Concorrência, Tecnologia e *Markups*: A Organização do Mercado de Água e Saneamento

Victor Gomes
Universidade de Brasília

9 de dezembro de 2024

Resumo

Neste artigo é quantificado o markup das empresas do setor de água e saneamento. As maiores empresas possuem markup elevado, enquanto que empresas pequenas não possuem markup maior do que um.

1 Introdução

Este estudo investiga o poder de mercado no setor de água e saneamento no Brasil, uma área de crescente relevância acadêmica e de política púplica. O objetivo deste trabalho é quantificar o poder de mercado das empresas do setor de água e saneamento no Brasil. A medida de poder de mercado comumente utilizada na literatura de organização industrial é o markup, que é razão entre preço e custo marginal. Essa métrica, amplamente utilizada na literatura de organização industrial, oferece um indicador preciso da capacidade de firmas em definir preços acima de seus custos marginais.

Diferentemente de métricas financeiras, o *markup* captura a lucratividade operacional na margem, sendo essencial para compreender a sustentabilidade econômica das operações e investimentos das firmas. Decisões financeiras, como a gestão de passivos e capital de giro, são mecanismos auxiliares que asseguram a continuidade das atividades empresariais, mas não substituem o papel central do *markup* como determinante da lucratividade estrutural.

O setor de saneamento no Brasil enfrenta desafios estruturais históricos, caracterizados por uma cobertura insuficiente de serviços e marcantes desigualdades regionais. Atualmente, apenas 56,0% da população têm acesso a redes de esgoto, o que equivale a aproximadamente 112,8 milhões de pessoas. As disparidades regionais são evidentes: enquanto o Sudeste apresenta uma cobertura de 80,9%, no Norte, o índice é de apenas 14,7%.[1] Esses indicadores refletem uma ausência de incentivos robustos para a ampliação da cobertura, com investimentos frequentemente direcionados para regiões já atendidas, como destacado pela Secretaria Nacional de Saneamento Ambiental (SNSA, 2023). Tal padrão pode perpetuar a exclusão de populações em áreas deficitárias.

O setor de saneamento no Brasil possui diagnóstico de baixa provisão de serviços, tanto de ligações à rede quanto de tratamento. As redes de esgotos atendem 56,0% da população total (112,8 milhões de habitantes) do Brasil. O maior valor do índice de atendimento total de esgoto é da macrorregião Sudeste (80,9%) e o menor, da macrorregião

[1] SNIS, p.70.

Norte (14,7%).[2] A sub-provisão é ligada diretamente a baixos incentivos para realização de investimentos. De acordo com a Secretaria Nacional de Saneamento Ambiental (SNSA, 2023), os investimentos não ocorrem onde existe maior deficit de cobertura e são concentrados em regiões com que já possuem redes de água e saneamento.

A principal hipótese em estudos sobre provisão de utilidade pública é do preço ser inferior ao custo marginal (Timmins, 2002). Mostro nesse artigo que grande parte das empresas do setor possuem markup mair do que 1, o que implica que o preço é superior ao custo marginal. Como destacado pela literatura de organização industrial,[3] a existência de markups elevados também está associada a existência de custos fixos, grande parte deles afundados, que são necessários para a oferta de produtos com custo marginal baixo (produtos tecnológicos, e.g.) e para manter barreiras à competição. Empresas de água e esgoto possuem investimentos significativos em ativos fixos, não são ativos tecnologicamente avançados mas são fixos e afundados. Sendo assim, para sustentar investimentos em infraestrutura as empresas devem possuir markup superior a unidade.

Existem duas implicações normativas da presença do markup (Syverson, 2024). A primeira está relacionada a perda do peso-morto, i.e. o produtor deixa de demandar mais unidades de insumo para ofertar para os consumidores não atendidos porque não seria lucrativo reduzir o preço das unidades inframarginais. No caso do saneamento, isto implica em incapacidade de universalização em função da existência de markups – naturalmente políticas públicas são adotadas para mitigar este efeito.

A segunda implicação normativa considera a distribuição do excedente. Na situação de concorrência imperfeita, o excedente é alocado para quem possui o maior poder de mercado, independentemente do tamanho da perda do peso-morto.

É importante deixar claro que a existência de markups leva a má alocação de recursos na presença de heterogeneidade de markups entre firmas. Dito de outra forma, se todos os produtores tem o mesmo markup pode existir perda do peso-morto mas não existe má-alocação. Markups diferentes entre produtores é o que à má-alocação, com perda de produtividade e bem-estar.[4]

A existência de markup significativo está associado a investimento em capital fixo. Firmas com significativos dispêndios em capital fixo precisam de markup para sustentar esta estrutura.[5]

Mudanças recentes na estrutura do setor, impulsionadas pela entrada de grandes grupos privados como Grupo Águas do Brasil, Aegea, Iguá e Brookfield, têm transformado o panorama da provisão de serviços. Essas transformações levantam questões fundamentais sobre a relação entre os níveis de investimento e a estrutura de mercado. Em particular, o presente estudo examina se o *markup* exerce um papel significativo na alocação de recursos e nos incentivos ao investimento no setor.

A literatura internacional oferece perspectivas relevantes. Coury et al. (2024) documentaram que a expansão da cobertura de água e saneamento em Chicago resultou em

[2] SNIS, p.70.

[3] Veja Sutton (1991), Berry, Gaynor, Morton (2019), Syverson (2024), entre outros.

[4] Veja por exemplo a análise de Boar e Midrigan (2024), que assumem um modelo em que o markup aumenta com o tamanho das firmas. Nesse caso, o problema do planejador social indica que firmas com markup mais baixo deveriam perder mercado para reduzir o problema de má-alocação (Syverson, 2024, p. 16).

[5] Como encontrado em diversas indústrias, os estudos encontram que grande parte dos investimentos são em custos fixos, em grande parte são de custo afundado que são construídos ao longo do tempo em redes, qualidade dos produtos, localização geográfica, etc. Veja as referências em Berry, Gaynor, e Morton (2019).

uma valorização das propriedades em 2,8 vezes entre 1874 e 1880. O benefício econômico dessa expansão foi estimado em 60 vezes o custo de construção das redes de saneamento.[6] A revisão bibliográfica conduzida por esses autores destaca a escassez de estudos empíricos que avaliem os efeitos econômicos e sociais da expansão da infraestrutura de saneamento.

No contexto brasileiro, Lucinda e Anuatti (2017) empregaram a função de custo translog[7] para avaliar a provisão de serviços de água e esgoto pela SABESP. Ao estimar custos marginais por município, os autores identificaram evidências modestas de economias de escala, sugerindo que a expansão da cobertura apresenta desafios específicos para a eficiência econômica das operações.

Este estudo busca avançar o entendimento sobre a interação entre *markups*, incentivos ao investimento e estrutura de mercado no setor de saneamento no Brasil. A análise oferece contribuições teóricas e empíricas ao debate regulatório, destacando as implicações econômicas do poder de mercado e suas consequências para a universalização dos serviços.

2 Mercado

2.1 Atores e Dados

Os dados utilizados neste estudo são provenientes do Sistema Nacional de Informações sobre Saneamento (SNIS), uma base de dados obrigatória, conforme estabelecido pela Lei nº 11.445/2007, art. 50, e reforçado pela Lei nº 14.026/2020. Essas legislações condicionam a alocação de recursos públicos federais e financiamentos com recursos da União à atualização regular das informações no SNIS.[8] A partir de 2024, o SNIS será substituído pelo Sistema Nacional de Informações em Saneamento Básico (SINISA), em conformidade com as atualizações do novo marco legal do saneamento. Este marco define saneamento básico como o conjunto de serviços públicos, infraestruturas e instalações operacionais relacionadas ao abastecimento de água potável, esgotamento sanitário, limpeza urbana e manejo de resíduos sólidos, além de drenagem e manejo de águas pluviais urbanas.

De acordo com a Lei nº 11.445/2007, cabe ao titular dos serviços a responsabilidade pelo planejamento, pela prestação e pela regulação dos serviços de saneamento. A prestação dos serviços pode ocorrer de forma direta, por meio de autarquias ou empresas públicas, ou de forma indireta, mediante concessões a empresas estatais ou privadas. Nos casos de concessão, o titular deve estabelecer metas de qualidade e custos para os serviços, supervisionados por um órgão regulador responsável por monitorar o desempenho.

