

NOTA TÉCNICA ARSAE

CRFEF 40/2016

Custos Operacionais Eficientes – Metodologia para a Revisão
Tarifária Periódica da Companhia de Saneamento de Minas Gerais
- COPASA

Consultores:

Ana Lúcia Miranda Lopes, Dr.

Marcelo Costa, Dr.

Bruno Vilela, doutorando CEPEAD/UFMG

Ricardo Santos, mestrando CEPEAD/UFMG

Neusa de Paula Antunes

Jan Morse Teixeira Koole

Janeiro 2017

INTRODUÇÃO

Este relatório faz uma análise da proposta da ARSAE-CRFEF/MG constante na Nota Técnica no. 40/2016 e que trata do cálculo dos custos operacionais eficientes da COPASA/MG, assim como propõe método para cálculo da mudança tecnológica ocorrida nos últimos anos.

Importante iniciar parabenizando o regulador mineiro pela decisão de mudar a metodologia de cálculo dos custos operacionais eficientes. A Fronteira Estocástica tem suas vantagens, mas também suas desvantagens das quais se pode ressaltar: a necessidade de se estabelecer, de antemão, uma forma funcional para a fronteira de produção, decisão de extrema importância e que pode levar a erros de cálculo dos escores de eficiência se a mesma não for a correta; dificuldade para as companhias reguladas e demais agentes do setor de entendimento de tão complexa metodologia. Por outro lado, *Data Envelopment Analysis* – DEA constrói uma fronteira sem a necessidade de estabelecimento prévio da forma funcional e a constrói utilizando os dados das próprias empresas. O entendimento é claro e muitos softwares de DEA estão disponíveis no mercado. Além disto, DEA permite que sejam calculados os escores de eficiência técnica, em um único modelo, empresas operando em diferentes retornos de escala e que produzem múltiplos produtos, utilizando múltiplos insumos.

Pode-se afirmar, também, que a ARSAE acompanha outros reguladores de abastecimento de água e esgotamento sanitário, assim como de energia elétrica na utilização cada vez maior de DEA na regulação de empresas.

MEDIÇÃO DE EFICIÊNCIA EM PRESTADORES DE SERVIÇO DE ABASTECIMENTO DE ÁGUA E ESGOTAMENTO SANITÁRIO

A ARSAE-MG irá calcular a distância da COPASA MG à fronteira eficiente por meio de um modelo DEA (*Data Envelopment Analysis*) orientado a insumo, corrigido para o viés, conforme Nota Técnica ARSAE-MG 40/2016.

Segundo o Regulador, primeiro será calculado o nível de custos eficientes por meio da metodologia Data Envelopment Analysis – DEA com orientação à redução de insumos e a meta de eficiência. As informações utilizadas virão do banco de dados SNIS. O modelo utilizado, de acordo com o código R disponibilizado pelo regulador, será o de retornos crescentes à escala. Neste modelo, a variável dependente será a despesa de exploração menos os impostos envolvidos na operação. Os produtos identificados como relevantes para a utilização no modelo DEA são:

- Total de economias ativas de água (nº);
- Total de economias ativas de esgoto (nº);
- Volume de esgoto tratado (m³)
- Perdas de água (m³)

Segundo a NT 40/2016 o produto perdas de água é considerado como produto indesejável no modelo DEA e, por este motivo, é utilizado como insumo.

Os dados serão a média dos anos de 2012 a 2014, mas a ARSAE salienta que, se na época da revisão tarifária da COPASA os dados de 2015 estiverem disponíveis, estes serão incluídos.

As empresas utilizadas no modelo DEA e contra a qual a COPASA será comparada são os prestadores nacionais constantes na base de dados do SNIS. Segundo a ARSAE, para compor a base do SNIS foram escolhidos

prestadores regionais de todos os estados, a partir de dados agregados por prestador (item 3.3.1, pág. 9). Com a base de dados construída, a ARSAE decidiu pela aplicação dos seguintes filtros (item 3.3.2, pág. 9):

- a) Seleção de prestadores que ofertam os dois tipos de serviço (abastecimento de água e esgotamento sanitário);
- b) Escala mínima de serviço (número de economias de água) de pelo menos 10% à da Copasa;
- c) Área de atuação que compreenda ao menos dois municípios, com o objetivo de excluir prestadores locais;
- d) Exclusão de autarquias;
- e) Nível mínimo de hidrometração = 70%.

Após a aplicação destes filtros a base de dados fica composta por 15 empresas de saneamento, a saber, AGESPISA, CAERN, CAESB, CAGECE, CAGEPA, CASAN, CESAN, COMPESA, COPASA, CORSAN, DESO, EMBASA, SABESP, SANEAGO e SANEPAR

A ARSAE propõe a aplicação do algoritmo de Simar e Wilson (2014) para correção dos escores de DEA, que, segundo ela, são viesados para cima. Para a identificação do fator de *catch-up* que será utilizado para a definição da receita requerida o escore de eficiência da COPASA, corrigido pelo viés, será dividido pelo maior escore, conforme equação abaixo.

Fator " catch - up " = 1- Índice de eficiência Copasa/ Maior índice de eficiência

CONSIDERAÇÕES SOBRE A PROPOSTA DA ARSAE PARA O MODELO DEA

Como já foi mencionando na introdução, a metodologia DEA é, realmente, a mais adequada e sua aplicação em modelos regulatórios vem crescendo rapidamente. Quanto aos insumos e produtos escolhidos pode-se afirmar o mesmo. Como o objetivo do regulador é o de identificar ineficiências nos gastos das concessionárias das empresas de serviço de abastecimento de água e esgotamento sanitário é muito comum que as variáveis propostas (economias ativas de água e esgoto e volume de esgoto tratado) sejam utilizadas. É, também, importante que a qualidade do abastecimento de água seja considerado por meio da variável perdas.

Quanto aos dados e aos filtros propostos consideramos que são adequados, porém alertamos a ARSAE que o código R, fornecido pela mesma, não gera a lista de empresas que deveria gerar para a realização do *benchmarking* da COPASA com a aplicação dos filtros propostos. Este item (erros no código R da ARSAE) será detalhadamente discutido posteriormente neste relatório.

Quanto ao cálculo do fator de *catch-up*, discordamos do regulador quando divide o escore de eficiência da COPASA pelo maior valor de escore identificado pelo modelo. Modelos modernos de regulação buscam gerar incentivos às empresas mais eficientes do setor e por este motivo utilizam uma normalização em que dividem o escore DEA das empresas pela média dos escores de todas elas, utilizando às vezes um corte. Este é o caso dos modelos DEA utilizados pela Agência Nacional de Energia Elétrica – ANEEL para a avaliação da eficiência das empresas de distribuição e transmissão de energia elétrica. Em todos os modelos utilizados pela ANEEL desde 2007, quando aplicou DEA no setor de transmissão de energia elétrica, a mesma utiliza a normalização como uma maneira de corrigir escores muito baixos e gerar incentivos para as empresas eficientes. Sendo assim, a COPASA alcançaria um escore acima de 100%, de acordo com as simulações realizadas. De acordo com a NT 40/2016 ganhos serão divididos com o

consumidor. Este mecanismo de normalização geraria incentivos à COPASA na busca de utilizar de forma cada vez mais eficiente os recursos, enquanto não penaliza demais o consumidor.

CÁLCULO DA VARIÁVEL *PERDAS DE ÁGUA*

A ARSAE-MG propõe utilizar o volume de perdas de água como variável indicadora de qualidade no modelo DEA. Para tanto, apura este volume conforme tabela a seguir:

Tabela 4 – Informações utilizadas para compor a variável Perdas

Informação	Unidade	Forma de cálculo	Informações primárias utilizadas
Perdas	1.000 m ³ /ano	((AG006 + AG018 - AG010))	AG006: Volume de água produzido AG010: Volume de água consumido AG018: Volume de água tratada importado

No entanto, para fechar o balanço de água é necessário subtrair também o volume de água exportado, para os prestadores em que ocorre.

Sugere-se utilizar no cálculo do volume de perdas de água a seguinte fórmula:

$$\text{Perdas} = \text{AG006} + \text{AG018} - \text{AG010} - \text{AG019}$$

Onde:

- AG006: Volume de água produzido (1000 m³/ano);
- AG010: Volume de água consumido 1000 m³/ano);
- AG018: Volume de água tratada importado (1000 m³/ano);
- AG019: Volume de água tratada exportado (1000 m³/ano).

Para tanto deve-se incluir a variável AG019 - Volume de água tratada exportado (1.000 m³/ano) do SNIS no banco de dados e modificar o código R para fazer o cálculo da maneira sugerida.

DESPESAS NÃO ADMINISTRÁVEIS (VPA)

A Lei Estadual 18.309/2009 em seu artigo 8º separa as despesas das concessionárias de saneamento em duas parcelas: Não administráveis (VPA) e Administráveis (VPB). As Despesas não Administráveis são o somatório das despesas com energia elétrica, material de tratamento, telecomunicação, combustíveis, lubrificantes, impostos e taxas.

LEI 18309, DE 03/08/2009 ASSEMBLEIA DE MINAS

*Art. 8º O reajuste e a revisão das tarifas cobradas pelos prestadores sujeitos à regulação e à fiscalização da ARSAE-MG serão autorizados mediante resolução da ARSAE-MG e **objetivarão assegurar o equilíbrio econômico-financeiro do ajuste e a modicidade** e o controle social das tarifas, observada, em todos os casos, a publicidade dos novos valores.*

*§ 1º Na composição dos valores de **reajuste e de revisão** das tarifas será garantida a geração de recursos para:*

- I. a realização dos investimentos;*
- II. a recuperação dos custos da **prestação eficiente do serviço**, entendendo-se como tais:
 - a. as despesas administráveis com mão de obra, materiais, serviços de terceiros e provisões;*
 - b. as despesas não administráveis com energia elétrica, material de tratamento, telecomunicação, combustíveis, lubrificantes, impostos e taxas;*
 - b) as quotas de depreciação e amortização;**
- III. a remuneração do capital investido pelos prestadores de serviços.[...]*

No entanto, a Agência está tratando essas despesas nessa 1ª Etapa da Revisão Tarifária Periódica de 2016 da COPASA, como se fossem gerenciáveis quando, além de incluí-las indevidamente nos estudos de análise comparativa, aplicou sobre as mesmas um fator de eficiência.

A Lei Estadual define que as despesas listadas como não administráveis sejam tratadas como tal, ou seja, repassadas aos usuários, visto que não há gerenciamento sobre seus custos. Entretanto, há alegações de que se podem gerenciar as quantidades utilizadas, o que não corresponde à realidade.

