	4
ITAIM BIBI	1,0000	74	80	0,0389	9	1,0000	74	80	0,0519	6
TRIANON	0,6009	84	100	0,0321	10	0,4380	83	100	0,0437	7
CORURIFE	0,6134	85	107	0,0281	11	0,1998	0	107	0,0000	1172
CAMARA CASCUDO	0,6602	87	100	0,0275	12	0,2841	0	100	0,0000	3268
HERACLITO GRACA	0,5294	82	101	0,0261	13	0,4162	71	101	0,0347	11
POSTALIS	1,0000	87	96	0,0255	14	1,0000	86	96	0,0312	13
CURITIBA	0,4480	81	102	0,0250	15	0,6725	85	102	0,0402	8
CAMPO VERDE	0,9849	76	85	0,0247	16	0,9320	76	85	0,0323	12
SAO DESIDERIO	0,8268	85	97	0,0227	17	0,5186	51	97	0,0113	34
SUDOESTE 303	0,9183	98	102	0,0223	18	0,2189	58	102	0,0026	72
VENANCIO AIRES	0,5895	80	113	0,0200	19	0,5105	6	113	0,0002	156
BOULEVARD SAO JOAO	0,6136	88	117	0,0197	20	0,5829	88	117	0,0381	10
CHAPADAO DO CEU	0,9494	83	94	0,0189	21	0,7613	87	94	0,0527	5
MONTIVIDIU	0,8320	86	103	0,0163	22	0,7634	81	103	0,0394	9
PARANAGUA	0,4571	74	121	0,0163	23	0,3086	55	121	0,0187	18
PRACA DO CARMO	0,5379	65	95	0,0136	24	0,3216	36	95	0,0084	44
SINOP	0,8453	80	95	0,0124	25	0,6734	71	95	0,0181	19
EUSEBIO	0,5885	66	107	0,0121	26	0,2929	0	107	0,0000	3086
SENADO FEDERAL	0,4921	65	109	0,0116	27	0,4707	62	109	0,0219	17
S. COMERCIAL SUL	0,5460	79	102	0,0115	28	0,4465	63	102	0,0115	32
AMERICA DOURADA	0,6251	86	119	0,0114	29	0,1546	0	119	0,0000	3258
PEDRA PRETA	0,8353	82	98	0,0112	30	0,6936	63	98	0,0069	49

Novamente observa-se a influência da variável “média da quantidade de contas de clientes” principalmente nas agências Vicente Machado (PR), Coruripe (AL), Câmara Cascudo – Natal (RN), Eusébio (CE) e América Dourada (BA) para as 30 agências selecionadas. Estas agências perderam, significativamente, grau de eficiência, bem como perderam influência e dominância nos limites da fronteira de produção (N Hits igual a zero). Em termos de eficiência, como ilustração, pode-se comentar sobre a agência Sudoeste 303 que diminuiu o seu grau de eficiência relativa de 0,9183 para 0,2189 o que

pode estar indicando um dimensionamento incorreto de clientes atendidos por essa unidade; exigindo, desta maneira, a possibilidade de abertura de outras unidades na região o que efetivamente foi feito migrando a unidade para a quadra 305 e colocando uma agência virtual na quadra 303. Por outro lado, as unidades Santa Cruz do Sul (RS), Entre Rios (BA), Sinimbu (RS) e Curitiba (PR) se destacam por aumentar a eficiência no modelo de 4 *outputs*, o que significa que o desempenho financeiro destas unidades e aproveitamento dos recursos é comparativamente eficiente, indicando que a quantidade de clientes é adequada.

As 30 unidades apresentadas na Tabela 3.17 fazem parte das 123 DMU's *outliers* eliminadas em Pessoa Física e que representam filiais com valores acima da média entre alguns itens de *output* ou abaixo da média, no que se refere ao *input*. Portanto, situadas no limite das fronteiras de produção e exercendo dominância relativa em algumas unidades. Também, no Apêndice 2, são mostrados os resultados identificando algumas unidades, entre as 123 *outliers* “eliminadas” pelo valor de *Leverage*, que foram geradas pelo modelo CRS – 5 *outputs*:

- As principais unidades com grande quantidade de clientes ou com indicador de *output* neste item acima da média são: UnB (DF), SMU (DF), Venâncio Aires (RS), Rio Verde (GO) e Prefeitura do Recife (PE).
- Outras apresentam forte concentração de aplicações (empréstimos bancários), principalmente no *agrobusiness*, mas também alguns casos na indústria ou serviços. Novamente Santa Cruz do Sul e Venâncio Aires (RS), Rio Verde (GO), Campo Verde, Sorriso e Sinop (MT), Rondonópolis (RO) e Guaicunus em Campo Grande (MS) entre outras.
- Quanto à captação destaca-se, como vimos anteriormente, a Agência de Itaim Bibi. Como foi ressaltado trata-se de agência que concentra os depósitos judiciais da capital de São Paulo.
- As que se destacam com tarifas e captações de fundos são: Agência Centro do Rio de Janeiro e Petrobrás (RJ), Afonso Pena de Campo Grande (MS), Rio Verde (GO), Planalto, Senado Federal e Tribunais em Brasília (DF).
- Finalmente aquelas que se destacam por *inputs* abaixo da média são, normalmente, agências situadas em municípios de pequeno porte e que atendem uma grande quantidade de clientes relativamente a outras regiões mais populosas, ou seja:

América Dourada e Lagoa Real (BA), Joviânia, Maurilândia, Montividiu e Chapadão do Céu (GO), Fronteira e Senador Firmino (MG), Tapuah (MT), Itaipulândia (PR) e Vale do Sol (RS).

Estas unidades poderiam ser avaliadas e serviriam de *benchmark*, tanto para exemplos de melhores práticas como para avaliação do perfil de mercado. Por outro lado, fariam parte de um conjunto de unidades selecionadas e que estariam disponíveis a avaliações sobre o aspecto do risco das aplicações e valores excessivos das tarifas. O valor excessivo de tarifas normalmente remunera unidades em pequenas localidades de baixa concorrência bancária, visto que uma política da empresa são os descontos de tarifas para grandes contas de captação e clientes fiéis e com vários produtos no Banco, o que normalmente não ocorre em pequenos municípios onde essas agências estão localizadas.

Tabela 3.18

Relação de Eficiência por Classe de Clientes - PF - CRS 5 e 4 *outputs*

Classe de Clientes	Número de Agências		Média da classe		Desvio padrão		Coeficiente de Assimetria		Mínimo	
	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav
CRS – 5										
0 a 3000	957	934	0,3375	0,6267	0,1017	0,1470	2,0858	-0,0053	0,0435	0,1013
3001 a 5000	951	932	0,3826	0,7181	0,0746	0,1172	1,8126	0,2708	0,1741	0,3354
5001 a 7000	566	553	0,3991	0,7541	0,0684	0,0997	2,6845	0,2515	0,2242	0,3703
7001 a 10000	488	462	0,4053	0,7663	0,0749	0,1091	2,2259	0,2109	0,2248	0,4544
10001 a 15000	339	319	0,4065	0,7750	0,0745	0,0984	3,0436	0,3785	0,3025	0,5682
15001 a 20000	150	140	0,4211	0,8066	0,0714	0,0958	2,4486	0,3529	0,3023	0,6127
acima de 20000	100	33	0,4552	0,8660	0,0688	0,0943	0,8248	-0,1564	0,3596	0,6733
CRS – 4										
0 a 3000	957	934	0,1932	0,4647	0,1261	0,1483	3,8973	0,9854	0,0210	0,0594
3001 a 5000	951	936	0,2155	0,5416	0,0872	0,1364	2,8284	0,8990	0,0903	0,2058
5001 a 7000	566	553	0,2371	0,5855	0,0817	0,1225	3,0700	0,6684	0,1149	0,3059
7001 a 10000	488	473	0,2557	0,6202	0,0831	0,1259	3,2241	0,7437	0,1349	0,3804
10001 a 15000	339	324	0,2684	0,6534	0,0788	0,1209	4,0115	0,6393	0,1380	0,3432
15001 a 20000	150	145	0,2784	0,6747	0,0827	0,1147	4,4176	0,7817	0,1807	0,4335
acima de 20000	100	96	0,2760	0,7123	0,0453	0,1180	0,8029	0,4070	0,1875	0,4652

A Tabela 3.18 acima compara algumas estatísticas (no caso classes de clientes) das medidas do grau de eficiência relativa dos modelos com 5 e 4 *outputs* e os modelos DEA normal e *Heaviside step* (após ter sido submetido a seleção dos *outliers*). Os resultados indicaram que a média das classes do grau de eficiência relativa é crescente em quase todas as situações e que, quando aplicada à distribuição *Heaviside step*, a tendência permanece a

mesma diminuindo a variabilidade e aumentando a eficiência relativa. Destacam-se, também, que as agências que mais impactaram a eficiência e alteraram o limite da fronteira de produção no modelo CRS de 5 *outputs* foram as unidades com número de clientes acima de 20.000, onde na seleção do modelo *Heaviside step* foram retiradas 67 unidades (de 100 a quantidade foi para 33 agências). Entretanto, no modelo de 4 *outputs* a seleção torna-se mais uniforme nas 7 classes de clientes. Novamente percebe-se a importância e o impacto da variável representada pela “Classe de Clientes” para os resultados de eficiência, independente de sua participação na composição dos *outputs* dos modelos. Isto é, nessa comparação pode-se afirmar que existe ganho de escala na alteração da base de clientes.

Para aprofundar detalhadamente os estudos do comportamento do grau de eficiência das agências do segmento Pessoa Física, além da comparação pelo número de clientes, foi feita comparação da variável eficiência com a distribuição da média de classe de renda desses clientes, considerando cinco classes distribuídas pelo intervalo discriminado na Tabela 3.19. Os resultados acompanham a tendência da Tabela 3.18, em que o grau de eficiência relativa é crescente com as maiores rendas médias, confirmando que a renda média dos clientes contribui positivamente e de forma efetiva para a eficiência relativa das unidades. Nota-se a diminuição do coeficiente de assimetria, quando da retirada dos *outliers*, indicando ajustamento à normalidade da distribuição da variável eficiência.

Tabela 3.19

Relação de Eficiência por Classe de Renda – PF - CRS 5 e 4 *outputs*

Classe de Renda	Número de Agências		Média de θ para Classe de Renda		desvio padrão		Coeficiente de Assimetria		Mínimo	
	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav
CRS – 5										
0 a 1000	2204	2161	0,3761	0,7090	0,0749	0,1309	0,5059	-0,2315	0,0435	0,1013
1000 a 1500	944	931	0,3771	0,7267	0,0712	0,1342	0,5359	-0,3336	0,0748	0,1241
1500 a 2000	247	229	0,4041	0,7550	0,1102	0,1446	1,8322	-0,6333	0,0888	0,1495
2000 a 3000	101	74	0,4561	0,7407	0,1823	0,1870	1,0139	-0,5667	0,1441	0,2670
acima de 3000	55	33	0,4790	0,7768	0,2021	0,1693	0,9659	-0,4905	0,1762	0,3792
CRS – 4										
0 a 1000	2204	2197	0,1913	0,5050	0,0612	0,1229	2,7648	0,4327	0,0363	0,0594
1000 a 1500	944	936	0,2591	0,6340	0,0776	0,1381	3,3230	-0,0542	0,0210	0,0820
1500 a 2000	247	226	0,3104	0,7076	0,1199	0,1584	2,2336	-0,2928	0,0696	0,1303
2000 a 3000	101	71	0,4108	0,7076	0,2065	0,1917	1,3185	-0,3875	0,1052	0,2606
acima de 3000	55	31	0,4518	0,7356	0,2080	0,1847	0,9810	-0,2692	0,1056	0,3828

Por fim, prosseguindo com a análise do grau de eficiência com as demais variáveis que podem influenciá-la foi feita comparação com o número de agências concorrentes (Tabela 3.20). Diferentemente das demais tabelas anteriores o grau de eficiência não apresenta um comportamento crescente com esta variável comparativa. Para a classe onde não existem agências concorrentes (existe apenas uma agência do próprio Banco na cidade) o grau de eficiência foi menor, indicando a presença da empresa em localidades de duvidosa viabilidade financeira. Porém, na medida em que o número de agências aumenta, até atingir um valor máximo entre 51 a 1000 agências concorrentes, o grau de eficiência cai. O que se pode concluir desse comportamento é que para uma determinada quantidade de concorrentes (normalmente situadas em grandes municípios ou metrópoles) o grau de eficiência sofre impacto negativo, indicando a dificuldade de se trabalhar em grandes centros com alto nível de competitividade. Portanto, isto provavelmente está associado à influência de fatores externos e de porte da localidade em que a agência está situada. Isto está acontecendo especificamente em agências de grandes centros metropolitanos, particularmente na cidade de São Paulo, onde o Banco tem uma *performance* abaixo do esperado comparativamente às demais cidades de grande potencial.

Tabela 3.20

Relação de Quantidade de Agências Concorrentes e Eficiência – PF com 5 e 4 *outputs*

Agências Concorrentes	Número de Agências		Média da Eficiência θ		desvio padrão		Coeficiente de Assimetria		Mínimo	
	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav	DEA	Heav
CRS – 5										
Nenhuma	142	138	0,3679	0,6632	0,0891	0,1531	0,4544	0,1821	0,1654	0,2779
1 a 3	1524	1487	0,3788	0,7078	0,0886	0,1305	1,7198	0,0735	0,1433	0,2846
4 a 10	742	728	0,3853	0,7361	0,0653	0,1103	1,5809	0,1345	0,1587	0,2726
11 a 50	318	302	0,3924	0,7390	0,1001	0,1401	1,9895	-0,8334	0,0435	0,1013
51 a 100	240	233	0,3871	0,7419	0,0829	0,1413	0,3998	-0,5730	0,1397	0,2508
101 a 1000	307	281	0,3879	0,7197	0,1043	0,1515	1,6908	-0,4768	0,1127	0,2002
acima de 1000	275	257	0,3770	0,7088	0,0947	0,1706	0,7623	-0,8081	0,0748	0,1241
CRS – 4										
Nenhuma	142	142	0,1354	0,4019	0,0292	0,0971	0,7128	0,5230	0,0686	0,1849
1 a 3	1524	1500	0,1968	0,4986	0,0910	0,1314	3,4773	1,0698	0,0859	0,2119
4 a 10	742	731	0,2508	0,6149	0,0756	0,1214	2,4573	0,6242	0,1126	0,2646
11 a 50	318	311	0,2692	0,6291	0,1188	0,1317	4,0492	-0,4656	0,0363	0,0594
51 a 100	240	236	0,2640	0,6355	0,0955	0,1449	4,0064	0,0302	0,0963	0,2058
101 a 1000	307	285	0,2798	0,6304	0,1216	0,1566	3,6044	-0,1104	0,0210	0,0820
acima de 1000	275	254	0,2498	0,6105	0,0995	0,1779	2,6485	-0,0228	0,0562	0,1047

Baseando-se nestes resultados comparativos de algumas variáveis com a eficiência estimada pelo modelo DEA e suas diversas alternativas técnicas e de variação de output, procura-se, no próximo capítulo, aprimorar o estudo utilizando a Regressão Quantílica como metodologia para apurar os efeitos que estas e outras variáveis podem ocasionar na eficiência, estabelecendo ou não o impacto de uma relação causal entre elas (nexo causal).

3.5.2. CLUSTERIZAÇÃO PELO CRITÉRIO DE MERCADO:

De acordo com Hair²⁸, a análise de cluster é o nome dado a um grupo de técnicas multivariadas cujo propósito primário é agrupar objetos (indivíduos, produtos, ou outras entidades) baseados em suas características. Para isso, essas técnicas buscam organizar as informações sobre variáveis de tal forma que, grupos relativamente homogêneos, possam ser formados, de maneira que o grau de associação entre dois elementos do mesmo grupo seja máximo e entre dois elementos de grupos distintos seja mínimo. Essa análise simplesmente descobre uma estrutura nos dados, mas não explica o porquê de sua existência. Assim, se a classificação resultante for bem sucedida, os objetos dentro de um mesmo *cluster* ficarão próximos quando “plotados” geometricamente, e entre *clusters* distintos, ficarão mais distantes. A análise de *cluster* é uma ferramenta de análise de dados bastante útil em diversas situações. Por exemplo, um pesquisador que faça coleta de dados por meio de um questionário pode se deparar com um número muito grande de observações que não possui nenhum significado a não ser quando esses dados são classificados em grupos menores, mais tratáveis. Neste sentido, a análise de cluster pode realizar esse procedimento de redução de dados de forma objetiva por meio da redução da informação de uma população inteira (ou de uma amostra) para informações sobre grupos específicos menores. No presente estudo os *clusters* são de municípios que tenham características semelhantes de dinamismo econômico e potencial bancário. Na análise de agrupamento, a similaridade entre duas amostras pode ser expressa em função da distância Euclidiana entre dois pontos representativos desta amostra no espaço n-dimensional (critério Hierárquico de ajustamento) ou através do critério *k-means*, cujo método utilizado é o da minimização da variabilidade dentro dos *clusters* e maximização da variabilidade entre os *clusters*. Este

²⁸ Para maiores detalhes, ver referência em Hair, Joseph F; Anderson, Rolph E.; Tatham, Ronald L.; Black, William C. *Multivariate Data Analysis*. Prentice Hall, 5a. Edition (2005) p. 87-138.

último critério foi o escolhido neste estudo uma vez que é o mais adequado para grandes amostras.

3.5.2.1 A técnica de clusterização k-means

A técnica de *clusterização K-means* é um procedimento de ajustamento não-hierárquico. Este procedimento procura identificar grupos relativamente homogêneos baseados em características selecionadas, usando um algoritmo capaz de trabalhar com grandes amostras. No entanto, o algoritmo requer que seja especificado o número de *clusters* a ser formado. Neste estudo, a quantidade de *clusters* escolhida foi $k = 4$ (escolhido de forma arbitrária). Em seguida, o programa seleciona uma semente inicial para cada *cluster*, as quais são usadas como “tentativas iniciais” para as médias dos *clusters*. Após as quatro sementes iniciais terem sido escolhidas, o programa atribui cada nova observação (município) ao *cluster* com a semente mais próxima. Os municípios vão sendo distribuídos de forma a minimizar a variabilidade dentro dos *clusters* e maximizar a variabilidade entre os *clusters*. O objetivo é movê-los para dentro e para fora dos grupos de forma a alcançar os resultados mais significantes. Usualmente, como resultado dessa análise de *clusterização*, deve-se examinar as médias obtidas em cada *cluster* para cada uma de suas dimensões a fim de avaliar o quão distintas são essas médias, em cada um dos *clusters*. A idéia é que, médias bastante diferentes são obtidas na maioria das dimensões usadas na análise. A magnitude dos *F-values* obtida por meio da análise de variância realizada em cada dimensão é uma outra indicação de como as respectivas dimensões são diferentes entre os *clusters*.

3.5.2.2 Dados utilizados e resultados obtidos com a análise de *clusterização*²⁹.

Conforme comentado, as variáveis utilizadas foram: (1) número de agências da praça; (2) operações de crédito da praça e (3) depósitos bancários na praça³⁰. Estas variáveis foram transformadas numa escala de 0 a 1 a fim de diminuir as diferenças absolutas que distorcem os resultados no processo de *clusterização*. Após este

²⁹ O *software* utilizado para Análise de *Cluster* foi o SPSS v.13.

procedimento submeteram-se as variáveis à técnica *k-means*. Inicialmente sugeriu-se 8 grupos, para que, no resultado final, chegasse a 4 grupos.

- 1) O primeiro grupo é composto por 566 agências espalhadas em todos os estados brasileiros. São agências pequenas que atendem os menores municípios, com agência do Banco do Brasil, em todo o país.
- 2) O segundo grupo tem em sua composição 2.724 agências situadas em capitais (exceto São Paulo, Rio de Janeiro e Distrito Federal) e cidades expressivas em seus estados (médias e grandes cidades)
- 3) O terceiro grupo é composto por 199 agências da cidade do Rio de Janeiro e Distrito Federal.
- 4) Finalmente o quarto grupo envolve 171 agências da cidade de São Paulo.

É importante destacar que as variáveis utilizadas com a finalidade de *clusterização* expressam o potencial bancário da região, na medida em que são volumes de valores aportados em determinado período, bem como a participação e indicação da importância avaliada pela indústria bancária (medido pelo número de agências daquela praça).

A Tabela 3.21 mostra algumas estatísticas do grau de eficiência relativa θ referentes aos 4 grupos estimados pelo método de Análise Multivariada utilizando clusterização para grandes amostras e comparando com a média de renda de cada grupo. O tamanho da amostra apresentado na tabela diferencia dos grupos clusterizados porque algumas agências não tinham informação de renda média dos clientes.

Tabela 3.21: Análise descritiva dos valores de θ para os Grupos *Clusterizados*. Estatística comparativa com número de Agências concorrentes, Renda média por agência e Média de Clientes por agência.

Grupo	N	Agências Concorrentes	Renda Média	Média Clientes	Eficiência CRS - 5			Eficiência CRS - 4		
					média	d-p	ass	média	d-p	ass
1	564	7	751,65	6943	0,3747	0,0673	0,583	0,0781	0,0368	5,816
2	2713	36	1040,19	6725	0,3423	0,0744	1,450	0,0868	0,0554	8,314
3	197	782	1654,41	9475	0,2948	0,1205	3,368	0,1026	0,1158	5,886
4	171	2031	1221,91	7879	0,3026	0,1345	3,464	0,1108	0,1572	5,002

³⁰ Estas informações foram conseguidas no SISBACEN – Sistema do Banco Central - valores de 2004. Lembrando que informações de municípios com menos de 3 agências não são fornecidas, sendo colocadas em um mesmo grupo.

Segundo a Tabela 3.21 observam-se, nos resultados, que o grupo de maior eficiência com 5 *outputs* compreende (grupo 1) agências situadas em pequenos municípios do interior do Brasil presentes em todos Estados e isoladas em termos de concorrência (até no máximo 7 agências concorrentes), contrastando com o menor valor de Renda Média e segunda menor média de clientes por agência. Na sequência tem-se o grupo 2 de agências situadas em municípios de certa importância no aspecto “potencial financeiro” e com menor média de clientes por agência, bem como em locais de pouca concorrência (até 36 agências). O grupo 3 é composto de agências das cidades de Brasília e Rio de Janeiro onde a renda média é maior, porém com baixa eficiência (considerando 5 *outputs*). Da mesma forma o grupo 4 (formado por agências da cidade de São Paulo) com o segundo menor grau de eficiência relativa contrastando com os maiores valores em número de agências concorrentes, renda média e quantidade média de clientes.