O novo marco regulatório do saneamento, instituído pela Lei nº 14.026 de 15 de julho de 2020, trouxe mudanças significativas ao setor. Estabeleceu metas ambiciosas, como a universalização do acesso a 99% da população com água potável e a 90% com coleta e tratamento de esgoto até 31 de dezembro de 2033. Para atingir esses objetivos, foram disponibilizadas ferramentas legais ao poder público, incluindo: (i) Reconhecimento de municípios, regiões metropolitanas e o Distrito Federal como titulares dos serviços, responsáveis pela prestação direta ou pela condução de licitações para escolha de prestadores. (ii) Incentivo à prestação regionalizada por meio de consórcios intermunicipais, convênios

[6] Os autores utilizaram características geográficas como variáveis instrumentais, explorando a variação quase-aleatória nas elevações para identificar efeitos causais sobre a expansão das redes de esgoto e os impactos econômicos subsequentes.

[7] A estimativa foi conduzida com base no arcabouço metodológico de Evans e Heckman (1983).

[8] A Secretaria Nacional de Saneamento Ambiental, vinculada ao Ministério das Cidades, emite anualmente um atestado de regularidade para os prestadores que atualizam as informações no SNIS.

de cooperação ou blocos de referência instituídos pela União. (iii) Proibição da renovação de contratos de programa entre titulares e companhias estaduais, exigindo licitações ao término desses contratos. Além disso, as empresas estaduais podem estabelecer parcerias público-privadas, delegando partes dos serviços à iniciativa privada, enquanto mantêm foco em atividades nas quais possuem maior eficiência.

Para promover a concorrência em áreas regionalizadas, o novo marco atribuiu à Agência Nacional de Águas e Saneamento Básico (ANA) a competência para editar normas de referência, harmonizando a atuação das agências reguladoras regionais. Os participantes do mercado incluem empresas privadas, estatais de economia mista (com gestão pública ou privada) e autarquias municipais. Essas empresas podem ser classificadas em: (a) Autarquias e administração pública direta (geralmente municipais); (b) Empresas privadas; (c) Empresas públicas; (d) Sociedades de economia mista com administração privada; (e) Sociedades de economia mista com administração pública. As companhias estaduais de saneamento geralmente se enquadram nas duas últimas categorias.

2.2 Definição do Mercado de Serviço

A definição de mercado é essencial para compreender a organização econômica do setor. No passado, serviços de água e saneamento eram considerados monopólios naturais devido a duas características principais. *Altos custos fixos:* a necessidade de infraestrutura extensa inibia a entrada de novos competidores, conferindo à empresa incumbente economias de escala significativas. *Custos afundados:* grande parte da infraestrutura não possui valor em um mercado secundário, já que sua utilidade depende da operação.

Historicamente, o arcabouço legal conferia exclusividade aos prestadores de serviços para garantir receitas suficientes para a expansão da infraestrutura. Essa exclusividade buscava evitar problemas econômicos como o *hold-up,* em que a indefinição de propriedade de ativos essenciais reduz investimentos e deteriora a estrutura produtiva.

O novo marco regulatório introduziu mudanças que incentivam a concorrência. Embora o setor tenha características de monopólio natural, atualmente se entende que é possível haver competição por meio de licitações para concessões locais. Essa abordagem também é adotada internacionalmente, como no Reino Unido, onde a agência reguladora *Ofwat* implementou a contestabilidade no mercado de água e saneamento desde 2006. A experiência britânica demonstra que o mercado pode evoluir de monopólios regulados para um cenário com maior concorrência, especialmente em: (i) designação de novas empresas para subáreas de exploração de serviços; (ii) fornecimento de novas conexões à infraestrutura existente das incumbetes; e (iii) competição na cadeia de suprimentos. Na prática, a mudança regulatória pró-concorrência no Reino Unido levou a Ofwat a recomendação de uso estrito do aparato legal de defesa da concorrência. Definições de mercado, análise de concentração e condutas passaram a ser ferramentas aplicadas no setor de água e saneamento.

No Brasil, o CADE tem analisado fusões e condutas no setor de saneamento, considerando o mercado geográfico como nacional devido às nova dinâmica das concessões regionais. A concorrência ocorre por meio de licitações, permitindo que novas entrantes ameacem incumbentes, especialmente empresas públicas ou autarquias municipais que enfrentam desafios para sustentar investimentos.[9]

[9]Devido a natureza da licitação, é de se esperar que os novos concessionários tenham margem elevada suficiente para realizar investimentos em ativos fixos e arcar com o processo de entrada no mercado.

2.3 Produção e Estrutura do Setor

Uma vez concedida a operação, a empresa decide como produzir, utilizando tecnologias para o tratamento e entrega de água e esgoto. A produção adotada neste estudo é a quantidade produzida de água e esgoto. Abastecimento de Água Potável é definido como atividades que disponibilizam e mantém infraestruturas e instalações operacionais necessárias ao abastecimento público de água potável, desde a captação até as ligações prediais e seus instrumentos de medição. Esgotamento sanitário é a disponibilização e manutenção de infraestruturas e instalações operacionais necessárias à coleta, ao transporte, ao tratamento e à disposição final adequados dos esgotos sanitários, desde as ligações prediais até sua destinação final para a produção de água de reúso ou seu lançamento de forma adequada no meio ambiente.

Uma característica interessante do setor é que os preços não são livres e, portanto, a produção em quantidades é diretamente observada. Isto implica que o preço do serviço não é uma variável estratégica para as empresas que operam no mercado. As decisões estratégicas são onde entrar ou sair (mercado local) e qual o volume ótimo a ser ofertado.

O SNINS é uma base de dados censitária, pois todos os prestadores de serviços que operam nos municípios precisam fornecer as informações. A extração utilizada da base de dados é chamada de *desagregada*, que é organizada por prestador de serviço local, regional ou microrregional que atua nos municípios.

A base de dados utilizada, derivada do SNIS, inclui apenas prestadores com receitas superiores a R$ duzentos mil anuais, excluindo empresas da Região Norte devido às características distintas de custos operacionais e micromedição.[10] A análise das empresas é orientada para o prestador de serviço de água e esgoto. Eles estão organizados por natureza jurídica (com número entre parêntesis indicando a quantidade de entidades em 2022) da seguinte forma: *Administração direta (816):* Órgão de prefeituras (secretaria, departamentos entre outros). *Autarquia (478):* Com autonomia administrativa e patrimônio próprio e sob controle municipal ou estadual. *Sociedades de economia mista (30):* Com capital público e privado. Gestão pública ou com participação dos sócios privados. *Empresa Pública (5):* Formada por uma ou várias entidades com capital exclusivamente público. *Empresa Privada (132):* com capital majoritário ou integralmente privado. Administrada por particulares Organização Social. *Organização social (17):* entidade civil sem fins lucrativos com delegação para administrar serviços.[11]

Produção. A análise considera a produção agregada de água e esgoto, i.e. firmas produzem soma das quantidades de água e esgoto. A agregação ocorre por dois motivos, primeiro faz parte das concessões a oferta de serviço de água e saneamento e, segundo, como a base não possui muitas empresas pequenas, todas proveem serviço de saneamento e água potável. Além disso, esta agregação faz sentido pela evidência de economia de escopo na oferta de água e esgoto conjuntamente (veja Lucinda e Annuati, 2017).

[10] Empresas pequenas não reportaram todas as variáveis ao SNIS e como o processo de estimação é dinâmico, é preciso ao menos dois períodos de tempo de presença na base de dados.

[11] No SNIS os prestadores também são organizados por abrangência: que são local, microrregional e regional. O primeira é definida como o prestador que atende apenas um município, no microrregional o prestador atende a pelo menos dois municípios, enquanto a última categoria é do prestador que atende diversos municípios.

3 Mensurando Markup

O *markup* é uma das medidas mais diretas sobre do poder de mercado, refletindo a capacidade das firmas em obter margens determinando a lucratividade. O markup é a decisão de precificação na margem. Ao menos em tese, as firmas não possuem poder direto de determinar preços, mas podem escolher o custo marginal para dada tarifa de remuneração do serviço. O markup é definido como a diferença ou razão entre preço e custo marginal, i.e.

$$\mu = p/c,$$

tal que p é o preço do produto e c o custo marginal.

Neste estudo é utilizada duas formas de mensurar o markup: uma pela demanda ótima de insumos e outra pela economia de escala. A seguir explico brevemente os dois métodos.

3.1 Markup pela Demanda Ótima de Insumos

Este método denominado de abordagem da produção é baseado na minimização de custos de um insumo variável. Este método foi proposto por Hall (1988) e aplicado para firmas por De Loecker e Warzinsky (2012). A principal hipótese desta abordagem diz que em cada período de tempo o produtor minimiza os custos pela escolha ótima de um insumo livre de fricções. Este fator precisa ser de escolha estática, tal como um insumo comum.