Em relação à energia elétrica, a maior das despesas não administráveis, o crescimento da demanda de água gera perda de escala no consumo da mesma (kWh/m³), pois esse crescimento exige a adição de mais bombas operando em paralelo nas elevatórias existentes e/ou a implantação de novas

elevatórias no sistema de distribuição. Isto pode acontecer devido ao alcance do limite de tempo de operação ou do limite da capacidade de bombeamento. Nestes casos, o consumo de energia aumenta mais do que o volume bombeado, sendo este acréscimo consequência do aumento da intensidade energética (kWh/m³) causada pela adição de equipamentos de bombeamento ao sistema.

Além disso, o aumento de consumo de energia elétrica devido ao aumento da intensidade energética é provocado em circunstâncias em que é necessária a redundância de fontes de abastecimento. Neste caso é natural que se opere primeiro aquelas fontes primárias de menor intensidade energética e, ocorrendo restrições hídricas, sejam operadas as fontes redundantes que, em geral são de maior intensidade energética. Deve ser ressaltado aqui que a demanda de água permanece a mesma enquanto o consumo de energia para atendê-la aumenta. Um exemplo recente desta situação é o da crise hídrica ocorrida nos últimos anos na Região Metropolitana de Belo Horizonte que provocou a necessidade de implantação de uma captação redundante no sistema Paraopeba de muito maior intensidade energética do que a fonte original do Rio Manso.

O aumento de intensidade energética acontece, ainda, em sistemas produtores onde, por restrições hídricas sazonais ou excepcionais, ocorre o aumento da altura manométrica necessária para o bombeamento. Nestes casos a intensidade energética é aumentada devido à maior potência requerida para se fazer o mesmo bombeamento, chegando ao caso de se necessitar o acréscimo de mais bombas em série no sistema. Isto é muito comum em captações superficiais, quando sazonalmente é necessário acrescentar balsas com bombas para ajudar o recalque da água até o sistema.

Assim, devido aos fatores expostos, pode ser observado no Gráfico abaixo que há uma tendência de crescimento da intensidade energética ao longo do tempo, sendo esta uma das causas do aumento da despesa com energia adicional provocando aumento de custos.

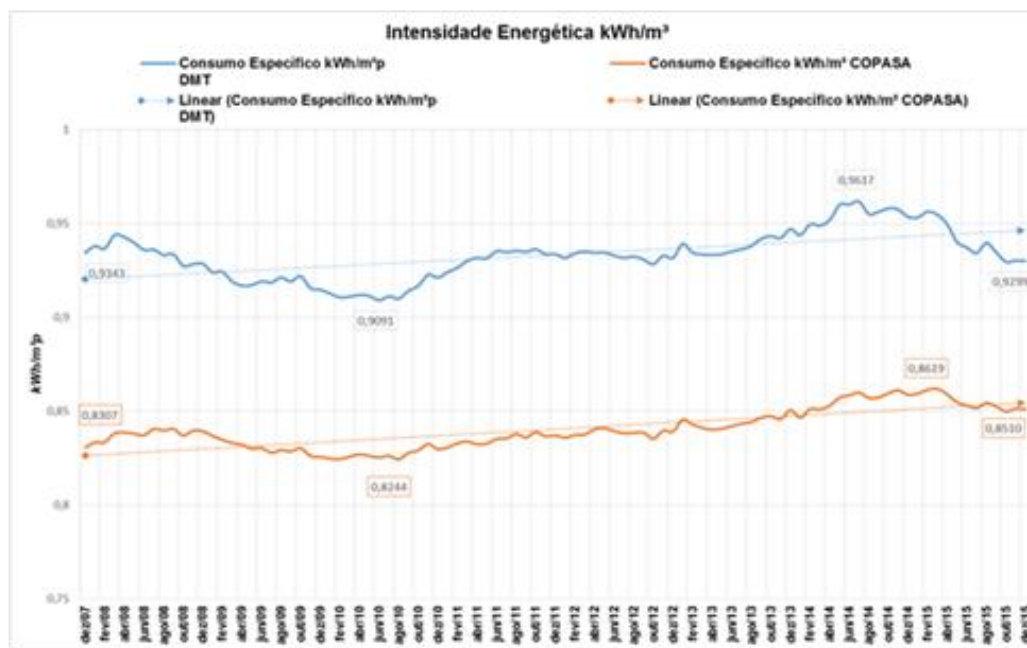


Gráfico 13 – Intensidade Energética (kWh/m³)

Do mesmo modo que energia elétrica, os gastos com o item material de tratamento também são influenciados por fatores sobre os quais não há controle por parte da Empresa, pois a aplicação de produtos químicos no processo de tratamento de água é feita de acordo com a composição físico-química e microbiológica da água bruta a ser tratada.

Nesse sentido, de acordo com as características da água, variam as doses e o número de produtos aplicados para a obtenção de água dentro do padrão de potabilidade.

A qualidade da água bruta apresenta variação distinta nos períodos seco e chuvoso, sendo este último o que demanda maior consumo de produtos, principalmente coagulantes, alcalinizantes e oxidantes. Entretanto, mesmo em condições de tempo seco, pode haver demanda de um número maior de produtos, para fazer frente às alterações físico-químicas, hidrobiológicas e microbiológicas da água.

Ao contrário do que possa parecer, o processo de tratamento pode enfrentar dificuldades maiores no período seco, comparado ao período chuvoso, em função da proliferação de microrganismos com algas, cianobactérias e

concentração de sais dissolvidos. Nessas condições, pode ser necessária a utilização de vários produtos, cada um com uma finalidade específica na remoção dos componentes indesejáveis da água. Observa-se, portanto, que não apenas os preços, mas também a quantidade consumida dos itens listados no VPA não são administráveis.

Pelo exposto, sugere-se excluir as despesas de Parcela A dos estudos de análise comparativa e não aplicar nenhuma eventual glosa advinda desses estudos sobre as despesas consideradas não administráveis (VPA).

Em termos do modelo DEA proposto, esta alteração significa que ao invés do Custo de Exploração retirando apenas os Impostos, seria utilizado o Custo de Exploração sem Impostos, Energia Elétrica e Material de Tratamento. Utilizando a nomenclatura do SNIS, a variável seria, pois calculada da seguinte maneira:

$$DEX = FN010 + FN014 + FN020 + FN027 + FN039$$

Onde:

FN010: Despesa com pessoal próprio (R\$/ano)

FN014: Despesa com serviços de terceiros (R\$/ano)

FN020: Despesa com água importada (bruta ou tratada) (R\$/ano)

FN027: Outras despesas de exploração (R\$/ano)

FN039: Despesa com esgoto exportado (R\$/ano)

Os valores de pessoal próprio (FN010) e serviços de terceiros (FN014) são normalizados pelo custo de vida em cada região, conforme proposto pela ARSAE.

REPLICAÇÃO E ANÁLISE DA PROPOSTA ARSAE PARA A COPASA MG – ÍNDICE DE *MALMQUIST*

A partir das informações contidas na NT CRFEF 40/2016 foi estimada a aplicação da metodologia proposta pela ARSAE-MG para o cálculo do índice de *Malmquist*, o índice de Mudança de Eficiência e o índice de Mudança Tecnológica. A principal referência bibliográfica para essa análise foi o artigo de Simar e Wilson (1999) - *Estimating and bootstrapping Malmquist indices*.

A ARSAE-MG disponibilizou a base de dados das empresas avaliadas e o código R da implementação do modelo de cálculo do índice de *Malmquist* e do procedimento de *bootstrapping*. Foram detectadas pequenas inconsistências no código R e na rotina de filtros para a seleção da base de dados final. As principais inconsistências são: (a) conflito na seleção das bases de dados, (b) inconsistência no limite de seleção de hidrometração, gerando uma base de dados de painel desequilibrada, (c) inconsistência no código para a definição da largura de banda e (d) códigos inconsistentes para o cálculo do índice de *Malmquist* ao final da simulação *bootstrap*. Além dessas inconsistências, partes do código foram simplificadas para melhor entendimento. A lista detalhada das inconsistências no código e suas correções encontram-se no Anexo I, e também estão indicadas no código revisado.

Um ponto importante da implementação do algoritmo de *bootstrap* é a estimação do parâmetro de largura de banda, h . Segundo Simar e Wilson (1999), pg. 466: $h \equiv (4/5N)^{1/6}$. Em geral, quanto maior a amostra menor o valor de h .

ESTIMATIVAS DO ÍNDICE DE *MALMQUIST* E INFERÊNCIA ESTATÍSTICA VIA *BOOTSTRAP*

Sejam empresas que produzem m produtos a partir de n insumos, sejam $x \in \mathbb{R}_+^n$ e $y \in \mathbb{R}_+^m$ os vetores de insumos e produtos. O conjunto de produção factível no tempo t pode ser definido como:

$$\mathfrak{X}^t(y) = \{x \in \mathbb{R}_+^n | (x, y) \in \mathcal{P}^t\} \quad (1)$$

onde $\mathcal{P}^t = \{(x, y) | x \text{ pode produzir } y \text{ no período } t\}$.

A distância de Shepard (1970), que define a distância dos insumos para a empresa i no tempo t_1 , com relação à tecnologia existente no tempo t_2 ($t_2 > t_1$) é definida como:

$$D_i^{t_1|t_2} \equiv \sup\{\theta > 0 | x_{it_1}/\theta \in \mathfrak{X}^{t_2}(y_{it_1})\}$$

Färe et al. (1992) define o índice de produtividade de *Malmquist* como:

$$\mathcal{M}_i(t_1, t_2) \equiv \frac{D_i^{t_2|t_2}}{D_i^{t_1|t_1}} \times \left(\frac{D_i^{t_2|t_1}}{D_i^{t_2|t_2}} \times \frac{D_i^{t_1|t_1}}{D_i^{t_1|t_2}} \right)^{1/2} \quad (2)$$

Valores $\mathcal{M}_i(t_1, t_2) < 1$ indicam aumento na produtividade e $\mathcal{M}_i(t_1, t_2) > 1$ indicam redução na produtividade.

A partir da equação (2) é possível definir o índice de mudança de Eficiência,

$$\mathcal{E}_i(t_1, t_2) = \frac{D_i^{t_2|t_2}}{D_i^{t_1|t_1}}, \text{ e o índice de mudança Tecnológica,}$$

$$\mathcal{T}_i(t_1, t_2) = \left(\frac{D_i^{t_2|t_1}}{D_i^{t_2|t_2}} \times \frac{D_i^{t_1|t_1}}{D_i^{t_1|t_2}} \right)^{1/2}.$$

É de interesse utilizar ferramentas estatísticas para avaliar a hipótese nula $H_0: \mathcal{M}_i(t_1, t_2) = 1$, ou seja, um cenário onde não ocorre a mudança na produtividade.