Comparando os resultados de eficiência com 4 *outputs* e as demais estatísticas da tabela (número de agências concorrentes, renda média das agências e média de clientes por agência) percebe-se que são diretamente proporcionais. Isto é, o grupo de maior renda média é, praticamente, o de maior grau de eficiência relativa e vice-versa (exceto o caso dos grupos 3 e 4 onde existe uma pequena diferença entre ambos). O mesmo pode-se observar a respeito do número de agências concorrentes na praça que cresce acompanhando o grau de eficiência.

Mais uma vez a adição ou retirada do output “média da quantidade de contas de clientes” influenciou os resultados do grau de eficiência invertendo o sentido de correlação das variáveis comparativas. Há de se esperar, como ocorreu com o modelo CRS – 4, que maiores médias de renda influenciariam diretamente a eficiência. O fato de isto não ter ocorrido, no modelo CRS – 5, deve-se, também, a mistura dos diversos segmentos do Banco (Governo, Pessoa Jurídica e Pessoa Física) que apontam resultados contrários, conforme observado nas tabelas de cada segmento quando comparados os modelos CRS com 5 e 4 *outputs*. Outra justificativa diz respeito à posição relativa da quantidade do número de clientes para agências de pequeno porte que influenciam (no modelo de 5 *outputs*) o grau de eficiência. De qualquer forma, isto coloca sob questão a eficácia dos modelos de *clusterização* para este tipo de segmentação.

Estes agrupamentos serão mais bem compreendidos quando, no capítulo seguinte, forem relacionados seus valores de eficiência com variáveis de mercado apuradas externamente e através de indicadores internos consolidados das agências representativas DMU.

De qualquer maneira, os resultados apresentados no Apêndice 3 - testes não paramétricos de comparação de k amostras independentes³¹ e Análise de Variância no teste paramétrico - identificaram a diferença significativa existente, em qualquer nível de confiança, entre as amostras formadas pelos segmentos empíricos do Banco (Pessoa Física, Pessoa Jurídica e Governo), bem como os 4 (quatro) grupos de *clusters* calculados pela Análise Multivariada. Isso significa que, numa medida de eficiência comparativa, tanto o modelo de gestão estratégico utilizado nos segmentos do Banco como a influência do mercado possivelmente são importantes no impacto dessa eficiência, o que será testado e confirmado no capítulo seguinte.

As unidades do segmento Pessoa Física, por terem uma participação relativamente grande no conjunto da amostra e, conseqüentemente, no total das agências do Banco em questão, serão analisadas separadamente no próximo capítulo. Como o modelo de 4 *outputs* apresentou-se mais adequado aos propósitos do Banco para medição da eficiência técnica na utilização de recursos para produção de resultados financeiros, o uso da variável como *output*: “média total de contas de clientes da agência”, comprovou exhaustivamente sua importância durante as seções anteriores, inclusive distorcendo resultados de eficiência financeira relativa (geração de lucro e produtividade financeira). Por isso, para a técnica comparativa de regressão quantílica, serão apresentados apenas os modelos Pessoa Física com 4 *outputs*, CRS e VRS, com e sem *outliers*.

³¹ Conforme Souza (2002) os testes utilizados foram de Kruskal-Wallis e da Mediana. Veja Siegel (1956) p.197 a 218.

CAPÍTULO 4 – ANÁLISE DO IMPACTO DO MERCADO NO GRAU DE EFICIÊNCIA RELATIVA: UMA APLICAÇÃO DE REGRESSÃO QUANTÍLICA

Os capítulos anteriores mostraram como se calcula eficiência em um modelo DEA e apresentaram os resultados empíricos desses indicadores aplicados em uma relação de agências bancárias, considerando algumas variáveis de *input* e *output*, bem como separando alguns segmentos de interesse. O objetivo deste capítulo é estabelecer a relação existente entre os graus de eficiência e o ambiente onde cada agência está inserida. O ambiente compreende: 1) as variáveis de mercado, tais como: população, renda per capita, região de localização, número de empresas e outras variáveis possíveis de expressar a situação econômica do município e sua influência no comportamento das agências quanto à eficiência; 2) algumas variáveis do cadastro das agências que podem expressar o perfil da localidade em que elas estão inseridas, tais como: renda média dos clientes da agência, quantidade de clientes, quantidade de clientes por estado civil, quantidade de clientes por situação escolar e demais informações do cadastro de clientes das agências; 3) além destas, algumas variáveis que expressam o estilo de gestão: grau de satisfação dos clientes e clima de satisfação dos funcionários quanto a situação de trabalho (estilo de gestão, companheirismo, sobrecarga de trabalho, instalações apropriadas, etc).

Para tal propósito, será usada a técnica de Regressão Quantílica³², a qual nos permite analisar a associação contemporânea entre a variável resposta (medida de eficiência) com as variáveis explicativas (variáveis sugeridas no parágrafo anterior) nos diversos quantis da distribuição condicional. Assim se obtém um mapeamento do impacto do ambiente sobre a eficiência, pois se consegue investigar como cada quantil responde, ao invés de se ter apenas um modelo de regressão para o plano linear no caso da média onde é utilizado o critério dos mínimos quadrados (MQO). De maneira geral, o ambiente é bastante heterogêneo ao longo da distribuição condicional do grau de eficiência corroborando para a utilização dessa técnica.

Algumas vantagens inerentes à regressão quantílica sobre os MQO podem ser listadas conforme Koenker & Bassett (1978):

³² O *software* utilizado para as Regressões foi o R 2.2.1 que é disponibilizado no *site* <http://cran.r-project.org>.

- A técnica de regressão quantílica permite caracterizar toda distribuição condicional de uma variável resposta a partir de um conjunto de regressores;
- Regressão quantílica é robusta a *outliers*, o que se ajusta bem aos modelos DEA;
- Por utilizar a distribuição condicional da variável resposta, podem-se estimar os intervalos de confiança dos parâmetros e do regressando diretamente dos quantis condicionais desejados;
- Como os erros não possuem uma distribuição normal, os estimadores provenientes da regressão quantílica podem ser mais eficientes que os estimadores por meio do MQO;
- A regressão quantílica pode ser representada como um modelo de programação linear, o que facilita a estimação dos parâmetros. A função objetivo da regressão quantílica é uma soma ponderada de desvios absolutos, fornecendo uma medida de locação robusta, de modo que o vetor de coeficientes estimado não é sensível a observações extremas na variável dependente.
- Soluções diferentes para quantis distintos podem ser interpretadas como diferenças na resposta da variável dependente às mudanças nos regressores em vários pontos da distribuição condicional desta mesma variável dependente.

4.1 MODELO DE REGRESSÃO QUANTÍLICA³³

A regressão quantílica pode ser vista como uma extensão natural dos quantis amostrais para o caso de um modelo linear $y_i = X_i\beta + \varepsilon_i$, o qual assume a seguinte forma:

$$\min_{\beta \in \mathfrak{R}} \frac{1}{n} \left\{ \sum_{i \in \{i: y_i \geq x_i \beta\}} \tau |y_i - x_i \beta| + \sum_{i \in \{i: y_i < x_i \beta\}} (1 - \tau) |y_i - x_i \beta| \right\} = \min \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(y_i - x_i \beta)$$

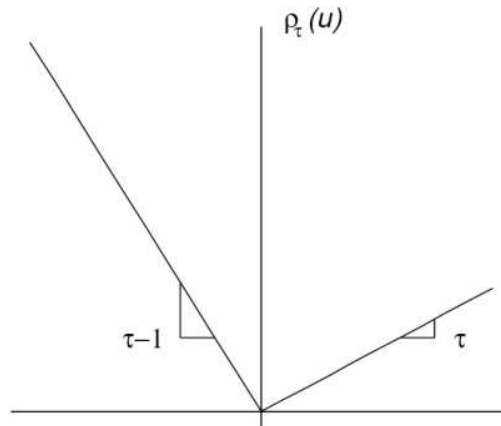
onde ρ é a função *check* definida por :

$$\rho_{\tau}(u) = \begin{cases} \tau u, & u \geq 0 \\ (\tau - 1)u, & u < 0 \end{cases}$$

³³ Maiores detalhes sobre as técnicas de Regressão Quantílica podem ser vistos em Buchinsky (1997), Capítulos 1 a 4 e no artigo de Koenker & Basset (1978).

A função ρ_τ multiplica os resíduos por τ se eles forem não-negativos e por $(\tau-1)$ caso contrário, para que haja um tratamento assimétrico. Na figura 4.0 é representada a função ρ_τ definida acima.

Figura 4.0 : Distribuição de $\rho_\tau(u)$



4.2 TEORIA ASSINTÓTICA PARA REGRESSÃO QUANTÍLICA

Para estudar o comportamento assintótico da estimativa de regressão quantílica são feitas algumas suposições adicionais:

- I. A distribuição dos erros, F_ε tem densidade contínua e estritamente positiva para todo z , tal que $0 < F_\varepsilon(z) < 1$.
- II. $\lim_{n \rightarrow \infty} n^{-1} X' X = D$ é uma matriz positiva definida.

Koenker e Basset (1978) mostraram que, para erros independentes e identicamente distribuídos (i.i.d.), tem-se :

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N(0, \Lambda_\tau)$$

Onde $\Lambda_\tau = \frac{\tau(1-\tau)}{f_\varepsilon(F_\varepsilon^{-1}\tau)} D^{-1}$. Assim, a precisão assintótica da estimativa de regressão

quantílica para erros i.i.d. depende basicamente da quantidade $s(\tau) = [f_\varepsilon F_\varepsilon^{-1}\tau]^{-1}$, comumente chamada de função *sparsity*³⁴.

Para erros não i.i.d., o limite da matriz de covariância assume a forma do “Huber Sandwich” :

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N(0, H_n^{-1} J_n H_n^{-1}) \quad \text{onde} \quad J_n(\tau) = \tau(1-\tau)n^{-1} X'X \quad \text{e}$$

$$H_n(\tau) = \lim_{n \rightarrow \infty} n^{-1} \sum x_i x_i' f_i(\xi_i(\tau)).$$

A função $f_i(\xi_i(\tau))$ é a densidade condicional da variável resposta y_i avaliada no τ -ésimo quantil condicional. No caso i.i.d. as funções $f_i(\xi_i(\tau))$ são idênticas e o “Huber Sandwich” se iguala à expressão Λ_τ para erros i.i.d..

4.3 DADOS E RESULTADOS

O objetivo empírico é analisar se há alguma relação positiva ou negativa entre o grau de eficiência de uma agência, calculado pela técnica DEA, e variáveis ambientais (interna e de mercado) nos diversos quantis da distribuição condicional da variável resposta. O fato é que existe uma forte tendência de alguns parâmetros representativos de variáveis que indicam alta variabilidade no grau de eficiência relativa. Isso, por si só, já recomenda a utilização da Regressão Quantílica. Se esses coeficientes são alterados com a mudança de τ , isso indica alguma forma de heterocedasticidade. A concentração sob determinada faixa quantílica provoca alterações dos estimadores representativos das variáveis, ocasionando mudanças significativas dessas variáveis no impacto à eficiência.

Para tal, utilizaram-se quatro medidas para representar a variável dependente: todas considerando o segmento Pessoa Física, vinculado a Diretoria de Varejo, com 4 *outputs*³⁵. As variações foram para o modelo com ou sem *outliers* utilizando as duas técnicas CRS e VRS. Portanto, as variáveis dependentes foram: eficiência com 4 *outputs* e 2 *inputs* – CRS – *Eficcrs4o* e a mesma situação com VRS - *Eficvrs4o*, e as duas variações sem *outliers*

³⁴ Segundo Koenker e Basset (1978) o termo “função *sparsity*” foi criado por Tukey em 1965, posteriormente, em 1979, foi chamado por Parzen de $s(\tau)$ ou “função densidade quantil”.

³⁵ Outros modelos e suas combinações também foram ajustados: Pessoa Física com 5 *outputs*, modelo geral considerando todos os segmentos (Pessoa Física, Pessoa Jurídica e Governo) com 4 e 5 *outputs* (CRS ou VRS).

(*Eficcrs4* e *Eficvrs4*). Diversas variáveis independentes, apuradas no final do ano de 2004, foram testadas para representar o mercado, tais como:

- 1) Variáveis representativas de características do município³⁶: população do município, população economicamente ativa - PEA, IDH – índice de desenvolvimento humano do município, IDH-educ – índice de desenvolvimento humano - fator educação - do município, renda *per capita* do município, número de empresas do município, número de agências bancárias da praça (esta variável expressa o nível de concorrência e presença da indústria financeira na localidade).
- 2) Variáveis que caracterizam o perfil do cliente e o tipo de agência³⁷: número de contas de clientes da agência, renda média dos clientes da agência, nível de escolaridade (analfabeto, fundamental, médio e superior), estado civil (solteiro, casado, separado e viúvo), resultado da pesquisa de “clima” e relacionamento interno das agências no ano de 2004, e pesquisa de grau de satisfação externa dos clientes por agência em 2004.
- 3) Além disso, mais três conjuntos de variáveis “dummies” a saber:
 - i. variáveis regionais que representam as 5 regiões do país;
 - ii. variáveis que identificam a predominância do tipo de cliente atendido pela agência (pessoa física, pessoa jurídica e governo ou pública) citado no capítulo 3 seção 3.1³⁸.
 - iii. variáveis representativas dos 4 grupos encontrados segundo o critério de clusterização do capítulo 3 seção 3.2 .

Algumas variáveis foram testadas e retiradas por não contribuírem para o modelo, ou pelo alto grau de correlação apresentado com as demais, ou por não apresentarem impacto significativo na eficiência na maioria dos parâmetros das regressões. Nesta situação estão PEA, população total do município, número de empresas e IDH. Permaneceram no modelo as variáveis “dummies”: sudeste (*se*), *sul*, norte (*no*), nordeste

³⁶ Essas variáveis têm origem das seguintes fontes: População e PEA – IBGE Censo 2000; IDH, IDH-educ e renda per capita – PNUD Censo 2000; número de empresas e número de agências no município – Unicad/BACEN 2004.

³⁷ Fonte: Controladoria / Banco do Brasil – 2004.

(*ne*) e grupos de *cluster* (*cl1*, *cl3* e *cl4*); a variável renda *per capita* (*rendapc*), IDH educação (*idheduc*), número de agências concorrentes na praça (*agencia*), média do número de contas da agência (*quantcli*), renda média dos clientes da agência (*rendacli*), pesquisa do “clima” interno da agência (*clima_2004*), grau de satisfação (*grausat*) e escolaridade (*analfabeto*, *fundamental*, *médio* e *superior*).

Inicialmente, para um melhor entendimento do comportamento das variáveis dependentes, em face do impacto das variáveis independentes, estabelece-se, visualmente, a relação destas variáveis em gráficos apresentados no Apêndice 4. Estes gráficos mostram a confrontação de uma variável dependente com uma independente³⁹. As linhas contínuas cinzas identificam, para cada τ (0,1; 0,25; 0,75; 0,9), a estimativa da relação linear entre estas duas variáveis, a linha contínua azul no centro representa a mediana com $\tau = 0,5$ e a linha pontilhada vermelha a média (ajuste dos Mínimos Quadrados Ordinários - MQO).

Os resultados apontam à relação das variáveis dependentes e independentes:

- 1) *Eficcrs4o*: crescimento em todos os quantis quando se considera o grau relativo de eficiência com 4 *outputs* e as onze variáveis independentes (*agencia*, *clima*, *grausat*, *idheduc*, *quantcli*, *rendacli*, *rendapc* e *nível de escolaridade*) decréscimo apenas quando confrontadas com o número de analfabetos e escolaridade de nível fundamental.
- 2) *Eficcrs4*: resultado semelhante ao item anterior.
- 3) *Eficvrs4o*: decrescente com quase todas as variáveis independentes *agencia*, *rendapc*, *idheduc*, *quantcli* e *clima_2004*. As variáveis de nível educacional: *ensino médio* e *ensino superior* são negativas, as demais apresentaram valores positivos. As variáveis *rendacli* e *grausat* são ascendentes.
- 4) *Eficvrs4o*: decréscimo em todos os quantis nas variáveis independentes: *clima* e nas variáveis de ensino *analfabeto* e *médio*. Quanto às variáveis *idheduc*, *quantcli*,

³⁸ Esta variável “*dummiesy*” não é aplicada quando a variável dependente de eficiência é proveniente das agências exclusivas de atendimento no segmento *Pessoa Física*, portanto não apresentada neste trabalho, embora calculada e estimada pelo autor.

³⁹ No Apêndice 4 estão os gráficos para todas as variáveis independentes exceto as “*dummiesy*”, visto que, nesta situação, não teria sentido achá-las. Observa-se, também, que o resultado dos parâmetros são condicionais às demais variáveis em todos os quantis da regressão. Portanto, a validade deste procedimento é limitada ao modelo univariado. O motivo de estar sendo exposto é para uma análise inicial mais detalhada da relação quantílica entre a variável explicada e suas explicativas.

grausat, *rendacli* e *rendapc* são positivas e crescentes. Com a variável *agência* observa-se decréscimo abaixo da linha mediana e crescimento acima dela.

Percebe-se que as variáveis de Renda (*rendacli* e *rendapc*) representadas no modelo, na sua grande maioria, por valores baixos comparativamente (concentradas nas abscissas), quando confrontadas (não condicionalmente) com as variáveis “eficiência” estão dispersas (nas ordenadas), indicando a concentração de grande número de agências localizadas em regiões de baixa renda e que, mesmo assim, oferecem possibilidades de resultados com alta eficiência.

Na seqüência estimou-se as quatro regressões quantílicas variando τ e considerando cinco possibilidades de separatrizes (0.1, 0.25, 0.5, 0.75, 0.9). Os resultados numéricos e os gráficos serão apresentados e analisados nas seções seguintes.

4.3.1 Análise da Regressão Quantílica da Eficiência em CRS como variável dependente – conservando os *outliers* :

Tabela 4.1: Eficiência Pessoa Física - CRS com 4 *outputs* e considerando *outliers*
N = 3551

Coefficients	T = 0.10	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	MQO
(Intercept)	-0.08866*	-0.06364*	0.00328	0.01631	0.07524	-0,001
se	0.00653*	-0.00420	-0.01239*	-0.03726*	-0.09687*	-0,043*
sul	0.01045*	0.00751	0.00173 *	-0.01774*	-0.06690*	-0,022*
no	-0.00883	-0.01929*	-0.03158*	-0.05775*	-0.11175*	-0,063*
ne	0.00186	-0.01048*	-0.02253*	-0.03968*	-0.08959*	-0,043*
Rendapc	0.30842 *	0.44726*	0.87237*	1.14009*	0.98488*	0,819*
Idheduc	0.03620 *	0.00701	-0.01238	-0.01947	0.03364	0,017
agencia	-0.01562	0.01621	0.01395	0.04160*	0.00998	0,025*
Quantcli	0.44388 *	0.43585*	0.45984 *	0.46155*	0.52224 *	0,492*
Rendacli	0.35065 *	0.47759 *	0.70487 *	1.13692*	1.49080*	0,838*
clima_2004	0.03662 *	0.04511 *	0.03840 *	0.02869*	0.02807 *	0,047*
grausat	0.13481 *	0.15782 *	0.11474*	0.13748*	0.08462 *	0,103*
cl1	0.00555 *	0.00456 *	0.00223	0.00075	0.00012	0,005
cl3	-0.01410*	-0.03821*	-0.04489*	-0.07493*	-0.05517*	-0,060*
cl4	0.02461	-0.03833	-0.03727	-0.09114*	-0.02576	-0,051*
Analfabeto	-0.03394	-0.05906*	-0.04829*	-0.08593*	-0.10195*	-0,097*
Fundamental	0.01929	0.00036	0.00392	0.01570	0.03066 *	0,021
medio	-0.02173	-0.01906*	-0.01874*	-0.03149*	-0.07064*	-0,025
Superior	0.07947*	0.05842 *	0.03448*	-0.01490	-0.02307	0,008

* Significativa $\rightarrow \text{Prob}(>|t|) < 0,1$

Observando a Tabela 4.1 e as figuras 4.1⁴⁰ percebe-se que as variáveis regionais (*dummies*) são significativas, negativas e decrescentes em quase todos os quantis.

As variáveis *rendapc* e *rendacl*, que expressam renda, são positivas e significativas em todos os quantis, sendo ambas praticamente crescentes até o último quantil, indicando que quanto maior a renda, existe tendência à maior eficiência.

O impacto das variáveis *idheduc* e *agencia* não tem significância sendo negativo com pequena oscilação decrescente indicando que estas variáveis contribuem negativamente para a eficiência estimada. É importante destacar que elas representam o nível de formação, educação e cultura e a expressão da concorrência bancária de uma mesma localidade, respectivamente. Portanto, teoricamente, em municípios onde existe maior conscientização e onde se instalam um número maior de representantes financeiros existe uma maior concorrência que provoca diminuição da eficiência relativa.