Esta abordagem leva a seguinte expressão para calcular o markup usando dados de produção e custos:

$$\mu_j = \theta_j^V \frac{p_j q_j}{p_j^V x_j^V}, \tag{1}$$

com θ_j^V sendo a elasticidade do produto do insumo x^V. V aqui significa que o insumo deve ser variável. Esta equação se aplica para cada produtor j em cada período de tempo t.

A princípio se tem múltiplas condições de primeira ordem, um para cada insumo variável na produção, que pode resultar na expressão do markup. Uma vez que se escolhe qual insumo variável utilizar, é preciso computar a razão da receita da firma j sobre o custo total do insumo X^V e multiplicar pela elasticidade da produção do referido insumo.

De Loecker e Syverson (2021, seção 7) detalham que a derivação padrão assume que as firmas tomam os preços dos insumos como dados. Não é necessário hipóteses adicionais sobre markups na oferta de insumos. A abordagem de produção também pode acomodar desvios da hipótese da firma ser tomadora de preços, considerando múltiplos insumos variáveis. A ideia da abordagem da produção para o markup sublinha que o custo marginal de produção é derivado de um único insumo variável, sem impor qualquer elasticidade de substituição particular em relação a outros insumos e sem assumir retornos de escala.

Para estimar o markup por este método é preciso se obter a elasticidade θ_j^V, que foi obtida pela estimação da função de produção.

3.2 Markups e Elasticidade de Escala

Syverson (2024, seção 1.2)[12] apresenta o cálculo do markup a partir da economia de escala. Este método pode ser uma forma alternativa de mensurar markups ou ao menos para checar a estimativa realizada por outra alternativa. Ele argumenta que esta decomposição

[12] Veja também Syverson (2019).

possui menos hipóteses do que as demais, mas por outro lado é baseado em conceito que é difícil de ser mensurado.

A derivação começa com a divisão e multiplicação do markup pelo custo médio, CM:

$$\mu_j \equiv \frac{p_j}{c_j} = \frac{p_j}{CM_j}\frac{CM_j}{c_j}. \tag{2}$$

Em seguida, $\frac{p_j}{CM_j}$ é multiplicado e divido pelo produto para chegar a razão das receita em relação custo total.[13]

A economia de escala, ϵ_{Cq}, pode ser escrita como a razão entre o custo médio e o custo marginal, que é o termo CM_j/c_j. De forma equivalente, este termo é o inverso da elasticidade do custo em relação a quantidade, que por sua vez é $\frac{\partial C}{\partial q}\frac{q}{C} = c\frac{1}{CM}$.[14] Quando o custo marginal é menor do que o custo médio, o CM está caindo na quantidade e a elasticidade da escala é maior do que um. Neste caso podemos esperar uma indústria mais concentrada quanto maior for a economia de escala. Se $c > CM$, então existem deseconomias de escala e o termo de elasticidade ϵ_{Cq} é menor do que um.

Portanto, o markup pode ser diretamente relacionado a economia de escala, ϵ_{Cq}, a relação da receita R_j com o custo total C_j:

$$\mu_j \equiv \frac{p_j}{c_j} = \frac{R_j}{C_j}\epsilon_{Cq}. \tag{3}$$

Syverson (2024) observa que esta relação se aplica a condições gerais, pois tudo que é requerido é a diferenciação da função que relaciona produto aos custos. Esta função nem mesmo precisa ser uma função de custo padrão de teoria da produção. Isto quer dizer que ela não precisa ser o custo total calculado para a demanda dos fatores minimizadoras de custo (que é uma hipótese utilizada no método da demanda ótima por insumos).

O aumento no markup expande a receita em relação aos custos (pelo menos) na unidade marginal. Esta receita extra deve levar a lucros mais elevados ou ser usada para pagar custos das unidades inframarginais quando o custo médio excede o custo marginal. De forma equivalente, se o processo de produção involve economia de escala na quantidade maximizadora de lucros, o produtor deve pagar de alguma forma o excesso do custo médio sobre o custo marginal. Se o preço é igual ao custo marginal não seria possível sustentar esta estrutura de custos, sendo assim, é necessário existir receita extra para sustentar a estrutura de custos.

A relação (**??**) implica em variações do markup e outros objetos de interesse. Se por exemplo, produtores são conhecidos por ter crescimento substancial do markup, tanto lucros puros ou a elasticidade de escala devem aumentar. Se por acaso a razão dos lucros é constante, precisa existir variação na economia de escala proporcional ao markup. Esta é a intuição econômica que se pode ter a partir da equação do markup (**??**).[15]

[13] $\frac{p_j}{C_j/q_j}\frac{q_j}{q_j} = \frac{p_j q_j}{C_j}$ sendo C_j o custo total da firma j e o numerador é naturalmente a receita total.

[14] Para uma função de produção homotética, a elasticidade de escala é igual ao retorno de escala da função de produção. Obs: qualquer função homogênea é uma função homotética.

[15] Aqui cabe uma discussão sobre estudos que observam dados de margem diretamente das firmas e/ou produtos. Pontos importantes sobre este tipo de mensuração levantados por Pakes (2018) são: o que deve ser incluído na margem? A mensuração pode variar muito e em casos de antitruste, onde se deseja avaliar sinergias, existem razões diferentes para incluir dados específicos de custos. Mas exatamente o que se deseja quando olhamos para a margem é comparar preços com custo marginal, e a informação direta deveria incluir depreciações e ativos fixos. O que é razoável de se incluir nos custos depende da flutuação do volume de vendas (quantidades) de um ano para outro. O problema no caso de avaliação de sinergias é que se deseja mais informações do que geralmente é capturado neste de tipo de informação.

4 Estimação da Função de Produção

Ambos os métodos utilizam elementos que são calculados pela estimativa da função de produção utilizando dados das firmas que operam no setor. Portanto, especificações e estrutura dos dados afetam diretamente o cálculo do markup.

Como descrito anteriormente a produção é mensurada pelos volumes de água e esgoto, isto é, são quantidades e não receita. Quando se somam os dois volumes implicitamente a mesma tecnologia transformadora para os dois. Esta hipótese é baseada na evidência empírica sobre existência de economia de escope em produzir e distribuir água e esgoto.[16] O uso de quantidade para a produção guia o método adotado para estimar a função de produção, como sera descrito a seguir, a estratégia empírica adotada é similar a proposta por DLGKP.

A abordagem de DLGKP segue o método da "função controle" de Olley e Pakes (1996, doravante OP) e aprimorada pelos trabalhos de Levinsohn e Petrin (2003, LP) e Ackergerg, Caves e Frazer (2015, ACF). A função de produção a ser estimada é:

$$Q_{jt} = Q^*_{jt} \exp(\varepsilon_{jt}) \tag{4}$$

$$Q^*_{jt} = F(K_{jt}, L_{jt}, M_{jt}, \theta) \exp(\omega_{jt})$$

tal que Q_{jt} é a produção da empresa j no tempo t, usando capital, K, trabalho, L, e materiais, M. ω_{jt} é o termo de produtividade observada pela firma no período t, mas não pelo econometrista e θ é o conjunto de parâmetros que precisam ser estimados. Seguindo OP, outras hipóteses sobre a produtividade são as que ela é Hicks-neutra e segue uma lei de movimento, tal que ω_{jt} é governada por um processo de Markov de primeira ordem:

$$\omega_{jt} = \mathrm{E}(\omega_{jt} \mid \omega_{jt-1}) + \xi_{jt} = g(\omega_{jt-1}, \cdot) + \xi_{jt} \tag{5}$$

tal que $g(\omega_{jt-1})$ é a produtividade esperada e ξ_{jt} é inovação sobre o nível de ω, i.e.,

$$\xi_{jt} = \omega_{jt} - g(\omega_{jt-1}, \cdot). \tag{6}$$

Nesta abordagem, o termo estocástico ε_{jt} contabiliza pela diferente entre a produção realizada pela firma, Q_{jt}, e o produto planejado, Q^*_{jt}, dada a escolha de insumos. O termo estocástico não é correlacionado ao longo do tempo e nem aos insumos. No caso, do setor de saneamento ele esta associado a paradas não esperadas nas redes de água ou esgoto, por exemplo. Uma vez que firmas e economistas não observam ε_{jt}, a produção planejada Q^*_{jt} também não é conhecida. Como a produção realizada é observada, controlar pelos choques não esperado equivale a estimar a produção planejada como $Q^*_{jt} = Q_{jt}/\exp(\varepsilon_{jt})$.

Existência de viés na estimação da função de produção. A estimativa da função de produção não deve ser realizada pois existem alguns vieses ao se utilizar métodos de regressão tradicionais, tais como MQO. A produtividade não-observada ω_{jt} em (4) leva a viés de simultaniedade e de seleção. A presença destes dois vieses é o foco predominante da literatura de estimação da função de produção de OP/LP/ACF.