Simar e Wilson (1999) apresentam um procedimento de simulação para gerar amostras *bootstrap* do índice de produtividade de *Malmquist*. Assim, são gerados B índices de Malmquist para cada empresa, $\widehat{\mathcal{M}}_i^*(t_1, t_2)$. A partir das diferenças entre os valores *bootstrap* e o índice de *Malmquist observado* ($\widehat{\mathcal{M}}_i(t_1, t_2)$) é possível construir intervalos de confiança e verificar se o valor unitário ($\mathcal{M}_i(t_1, t_2) = 1$) se encontra no intervalo.

Segundo Simar e Wilson (1999), pág 464 (equações 16 a 17), se $B \rightarrow \infty$ é possível assumir que:

$$[\widehat{\mathcal{M}}_i(t_1, t_2) - \mathcal{M}_i(t_1, t_2)] \sim [\widehat{\mathcal{M}}_i^*(t_1, t_2) - \widehat{\mathcal{M}}_i(t_1, t_2)] | \mathcal{J}$$

o que permite obter a seguinte aproximação bootstrap:

$$\text{Prob}(-b_\alpha^* \leq \widehat{\mathcal{M}}_i(t_1, t_2) - \mathcal{M}_i(t_1, t_2) \leq -a_\alpha^* | \mathcal{J}) \approx 1 - \alpha$$

Portanto, utilizando as estimativas bootstrap, define-se um intervalo de confiança $(1-\alpha)$ como:

$$\widehat{\mathcal{M}}_i(t_1, t_2) + a_\alpha^* \leq \mathcal{M}_i(t_1, t_2) \leq \widehat{\mathcal{M}}_i(t_1, t_2) + b_\alpha^*$$

É dito que a estimativa do índice de *Malmquist* é estatisticamente diferente do valor unitário (o que indica ausência de mudança de produtividade) se o intervalo de confiança não inclui o valor unitário, 1. Neste caso, o resultado do processo de *bootstrapping* confirma que não houve mudança de produtividade. Utilizando o mesmo princípio, é possível definir intervalos de confiança para o índice de mudança de eficiência (\mathcal{E}_i) e o índice de mudança de tecnologia (\mathcal{T}_i).

Para o cálculo do índice de *Malmquist* a base de dados utilizada é a mesma do cálculo dos escores de eficiência, ou seja: AGESPISA, CAERN, CAESB, CAGECE, CAGEPA, CASAN, CESAN, COMPESA, COPASA, CORSAN, DESO, EMBASA, SABESP, SANEAGO e SANEPAR. As variáveis de insumo também continuam as mesmas, assim como o período analisado.

Para entender melhor a proposta da ARSAE, constante na Nota Técnica nº. CRFEF 40/2016, foram realizadas replicações do modelo de bootstrapping proposto. O número de replicações foi definido como $B = 10.000$ e mostra que para a maior parte das empresas não há evidência estatística de ocorrência de ganhos tecnológicos no período, que é o que estamos procurando. A Tabela 1 mostra os resultados.

Salientamos aqui que a proposta de Simar e Wilson (1999) estabelece um procedimento para o cálculo de intervalos de confiança para o índice de *Malmquist* e suas decomposições. Conforme Simar e Wilson (1999), diz-se que há evidência estatística de que o índice é diferente da unidade se o intervalo estabelecido pela equação 17 (pg. 464) não inclui a unidade. A Tabela 1 mostra que não há evidência estatística de que ocorreu mudança tecnológica no período analisado, pois a maior parte delas inclui o valor unitário em seu intervalo. Afirma-se também que a média aritmética ou geométrica dos índices deve ser interpretada como uma medida da centralidade dos índices e não como uma única medida aplicável a todos os prestadores. Para cada prestador foi estimado um intervalo de confiança.

Importante salientar que, na Tabela 1, valores acima de 1 indicam mudança de produtividade, eficiência ou tecnológica positivas e abaixo de 1 negativas. Isto se dá porquê, ao final do artigo citado e utilizado pela ARSAE (Simar & Wilson, 1999, pgs. 467) os mesmos afirmam que o recíproco dos resultados devem ser considerados, mudando assim a interpretação apresentada na página 461.

Quando se olha a mesma simulação para o período de 2012 a 2015, no entanto, vê-se que somente três empresas apresentam valores de evolução tecnológica não significativos. Para as demais empresas há evidência de um aumento de custos ou de perda de produtividade. Isto mostra que o ano de 2015 aprestou custos operacionais relativamente maiores para a maior parte do setor. Este aumento pode ser resultado das condições climáticas desfavoráveis em 2015, como também pode ser indício de um aumento de custo mais duradouro.

Tabela 1 – Resultados Modelo DEA com banco de dados ARSAE 2012-2014

prest	DEA	DEA_Unbiased	DEA_Final	Malmquist	LI95.Malmquist	LS95.Malmquist	Eff.change	LI95.Eff.change	LS95.Eff.change	Tech.change	LI95.Tech.change	LS95.Tech.change
AGESPISA	0.8240	0.7863	0.8318	1.0391	0.9540	1.2135	0.9953	0.8943	1.1587	1.0440	1.0035	1.1100
CAERN	0.9642	0.9231	0.9765	1.0774	1.0320	1.2883	1.0320	0.9714	1.2316	1.0440	1.0031	1.1029
CAESB	1.0000	0.9182	0.9712	1.1050	1.0218	1.2782	1.0000	0.9633	1.0000	1.1050	1.0218	1.2795
CAGECE	0.8785	0.8450	0.8938	1.0954	1.0732	1.3299	1.0491	0.9910	1.2604	1.0441	0.9980	1.1368
CAGEPA	0.8477	0.8067	0.8533	1.0527	1.0169	1.2670	1.0258	0.9729	1.2353	1.0262	0.9598	1.1164
CASAN	0.9580	0.9143	0.9672	1.0090	0.8851	1.1432	0.9664	0.8164	1.0868	1.0441	1.0008	1.1232
CESAN	1.0000	0.9206	0.9738	1.0385	0.9633	1.2136	0.9947	0.8981	1.1580	1.0440	0.9926	1.1293
COMPESA	0.8459	0.8201	0.8676	0.9415	0.7558	1.0156	0.9019	0.7090	0.9687	1.0440	1.0026	1.1026
COPASA	1.0000	0.9454	1.0000	0.9612	0.8872	1.1190	1.0000	0.8951	1.1363	0.9612	0.8917	1.0788
CORSAN	0.7502	0.7205	0.7621	0.8035	0.5520	0.7906	0.8226	0.5380	0.8058	0.9767	0.9065	1.0635
DESO	0.9676	0.9355	0.9895	1.0287	0.9325	1.1853	0.9853	0.8768	1.1321	1.0440	1.0034	1.1030
EMBASA	0.8373	0.8058	0.8524	0.9362	0.7051	0.9800	0.8966	0.6104	0.9295	1.0442	0.9792	1.1447
SABESP	1.0000	0.9300	0.9838	0.9500	0.8716	1.0911	1.0000	0.9489	1.1595	0.9500	0.8384	1.0472
SANEAGO	1.0000	0.9250	0.9785	0.9620	0.9020	1.1334	1.0000	0.9833	1.1805	0.9620	0.8655	1.0615
SANEPAR	1.0000	0.9187	0.9718	1.0110	0.9350	1.1698	1.0000	0.9522	1.0355	1.0110	0.9357	1.1647
em vermelho estão marcadas as prestadoras que apresentam evidência estatística de ganho ou perda de produtividade e/ou eficiência e/ou mudança tecnológica												

Tabela 2 – Resultados Modelo DEA com banco de dados Escher e 2012-2015

prest	DEA	DEA_Unbiased	DEA_Final	Malmquist	LI95 Malmquist	LS95 Malmquist	Eff.change	LI95 Eff.change	LS95 Eff.change	Tech change	LI95 Tech change	LS95 Tech change
AGESPISA	0,8562	0,8228	0,9229	0,7144	0,3092	0,8299	0,8752	0,3863	1,0256	0,8163	0,6987	0,8742
CAERN	0,9704	0,9351	1,0488	1,1061	1,0946	1,6141	1,3550	1,3618	1,9864	0,8163	0,6990	0,8686
CAESB	1,0000	0,9320	1,0454	1,2101	0,7648	1,4761	1,0000	0,5867	1,0000	1,2101	0,9144	1,4897
CAGECE	0,8385	0,8109	0,9096	1,0311	0,9021	1,4378	1,2383	1,1168	1,7441	0,8326	0,6914	0,8995
CAGEPA	0,9095	0,8748	0,9812	0,8985	0,5944	1,1512	1,0371	0,7265	1,3437	0,8663	0,7055	0,9463
CASAN	1,0000	0,9406	1,0550	0,9728	0,7980	1,3329	1,1469	0,9647	1,5772	0,8482	0,7204	0,9150
CESAN	1,0000	0,9316	1,0449	0,9684	0,8288	1,3449	1,1863	1,0343	1,6576	0,8163	0,6306	0,8996
COMPESA	0,8452	0,8189	0,9185	0,9882	0,8557	1,3757	1,2105	1,0809	1,6992	0,8163	0,6978	0,8680
COPASA	1,0000	0,9612	1,0781	0,9260	0,7093	1,2332	1,1069	0,8615	1,4572	0,8366	0,6501	0,9528
CORSAN	0,8636	0,8313	0,9324	0,5733	0,0214	0,5370	0,6537	0,0597	0,5990	0,8770	0,7329	0,9883
DESO	0,9616	0,9311	1,0444	1,0112	0,9190	1,4323	1,2388	1,1515	1,7658	0,8163	0,6964	0,8706
EMBASA	0,8187	0,7902	0,8863	0,8811	0,5689	1,1169	1,0302	0,7010	1,3169	0,8553	0,6709	0,9425
SABESP	1,0000	0,9304	1,0436	0,6966	0,4652	0,8754	1,0000	1,0000	1,2739	0,6966	0,4053	0,8040
SANEAGO	1,0000	0,9315	1,0448	0,9329	0,6295	1,1774	1,0000	0,7185	1,1747	0,9329	0,7260	1,0981
SANEPAR	1,0000	0,9309	1,0441	0,9304	0,5638	1,1190	1,0000	0,5063	1,0459	0,9304	0,6772	1,1373

Em vermelho estão os valores que indicam resultados estatisticamente significativos

APLICAÇÃO DO FATOR DE EFICIÊNCIA

A ARSAE-MG está propondo aplicação dos custos operacionais considerando os itens abaixo:

- I. Referência de eficiência a ser utilizada;
- II. Impacto do Programa de Desligamento Voluntário no Fator de Produtividade;
- III. Aplicação da meta estabelecida em uma trajetória anual;
- IV. Prêmio de eficiência.

Propõe ainda que a definição dos custos operacionais eficientes parta da determinação de dois componentes:

- i. custos operacionais representativos para o período de referência aplicando-se a Classificação Regulatória (NT CRFEF 38/2016)
- ii. meta regulatória para redução dos custos ao longo do ciclo tarifário.