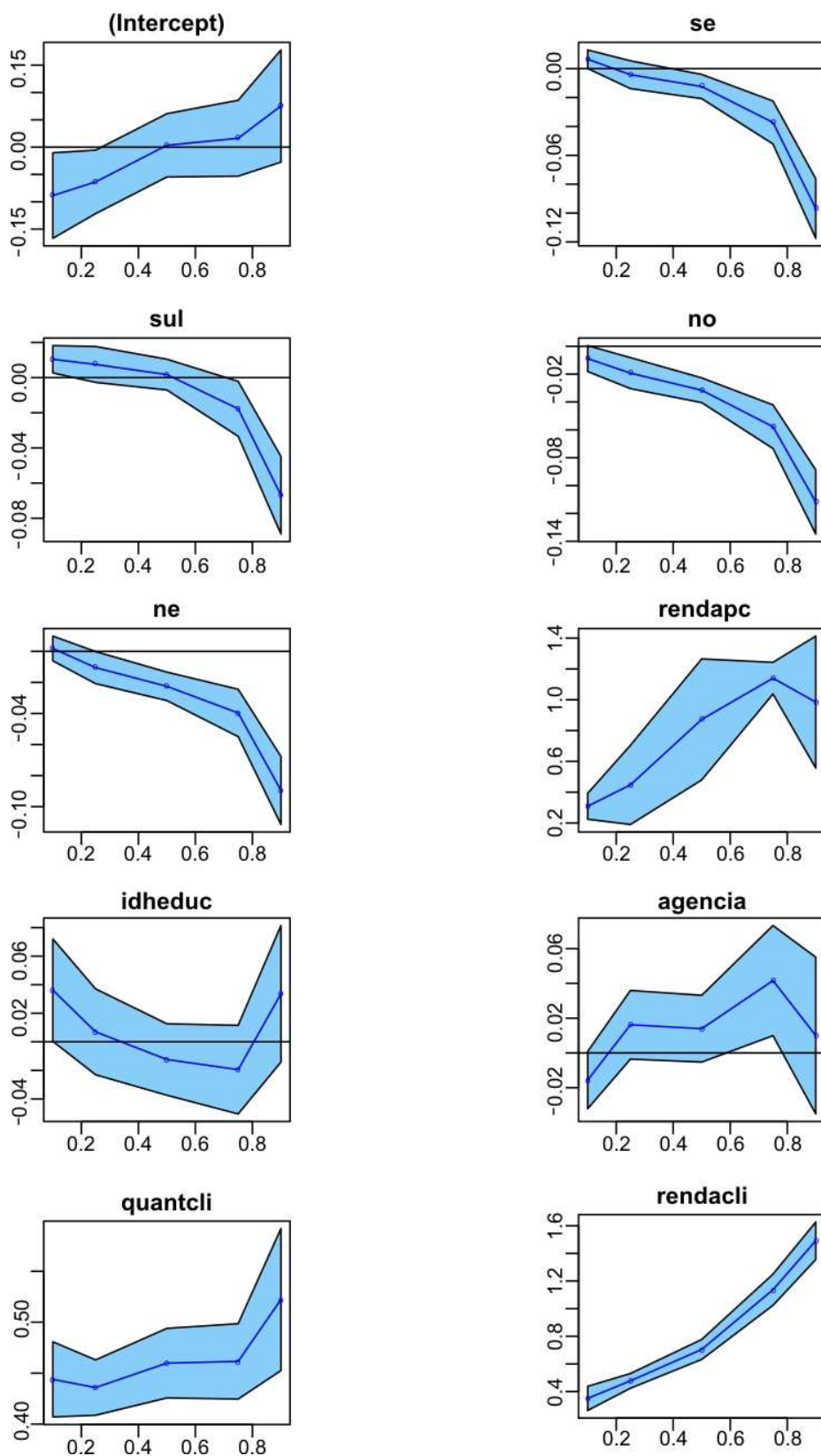
As variáveis que identificam as características individuais de cada DMU mostram valores positivos em *quantcli*, *grausat* e *clima_2004* (todas praticamente estáveis nos quantis). Indicando que todas estas três variáveis contribuem positivamente e na mesma proporção para o grau de eficiência em qualquer nível (ou separatriz) da variável dependente. Isto é, elas são importantes para o impacto à eficiência, no entanto não apresentam grandes variações de comportamento nos quantis.

As indicações dos níveis de escolaridade expressam tendência negativa com a variável *analfabetismo*, mostrando um impacto negativo na eficiência em decorrência de um crescente número de analfabetos nas agências. A escolaridade oscila entre valores negativos e positivos e na sua maioria não são significativos, talvez indicando um problema de atualização cadastral.

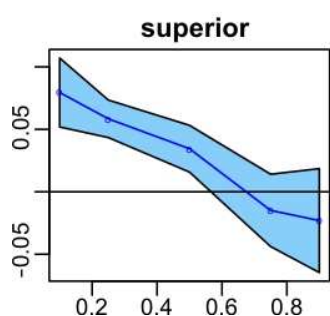
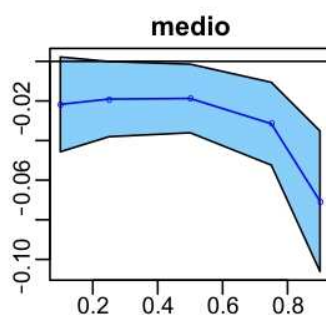
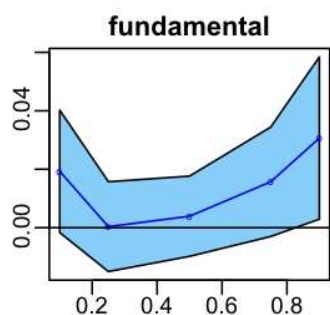
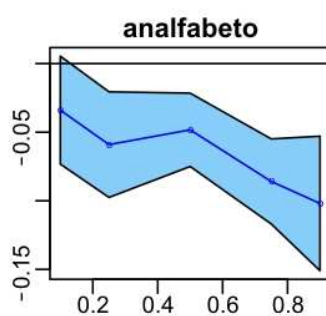
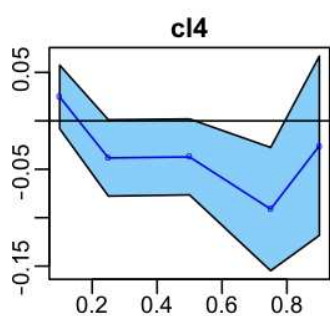
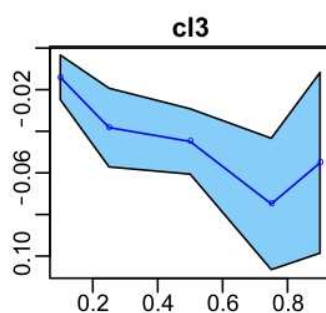
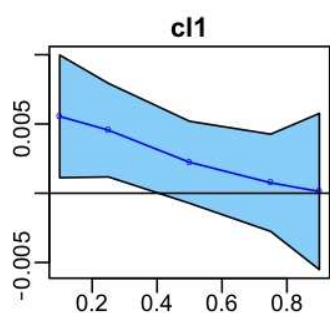
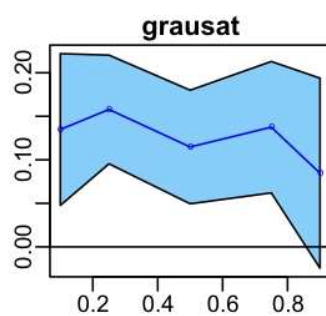
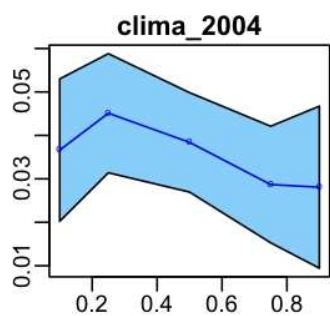
Quanto aos grupos de cluster *cl1*, *cl3* e *cl4* na sua maioria não apresentaram resultados positivos e algumas foram não significativas, indicando que a região de localidade das agências aparentemente não contribui ou altera o grau de eficiência.

⁴⁰ As figuras indicam nas abcissas os cinco quantis definidos na regressão quantílica e nas ordenadas os valores estimados pela regressão para cada variável independente mais o comportamento do intercepto na regressão.

Figuras 4.1: PF com 4 *outputs* e *outliers* - CRS: β do τ quantílico $X \theta$



Figuras 4.1 continuação



4.3.2 Análise da Regressão Quantílica da Eficiência em CRS como variável dependente - retirando os *outliers* :

Tabela 4.2 : Eficiência Pessoa Física - CRS com 4 output retirando os *outliers*
N = 3459

Coefficients	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	MQO
(Intercept)	0.01723	-0.08932	0.10083	0.37321*	0.50507*	0,225*
se	0.02936*	0.02323*	0.01475*	-0.01483	-0.03626*	-0,001
sul	0.01426*	0.03348*	0.02832*	0.00661	-0.00276	0,011
no	0.00177	0.00265	-0.02237*	-0.06280*	-0.09460*	-0,042*
ne	-0.00403	0.00030	-0.02609*	-0.05901*	-0.09979*	-0,046*
rendapc	0.88356*	1.15771*	2.02744*	3.31729*	2.84007*	1,806*
idheduc	-0.03827	-0.06180*	-0.13755*	-0.21781*	-0.20811*	-0,134*
agencia	-0.02358	-0.02599	0.02418	0.09545*	0.09788*	0,036
quantcli	1.07118*	1.05152*	1.01909*	0.93061*	0.79359*	1,065*
rendacli	0.43659*	0.69741*	1.12691*	1.46014*	1.73941*	0,972*
clima_2004	0.10940*	0.10783 *	0.09255*	0.06159*	0.02265	0,087*
grausat	0.24823*	0.43430 *	0.31739*	0.20241*	0.18474	0,221*
cl1	0.01632*	0.01101*	0.00419	-0.00225	-0.00514	0,002
cl3	-0.02075	-0.01869	-0.07539*	-0.12148*	-0.08540*	-0,065*
cl4	0.01355	0.03608	-0.07402*	-0.19207*	-0.20085*	-0,086*
analfabeto	-0.27338*	-0.23651*	-0.26191*	-0.39751*	-0.38512*	-0,344*
fundamental	0.05804*	0.03695 *	0.04941*	0.02949	0.05180*	0,071*
medio	-0.10001*	-0.08653*	-0.03991*	-0.05373*	-0.09903*	-0,077*
superior	0.19323*	0.24173*	0.20557*	0.14534*	0.12072*	0,188*

* Significativa $\rightarrow \text{Prob}(>|t|) < 0,1$

Na Tabela 4.2 é analisada a regressão quantílica confrontando o grau de eficiência relativa das agências Pessoa Física gerado com quatro variáveis de *output* e as demais variáveis independentes citadas no início desta seção (neste caso foram retirados os *outliers* utilizando a distribuição *Heaviside step* para o cálculo do grau de eficiência relativa). Os resultados, mostrados nesta tabela e nas figuras 4.2, indicam que as variáveis *dummies* regionais não são significativas, negativas nos três últimos quantis e decrescentes, isto é, quanto maior o grau de eficiência menor a influência condicional das variáveis regionais.

Também, neste modelo a renda per capita (*rendapc*) e a renda média dos clientes por agência (*rendacli*) mostraram parâmetros significativos e crescentes, indicando, como era de se esperar, que o maior grau de eficiência aconteça nos municípios de maior renda, bem como confirmando a importância da média de renda dos clientes na evolução da eficiência nas unidades. O *idheduc* continua gerando parâmetros negativos, sendo que em $\tau = 0,1$ a

estimativa não foi significativa. Os parâmetros que expressam a concorrência (*agencia*) foram crescentes e apresentaram resultados significativos apenas nos dois últimos quantis indicando maior impacto da concorrência nas localidades de maior eficiência.

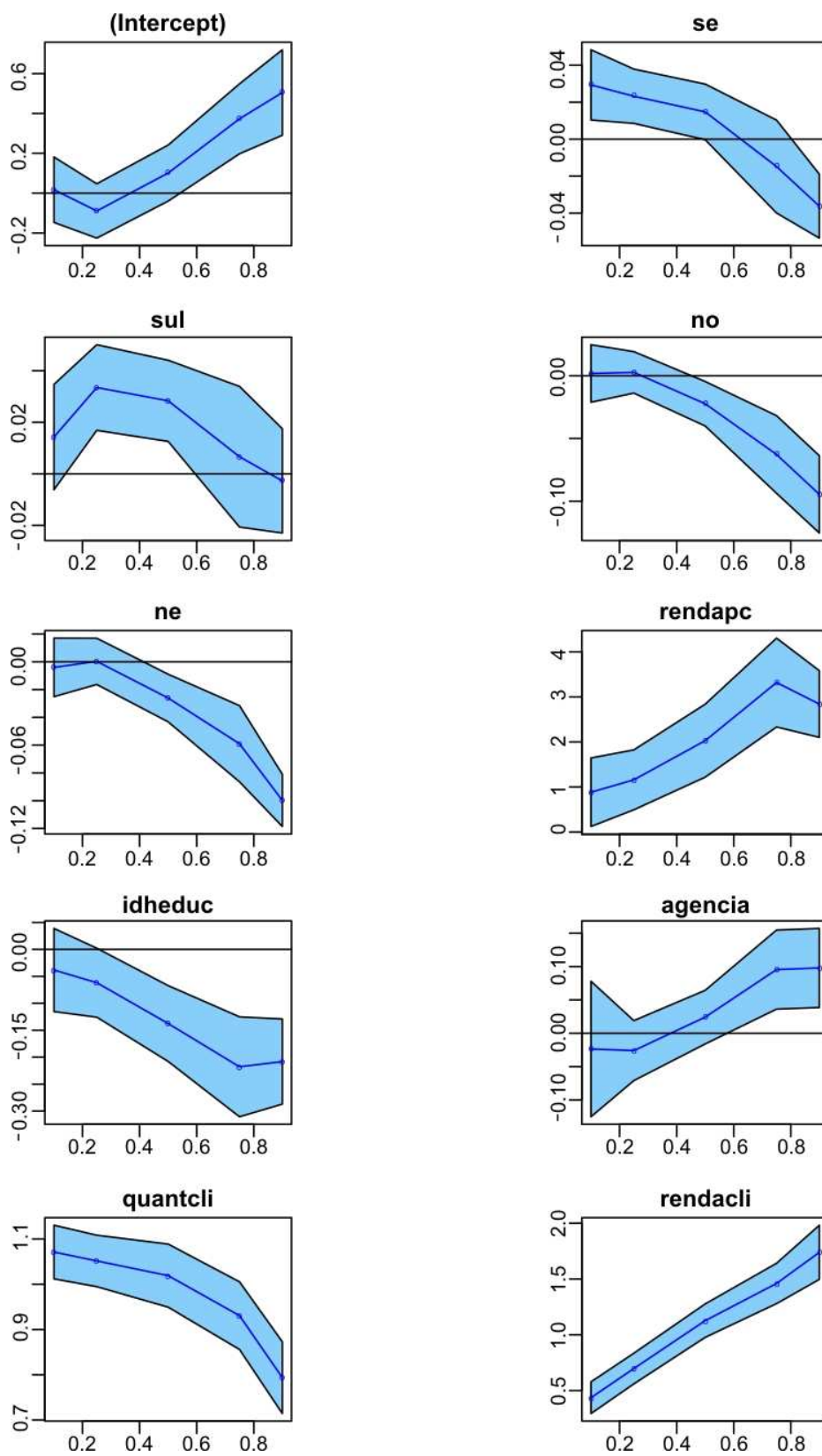
Neste mesmo modelo as variáveis *quantcli* e *clima_2004* foram positivas e decrescentes. Uma explicação plausível para isso é que o impacto condicional à eficiência causado pelo número de contas de clientes é importante para todas as agências. No entanto, naquelas onde estão agências de maior eficiência, o impacto ou influência da quantidade de contas de clientes torna-se menos importante. Em resumo, da mesma forma que o modelo anterior, o grau de eficiência relativa é influenciado pela variável *quantcli*. Isto é, embora o número de clientes e de contas influencie o grau de eficiência, este impacto não é tão grande quanto nas agências de menor eficiência relativa (o que caracteriza para as agências menos eficientes a necessidade de um ganho de escala). A variável *clima_2004* que expressa o nível de relacionamento interno dos funcionários, em determinada unidade, apresentou as mesmas características da variável *quantcli*. Uma das explicações dadas pela empresa é que as agências menores e, portanto, com menor grau de eficiência, normalmente pela grande proximidade entre funcionários e gerentes, podem apresentar um viés na avaliação anual de relacionamento interno. Porém, não se pode descartar a possibilidade de que em agências ineficientes o descontentamento seja maior.

Os parâmetros da variável *grausat* são positivos e significativos (exceto $\tau = 0,90$), embora sem definição de seqüência de crescimento ou decrescimento (especificamente neste caso a satisfação dos clientes impacta a eficiência).

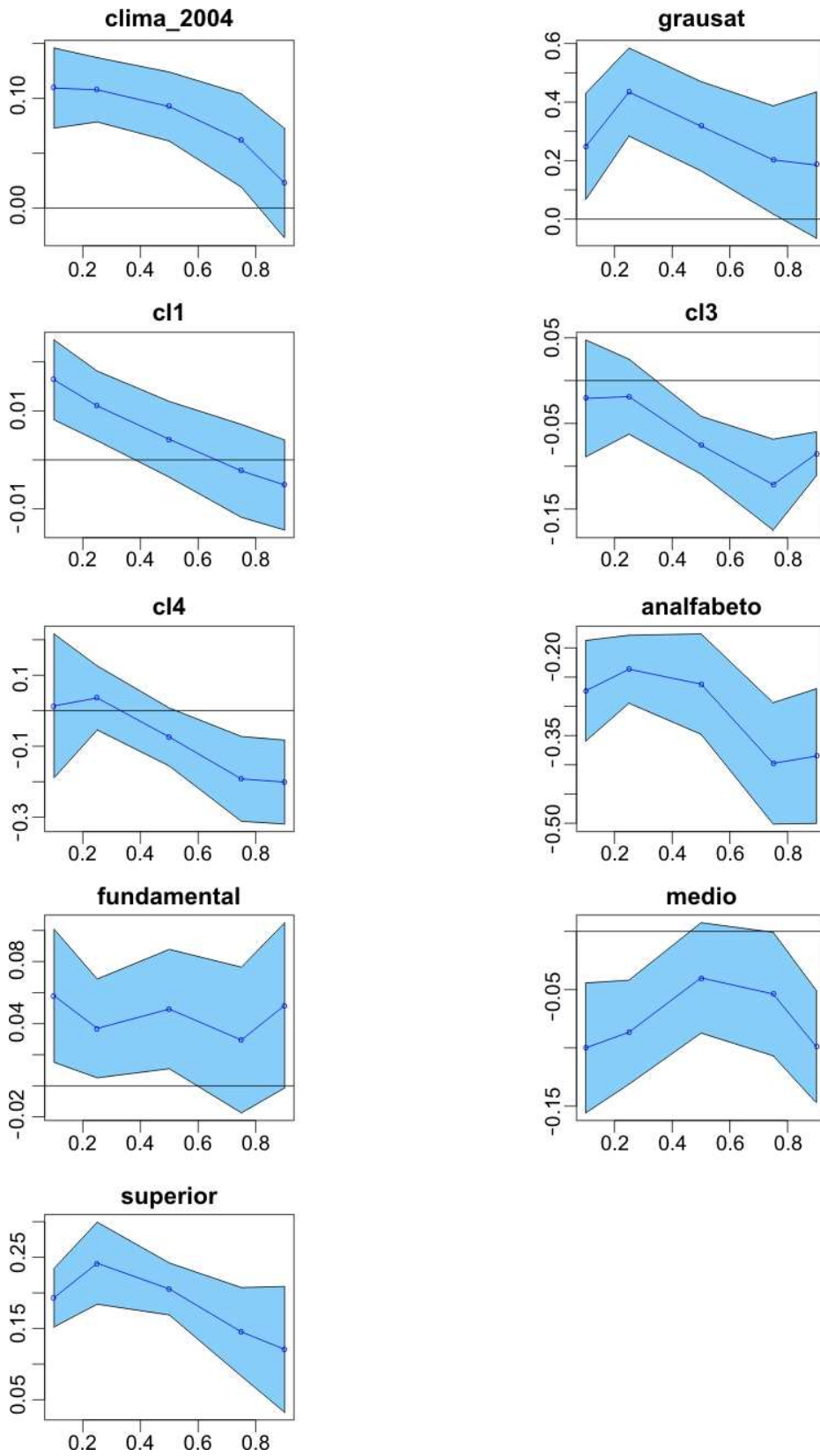
As demais variáveis *cl1*, *cl3* e *cl4* não apresentaram resultados significativos na maioria dos seus quantis mostrando valores negativos e próximos de zero.

Quanto ao nível de ensino, apresentou resultados significativos e interessantes nos parâmetros das variáveis, sendo que o parâmetro da variável *analfabeto* mostra valores negativos e da variável *superior* positivo. Indicando a importância do número de clientes com formação superior para aumentar o grau de eficiência da unidade. Da mesma forma, acontece o contrário com unidades que predomina o índice de analfabetos.

Figuras 4.2: PF com 4 *outputs* sem *outliers* - CRS: β do τ quantílico $X \theta$



Figuras 4.2 continuação



4.3.3 Análise da Regressão Quantílica da Eficiência em VRS como variável dependente – considerando os outliers:

Na Tabela 4.3 e figuras 4.3 são analisados o grau de eficiência do modelo VRS com as mesmas variáveis independentes.

O impacto das variáveis regionais continua negativo em quase todos os quantis e com valores oscilando sem característica de crescimento ou decréscimo.

Tabela 4.3 : Eficiência Pessoa Física - VRS com 4 output considerando os outliers
N = 3551

Coefficients	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	MQO
(Intercept)	0.31521*	0.49475*	0.61398*	0.95083*	1.1285*	0,799*
se	-0.00586*	-0.00982	-0.02140*	-0.03658*	-0.0663*	-0,041*
sul	0.00921*	0.00715	-0.00214	-0.00830	-0.03602	-0,011*
no	-0.03039*	-0.03261*	-0.03466 *	-0.05789*	-0.0932*	-0,066*
ne	-0.01889 *	-0.02459*	-0.03655 *	-0.04719*	-0.053*	-0,042*
rendapc	0.03391	0.37019	0.96214 *	0.81332*	1.3354	0,755*
idheduc	-0.12728*	-0.15226*	-0.19914*	-0.30092*	-0.2154*	-0,228*
agencia	-0.00500	0.00437	0.02420	0.05548*	0.06906	0,038*
quantcli	-0.58513*	-0.58702*	-0.62936*	-0.59383*	-0.5361	-0,528*
rendacl	0.32866*	0.38669*	0.57715 *	0.83659 *	1.0650*	0,625*
clima_2004	-0.05247*	-0.09611*	-0.18484 *	-0.32153*	-0.5502*	-0,331*
grausat	0.21952 *	0.10934*	0.17017 *	0.07562	0.1252*	0,173*
cl1	-0.01960 *	-0.02174*	-0.02560*	-0.02758*	-0.0386*	-0,027*
cl3	-0.02710 *	-0.04407*	-0.06804 *	-0.09941*	-0.1239*	-0,075*
cl4	-0.01067	-0.03582	-0.08818*	-0.15717*	-0.1889	-0,108*
Analfabeto	0.05119	0.09191*	0.11775*	0.09315	0.1819*	0,083
Fundamental	0.03011*	0.02319	0.01671	0.01851	0.02419	0,021
medio	-0.11322*	-0.13667*	-0.16232*	-0.19805*	-0.2321*	-0,177*
superior	-0.03354*	-0.02042	-0.07085*	-0.07775*	-0.1163*	-0,071*

* Significativa $\rightarrow \text{Prob}(>|t|) < 0,1$

Nesta tabela são apresentados os resultados para o modelo VRS com 4 outputs e percebe-se que o comportamento, de uma forma geral, para algumas variáveis não difere muito dos outros modelos mostrados até agora.

