

Análise de Eficiência e Fronteira de Produção para companhias aéreas brasileiras

Gabriel Lima Novais

1 Introdução

O trabalho em questão procurou estabelecer a análise de eficiência de companhias aéreas com atuação em nível nacional. Para isso foi necessário estabelecer a coleta e o tratamento de dados relevantes, para posterior aplicação dos métodos selecionados, de maneira a obter os resultados comparativos consequentes.

Devido a disponibilidade dos dados e a grande variedade de empresas com períodos de vigência diferentes, em que houve processos específicos de fusões e falências, o estudo focou em 6 empresas aéreas no período de 2010 até 2014. Diversos determinantes econômicos e políticos ocorreram no Brasil, e por este motivo procurou-se contextualizar os fatos adversos para melhor entendimento dos resultados obtidos pelos métodos econométricos.

Os métodos em questão foram o Data Envelopment Analysis (DEA), o Deterministic Frontier Analysis (DFA) e o Stochastic Frontier Analysis (SFA). Todos os métodos serão explicados de maneira mais detalhada à frente. Através dos métodos procurou-se observar a evolução da eficiência de cada uma das empresas selecionadas nesse período de tempo, por meio do estudo de cross section para cada ano do período citado.

2 Dados

Para a realização do estudo foram utilizadas duas bases de dados disponíveis na mesma fonte, obtidos no site da Agência Nacional de Aviação Civil (ANAC). Após coleta, ambas as bases foram agrupadas em apenas uma utilizando-se para isso o procedimento de "merge" disponível no pacote estatístico do Stata. As duas bases de dados utilizadas referem-se a aspectos inseridos no contexto do mercado de transporte aéreo. As suas descrições são feitas a seguir.

2.1 Dados Estatísticos

Os dados foram retirados do site da Agência Nacional de Aviação Civil (ANAC). Os dados brutos compreendem o período de Janeiro de 2000 até Março de 2018. Eles estão dispostos de maneira a informar para cada empresa, em cada ano, mês e voo, os respectivos valores de variáveis inerentes ao percurso determinado. Além disso, os dados são relativos tanto a etapa básica de voo, que representam os dados estatísticos de cada aeronave em cada etapa do voo entre dois aeródromos, assim como os demais dados disponíveis na base de dados estatísticos. Desta maneira, nesse estudo, não se procurou focar em um tipo exclusivo de voo (como os voos regulares por exemplo). O que procurou-se fazer foi apenas considerar as empresas brasileiras, que efetuaram no período disponível, serviços associados ao transporte aéreo. Não houve discriminação de voos domésticos ou não.

Inicialmente a base de dados possuía um número de 38 variáveis constituindo 832.443 observações. Após a realização dos filtros descritos no parágrafo anterior a base diminuiu para 21

variáveis e 507.875 observações. Foram feitos também renomeações para facilitar a programação de tais variáveis no pacote estatístico do Stata. Após a realização dos filtros e de correções, procurou-se reduzir a periodicidade da base para os anos de 2010 até 2014 inclusive, pois como será visto na subseção à frente, este era o período que fora escolhido para o estudo das 6 empresas selecionadas através dos dados coletados dos Demonstrativos de Relatório Operacional. Assim o resultado obtido foi de 21 variáveis e 145.381 observações.

Após essa redução, procurou-se agregar todos os dados em periodicidade anual, de forma que foram obtidas 18 variáveis e 97 observações. Além disso, vale ressaltar que determinadas variáveis foram sendo descartadas por não cumprirem função alguma no estudo em questão. Finalmente selecionaram-se apenas as empresas aéreas desejadas obtendo-se 18 variáveis e 30 observações.

As 18 variáveis são as que seguem abaixo, assim como suas respectivas descrições conforme disposto no site da ANAC¹:

- **Sigla (Sigla):** Sigal da empresa aérea em questão.
- **Empresa Aérea (nome):** Empresa Aérea responsável por operar as etapas.
- **Ano (anos):** Ano da observação.
- **Passageiros Pagos (PP):** Refere-se aos passageiros que ocupam assentos comercializados ao público e que geram receita, com a compra de assentos, para a empresa de transporte aéreo. Incluem-se nesta definição as pessoas que viajam em virtude de ofertas promocionais, as que se valem dos programas de fidelização de clientes, as que se valem dos descontos concedidos pelas empresas, as que viajam com tarifas preferenciais, as pessoas que compram passagem no balcão ou através do site de empresa de transporte aéreo e as pessoas que compram passagem em agências de viagem.
- **Passageiros Grátis (PG):** Refere-se aos passageiros que ocupam assentos comercializados ao público mas que não geram receita, com a compra de assentos, para a empresa de transporte aéreo. Incluem-se nesta definição as pessoas que viajam gratuitamente, as que se valem dos descontos de funcionários das empresas aéreas e seus agentes, os funcionários de empresas aéreas que viajam a negócios pela própria empresa e os tripulantes ou quem estiver ocupando assento destinado a estes.
- **Combustível (combL):** Refere-se à quantidade, em litros, de combustível consumida pela aeronave na execução da referida etapa. Informação disponível apenas para empresas brasileiras.
- **Distância (D):** Refere-se à distância, expressa em quilômetros, entre os aeródromos de origem e destino da etapa, considerando a curvatura do planeta Terra.
- **Horas Voadas (H) :** Refere-se ao número de horas de voo entre os aeródromos de origem e destino da etapa.
- **Decolagens (Dec):** Refere-se ao número de decolagens que ocorreram entre os aeródromos de origem e destino da etapa.
- **Carga Paga Km (CP_km):** Refere-se ao volume de Carga Paga (kg) em cada quilômetro, ou seja, a soma do produto entre a quantia (kg) de carga paga e a distâncias das etapas.

¹(<http://www.anac.gov.br/assuntos/dados-e-estatisticas/descricao-de-variaveis>)

- **Carga Grátis Km (CG_km):** Refere-se ao volume de Carga Grátis (kg) em cada quilômetro, ou seja, a soma do produto entre a quantia (kg) de carga grátils e a distâncias das etapas.
- **Correio Km (Corr_km):** Refere-se ao volume de Correio (kg) em cada quilômetro, ou seja, a soma do produto entre a quantia (kg) de correio e a distâncias das etapas.
- **Assentos (A):** é o número de assentos disponíveis em cada etapa de voo de acordo com a configuração da aeronave na execução da etapa.
- **Payload (Pyl):** é a capacidade total de peso na aeronave, expressa em quilogramas, disponível para efetuar o transporte de passageiros, carga e correio.
- **Carga Paga Kg (CP_kg)::** Refere-se à quantidade total, expressa em quilogramas, de todos os bens que tenham sido transportados na aeronave, exceto correio e bagagem, e tenham gerado receitas direta ou indireta para a empresa aérea.
- **Carga Grátis Km (CG_kg)::** Refere-se à quantidade total, expressa em quilogramas, de todos os bens que tenham sido transportados na aeronave, exceto correio e bagagem, e não tenha gerado receitas diretas ou indiretas para a empresa aérea.
- **Correio Kg (Corr_kg):** Refere-se à quantidade de objetos transportados de rede postal em cada trecho de voo realizado, expresso em quilogramas.
- **Bagagem(B_kg):** Refere-se à quantidade total de bagagem despachada, expressa em quilogramas.

O resumo estatístico dessas variáveis podem ser observadas na tabela abaixo:

Figura 1: Sumário da base de Dados Estatísticos

Variable	Obs	Mean	Std. Dev.	Min	Max
Sigla	0				
Nome	0				
anos	30	2012	1,43839	2010	2014
PP	30	1,36e+07	1,46e+07	26206	3,65e+07
PG	30	304771,7	366114,8	322	948706
CP_kg	30	6,45e+07	9,62e+07	1726	3,05e+08
CG_kg	30	1098688	2242843	0	6604094
Cor_kg	30	199836,3	547926,6	0	2277503
combL	30	7,55e+08	9,22e+08	1734976	2,60e+09
D	30	1,39e+08	1,47e+08	2068730	4,02e+08
Dec	30	138711,4	129627,3	6816	319234
CP_km	30	1,67e+11	3,29e+11	853413	1,11e+12
CG_km	30	9,40e+08	1,78e+09	0	4,98e+09
Cor_km	30	3,95e+08	8,75e+08	0	3,86e+09
A	30	2,09e+07	2,27e+07	100335	5,49e+07
H	30	250173,4	248185,7	7980,95	652691
Pyl	30	2,25e+09	2,34e+09	1,23e+07	5,72e+09
B_kg	30	1,75e+08	1,87e+08	224730	4,80e+08

2.2 Dados de Demonstrativo de Relatório Operacional (DRO)

A divulgação das demonstrações contábeis das empresas aéreas encontram-se disponíveis também no site da ANAC, em que constam os anos de 2009 até 2017. Destaca-se que o ano de 2017, até o presente momento em que este trabalho foi realizado, não possuía disponível o DRO do quarto trimestre de 2017. Como o objetivo era iniciar a partir de 2010, com a função de excluir efeitos maiores da crise de 2008 o trabalho possuia a capacidade de abordar os dados de 2010 até 2016, inclusive. Entretanto, houve mudança na organização do DRO no ano de 2015. Com isso alguns itens, importantes para o estudo, foram ou anexados ou excluídos nos novos itens dos demonstrativos posteriores a 2014. A decisão seria ou compreender o estudo de 2010 até 2014 ou então fazer o mesmo para os anos de 2015 e 2016, uma vez que adaptar ou tentar unir ambos os períodos poderiam gerar distorções e efeitos artificiais nos resultados em questão. Para abranger mais dados e fazer uma melhor análise de evolução dos coeficientes de eficiência procurou-se estabelecer a análise para os anos de 2010 até 2014.

Desta maneira, com base nas 6 empresas selecionadas, inclusive por motivos de completude nos dados do DRO, como já foi mencionado, foi realizada mediante coleta manual dos dados, uma base em Excel contendo tanto os itens de Receita como de Custos, obtido por cada uma das empresas no período selecionado de forma que resumisse as contabilidades anuais. Maiores informações sobre as variáveis, inclusive aquelas que serão de fato utilizadas no estudo, serão feitas mais a frente na seções que analisam os resultados e o contexto das empresas.

2.3 Base de dados final

Após obtenção e realização das renomeações das variáveis em cada base de forma a possibilitar a associação delas, como a padronização no nome das empresas aéreas (Sigla), foi realizado mediante pacote estatístico Stata, a aplicação do procedimento de "merge" de maneira que ambas as bases anteriormente descritas formassem uma só. Nela, encontram-se 56 variáveis e 30 observações. Após essa organização, foram criadas as variáveis que serão caracterizadas como insumos e produtos. Algumas delas são variáveis simplesmente obtidas diretamente de uma das bases, mas outras foram construídas por cálculos simples a partir de ambas as bases.

3 Contexto das Empresas selecionadas

Observando a conjuntura econômica do país, é possível verificar que o mesmo experimentou certo crescimento durante o período de 2002 até 2010, a economia apresentava indicativos positivos, tais como a redução da Selic a patamares de 9,73%, inflação controlada e crescimento do PIB. Alguns fatores importantes contribuiram para isto como termos de trocas favoráveis para a exportação, as taxas de juros internacionais baixas que favoreceram os fluxos de investimentos aos países emergentes, o boom das commodities entre outros. Entretanto depois de 2010, sob a presidência de Dilma Rousseff, as políticas fiscais do governo ficaram cada vez mais distantes do ideal tripé econômico, iniciando uma escalada cada vez maior da dívida pública, do aumento da inflação e certo controle artificial do câmbio, sem contar com o aumento progressivo da Selic. Tal fato se torna mais evidente no quadro político e econômico que se estabelece no pós 2014.

Sob a ótica do nosso trabalho alguns fatores devem ser levado em consideração para o período de 2010 até 2014. O primeiro diz respeito aos acontecimentos turísticos como as olimpíadas e a Copa do Mundo (esta ocorrida no Brasil). A implicação disto torna-se evidente no aumento das vendas de curto prazo de passagens aéreas pelas companhias. Outro importante fator, foi que devido ao controle da inflação e do câmbio no período da presidência do governo Lula, tal fato contribuiu para os melhores preços das passagens e do combustível aeronáutico. Tanto é evidente este ponto que os preços das passagens aéreas como pode ser observado no site

da ANAC, caíram em termos substanciais desde 2003 até 2013, quando em 2014 voltou a crescer. Além disso, como maneira de preparar os aeroportos para a recepção de turistas na Copa, houve projetos de desestatização de aeroportos, com a concessão dos mesmos a iniciativa privada, seja ela de maneira parcial ou não, iniciando-se em 2011. Isso permitiu um aumento da infraestrutura aeroportuária, capaz de atrair mais turistas e permitir melhor qualidade dos serviços no setor.

Após essa breve contextualização, vale a pena indicar como estavam cada uma das empresas selecionadas neste período, em termos de administração e gestão logística. As 6 empresas selecionadas foram : TAM (TAM), GOL (GLO), Avianca (ONE), Azul (AZU), Sete (SLX) e Passaredo (PTB).

Em 2010 foi anunciado pelos dirigentes da TAM e da LAN uma fusão entre as duas empresas, criando-se uma holding denominada LATAM Airlines Group. Essa fusão tinha como objetivo garantir um maior desenvolvimento das economias de escala entre ambas as empresas e que por outro lado seria benéfico também para os seus clientes uma vez que haveria o aumento das opções de voos e destinos disponíveis. A associação entre LAN e TAM resultou no transporte de 60,3 milhões de passageiros através da capacidade de uma frota composta de 310 aeronaves.

A GOL foi uma companhia aérea fundada em 2000, sob o conceito principal de fornecer serviços a baixos custos, diminuindo o preço das passagens aéreas. Em 28 de março de 2007, a Gol comprou a marca da Varig, pagando o equivalente a 275 milhões de dólares. A partir de então, a Gol mudou sua razão social para "VRG Linhas Aéreas S/A", adquiriu um Boeing 767 remanescente da companhia e pintou algumas aeronaves com a marca da Nova Varig.

Enquanto isso a companhia aérea Azul, em maio de 2012, anunciava a sua fusão com a empresa TRIP. A TRIP era até então a maior empresa aérea regional na América Latina, e juntas iriam formar uma parcela considerável de voos domésticos. Durante a espera da aprovação da fusão pela ANAC e pelo CADE, as companhias operaram em Code Share, através da união de suas malhas e do serviço de bordo. Resultado disso pode ter sido observado no maior aumento de mercado na região Norte e em alguns aeroportos regionais. A aprovação pela ANAC só ocorreu ao final de 2012 e pelo CADE em 2013.

Para a companhia Passaredo (PTB), foi realizado em 2008 um projeto de expansão da empresas que incluíam compras de aeronaves e renovação do marketing da marca. Em 2009 apresentou um número maior de assentos disponíveis e uma oferta de serviços cada vez maiores, junto com o processo de expansão acelerado de sua malha. Em 2011, a Passaredo aposentou algumas de suas aeronaves e deixou de realizar determinados tipos de rotas. De 2008 até 2012, a companhia operou algo em torno de 16 aeronaves. Em outubro de 2012, ela ingressou com pedido de Recuperação Judicial com o objetivo de viabilizar o pagamento do passivo existente devido a diversos fatores, como o preço do combustível que era relativamente alto para a companhia, e assim, tentar promover a conservação de suas atividades, nas quais já atuava há mais de dezessete anos. Ao término de 2014, a dívida já havia sido reduzida para 90 milhões.