A meta de redução dos custos operacionais contempla o fator “catch-up” e fator de Deslocamento da Fronteira.

Para o fator catch-up, a referência de eficiência utilizada pela ARSAE foi o prestador mais eficiente da amostra. As eficiências foram normalizadas para o melhor prestador, isto é, foram divididas pelo maior índice.

A distância da COPASA para o mais eficiente será igual ao fator de “catch-up” (em percentual), parcela do fator de produtividade.

$$Fator\ catch\ up = 1 - \frac{\text{Índice de Eficiência COPASA}}{\text{Maior Índice de Eficiência}}$$

Além do fator catch-up, a ARSAE propõe que o fator deslocamento da fronteira, seja mensurado por meio da estimação do Índice de Malmquist, do qual considerou somente o deslocamento médio da fronteira de produção, sendo que, se o fator de deslocamento for negativo, a ARSAE o considerará igual a zero.

A ARSAE ainda coloca que é preciso considerar no Fator de Produtividade a recente reestruturação de custos promovida pela COPASA, em 2015, com os Programas de Desligamento Voluntário (PDV's), pois a redução de custos

proveniente destes programas não será observável nos dados do SNIS em 2015.

Se aplicarmos o índice de produtividade obtido sem considerar os PDV's, poderíamos estar realizando uma "dupla glosa". Ou seja, considerar os custos operacionais reduzidos pelos programas na construção das tarifas e aplicar a meta regulatória obtida com as despesas antes dos PDV's. Desta forma, a ARSAE propõe que seja calculada a redução de custos provocada pelos programas de desligamento e que o Fator de Produtividade seja subtraído desta redução.

Assim propõe que meta regulatória dos custos operacionais seja:

$$\text{Fator de Produtividade} = \text{Fator catch up} + \text{Fator Deslocamento} - \frac{\text{redução dos PDVs}}{\text{Custos Operacionais}}$$

O Fator de Produtividade será aplicado sobre os custos operacionais representativos para o período de referência.

Além disso, a ARSAE propõe um "prêmio" quando a soma dos fatores catch-up e deslocamento for menor que a redução provocada pelos programas de desligamento voluntário da seguinte forma:

Por fim, deve-se considerar a possibilidade da COPASA já ter reduzido o seu custo além da meta regulatória, ou seja, a soma dos fatores catch-up e deslocamento é menor que a redução provocada pelos programas de desligamento voluntário. Neste caso, a Companhia deve ser premiada pela eficiência alcançada. Metade desta distância será aplicada aumentando os custos operacionais utilizados na construção das tarifas, sendo que a outra metade será compartilhada para os usuários.

1.1.1. Análise

Em primeiro lugar, entende-se que o cálculo do fator "catch-up" tendo como referência o maior índice de eficiência, não incentiva os prestadores de maior eficiência a manterem os seus custos relativamente mais baixos, nem tampouco a melhorarem ainda mais sua gestão.

Tal qual se faz no setor elétrico brasileiro o cálculo do índice deve ser feito com a eficiência média do setor, ou com o 3º quartil, estimulando as empresas com eficiência acima da média ou do 3º quartil a continuarem atuando na melhoria dos processos, lembrando que a própria ARSAE utilizou nos processos

tarifários da CESAMA e do SAAE de Passos o 3º Quartil como referência de eficiência.

Para isto é fundamental reconhecer, até um certo limite, custos eficientes acima dos custos reais como um necessário incentivo.

A Agência Nacional de Energia Elétrica Brasileira – ANEEL, que também utiliza bootstrap no cálculo dos intervalos de variação do custo operacional eficiente calcula os custos operacionais da seguinte maneira (Submódulo 2.2 do PRORET – Procedimentos de Regulação Tarifária – Custos Operacionais e Receitas irrecuperáveis):

17. O valor do custo operacional eficiente será definido a partir da comparação da receita de custos operacionais no momento da revisão e o intervalo de custos operacionais eficientes para a concessionária em análise, conforme abaixo.

$$CO_{Ef} = \min(\max(CO_{At}; LI); LS)$$

CO_{At}: receita de custos operacionais no Ano Teste

CO_{ef}: valor dos custos operacionais regulatórios eficientes;

LS: limite superior dos custos operacionais regulatórios eficientes;

LI: limite inferior dos custos operacionais regulatórios eficientes;

18. Como regra geral, a meta será ajustada de tal forma que não implique uma trajetória que exceda uma variação de custos operacionais regulatórios superior a 5% a.a.

$$\Delta CO = \min\left(\left|\sqrt[N]{\frac{CO_{Ef}}{CO_{At}}} - 1\right|; 5\%\right)$$

$$CO_{meta} = CO_{At}(1 \pm \Delta CO)^N$$

ΔCO: variação anual dos custos operacionais regulatórios;

CO_{meta}: meta de custos operacionais ajustada ao limite máximo de variação anual.

19. Para os casos em que a razão entre o custo definido pela Eq. 3 (CO_{meta}) e os custos operacionais reais for superior a 120%, a meta de custos operacionais será ajustada conforme expressão a seguir.

$$CO'_{meta} = \frac{1,2 \cdot Opex_{médio} + CO_{meta}}{2}$$

CO_{meta}' : meta de custos operacionais regulatórios com compartilhamento;

Opexmédio: média dos custos operacionais reais.

21. Para o cálculo de *Opexmédio* será considerado o valor médio dos custos operacionais reais nos dois anos consecutivos mais próximos da revisão tarifária, considerando a informação mais recente disponível, atualizados monetariamente pelo IPCA até a data da revisão tarifária.

23. Os intervalos de custos operacionais eficientes são calculados conforme expressão a seguir:

$$LS = \alpha \cdot \frac{\theta_{sup}}{\theta_{ref}} \cdot Opex$$

$$LI = \alpha \cdot \frac{\theta_{inf}}{\theta_{ref}} \cdot Opex$$

LS: limite superior dos custos operacionais regulatórios;

LI: limite inferior dos custos operacionais regulatórios;

θsup: limite superior do intervalo de eficiência apurado para a empresa;

θinf: limite inferior do intervalo de eficiência apurado para a empresa;

θref: referência de eficiência;

Opex: custo operacional real da empresa usado no cálculo de eficiência;

α: fator de atualização.

25. A referência de eficiência (*θref*) é de 76%.

27. Para o cálculo do fator de atualização *α*, considera-se a seguinte equação:

$$Opex_{Ef} = \frac{1}{u} \left(\sum_{j=1}^m v_j \gamma_j + \varphi \right)$$

Opex_{ef}: o custo eficiente estimado para a concessionária;

γ_j: produto “j” da empresa;

v_j: “peso” atribuído ao produto “j”;

u: “peso” atribuído ao insumo;

m: total de produtos;

φ: “fator de escala” da empresa;

29. O fator de atualização será calculado conforme a equação a seguir.

$$\alpha = \frac{Opex_{Ef}^{t+1}}{Opex_{Ef}^t} \times \frac{IPCA_{Ef}^{t+1}}{IPCA_{Ef}^t}$$

Opex_{ef}^{t+1}: custo eficiente estimado na data base da revisão tarifária;

Opex_{eft}: custo eficiente estimado na data base do cálculo da eficiência;
IPCA_{t+1}: número índice do IPCA no mês anterior à data base da revisão tarifária;
IPCA_t: número índice do IPCA no mês anterior à data base do cálculo da eficiência;

Percebe-se assim, que para o cálculo do intervalo de Custos Operacionais Eficientes, de acordo com a ANEEL, a primeira etapa é o cálculo do fator de atualização. Para isto se faz necessário atualizar o custo eficiente estimado na data base do cálculo da eficiência para a data base da revisão tarifária e além disso, fazer atualização monetária.

Além do Fator de Atualização nos cálculos da ANEEL, para o cálculo de LS e LI tem como referência de 76% que é a Eficiência média do setor elétrico apurada no modelo DEA.

Assim, por exemplo, para uma empresa com a ELFSM, a ANEEL calculou uma receita de Custos Operacionais Eficientes em sua Revisão 24,5% acima dos custos reais. Isto induzirá as empresas acima da eficiência média do setor a melhorarem ainda mais, e sempre, sua eficiência, contribuindo, também, para o deslocamento da fronteira de eficiência.

Portanto, identifica-se vários problemas nos procedimentos propostos pela ARSAE, que precisam ser terminantemente sanados.

A primeira questão diz respeito à Classificação Regulatória, ou reconhecimento dos custos na Receita Requerida, e os estudos de Análise Comparativa.

A despeito das contribuições feitas no âmbito da AP12, sobre a Classificação Regulatória, na qual mostramos a necessidade de reconhecimento de várias despesas totalmente necessárias à prestação do serviço, existe uma grande distância entre os custos considerados na Análise Comparativa e os custos operacionais do período de referência. E maior ainda é a distância entre os custos considerados na Análise Comparativa e os custos operacionais “representativos”, diga-se glosados, do período de referência

Não se pode calcular fator de eficiência com determinados custos, de um período, e simplesmente aplicar em outros custos, de outro ano (que podem ter reduzido ou até aumentado), com outro mercado associado, e ainda glosados.

Assim a primeira correção no método de aplicação dos escores de eficiência é atualizar o valor dos Custos Operacionais Totais da Empresa Eficiente, pois os estudos de Análise Comparativa foram feitos com a totalidade dos custos operacionais, e recalculer a distância à fronteira no momento da Revisão Tarifária, pois:

- i. No momento da Revisão, deve-se atualizar o valor dos Custos Operacionais da Empresa Eficiente – COPASA MG, a partir da função de custo definida no modelo, trazendo-os para o Período de Referência, atualizando pela inflação e crescimento dos produtos. Se houve crescimento de mercado e rede, necessariamente haverá crescimento dos Custos Operacionais da Empresa Eficiente – COPASA MG, pois os quantitativos dos produtos se alteraram.
- ii. A Empresa pode ter reduzido, ou até aumentado, os seus Custos Operacionais reais entre o período dos estudos de análise comparativa e a data da Revisão.

Somente após esta atualização é que se pode comparar o Custo Realizado com o Eficiente no Período de Referência e calcular a verdadeira distância da empresa Real à empresa Eficiente, bem como o correto fator catch-up.

Caso não houvesse necessidade de aplicar um fator catch-up no momento da Revisão, ou seja, se fosse possível um ajuste total e imediato dos custos, bastaria simplesmente considerar na composição da Receita Requerida o Custo Eficiente atualizado da Empresa (incluindo a atualização monetária).

Além disso, os Custos Operacionais da COPASA MG só poderiam ser considerados sem os custos comumente denominados inelegíveis, se a Análise Comparativa tivesse sido feita somente com os custos elegíveis de todas as empresas da amostra utilizada. De fato, esta é a forma correta de se fazer análise comparativa, mas isto não foi feito, pois praticamente todos os custos estão na base do SNIS, de forma que, mantendo essa base, não se pode aplicar tal glosa nesta Revisão.