As variáveis que expressam as regiões são, na sua maioria, representadas por parâmetros significativos, negativos e decrescentes, exceto a região sul que é não significativa.

Os parâmetros das variáveis que expressam renda (*rendapc* e *rendaclí*) são significativos, positivos e crescentes, apenas o primeiro quantil do parâmetro da variável *rendapc* não é significativo. Novamente indicando que as agências de maior eficiência estão situadas em regiões de maior renda per capita com maior renda média entre seus clientes.

Os parâmetros da variável *agencia* são, na sua maioria, não significativos.

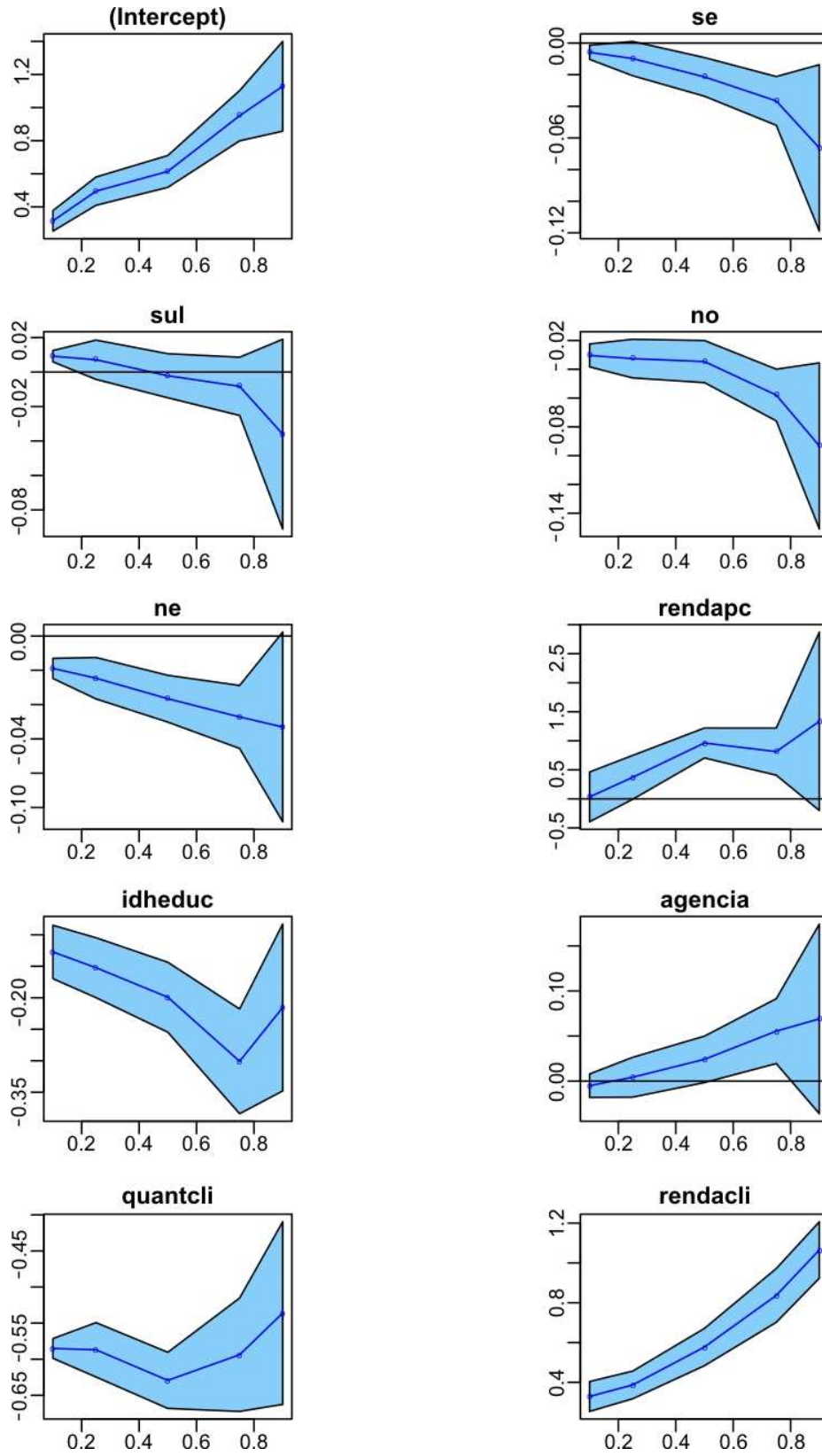
Os resultados para as variáveis *idheduc* e *clima_2004* são significativos, negativos e decrescentes, continuam acompanhando o modelo anterior, muito embora esta última apresentasse uma queda mais acentuada e sempre negativa.

A variável *quantcli* inverteu o sentido do sinal dos parâmetros de positivo para negativo, no entanto, permanecendo estáveis e sem muitas oscilações. Isto porque, o modelo VRS tem indicado uma quantidade maior de agências eficientes, normalmente agências de menor porte com reduzido número absoluto de clientes.

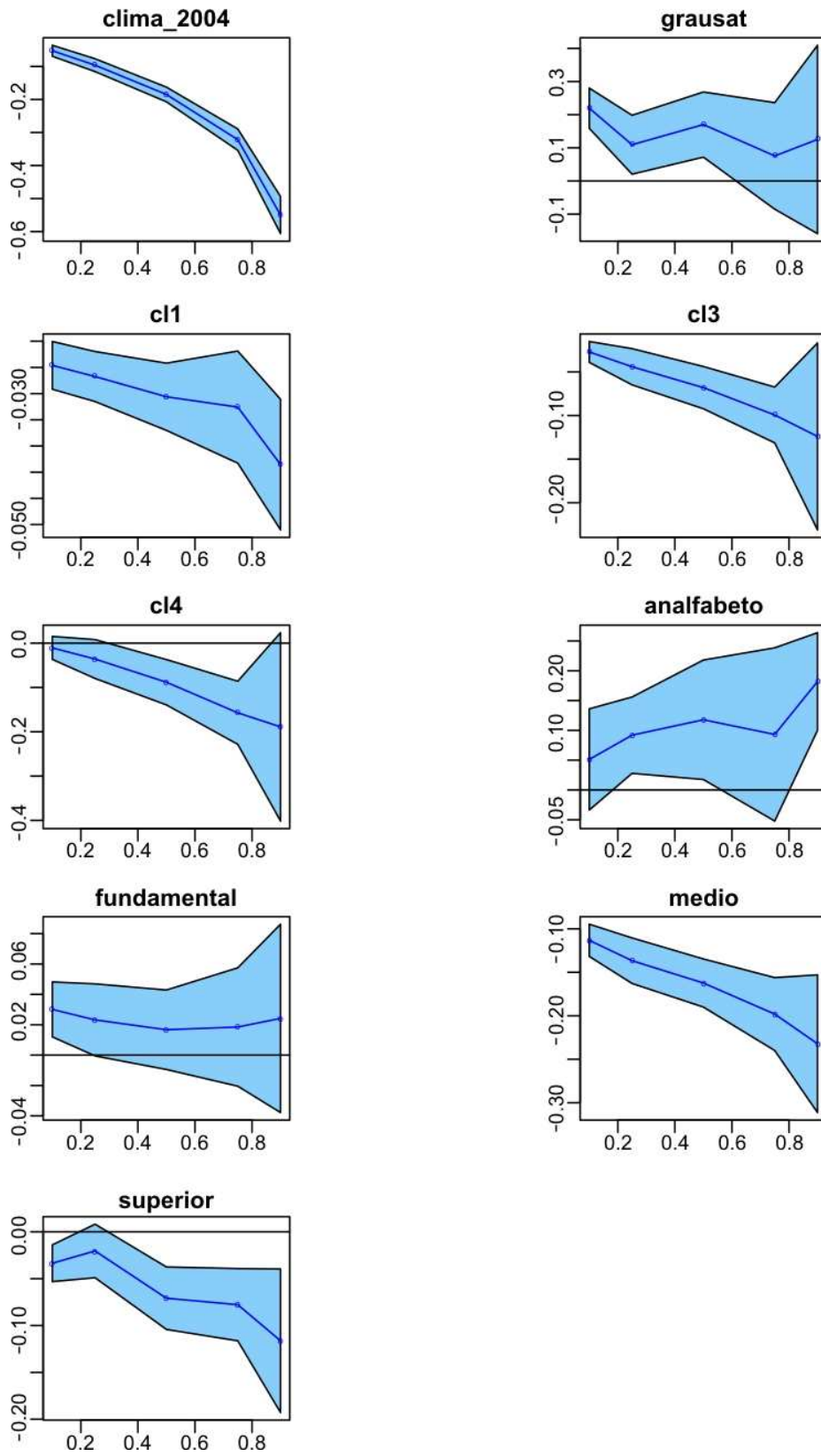
Nos grupos cluster os resultados mostraram valores negativos e significativos.

Os parâmetros das variáveis representativas do nível de escolaridade, em sua maioria, não foram significativos, a não ser o parâmetro da variável ensino médio que apresentou valores negativos e decrescentes. Uma das hipóteses prováveis deste fato diz respeito ao desinteresse de atualização cadastral dessa informação. Diferentemente das variáveis de renda e quantidade. A última pela própria facilidade de apuração, visto que consta nos cadastros automaticamente, e a primeira pela necessidade de atualização para concessão e estabelecimento de limites dos produtos como: renovação de cheque especial, cdc e outros empréstimos. As observações dos valores da Tabela 4.3 podem ser visualizadas nas figuras 4.3 apresentadas nas duas próximas páginas.

Figuras 4.3: PF com 4 *outputs* e *outliers* - VRS: β do τ quantílico $X \theta$



Figuras 4.3 continuação



4.3.4 Análise da Regressão Quantílica da Eficiência em VRS como variável dependente – retirando os *outliers*:

Os resultados da Tabela 4.4 e figuras 4.4 nas variáveis *dummies* que indicam o impacto da região foram negativos, decrescentes e significativos apenas no nordeste. Sinalizando que agências de baixa eficiência relativa estão concentradas nessa região. As demais regiões não apresentaram resultados consistentes.

As variáveis que expressam renda (*rendapc* e *rendacl*) novamente apresentam seus parâmetros positivos e crescentes acompanhando a medida do grau de eficiência (não modificam inclusive sua tendência e escala com relação aos demais resultados analisados nos modelos anteriores).

O parâmetro da variável *idheduc* (IDH– educação) foi negativo e decrescente. Pelo mesmo raciocínio anterior pode-se dizer que o modelo VRS introduz, através da restrição de raio limitado, possibilidades de um número maior de unidades eficientes, principalmente agências de menor porte situadas em pequenos municípios e que podem ter contribuído para esse resultado.

A variável *quantcli* com parâmetros relativamente estáveis e significativos confirma, comparativamente aos demais resultados de outros modelos e nas análises descritivas, que o número de clientes impacta positivamente a eficiência, na maioria dos modelos aplicados (exceto no modelo VRS com *outliers*).

O parâmetro da variável que mede o relacionamento interno da agência (*clima_2004*) foi negativo e decrescente, também no modelo sem *outliers*. Quanto à medida do grau de satisfação dos clientes (*grausat*) o resultado foi positivo apenas nos três primeiros quantis, os demais não foram significativos.

O impacto da diferenciação de mercado causado pela *clusterização* do conjunto de dados, neste caso, foi significativo (com valores negativos) apenas nas agências localizadas nos 566 pequenos e médios municípios do grupo *cl1*. Neste cluster concentram-se as menores agências do Banco situadas em cidades que detém apenas uma agência bancária.

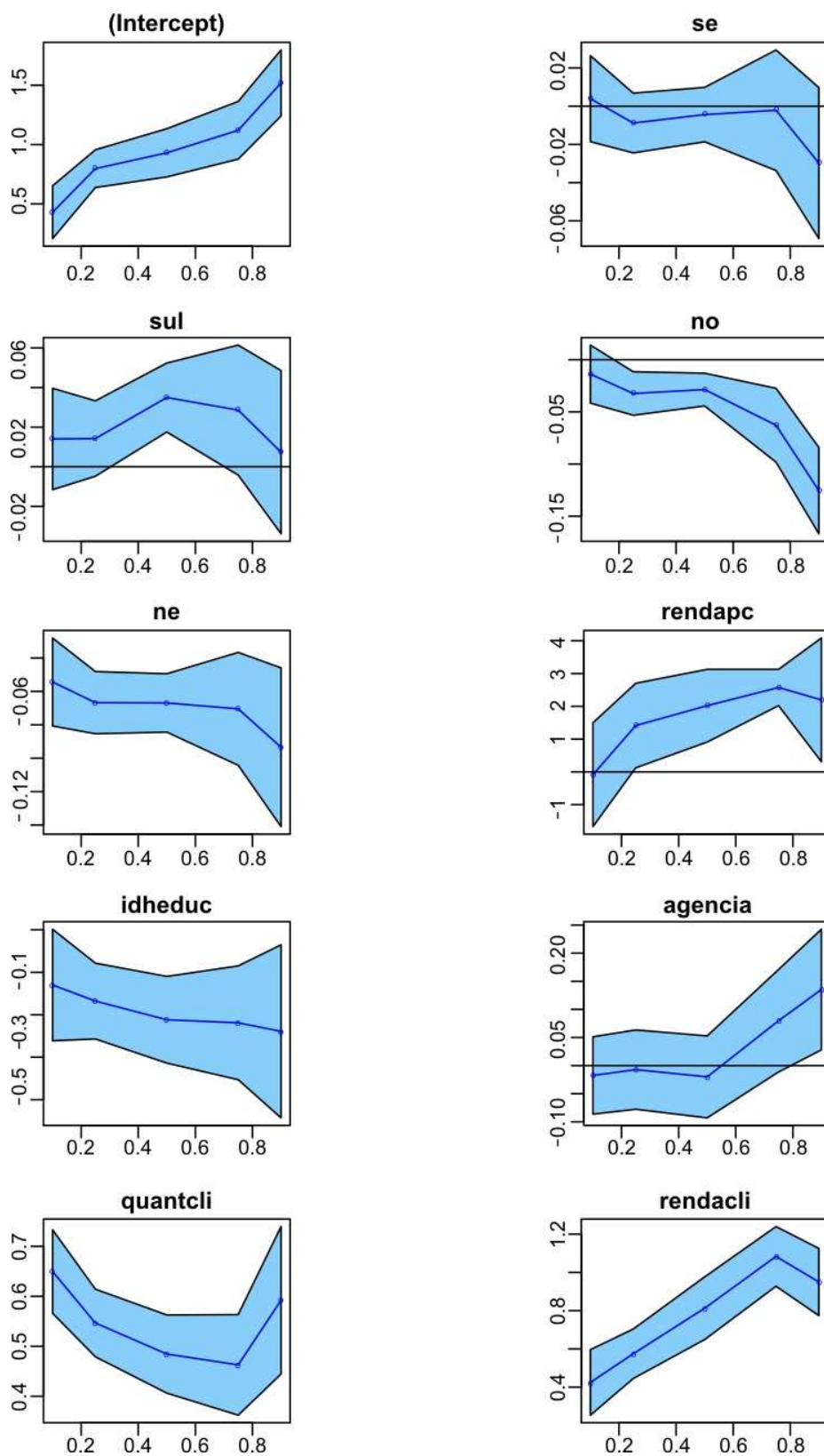
Finalmente, as variáveis indicativas do grau de instrução dos clientes apresentaram resultados significativos e negativos no parâmetro de nível educacional médio.

Tabela 4.4 : Eficiência Pessoa Física - VRS com 4 output retirando os *outliers*
 N = 3400

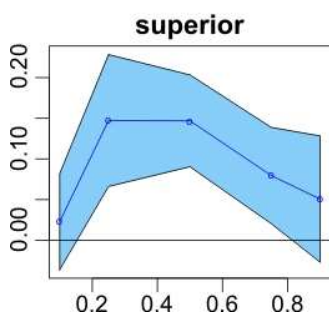
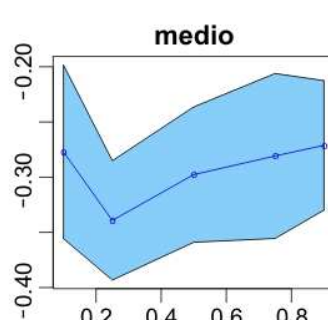
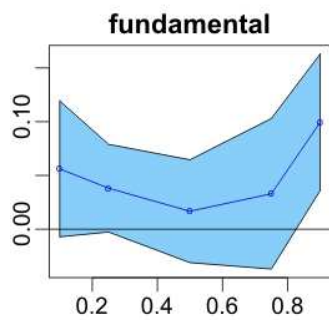
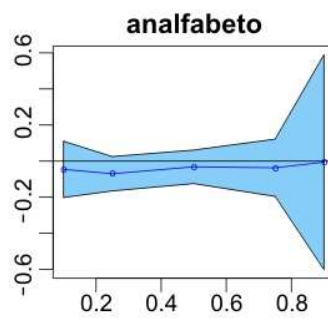
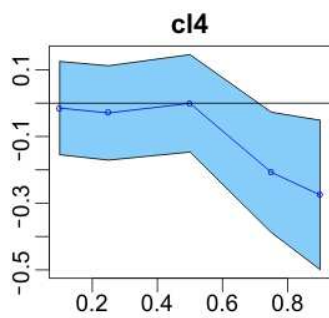
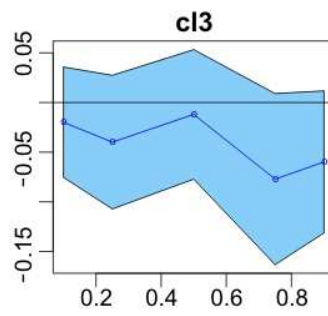
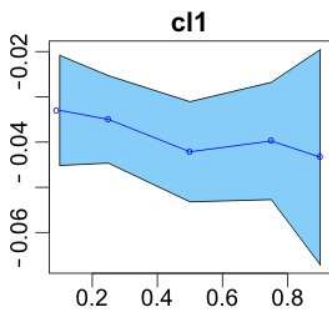
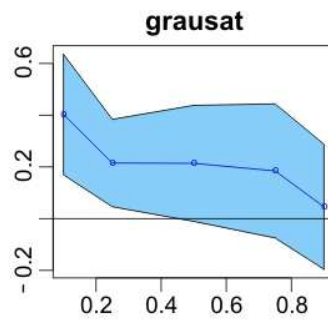
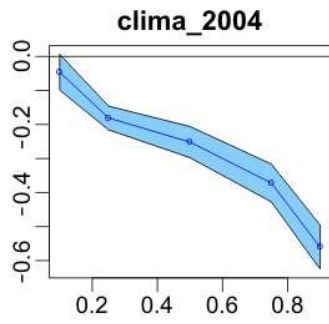
Coefficients	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	MQO
(Intercept)	0.42975*	0.79686*	0.93044*	1.1202*	1.5190*	0,910*
Se	0.00391	-0.00881	-0.00441	-0.00214	-0.02978	0,007
Sul	0.01404	0.01426	0.03496*	0.0286*	0.00739	0,034*
No	-0.01383	-0.03243*	-0.02866*	-0.0628*	-0.1255*	-0,046*
Ne	-0.03435 *	-0.04674*	-0.04690*	-0.0504*	-0.0734*	-0,041 *
Rendapc	-0.08562	1.41411*	2.01966*	2.5742*	2.1960*	1,512*
Idheduc	-0.23002 *	-0.26779*	-0.31185*	-0.3190*	-0.3391*	-0,336*
Agencia	-0.01757	-0.00715	-0.02005	0.0798	0.1353*	0,014
Quantcli	0.64963*	0.54667*	0.48462*	0.4629*	0.5921*	0,543*
Rendacli	0.42473 *	0.57456*	0.81251*	1.0836*	0.9497*	0,755*
clima_2004	-0.04637	-0.18082*	-0.25174*	-0.3718*	-0.5607*	-0,281 *
Grausat	0.40232 *	0.21498*	0.21377	0.18459	0.04590	0,257*
cl1	-0.03299*	-0.03499*	-0.04214*	-0.0398*	-0.0433*	-0,028*
cl3	-0.01986	-0.03989	-0.01197	-0.0771	-0.0598	-0,060*
cl4	-0.01478	-0.02935	-0.00061	-0.2075*	-0.2753*	-0,086*
analfabeto	-0.04586	-0.07015	-0.03171	-0.03712	-0.00570	0,006
fundamental	0.05609	0.0382	0.01671	0.03294	0.0995*	0,061*
medio	-0.27709*	-0.3392*	-0.29766*	-0.2808*	-0.2712*	-0,215*
superior	0.02293	0.1472*	0.14686*	0.0795*	0.05061	0,135*

* Significativa $\rightarrow \text{Prob}(>|t|) < 0,1$

Figuras 4.4: PF com 4 *outputs* sem *outliers* - VRS: β do τ quantílico $X \theta$



Figuras 4.4 continuação



4.3.5 Análise Consolidada

Embora, muitas vezes, os valores relativamente sejam diferentes na apuração dos parâmetros entre os diversos modelos aplicados, ocorrem resultados que obedecem a uma tendência. Os detalhes de cada modelo e tabelas da regressão quantílica apresentada estão representados no Apêndice 5 com seus parâmetros e testes.

Os indicadores regionais, representados pelas variáveis *dummiesy* (*se, sul, no e ne*) não apresentaram grandes diferenças de um modelo para outro, normalmente negativos e decrescentes ou oscilando. Esta situação é um indicador de que essas variáveis, num modelo condicional, não diferenciam a eficiência regionalmente, independente do modelo ajustado.

Nesta mesma linha pode-se concluir a respeito das variáveis *cl1, cl3 e cl4* de onde foram *clusterizados* quatro grupos homogêneos por variáveis de mercado. No contexto, sobre esse assunto no capítulo 3 (seção 3.5.2.2), os agrupamentos foram gerados por semelhança de municípios. Pelo resultado nenhum dos grupos apresentou destaque, de tal forma a concluir que não existe qualquer município exercendo influência maior ou menor no grau de eficiência relativa a não ser, como será visto adiante, pela sua renda. Apenas no modelo VRS o conjunto dos 566 menores municípios, interfere significativamente e negativamente, no grau de eficiência relativa.

Para corroborar com essa afirmativa, retirando-se as variáveis *dummiesy* regionais e *cluster* e posteriormente aplicando a regressão quantílica altera-se apenas, e positivamente, o intercepto dos modelos.