Se o economista possui dados de produção e insumos em quantidades, estes procedimentos de "função controle" são suficientes para estimar de forma consistente os parâmetros da função de produção. Todavia, na base de dados do SNIS que é utilizada

[16] Veja Lucinda e Anuatti (2017) e as referências que eles citam sobre economias de escopo no setor.

para estimar a função de produção não é possível obter quantidades físicas de materiais.[17] Embora se tenha quantidades da produção e dos demais fatores, insumos intermediários são deflacionados por um índice de preços setorial. Isto não é classificado apenas como um problema de mensuração, uma vez que as firmas tipicamente usam insumos diferentes para ofertar os serviços de água e esgoto. Portanto, o produto físico não pode ser diretamente comparável ao insumo deflacionado. Por exemplo, a SABESP pode utilizar produtos químicos mais eficientes e provavelmente mais caros do que um departamento de água e esgoto municipal.

Para entender a implicação deste viés represente $\tilde{v}_{jt}$ como sendo o log do dispêndio com um insumo v qualquer deflacionado por um índice de preços setorial. O log do dispêndio com o insumo v_{jt} corretamente deflacionado é $v_{jt} = \tilde{v}_{jt} - w^v_{jt}$, com w^v_{jt} representando o desvio do (log) preço não-observado do insumo da firma j no período t em relação ao (log) do deflator agregado. O problema aqui é que sem o controle adequado w^v_{jt} se confunde com a produtividade ω_{jt}, causando novo viés na medida de quantidade da produtividade. Substituindo a fórmula para v_{jt} na função de produção e definindo $\mathbf{w}_{jt}$ como o vetor dos (log) preços específicos dos insumos, a versão log-linear da função de produção (4) é:

$$q_{jt} = f(k_{jt}, l_{jt}, m_{jt}, , \theta) + B\left(\mathbf{w}_{jt}, \tilde{m}_{jt}, \theta\right) + \omega_{jt} + \varepsilon_{jt}, \tag{7}$$

com as letras em minúsculas representando o log natural das variáveis. $B(\cdot)$ captura o efeito dos preços não-observados dos materiais e leva a outro viés na estimativa da função de produção. Na base de dados do SNIS é possível observar quantidades de trabalho mas não é possível observar quantidades de produtos químicos, carros, computadores, etc. Portanto, dispêndio de materiais foi agregado e deflacionado pelo índice de preços ao consumidor.

A solução deste viés faz parte da inovação proposta por DLGKP ao procedimento usual de função controle. Na abordagem deles, a variação de preços específico dos insumos por surgir por característica local (capturado por $\tilde{G}_j$) e/ou por variação na qualidade dos insumos (ν_{jt}). Isto implica que duas firmas com a mesma característica geográfica operacional apenas possuem os mesmos preços de insumos se eles tem a mesma qualidade/especificação. DLGKP se baseiam em modelo formal para fundamentar a solução de controlar não-parametricamente o preço não-observado dos insumos pelo preço do produto final e por características geográficas. A racionalidade é que preço do produto final contém informação sobre preços dos insumos.[18] Eles também assumem complementariedade: produção de maior qualidade requer combinar insumos de maior qualidade com trabalho e capital mais qualificados, significando que os preços de todos os insumos podem ser expressados como função de um único índice de qualidade.[19]

Dado que preços dos insumos são função crescente da qualidade dos insumos, que por sua vez são relacionados com maior qualidade na oferta, DLGKP usam variáveis para aproximar o índice de preço correto dos insumos. Formalmente eles escrevem os preços dos insumos como função da qualidade do produto ν_{jt} e da localização das firmas $\tilde{G}_j$:

$$w^v_{jt} = w_t(\nu_{jt}, \tilde{G}_j). \tag{8}$$

Entre os controles utilizados por DLGKP estão preço do produto final, market shares, características geográficas, dummy de empresa exportadora e dummies de tipos de produtos. Seguindo o conceito de complementariedade adotado no modelo de DLGKP, o preço

[17] Não é observado a quantidade física de produtos químicos, por exemplo. Nem é possível recuperar a quantidade a partir de deflatores específicos para produtos químicos utilizados no processo produtivo.

[18] Kugler e Verhoogen (2011) documentam que produtores de bens mais caros usam insumos mais caros.

[19] Não significa que eles adotam diferenciação horizontal e não vertical.

do produto final foi substituído pelos impostos pagos. Na base de dados do SNIS impostos pagos são mais estáveis além de serem considerados variáveis exógenas.[20] Como controle geográfico foi utilizado apenas a distância entre um município e a capital do estado em km. Para concessionárias com mais de um município foi calculada a média ponderada da distância entre a sede do município e a capital.[21] Devido ao caracter de homogeneidade do produto e a natureza da concessão, utilizamos número menor de controles.

A função controle não-paramétrica para o preço dos materiais proposta por DLGKP é

$$B\left(\mathbf{w}_{jt}, \tilde{m}_{jt}, \theta\right) = B\left((\tau_{jt}, \tilde{G}_j)' \times [1, \tilde{m}_{jt}]; \theta, \delta\right). \tag{9}$$

Esta equação é substituída na função de produção (7). A função $B(\cdot)$ é diferente da função de preços $w_t(\cdot)$ em (8). Ela contém os elementos de $w_t(\cdot)$, mas também possui termos de interações com o vetor deflacionado de materiais ($\tilde{m}_{jt}$). Por isso é usado o termo $[1, \tilde{m}_{jt}]$. A notação também deixa claro que o uso da função controle requer que se estime vetor adicional de parâmetros δ em conjunto com os parâmetros da função de produção θ.

4.1 Função Controle

A única fonte de viés potencial na função de produção é a produtividade ω_{jt} após a correção do viés de preços de insumos. O procedimento de estimação OP/LP começa com a estimação do primeiro estágio onde se deseja controlar por choques não esperados sobre a produção. A hipótese crucial utilizada por OP/LP é de que a demanda por investimento, no caso OP, ou a demanda por materiais, na ideia de LP, é função da produtividade ω_{jt} e de variáveis de estado $\mathcal{S}_{jt}$. As variáveis de estado da firma j são

$$\mathcal{S}_{jt} = \left\{K_{jt}, \exp(\omega_{jt}), \tilde{G}_j, \tau_{jt}, \tilde{D}_{jt}\right\}$$

Além dos elementos da função de produção, demais fatores exógenos são considerados variáveis de estado. Estes fatores são variáveis geográficas do municípios que as empresas possuem operações, $\tilde{G}_j$, a alíquota de imposto efetivamente pago, τ_{jt}, e *dummy* para indicar se a empresa atua em única UF (unidade da Federação), $\tilde{D}_j t$.

A função utilizada por LP/ACF é de que a demanda por materiais, $\tilde{m}_{jt}$, é

$$\tilde{m}_{jt} = m_t(\omega_{jt}, \mathcal{S}_{jt}) \tag{10}$$

tal que v_t é estritamente monotônica no escalar não-observável ω_{jt}. No caso original de OP, eles utilizam equação similar para a demanda por investimento. Se as condições que estão presentes na equação (10) valem, então ela pode ser invertida para identificar ω_{jt}. Isto é

$$\omega_{jt} = h_t(\tilde{m}_{jt}, \mathcal{S}_{jt}). \tag{11}$$

[20] Veja Dearing (2022).

[21] Também foi estimado uma versão do modelo incluindo a média da altura de cada município. De acordo com Coury et al (2024), o custo de construção de rede de esgoto depende de variações na elevação, e tais mudanças na altura aumentam o custo de investimento. Em cidades com muita variação de altitude é mais custoso o investimento em saneamento. Entretanto, esta variável em conjunto com a distância das capitais não foi muito importante como controle, assim optamos por manter apenas distância da capital para reduzir o número de momentos a serem estimados. Também não foi incluindo o market share uma vez que existe pouca competição por preço, isto é, as concessionárias não podem ganhar mercado das concorrentes.

Substituindo ω_{jt} na equação de função de produção (7), excluindo $\tilde{m}_{jt}$ para evitar colinearidade com o termo da função de produção e representando as minúsculas como log natural, se tem o primeiro passo para a estimação OP/LP/ACF:

$$q_{jt} = F(k_{jt}, l_{jt}, \tilde{m}_{jt}) + B\left(\mathbf{w}_{jt}, \tilde{m}_{jt}, \theta\right) + h_t(\tilde{m}_{jt}, \mathcal{S}_{jt}) + \varepsilon_{jt}. \tag{12}$$

Assumindo que ε_{jt} não é correlacionado com os demais termos e estimando a equação (12) resulta na estimativa de $\phi(\cdot)$ e ε_{jt}: $\omega_{jt} = \hat{\phi}_{jt} - F(\cdot) - B(\cdot)$. Como comentado por DJ, esta equação separa o produto realizado do produto planejado ($q^*_{jt} = \phi(\cdot)$).