Como já colocado exaustivamente, não cabe glosar os custos operacionais do prestador, quando o regulador já está definindo por análise comparativa os seus Custos Eficientes, calculados com todos os custos operacionais dos demais prestadores, e quando o regulador está calculando um fator catch-up com os Custos Operacionais totais da Empresa. O próprio escore de eficiência, ou melhor, a distância à fronteira estabelecida é o correto ajuste que se faz para reconhecer nas tarifas o custo considerado eficiente

Todavia, a forma correta de se fazer Análise Comparativa é criar uma base de custos ditos elegíveis, realmente necessários à prestação dos serviços e retirar aqueles que tornam as empresas incomparáveis, como as despesas com aluguéis, veículos e softwares, bem como excluir os Custos Não Gerenciáveis.

Faz-se necessário iniciar essa discussão para que na próxima Revisão se ajuste a base de dados para permitir uma correta comparação entre empresas.

Propõe-se, portanto, primeiramente calcular o fator “catch-up” tendo como referência o índice de eficiência média do setor para incentivar os prestadores de maior eficiência a manterem os seus custos relativamente mais baixos, e reduzirem ainda mais os seus custos operacionais.

Para avaliar qual seria o impacto do uso da eficiência média, apresenta-se na tabela a seguir o resultado do modelo DEA proposto pela ARSAE, com informações de 2012 a 2015, provenientes do SNIS e com as alterações propostas neste documento, que são:

- 1) Consideração do balanço de água com retirada do volume de água exportada;
- 2) Utilização dos custos de Exploração são itens da Parcela A, ou seja, retirando os custos com Energia Elétrica e Material de Tratamento;

Prestador	Resultados DEA		
	DEA	DEA_Unbiased	DEA_Final
AGESPISA	85,62%	82,32%	92,33%
CAERN	97,04%	93,38%	104,75%
CAESB	100,00%	93,09%	104,42%
CAGECE	83,85%	81,15%	91,02%
CAGEPA	90,95%	87,58%	98,24%
CASAN	100,00%	93,98%	105,41%
CESAN	100,00%	93,11%	104,44%
COMPESA	84,52%	81,88%	91,84%
COPASA	100,00%	96,14%	107,83%
CORSAN	86,36%	83,20%	93,32%
DESO	96,16%	93,16%	104,50%
EMBASA	81,87%	79,01%	88,62%
SABESP	100,00%	92,92%	104,23%
SANEAGO	100,00%	93,23%	104,57%
SANEPAR	100,00%	93,16%	104,49%

Nota-se que, uma vez que os resultados de eficiência têm dispersão relativamente baixa, os valores de prêmio de eficiência são baixos, principalmente se considerado que na metodologia da ARSAE eles serão aplicados gradativamente ao longo dos 4 anos do ciclo tarifário seguinte. No caso da COPASA este prêmio seria de 1,9% ao ano para os próximos 4 anos. O escore médio de eficiência, depois da correção de viés é de 89,15%, enquanto o escore da COPASA é de 96,14%.

Em segundo lugar, propõe-se atualizar o valor dos Custos Operacionais da Empresa Eficiente – COPASA MG, da data base dos custos utilizados nos estudos de análise comparativa, trazendo-os para o Período de Referência (maio/16 a abril/17), a partir da função de custo definida no DEA que percebe o crescimento dos produtos e atualizando pela inflação, para em seguida multiplicar pelo Fator de Produtividade, assim calculado:

$$CO_{Ef} = Opex_{AC} \cdot \alpha \cdot \left(\frac{\theta_{COPASA}}{\theta_{m\u00e9dio}} \right)$$

Onde:

CO_{ef} = custo operacional eficiente

$Opex_{AC}$ = Custo Operacional utilizado nos estudos de Análise Comparativa

α = Fator de Atualização

θ_{COPASA} = escore de Eficiência da COPASA

$\theta_{médio}$ = escore de Eficiência médio

O fator de atualização será calculado conforme a equação a seguir.

$$\alpha = \frac{Opex_{Ef}^{t+1}}{Opex_{Ef}^t} \times \frac{IPCA_{Ef}^{t+1}}{IPCA_{Ef}^t}$$

$Opex_{Ef}^{t+1}$ = custo eficiente estimado na data base da revisão tarifária;

$Opex_{Ef}^t$ = custo eficiente estimado na data base do cálculo da eficiência;

$IPCA_{Ef}^{t+1}$ = número índice do IPCA no mês anterior à data base da revisão tarifária;

$IPCA_{Ef}^t$ = número índice do IPCA no mês anterior à data base do cálculo da eficiência;

Como exemplo, segue o cálculo do resultado obtido com o modelo proposto neste documento.

A despesa de exploração média do período, em valores de dezembro de 2015, considerado no modelo foi de R\$ 1.860 milhões. Este mesmo valor, sem Impostos, Energia Elétrica e Material de Tratamento, apurado em 2016 foi de R\$ 1.624 milhões, calculado conforme tabela a seguir.

Despesas de Exploração (Total - R\$)	2.355.319.977
Impostos	290.998.435
Despesas com Energia Elétrica	374.919.814
Despesas com Materiais Tratamento	64.548.185
Despesas para comparação com empresa eficiente	1.624.296.294

Os demais valores considerados na Análise Comparativa (Valores AC) e as informações para o ano de 2016 (ano civil), assim como os pesos das variáveis dadas pelo modelo estão na tabela abaixo.

	opex_sem_vpa_defl	econativas_a	econativas_e	voledt	perdas
Valores AC	1.860.462.927	4.741.852	3.106.373	224.392	321.538
Valores 2016	1.624.296.294	4.959.461	3.322.262	299.567	334.697
Pesos AC	1,94539E-10	1,98443E-06	1,44997E-07	1,00583E-07	0,00000E+00

O valor estimado de custos operacionais atualizado, é dado pela seguinte fórmula:

$$Opex_{Ef} = \frac{1}{u_1} \left(\sum_{j=1}^m \vartheta_j \gamma_j - u_2 p \right)$$

Onde:

γ_j : produto "j" da empresa;
 θ_j : "peso" atribuído ao produto "j" da empresa;
 v_j : "peso" atribuído ao produto "j";
 u_1 : "peso" atribuído ao insumo "Despesa de Exploração";
 u_2 : "peso" atribuído ao insumo "Volume de Perdas";
 p : volume de perdas;
 m : total de produtos;

Este cálculo retorna para a COPASA uma despesa de exploração eficiente de R\$ 2.000,045 milhões. Considerando uma inflação no ano de 2016 de 6,2881%, isto resulta em um fator de atualização α de 114,26%

O valor de custo eficiente atualizado para a COPASA resultante da fórmula abaixo, é de R\$ 2.292.315.391.

$$CO_{Ef} = 1.860 \text{ milhões} \cdot 114,26 \cdot \left(\frac{96,14\%}{89,15\%} \right)$$

Esta fórmula resulta em um custo operacional a ser reconhecido de R\$ 2.120.735.472. Comparando este valor com o custo operacional real da COPASA no período de 2016, vemos que a diferença é de 41%. No entanto, este valor está sem as despesas com Impostos, Energia Elétrica e Material de Tratamento. Estes custos devem ser adicionados ao valor a ser reconhecido de custos operacionais, o que daria um valor a ser reconhecido de R\$ 2.565.261.248 se não considerarmos o último termo da fórmula (prêmio de eficiência), ou de R\$ 2.731.895.007, já com o prêmio de eficiência.

Conforme mostrado anteriormente, os custos operacionais reais de 2016 da COPASA foram de R\$ 2.064 milhões, enquanto o custo eficiente a ser considerado é de R\$ 2.565 milhões, o que dá uma diferença de R\$ 501 milhões. Isto mostra claramente que a simples consideração do PDV como

redução de custo promovido pela COPASA entre o período utilizado para a análise comparativa e o atual não é suficiente. A comparação mostra que o esforço de eficiência da COPASA no último ano foi de alcançar um custo R\$ 501 milhões menor que o custo eficiente, o que é mais que o dobro do impacto do PDV.

A atualização dos custos operacionais usados na análise comparativa de eficiência pelos pesos encontrados no estudo pode ser entendida de outra forma. A tabela abaixo apresenta novamente os valores considerados no estudo de análise comparativa, que são valores médios de 2012 a 2015. Vemos que apesar da COPASA atender um mercado de economias de água 4,59% maior que na época do estudo, assim como um mercado de economias de esgoto 6,95% maior e ter tratado um volume de esgoto 33,5% maior, teve um custo operacional 12,69% menor, sem considerar a inflação de 2016, já que os valores do estudo estão com atualização monetária apenas até dezembro de 2015.

	opex_sem_vpa_defl	econativas_a	econativas_e	voledt	perdas
Valores AC	1.860.462.927	4.741.852	3.106.373	224.392	321.538
Valores 2016	1.624.296.294	4.959.461	3.322.262	299.567	334.697
Evolução	-12,69%	4,59%	6,95%	33,50%	4,09%

Por outro lado, o volume de perdas cresceu em 4,09%. Isto não significa necessariamente que o índice de perdas da COPASA piorou, porque o volume de água entregue também é diferente. Inclusive, o volume de perdas cresceu menos que o número de ligações, o que pode indicar uma melhora no índice. Mesmo assim, é difícil dar um peso para este efeito potencialmente negativo. É por isto que o uso dos pesos obtidos no modelo DEA se mostra um solução interessante.

Para verificar se o resultado da atualização dos custos eficientes faz sentido, pode-se calcular o crescimento de custos que este indica e comparar com o crescimento das variáveis que são de interesse ao estudo. Quando se calcula o custo eficiente com os produtos de 2016, chega-se a um valor de R\$ 2.000 milhões, que é 7,5% maior que a média do período 2012 a 2015. Este crescimento leva em conta o crescimento do número de economias tanto de

água, quanto de esgoto e o crescimento relativamente menor de perdas, que é considerada uma variável negativa.

Há quem questione o uso dos pesos dados pelo modelo DEA para esta atualização, já que o DEA é calculado com programação linear e os pesos obtidos são resultado de uma entre possivelmente várias outras soluções. No entanto, consideramos ser uma solução satisfatória para esta atualização, já que o ano para o qual se está projetando o custo eficiente é imediatamente posterior ao estudo, o que impede que mudanças estruturais muito grandes na companhia distorçam o resultado obtido com a atualização. Adicionalmente, o aumento projetado desta forma está condizente com o crescimento no número de economias de água e de esgoto, principais direcionadores de custo da companhia.

De toda forma, caso a aplicação do fator de produtividade como proposto acima seja algo de questionável utilização, poder-se-ia adotar proposta alternativa a seguir que se aproxima do conceito proposta pela ARSAE-MG.