Os estimadores das variáveis *rendapc* e *rendacli* iniciaram positivos e terminaram no último quantil com valores maiores e positivos, portanto foram crescentes durante toda a variação de τ (para todos os modelos apurados). Isso caracteriza que quanto maior a renda per capita do município, maior a renda média dos clientes da agência e maior o grau de eficiência relativa.

No entanto, aparentemente paradoxal, a variável *quantcli* apresentou-se significativa em todos os modelos com importantes diferenças conceituais entre eles. No modelo *CRS 4 retirando os outliers* indicou decréscimo positivo. Isto é, as unidades mais eficientes são impactadas pela quantidade de clientes. No entanto, as demais unidades, para aumentar sua eficiência, necessitam de expandir mais significativamente sua base de clientes.

A variável *agência* (número de concorrentes na praça) não se mostrou significativa na maioria dos quantis dos quatro modelos analisados (apenas nos maiores quantis), porém acompanha a variação da *rendapc* e *rendacli*. O que é normal e conhecido no mercado financeiro a estratégia de determinadas instituições de abrir ou adquirir agências em localidades onde a renda é alta e onde já existem outras instituições.

Os estimadores da variável *idheduc* são negativos diante do crescimento do grau de eficiência relativa e em muitos pontos não significativos, mostrando que não existe ou é difícil de determinar uma associação entre essa variável e o grau de eficiência relativa destas unidades. Refletindo, talvez, uma péssima política educacional ou de distribuição de renda em localidades de maior índice de eficiência e renda.

Os indicadores de escolaridade, componentes do cadastro das unidades, não foram muito conclusivos, talvez indicando o pouco interesse que se tem para atualização cadastral dessas variáveis. Porém, os modelos *CRS* e *VRS* sem *outliers* mostraram resultados significativos e positivos para eficiência nos parâmetros referentes à variável de ensino superior e negativo nos de analfabetos.

Para os parâmetros das variáveis *clima_2004* e *grausat* que medem a satisfação dos funcionários e clientes, respectivamente, são indicadores que contribuem para a eficiência relativa em todos os níveis, exceto nos modelos *VRS* para o indicador de relacionamento interno (*clima_2004*) que, quando significativo, é negativo.

De todos os modelos analisados os melhores resultados de ajustamento, tanto na regressão quantílica como nos mínimos quadrados ordinários (MQO), foram encontrados nos casos do modelo *CRS* com clientes Pessoa Física⁴¹ que, no cálculo da eficiência, utilizaram 4 *outputs*. Para o MQO os resultados do ajustamento das quatro regressões foram:

- a) *CRS-4* com *outliers* : $R=0,677$ e $F=165,864$ com $N=3551$,
- b) *CRS-4* sem *outliers* : $R=0,682$ e $F=166,414$ com $N=3459$,
- c) *VRS-4* com *outliers* : $R=0,577$ e $F=97,652$ com $N=3551$ e
- d) *VRS-4* sem *outliers* : $R=0,469$ e $F=52,968$ com $N=3400$.

⁴¹ Este foi um dos motivos pelos quais as análises foram limitadas aos casos de 4 *outputs* do segmento Pessoa Física com e sem *outliers* das metodologias *CRS* e *VRS*.

O que reforça o fato do modelo CRS se ajustar melhor aos dados. No Apêndice 6 essas regressões estão mais detalhadas com a ANOVA e os testes dos parâmetros para cada variável.

Isso é um indicativo de que o cálculo da eficiência com cinco outputs (incluindo a variável “média da quantidade de contas dos clientes”) não retrata muito bem as expectativas dos indicadores de mercado (neste caso as variáveis exógenas do modelo), isto porque os resultados são influenciados fortemente pela influência importante dessa variável.

Da mesma forma, embora esse trabalho tenha se limitado a apresentar apenas os modelos de regressão quantílica para o segmento de Pessoa Física, as regressões com os demais modelos (calculadas nessa pesquisa) comprovaram a importância das variáveis *dummies jurídica e pública* para identificar modelos diferenciados de gestão que apresentaram um maior grau de eficiência relativa conforme se aumentavam os casos de unidades desses dois segmentos.

A seguir, no capítulo 5, serão mostrados os resultados e conclusões gerais sobre esse trabalho.

CAPÍTULO 5 – CONCLUSÕES E SUGESTÕES

Com as aplicações utilizando variáveis de *input* e *output* que geram o índice de eficiência relativa do DEA, apurou-se um melhor resultado para o modelo CRS em comparação com o modelo VRS (no caso desta aplicação específica) pelas seguintes evidências:

- 1) Há diferenças significativas avaliadas nos testes paramétricos e não paramétricos entre os dois modelos, tanto para a diferença dos valores de eficiência quanto para o ordenamento das unidades. Portanto, os modelos apresentam diferenças significativas no *ranking* de agências para o caso de cinco *outputs* bem como no de quatro *outputs*.
- 2) Esses dois modelos quando submetidos à seleção de *outliers*, através dos cálculos dos *leverages* e aplicação das distribuições, geram unidades diferentes de impacto. O modelo CRS com a distribuição *Heaviside step* e quatro *outputs* detectou, como *outliers*, 100% das unidades que compõem o segmento pessoa jurídica e governo (pública). Esses segmentos têm uma característica de gestão totalmente diferenciada e adaptada pelo tipo de cliente atendido, em conformidade com o resultado CRS, identificados, também, na regressão quantílica. Os segmentos em questão são notabilizados pela utilização de poucos recursos com pequenas agências semelhantes a escritórios que, por consequência, apresentam número reduzido de funcionários (*inputs*). Por outro lado, gerando grandes resultados (*outputs*) característicos da capacidade e potencial financeiro elevado de seus clientes. As variáveis *publica* e *jurídica* são significativas e crescentes confirmando que os segmentos de clientes do Banco são importantes para o impacto no grau de eficiência, independente de se considerar o número de contas de clientes como uma variável de influência no cálculo de θ .
- 3) Quando da utilização da regressão quantílica o modelo CRS mostrou-se mais adequado, visto que variáveis conhecidamente influentes apresentaram parâmetros de significativo impacto e direcionamento coerente com sua influência. Esse resultado isoladamente não justifica a escolha do modelo. No entanto, associado com as justificativas anteriores, consolida a afirmativa.

Evidentemente que esta afirmativa está limitada ao tipo de aplicação empírica utilizada neste trabalho. No entanto, há fortes indícios de que se possa ampliar essa conclusão, cabendo a outros estudos teóricos e aplicados sua confirmação.

Embora o modelo com 5 outputs tenha gerado um resultado com maior média, mais homogêneo e com menor assimetria, aproximando-se mais da distribuição Normal, o modelo com 4 outputs está mais ajustado a realidade dos interesses de uma instituição financeira. As unidades mais rentáveis são privilegiadas quando não é considerada a variável “média da quantidade de contas de clientes” como *output*.

A influência de variáveis como renda *per capita* do município (*rendapc*) e renda média dos clientes da agência (*rendacli*) são importantes e obedecem uma coerência para futuros investimentos. Isto é, em municípios de maior renda per capita, tanto maior será a média de renda dos clientes das agências ali localizadas e, conseqüentemente, mais eficientes essas agências serão. Sobre a *performance* das agências em relação a essas duas variáveis, interessante seria uma comparação entre elas e o grau de eficiência relativa. Isso permite apurar o segmento de renda que compõe a carteira de clientes do Banco do Brasil e se essa carteira está em conformidade com a renda *per capita* do município em que ela está inserida.

Nota-se, também, que a variável quantidade de clientes da agência (*quantcli*), na análise descritiva e na regressão quantílica, provoca impacto positivo na maioria dos modelos (com exceção do modelo VRS com *outliers*), o que confirma o fato de que, para estas unidades, qualquer estratégia de ampliação da base de clientes é correta. No entanto, se a estratégia da empresa é atender diversas localidades de menor recurso uma atitude adequada é buscar o maior número de clientes possíveis ganhando em escala e equilibrando a diferença de rentabilidade individual.

A eficiência pode ser encontrada em qualquer unidade independente da região considerada. O que faz a diferença no crescimento da eficiência é a renda *per capita* da localidade de atuação da unidade e o quanto essa renda é “transferida” para cada unidade compondo sua renda média. A exceção se faz em localidades onde a concorrência é acentuada e com grande número de agências de outros Bancos. Isso ocorre, principalmente, em São Paulo, Brasília, Rio de Janeiro e Belo Horizonte, localidades onde o Banco deveria

intensificar sua atuação de expansão da sua base de clientes e, principalmente, aumento da margem individual de contribuição de seus clientes.

A *clusterização* por variáveis de mercado é necessária apenas por uma melhoria administrativa e facilidade de gestão e não para uma melhoria na eficiência econômica e financeira. Isto porque existe homogeneidade nos indicadores de eficiência relativa avaliados separadamente dentro dos grupos de clusters, porém quando avaliados sob o mesmo modelo DEA e submetidos à regressão quantílica os resultados não são tão conclusivos.

As informações cadastrais da instituição devem ser mais cuidadosas, principalmente naquelas que não indicam renda. Estas informações são de grande importância quando se deseja conhecer o perfil e tendências ao potencial bancário dos clientes.

O modelo estratégico de três segmentos, adotado pelo Banco, é adequado e eficiente. Embora não mostrado nesse trabalho as regressões quantílicas desenvolvidas com variáveis dummies, apresentaram resultados onde o grau de eficiência relativa é maior nos segmentos Governo, Pessoa Jurídica e Pessoa Física, respectivamente, corroborando com a análise descritiva da distribuição de θ na tabela 3.13.

A utilização das distribuições dos *leverages*, principalmente a distribuição *Heaviside step* (com menor grau de arbitrariedade), é extremamente eficiente para selecionar *outliers* em indicadores DEA.

A aplicação deste tipo de técnica DEA para seleção de *outliers* e a metodologia de regressão quantílica se encaixam muito bem nas necessidades de instituições desta natureza, cabendo a estudos posteriores a ampliação para utilização em atividades de auditoria e controladoria, e que influenciam diretamente na concessão e acompanhamento do crédito, distribuição e alocação de unidades de produtos e clientes, avaliação e acompanhamento de risco.

Estes estudos direcionam para a observação e acompanhamento de melhores práticas (*benchmark*) ou, por outro lado, tratando-se de instituições financeiras, o controle e acompanhamento de possíveis atividades que fogem dos padrões aceitáveis fazendo com que os agentes possam interferir em tempo hábil.

Após pesquisa teórica e aplicada no Banco do Brasil para o ano de 2004 algumas sugestões são recomendáveis para futuros trabalhos utilizando modelos DEA e seleção de *outliers*:

1. No que se refere ao Planejamento Estratégico o modelo utilizado contribui no sentido de motivar para a necessidade de reavaliar os seguintes pontos:
 - a) Abertura ou remanejamento de agências para compartilhar clientes de outras unidades menos eficientes e que, pelo resultado, não justificam o funcionamento sem vínculo a outra agência com aproximação geográfica e, ao mesmo tempo, semelhança na fronteira de produção. Para isso deve-se analisar com mais detalhes a dominância na fronteira de produção, detectar as dominantes e dominadas para a separação homogênea e sugestão de adequação.
 - b) Da mesma forma, aquelas unidades caracterizadas por grande impacto no *output* “média do número de contas de clientes”, deveriam ser reavaliadas para o remanejamento de máquinas e infra-estrutura de atendimento on-line, inclusive no que se refere à adequação do número de pessoal – gerentes e funcionários -, bem como do *mix* ótimo de produtos oferecidos para esses clientes.
2. Definição de um modelo para Análise de Risco das agências no momento em que se apura o perfil da unidade (aplicadora ou captadora) na comparação com as demais unidades. Geralmente as agências de perfil aplicadora obedecem um limite de valores para empréstimos de acordo com o número de clientes e esse aspecto é facilmente perceptível no modelo de eficiência DEA. Quando o limite é ultrapassado é imediatamente detectado na seleção dos *leverages* com utilização das distribuições apropriadas. Desta maneira, monitora os processos de gestão de risco, auxilia na inibição de ações irregulares e na indicação de eventuais falhas para que sejam sanadas tempestivamente.
3. Medir a dinâmica da variável eficiência utilizando valores de *input* ou *output* com variabilidade temporal. A possibilidade e facilidade de se levantar dados de anos anteriores, bem como de 2005, viabiliza a estimativa temporal do comportamento do grau de eficiência de cada unidade e sua variação no decorrer desses anos,

estabelecendo um comportamento crescente ou decrescente de eficiência, o que é muito interessante para uma avaliação mais detalhada.

4. Principalmente deve-se analisar melhor, dentro dos segmentos, o motivo da posição de impacto para cada unidade, trabalhando profundamente as alterações no grau de eficiência e nas comparações dos parâmetros no uso da regressão quantílica quando se modifica (retirando ou acrescentando) os *inputs* e *outputs* do modelo.

Como descrito acima o modelo colabora para uma reavaliação dos recursos, avaliação de risco (da agência), redimensionamento das agências, abertura e fechamento de unidades de produção, dotação, classificação ou “*ranqueamento*”, bem como outras aplicações desde que adequadamente definidos os *inputs* e *outputs* de interesse.

A partir desse trabalho conclui-se que o modelo DEA torna-se um eficaz e competente instrumento, não paramétrico, de gestão corporativa para a indústria financeira.

APÊNDICE 1

NPar Tests

Friedman Test

Ranks

	Mean Rank
ord_crs	2.38
ord_vrs	2.58
ord_crs4o	2.44
ord_vrs4o	2.60

Test Statistics^a

N	3663
Chi-Square	77.848
df	3
Asymp. Sig.	.000

a. Friedman Test

Kendall's W Test

Ranks

	Mean Rank
ord_crs	2.38
ord_vrs	2.58
ord_crs4o	2.44
ord_vrs4o	2.60

Test Statistics

N	3663
Kendall's W ^a	.007
Chi-Square	77.848
df	3
Asymp. Sig.	.000

a. Kendall's Coefficient of Concordance

NPar Tests

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
ord_vrs - ord_crs	Negative Ranks	1554 ^a	1971.06	3063025.50
	Positive Ranks	2094 ^b	1715.74	3592750.50
	Ties	15 ^c		
	Total	3663		
ord_crs4o - ord_crs	Negative Ranks	1783 ^d	1837.44	3276149.00
	Positive Ranks	1864 ^e	1811.15	3375979.00
	Ties	16 ^f		
	Total	3663		
ord_vrs4o - ord_vrs	Negative Ranks	1736 ^g	1810.43	3142900.50
	Positive Ranks	1853 ^h	1780.55	3299354.50
	Ties	74 ⁱ		
	Total	3663		

a. ord_vrs < ord_crs

b. ord_vrs > ord_crs

c. ord_vrs = ord_crs

d. ord_crs4o < ord_crs

e. ord_crs4o > ord_crs

f. ord_crs4o = ord_crs

g. ord_vrs4o < ord_vrs

h. ord_vrs4o > ord_vrs

i. ord_vrs4o = ord_vrs

Test Statistics^b

	ord_vrs - ord_crs	ord_crs4o - ord_crs	ord_vrs4o - ord_vrs
Z	-4.163 ^a	-.785 ^a	-1.260 ^a
Asymp. Sig. (2-tailed)	.000	.432	.208

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Sign Test

Frequencies

		N
ord_vrs - ord_crs	Negative Differences ^{a,b,c}	1554
	Positive Differences ^{d,e,f}	2094
	Ties ^{g,h,i}	15
	Total	3663
ord_crs4o - ord_crs	Negative Differences ^{a,b,c}	1783
	Positive Differences ^{d,e,f}	1864
	Ties ^{g,h,i}	16
	Total	3663
ord_vrs4o - ord_vrs	Negative Differences ^{a,b,c}	1736
	Positive Differences ^{d,e,f}	1853
	Ties ^{g,h,i}	74
	Total	3663

- a. ord_vrs < ord_crs
- b. ord_crs4o < ord_crs
- c. ord_vrs4o < ord_vrs
- d. ord_vrs > ord_crs
- e. ord_crs4o > ord_crs
- f. ord_vrs4o > ord_vrs
- g. ord_vrs = ord_crs
- h. ord_crs4o = ord_crs
- i. ord_vrs4o = ord_vrs

Test Statistics^a

	ord_vrs - ord_crs	ord_crs4o - ord_crs	ord_vrs4o - ord_vrs
Z	-8.924	-1.325	-1.936
Asymp. Sig. (2-tailed)	.000	.185	.053

- a. Sign Test

APÊNDICE 2

Nome	UF	lev. hits	total hits	lev	Eficiência
VERA CRUZ	RS	98	122	0,1304	1,0000
PLANALTO	DF	89	100	0,0556	1,0000
VICENTE MACHADO	PR	80	110	0,0502	0,6848
SANTA CRUZ DO SUL	RS	78	99	0,0486	0,9219
ENTRE RIOS	PR	97	102	0,0485	0,9520
ACREUNA	GO	85	91	0,0447	1,0000
SINIMBU	RS	85	102	0,0420	0,6053
PETROBRAS	RJ	78	86	0,0406	0,6848
ITAIM BIBI	SP	74	80	0,0389	1,0000
TRIANON	SP	84	100	0,0321	0,6009
CORURIBE	AL	85	107	0,0281	0,6134
CAMARA CASCUDO	RN	87	100	0,0275	0,6602
HERACLITO GRACA	CE	82	101	0,0261	0,5294
POSTALIS	DF	87	96	0,0255	1,0000
CURITIBA	PR	81	102	0,0250	0,4480
CAMPO VERDE	MT	76	85	0,0247	0,9849
SAO DESIDERIO	BA	85	97	0,0227	0,8268
SUDOESTE 303	DF	98	102	0,0223	0,9183
VENANCIO AIRES	RS	80	113	0,0200	0,5895
BOULEVARD SAO JOAO	SP	88	117	0,0197	0,6136
CHAPADAO DO CEU	GO	83	94	0,0189	0,9494
MONTIVIDIU	GO	86	103	0,0163	0,8320
PARANAGUA	PR	74	121	0,0163	0,4571
PRACA DO CARMO	CE	65	95	0,0136	0,5379
SINOP	MT	80	95	0,0124	0,8453
EUSEBIO	CE	66	107	0,0121	0,5885
SENADO FEDERAL	DF	65	109	0,0116	0,4921
SETOR COMERCIAL SUL	DF	79	102	0,0115	0,5460
AMERICA DOURADA	BA	86	119	0,0114	0,6251
PEDRA PRETA	MT	82	98	0,0112	0,8353
CAMPOS DE JULIO	MT	85	102	0,0111	0,9352
PALACIO DAS ARTES	MG	74	101	0,0109	0,5028
RIO VERDE	GO	92	110	0,0107	0,8597
AVENIDA IMPERADOR	CE	63	107	0,0101	0,5915
CAAPORA	PB	53	100	0,0101	0,5707
VILA MILITAR	RJ	58	87	0,0098	0,5681
BNDDES	RJ	75	87	0,0096	0,5250
ASA BRANCA	RR	78	107	0,0093	0,5981
BACEN	RJ	77	103	0,0093	0,5818
PONTAL	SP	58	107	0,0092	0,5637
SAO GOTARDO	MG	63	90	0,0091	0,6317
NOVA MUTUM	MT	68	118	0,0091	0,8004
SILVANIA	GO	67	102	0,0087	0,6529

Nome	UF	lev. hits	total hits	lev	Eficiência
PREFEITURA DO RECIFE	PE	89	116	0,0084	0,6355
EDEIA	GO	56	101	0,0082	0,6561
TRIBUNAIS	DF	73	99	0,0081	0,6030
JOVIANIA	GO	42	80	0,0078	0,7367
ESCOLA PLTA.MEDICINA	SP	68	97	0,0075	0,6089
POJUCA	BA	45	93	0,0075	0,4006
PETROBRAS	SP	67	93	0,0074	0,6486
CHAPADAO DO SUL	MS	61	95	0,0072	0,7442
CAMINHO DO MEIO	RS	75	105	0,0071	0,6162
AV.ESTADOS UNIDOS	BA	35	94	0,0070	0,3785
VALE DO SOL	RS	62	95	0,0068	0,4516
MAJESTIC	RS	61	100	0,0068	0,5568
RIO VERMELHO	MT	86	124	0,0067	0,8458
SAO MIGUEL DO TAPUIO	PI	77	109	0,0066	0,6033
NOVA SUICA	GO	58	102	0,0065	0,6088
PARLAMENTO	DF	65	103	0,0064	0,5643
CANGUCU	RS	41	91	0,0058	0,5307
SENADOR FIRMINO	MG	45	79	0,0055	0,4206
SOBRADINHO	RS	57	111	0,0055	0,5403
EDIFICIO SEDE BACEN	DF	58	94	0,0055	0,5755
RONDONOPOLIS	MT	58	93	0,0055	0,6540
REPUBLICA DO LIBANO	GO	54	116	0,0054	0,5673
PRIMAVERA DO LESTE	MT	83	116	0,0054	0,8072
CASA NOVA	BA	68	112	0,0054	0,5351
POTIGUAR	RN	75	118	0,0053	0,7582
IPIUA	SP	56	87	0,0051	0,6728
PALACIO DO TRABALHO	RJ	54	113	0,0051	0,4595
IPANGUACU	RN	39	94	0,0050	0,5964
ITAIPULANDIA	PR	56	75	0,0050	0,6851
TRIBUNAL JUSTICA DF	DF	51	99	0,0049	0,5305
CAMPO GRANDE	RN	40	86	0,0049	0,5453
MAURILANDIA	GO	58	107	0,0048	0,6515
RUA TREZE DE MAIO	PB	43	106	0,0047	0,5117
TUPANCIRETA	RS	61	113	0,0045	0,6879
CASCAVEL	CE	35	98	0,0043	0,5711
QUIRINOPOLIS	GO	53	103	0,0042	0,5935
PARQUE TOMAS EDSON	SP	69	101	0,0041	0,5911
PALACIO DA FAZENDA	RJ	53	103	0,0041	0,5249
MINEIROS	GO	43	91	0,0041	0,6183
MARACANAU	CE	43	94	0,0041	0,5812
TIJUCA	RJ	68	108	0,0039	0,5115
VILA HELENA	MS	77	101	0,0039	0,9088
UFSCAR	SP	56	102	0,0036	0,5667
BARUERI	SP	25	86	0,0036	0,4745