Observando a companhia aérea SETE (Serviços Especiais de Transportes Executivos) verifica-se que a mesma suspendeu suas operações em 2016 devido a conjuntura econômica do país. Em 2014 a companhia se firmou como regional de destaque no elo Centro-Oeste para o Norte do Brasil. Entretanto em meados de dezembro de 2015 a empresa anunciou o encerramento de suas operações regulares no último dia de 2015, focando somente com as operações de táxi aéreo. De maneira resumida para a Avianca (Sigla:ONE) a mesma manteve uma trajetória relativamente estável, aposentando em 2011 certas aeronaves e renovando a marca em 2013, após a sua união com a empresa aérea Star Alliance em 2012.

Tendo em vista as ações expressivas apresentadas durante esse período pelas companhias em questão, torna-se ideal a análise mais próxima do DRO, apresentados oficialmente no site da ANAC. Começando pela observância das receitas e do número de passageiros de cada uma delas, tem-se o exposto a seguir:

Figura 2: Fração de Receitas provenientes de Passagem

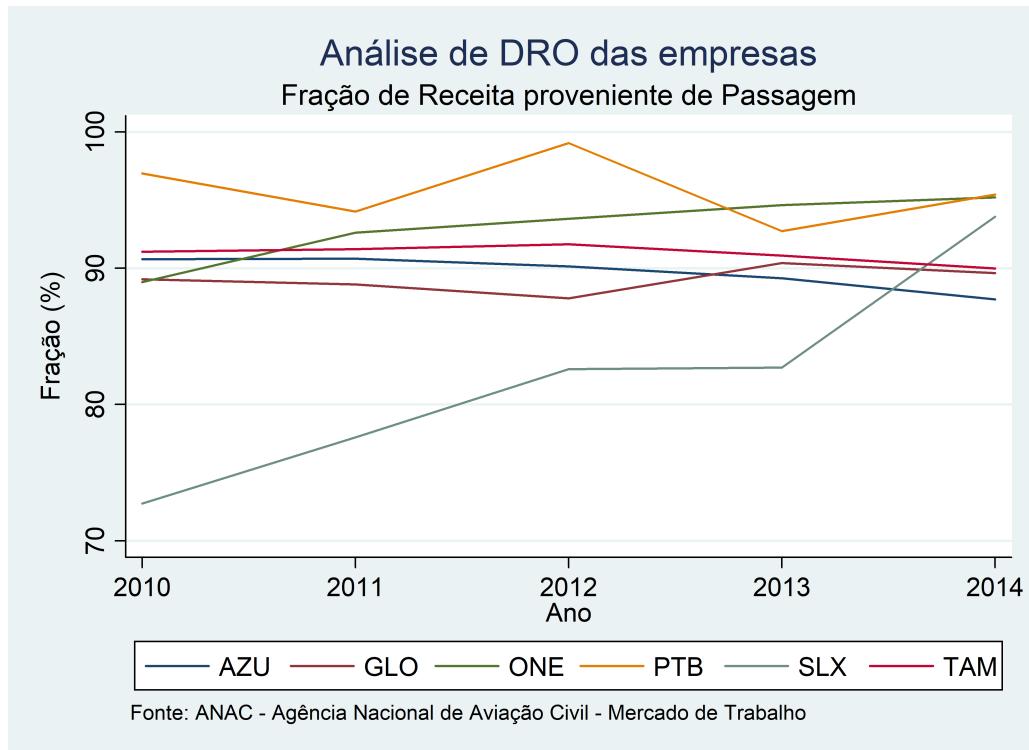


Figura 3: Receitas provenientes de Passagem

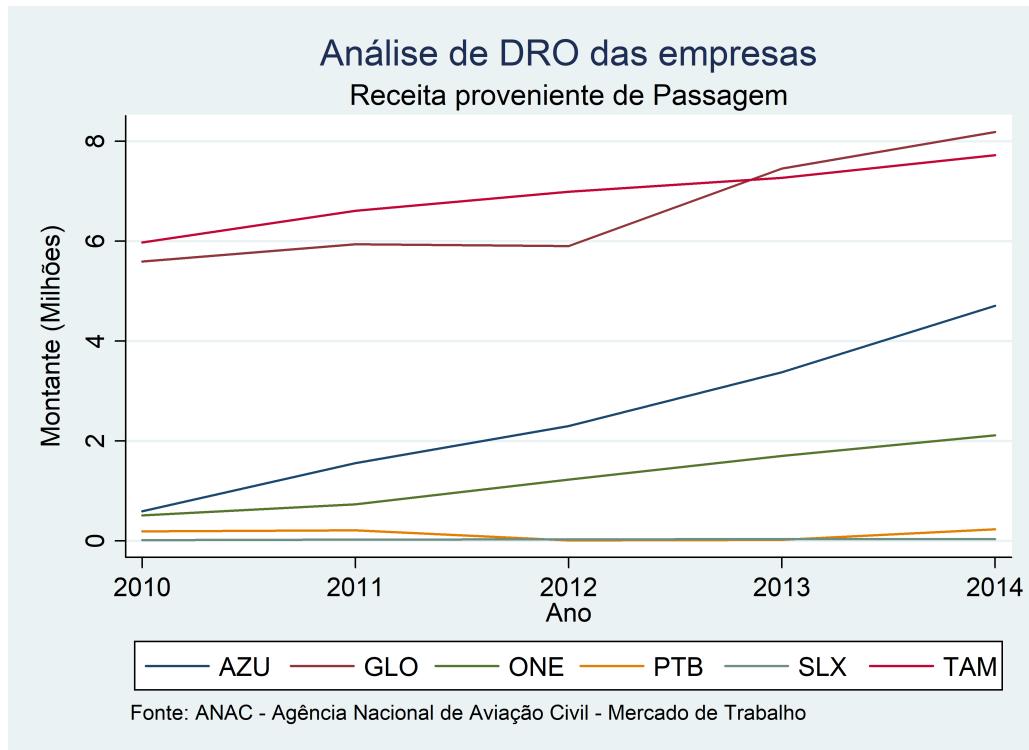


Figura 4: Total da Receita de Voo

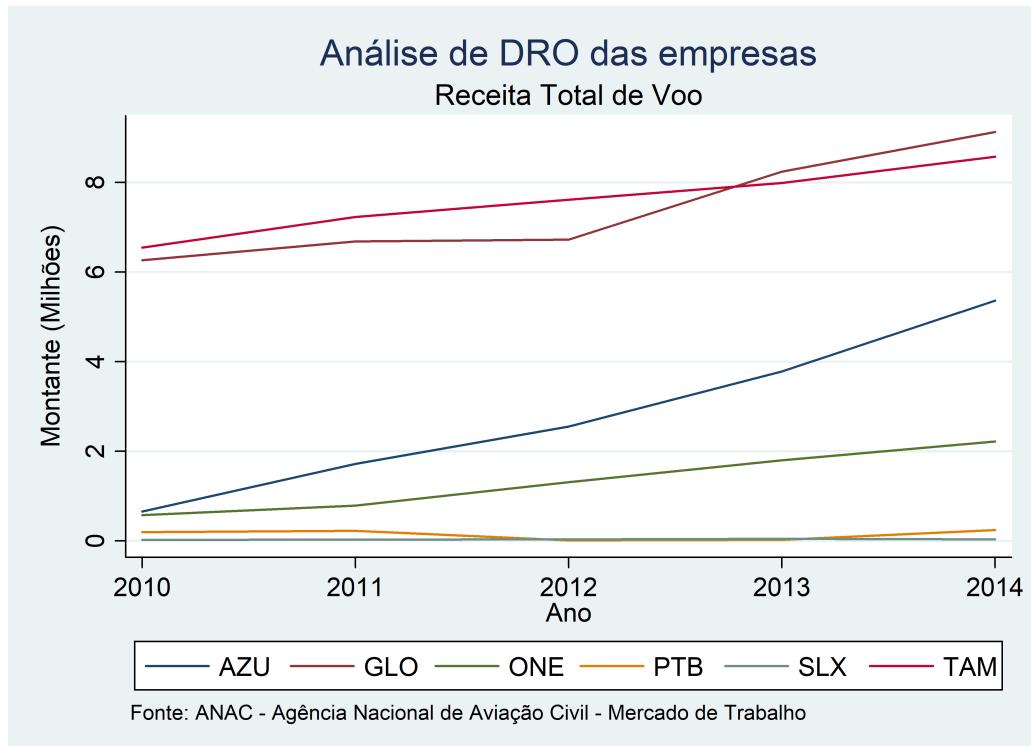
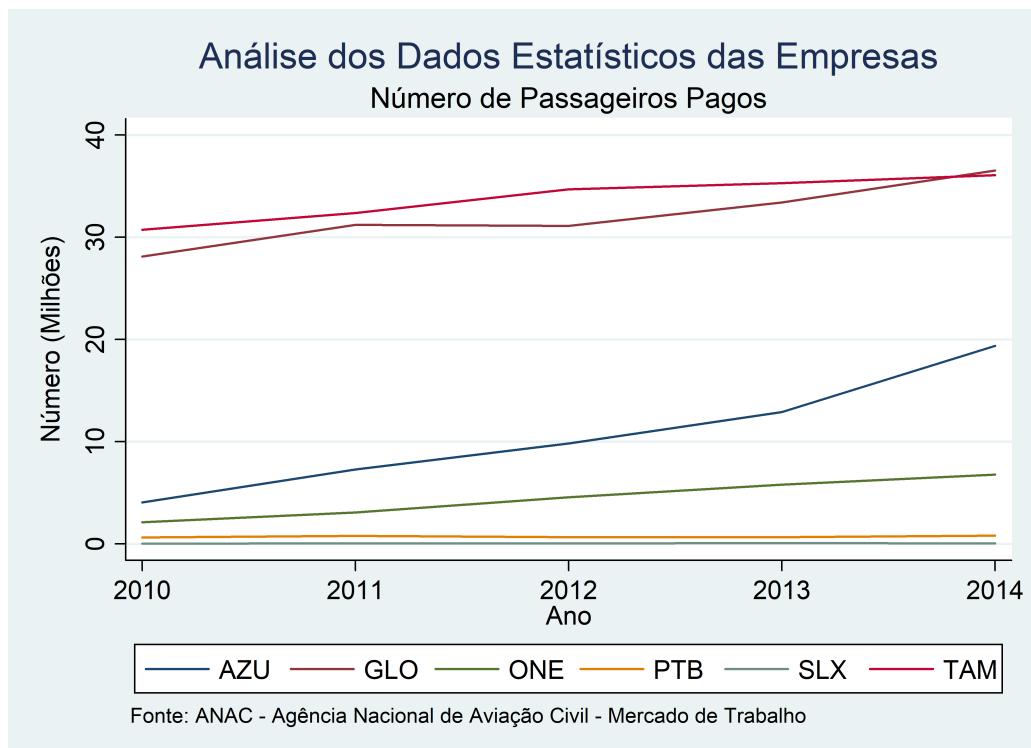


Figura 5: Quantidade de Passageiros Pagos



O ponto mais óbvio que pode ser observado, é justamente a existência de 3 grupos de companhias: pequenas, médias e grandes. O primeiro composto pela SETE e Passaredo, o segundo pela Avianca e Azul e o terceiro pela GOL e TAM. Inclusive tal fato pode ser visto quando são analisadas outras estatísticas destas companhias, bastando observar o número de assentos, a distância percorrida em Km, o número de litros de combustível consumido, o total de

carga paga transportada dentre outras. E embora sejam diferentes em termos de escala nessas diversas estatísticas, as empresas apresentam uma mesma proporção semelhante de receitas obtidas em passagens aéreas, caracterizando um perfil parecido em termos de operação. Inclusive ressalta-se que o montante e a fração obtidos pela venda de passagens aéreas são de fato responsáveis pelos ganhos substanciais dessas companhias.

Analisando os custos, observa-se dentro dos custos totais, subgrupos principais de custos. O primeiro deles são os custos operacionais que são compostos por organização de tráfego de passageiros, organização de cargas, além de despesas administrativas e gerais, e por fim outras despesas genéricas. No gráfico a seguir seguem a evolução destes custos no período. O último gráfico elucida justamente a fração dos custos representados nos outros gráficos diante do custo total obtido em cada ano.