Para proceder à correta aplicação do resultado da Análise de Eficiência Comparativa, há que se considerar, além do score de eficiência obtido pela empresa na comparação, a evolução dos custos reais da empresa entre o período analisado pelo estudo de eficiência comparativa e o período atual. Isto porque entre o período analisado e o da aplicação, a empresa pode ter feito esforços para reduzir os custos operacionais ou incorrido em gastos maiores do que no período analisado. Em ambos os casos, o score calculado na análise de eficiência não estaria mais corretamente dimensionado.

No caso específico da COPASA, houve no ano de 2015 e 2016 um esforço muito grande de redução de custos operacionais. Entre estes esforços, o mais proeminente foi o plano de demissões – PDV, cujo efeito aparece nos custos operacionais a partir do ano de 2016. Por isso mesmo, a ARSAE, agindo corretamente, considerou o efeito do PDV na aplicação do resultado da análise de eficiência. No entanto, o PDV não foi a única ação de eficiência promovida pela COPASA em 2016. A redução de custos foi bem maior.

Para se fazer uma avaliação de quanto foi, de fato, a redução de custos da COPASA em 2016 pode-se comparar o custo usado na análise comparativa, atualizá-lo para 2016 e, então, verificar qual é a diferença entre os custos da análise comparativa e o real.

Para se fazer esta atualização, deve-se levar em conta duas coisas. Primeiramente, o mercado da empresa evoluiu, ou, de outra forma, os produtos utilizados na análise comparativa, não são mais os mesmos. O número de ligações de água e esgoto mudou e as perdas e o volume de esgoto tratado mudaram. Em segundo lugar, deve-se levar em conta a atualização monetária entre o período da análise comparativa e o atual. Feitas estas duas atualizações, a comparação entre o custo eficiente da análise comparativa e o custo real atual nos mostrará qual foi a redução (ou aumento) efetivo dos custos operacionais.

Iremos primeiramente fazer a atualização de mercado. Esta atualização pode ser feita substituindo-se os produtos utilizados na análise comparativa com os valores destes mesmos produtos no ano de referência.

A fórmula para obter o valor de custos atualizado, no caso do estudo proposto pela ARSAE, é a seguinte:

$$Opex_{Ef} = \frac{1}{u_1} \left(\sum_{j=1}^m \vartheta_j \gamma_j - u_2 p \right)$$

Onde:

Opex_{Eff}: Custo operacional eficiente;
y_j: produto “j” da empresa;
θ_j: “peso” atribuído ao produto “j” da empresa;
v_j: “peso” atribuído ao produto “j”;
u₁: “peso” atribuído ao insumo “Despesa de Exploração”;
u₂: “peso” atribuído ao insumo “Volume de Perdas”;
p: volume de perdas;
m: total de produtos;

Os valores dos produtos e os coeficientes da fórmula (pesos) são os da tabela abaixo. Deve-se atentar para o fato que os custos aqui apresentados não

incluem os gastos com Energia Elétrica e Material de Tratamento, que a COPASA considera que não devem fazer parte da análise comparativa.

	opex_sem_vpa_defl	econativas_a	econativas_e	Voledt	perdas
Valores AC	1.860.462.927	4.741.852	3.106.373	224.392	321.538
Valores 2016	1.624.296.294	4.959.461	3.322.262	299.567	334.697
Pesos AC	1,94539E-10	1,44997E-07	1,00583E-07	0,00000E+00	1,98443E-06

Este cálculo retorna para a COPASA uma despesa de exploração eficiente atualizada de R\$ 2.000 milhões. Considerando uma inflação no ano de 2016 de 6,2881%, custo eficiente atualizado para a COPASA resultante seria de R\$ 2.126 milhões.

Comparando este custo com o real da COPASA, sem energia elétrica e material de tratamento, vê-se que há uma diferença de R\$ 501 milhões, que corresponde ao esforço de eficiência da COPASA, que inclui o PDV e demais medidas de eficiência praticadas pela COPASA depois do período de análise comparativa. Ainda haverá alguma diferença quando for considerado o período de referência de 12 meses anteriores à data da revisão, mas o valor deverá ser próximo a este.

Assim, propõe-se que a ARSAE considere no cálculo dos custos operacionais todo o esforço de eficiência da COPASA posterior à análise comparativa e não somente o PDV. A fórmula de cálculo do Fator de produtividade seria, portanto dada da seguinte forma:

Fator de Produtividade

$$= \text{Fator cach up} + \text{Fator Deslocamento} - \frac{\text{Redução CO posterior à AC}}{\text{Custos Operacionais}}$$

Onde o fator “Redução CO posterior à AC” se refere à redução dos custos operacionais posterior à análise comparativa, incluindo o PDV e os demais esforços de redução de custos, conforme proposta anteriormente.

PROPOSTAS

- 1) Dividir o escore de eficiência da COPASA pela média de todas as empresas ou pelo valor do 3º. Quartil gerando assim incentivos à eficiência, conforme toda boa regulação;
- 2) Como foi demonstrado nos resultados, não há evidências estatísticas de evolução tecnológica para o período de 2012 a 2014 e por este motivo o valor do deslocamento constante na equação do fator de produtividade (pág. 13 – NT 40/2016 – CRFEF/ARSAE) seria nulo. Por outro lado, quando se considera também o ano de 2015, nota-se que neste último ano há um incremento significativo nos custos operacionais de praticamente todas as empresas. Ou seja, há um retrocesso da fronteira de eficiência. Há de se avaliar se este efeito de aumento de custos foi devido às condições excepcionais do ano de 2015 (escassez de água), ou se também há indícios de outros fatores que podem ter levado a um aumento mais duradouro dos custos operacionais no setor. Neste último caso, como os resultados são estatisticamente significantes, seria o caso de aplicar um fator de deslocamento da fronteira negativo.
- 3) Fazer as correções propostas no Script do R constante no Anexo I.
- 4) Considerar os efeitos da redução de custos operacionais da COPASA entre o período da análise comparativa e o atual.
- 5) Considerar na análise comparativa o balanço de água deduzindo a água tratada exportada;
- 6) Não considerar na análise comparativa os custos operacionais não gerenciáveis – energia elétrica e material de tratamento;

ANEXO I - Inconsistências nos filtros da base de dados

A Tabela 1 apresenta uma lista de comandos de filtro da base de dados que apresentaram inconsistências e suas respectivas modificações.

Tabela 1. Inconsistências e correções no código R da seleção da base de dados.

Tipo	Alteração
simplificação	De: <pre>base_corte1 <- base_dados[base_dados\$serv == 'Água e Esgoto',]</pre> Para: <pre>base_corte1 <- subset(base_dados, serv == 'Água e Esgoto')</pre>
simplificação	De: <pre>base_corte2 <- base_corte1[!(base_corte1\$natjur == 'Autarquia'),]</pre> Para: <pre>base_corte2 <- subset(base_corte1, natjur != 'Autarquia')</pre>
Inconsistência	De: <pre>econ_agua_med_copasa <- mean (base_corte2[base_corte1\$prest == "COPASA",25],na.rm=TRUE)</pre> Para: <pre>base_corte2B <- subset(base_corte2, prest == 'COPASA') econ_agua_med_copasa <- mean(base_corte2B\$econativas_a)</pre>
simplificação	De: <pre>hidrom_med <- with(base_corte3, tapply(hidrom, prest,mean,na.rm=TRUE)) hidrom_med <- data.frame(row.names(hidrom_med),hidrom_med) names(hidrom_med)[1] <- 'prest'</pre> Para: <pre>hidrom_med <- aggregate(hidrom ~ prest, data=base_corte3, FUN=mean) names(hidrom_med)[2] <- 'hidrom_med'</pre>
Inconsistência	De: <pre>base_corte4 <- base_corte3[base_corte3\$hidrom_med >= 70,]</pre> Para: <pre>base_final <- subset(base_corte3, base_corte3\$hidrom_med >= 69.54) ## corrigido p/ "CAERN" que tem em 2012 hidrom_med == 69.55</pre>
simplificação	De: <pre>opex_sem_imp_defl_med <- with(base_final, tapply(opex_sem_imp_defl, prest,mean,na.rm=TRUE)) econativas_a_med <- with(base_final, tapply(econativas_a,prest,mean,na.rm=TRUE)) econativas_e_med <- with(base_final, tapply(econativas_e,prest,mean,na.rm=TRUE)) voledt_med <- with(base_final, tapply(voledt,prest,mean,na.rm=TRUE)) perdas_med <- with(base_final, tapply(perdas,prest,mean,na.rm=TRUE)) prestadores <- unique(base_final\$prest) ## Escolhendo insumos e produtos x = cbind (opex_sem_imp_defl_med ,perdas_med) y = cbind (econativas_a_med, econativas_e_med, voledt_med)</pre> Para: <pre>base_medias <- aggregate(cbind(opex_sem_imp_defl, econativas_a, econativas_e, voledt, perdas) ~ prest, data=base_final, FUN=mean) x <- as.matrix(base_medias[,c("opex_sem_imp_defl", "perdas")]) y <- as.matrix(base_medias[,c("econativas_a", "econativas_e", "voledt")])</pre>
Inconsistência	De: $h = (4/(5*n))^{1/6}$ Para: $h = (4/(5*n))^{(1/6)}$
Inconsistência	De: <pre>if (index[i] <= n index[i] > 3*n) {</pre>

	Para: <pre>if (index[i] %in% c(1:n, 31:45)) {</pre>
Inconsistência	De: <pre>x_star1 = gamma_star[,1]/original[,1]*x1 x_star2 = gamma_star[,2]/original[,2]*x2 Dt1t1_boot = dea(x_star1,y1, RTS = rts,ORIENTATION = orientation) Dt2t2_boot = dea(x_star2,y2, RTS = rts,ORIENTATION = orientation) Dt1t2 = dea(x_star1,y1, XREF = x_star2, YREF = y2, RTS = rts,ORIENTATION = orientation) Dt2t1 = dea(x_star2,y2,XREF = x_star1, YREF = y1, RTS = rts,ORIENTATION = orientation) tal[,N] = sqrt((Dt1t1_boot\$eff*Dt2t1\$eff)/(Dt1t2\$eff*Dt2t2_boot\$eff)) epsilon[,N] = Dt2t2_boot\$eff/Dt1t1_boot\$eff</pre> Para: <pre>x_star1 = (gamma_star[,1]/original[,1])*x1 x_star2 = (gamma_star[,2]/original[,2])*x2 Dt1t1_boot = 1/dea(x_star1, y1, RTS = rts, ORIENTATION = orientation)\$eff Dt2t2_boot = 1/dea(x_star2, y2, RTS = rts, ORIENTATION = orientation)\$eff Dt1t2_boot = 1/dea(x_star1, y1, XREF = x_star2, YREF = y2, RTS = rts, ORIENTATION = orientation)\$eff Dt2t1_boot = 1/dea(x_star2, y2, XREF = x_star1, YREF = y1, RTS = rts, ORIENTATION = orientation)\$eff tal[,N] = sqrt(((Dt2t1_boot/Dt2t2_boot)*(Dt1t1_boot/Dt1t2_boot))) epsilon[,N] = Dt2t2_boot/Dt1t1_boot</pre>

Além das correções das inconsistências na base, também foi implementada uma rotina independente, para validação dos resultados. O código corrigido e o novo código da implementação estão disponíveis nos anexos.