Nome	UF	lev. hits	total hits	lev	Eficiência
CAMPO NOVO PARECIS	MT	70	108	0,0035	0,7014
SORRISO	MT	61	111	0,0034	0,7037
SHOP.LITORAL NORTE	BA	51	100	0,0034	0,6457
IPANEMA	RJ	52	112	0,0034	0,4898
AVENIDA RIO BRANCO	RN	39	94	0,0034	0,4249
FRONTEIRA	MG	37	90	0,0033	0,5496
SAO JORGE DO IVAI	PR	46	103	0,0032	0,7368
SIDROLANDIA	MS	49	110	0,0031	0,6034
ASA SUL 302	DF	59	113	0,0031	0,4774
PRACA TIRADENTES	PR	44	102	0,0029	0,4996
MARILANDIA DO SUL	PR	38	99	0,0029	0,4552
RECEITA FEDERAL	DF	37	90	0,0028	0,5286
CANGUARETAMA	RN	33	100	0,0026	0,5644
LUCAS DO RIO VERDE	MT	35	101	0,0024	0,7127
PRESIDENTE DUTRA	RO	14	94	0,0024	0,5086
POCOS DE CALDAS	MG	23	97	0,0022	0,3807
SETOR MILITAR URBANO	DF	39	95	0,0022	0,6053
GRAPIUNA	BA	40	108	0,0022	0,5633
ARACRUZ	ES	30	109	0,0020	0,4375
SAPEZAL	MT	54	90	0,0020	0,6269
UNIV.DE SAO PAULO	SP	54	113	0,0020	0,5345
LEBLON	RJ	23	96	0,0020	0,4477
PRACA PIO XII	ES	54	99	0,0020	0,5499
RUA URUGUAI	RS	35	103	0,0019	0,4771
GUAICURUS	MS	46	109	0,0019	0,6611
TAPURAH	MT	62	110	0,0018	0,5980
LAGOA REAL	BA	47	99	0,0017	0,5708
UNIV.DE BRASILIA	DF	36	85	0,0017	0,5634
MINISTERIO FAZENDA	SP	37	112	0,0017	0,4719
LUIS E.MAGALHAES	BA	32	96	0,0016	0,5753
PRACA VINTE	RS	25	102	0,0016	0,5028
VINTE E CINCO AGOSTO	RJ	27	113	0,0016	0,4501
AFONSO PENA	MS	27	103	0,0015	0,5920
RIO	RJ	28	96	0,0015	0,4977
IATI	PE	35	100	0,0015	0,5642
SAO GABRIEL DO OESTE	MS	20	118	0,0014	0,6475

APÊNDICE 3

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
eficiencia	3663	.38412329	.098704583	.043495	1.000000
codigo	3663	1.04	.245	1	3

Kruskal-Wallis Test

Ranks

	codigo	N	Mean Rank
eficiencia	1	3552	1832.61
	2	73	1436.91
	3	38	2534.03
	Total	3663	

Test Statistics^{a,b}

	eficiencia
Chi-Square	26.934
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: codigo

Median Test

Frequencies

		codigo		
		1	2	3
eficiencia	> Median	1777	28	26
	<= Median	1775	45	12

Test Statistics^b

	eficiencia
N	3663
Median	.37554600
Chi-Square	9.118 ^a
df	2
Asymp. Sig.	.010

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 19.0.

b. Grouping Variable: codigo

Oneway

ANOVA

eficiencia					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.170	2	.585	62.069	.000
Within Groups	34.507	3660	.009		
Total	35.677	3662			

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
eficiencia	3660	.40377351	.141437080	.089432	1.000000
código	3660	1.99	.629	1	4

Kruskal-Wallis Test

Ranks

	código	N	Mean Rank
eficiencia	1	566	2970.54
	2	2724	1478.52
	3	199	2347.12
	4	171	3062.76
	Total	3660	

Test Statistics^{a,b}

	eficiencia
Chi-Square	1241.146
df	3
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: código

Median Test

Frequencies

	código			
	1	2	3	4
eficiencia > Median	555	973	149	153
eficiencia <= Median	11	1751	50	18

Test Statistics^b

	eficiencia
N	3660
Median	.36758250
Chi-Square	900.889 ^a
df	3
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 85.5.

b. Grouping Variable: código

Oneway

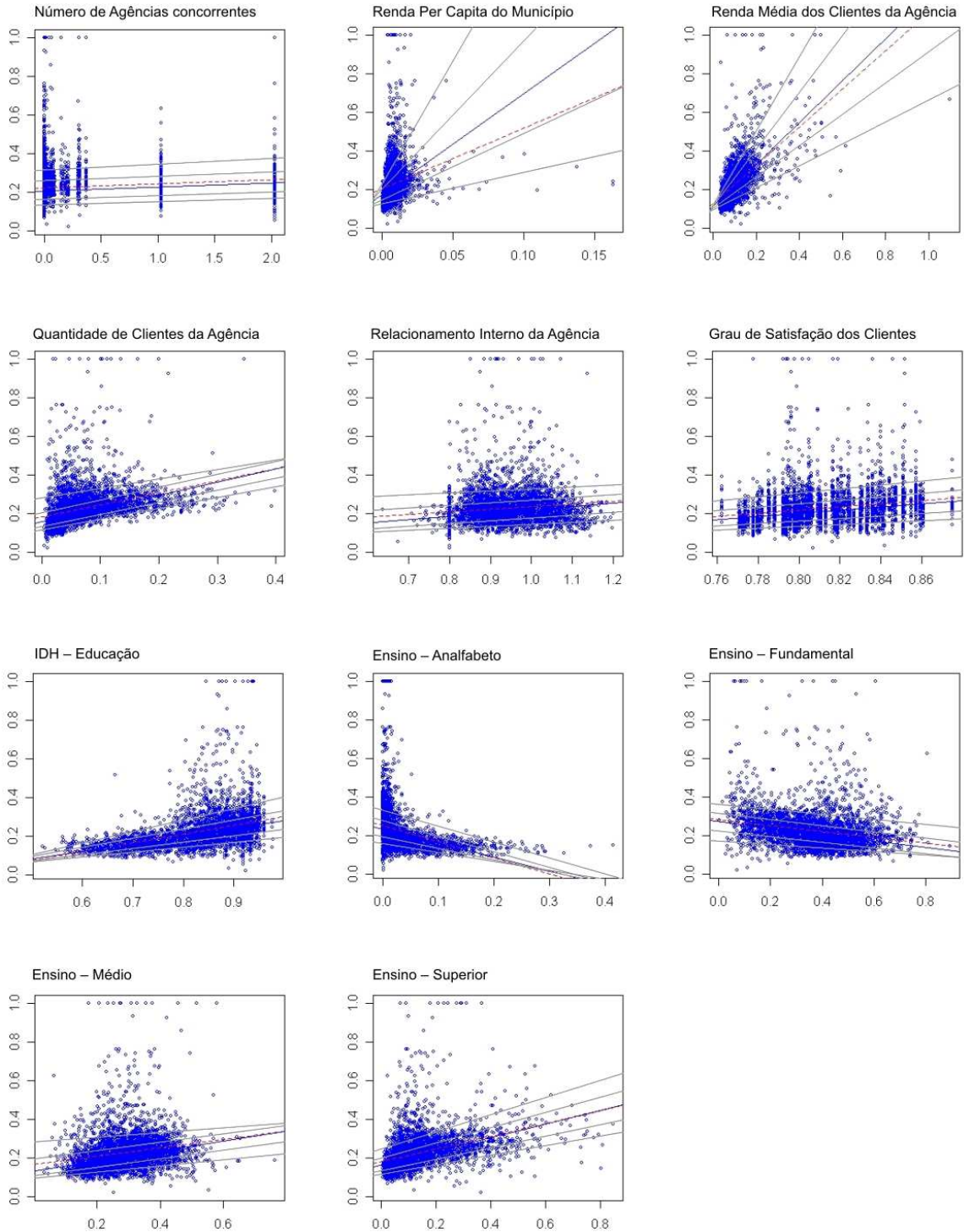
ANOVA

eficiencia

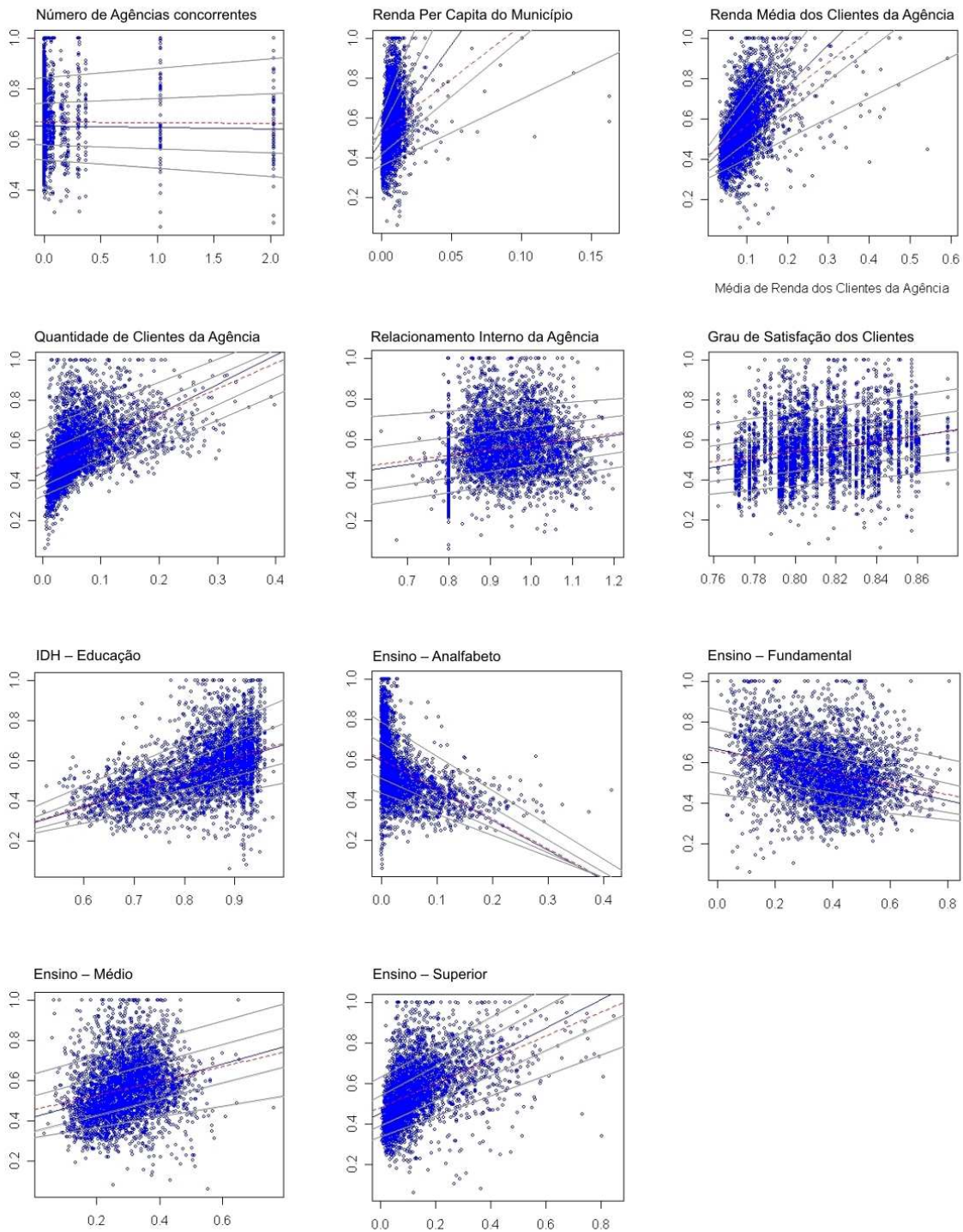
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.393	3	8.798	687.222	.000
Within Groups	46.803	3656	.013		
Total	73.196	3659			

APÊNDICE 4

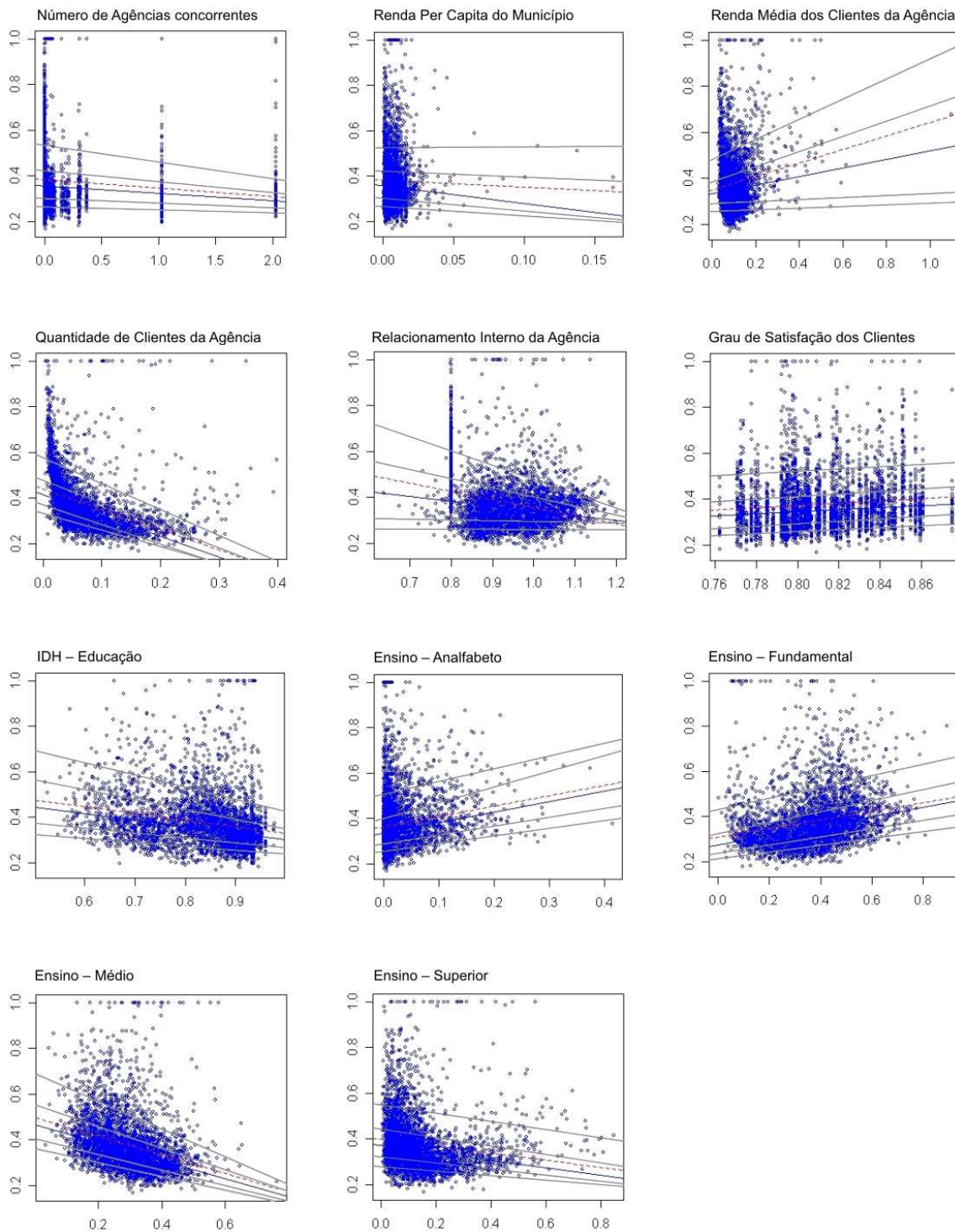
Eficiência CRS – PF com 4 outputs



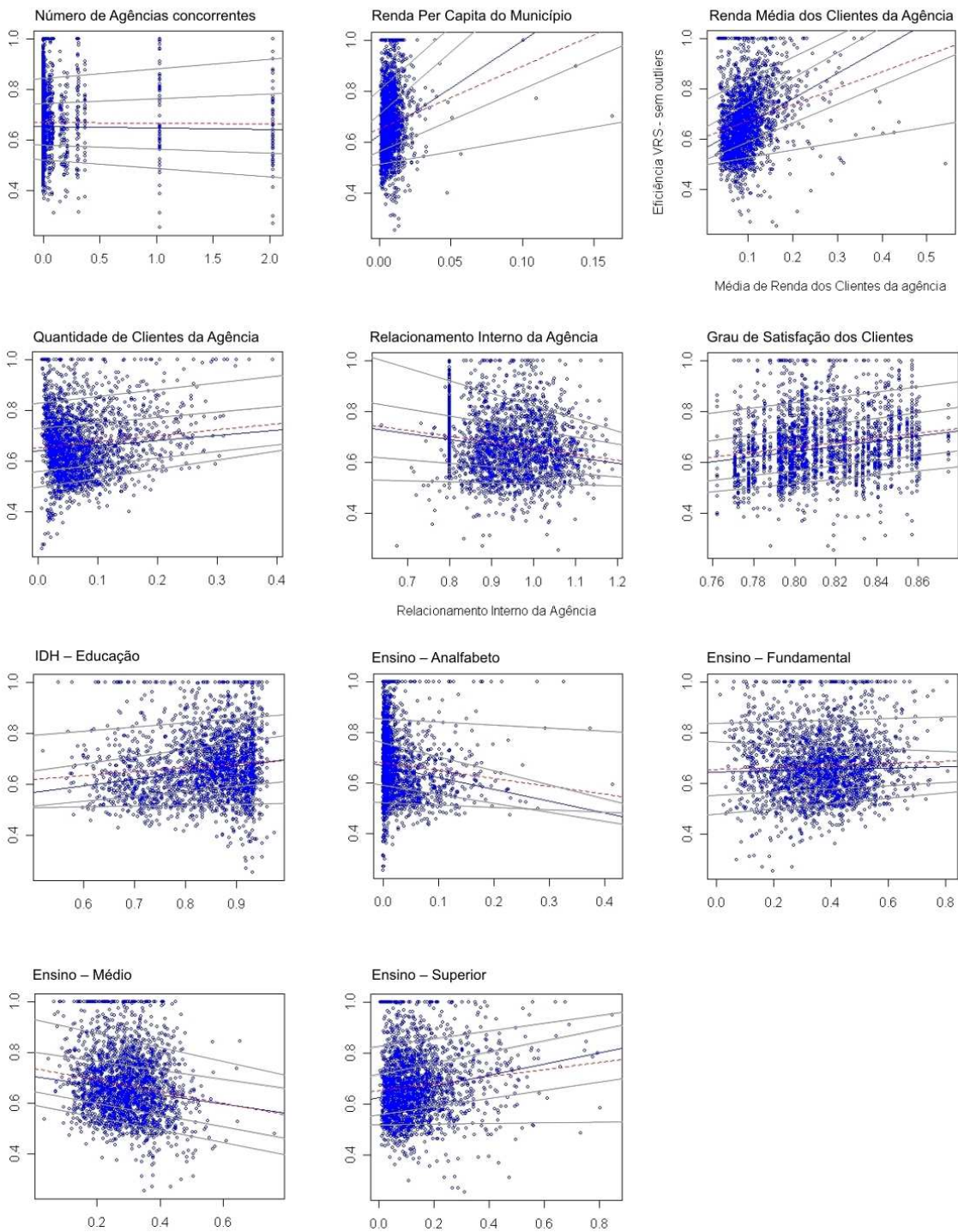
Eficiência CRS – sem outliers



Eficiência VRS – PF com 4 outputs



Eficiência VRS – sem outliers



APÊNDICE 5

Tabela Pessoa Física – CRS com 4 *outputs*

N = 3551

tau: 0.1

Coefficients	Value	Std. Error	T value	Pr(> t)
(Intercept)	-0.08866	0.04759	-1.86282	0.06257
se	0.00653	0.00391	1.66849	0.09531
sul	0.01045	0.00473	2.20859	0.02727
no	-0.00883	0.00565	-1.56084	0.11865
ne	0.00186	0.00485	0.38457	0.70058
rendapc	0.30842	0.05105	6.04128	0.00000
idheduc	0.03620	0.02180	1.66037	0.09693
agencia	-0.01562	0.01000	-1.56168	0.11845
quantcli	0.44388	0.02245	19.77259	0.00000
rendacl	0.35065	0.05308	6.60563	0.00000
clima_2004	0.03662	0.00996	3.67688	0.00024
grausat	0.13481	0.05297	2.54498	0.01097
cl1	0.00555	0.00269	2.06059	0.03942
cl3	-0.01410	0.00651	-2.16381	0.03055
cl4	0.02461	0.01998	1.23205	0.21801
analfabeto	-0.03394	0.02393	-1.41783	0.15633
fundamental	0.01929	0.01273	1.51458	0.12997
medio	-0.02173	0.01454	-1.49412	0.13523
superior	0.07947	0.01678	4.73656	0.00000

Tabela Pessoa Física – CRS com 4 outputs

N = 3551

tau: 0.25

Coefficients	Value	Std. Error	T value	Pr(> t)
(Intercept)	-0.06364	0.03533	-1.80152	0.07171
se	-0.00420	0.00585	-0.71701	0.47342
sul	0.00751	0.00622	1.20783	0.22719
no	-0.01929	0.00677	-2.85034	0.00439
ne	-0.01048	0.00628	-1.66782	0.09544
rendapc	0.44726	0.15624	2.86270	0.00423
idheduc	0.00701	0.01828	0.38368	0.70124
agencia	0.01621	0.01200	1.35024	0.17703
quantcli	0.43585	0.01651	26.40170	0.00000
rendacl	0.47759	0.03244	14.72317	0.00000
clima_2004	0.04511	0.00833	5.41318	0.00000
grausat	0.15782	0.03795	4.15814	0.00003
cl1	0.00456	0.00205	2.21829	0.02660
cl3	-0.03821	0.01149	-3.32476	0.00089
cl4	-0.03833	0.02396	-1.59975	0.10974
analfabeto	-0.05906	0.02340	-2.52392	0.01165
fundamental	0.00036	0.00935	0.03869	0.96914
medio	-0.01906	0.01154	-1.65147	0.09873
superior	0.05842	0.00913	6.39882	0.00000

tau: 0.5

Coefficients	Value	Std. Error	T value	Pr(> t)
(Intercept)	0.00328	0.03526	0.09298	0.92593
se	-0.01239	0.00510	-2.42916	0.01518
sul	0.00173	0.00533	0.32497	0.74522
no	-0.03158	0.00540	-5.84608	0.00000
ne	-0.02253	0.00558	-4.03928	0.00005
rendapc	0.87237	0.23881	3.65302	0.00026
idheduc	-0.01238	0.01521	-0.81422	0.41557
agencia	0.01395	0.01168	1.19412	0.23251
quantcli	0.45984	0.02077	22.14002	0.00000
rendacl	0.70487	0.04398	16.02662	0.00000
clima_2004	0.03840	0.00695	5.52191	0.00000
grausat	0.11474	0.03966	2.89301	0.00384
cl1	0.00223	0.00180	1.24201	0.21431
cl3	-0.04489	0.00955	-4.70304	0.00000
cl4	-0.03727	0.02383	-1.56437	0.11782
analfabeto	-0.04829	0.01625	-2.97118	0.00299
fundamental	0.00392	0.00834	0.46939	0.63882
medio	-0.01874	0.01054	-1.77878	0.07536
superior	0.03448	0.01138	3.02920	0.00247