Figura 6: Fração de Custos Administrativos

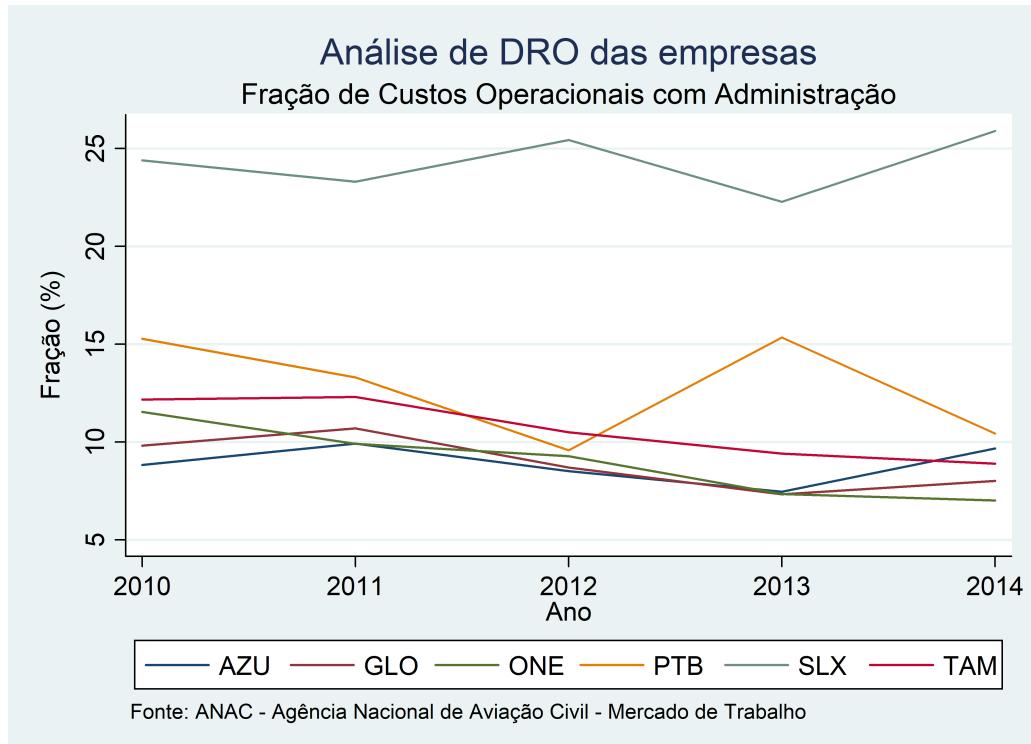


Figura 7: Fração de Custos com Cargas

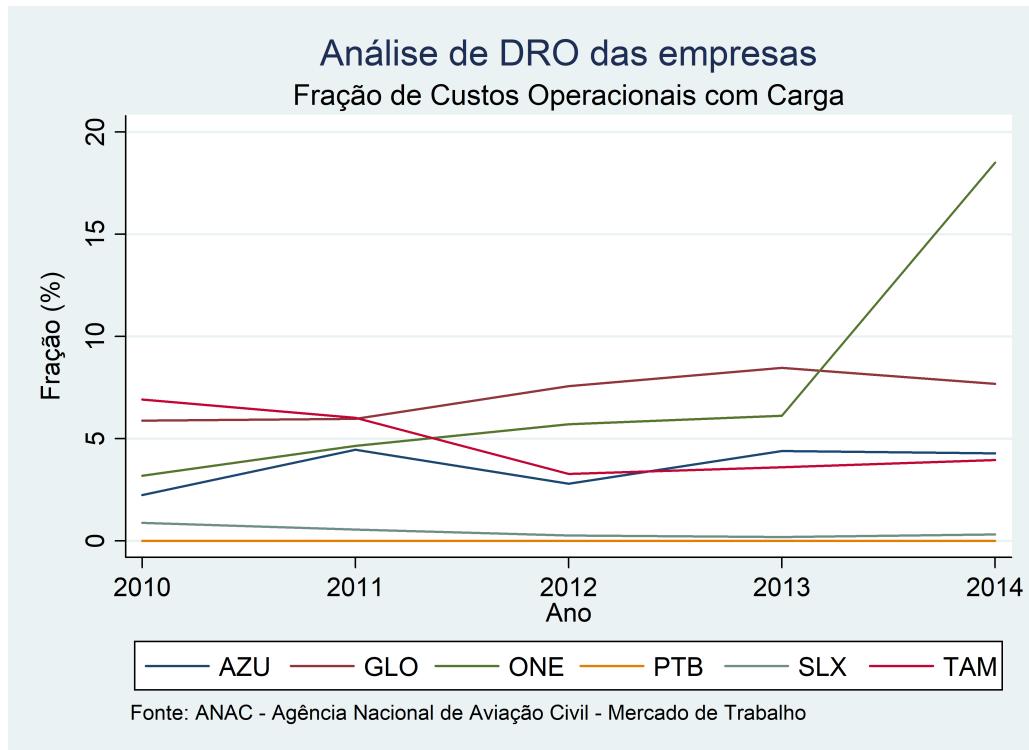


Figura 8: Fração de Custos com Tráfego

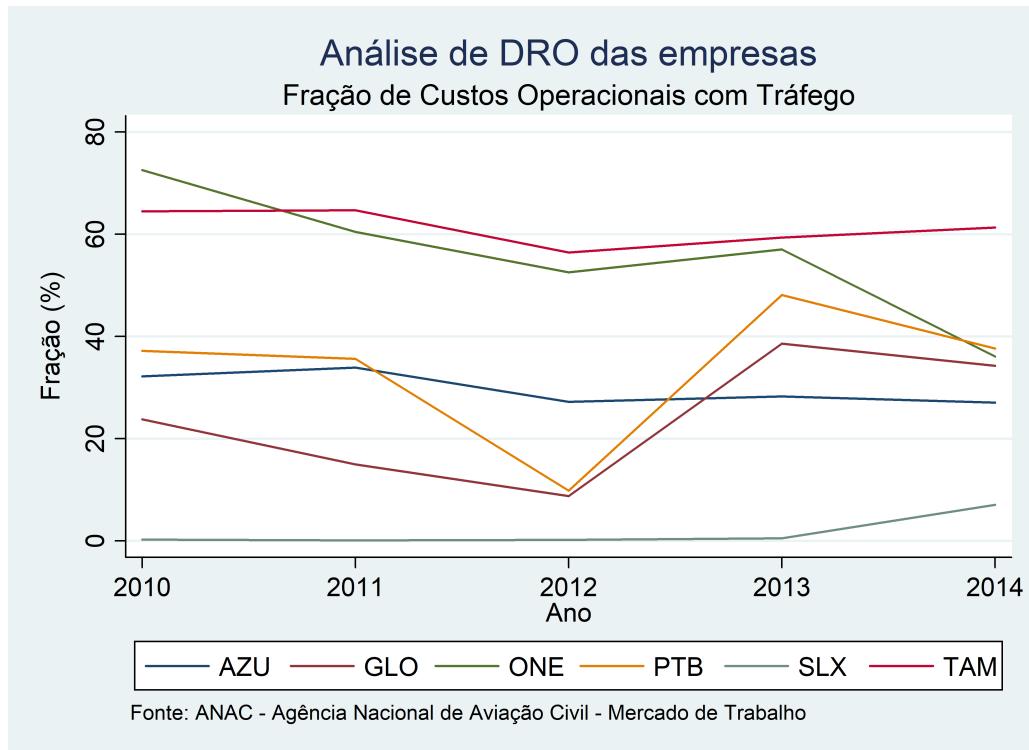
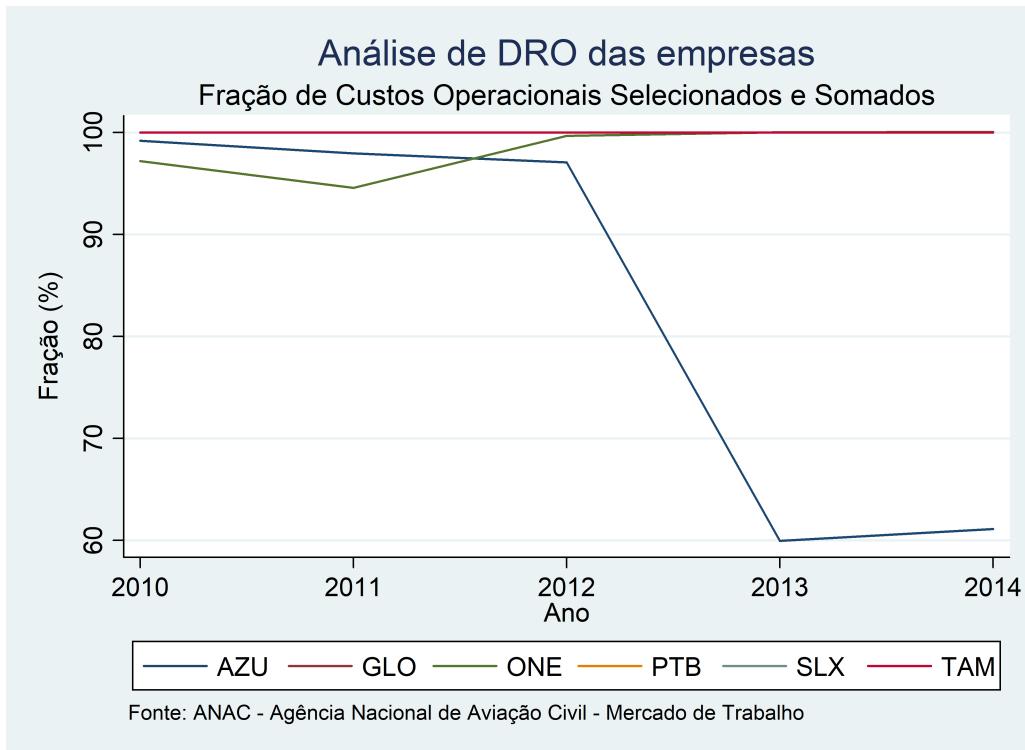


Figura 9: Fração de Custos Operacionais



O segundo e mais importante são os custos diretos, em que os mesmos são compostos por custos com tripulantes técnicos, com comissários de bordo, com combustível, com a depreciação dos equipamentos de voo, com manutenções e revisões das aeronaves, com o seguro e arrendamento delas também e por fim por custos oriundos de tarifas tanto de navegação aérea quanto de aeroportuárias. O principal componente dos custos, e que será sem dúvida levado em consideração na análise de eficiência reside naquele obtido na compra de combustíveis e nos gastos em comissários e tripulantes técnicos. Vale a pena destacar, que nestes tripulantes técnicos incluem-se os pilotos das companhias.

Figura 10: Fração de Custos com Combustível

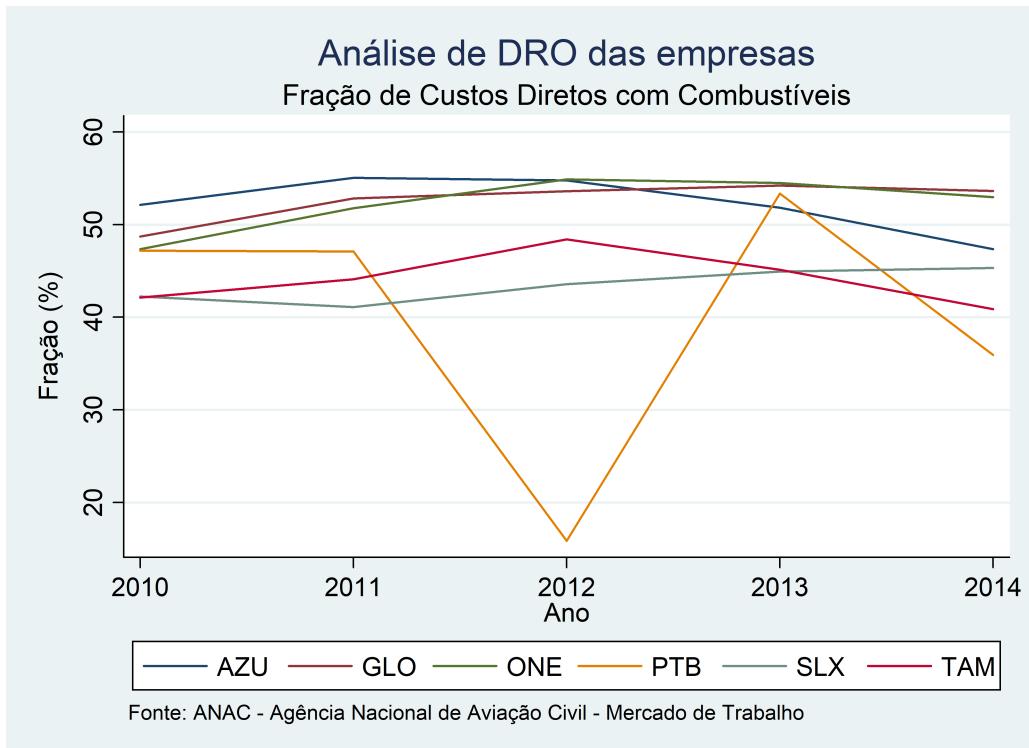


Figura 11: Quantidade em Litros de Combustível

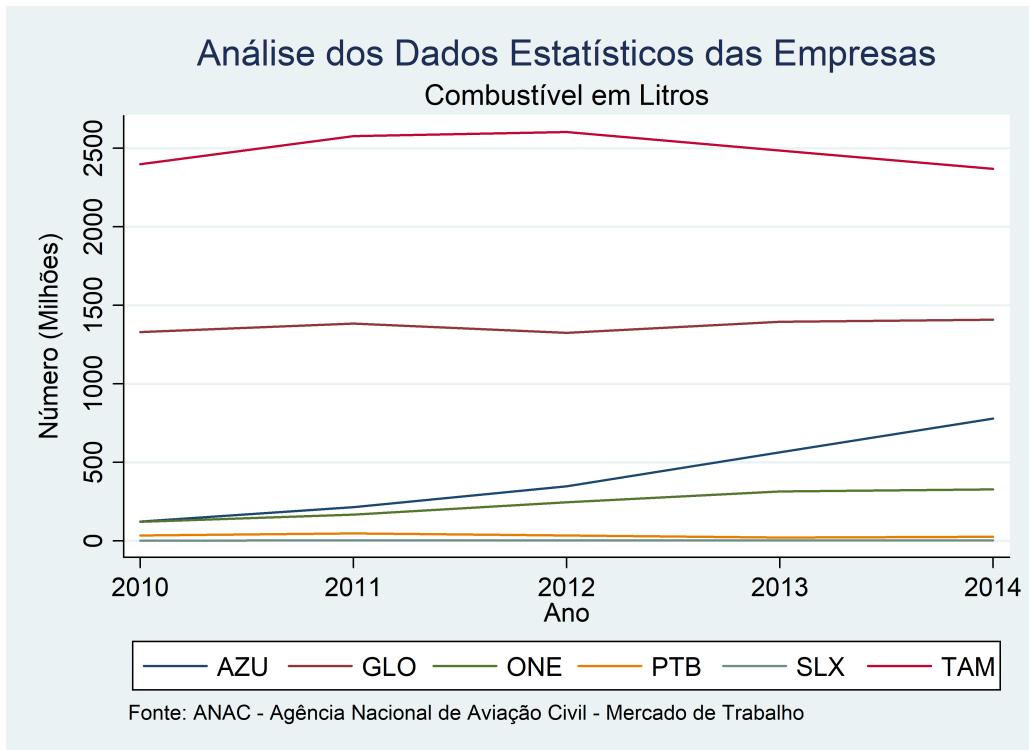


Figura 12: Fração de Custos com Comissários

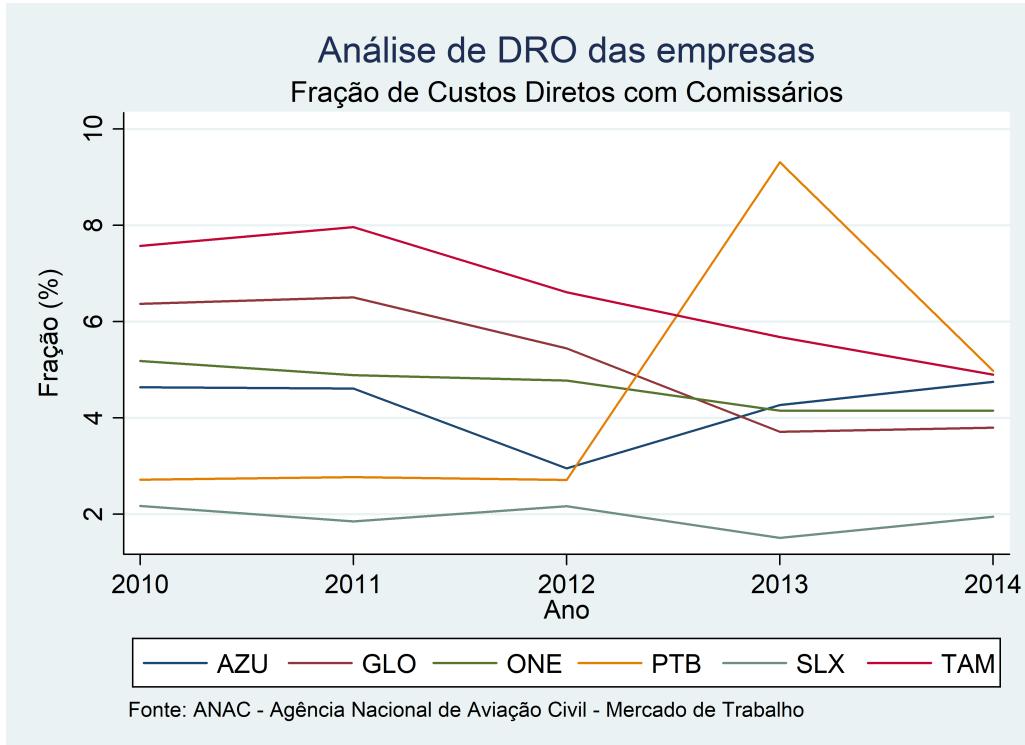
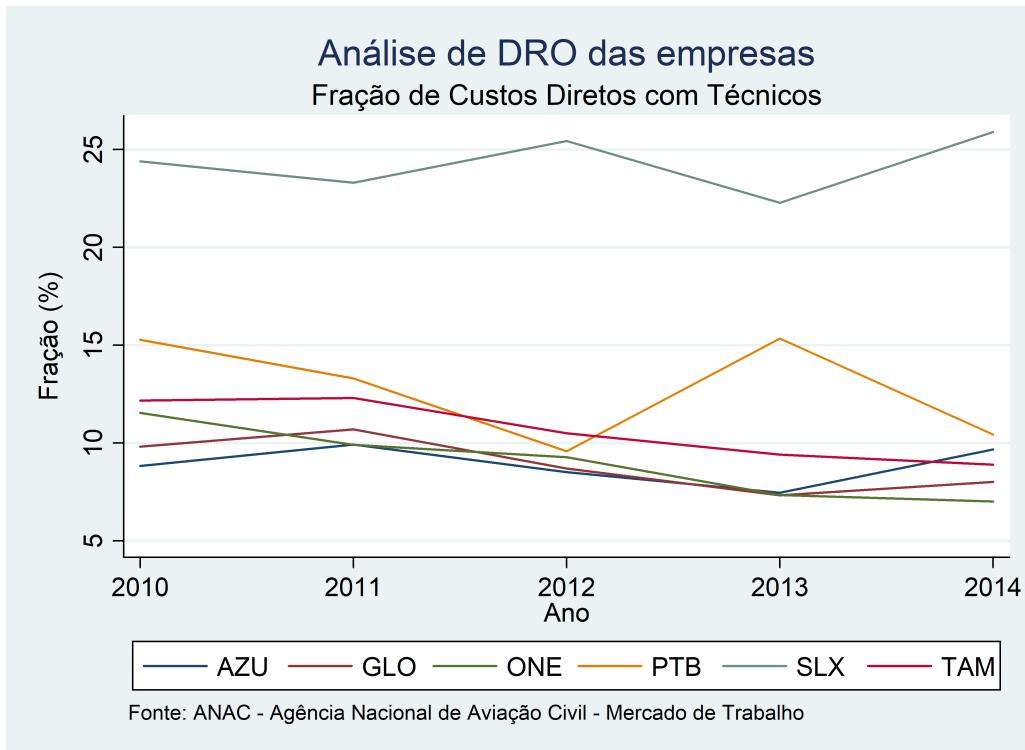


Figura 13: Fração de Custos com Técnicos



O terceiro e último tipo de custo são os indiretos. Neles existem apenas dois custos que chamam a atenção o de serviços de bordo e o de organização terrestre. Embora o de organização terrestre tenha uma magnitude maior, o primeiro custo apresenta maior significado em termos de qualidade do serviço prestado aos passageiros que representam, como visto anteriormente, em grande parte das receitas das companhias aéreas. Entretanto, não nos aprofundaremos nas

observações em torno destes custos, uma vez que são bem menores quando em comparação com os dois outros tipos.