Como discutido em OP, DLGKP e Ackerberg e De Loecker (2024), a abordagem da função controle não requer conhecimento da estrutura de mercado de insumos. Esta equação apenas diz que a demanda por insumos depende das variáveis de estado e das variáveis que afetam a demanda por insumos. Ao se usar o controle estático como *proxy* para a produtividade, não se tem que revisitar o modelo dinâmico e provar invertibilidade em comparação com a abordagem que utiliza insumos como controle quando se inclui demais variáveis de estado exógenas.

Um passo importante no processo de estimação é a lei de movimento da produtividade, equação (5). A análise da base de dados mostra que raramente as empresas deixam a base de dados, exceto quando perdem uma área de concessão. São poucas observações, menos de 3% do painel, mas podem causar viés de seleção. Para corrigir para este problema foi utilizada a estimativa não-paramétrica de OP. Além de resolver o problema de viés de seleção o uso desta correção também torna o procedimento de função controle mais eficiente.

Defina uma função indicadora χ_{jt} ser igual a um se a firma continua ativa e zero se ela deixa o mercado. Faça $\underline{\omega}_{jt} = \underline{\omega}_t(k_{jt}, \phi_j)$ seja o mínimo de produtividade que uma firma precisa ter para se manter ativa. Tradicionalmente a regra de seleção é escrita como

$$\begin{aligned} Prob(\chi_{jt} = 1) &= Prob\left[\omega_{jt} \geq \underline{\omega}_t(\cdot) \mid \underline{\omega}_t(\cdot), \omega_{jt-1}\right] \\ &= \kappa_{t-1}(k_{jt-1}, i_{t-1}, \phi_j) \equiv \mathbf{P_{jt}}. \end{aligned} \tag{13}$$

É utilizado o fato de que a produtividade mínima em t é prevista utilizando as variáveis de estado da firma, $\mathcal{S}_{jt}$, em $t-1$. Como em OP, existem dois índices de heterogeneidade das firmas: a produtividade e a produtividade mínima de operação. Portanto, se $\mathbf{P_{jt}} = \kappa_{t-1}(\omega_{jt-1}, i_{t-1}, \underline{\omega}_{jt}, \phi_j)$ então $\underline{\omega}_{jt} = \kappa_{t-1}(\omega_{jt-1}, i_{t-1}, \mathbf{P}_{jt}, \phi_j)$.

O procedimento de saída do mercado afeta diretamente a regra de movimento da produtividade (5). Isto significa que a dinâmica da produtividade vale enquanto a empresa continua no mercado, implicando que um dos argumentos da dinâmica da produtividade deve ser $\underline{\omega}_{jt}$. Em termos práticos, $\underline{\omega}_{jt}$ é substituído por $\mathbf{P}_{jt}$. De Loecker (2013) diz que fatores que podem afetar a dinâmica da produtividade devem entrar na regra de movimento (5) mesmo não tendo certeza do impacto sobre a produtividade. Se eles não forem importantes a dinâmica da produtividade não será afetada. Por outro lado, a omissão destas variáveis pode viesar o procedimento de estimação. Portanto, a regra de dinâmica da produtividade é alterada para incluir o valor mínimo de produtividade para cada firma, bem como outras variáveis de estado que podem afetar ω_{jt}:

$$\omega_{jt} = g(\omega_{jt-1}, \mathbf{P}_{jt}, \tilde{G}_j, \tau_{jt}, \tilde{D}_{jt}) + \xi_{jt}. \tag{14}$$

Variáveis de estado que não são endógenas geralmente podem afetar a dinâmica da produtividade. Por exemplo, mudanças na alíquota de imposto estadual pode afetar o crescimento da firma, portanto afetando ω_{jt}. As empresas privadas a partir do novo marco

regulatório tem crescido mais do que empresas regionais, nesse sentido a dummy $\tilde{D}_j t$ controla pela dinâmica deste tipo de empresa. Observe que a dummy $\tilde{D}_j t$ captura empresas que operam em única UF, logo empresas de capital privado que são pequenas não tem a mesma dinâmica das empresas que operam em várias UFs. Por outro lado, se a empresa não-regional adquire mais direitos de operações em municípios mais distantes das capitais, ela pode experimentar redução no crescimento da produtividade.

Momentos para estimação. O terceiro estágio é uma estimação GMM com os seguintes momentos:

$$\mathrm{E}\left(\xi_{jt}(\theta,\delta)Z \mid I_{t-1}\right) = 0, \tag{15}$$

com Z sendo a matriz dos instrumentos, e os momentos são condicionais ao conjunto de informação do período anterior I_{t-1}. Os objetos que formam a equação de momentos (15) são:

$$\xi_{jt} = \omega_{jt} - \rho R_t(\omega_{jt-1}, \hat{\mathbf{P}}_{jt}, \tilde{G}_j, \tau_{jt}, \tilde{D}_j t), \tag{16}$$

reescrevendo a (14) com estrutura autoregressiva de ordem um (ver ACF) e os demais elementos seguem abaixo

$$\omega_{jt} = \hat{\phi}_{jt} - F(k_{jt}, l_{jt}, \tilde{m}_{jt}) - B\left(\mathbf{w}_{jt}, \tilde{m}_{jt}, \theta\right), \tag{17}$$

$$\omega_{jt-1} = \left(\hat{\phi}_{t-1} - F(k_{jt-1}, l_{jt-1}, \tilde{m}_{jt-1}) - B\left(\mathbf{w}_{jt-1}, \tilde{m}_{jt-1}, \theta\right)\right). \tag{18}$$

$$R_t(\omega_{jt-1}, \cdot) = (\gamma_0 + \gamma_1 \omega_{jt-1}\hat{\mathbf{P}}_t + \gamma_2 \omega_{jt-1} T_j + \gamma_3 T_j + \gamma_4 \tilde{G}_j + \gamma_5 \tau_{jt} + \gamma_6 \tilde{D}_{jt-1}). \tag{19}$$

T_j é uma tendência temporal para cada firma.

O termo de persistência da regra de movimento da produtividade, ρ, não é estimado com os momentos GMM. Como explicado em ACF, a regressão que estima ρ é incluída dentro do algoritmo GMM, mas é uma estimação polinomial por mínimos quadrados ordinários. Os parâmetros γ não são utilizados na análise, mas os demais regressores podem alterar a estimativa de ρ.

Os momentos a serem estimados são os necessários para obter os parâmetros da função de produção, θ, e os necessários para função controle de preço dos insumos, δ. Os parâmetros da função controle são importantes para identificar θ, mas não são utilizados.

Formas funcionais. As formas funcionais utilizada na estimação da função de produção são descritas a seguir. O primeiro passo da função controle é estimado utilizando polinômio de terceira ordem para os fatores de produção. Os demais fatores exógenos e controles de tempo entram de forma linear. O modelo probit é utilizado no segundo passo e contém apenas as variáveis de estado $\mathcal{S}_{jt-1}$.

A forma funcional para o controle $B(\cdot)$ é a mesma da (9) enquanto que para a função de produção $F(\cdot)$ é utilizada a *translog*. A *translog*, que foi introduzida por Christensen, Jorgenson, e Lau (1973), é uma aproximação de segunda ordem de qualquer função de produção geral. A forma funcional dela não gera problemas de identificação. A hipótese crucial aqui é a produtividade ser Hicks-neutra.[22] De Loecker e Warzynski (2012) e DLGKP advogam o uso da *translog* uma vez que ela possui flexibilidade e produz elasticidades do produto que são especificas para cada empresa (embora os parâmetros sejam

[22]Diferente da função Cobb-Douglas, a translog não requer a hipótese de substituição estável entre os fatores produtivos.

constantes ao longo do tempo e entre empresas).[23] Por exemplo, firmas grandes podem ter elasticidades diferentes das pequenas.

A especificação translog adotada para $f(k, l, m)$, sem os subscritos t e j, tem a seguinte forma

$$f(k,l,m) =$$
$$\beta_k k + \beta_l l + \beta_m \tilde{m} + \beta_{kk} k^2 + \beta_{ll} l^2 + \beta_{mm} \tilde{m}^2 + \beta_{km} k.\tilde{m} + \beta_{kl} k.l + \beta_{lm} l.\tilde{m} + \beta_{klm} k.l.\tilde{m} + \omega. \quad (20)$$

Com esta forma funcional a elasticidade do produto em relação a cada insumo depende do nível de todos os insumos, lembrando que os termos em $B(\cdot; \delta)$ são usados para controlar pela ausência dos preços dos insumos. Isto faz que firmas diferentes tem elasticidades diferentes para o mesmo conjunto de parâmetros. Por exemplo, a elasticidade do produto em relação aos materiais (m) para uma dada firma é:

$$\hat{\theta}^m = \hat{\beta}_m + 2\hat{\beta}_{mm} m + \hat{\beta}_{km} k + \hat{\beta}_{lm} l + \hat{\beta}_{klm} k.l. \quad (21)$$

A elasticidade do produto utilizando a função de produção *translog* muda ao longo do tempo devido a mudanças nos fatores m, k e l. Por sua vez, os retornos de escala são calculados como a soma das elasticidades para capital, trabalho e materiais.