Tabela Pessoa Física – CRS com 4 outputs

N = 3551

tau: 0.75

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.01631	0.04221	0.38647	0.69918
se	-0.03726	0.00904	-4.12097	0.00004
sul	-0.01774	0.00951	-1.86496	0.06227
no	-0.05775	0.00951	-6.06973	0.00000
ne	-0.03968	0.00928	-4.27699	0.00002
rendapc	1.14009	0.06322	18.03290	0.00000
idheduc	-0.01947	0.01885	-1.03287	0.30174
agencia	0.04160	0.01926	2.16023	0.03082
quantcli	0.46155	0.02250	20.51559	0.00000
rendacl	1.13692	0.06864	16.56332	0.00000
clima_2004	0.02869	0.00815	3.51904	0.00044
grausat	0.13748	0.04588	2.99662	0.00275
cl1	0.00075	0.00213	0.35311	0.72402
cl3	-0.07493	0.01919	-3.90456	0.00010
cl4	-0.09114	0.03869	-2.35573	0.01854
analfabeto	-0.08593	0.01887	-4.55401	0.00001
fundamental	0.01570	0.01137	1.38057	0.16750
medio	-0.03149	0.01270	-2.47899	0.01322
superior	-0.01490	0.01760	-0.84650	0.39733

tau: [1] 0.9

Coefficients	Value	Std. Error	T value	Pr(> t)
(Intercept)	0.07524	0.06240	1.20585	0.22796
se	-0.09687	0.01253	-7.73039	0.00000
sul	-0.06690	0.01341	-4.98752	0.00000
no	-0.11175	0.01399	-7.98834	0.00000
ne	-0.08959	0.01333	-6.72351	0.00000
rendapc	0.98488	0.26058	3.77961	0.00016
idheduc	0.03364	0.02904	1.15836	0.24679
agencia	0.00998	0.02740	0.36414	0.71578
quantcli	0.52224	0.04238	12.32393	0.00000
rendacl	1.49080	0.08233	18.10688	0.00000
clima_2004	0.02807	0.01133	2.47721	0.01329
grausat	0.08462	0.06633	1.27581	0.20211
cl1	0.00012	0.00343	0.03596	0.97131
cl3	-0.05517	0.02638	-2.09093	0.03661
cl4	-0.02576	0.05628	-0.45779	0.64713
analfabeto	-0.10195	0.02981	-3.41997	0.00063
fundamental	0.03066	0.01686	1.81882	0.06902
medio	-0.07064	0.02157	-3.27440	0.00107
superior	-0.02307	0.02526	-0.91356	0.36101

Tabela Pessoa Física retirados os outliers – CRS com 4 outputs

N = 3459

tau: 0.1

Coefficients	Value	Std. Error	T value	Pr(> t)
(Intercept)	0.01723	0.09986	0.17250	0.86305
se	0.02936	0.01155	2.54315	0.01103
sul	0.01426	0.01242	1.14778	0.25114
no	0.00177	0.01394	0.12718	0.89880
ne	-0.00403	0.01281	-0.31497	0.75280
rendapc	0.88356	0.46243	1.91070	0.05613
idheduc	-0.03827	0.04692	-0.81568	0.41474
agencia	-0.02358	0.06174	-0.38186	0.70259
quantcli	1.07118	0.03607	29.69616	0.00000
rendacl	0.43659	0.08637	5.05515	0.00000
clima_2004	0.10940	0.02220	4.92839	0.00000
grausat	0.24823	0.11044	2.24770	0.02466
cl1	0.01632	0.00495	3.29514	0.00099
cl3	-0.02075	0.04153	-0.49966	0.61735
cl4	0.01355	0.12334	0.10989	0.91250
analfabeto	-0.27338	0.05237	-5.22034	0.00000
Fundamental	0.05804	0.02608	2.22583	0.02609
medio	-0.10001	0.03389	-2.95088	0.00319
superior	0.19323	0.02495	7.74364	0.00000

tau: 0.25

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	-0.08932	0.08294	-1.07696	0.28158
se	0.02323	0.00894	2.59902	0.00939
sul	0.03348	0.01011	3.31144	0.00094
no	0.00265	0.01005	0.26374	0.79200
ne	0.00030	0.01012	0.02922	0.97669
rendapc	1.15771	0.40438	2.86290	0.00422
idheduc	-0.06180	0.03890	-1.58843	0.11228
agencia	-0.02599	0.02733	-0.95100	0.34167
quantcli	1.05152	0.03457	30.41842	0.00000
rendacl	0.69741	0.08351	8.35126	0.00000
clima_2004	0.10783	0.01781	6.05570	0.00000
grausat	0.43430	0.09154	4.74437	0.00000
cl1	0.01101	0.00432	2.55095	0.01079
cl3	-0.01869	0.02662	-0.70196	0.48275
cl4	0.03608	0.05482	0.65818	0.51047
analfabeto	-0.23651	0.03533	-6.69363	0.00000
Fundamental	0.03695	0.01934	1.91048	0.05615
medio	-0.08653	0.02714	-3.18820	0.00144
superior	0.24173	0.03499	6.90759	0.00000

Tabela Pessoa Física retirados os outliers – CRS com 4 outputs

N = 3459

tau: 0.5

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.10083	0.08552	1.17908	0.23845
se	0.01475	0.00915	1.61221	0.10701
sul	0.02832	0.00958	2.95556	0.00314
no	-0.02237	0.01078	-2.07430	0.03813
ne	-0.02609	0.01043	-2.50188	0.01240
rendapc	2.02744	0.49232	4.11813	0.00004
idheduc	-0.13755	0.04279	-3.21428	0.00132
agencia	0.02418	0.02444	0.98902	0.32272
quantcli	1.01909	0.04239	24.04135	0.00000
rendacl	1.12691	0.09119	12.35791	0.00000
clima_2004	0.09255	0.01912	4.84128	0.00000
grausat	0.31739	0.09273	3.42261	0.00063
cl1	0.00419	0.00470	0.89246	0.37221
cl3	-0.07539	0.02050	-3.67695	0.00024
cl4	-0.07402	0.04965	-1.49081	0.13610
analfabeto	-0.26191	0.05221	-5.01668	0.00000
fundamental	0.04941	0.02336	2.11517	0.03449
medio	-0.03991	0.02879	-1.38651	0.16568
superior	0.20557	0.02219	9.26608	0.00000

tau: 0.75

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.37321	0.10677	3.49554	0.00048
se	-0.01483	0.01528	-0.97059	0.33182
sul	0.00661	0.01660	0.39846	0.69031
no	-0.06280	0.01881	-3.33869	0.00085
ne	-0.05901	0.01667	-3.53945	0.00041
rendapc	3.31729	0.59940	5.53432	0.00000
idheduc	-0.21781	0.05636	-3.86479	0.00011
agencia	0.09545	0.03604	2.64832	0.00813
quantcli	0.93061	0.04547	20.46850	0.00000
rendacl	1.46014	0.10990	13.28591	0.00000
clima_2004	0.06159	0.02583	2.38481	0.01714
grausat	0.20241	0.11233	1.80189	0.07165
cl1	-0.00225	0.00576	-0.39065	0.69608
cl3	-0.12148	0.03226	-3.76524	0.00017
cl4	-0.19207	0.07267	-2.64300	0.00825
analfabeto	-0.39751	0.06312	-6.29749	0.00000
Fundamental	0.02949	0.02852	1.03413	0.30115
medio	-0.05373	0.03223	-1.66708	0.09559
superior	0.14534	0.03782	3.84319	0.00012

Tabela Pessoa Física retirados os outliers – CRS com 4 outputs
N = 3459

tau: 0.9

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.50507	0.13046	3.87149	0.00011
se	-0.03626	0.01051	-3.45074	0.00057
sul	-0.00276	0.01230	-0.22461	0.82230
no	-0.09460	0.01872	-5.05316	0.00000
ne	-0.09979	0.01142	-8.73869	0.00000
rendapc	2.84007	0.44967	6.31588	0.00000
idheduc	-0.20811	0.04802	-4.33339	0.00002
agencia	0.09788	0.03604	2.71585	0.00664
quantcli	0.79359	0.04820	16.46364	0.00000
rendacli	1.73941	0.14764	11.78132	0.00000
clima_2004	0.02265	0.03036	0.74592	0.45577
grausat	0.18474	0.15175	1.21743	0.22353
cl1	-0.00514	0.00558	-0.92025	0.35751
cl3	-0.08540	0.01557	-5.48395	0.00000
cl4	-0.20085	0.07191	-2.79329	0.00525
analfabeto	-0.38512	0.07004	-5.49863	0.00000
fundamental	0.05180	0.03234	1.60189	0.10927
medio	-0.09903	0.02930	-3.37975	0.00073
superior	0.12072	0.05372	2.24736	0.02468

Tabela Pessoa Física – VRS com 4 outputs

N = 3551

tau: 0.1

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.31521	0.03771	8.35866	0.00000
se	-0.00586	0.00271	-2.15846	0.03096
sul	0.00921	0.00200	4.59817	0.00000
no	-0.03039	0.00484	-6.27801	0.00000
ne	-0.01889	0.00358	-5.28372	0.00000
rendapc	0.03391	0.26057	0.13012	0.89648
idheduc	-0.12728	0.02578	-4.93637	0.00000
agencia	-0.00500	0.00790	-0.63300	0.52677
quantcli	-0.58513	0.00831	-70.3815	0.00000
rendacl	0.32866	0.04563	7.20244	0.00000
clima_2004	-0.05247	0.01012	-5.18473	0.00000
grausat	0.21952	0.03691	5.94719	0.00000
cl1	-0.01960	0.00277	-7.07417	0.00000
cl3	-0.02710	0.00729	-3.71748	0.00020
cl4	-0.01067	0.01574	-0.67794	0.49785
analfabeto	0.05119	0.05163	0.99153	0.32149
fundamental	0.03011	0.01100	2.73801	0.00621
medio	-0.11322	0.01127	-10.04557	0.00000
superior	-0.03354	0.01193	-2.81136	0.00496

tau: 0.25

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.49475	0.05197	9.51959	0.00000
se	-0.00982	0.00659	-1.49091	0.13607
sul	0.00715	0.00694	1.03047	0.30286
no	-0.03261	0.00816	-3.99637	0.00007
ne	-0.02459	0.00734	-3.34991	0.00082
rendapc	0.37019	0.22989	1.61030	0.10742
idheduc	-0.15226	0.02892	-5.26433	0.00000
agencia	0.00437	0.01343	0.32554	0.74479
quantcli	-0.58702	0.02291	-25.6225	0.00000
rendacl	0.38669	0.04225	9.15293	0.00000
clima_2004	-0.09611	0.01184	-8.11618	0.00000
grausat	0.10934	0.05403	2.02383	0.04306
cl1	-0.02174	0.00290	-7.48483	0.00000
cl3	-0.04407	0.01263	-3.49071	0.00049
cl4	-0.03582	0.02670	-1.34181	0.17974
analfabeto	0.09191	0.03893	2.36065	0.01830
fundamental	0.02319	0.01443	1.60746	0.10804
medio	-0.13667	0.01592	-8.58419	0.00000
superior	-0.02042	0.01737	-1.17551	0.23987

Tabela Pessoa Física – VRS com 4 outputs

N = 3551

tau: 0.5

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.61398	0.05858	10.48035	0.00000
se	-0.02140	0.00744	-2.87745	0.00403
sul	-0.00214	0.00778	-0.27558	0.78288
no	-0.03466	0.00889	-3.89985	0.00010
ne	-0.03655	0.00830	-4.40144	0.00001
rendapc	0.96214	0.15682	6.13550	0.00000
idheduc	-0.19914	0.03377	-5.89650	0.00000
agencia	0.02420	0.01565	1.54620	0.12215
quantcli	-0.62936	0.02366	-26.6022	0.00000
rendacl	0.57715	0.05701	10.1235	0.00000
clima_2004	-0.18484	0.01350	-13.6902	0.00000
grausat	0.17017	0.05976	2.84756	0.00443
cl1	-0.02560	0.00390	-6.56224	0.00000
cl3	-0.06804	0.01473	-4.61744	0.00000
cl4	-0.08818	0.03111	-2.83494	0.00461
analfabeto	0.11775	0.06094	1.93212	0.05342
fundamental	0.01671	0.01591	1.05025	0.29368
medio	-0.16232	0.01682	-9.64866	0.00000
superior	-0.07085	0.02026	-3.49647	0.00048

tau: 0.75

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.95083	0.09246	10.28393	0.00000
se	-0.03658	0.00937	-3.90395	0.00010
sul	-0.00830	0.01027	-0.80739	0.41949
no	-0.05789	0.01092	-5.30156	0.00000
ne	-0.04719	0.01117	-4.22474	0.00002
rendapc	0.81332	0.24708	3.29175	0.00101
idheduc	-0.30092	0.05055	-5.95298	0.00000
agencia	0.05548	0.02186	2.53719	0.01122
quantcli	-0.59383	0.04776	-12.4338	0.00000
rendacl	0.83659	0.08219	10.1787	0.00000
clima_2004	-0.32153	0.01967	-16.3499	0.00000
grausat	0.07562	0.09769	0.77409	0.43893
cl1	-0.02758	0.00651	-4.23811	0.00002
cl3	-0.09941	0.01946	-5.10796	0.00000
cl4	-0.15717	0.04330	-3.62975	0.00029
analfabeto	0.09315	0.08835	1.05432	0.29181
fundamental	0.01851	0.02369	0.78111	0.43479
medio	-0.19805	0.02549	-7.77014	0.00000
superior	-0.07775	0.02345	-3.31570	0.00092

Tabela Pessoa Física – VRS com 4 outputs

N = 3551

tau: 0.9

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.12852	0.16487	6.84482	0.00000
se	-0.06629	0.03198	-2.07254	0.03829
sul	-0.03602	0.03346	-1.07665	0.28171
no	-0.09323	0.03505	-2.65958	0.00786
ne	-0.05301	0.03363	-1.57643	0.11502
rendapc	1.33542	0.93417	1.42953	0.15294
idheduc	-0.21544	0.08056	-2.67416	0.00753
agencia	0.06906	0.06392	1.08051	0.27999
quantcli	-0.53612	0.07693	-6.96882	0.00000
rendacli	1.06500	0.08619	12.3567	0.00000
clima_2004	-0.55020	0.03413	-16.1204	0.00000
grausat	0.12522	0.17274	0.72492	0.46855
cl1	-0.03855	0.00758	-5.08810	0.00000
cl3	-0.12397	0.06516	-1.90255	0.05718
cl4	-0.18894	0.12926	-1.46173	0.14390
analfabeto	0.18198	0.04979	3.65473	0.00026
fundamental	0.02419	0.03771	0.64156	0.52120
medio	-0.23212	0.04813	-4.82260	0.00000
superior	-0.11626	0.04659	-2.49549	0.01262

Tabela Pessoa Física retirados os outliers – VRS com 4 outputs

N = 3400

tau: 0.1

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.42975	0.13521	3.17844	0.00150
se	0.00391	0.01367	0.28627	0.77470
sul	0.01404	0.01556	0.90272	0.36678
no	-0.01383	0.01690	-0.81865	0.41308
ne	-0.03435	0.01601	-2.14596	0.03199
rendapc	-0.08562	0.96309	-0.08890	0.92917
idheduc	-0.23002	0.07988	-2.87953	0.00402
agencia	-0.01757	0.04184	-0.41983	0.67465
quantcli	0.64963	0.05077	12.79464	0.00000
rendacl	0.42473	0.10413	4.07872	0.00005
clima_2004	-0.04637	0.03197	-1.45061	0.14704
grausat	0.40232	0.14166	2.84001	0.00455
cl1	-0.03299	0.00739	-4.46227	0.00001
cl3	-0.01986	0.03379	-0.58784	0.55670
cl4	-0.01478	0.08507	-0.17374	0.86209
analfabeto	-0.04586	0.09477	-0.48394	0.62848
fundamental	0.05609	0.03846	1.45852	0.14485
medio	-0.27709	0.04780	-5.79745	0.00000
superior	0.02293	0.03587	0.63904	0.52287

tau: 0.25

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.79686	0.09688	8.22495	0.00000
se	-0.00881	0.00950	-0.92703	0.35402
sul	0.01426	0.01158	1.23193	0.21811
no	-0.03243	0.01268	-2.55658	0.01064
ne	-0.04674	0.01132	-4.12921	0.00004
rendapc	1.41411	0.78432	1.80299	0.07153
idheduc	-0.26779	0.05429	-4.93254	0.00000
agencia	-0.00715	0.04291	-0.16672	0.86761
quantcli	0.54667	0.04111	13.29711	0.00000
rendacl	0.57456	0.07866	7.30460	0.00000
clima_2004	-0.18082	0.02119	-8.53505	0.00000
grausat	0.21498	0.10272	2.09289	0.03648
cl1	-0.03499	0.00587	-5.95710	0.00000
cl3	-0.03989	0.04088	-0.97572	0.32931
cl4	-0.02935	0.08603	-0.34121	0.73298
analfabeto	-0.07015	0.05797	-1.21004	0.22640
fundamental	0.03815	0.02481	1.53792	0.12422
medio	-0.33914	0.03287	-10.3184	0.00000
superior	0.14721	0.04929	2.98633	0.00286

Tabela Pessoa Física retirados os outliers – VRS com 4 outputs

N = 3400

tau: 0.5

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.93044	0.12348	7.53538	0.00000
se	-0.00441	0.00866	-0.50931	0.61059
sul	0.03496	0.01059	3.30226	0.00098
no	-0.02866	0.00954	-3.00262	0.00271
ne	-0.04690	0.01062	-4.41694	0.00001
rendapc	2.01966	0.67453	2.99418	0.00278
idheduc	-0.31185	0.06207	-5.02387	0.00000
agencia	-0.02005	0.04433	-0.45241	0.65102
quantcli	0.48462	0.04751	10.20133	0.00000
rendacl	0.81251	0.09965	8.15332	0.00000
clima_2004	-0.25174	0.02787	-9.03392	0.00000
grausat	0.21377	0.13639	1.56737	0.11718
cl1	-0.04214	0.00673	-6.26011	0.00000
cl3	-0.01197	0.03963	-0.30205	0.76264
cl4	-0.00061	0.08883	-0.00689	0.99451
analfabeto	-0.03171	0.05650	-0.56127	0.57468
fundamental	0.01671	0.02910	0.57442	0.56575
medio	-0.29766	0.03727	-7.98688	0.00000
superior	0.14686	0.03433	4.27765	0.00002

tau: 0.75

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.12018	0.14781	7.57870	0.00000
se	-0.00214	0.01917	-0.11143	0.91129
sul	0.02861	0.01995	1.43410	0.15169
no	-0.06283	0.02152	-2.91972	0.00354
ne	-0.05044	0.02054	-2.45552	0.01415
rendapc	2.57418	0.33724	7.63297	0.00000
idheduc	-0.31904	0.08147	-3.91611	0.00009
agencia	0.07982	0.05540	1.44076	0.14980
quantcli	0.46294	0.06117	7.56820	0.00000
rendacl	1.08356	0.09484	11.4252	0.00000
clima_2004	-0.37184	0.03374	-11.0195	0.00000
grausat	0.18459	0.15761	1.17119	0.24166
cl1	-0.03975	0.00788	-5.04574	0.00000
cl3	-0.07709	0.05236	-1.47223	0.14111
cl4	-0.20749	0.10963	-1.89265	0.05854
analfabeto	-0.03712	0.09645	-0.38480	0.70043
fundamental	0.03294	0.04256	0.77394	0.43906
medio	-0.28081	0.04541	-6.18333	0.00000
superior	0.07950	0.03600	2.20826	0.02733

Tabela Pessoa Física retirados os outliers – VRS com 4 outputs

N = 3400

tau: 0.9

Coefficients	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.51901	0.16922	8.97676	0.00000
se	-0.02978	0.02401	-1.24020	0.21504
sul	0.00739	0.02503	0.29513	0.76792
no	-0.12549	0.02512	-4.99483	0.00000
ne	-0.07341	0.02889	-2.54116	0.01112
rendapc	2.19599	1.14852	1.91202	0.05601
idheduc	-0.33909	0.12388	-2.73716	0.00625
agencia	0.13533	0.06522	2.07489	0.03812
quantcli	0.59205	0.08963	6.60542	0.00000
rendacli	0.94971	0.10678	8.89364	0.00000
clima_2004	-0.56068	0.03948	-14.2026	0.00000
grausat	0.04590	0.14696	0.31232	0.75483
cl1	-0.04333	0.01445	-2.99915	0.00274
cl3	-0.05979	0.04348	-1.37526	0.16920
cl4	-0.27525	0.13616	-2.02152	0.04335
analfabeto	-0.00570	0.36227	-0.01573	0.98745
fundamental	0.09947	0.03858	2.57825	0.01000
medio	-0.27118	0.03566	-7.60476	0.00000
superior	0.05061	0.04722	1.07194	0.28387