ANEXO II – CÓDIGO R DA ARSAE CORRIGIDO COM COMENTÁRIOS

```

## Codigo ARSAR revisado e corrigido
## NESP/UFMG - 28/12/2016

# Limpar os dados do R
rm(list=ls(all=TRUE))

## O script utilizará os seguintes pacotes
if (!require(openxlsx)) { install.packages('openxlsx');
library(openxlsx) }
if (!require(Benchmarking)) { install.packages('Benchmarking');
library(Benchmarking) }
if (!require(svDialogs)) { install.packages('svDialogs');
library(svDialogs) }

## - - - - -
## NTCRFEF_40_2016_COE_Copasa_Banco_de_dados2
dlgMessage('Selecione o arquivo com as informações do Snis!')
base_dados <- read.xlsx(file.choose(), sheet = "Snis", startRow =
15)

base_corte1 <- subset(base_dados, serv == 'Água e Esgoto')
base_corte2 <- subset(base_corte1, natjur != 'Autarquia')
base_corte2B <- subset(base_corte2, prest == 'COPASA')
econ_agua_med_copasa <- mean(base_corte2B$econativas_a)

econativas_a_med <- aggregate(econativas_a ~ prest,
data=base_corte2, FUN=mean)
names(econativas_a_med)[2] <- 'econativas_a_med'
base_corte2 <- merge(base_corte2, econativas_a_med, by='prest')
base_corte3 <- subset(base_corte2, econativas_a_med >=
econ_agua_med_copasa*0.1)

# Nível mínimo de hidrometração = 70%.
hidrom_med <- aggregate(hidrom ~ prest, data=base_corte3,
FUN=mean)
names(hidrom_med)[2] <- 'hidrom_med'

base_corte3 <- merge(base_corte3, hidrom_med, by='prest')

## base_final
base_final <- subset(base_corte3, base_corte3$hidrom_med >= 69.54)
## corrigido p/ "CAERN" que tem em 2012 hidrom_med == 69.55

## Construção das perdas (1000 m³/ano)
base_final$perdas = with(base_final, volprod_a + volimp_a -
volcons_a)
## write.csv2(base_final, "base_final.csv")

## Agrega a base para as médias
base_medias <- aggregate(cbind(opex_sem_imp_defl, econativas_a,
econativas_e, voledt, perdas) ~
prest, data=base_final, FUN=mean)

x <- as.matrix( base_medias[,c("opex_sem_imp_defl", "perdas")] )
y <- as.matrix( base_medias[,c("econativas_a", "econativas_e",
"voledt")] )
base_medias$DEA <- dea(X=x, Y=y, RTS="irs", ORIENTATION="in")$eff
prestadores <- base_medias$prest ## Para usar no código da ARSAE

```

```

## - - - - -
- -
## Calculo do Indice de Malmquist
  x.t1 <- subset(base_final, ano == 2012,
select=c("opex_sem_imp_defl", "perdas"))
  y.t1 <- subset(base_final, ano == 2012, select=c("econativas_a",
"econativas_e", "voledt"))
  x.t2 <- subset(base_final, ano == 2014,
select=c("opex_sem_imp_defl", "perdas"))
  y.t2 <- subset(base_final, ano == 2014, select=c("econativas_a",
"econativas_e", "voledt"))

  x.t1 <- as.matrix(x.t1)
  y.t1 <- as.matrix(y.t1)
  x.t2 <- as.matrix(x.t2)
  y.t2 <- as.matrix(y.t2)

  ## t2 > t1
  Dt1_t1 <- 1/dea(X=x.t1, Y=y.t1, RTS="crs", ORIENTATION="in")$eff
  Dt2_t2 <- 1/dea(X=x.t2, Y=y.t2, RTS="crs", ORIENTATION="in")$eff

  Dt2_t1 <- 1/dea(X=x.t2 ,Y=y.t2, RTS="crs", ORIENTATION="in",
XREF=x.t1, YREF=y.t1)$eff

  Dt1_t2 <- 1/dea(X=x.t1 ,Y=y.t1, RTS="crs", ORIENTATION="in",
XREF=x.t2, YREF=y.t2)$eff

  Malmquist <- (Dt2_t2/Dt1_t1)*sqrt(
((Dt2_t1/Dt2_t2)*(Dt1_t1/Dt1_t2)) )
  Eff.change <- Dt2_t2/Dt1_t1
  Tech.change <- sqrt( ((Dt2_t1/Dt2_t2)*(Dt1_t1/Dt1_t2)) )
  ## Eff.change * Tech.change

  base_medias$Eff.change <- Eff.change
  base_medias$Tech.change <- Tech.change
  base_medias$Malmquist <- Malmquist

## - - - - -
- -
## - - - - -
- -
## ALGORITMO ARSAE...

## CALCULAR O DEA
  dea_model <- dea(x, y, RTS="irs", "in")
  eficiencia <- eff(dea_model)
  is.vector(eficiencia)
  tab_eff <- data.frame(prestadores, eficiencia)

## - - - - -
- - - - -
## Correção de viés com o algoritmo de Simar & Wilson (1998)
  dea_model_ub <- dea.boot(x, y, NREP = 2000, EFF = NULL, RTS =
"irs")
  tab_eff_ub <-
data.frame(base_medias$prest, dea_model_ub$eff.bc, dea_model_ub$conf.int
)
  names(tab_eff_ub) <- c("Prestadores", "Eficiências", "LI
Eficiências", "LS Eficiências")
  max(dea_model_ub$eff.bc)

```

```

summary(tab_eff_ub$Eficiências)
## -----
-----

base_12 <- subset (base_final,base_final$ano == 2012)
base_14 <- subset (base_final,base_final$ano == 2014)
x_12 = with(base_12,cbind (opex_sem_imp_defl, perdas))
y_12 = with(base_12,cbind (econativas_a, econativas_e, voledt))
x_14 = with(base_14,cbind (opex_sem_imp_defl, perdas))
y_14 = with(base_14,cbind (econativas_a, econativas_e, voledt))

x1 = x_12; y1 = y_12
x2 = x_14; y2 = y_14

## Fazendo o bootstrapping do Malmquist
f.Malmquist.boot <- function (x1, y1, x2, y2, rts="crs",
orientation="in", NREP = 2000){
  n      <- nrow(x1)
  tal    <- matrix(NA, n, NREP)
  epsilon <- matrix(NA, n, NREP)
  Malm   <- matrix(NA, n, NREP)
  Dt1_t1 <- 1/dea(x1, y1, RTS = rts, ORIENTATION =
orientation)$eff
  Dt2_t2 <- 1/dea(x2, y2, RTS = rts, ORIENTATION =
orientation)$eff
  A      <- matrix(Dt1_t1, ncol = 1)
  B      <- matrix(Dt2_t2, ncol = 1)
  original <- cbind(A,B)
  delta  <- cbind(rbind(A, 2-A, 2-A, A), rbind(B, B, 2-B, 2-B))
  ## Bandwith recomendado por Simar & Wilson (1999)
  h = (4/(5*n))^(1/6) ## antes: (4/(5*n))^(1/6) ### Erro encontrado
aqui !!!!

  ## Matrizes de correlação dos dados originais e dos reflexos
  sigma <- cov(cbind(A,B))
  sigma_r <- cov(cbind(2-A,B))
  C = matrix (1, ncol = 2, nrow = n)

  for (N in 1:NREP) {
    e <- matrix(NA, n, 2)
    index <- sample(1:(4*n), n, replace = TRUE)
    delta_star = delta[index,]
    delta_mean_matrix = diag(colMeans(delta_star))
    ## Fazendo o "random draw" da normal bivariada, utilizamos a
decomposição de Cholesky
    for (i in 1:n) {
      #####if ( index[i] <= n | index[i] > 3*n) { ##[A,B] ou [2-
A,2-B] ## ERRO AQUI!!!
      if ( index[i] %in% c(1:n, 31:45) ) {
        M <- t(chol(sigma))
        Z <- matrix(rnorm(2), 2, 1)
        bvn2 <- t(M %*% Z)
      } else {
        M <- t(chol(sigma_r))
        Z <- matrix(rnorm(2), 2, 1)
        bvn2 <- t(M %*% Z)
      }
      e[i,] <- bvn2
    }
  }
}

```

```

    gamma = (delta_star + h*e - C %*%
delta_mean_matrix)/sqrt(1+h^2) +
    C%*% delta_mean_matrix

    gamma_star = matrix (NA, n, 2)
    for (i in 1:n) {
      for(j in 1:2){
        if (gamma[i,j]>= 1) { gamma_star[i,j] = gamma[i,j]
        } else {
          gamma_star[i,j] = 2 - gamma[i,j]
        }
      }
    }

    x_star1 = (gamma_star[,1]/original[,1])*x1
    x_star2 = (gamma_star[,2]/original[,2])*x2

    ### VERIFICAR - POSSIVEL ERRO ARSAE !!!!!!!
    Dt1t1_boot = 1/dea(x_star1, y1, RTS = rts, ORIENTATION =
orientation)$eff
    Dt2t2_boot = 1/dea(x_star2, y2, RTS = rts, ORIENTATION =
orientation)$eff
    Dt1t2_boot = 1/dea(x_star1, y1, XREF = x_star2, YREF = y2,
RTS = rts,
    ORIENTATION = orientation)$eff
    Dt2t1_boot = 1/dea(x_star2, y2, XREF = x_star1, YREF = y1,
RTS = rts,
    ORIENTATION = orientation)$eff

    tal[,N] = sqrt(
((Dt2t1_boot/Dt2t2_boot)*(Dt1t1_boot/Dt1t2_boot)) )
    epsilon[,N] = Dt2t2_boot/Dt1t1_boot
  }
  Malm = epsilon * tal
  resultados <- list(malmquist = rowMeans(Malm),
    catch = rowMeans(epsilon),
    desloc = rowMeans(tal),
    malm.boot = Malm,
    catch.boot = epsilon,
    desloc.boot = tal)

  return (resultados)
}
## -----
-----

    modelo_Malmquist <- f.Malmquist.boot (x_12, y_12, x_14, y_14,
"crs", "in", NREP=2000)
    deslocamento_frenteira <- data.frame(prestadores,
    modelo_Malmquist$desloc,

t(apply(modelo_Malmquist$desloc.boot,1,quantile,c(0.025,0.975))))
  names(deslocamento_frenteira)<- c("Prestadores","Deslocamento da
Frenteira", "LI Deslocamento da Frenteira","LS Deslocamento da
Frenteira")

## -----
-----