4 Especificação do Modelo de Eficiência

A especificação da função de produção seguiu o formato de uma Cobb-Douglas. Os insumos que foram caracterizados como importantes para o processo de geração do output, encontram-se listados abaixo:

- insumo 1 : representando o número de assentos (A).
- insumo 2 : representando as despesas administrativas gerais retiradas do DRO.
- insumo 3 : representando o payload total dos voos de um ano (Pyl).
- insumo 4: razão do custo de combustível retirado do DRO pela quantidade em Litros de combustível, de querosene de aviação (combL).
- insumo 5 : a razão do custo com comissários de bordo somados ao custo com tripulantes técnicos dividido pela quantidade de horas de voo.

Através destes 5 insumos, verifica-se que a disponibilidade de assentos, assim como o peso total da aeronave associados aos gastos tanto em serviços administrativos como em tripulantes técnicos e comissários de bordo influenciam no output gerado. A descrição estatística desses insumos, dos 5 anos ao todo, se encontram abaixo:

Figura 14: Descrição dos Insumos

Variable	Obs	Mean	Std. Dev.	Min	Max
insumo_1	30	2,09e+07	2,27e+07	100335	5,49e+07
insumo_2	30	267584	314472,6	1677,612	953762,1
insumo_3	30	2,25e+09	2,34e+09	1,23e+07	5,72e+09
insumo_4	30	,0020975	,0009032	,0001484	,0037604
insumo_5	30	1,120809	,4615447	,0928041	1,681982

Um dos principais problemas que foram encontrados no momento da execução dos métodos econômétricos foi justamente a alta correlação entre esses insumos. Tal fato decorre do princípio de que muitos insumos dependem do peso da aeronave para poderem ter seus níveis estabelecidos, como por exemplo a quantidade, em Litros, de combustível a ser utilizada. A forma encontrada para resolver este problema foi justamente a aplicação da análise de componentes principais e extrair componentes que fossem capazes de representar tais insumos. O correograma dos insumos ao longo dos 5 anos pode ser observado abaixo:

Figura 15: Correlograma entre os insumos

	insumo_1	insumo_2	insumo_3	insumo_4	insumo_5
insumo_1	1,0000				
insumo_2	0,9550	1,0000			
insumo_3	0,9969	0,9313	1,0000		
insumo_4	-0,7890	-0,6519	-0,8114	1,0000	
insumo_5	0,6675	0,5803	0,6755	-0,9448	1,0000

Os mesmos resultados podem ser obtidos quando separam-se os dados por ano. Através da aplicação da análise de componentes principais, foram obtidos componentes, que seguindo a sequência correta de seleção da quantidade de componentes, apontavam para um ou dois componentes dependendo do ano. Nas figuras abaixo é possível verificar os resultados obtidos via análise de componentes principais para cada ano:

Figura 16: Resultado PCA - 2010

```
Principal components/correlation
Number of obs      =       6
Number of comp.   =       1
Trace              =       5
Rotation: (unrotated = principal) Rho          =  0,8425
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Compl	4,21226	3,52757	0,8425	0,8425
Comp2	,684691	,589516	0,1369	0,9794
Comp3	,095175	,0876126	0,0190	0,9984
Comp4	,00756248	,0072518	0,0015	0,9999
Comp5	,000310676	.	0,0001	1,0000

Figura 17: Resultado PCA - 2011

```
Principal components/correlation
Number of obs      =       6
Number of comp.   =       1
Trace              =       5
Rotation: (unrotated = principal) Rho          =  0,7934
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Compl	3,96709	3,34141	0,7934	0,7934
Comp2	,625674	,301406	0,1251	0,9186
Comp3	,324268	,241463	0,0649	0,9834
Comp4	,0828043	,0826372	0,0166	1,0000
Comp5	,000167017	.	0,0000	1,0000

Figura 18: Resultado PCA - 2012

Principal components/correlation	Number of obs = 6
	Number of comp. = 2
	Trace = 5
Rotation: (unrotated = principal)	Rho = 0,9302

Component	Eigenvalue	Difference	Proportion	Cumulative
Compl	3,50427	2,35763	0,7009	0,7009
Comp2	1,14665	,836914	0,2293	0,9302
Comp3	,309733	,273011	0,0619	0,9921
Comp4	,0367219	,0340978	0,0073	0,9995
Comp5	,00262408	.	0,0005	1,0000

Figura 19: Resultado PCA - 2013

Principal components/correlation	Number of obs = 6
	Number of comp. = 2
	Trace = 5
Rotation: (unrotated = principal)	Rho = 0,9328

Component	Eigenvalue	Difference	Proportion	Cumulative
Compl	3,52548	2,3868	0,7051	0,7051
Comp2	1,13868	,822006	0,2277	0,9328
Comp3	,316671	,298095	0,0633	0,9962
Comp4	,0185756	,0179779	0,0037	0,9999
Comp5	,000597738	.	0,0001	1,0000

Figura 20: Resultado PCA - 2014

Principal components/correlation	Number of obs = 6
	Number of comp. = 1
	Trace = 5
Rotation: (unrotated = principal)	Rho = 0,7556

Component	Eigenvalue	Difference	Proportion	Cumulative
Compl	3,77793	3,0841	0,7556	0,7556
Comp2	,693828	,200226	0,1388	0,8944
Comp3	,493603	,459181	0,0987	0,9931
Comp4	,0344219	,0342078	0,0069	1,0000
Comp5	,000214101	.	0,0000	1,0000

Pelos resultados acima, verifica-se que para os anos de 2010, 2011 e 2014, o ideal é a utilização de um componente principal, enquanto que para os anos de 2012 e 2013 temos que o indicado

é a utilização de dois componentes. Porém, verificam-se duas condições que nos induzem a utilizar apenas um componente principal para todos os anos. O primeiro se refere ao fato de que a proporção explicativa da primeira componente principal já é suficientemente alta para não precisar de uma segunda componente, perceba que a proporção no pior caso chega a 70%. Além disso, de forma a corroborar esse ponto, temos que a diferença ainda continua alta entre o primeiro e o segundo componente principal.

O segundo ponto se refere ao fato de que ao se testar mais de um componente no modelo via OLS, os resultados pioravam muito quando em comparação com apenas um componente, não em termos de fitness do modelo, mas em termos da presença de coeficientes estatisticamente insignificantes. Somado a isso, a interpretação dos sinais e valores da componente principal normalizada foram apenas coerentes para o primeiro componente principal. Portanto, pelos motivos apresentados, foi estabelecido que a função de produção ficou caracterizada por apenas um output e um input, sendo este último representado por apenas um componente principal.

O input então tornou-se a associação dos 5 insumos apresentados anteriormente, na forma do componente principal, em que constam sinais positivos para todos os insumos com exceção do insumo relacionado ao combustível, cujo sinal é negativo. A interpretação dos sinais positivos explicam justamente que uma maior quantidade do insumo implicam em mais input para gerar uma quantidade maior de output, enquanto que o sinal negativo do insumo 4 expressa em termos econômicos o ganho de escala, pois o combustível relaciona-se não apenas com a quantidade de peso, mas também com a qualidade e quantidade de aeronaves que existem nas companhias. Além disso, o preço do combustível sofre "diminuição" de magnitude quando a compra é realizada em volumes maiores.

A definição do output foi estabelecida mediante a soma da normalização das distâncias (em Km) e dos pesos (em Kg) sobre a quantidade de passageiros, de correio e de carga todos eles pagos. Vale a pena ressaltar que o peso médio para o cálculo da quantidade de passageiros em Kg foi de 75.00 Kg. Desta maneira, foi garantido as três principais formas de obtenção de renda pelas companhias aéreas de maneira que a análise de eficiência não ficasse subjugada ou enviesada a apenas uma destas formas. Por outro lado, pela similaridade de comportamento entre as empresas como pode ter sido observado na sessão anterior, tal evento não seria muito expressivo caso viesse a ocorrer. A forma matemática para descrever o Output é dada pela equação abaixo:

$$Output = \frac{Qtd.Passageiros(Kg)}{D} + \frac{Qtd.Carga(Kg)}{D} + \frac{Qtd.Correio(Kg)}{D}$$

A função de produção então ficou descrita como segue abaixo nas equações:

$$Output_t = A_t \cdot (Input_t)^{\alpha_t}$$

onde temos que

$$Input_t = c_{1t} \cdot (insumo_{1t}) + c_{2t} \cdot (insumo_{2t}) + c_{3t} \cdot (insumo_{3t}) + c_{4t} \cdot (insumo_{4t}) + c_{5t} \cdot (insumo_{5t})$$

Em que t assume os valores dos anos contidos entre 2010 e 2014 inclusive. Como já foi esclarecido anteriormente, serão feitos estudos de eficiência em cada ano e depois será observado a evolução de cada empresa no período. Existe uma observação importante a ser realizada em torno da obtenção dos resíduos das regressões feitas em OLS, na primeira equação (após simples aplicação do logaritmo para efetuar a linearização da equação) que é o fato de que por serem poucas observações a cada ano, da ordem de 6 observações, os resultados podem por vezes apresentar certos coeficientes fora do nível mínimo de 5% de confiança. Entretanto optou-se por proseguir nesta direção apesar deste problema.

Ainda sobre a especificação dos insumos antes mesmo de aplicá-los em qualquer regressão, procurou-se normalizar os resultados do componente principal escolhido uma vez que tais insumos devem possuir um valor positivo para que haja quaisquer tipo de interpretação para o estudo. Assim o que se fez foi uma normalização simples através de uma distribuição uniforme no intervalo de 1 a 2, pois valores próximos a 0 terminam por gerar problemas de multicolinearidade ou simplesmente prejudicam o procedimento de inversão de matrizes, assim como é necessário na aplicação do método de DEA. Não apenas os insumos foram normalizados desta maneira, mas também o output, para que a magnitude de ambos os termos ficassem próximos e não gerassem distorções indesejadas. Após esta definição da especificação do modelo e dos tratamentos essenciais dos dados assim como das variáveis a serem utilizadas, realizou-se aplicação dos três métodos. Os resultados podem ser visualizados na sessão a seguir.

5 Resultados

5.1 Resultados pelo método de DEA

Neste método foram consideradas três hipóteses principais: Retornos Constantes de Escala, Retornos Variantes de Escala e Retornos Decrescentes de Escala. Para cada uma destas hipóteses os resultados apontaram para os coeficientes de eficiência descritos nos gráficos abaixo:

Figura 21: Hipótese de Retornos Constantes de Escala

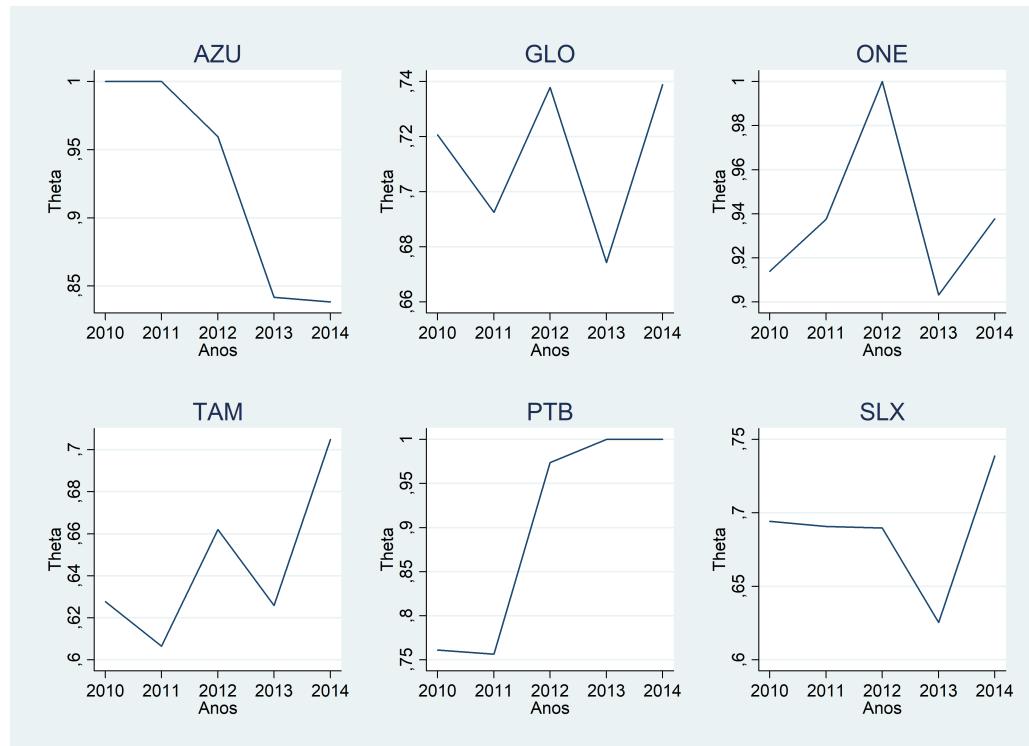


Figura 22: Hipótese de Retornos Variantes de Escala

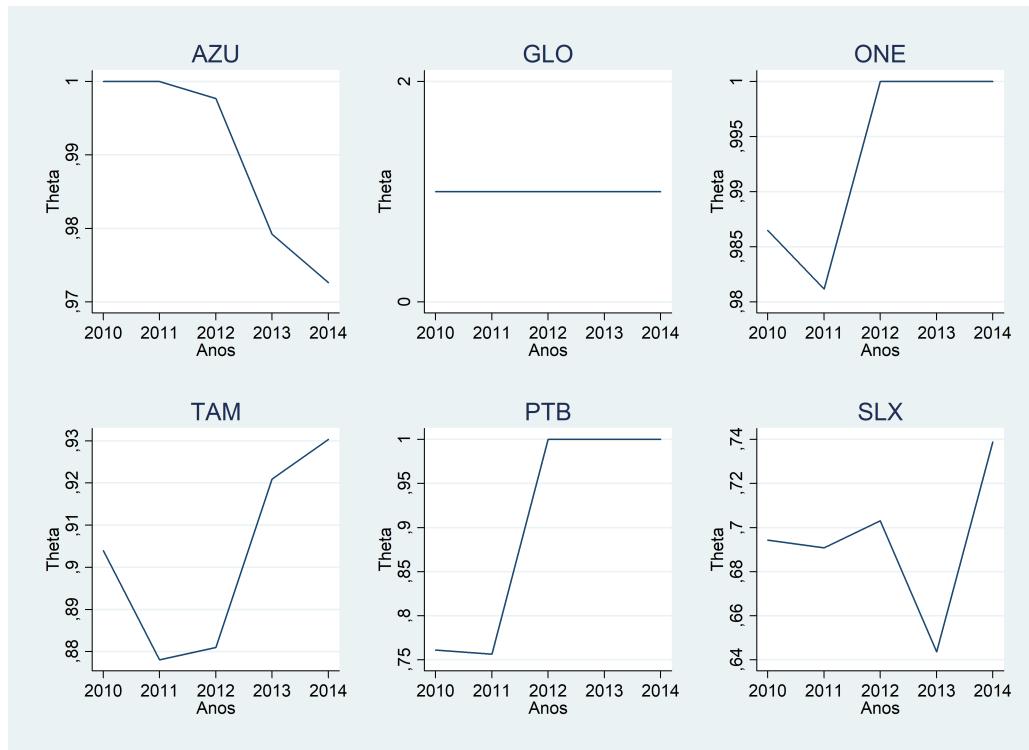
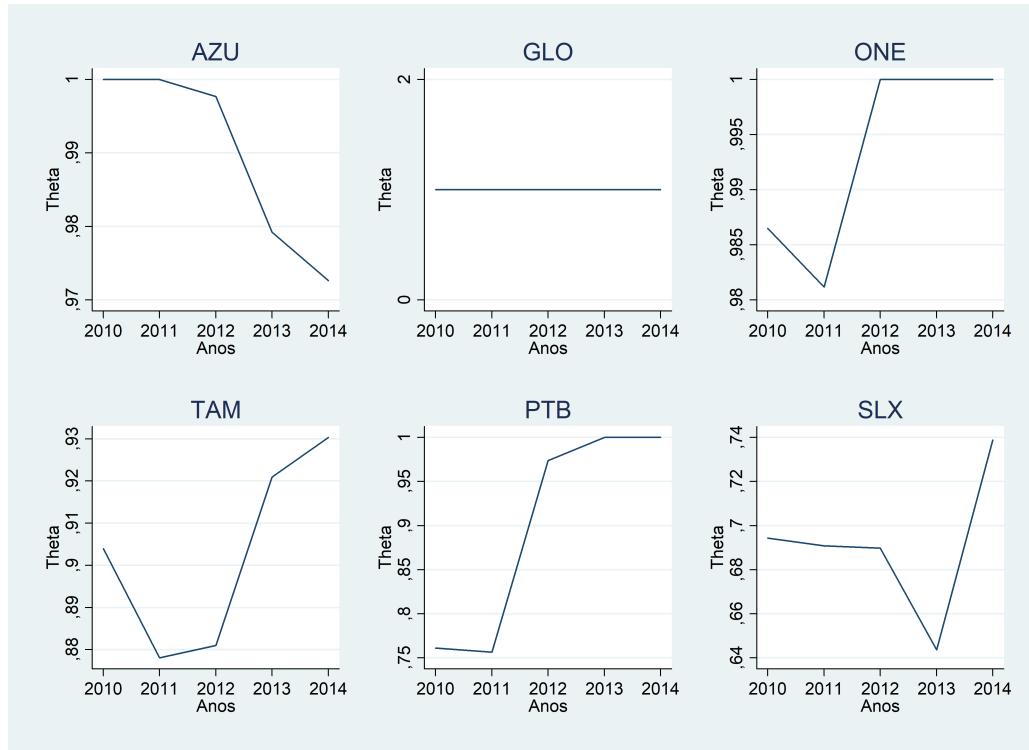


Figura 23: Hipótese de Retornos Decrescentes de Escala



Algumas importantes observações podem ser feitas sobre os gráficos acima. A primeira delas é que a variação das hipóteses de VRS para DRS não foi suficiente para deslocar o rank entre as empresas, tal fato apenas ocorreu quando em comparação com a hipótese de CRS. Segundo é que as mesmas duas variações comentadas não foram capazes de modificar substancialmente o formato das curvas. Outro ponto é que tais resultados demonstram certo nível de aderência

com relação ao contexto das empresas que já foi discutido em sessão anterior. Por exemplo, a recuperação da TAM após fusão com outra empresa, e a recuperação da dívida promovida pela empresa Passaredo que inicialmente apresentava baixa eficiência mas que com o tempo se levou de maneira significativa. É possível evidenciar também que o fato de se ter uma empresa com porte maior (em termo de escala de produção) não implica em maiores níveis de eficiência, uma vez que a locação de recursos de forma eficiente não necessariamente se correlaciona com a quantidade utilizada de inputs e nem a gerada de output, e sim com a razão delas como discutido na sessão do método.

Além daquilo que foi citado acima, aparentemente as curvas de eficiência da Azul e da TAM parecem ser coerentes com o método. Entretanto, a forma como foi padronizada os outputs e os inputs podem ter feito com que a curva da empresa SLX ficasse invariante aos diferentes métodos (inclusive aos que serão apresentados mais à frente). Embora isso possa ser considerada uma falha metodológica, ressalta-se que pelo fato do estudo apresentar uma comparação relativa entre os diferentes níveis de eficiência das empresas sempre haverá uma escolhida como a pior para determinar a eficiência dos demais. Logo isso não produz problemas ao objetivo de comparação relativo que aqui é estipulado.

Logo, as fronteiras de produção para cada ano podem ser vistas, para cada um desses métodos de acordo como o exposto no gráfico abaixo, em que a disposição dos mesmos seguem em ordem crescente de ano da esquerda para direita e de cima para baixo:

Figura 24: Descrição da Fronteira de Eficiência : 2010

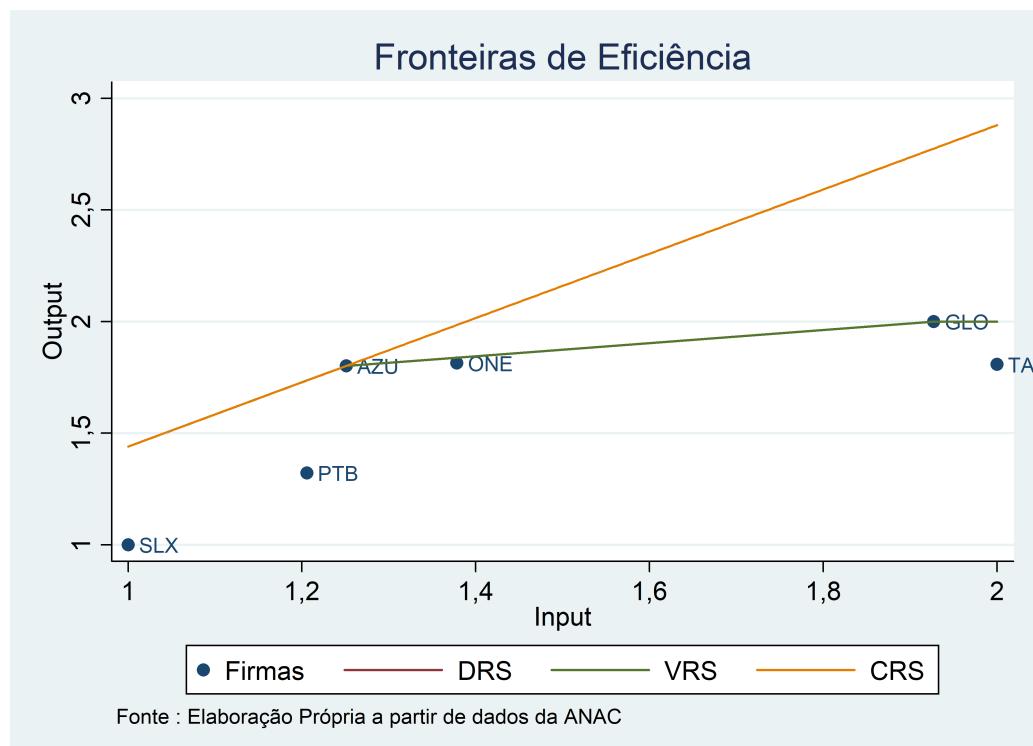


Figura 25: Descrição da Fronteira de Eficiência : 2011

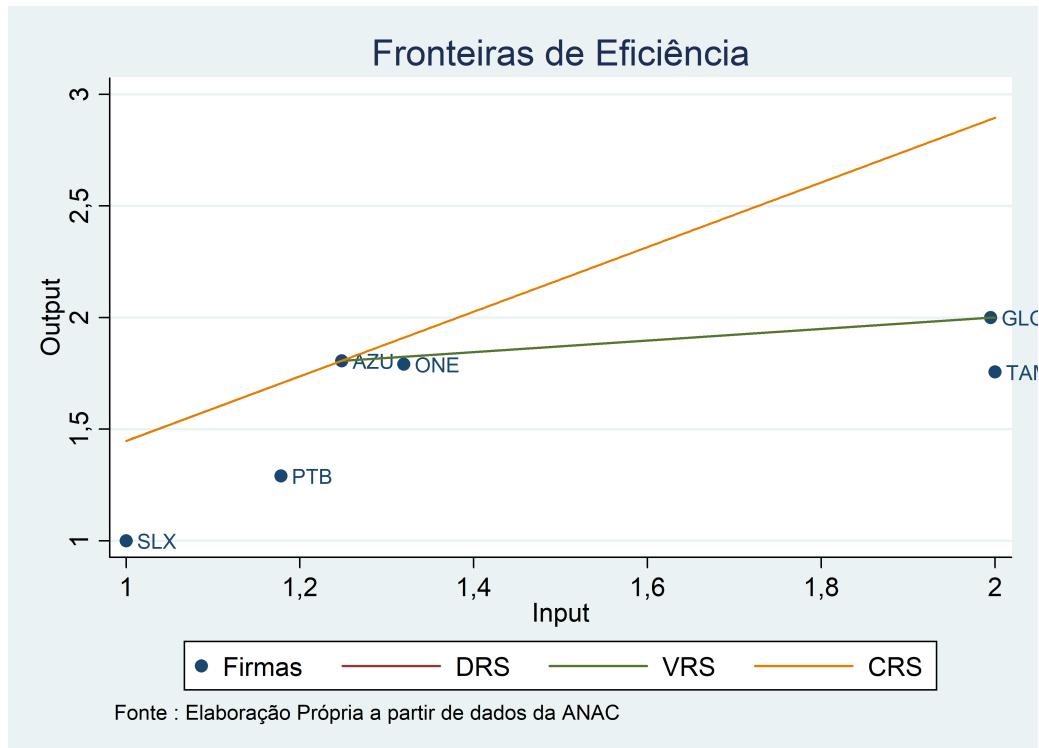


Figura 26: Descrição da Fronteira de Eficiência : 2012

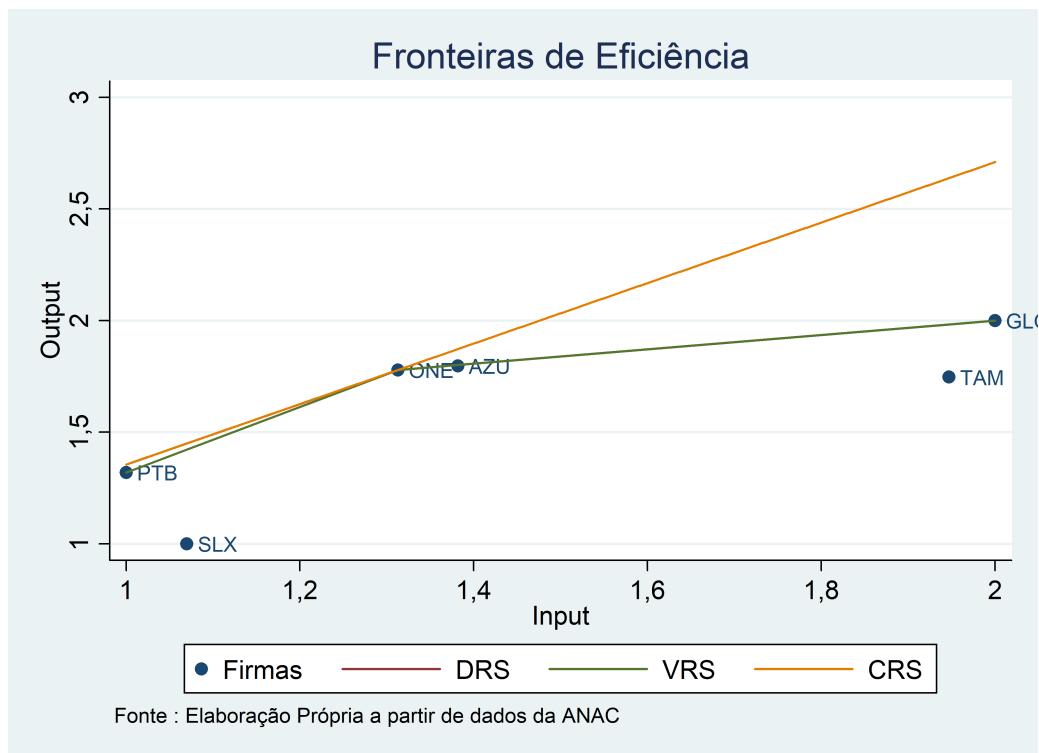


Figura 27: Descrição da Fronteira de Eficiência : 2013

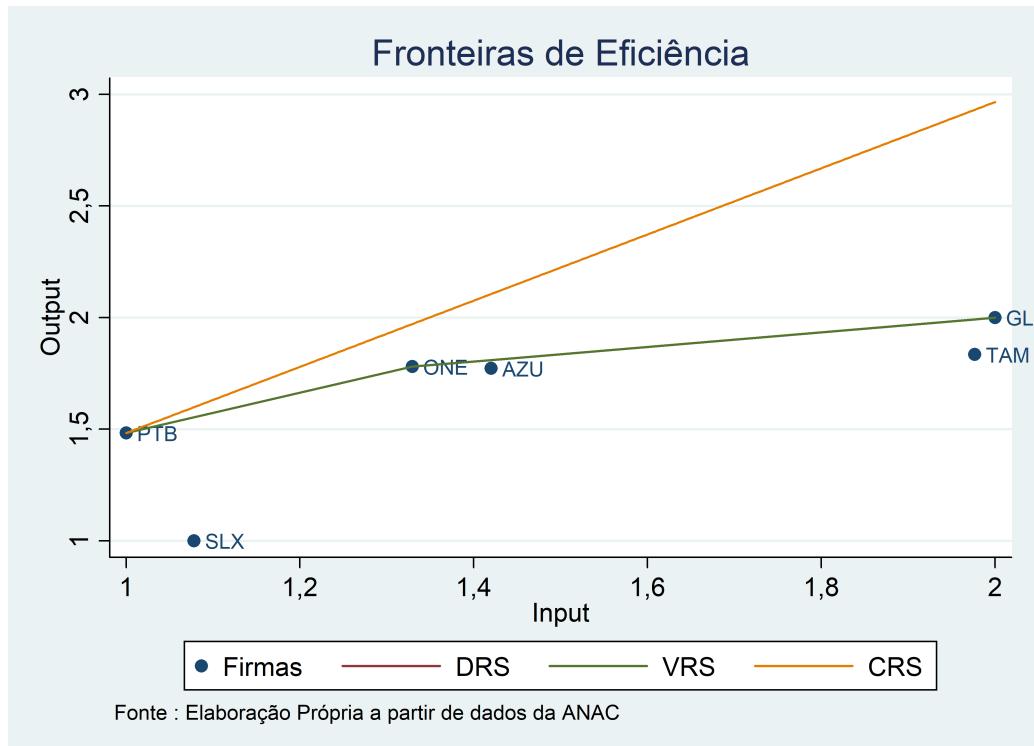
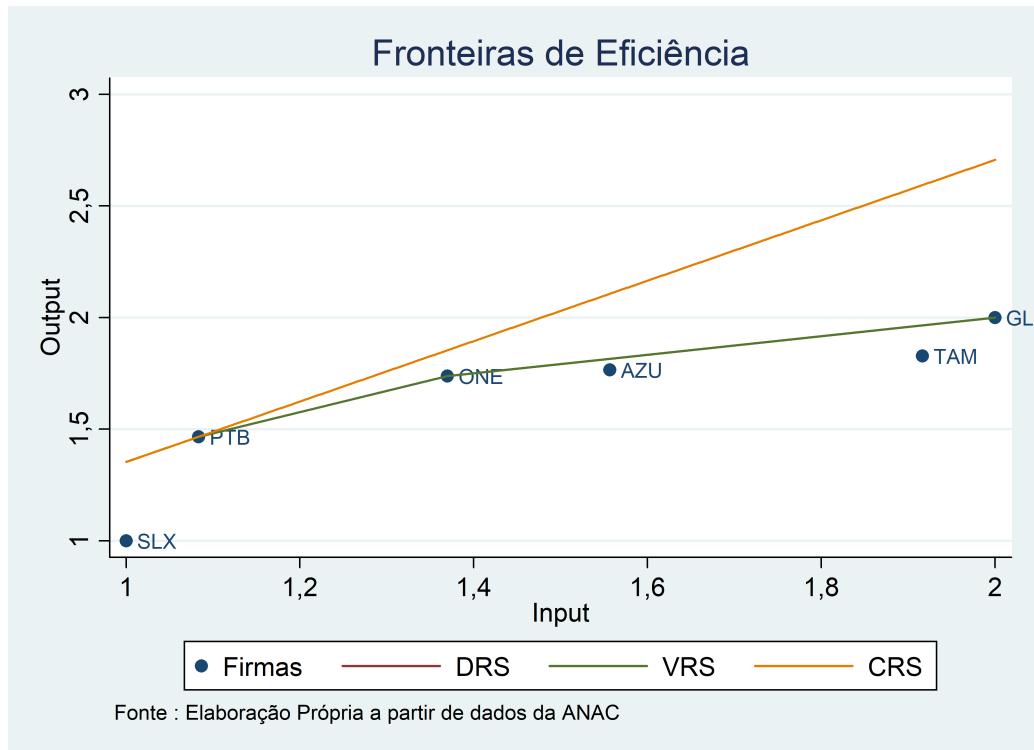


Figura 28: Descrição da Fronteira de Eficiência : 2014



5.2 Resultados pelo método de DFA

Conforme discutido nas seções anteriores, o método não inclui choques exógenos na função de produção de maneira que tudo se passa em um ambiente determinístico. A função de produção considerada para efeitos comparativos entre os métodos é o mesmo. Os resultados

das fronteiras de eficiência para cada ano podem ser visualizadas nos 5 gráficos abaixo.

Figura 29: Resultado de DFA - cross section ano de 2010

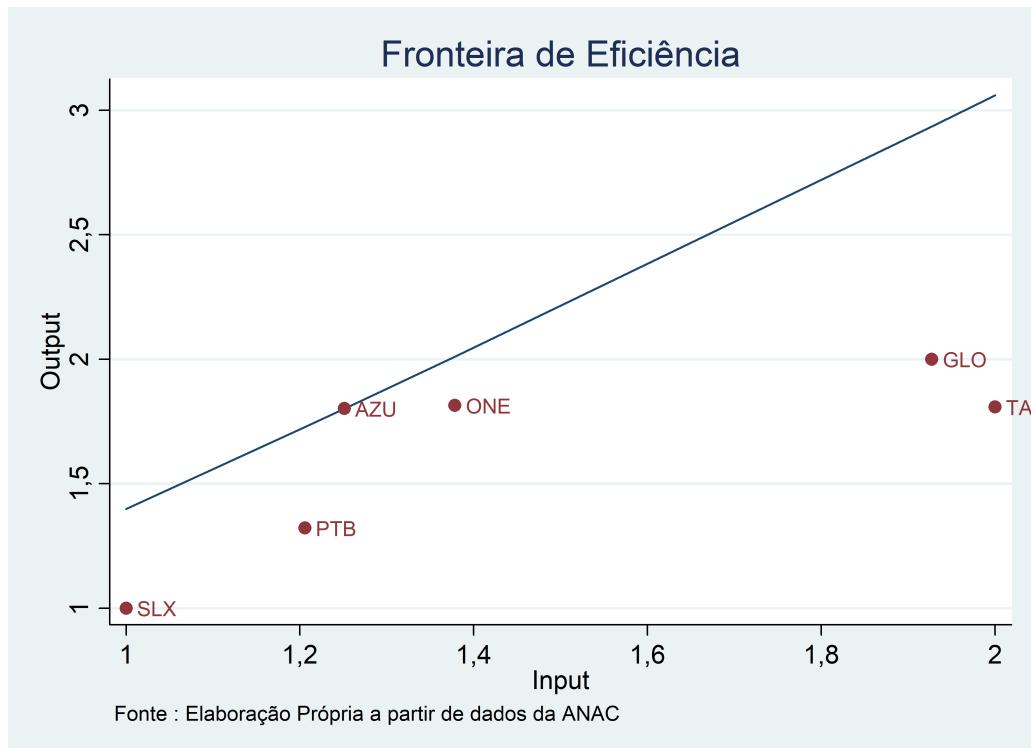


Figura 30: Resultado de DFA - cross section ano de 2011

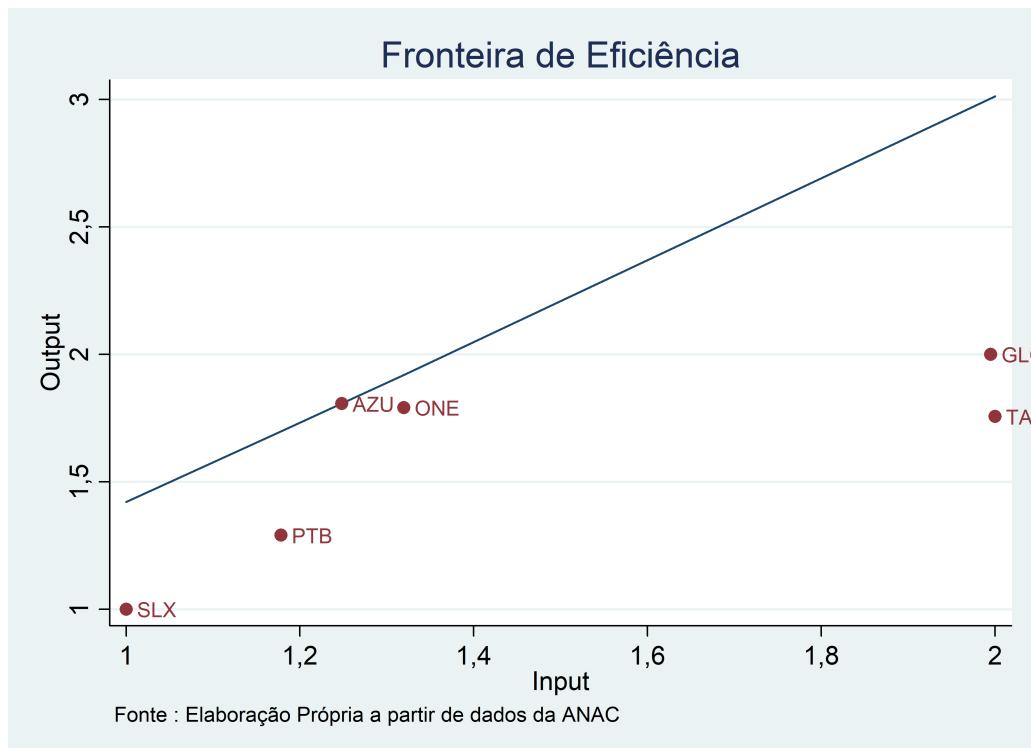


Figura 31: Resultado de DFA - cross section ano de 2012

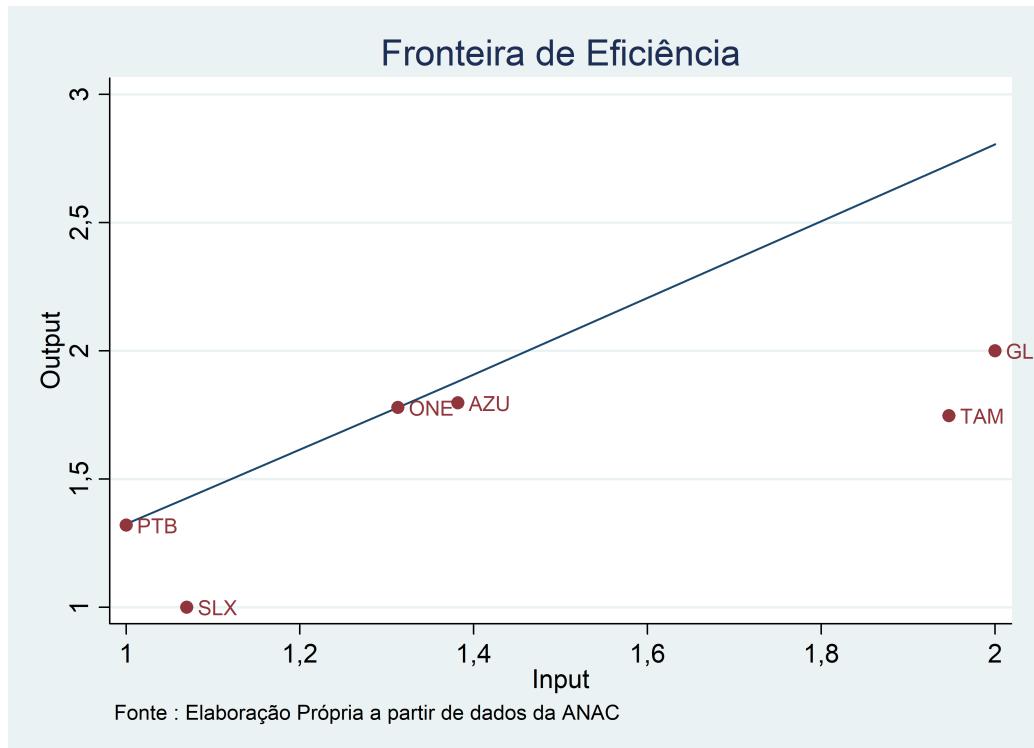


Figura 32: Resultado de DFA - cross section ano de 2013

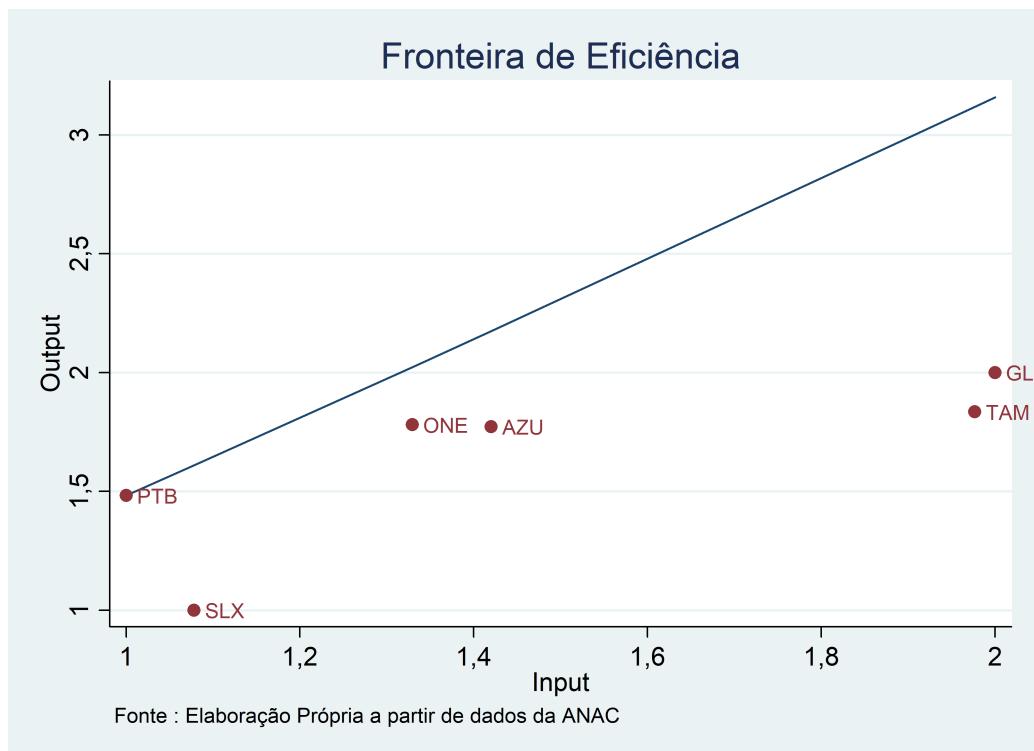
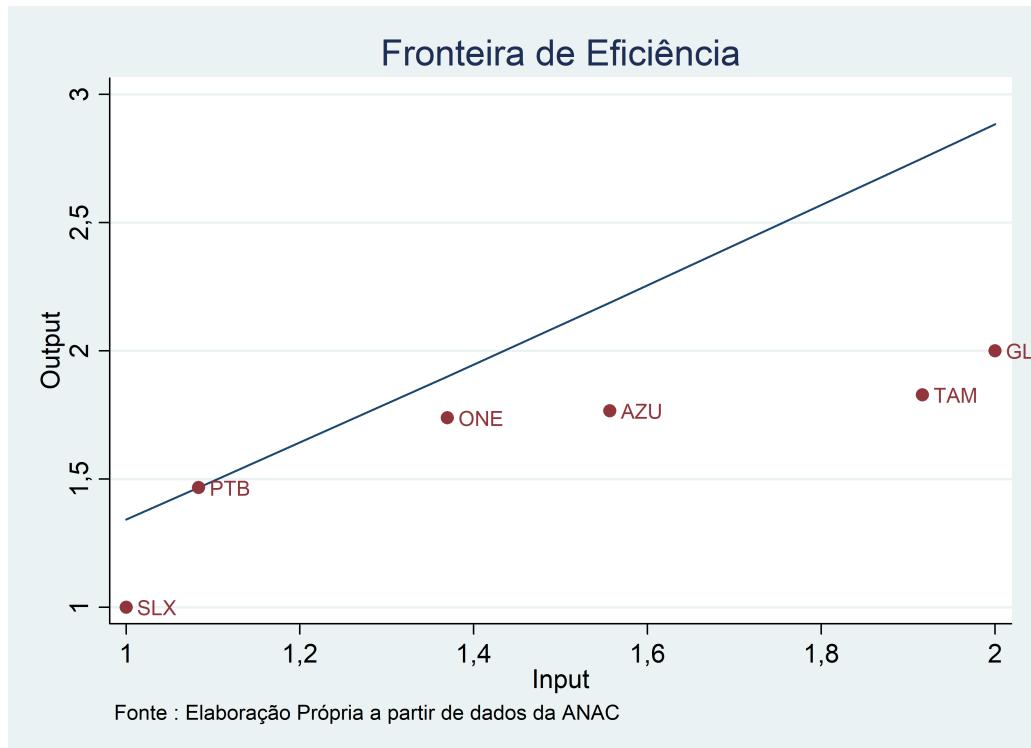