APÊNDICE 6

Modelo Eficiência Pessoa Física CRS – 4 outputs com outliers

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,677(a)	,458	,455	,074079497

a Predictors: (Constant), superior, sul, clima_2004, no, cl4, cl1, rendapc, quantcli, cl3, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16,384	18	,910	165,864	,000 ^a
	Residual	19,372	3530	,005		
	Total	35,756	3548			

a. Predictors: (Constant), superior, sul, clima_2004, no, cl4, cl1, rendapc, quantcli, cl3, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

b. Dependent Variable: eficfiscrs4

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,001	,058		-,019	,985
	se	-,043	,005	-,205	-8,350	,000
	sul	-,022	,006	-,090	-3,979	,000
	no	-,063	,007	-,150	-8,790	,000
	ne	-,043	,006	-,194	-7,045	,000
	rendapc	,819	,181	,064	4,520	,000
	idheduc	,017	,032	,016	,540	,589
	agencia	,025	,014	,109	1,723	,085
	quantcli	,492	,028	,267	17,617	,000
	rendacli	,838	,028	,492	29,996	,000
	clima_2004	,047	,014	,043	3,326	,001
	grausat	,103	,062	,025	1,656	,098
	cl1	,005	,004	,019	1,369	,171
	cl3	-,060	,012	-,133	-5,076	,000
	cl4	-,051	,028	-,104	-1,800	,072
	analfabeto	-,097	,045	-,043	-2,185	,029
	fundamental	,021	,016	,030	1,328	,184
	medio	-,025	,019	-,023	-1,327	,185
	superior	,008	,019	,009	,403	,687

a. Dependent Variable: eficfiscrs4

Modelo Eficiência Pessoa Física – CRS 4 outputs retirados os outliers

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,682(a)	,465	,463	,111088455

a Predictors: (Constant), superior, sul, cl4, no, clima_2004, cl1, rendapc, cl3, quantcli, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	36,966	18	2,054	166,414	,000 ^a
	Residual	42,452	3440	,012		
	Total	79,418	3458			

a. Predictors: (Constant), superior, sul, cl4, no, clima_2004, cl1, rendapc, cl3, quantcli, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

b. Dependent Variable: efch4

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,225	,088		2,553	,011
	se	-,001	,008	-,005	-,183	,855
	sul	,011	,009	,030	1,318	,187
	no	-,042	,011	-,068	-3,869	,000
	ne	-,046	,010	-,137	-4,830	,000
	rendapc	1,806	,274	,093	6,583	,000
	idheduc	-,134	,048	-,082	-2,803	,005
	agencia	,036	,023	,104	1,588	,112
	quantcli	1,065	,043	,382	24,928	,000
	rendacli	,972	,054	,307	17,960	,000
	clima_2004	,087	,021	,053	4,056	,000
	grausat	,221	,095	,036	2,326	,020
	cl1	,002	,006	,005	,328	,743
	cl3	-,065	,019	-,091	-3,432	,001
	cl4	-,086	,045	-,114	-1,908	,057
	analfabeto	-,344	,067	-,102	-5,133	,000
	fundamental	,071	,024	,068	2,991	,003
	medio	-,077	,029	-,048	-2,676	,007
	superior	,188	,030	,142	6,200	,000

a. Dependent Variable: efch4

Modelo Eficiência Pessoa Física – VRS 4 outputs com outliers

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,577(a)	,332	,329	,102119342

a Predictors: (Constant), superior, sul, clima_2004, no , cl4, cl1, rendapc, quantcli, cl3, medio, grausat, se, rendacl, analfabeto, fundamental, ne, idheduc, agencia

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18,330	18	1,018	97,652	,000 ^a
	Residual	36,812	3530	,010		
	Total	55,142	3548			

a. Predictors: (Constant), superior, sul, clima_2004, no , cl4, cl1, rendapc, quantcli, cl3, medio, grausat, se, rendacl, analfabeto, fundamental, ne, idheduc, agencia

b. Dependent Variable: eficisvrs4

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,799	,080		9,986	,000
	se	-,041	,007	-,158	-5,807	,000
	sul	-,011	,008	-,036	-1,434	,152
	no	-,066	,010	-,126	-6,651	,000
	ne	-,042	,008	-,151	-4,934	,000
	rendapc	,755	,250	,047	3,024	,003
	idheduc	-,228	,044	-,170	-5,243	,000
	agencia	,038	,020	,137	1,949	,051
	quantcli	-,528	,038	-,231	-13,713	,000
	rendacl	,625	,039	,295	16,228	,000
	clima_2004	-,331	,019	-,244	-17,005	,000
	grausat	,173	,086	,034	2,004	,045
	cl1	-,027	,005	-,080	-5,179	,000
	cl3	-,075	,016	-,135	-4,613	,000
	cl4	-,108	,039	-,177	-2,766	,006
	analfabeto	,083	,061	,030	1,349	,177
	fundamental	,021	,022	,024	,960	,337
	medio	-,177	,026	-,132	-6,715	,000
	superior	-,071	,027	-,067	-2,626	,009

a. Dependent Variable: eficisvrs4

Modelo Eficiência Pessoa Física – VRS 4 outputs sem outliers

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,469(a)	,220	,216	,112802671

a Predictors: (Constant), superior, sul, cl4, no , clima_2004, cl1, rendapc, cl3, quantcli, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12,132	18	,674	52,968	,000 ^a
	Residual	42,996	3379	,013		
	Total	55,128	3397			

a. Predictors: (Constant), superior, sul, cl4, no , clima_2004, cl1, rendapc, cl3, quantcli, medio, grausat, se, rendacli, analfabeto, fundamental, ne, idheduc, agencia

b. Dependent Variable: eficiencia

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,910	,090		10,105	,000
	se	,007	,008	,028	,913	,361
	sul	,034	,009	,106	3,741	,000
	no	-,046	,011	-,087	-4,068	,000
	ne	-,041	,010	-,146	-4,205	,000
	rendapc	1,512	,282	,092	5,356	,000
	idheduc	-,336	,049	-,246	-6,863	,000
	agencia	,014	,023	,050	,627	,531
	quantcli	,543	,045	,227	12,164	,000
	rendacli	,755	,060	,266	12,663	,000
	clima_2004	-,281	,022	-,202	-12,788	,000
	grausat	,257	,097	,050	2,647	,008
	cl1	-,028	,006	-,082	-4,797	,000
	cl3	-,060	,019	-,100	-3,098	,002
	cl4	-,086	,046	-,136	-1,869	,062
	analfabeto	,006	,069	,002	,089	,929
	fundamental	,061	,024	,069	2,497	,013
	medio	-,215	,030	-,156	-7,196	,000
superior	,135	,031	,119	4,325	,000	

a. Dependent Variable: eficiencia

APÊNDICE 7

Modelo DEA

Para ilustração de um exemplo, suponha calcular a eficiência da agência número 3 para um modelo com 5 *outputs* e 2 *inputs*. Isto é:

Minimizar $\theta \lambda$ θ_3

Sujeito as seguintes restrições:

OUTPUT

$$-y_{31} + (y_{11} \lambda_1 + y_{21} \lambda_2 + y_{31} \lambda_3 + \dots + y_{3663\ 1} \lambda_{3663}) \geq 0$$

$$-y_{32} + (y_{12} \lambda_1 + y_{22} \lambda_2 + y_{32} \lambda_3 + \dots + y_{3663\ 2} \lambda_{3663}) \geq 0$$

$$-y_{33} + (y_{13} \lambda_1 + y_{23} \lambda_2 + y_{33} \lambda_3 + \dots + y_{3663\ 3} \lambda_{3663}) \geq 0$$

$$-y_{34} + (y_{14} \lambda_1 + y_{24} \lambda_2 + y_{34} \lambda_3 + \dots + y_{3663\ 4} \lambda_{3663}) \geq 0$$

$$-y_{35} + (y_{15} \lambda_1 + y_{25} \lambda_2 + y_{35} \lambda_3 + \dots + y_{3663\ 5} \lambda_{3663}) \geq 0$$

Input

$$\theta_3 x_{31} + (x_{11} \lambda_1 + x_{21} \lambda_2 + x_{31} \lambda_3 + \dots + x_{3663\ 1} \lambda_{3663}) \geq 0$$

$$\theta_3 x_{32} + (x_{12} \lambda_1 + x_{22} \lambda_2 + x_{32} \lambda_3 + \dots + x_{3663\ 2} \lambda_{3663}) \geq 0$$

onde $\lambda = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_{3663})' \geq 0$

para o modelo VRS existe outra restrição: $\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_{3663} \leq 1$

Para o cálculo do *Leverage* suponha que seja retirada a primeira unidade, portanto $j = 1$, então :

Minimizar $\theta^* \lambda$ θ^*_{31}

Sujeito as restrições

OUTPUT

$$-y_{31} + (y_{21} \lambda_2 + y_{31} \lambda_3 + \dots + y_{3663\ 1} \lambda_{3663}) \geq 0$$

$$-y_{32} + (y_{22} \lambda_2 + y_{32} \lambda_3 + \dots + y_{3663\ 2} \lambda_{3663}) \geq 0$$

$$-y_{33} + (y_{23} \lambda_2 + y_{33} \lambda_3 + \dots + y_{3663\ 3} \lambda_{3663}) \geq 0$$

$$-y_{34} + (y_{24} \lambda_2 + y_{34} \lambda_3 + \dots + y_{3663\ 4} \lambda_{3663}) \geq 0$$

$$-y_{35} + (y_{25} \lambda_2 + y_{35} \lambda_3 + \dots + y_{3663\ 5} \lambda_{3663}) \geq 0$$

Input

$$\theta^*_{31} x_{31} + (x_{21} \lambda_2 + x_{31} \lambda_3 + \dots + x_{3663\ 1} \lambda_{3663}) \geq 0$$

$$\theta^*_{31} x_{32} + (x_{22} \lambda_2 + x_{32} \lambda_3 + \dots + x_{3663\ 2} \lambda_{3663}) \geq 0$$

onde $\lambda = (\lambda_2, \lambda_3, \dots, \lambda_{3663})' \geq 0$

para o modelo VRS existe outra restrição: $\lambda_2 + \lambda_3 + \dots + \lambda_{3663} \leq 1$

Programa de Parametrização para o cálculo do DEA e do *LEVERAGE*

Base_io_01.txt DATA FILE NAME
Baseiocrs.out OUTPUT FILE NAME
iomeancrs.out OUTPUT FILE NAME FOR MEAN VALUES
iocrs_leverage.dat LEVERAGE FILE NAME
3663 NUMBER OF DMU's
1 NUMBER OF TIME PERIODS
5 NUMBER OF OUTPUTS
2 NUMBER OF INPUTS
0 0=INPUT, 1=OUTPUT ORIENTATED
0 0=CRS, 1=VRS, 2=NRS
0 0=DEA(MULTI-STAGE), 1=COST-DEA, 2=MALMQUIST-DEA,
3=DEA(1-STAGE), 4=DEA(2-STAGE)
1000 NUMBER OF BOOTSTRAP (or JACKSTRAPP) PASSES
370 NUMBER OF DMU'S TO BOOTSTRAP (USED FOR TESTING,
DEFAULT=#DMU's, for JACKSTRAPP, bubble size)
1 SPAWN: 0=INTERNAL ALGORITHM, 1=SPAWN DEA.EXE 2=FDH
3=detalhado
0.00 BIN SIZE: 0=WRITE ALL DATA (DEFAULT=0.01)
0 METHOD: 0=BOOTSTRAP, 1=MONTE CARLO (DEFAULT=0)
1.0 COMPRESSION ORIGINAL DISPERSION (FOR MONTE CARLO
ONLY, DEFAULT=1.00)
10 # of PASSES BETWEEN PRINT INFO TO SCREEN (DEFAULT 100)
3.0 LEVEL OF SIGNIFICANCY (0-100%)
0 JACKSTRAP SCHEME: 0=UNIFORM; 1=LINEAR; 2=INVERSE;
3=EXP; 4=STEP (DEFAULT=0); 5=SEQUENCE

REFERÊNCIAS BIBLIOGRÁFICAS

1. R 2.2.1 . The R Foundation Statistical Computing version 2.2.1. 20/12/2005 – Software livre. Site <http://cran.r-project.org>.
2. SPSS / versão 13.0 for Windows license UnB/2004.
3. BANKER, R.D. (1993). *Maximum likelihood, consistency and DEA: a statistical foundation*. Management Science, vol 39, no.10, pp. 1265-1273.
4. BANKER, R.D., CHARNES A., COOPER W. W., (1984). *Some models for estimating technical and scale inefficiencies in data envelopment analysis*. Management Science, vol.30, no.9, pp. 1370-1382.
5. BARBA-ROMERO, S., POMEROL, J.C., (1997). *Decisiones Multicriterio: Fundamentos Teóricos e Utilización Práctica*, Colección de Economía, Universidad de Alcalá.
6. BELTON, V., VICKERS, S.P., (1993) *Demistifying DEA - A Visual Interactive Approach based on Multiple Criteria Analysis*, Journal of Operational Research Society, v.44, n. 9, pp. 883-896.
7. BERGER, Allen N.; HUMPHREY, David B. (1991). *The dominance of inefficiencies over scale and product mix economies in banking*. Journal of Productivity Analysis, v. 07, n. 2/3, pp 139-160.
8. BERGER, Allen N.; HUMPHREY, David B. (1997). *Efficiency of financial institutions: internacional survey and directions for future research*. European Journal of Operational Research, v. 98, n. 2, p. 175-212.
9. BIRKIN, M., CLARKE, G., CLARKE, M., Wilson, A, (1996) *Intelligent GIS: Location Decisions and Strategic Planning*, GeoInformation International, London.
10. BUCHINSKY, M. (1997). *Recent advances in quantile regression: A practical guideline for empirical research*. Brown University and NBER.
11. CAZALS, C., FLORENS, J.P., and SIMAR L. (2002). *Nonparametric frontier estimation: A robust approach*. Journal of Econometrics 106, 1-25.
12. CHARNES, A., COOPER W.W., GOLANY, B., SEIFORD, L.M., STUTZ, J. (1985). *Foundation of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions*. Journal of Econometrics, vol.30, no. 1/2 , pp. 91-107.
13. CHARNES, A., COOPER W.W., LEWIN, A Y. & SEIFORD, L.M. (1996). *Data Envelopment Analysis*, 1996, KAP, London.

14. CHARNES A., COOPER W. W., RHODES, E., (1978). *Measuring the Efficiency of Decision-making Units*, European Journal of Operational Research, v.2, pp. 429-444.
15. CHERCHYE, L., KUOSMANEN, T. and POST, G.T. (2000). "New tools for Dealing with Errors in Variables in DEA", CES Discussion paper 00.06.
16. CRIBARI, F and ZARKOS, S. (2004). *Leverage adjusted heteroskedastic bootstrap methods*. Journal of Statistical Computation and Simulation, v. 74, pp. 215-232.
17. COOK, W.D., KRESS, M., SEIFORD, L. M., (1996). *Priorization Models for Frontier Decision making Units in DEA*, European Journal of Operational Research, v. 59, pp.319-323.
18. DEBREU, G. (1951). *The Measurement of Productive Efficiency*, Econometrica, vol.19, no. 3, pp. 273-292.
19. DEPRINS, D., L. SIMAR, and H. TULKENS. (1984). "Measuring Labor Efficiency in Post Offices." In M. Marchand, P. Pestieau and H. Tulkens (eds) *The Performance of Public Enterprises: Concepts and Measurements*. Elsevier, 345-367.
20. ECO, Umberto (1996). *Como se faz uma tese*. 14^a. Ed., São Paulo, Perspectiva.
21. ESTELLITA LINS, M., MEZA, L. A, (2000). *Análise Envoltória de Dados e perspectivas de integração no ambiente de Apoio à Decisão*. Editora da COPPE / UFRJ, Rio de Janeiro.
22. ESTELLITA LINS, M., MOREIRA, M. C. B. (1999). *Método I-O Stepwise para seleção de variáveis em Modelos de Análise Envoltória de Dados*. Pesquisa Operacional, vol.19 no.1, pp. 39-50.
23. FARREL, M.J., (1957). *The measurement of productive efficiency*, Journal of the Royal Statistical Society, Series A, General120, Part 3, pp. 253-281.
24. FÄRE , R., GROSSKOPF, S. e LOVEL C.A.K. (1994). *Production Frontiers*. Cambridge University Press, New York.
25. FREIXAS, Xavier; ROCHET, Jean-Charles. *Microeconomics of banking*. Massachusetts Institute of Technology, 1999.
26. GOLANY, B., STORBECK, J. E., (1999). *A Data Envelopment Analysis of the Operational Efficiency of Bank Branches*, INTERFACES, v.29, pp.14-26.
27. HAIR, J.F., ANDERSON, R.E., TATHAM, R.L., BLACK,W.C., (2005). *Análise Multivariada de Dados*, 2005, 5^a. edição. Bookman – SP.

28. JANKOWSKI, P., (1995). Integration geographical information systems and multicriteria decision decision-making methods, *International Journal of Geographical Informations Systems*, v.9, n.3, pp. 251-273
29. KANTOR, J., MAITAL, S., (1999). Measuring Efficiency by Product Group: Integrating DEA with Activity-Based Accounting in a Large Mideast Bank, *INTERFACES* v.29, pp. 27-36.
30. KAPLAN, Robert & NORTON, David, (1997). *Balanced scorecard: a estratégia em ação*. 1997, São Paulo, Campus.
31. KOENKER, R., & BASSET, G. (1978). *Regression Quantiles*. *Econometrica*, 46, pp. 33-50.
32. KRAUSE, Kathleen; PORTELLA, Gualter R.; TABAK, Benjamin M. (2003). In: Seminário de Eficiência Bancária – Banco Central, Rio de Janeiro. *Eficiência Bancária: O Valor Intrínseco na Função de Produção*. Universidade de Brasília e Banco Central do Brasil.
33. KUOSMANEN, T. and POST, G.T. (1999). *Robust Efficiency Measurement*. Rotterdam Institute for Business Economic Studies (RIBES). Report 9911. Rotterdam.
34. LAKATOS, Eva Maria, (1992). *Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos*, 1992, 4^a. Ed., São Paulo, Atlas.
35. MAGUIRE, D.J., (1991). *An overview and definition of GIS*. In: Maguire, D.J., Goodchild, M.F., Rhind, D.W., *Geographical Information Systems : Principles and Applications*, v.1, Longman, pp. 9-20.
36. MUSGRAVE, R.A, MUSGRAVE, P.B., (1980). *Finanças Públicas - Teoria e Prática*. 1980, Editora Campus - USP São Paulo.
37. NUNES DA SILVA, E., PORTO JÚNIOR, S. da S., (2005). *Sistema Financeiro e Crescimento Econômico: Uma aplicação de regressão quantílica*. UFRGS.
38. ROSA, Fernando, MAZZON, J.A., (2003). *Análise de eficiência operacional, mercadológica e de resultados baseada no conceito de cadeia de serviços-lucro*. In: ENANPAD, 27^o. Atibaia. Rio de Janeiro.
39. SALVADOR, Ângelo Domingos, (1980). *Métodos e técnicas de pesquisa bibliográfica: elaboração de trabalhos científicos*. Porto Alegre, Sulina.
40. SAMUELSON, P.A and NORDHAUS, W.D. (2004) *Economia*. McGraw Hill1. 2004, 17^a. edição.

41. SEAVER, B. and K. TRIANTIS (1992). "A Fuzzy Clustering Approach Used in Evaluating Technical Efficiency Measures in Manufacturing." *Journal of Productivity Analysis* 3, 337-363.
42. SEAVER, B., and K. TRIANTIS (1995). "The Impact of Outliers and Leverage Points for Technical Efficiency Measurement Using High Breakdown Procedures." *Management Science* 41, 937-956.
43. SEIFORD, L. M., THRALL, R.M., (1990). *Recent Developments in DEA*. *Journal of Econometrics*, vol.46, pp. 7-38.
44. SIEGEL, Sidney, (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.
45. SILVA, Antonio C.M., AZEVEDO, Gustavo H.W. (2004). *Eficiência e Sobrevivência: binômio fundamental para a previdência privada aberta*. *Revista Brasileira de Risco e Seguro* v.1 n. 0.
46. SIMAR, L. (2003). *Detecting outliers in frontier models: A simple approach*, *Journal of Productivity Analysis*, forthcoming.
47. SIMAR, L. and WILSON, P. (2007). *Estimation and inference in two-stage, semi-parametric models of production models*. Forthcoming on "International Transactions of Operations Research".
48. SOUSA, M.C.S and STOSIC, B. D. (2005). *Technical Efficiency of the Brazilian Municipalities: Correcting Nonparametric Frontier Measurements for Outliers*. *Journal of Productivity Analysis*. Springer-Netherlands, v. 24, p. 155-179.
49. SOUSA, M.C.S; CRIBARI, F.; STOSIC, B. D. (2005). *Explaining DEA Technical Efficiency Scores in an Outlier Corrected Environment: The Case of Public Services in Brazilian Municipalities*. *Review of Econometrics*. Forthcoming.
50. SOUZA, G. S. (2001). *Statistical properties of data envelopment analysis estimators of production functions*. *Brazilian Journal of Econometrics*, 21(2) 291-322.
51. SOUZA, G. S. (2002). *Funções de Produção: Uma Abordagem Estatística com o Uso de Modelos de Encapsulamento de Dados*. EMBRAPA.
52. SOUZA, G. S. (2005). *Significância de efeitos técnicos na eficiência de produção da pesquisa agropecuária brasileira*. Forthcoming on 'Revista Brasileira de Economia', FGV, Rio.
53. SOUZA, G.S., ALVES, E., ÁVILA, A.F.D. and CRUZ, E.R., (1997). *Produtividade e eficiência relativa de produção em sistemas de produção de pesquisa agropecuária*, *Revista Brasileira de Economia*, 51 (3), 281-307.

54. SOUZA, Geraldo S.; TABAK, Benjamin M.; STAUB, Roberta B (2003). *Modelos tobit aplicados à eficiência bancária no Brasil*. In: SEMINÁRIO DE EFICIÊNCIA, PRODUTIVIDADE E ESTABILIDADE BANCÁRIA, Rio de Janeiro, Banco Central do Brasil.
55. SOUZA G. S. and STAUB R. B. (2006). *Two stage inference using dea efficiency measurements in univariate production models*. Forthcoming on 'International Transactions of Operations Research'.
56. SOUZA, Geraldo S.; TABAK, Benjamin M., STAUB R.B. (2006). *Assessing the significance of factors effects in output oriented DEA measures of efficiency: An application to Brazilian banks*. Revista Brasileira de Economia de Empresas. Vol 6(1) pp. 7-20.
57. THANASSOULIS, E., (1999). *Data Envelopment Analysis and Its Use in Banking*, Interfaces v.29, pp.1-13.
58. WHEELOCK, David C.; WILSON, Paul W. (2003). *Robust Nonparametric Estimation of Efficiency and Technical Change in U.S. Commercial Banking*. Working Paper Series. Fedral Reserve Bank of St. Louis. November.
59. WILSON, P. (1993). "Detecting Influential Observations in Data Envelopment Analysis." Journal of Productivity Analysis 6, 27-45.
60. WILSON, P. (1995). "Detecting Influential Observations in Deterministic Non-Parametric Frontiers models." Journal of Business and Economic Statistics 11, 319-323.
61. YEH, Q. J. (1996). *The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation*. The Journal of the Operational Research Society, v. 47, n. 8, p. 980-988.
62. ZENIOS, C.V., ZENIOS, S.A, AGATHOCLEOUS, K., SOTERIOU, A C., (1999). *Benchmarks of the Efficiency of Bank Branches*, Interfaces v.29, pp. 37-51.