4.1.1 Hipóteses sobre *Timing*

A relação temporal da variáveis no modelo é determinante da estimação e, naturalmente, dos resultados. Na função controle do primeiro estágio todas as variáveis são contemporâneas em relação a produção.

Na regra de movimento da produtividade, (14), a produtividade é relacionada com a sua defasagem. Como se trata de concessão de serviços, as demais variáveis são conhecidas contemporaneamente: área de concessão, impostos e localização.[24]

Assumimos que em cada período todos os insumos produtivos são fixos, i.e., a firma não ajusta os seus volumes no ano corrente. Devido a estabilidade e previsibilidade da demanda, as empresas possuem contratos fixos com trabalhadores e outros materiais. DLGKP para empresas grandes da manufatura da Índia assuem que capital e trabalho são fixo, mas materiais são de ajuste livre em t. Se assume que insumos dinâmicos não são correlacionados com a inovação no termo de produtividade ξ_{jt}.[25]

O capital é tipicamente assumido por ser um fator dinâmico em virtude da regra de movimento do estoque de capital.[26] O trabalho assumimos ser um fator dinâmico pois as empresas do setor são estatais e as privadas oferecem contratos para os funcionários. É assumido que materiais também é um fator dinâmico, uma vez que o consumo de fatores intermediários depende da demanda, que é estável.

Outra questão importante é de como variáveis que entram na função controle $B(\cdot)$ são identificadas. A variável não-geográfica que entra na dinâmica da produtividade e na controle dos preços de insumos é a tarifa. Na produtividade ela entra defasada e em $B(\cdot)$ é contemporânea. Como discutido anteriormente, impostos defasados são bons instrumentos para o lado da demanda e devido a alguma potencialmente correlação com ξ_{jt} as momentos com impostos são formados com uma defasagem ($\xi_{jt}.\tau_{jt-1}$, etc.).

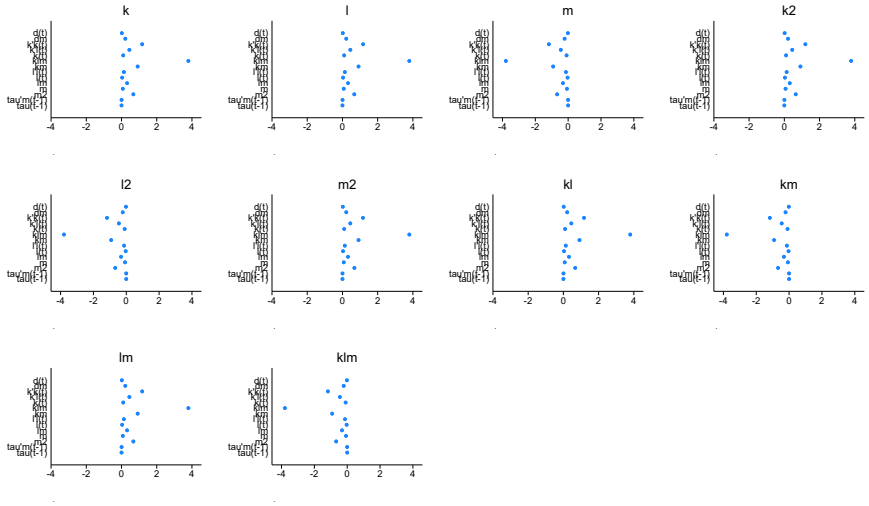
[23] Veja também Raval (2023).

[24] Incluir defasagem em impostos não alterou significativamente o resultado, por exemplo.

[25] Aqui pode existir discussão relevante de problema de mensuração do estoque de capital. Veja De Loecker e Syverson (2021), subsection 5.6.1 e a referência citada.

[26] $k_{t+1} = (1 - \underline{\delta})k_{t-1} + i_{t-1}$

Figura 1: Sensitividade dos Momentos



4.2 Identificação

Na Figura 1 são apresentadas as sensibilidades dos momentos da estimação da função de produção de acordo com a definição de Andrews, Gentzkow, e Shapiro (2017). A sensibilidade dos momentos captura o quanto que uma violação em uma condição de exclusão pode provocar viés nos parâmetros estimadas.

O momento com maior sensitividade é o associado ao termo $k.l.\tilde{m}$. Em geral, momentos que incluem o estoque de capital apresentam sensitividade em relação aos demais momentos. Por exemplo, esta sensitividade pode ensejar estudos similares que modelem o markup com modelos que usem outras medidas de capital como variáveis instrumentais.

5 Elasticidades e Retornos de Escala

Aqui são discutidos resultados sobre as estimativas das elasticidades e da economia de escala.[27] Como discutido anteriormente, uma característica interessante da função *translog* é a que a variação das elasticidades e da economia de escala são por produtor. Assumindo uma função de produção Cobb-Douglas com parâmetros constantes ao longo do tempo e entre empresas, a elasticidade resultante seria a mesma entre todos os produtores.

Estatísticas das elasticidades entre produtores são apresentadas na Tabela 1. As estatísticas são: média, mediana, percentil 25 e 75. As elasticidades dos fatores trabalho e intermediários são similares, exceto para o percentil 25, com trabalho maior. A elasticidade do capital é a menor, mas a diferença entre o percentil 25 e 75 é o dobro e este fato tem implicações para a medida de economia de escala.

Na Tabela 2 são apresentadas a média, mediana, percentil 25 e 75 dos retornos de escala para três períodos: 2016, 2020, 2022. Na média o retorno de escala tem aumentado, mas com retornos decrescentes de escala. A mediana é menor do que a média, o que mostra que existem valores elevados acima da média. Como os valores do percentil 75 tem aumentado,

[27] Não discuto como o procedimento da função controle de OP/LP/ACF melhora a estimação em relação a outros méotodos – OLS, painel dinâmico, etc. Para uma comparação dos resultados da modelagem de DLGKP com outros métodos veja a seção 4 do artigo.

Tabela 1: Elasticidades Estimadas

Estatísticas	$\hat{\theta}_k$	$\hat{\theta}_l$	$\hat{\theta}_m$
Média	0.0947	0.4129	0.4037
p25	0.0617	0.3502	0.2984
Mediana	0.0923	0.4153	0.4054
p75	0.1298	0.4735	0.4731

Tabela 2: Retornos de Escala

Estatísticas	2016	2020	2022
Média	0.9063	0.9124	0.9354
p25	0.8525	0.8438	0.8607
Mediana	0.8927	0.8930	0.9004
p75	0.9617	0.9651	0.9865
Min	0.7435	0.7590	0.7805
Max	1.2254	1.2142	1.2658
N	45	51	52

mas ainda inferiores a 1, as empresas com economia de escala estão concentradas acima deste percentil. Portanto, embora exista economia de escala, a maioria das firmas operam com retornos decrescentes de escala.

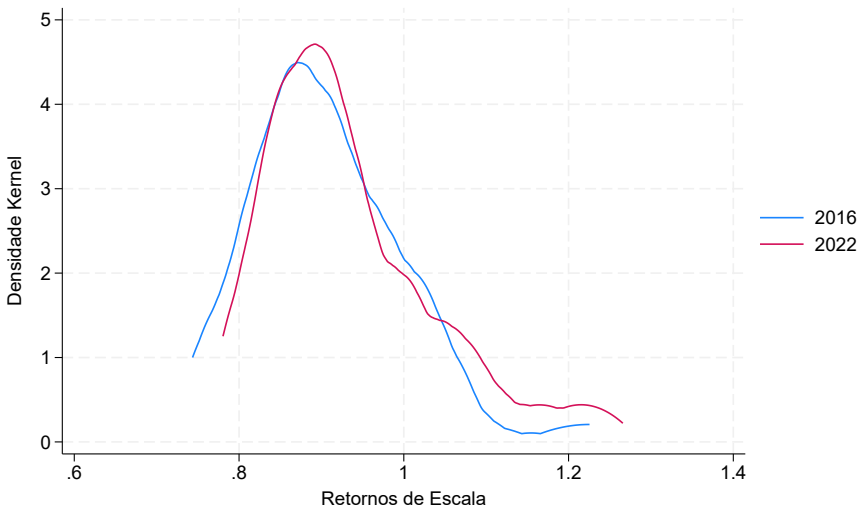
Como na média o retorno de escala tem aumentado é importante comparar com mais detalhe a distribuição dos retornos de escala. Isto é realizado por duas estimativas kernel da distribuição do retorno de escala para 2016 e 2022, estes resultados estão na Figura 2. Como descrito anteriormente existe o deslocamento da distribuição para a direita, significando aumento do retorno de escala, mas com a mediana menor do que um. Na cauda direta é possível observar que ocorre um deslocamento de densidade de valores abaixo de 1 para acima. Dito de outra forma, existe muito mais densidade de observações acima de 1, retorno constante de escala, em 2022 do que em 2016.