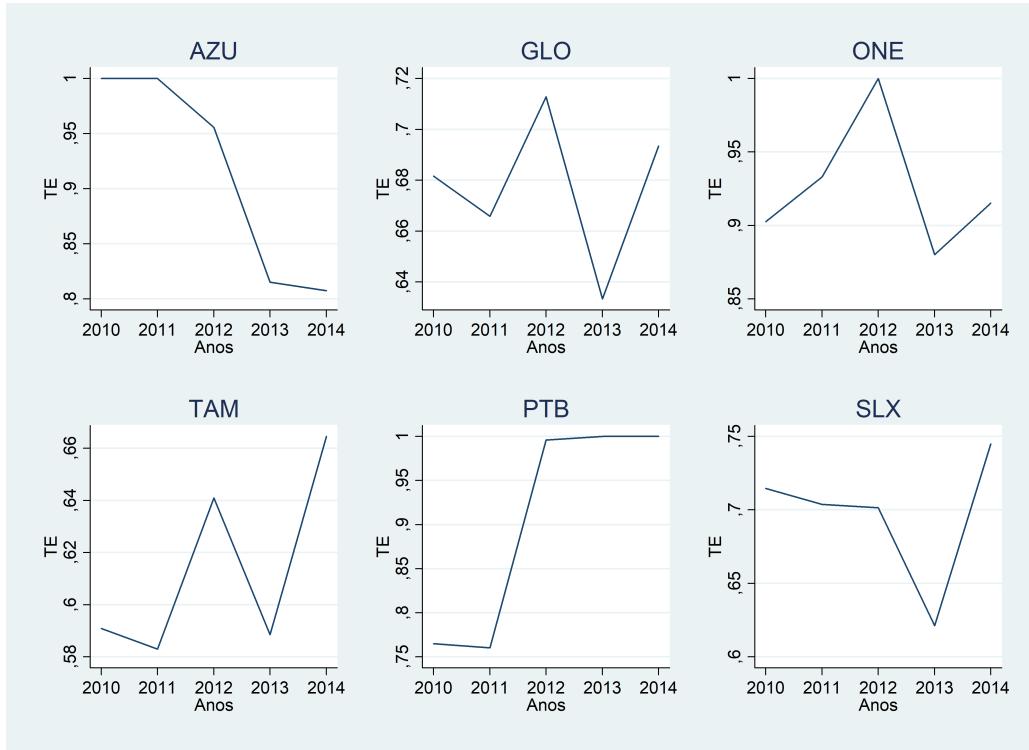
Figura 33: Resultado de DFA - cross section ano de 2014



Dois pontos devem ser percebidos nos gráficos. O primeiro deles é justamente a percepção de uma fronteira de eficiência com aparência linear, embora não seja. Tal fato advém da consequência da normalização dos inputs e outputs, entre os valores de 1 e 2. Isso faz com que a função Cobb-Douglas apresente uma curvatura muito pequena dando a aparência linear. O segundo ponto interessante é justamente o agrupamento do tamanho de cada uma dessas empresas pelo seu tamanho (escala de produção). Os dois pontos de cada tipo de escala de produção estão distribuídos conforme o esperado em termos de magnitude do input, mas em termos de proximidade da fronteira de produção, é possível perceber que a escala da empresa não necessariamente associa-se a uma maior ou menor eficiência alocativa dos recursos. Exemplo deste fato é o caso da empresa Passaredo que apresenta baixa escala de produção mas em consequência apresenta um nível de eficiência bem alto. Além disso, vale ressaltar que as eficiências estão associadas também ao contexto que cada empresa está situada, e por este motivo empresas que sofreram fusões, como no caso das maiores (TAM, GOL) podem apresentar perda de eficiência pode estar associada ao processo de acomodação financeira de tal evento.

Então, pelo gráfico apresentado a seguir pode ser visualizado a evolução dos coeficientes de eficiência (TE) de cada empresa, para os anos em questão:

Figura 34: Evolução da eficiência pelo método DFA - cross section



5.3 Resultados pelo método de SFA

Neste caso, adianta-se que os resultados obtidos não obtiveram evidência capaz de sustentar o método utilizado de forma que os testes apontaram para o resíduo não proveniente do choque exógeno como possível de ser zero, ao não rejeitar tal hipótese.

Adicionalmente, nos anos em que não foi possível rejeitar a hipótese de que o resíduo da regressão via OLS eram iguais a zero, o coeficiente do resíduo que representaria o choque exógeno na função de produção apresentou evidência de ser estatisticamente não significativo, o que se traduz em um modelo claramente determinístico tentando ser estimado via método estocástico. Além do mais o número pequeno de observações da base, para cada ano, pode estar influenciando nesses resultados.

Iniciando-se pelo ano de 2010 e indo para os demais anos temos os seguinte resultados:

Figura 35: Ano de 2010 - Resultados

	Percentiles	Smallest				
1%	-,2132598	-,2132598				
5%	-,2132598	-,1100757				
10%	-,2132598	,0753083	Obs		6	
25%	-,1100757	,441141	Sum of Wgt.		6	
50%	,2582246		Mean		,3518786	
		Largest	Std. Dev.		,5351286	
75%	,7626271	,0753083	Variance		,2863626	
90%	1,155531	,441141	Skewness		,412931	
95%	1,155531	,7626271	Kurtosis		1,753254	
99%	1,155531	1,155531				
Stoc. frontier normal/half-normal model				Number of obs	=	6
Log likelihood = -3,7493233				Wald chi2(1)	=	3032,32
				Prob > chi2	=	0,0000
lnoutput	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lninput1 _cons	1,014512 ,4633996	,0184234 1,160696	55,07 0,40	0,000 0,690	,9784024 -1,811523	1,050621 2,738322
/lnsig2v /lnsig2u	-1,588644 -8,039606	,584798 159,3656	-2,72 -0,05	0,007 0,960	-2,734827 -320,3904	-,4424606 304,3112
sigma_v sigma_u sigma2 lambda	,4518876 ,0179565 ,2045248 ,0397367	,1321315 1,430824 ,1224467 1,457221			,2547651 2,68e-70 -,0354663 -2,816364	,8015321 1,20e+66 ,4445159 2,895837
Likelihood-ratio test of sigma_u=0: chibar2(01) = 0,00 Prob>chibar2 = 1,000						

Figura 36: Ano de 2011 - Resultados

Percentiles		Smallest				
1%	- ,2795921	- ,2795921				
5%	- ,2795921	- ,1264111				
10%	- ,2795921	,0050458	Obs			6
25%	- ,1264111	,4809919	Sum of Wgt.			6
50%	,2430189		Mean			,3589954
		Largest	Std. Dev.			,5898977
75%	,9009119	,0050458				
90%	1,173026	,4809919	Variance			,3479793
95%	1,173026	,9009119	Skewness			,3001809
99%	1,173026	1,173026	Kurtosis			1,512413
Stoc. frontier normal/half-normal model				Number of obs	=	6
Log likelihood = -4,3972965				Wald chi2(1)	=	2435,39
				Prob > chi2	=	0,0000

lnoutput	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lninput1	1,014507	,0205575	49,35	0,000	,9742153 1,054799
_cons	,486186	3,785134	0,13	0,898	-6,93254 7,904912
/lnsig2v	-1,373611	,7259584	-1,89	0,058	-2,796463 ,0492412
/lnsig2u	-6,862175	292,5978	-0,02	0,981	-580,3432 566,6189
sigma_v	,5031809	,1826442			,2470334 1,024926
sigma_u	,0323517	4,733023			9,6e-127 1,1e+123
sigma2	,2542376	,2439227			-,2238421 ,7323173
lambda	,0642944	4,845643			-9,432992 9,561581

Likelihood-ratio test of sigma_u=0: chibar2(01) = 0,00 Prob>=chibar2 = 1,000

Figura 37: Ano de 2012 - Resultados

Percentiles		Smallest		
1%	-27,26731	-27,26731		
5%	-27,26731	-,2843223		
10%	-27,26731	-,0953897	Obs	6
25%	-,2843223	-,0910357	Sum of Wgt.	6
50%	-,0932127		Mean	-4,183872
		Largest	Std. Dev.	11,36238
75%	0	-,0953897		
90%	2,634831	-,0910357	Variance	129,1036
95%	2,634831	0	Skewness	-1,749432
99%	2,634831	2,634831	Kurtosis	4,135173
Stoc. frontier normal/half-normal model			Number of obs	= 6
			Wald chi2(1)	= 305273,71
Log likelihood = -18,869839			Prob > chi2	= 0,0000

lnoutput	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lninput1	,0412642	,0000747	552,52	0,000	,0411179 ,0414106
_cons	,0000167	,0039892	0,00	0,997	-,007802 ,0078354
/lnsig2v	-25,16547	486,0447	-0,05	0,959	-977,7956 927,4646
/lnsig2u	4,838363	,5773503	8,38	0,000	3,706778 5,969949
sigma_v	3,43e-06	,0008337			4,7e-213 2,5e+201
sigma_u	11,23666	3,243745			6,381408 19,786
sigma2	126,2626	72,89773			-16,61437 269,1395
lambda	3275294	3,243745			3275288 3275301

Likelihood-ratio test of sigma_u=0: chibar2(01) = 7,06 Prob>=chibar2 = 0,004

Figura 38: Ano de 2013 - Resultados

	Percentiles	Smallest			
1%	-27,33038	-27,33038			
5%	-27,33038	-,177748			
10%	-27,33038	-,1557709	Obs		6
25%	-,177748	-,1169451	Sum of Wgt.		6
50%	-,136358		Mean		-4,208305
		Largest	Std. Dev.		11,37688
75%	0	-,1557709			
90%	2,531016	-,1169451	Variance		129,4333
95%	2,531016	0	Skewness		-1,752751
99%	2,531016	2,531016	Kurtosis		4,140448
Stoc. frontier normal/half-normal model			Number of obs	=	6
Log likelihood = -18,879044			Wald chi2(1)	=	279486,33
			Prob > chi2	=	0,0000

lnoutput	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lninput1	,0263424	,0000498	528,66	0,000	,0262447 ,02644
_cons	9,86e-06	,0027576	0,00	0,997	-,005395 ,0054147
/lnsig2v	-26,24655	568,2333	-0,05	0,963	-1139,963 1087,47
/lnsig2u	4,841349	,5773264	8,39	0,000	3,70981 5,972888
sigma_v	2,00e-06	,0005677			2,9e-248 1,4e+236
sigma_u	11,25345	3,248456			6,391091 19,81509
sigma2	126,6401	73,11265			-16,6581 269,9382
lambda	5631842	3,248456			5631836 5631848

Likelihood-ratio test of sigma_u=0: chibar2(01) = 7,06 Prob>chibar2 = 0,004

Figura 39: Ano de 2014 - Resultados

	Percentiles	Smallest				
1%	-,1842265	-,1842265				
5%	-,1842265	-,1015565				
10%	-,1842265	0	Obs			6
25%	-,1015565	,3155698	Sum of Wgt.			6
50%	,1577849		Mean			,4031849
		Largest	Std. Dev.			,7130663
75%	,6850548	0	Variance			,5084635
90%	1,704268	,3155698	Skewness			1,104242
95%	1,704268	,6850548	Kurtosis			2,861404
99%	1,704268	1,704268				
Stoc. frontier normal/half-normal model				Number of obs	=	6
Log likelihood = -5,5974804				Wald chi2(1)	=	1639,75
				Prob > chi2	=	0,0000
lnoutput	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lninput1 _cons	1,014597 ,5251402	,0250556 ,9083186	40,49 0,58	0,000 0,563	,9654887 -1,255132	1,063705 2,305412
/lnsig2v /lnsig2u	-,9721753 -8,742207	,5780187 171,1272	-1,68 -0,05	0,093 0,959	-2,105071 -344,1453	,1607205 326,6609
sigma_v sigma_u sigma2 lambda	,6150279 ,0126373 ,378419 ,0205475	,1777488 1,081292 ,2191325 1,10358			,3490516 1,86e-75 -,0510728 -2,142429	1,083677 8,58e+70 ,8079108 2,183524
Likelihood-ratio test of sigma_u=0: chibar2(01) = 0,00 Prob>chibar2 = 1,000						

Os testes descritos na seção que detalha o método do SFA, discorre sobre a necessidade de verificar a assimetria da distribuição dos resíduos (Skewness), estes provenientes do OLS aplicado na equação que descreve a função de produção. Além disso, tem-se o teste proposto por Coelli (1995) também não aprova a aplicação do método como veremos a seguir para cada um dos anos que a base de dados possui. Desta forma, o que se verifica é justamente que o modelo de análise determinístico é melhor para o estudo que aquele realizado sobre a forma estocástica.

Para executar o método, foi utilizado o comando "frontier" disponível no pacote estatístico do Stata. Vale ressaltar que o comando "summarize" ao analisar a Skewness dos resíduos se utiliza do método proposto por Coelli (1995). Os gráficos de fronteira de eficiência apresentados mostram justamente a melhor produção possível para cada firma dado a quantidade de insumo utilizada.

6 Conclusões

Através do estudo apresentado, podemos concluir que dentre as empresas selecionadas e do contexto de cada uma diante da conjuntura econômica do país, a evolução da eficiência dos recursos utilizados para produzir o produto final tal como descrito no modelo foram justificadas de maneira coerente pelos modelos de DEA e DFA. Entretanto, devido pela impossibilidade de aplicação do método de SFA, os resultados se mantiveram sob a ótica determinística sem possibilidade de extensão para o caso estocástico. Além disso, vale ressaltar que o estudo ficou relativamente restrito, devido ao fato de que o tamanho da base de dados utilizado foi pequena demais apresentando problemas de estimativa, em alguns casos, e também problemas ao construir as fronteiras.

Sobre as empresas, foi possível notar que tanto pelo DEA quanto pelo DFA foi obtiveram-se alguns pontos interessantes sobre a evolução da eficiência. Primeiramente é possível concluir que dos anos de 2010 para 2014 a companhia aérea Azul perdeu eficiência relativa as outras empresas enquanto que a empresa Passaredo ganhou sobremaneira níveis maiores de eficiência. Para as outras companhias ficou claro que devido a problemas estruturais internos, seja por fusões recentes ou compra e venda de aeronaves, houve quedas parciais de eficiência em especial para os anos de 2011 e 2013, mas de forma que após estes anos as eficiências retomaram níveis iguais ou maiores que os anteriores ao final de 2014.