```

ANEXO III – CÓDIGO R IMPLEMENTADO DE ACORDO COM O ARTIGO SIMAR WILSON (1999) – PARA VALIDAÇÃO

```
## Implementacao do algoritmo do artigo:
## Estimating and bootstrapping Malmquist indices
## Prof. Marcelo A Costa - 27/12/2016

# Limpar os dados do R
rm(list=ls(all=TRUE))

## O script utilizará os seguintes pacotes
if (!require(openxlsx)) { install.packages('openxlsx');
library(openxlsx) }
if (!require(Benchmarking)) { install.packages('Benchmarking');
library(Benchmarking) }
if (!require(svDialogs)) { install.packages('svDialogs');
library(svDialogs) }
if (!require(mvtnorm)) { install.packages('mvtnorm');
library(mvtnorm) }

## - - - - -
## Marcelo: NTCRFEF_40_2016_COE_Copasa_Banco_de_dados2
dlgMessage('Selecione o arquivo com as informações do Snis!')
base_dados <- read.xlsx(file.choose(), sheet = "Snis", startRow =
15)

## RETIRA A "CAESB"
## base_dados <- subset(base_dados, prest != "CAESB")

base_corte1 <- subset(base_dados, serv == 'Água e Esgoto')
base_corte2 <- subset(base_corte1, natjur != 'Autarquia')
base_corte2B <- subset(base_corte2, prest == 'COPASA')
econ_agua_med_copasa <- mean(base_corte2B$econativas_a)

econativas_a_med <- aggregate(econativas_a ~ prest,
data=base_corte2, FUN=mean)
names(econativas_a_med)[2] <- 'econativas_a_med'
base_corte2 <- merge(base_corte2, econativas_a_med, by='prest')
base_corte3 <- subset(base_corte2, econativas_a_med >=
econ_agua_med_copasa*0.1)

# Nível mínimo de hidrometração = 70%.
hidrom_med <- aggregate(hidrom ~ prest, data=base_corte3,
FUN=mean)
names(hidrom_med)[2] <- 'hidrom_med'

base_corte3 <- merge(base_corte3, hidrom_med, by='prest')

## base_final
base_final <- subset(base_corte3, base_corte3$hidrom_med >= 69.54)
## corrigido p/ "CAERN" que tem em 2012 hidrom_med == 69.55

## Construção das perdas (1000 m³/ano)
base_final$perdas = with(base_final, volprod_a + volimp_a -
volcons_a)
## write.csv2(base_final, "base_final.csv")

## Agrega a base para as médias
base_medias <- aggregate(cbind(opex_sem_imp_defl, econativas_a,
```

```

econativas_e, voledt, perdas) ~
prest, data=base_final, FUN=mean)

x <- as.matrix( base_medias[,c("opex_sem_imp_defl", "perdas")] )
y <- as.matrix( base_medias[,c("econativas_a", "econativas_e",
"voledt")] )
base_medias$DEA <- dea(X=x, Y=y, RTS="irs", ORIENTATION="in")$eff

dea_model_ub <- dea.boot(X=x, Y=y, NREP = 2000, EFF =
NULL, RTS = "irs")
base_medias$DEA_Unbiased <- dea_model_ub$eff.bc ## Bias Corrected
base_medias$DEA_Final <-
base_medias$DEA_Unbiased/max(base_medias$DEA_Unbiased)

## - - - - -
## Calculo do Indice de Malmquist
x.t1 <- subset(base_final, ano == 2012,
select=c("opex_sem_imp_defl", "perdas"))
y.t1 <- subset(base_final, ano == 2012, select=c("econativas_a",
"econativas_e", "voledt"))
x.t2 <- subset(base_final, ano == 2014,
select=c("opex_sem_imp_defl", "perdas"))
y.t2 <- subset(base_final, ano == 2014, select=c("econativas_a",
"econativas_e", "voledt"))

x.t1 <- as.matrix(x.t1)
y.t1 <- as.matrix(y.t1)
x.t2 <- as.matrix(x.t2)
y.t2 <- as.matrix(y.t2)

## t2 > t1
Dt1_t1 <- 1/dea(X=x.t1, Y=y.t1, RTS="crs", ORIENTATION="in")$eff
Dt2_t2 <- 1/dea(X=x.t2, Y=y.t2, RTS="crs", ORIENTATION="in")$eff

Dt2_t1 <- 1/dea(X=x.t2 ,Y=y.t2, RTS="crs", ORIENTATION="in",
XREF=x.t1, YREF=y.t1)$eff

Dt1_t2 <- 1/dea(X=x.t1 ,Y=y.t1, RTS="crs", ORIENTATION="in",
XREF=x.t2, YREF=y.t2)$eff

Eff.change <- Dt2_t2/Dt1_t1
Tech.change <- sqrt( ((Dt2_t1/Dt2_t2)*(Dt1_t1/Dt1_t2)) )
Malmquist <- Eff.change * Tech.change

## - - - - -
## Elementos do algoritmo
A <- Dt1_t1
B <- Dt2_t2
N <- length(A) ## numero de DMUs
C <- cbind(rep(1,N), rep(1,N))
## h <- ((4/5)*N)^(1/6)
h <- (4/(5*N))^(1/6) ## ARSAE
NREP <- 10000 ## Replicas bootstrap.

Delta <- rbind(cbind( A, B),
cbind(2-A, B),
cbind(2-A, 2-B),
cbind(A, 2-B))

```

```

Delta.type <- rep(c("A:B", "2-A:B", "2-A:2-B", "A:2-B"), each=N)

COV <- cov( cbind( A, B ) ) ## cov( cbind(2-A, 2-B) )
COV.R <- cov( cbind(2-A, B) ) ## cov( cbind(A, 2-B) )

Malmquist.boot <- matrix(NA, nrow=N, ncol=NREP)
diff.Malmquist.boot <- matrix(NA, nrow=N, ncol=NREP)

Eff.change.boot <- matrix(NA, nrow=N, ncol=NREP)
diff.Eff.change.boot <- matrix(NA, nrow=N, ncol=NREP)

Tech.change.boot <- matrix(NA, nrow=N, ncol=NREP)
diff.Tech.change.boot <- matrix(NA, nrow=N, ncol=NREP)

for(b in 1:NREP){
  ## First, we randomly draw with replacement N rows from
  ## Delta to form (N x 2) matrix Delta.star
  linhas <- sample.int(4*N, size=N, replace=TRUE)
  Delta.star <- Delta[linhas,]
  D.type <- Delta.type[linhas]
  delta.bar <- colMeans(Delta.star)

  ## Gerar eps.star
  eps.star <- matrix(NA, nrow=N, ncol=2)
  for(i in 1:N){
    if(D.type[i] %in% c("A:B", "2-A:2-B")){
      eps.star[i,] <- rmvnorm(n=1, sigma=COV)
    } else { ## in ("2-A:B", "A:2-B")
      eps.star[i,] <- rmvnorm(n=1, sigma=COV.R)
    }
  }

  ## Calcula o Gama
  Gama <- (Delta.star + h*eps.star -
C%*%diag(delta.bar))/sqrt(1+h^2) + C%*%diag(delta.bar)
  ## Aplica a equacao (28)
  auxiliar <- Gama >= 1
  Gama.star <- Gama * auxiliar + (2 - Gama) * (!auxiliar)

  ## Calcula os novos insumos
  x.t1.star <- cbind(Gama.star[,1], Gama.star[,1])*(x.t1/Dt1_t1)
  x.t2.star <- cbind(Gama.star[,2], Gama.star[,2])*(x.t2/Dt2_t2)

  ## Calcula o índice de Malmquist "bootstrap"...
  Dt1_t1.star <- 1/dea(X=x.t1.star, Y=y.t1, RTS="crs",
ORIENTATION="in")$eff
  Dt2_t2.star <- 1/dea(X=x.t2.star, Y=y.t2, RTS="crs",
ORIENTATION="in")$eff

  Dt2_t1.star <- 1/dea(X=x.t2.star ,Y=y.t2, RTS="crs",
ORIENTATION="in",
XREF=x.t1.star, YREF=y.t1)$eff

  Dt1_t2.star <- 1/dea(X=x.t1.star ,Y=y.t1, RTS="crs",
ORIENTATION="in",
XREF=x.t2.star, YREF=y.t2)$eff

  Eff.change.star <- (Dt2_t2.star/Dt1_t1.star)
  Tech.change.star <- sqrt(
((Dt2_t1.star/Dt2_t2.star)*(Dt1_t1.star/Dt1_t2.star)) )
  Malmquist.star <- Eff.change.star*Tech.change.star

```



```

Malmquist.boot[,b]      <- Malmquist.star
diff.Malmquist.boot[,b] <- Malmquist.star - Malmquist

Eff.change.boot[,b]    <- Eff.change.star
diff.Eff.change.boot[,b] <- Eff.change.star - Eff.change

Tech.change.boot[,b]   <- Tech.change.star
diff.Tech.change.boot[,b] <- Tech.change.star - Tech.change
}
## FIM ALGORITMO
## - - - - -
- -

## Processamento dos resultados p/ calculo de intervalos de confianca
"bootstrap"
## Equacoes (15) a (17)
## Malmquist
  b = - apply(diff.Malmquist.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.025))
  a = - apply(diff.Malmquist.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.975))

  base_medias$Malmquist      <- Malmquist
  base_medias$LI95.Malmquist <- Malmquist + a
  base_medias$LS95.Malmquist <- Malmquist + b

## Eff.change
  b = - apply(diff.Eff.change.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.025))
  a = - apply(diff.Eff.change.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.975))

  base_medias$Eff.change      <- Eff.change
  base_medias$LI95.Eff.change <- Eff.change + a
  base_medias$LS95.Eff.change <- Eff.change + b

## Tech.change
  b = - apply(diff.Tech.change.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.025))
  a = - apply(diff.Tech.change.boot, MARGIN=1, FUN=function(x)
quantile(x, probs=0.975))

  base_medias$Tech.change      <- Tech.change
  base_medias$LI95.Tech.change <- Tech.change + a
  base_medias$LS95.Tech.change <- Tech.change + b

## write.csv(base_medias, "saida.csv")

```

Bibliografia

Simar, Leopold, and Paul W. Wilson. "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models." *Management science* 44.1 (1998): 49-61.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-Economic Planning Sciences*, 22(4), 167-176.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21-37.

Johnson, A. L., & Kuosmanen, T. (2012). One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research*, 220(2), 559-570.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis. John Wiley & Sons.