Os números encontrados para o setor de saneamento no Brasil estão em linha com as estimativas realizadas por DLGKP para firmas (médias e grandes) da manufatura indiana. A mediana das estimavas do retorno de escala para diversos setores estão entre 1.02 e 1.09.[28]

Kim e Clark (1988) não encontraram economia de escala significante na oferta de água. Existira economia de escala na oferta não-residencial e existira deseconomia de escala na

[28]Evidência de estimativa de retornos de escala para o setor no Brasil é dada por Seroa da Motta e Moreira (2006) que encontram números elevados, 3 por exemplo, usando a estimação de um modelo de fronteira estocástica. Embora não seja explicito no texto, o modelo empírico não possui estoque de capital, o que faz com que o retorno de escala seja muito elevado. Lucinda e Anuatti (2017) econtram economias de escala para a SABESP ofertando água e esgoto nos municípios de SP, utilizando uma abordagem usando uma função custo *translog*. Neste trabalho também se tem evidência de economia de escala para a SABESP.

Figura 2: Densidade da Economia de Escala: 2016 e 2022



oferta residencial. Portanto, serviços associados a oferta de água não se comportariam como monopólio natural. Lucinda e Anuatti utilizando estimativas da função custo translog para o estado de São Paulo comparam a oferta de água e esgoto para os diferentes municípios servidos pela SABESP. Eles estimam custo marginal por município. Mostram que existe evidência fraca de economia de escala por município coberto pela SABESP. Abbott e Cohen (2009) fazem uma resenha sobre a literatura e mostram que existe economias de escala na distribuição de água até certo ponto a partir do qual, o custo unitário começa a subir. Entretanto, a literatura não é consensual sobre isto, sendo dependente de como a variável produto é definida. Para oferta de coleta e tratamento de esgoto a literatura é consensual, com resultados mostrando a existência apenas de economias de escopo e sem economias de escala.

5.1 Economia de Escala e Monopólio Natural

A definição clássica de monopólio natural passa pela definição tecnológica da produção.[29] Uma firma produzindo um produto homogêneo é um monopólio natural se o custo de produção é menor quando apenas uma firma produz qualquer quantidade em comparação com duas ou mais firmas. Além disso, esta relação de dominância do custo deve ser verdade para vários mercados que demandam a referida produção.

Assuma uma função custo simples com dois componentes: custo fixo e custo variável. Por definição, apenas o custo variável muda com o volume produzido. Uma fórmula relevante para esta função é

$$C = F + c.q,$$

tal que F é o custo fixo, $c.q$ é o custo variável, q a quantidade produzida por uma firma e C o custo total. Cabe observar que q é medida física da produção e as demais variáveis são monetárias. Dividindo o custo total, C, pelo volume produzido se tem o custo médio de produção de uma firma qualquer que opera no mercado. A representação do custo

[29] A definição tecnológica segue a exposição de Paul L. Joskow (2007).

médio de uma firma j, CM_j, é

$$CM_j = \frac{F}{q_j} + c.$$

O custo médio de produção da firma representativa declina à medida que a produção aumenta. Isto é fácil de observar para este custo médio, uma vez que o custo fixo médio sempre cai quando a produção aumenta, para um dado custo variável médio constante. Uma função custo para uma firma com produto único caracterizada por custo médio declinante ao longo da produção relevante é subaditiva ao longo da capacidade produtiva. Neste contexto, a economia de escala sobre a produção relevante é uma condição suficiente para a definição tecnológica de monopólio natural. O retorno crescente de escala é definido pela operação empresarial com custo médio decrescente. Por outro lado, se o custo médio for crescente, então existe deseconomia de escala. Por exemplo, a função custo adotada nesta discussão teórica é caracterizada por custo médio decrescente, mas esta racionalidade é generalizada pelos economistas para outras funções.

A existência de economia de escala é uma condição necessária, mas não suficiente para a definição de monopólio natural. Mas isto implica que precisa existir economia de escala para existir monopólio natural, mas nem toda produção com retornos crescentes recebe esta classificação. Joskow (2007) mostra que pode existir uma função de custo com custo médio decrescente até certa quantidade produzida, passando a ser constante e em seguida o custo médio é crescente. Neste seguimento de custo médio constante pode existir um produtor sem a necessidade de existir custo médio estritamente decrescente. Isto poderia ocorrer porque o mercado não é grande o suficiente para suportar dois competidores. Este caso é chamado de monopólio natural de curto prazo, pois à medida que o mercado aumenta, também cresce a probabilidade de entrada de novos competidores para operar na região de custo médio constante.

Joskow (2007) diz que existem na história da análise econômica outras explicações para monopólio natural, mas todas devem se conectar com a tecnologia que é operada, i.e. com a economia de escala. É o que defende Kahn (1971, p. 123) em seu livro clássico: "o princípio geral é de que custos decrescentes de longo prazo são uma condição indispensável para o monopólio natural."[30]

Ao menos no Brasil a visão convencional é de que a tecnologia do setor de saneamento é dominada por economias de escala em virtude da existência de custos fixos (veja por exemplo, Galvão Jr. e Paganini, 2009). Como argumentado acima, a presença de custos fixos elevados, totalmente ou parcialmente afundado, não é suficiente para caracterizar monopólio natural. É preciso que exista economia de escala.

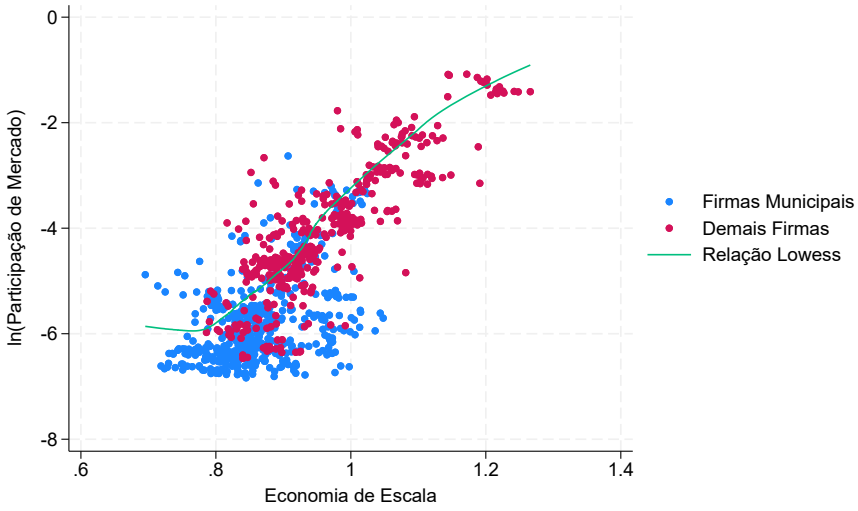
Seguindo o resultado relatado por Abbott e Cohen (2009), de que existe economia de escala até certo valor. Schmalensee (1978) também sugere que pode surgir economia de escala quando a demanda não é suficiente para sustentar determinada estrutura de custos. Para explorar um pouco esta tipo de relação, analisamos brevemente a economia de escala com uma medida de tamanho das firmas. Esta relação esta descrita na Figura 3, onde esta o gráfico do log natural da participação de mercado com a medida de economia de escala.[31] A linha continua é uma estimativa não-paramétrica logit por Lowess entre a participação de mercado a economia de escala. Esta estimativa mostra que existe relação positiva entre economia de escala e tamanho das empresas, a partir de 0,8 de economia

[30] Veja também Schmalensee (1978, p. 271) que argumenta que a economia de escala às vezes é importante na distribuição/entrega do serviço e não necessariamente na transformação produtiva.

[31] A participação de mercado é definida como a quantidade total produzida em relação a quantidade total produzida no mercado brasileiro por ano.

de escala, aproximadamente.

Figura 3: Economia de Escala vs Participação de Mercado



No gráfico estão separadas as empresas municipais, que somente podem operar em uma localidade, e as demais. Observe que como as entidades municipais não operam em mais de uma localidade, tipicamente elas possuem deseconomias de escala. Este tipo de explicação deveria ser aprofundado em estudos sobre o setor.