7 Apêndice : Econometria Utilizada

Esta seção procura estabelecer os métodos que foram utilizados para realizar o estudo de eficiência e de fronteira de produção para as companhias aéreas brasileiras. Trata-se portanto de um resumo dos métodos de envoltória, em especial de três principais: DEA, DFA e o SFA. Suas descrições encontram-se dispostas nas seções abaixo.

7.1 DEA: Data Envelopment Analysis

A metodologia DEA foi primeiramente introduzida por Charnes, Cooper e Rhodes em 1978, e posteriormente aprofundada e extendida por Banker, Charnes e Cooper em 1984. Trata-se de uma metodologia que tem como objetivo estimar uma medida de eficiência, a Technical Efficiency (TE) para cada uma das Decisions Making Units (DMU). O cálculo por trás do método é realizado via programação dinâmica, que consegue promover uma medida única de eficiência. A programação dinâmica se faz pela realização do cálculo com inputs e outputs gerando uma fronteira eficiente, que representaria a melhor forma de se produzir. A partir desta fronteira é então calculada para cada DMU a eficiência relativa. Desta maneira, para cada DMU o método DEA informa um valor de eficiência (efficiency score), que geralmente varia no intervalo de 0 a 1, mostrando quando existem ou não ineficiências na produção. Logo o propósito deste método é o de apresentar quando existem ineficiências e informar as companhias ou indústrias possibilidades de aumentar sua produtividade e consequentemente sua eficiência.

Existem 4 modelos principais para realizar o método DEA, pois existem hipóteses diferentes sobre os retornos de escala. Todos os modelos calculam o valor de eficiência (efficiency score) para cada DMU j , de acordo com a hipótese adotada sobre os retornos de escala. Seja θ o efficiency score de determinada DMU j . Seja y_{rk} a variável que representa o output r ($r = 1, \dots, R$) e x_{ik} a avriável que representa o input i ($i = 1, \dots, I$) para cada DMU k . Seja λ_k os pesos das observações utilizadas. Desta maneira enumeremos os modelos nas subseções abaixo.

7.1.1 Modelo 1 : Constant Return of Scale (CRS)

O problema a ser resolvido é :

$$\theta_j^{CRS} = \max \theta$$

sujeito a:

$$\sum_{k=1}^N \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1, \dots, R$$

$$\sum_{k=1}^N \lambda_k x_{rk} \leq \theta x_{rj}, \forall i = 1, \dots, I$$

$$\lambda_k \geq 0, \forall k = 1, \dots, N$$

7.1.2 Modelo 2 : Variable Return of Scale (VRS)

O problema a ser resolvido é :

$$\theta_j^{VRS} = \max \theta$$

sujeito a:

$$\sum_{k=1}^N \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1, \dots, R$$

$$\sum_{k=1}^N \lambda_k x_{rk} \leq \theta x_{rj}, \forall i = 1, \dots, I$$

$$\sum_{k=1}^N \lambda_k = 1, \lambda_k \geq 0, \forall k = 1, \dots, N$$

7.1.3 Modelo 3 : Non-Increasing Return of Scale (NIRS)

O problema a ser resolvido é :

$$\theta_j^{NIRS} = \max \theta$$

sujeito a:

$$\sum_{k=1}^N \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1, \dots, R$$

$$\sum_{k=1}^N \lambda_k x_{rk} \leq \theta x_{rj}, \forall i = 1, \dots, I$$

$$\sum_{k=1}^N \lambda_k \leq 1, \lambda_k \geq 0, \forall k = 1, \dots, N$$

7.1.4 Modelo 4 : Non-Decreasing Return of Scale (NDRS)

O problema a ser resolvido é :

$$\theta_j^{NDRS} = \max \theta$$

sujeito a:

$$\sum_{k=1}^N \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1, \dots, R$$

$$\sum_{k=1}^N \lambda_k x_{rk} \leq \theta x_{rj}, \forall i = 1, \dots, I$$

$$\sum_{k=1}^N \lambda_k \geq 1, \lambda_k \geq 0, \forall k = 1, \dots, N$$

7.2 Stochastic Frontier Analysis (SFA) e Deterministic Frontier Analysis (DFA)

Os modelos de SFA foram criados de maneira simultânea por Aigner, Lovell e Schmidt (ALS) (1977) e por Meeusen e Van der Broeck (MB) (1977). Tais modelos apresentam, resumidamente, dois principais objetivos: o primeiro de medir a eficiência técnica da produção (technical efficiency, TE) e segundo de proporcionar uma análise capaz de introduzir choques exógenos na função de produção, responsáveis por alterar o output.

Iniciando pela forma determinística, vamos assumir que a função de produção $f(x_i; \beta)$ seja tal que represente uma função de produção, normalmente, descrita como uma Cobb-Douglas. A fronteira de produção determinística pode ser descrita como:

$$y_i = f(x_i; \beta) \cdot TE_i$$

Na forma determinística não existem choques exógenos que modifiquem o output gerado pela firma. Nesta equação teremos que $TE_i = \exp(-u_i)$ onde $u_i \geq 0$ para que tenhamos sempre $TE_i \leq 1$. Para estimar a equação acima existem diversos métodos, passando tanto por métodos de programação, assim como o método comum de OLS. Geralmente ao se deparar com erros não positivos, torna-se interessante para viabilizar a aplicação do DFA, a aplicação do Corrected Ordinary Least Squares (COLS), em que se padronizam os erros originados pelo OLS, tal como descrito na equação abaixo, modificando a estimativa do intercepto também:

$$-\hat{u}_i^* = \hat{u}_i - \max(\hat{u}_i)$$

Existem outras formas de corrigir o erro e deixá-lo pronto para a aplicação do DFA, como o Modified Ordinary Least Squares (MOLS), que se utiliza, ao invés do máximo, do valor esperado dos erros para fazer tal correção. Entretanto neste trabalho optou-se pelo método de COLS.

O modelo determinístico serve de base para a aplicação do modelo estocástico. A diferença entre ambos reside no choque exógeno de forma que a equação se transforma em :

$$y_i = f(x_i; \beta) \cdot \exp(\varepsilon_i)$$

Na equação acima, teremos tal como na forma determinística $TE_i = \exp(-u_i) \leq 1$, onde temos que $\varepsilon = v_i - u_i$, sendo que v_i representa o choque exógeno comentando anteriormente. Para a aplicação do método devem-se aplicar algumas hipóteses sobre as distribuições destes erros. A mais comum e que será utilizada neste trabalho reside na caracterização do modelo como Normal e Half-Normal, descrito a seguir:

- $v_i \sim iidN(0, \sigma_v^2)$
- $u_i \sim iidN^+(0, \sigma_u^2)$
- v_i e u_i são distribuídos independentemente entre eles e entre os regressores.

A partir destas hipóteses, serão feitos os cálculos necessários para obter a eficiência técnica desejada. Primeiramente se obtém as densidades marginais descritas a seguir:

$$f(\varepsilon) = \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon}{\sigma}\right) \cdot \Phi\left(\frac{-\varepsilon\lambda}{\sigma}\right)$$

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp\left(\frac{-v^2}{2\sigma_v^2}\right)$$

onde temos que $\sigma = \sigma_u^2 + \sigma_v^2$ e $\lambda = \frac{\sigma_u}{\sigma_v}$. Desta maneira seguindo as hipóteses apresentadas anteriormente conseguimos obter os seguintes resultados abaixo:

$$E(\varepsilon) = -E(u) = -\sigma_u \cdot \sqrt{\frac{2}{\pi}}$$

$$V(\varepsilon) = \frac{\pi - 2}{\pi} \cdot \sigma_u^2 + \sigma_v^2$$

Pela sugestão em ALS, verifica-se que uma maneira de se medir a média da eficiência técnica de todos os produtores é dada por $[1 - E(u)]$. Porém, Lee e Tyler (1978) propõem que a mesma medida deveria ser dada por:

$$E(\exp(-u)) = 2 \cdot [1 - \Phi(\sigma_u)] \cdot \exp\left(\frac{\sigma_u^2}{2}\right)$$

Esta medida apresenta maior consistência em termos de definição ao que foi apresentando anteriormente. A estimativa pode ser feita pelo método de verossimilhança de forma que se obtenham todos os parâmetros desejados. Entretanto encontramos um problema que reside na possibilidade de ε_i conter informações em u_i , assim uma maneira de resolver essa situação seria realizar a distribuição de u_i condicional a ε_i . Logo:

$$f(u|\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp\left(\frac{-(u - \mu_*)^2}{2\sigma_*^2}\right) / [1 - \Phi(-\frac{\mu_*}{\sigma_*})]$$

onde temos que $\mu_* = -\varepsilon \cdot \frac{\sigma_u^2}{\sigma^2}$ e que $\sigma_*^2 = \sigma_u^2 \cdot \frac{\sigma_v^2}{\sigma^2}$. Por este motivo, temos que a distribuição de $f(u|\varepsilon)$ será dada por :

$$f(u|\varepsilon) \sim N^+(\mu_*, \sigma_*^2)$$

assim como a sua média e moda podem ser descritas como:

$$E(u_i|\varepsilon_i) = \mu_{*i} + \sigma_* \left[\frac{\Phi(-\mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)} \right]$$

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \cdot \frac{\sigma_v^2}{\sigma^2}, & \text{se } \varepsilon_i \leq 0 \\ 0, & \text{caso contrário.} \end{cases}$$

contudo, a média é mais utilizada que a moda, embora a moda tenha uma interpretação como estimador de máxima verossimilhança. Após a estimativa de u_i ter sido obtida, teremos que a eficiência técnica de cada produtor poderá ser obtida da seguinte maneira:

$$TE_i = \exp(-\hat{u}_i)$$

em que \hat{u}_i pode ser tanto a média como a moda.

Outra forma apresentada por Battese e Coelli (1988) reside num estimador alternativo para a eficiência técnica dado por :

$$TE_i = E(\exp(-u_i)|\varepsilon_i) = \left[\frac{\Phi(\sigma_* - \mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)} \right] \exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right)$$

A partir destes resultados torna-se possível criar intervalos de confiança para as estimativas da eficiência técnica, tal como aqueles obtidos por Horrace e Schmidt (1995,1996). Para uma confiança de $(1 - \alpha)100\%$ tem-se o intervalo (L_i, U_i) para $(\exp(-u_i)|\varepsilon_i)$, em que temos:

$$L_i = \exp(-\mu_{*i} - z_L \sigma_*)$$

$$U_i = \exp(-\mu_{*i} - z_U \sigma_*)$$

de tal forma que Z é distribuído na normal padrão e em que :

$$z_L = \Phi^{-1}\left([1 - \frac{\alpha}{2}][1 - \Phi(\frac{-\mu_{*i}}{\sigma_*})]\right)$$

$$z_U = \Phi^{-1}\left([1 - (1 - \frac{\alpha}{2})][1 - \Phi(\frac{-\mu_{*i}}{\sigma_*})]\right)$$

Como foi dito anteriormente este trabalho se utilizará do modelo com hipóteses sobre a distribuição dos erros serem normais. Existem outras hipóteses como a distribuição Normal e Exponencial, a Normal e a Normal Truncada entre outras, mas todas elas apresentam similaridades na realização das contas, procurando a distribuição condicional, estimando o modelo e calculando a eficiência técnica. Um ponto importante para realizar a aplicação deste método, diferentemente do DFA, que pelo COLS já corrige os erros de maneira a garantir a aplicação da técnica, é que o SFA necessita de alguns testes preliminares de maneira a verificar se este é coerente ou não. O argumento dos testes residem na seguinte análise, feita através de um simples OLS: Se $u_i = 0$ então $\varepsilon_i = v_i$ e o termo de erro será simétrico, fazendo com que os dados não suportem a aplicação do método. Entretanto se $u_i \geq 0$, então teremos que $\varepsilon = v_i - u_i$ seja negativamente viésado e desta maneira torna-se possível a aplicação do método. Schmidt e Lin (1984) propuseram o teste estatístico abaixo:

$$(b_1)^{\frac{1}{2}} = \frac{m_3}{(m_2)^{\frac{3}{2}}}$$

onde m_2 e m_3 são justamente o segundo e o terceiro momentos dos resíduos gerados pelo método de OLS. Através de Coelli (1995) propõem-se um método alternativo de realizar o teste, distribuído assintoticamente em uma Normal padrão, baseado na seguinte equação:

$$b_1 = \frac{m_3}{(6m_2^3/I)^{\frac{1}{2}}}$$

Embora sejam testes muito bem estruturados, existe um problema em termos de tamanho da base de dados, inclusive uma crítica bem comum em trabalhos econôméticos, que reside na suposição de distribuição assintótica. A análise de eficiência, de maneira frequente, termina por possuir uma base de dados pequenos que promovem regressões com problemas de fit e também com problemas de credibilidade dos testes aplicados.

Excluindo-se estes problemas, torna-se interessante enumerar os procedimentos para verificar a possibilidade de aplicação do método em questão: (i) executar a regressão, verificar significância dos coeficientes e o R^2 (ii) verificar a Skewness e a sua probabilidade associada (iii) aplicar o teste de Coelli (1995). Ressalta-se que neste trabalho procurou-se garantir resultados que estivessem com pelo menos um $\alpha = 5\%$.

8 Referências

ANAC - Agência Nacional de Aviação Civil (<http://www.anac.gov.br/>)

- AIGNER, Dennis; LOVELL, CA Knox; SCHMIDT, Peter. Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, v.6, n.1, p.21-37, 1977.
- BANKER, Rajiv D.; CHARNES, Abraham; COOPER, William Wager. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, v. 30, n. 9, p. 1078-1092, 1984.
- BATTESE, George Edward; COELLI, Tim J. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, v. 20, n. 2, p. 325-332, 1995.
- BATTESE, George Edward; COELLI, Prediction of firm-level technical efficiencies with a generalized frontier production function and painel data. *Journal of Econometrics*, v. 38, n. 3, p. 387-399, 1988.
- COELLI, Tim J. A guide to FRONTIER version 4.1: a computer program for stochastic frontier production and cost function estimation. *CEPA Working papers*, 1996.
- COELLI, Tim; PERELMAN, Sergio; ROMANO, Elliot. Accounting for environmental influences in stochastic frontier models: with application to international airlines. *Journal of productivity analysis*, v. 11, n. 3, p. 251-273, 1999.
- COELLI, T. J. A Monte Carlo analysis of the stochastic frontier production function. *Journal of Productivity Analysis*, v. 6, p. 247-268, 1995.
- CHARNES, Abraham; COOPER, William W.; RHODES, Edwardo. Measuring the efficiency of decision making units. *European journal of operational research*, v. 2, n. 6, p. 429-444, 1978.
- HORRACE, William C.; SCHMIDT, Peter; Witte, Ann Dryden. Sampling errors and confidence intervals for order Statistics: Implementing the Familiy Support Act. National Bureau of Economics Research, 1995.
- LEE, Lung- Fei; TYLER, William G. The stochastic frontier production function and average efficiency: An empirical analysis. *Journal of Econometrics*, v.7, n.3, p.385-389, 1978.
- MEEUSEN, Wim. VAN DEN BROECK, Julien. Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, p.435-444, 1977.
- KUMBHAKAR, Subal C.; LOVELL, CA Knox. Stochastic frontier analysis. Cambridge university press, 2003.
- SCHMIDT, Peter; LIN, Tsai-Fen. Simple tests of alternative specificationsin stochastic frontier models. *Journal of Econometrics*, v.24, n.3, p.349-361, 1984.