Economia de escala é crescente com o tamanho da empresa. Isto significa que a economia de escala não é apenas função da tecnologia, mas também do tamanho da demanda (receita). Empresas que podem acessar mais mercados locais podem usufruir de retornos de escala.

6 Markups

Como utilizamos a função de produção *translog,* as elasticidades e os retornos de escala variam por firma. O uso da função permite maior heterogeneidade no cálculo do markup.

A medida de markup captura a lucratividade operacional na margem quando as empresas ofertam produtos e serviços. Ela não é medida financeira e sim de margem operacional, mas ela é que sustenta a lucratividade das firmas. Decisões financeiras das empresas tratam de como a empresa sustenta o passivo e o circulante ao longo do tempo, sendo ferramentas de sustentação das operações e investimentos.

Na Tabela 3 são apresentadas estatísticas descritivas do markup para o setor em três anos selecionados. Observe que o markup médio era de 1,25 até 2020 e em 2022 se reduz para 1,21. As duas medidas de percentis também se reduzem e com a percentil 25 passando a ser menor do que um em 2022. Como a mediana é menor do que a média, a distribuição é caudal à direita, com empresas com markup elevado.[32] Algumas empresas possuem markup menor do que a unidade (provavelmente estas são empresas com pouca capacidade de investimento).

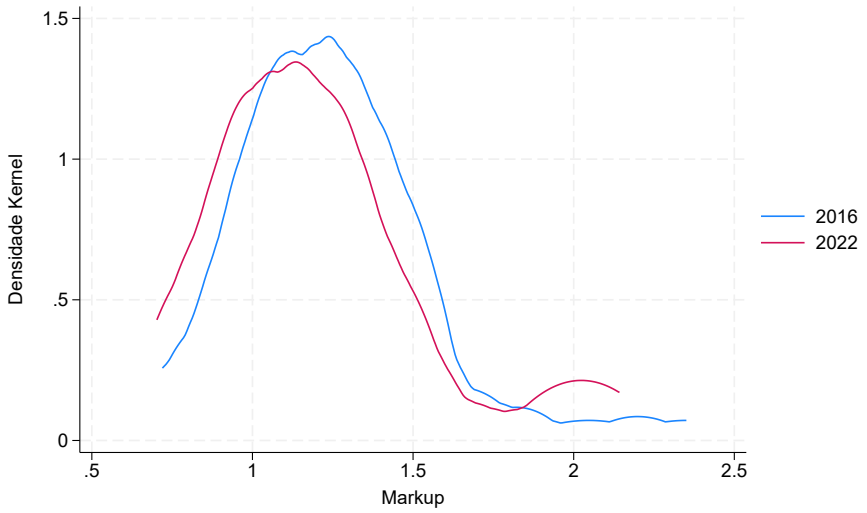
[32]DLGKP reportam markups de 2 para a maédia ponderada dos setores e especificamente para o setor químicos e maquinaria & equipamento para as empresas médias e grande da Índia. O setor com maior markup é o de máquinas elétricas e comunicações, com 5,66. Veja Tabela VI de DLGKP.

Tabela 3: Markups

Estatísticas	2016	2020	2022
Média	1.2551	1.2583	1.2129
SD	0.3081	0.3813	0.3390
p25	1.0582	1.0144	0.9597
Mediana	1.2156	1.1846	1.1325
p75	1.3969	1.4410	1.3694

Na Figura 4 é descrita a distribuição do markup a partir de estimativa kernel. Na distribuição do markup fica evidente mudança para a esquerda. O volume de markup em torno de 2 não é pequeno com as empresas acima do percentil 75 fazendo markup entre 1,4 e 2,3, aproximadamente.

Figura 4: Distribuição do Markup, 2016 e 2022



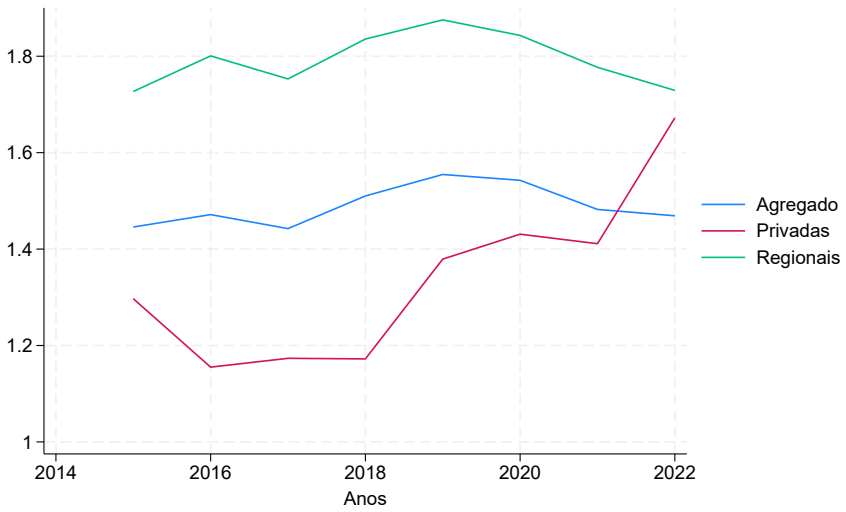
Na Figura 5 é apresentado o markup agregado do setor. A agregação é realizada utilizando a soma dos markups individuais ponderado pela respectiva participação na receita total. O markup total entre 2016 e 2022 varia entre 1.44 e 1.57, com moderado crescimento até 2020 e decréscima para 1.50 em 2022.

Na Tabela 4 são apresentados os investimentos como função do markup defasado em um período. O markup implica em .3 de aumento de investimentos um período à frente. Empresas privadas também possuem relação positiva entre investimento em markup um período defasado.

7 Conclusões

Este estudo analisou o poder de mercado das empresas do setor de água e saneamento no Brasil, empregando a medida de *markup* para mensurar a capacidade dessas firmas de

Figura 5: Markup Agregado



praticar preços superiores aos custos marginais. A investigação revelou que os *markups* elevados desempenham um papel central na dinâmica econômica do setor, refletindo tanto as barreiras estruturais quanto os incentivos à realização de investimentos em infraestrutura.

Os resultados apresentados destacam a dualidade dos *markups*. Por um lado, sua presença é necessária para viabilizar os altos custos fixos e afundados que caracterizam o setor, assegurando a manutenção e expansão da cobertura de serviços. Por outro lado, *markups* excessivos podem gerar ineficiências alocativas e limitar o acesso universal aos serviços, perpetuando disparidades regionais significativas, como evidenciado nas diferenças de cobertura entre as regiões Nordeste e Sudeste do Brasil.

Por fim, a análise reforça a importância de futuros estudos empíricos que investiguem os impactos de mudanças regulatórias e a evolução da estrutura de mercado no setor. A combinação de abordagens teóricas e dados empíricos pode oferecer subsídios fundamentais para a formulação de políticas públicas mais eficazes e para a promoção de um sistema de saneamento mais inclusivo e eficiente.

Referências

[1] Abbott, Malcolm e Bruce Cohen. "Productivity and Efficiency in the Water Industry." *Utilities Policy,* 17(3), 2009.

[2] Ackerberg, Daniel A., Kevin Caves, e Garth Frazer. "Identification Properties of Recent Production Function Estimators." *Econometrica,* 83(6), 2015.

[3] Ackerberg, Daniel A., e Jan De Loecker. "Production Function Identification Under Imperfect Competition." 2024.

[4] Andrews, Isiah, Matthew Gentzkow, e Jesse M. Shapiro, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics,* 132 (4), 2017.

Tabela 4: Investimento, Markup Defasado e Classe de Firmas

	(1)	(2)	(3)	(4)
Markup $(t-1)$	0.331***	0.359***	0.330***	0.370***
	(0.0302)	(0.0470)	(0.0331)	(0.0523)
Privadas	0.0662**	-0.0720	0.0652**	-0.0751
	(0.0262)	(0.0691)	(0.0279)	(0.0750)
Regionais	0.183***	0.199***	0.186***	0.206***
	(0.0167)	(0.0203)	(0.0180)	(0.0222)
Markup $(t-1)$ × Privadas		0.304*		0.297*
		(0.159)		(0.172)
Markup $(t-1)$ × Regionais		-0.0709		-0.0908
		(0.0620)		(0.0679)
Constante	2.267***	2.264***	2.269***	2.265***
	(0.00888)	(0.00954)	(0.00974)	(0.0106)
Observations	323	323	267	267

Erro-padrão entre parêntesis. Variável dependente é o ln do investimento. O investimento é previamente controlado por dummies de tempo e firma. Modelos (1) e (2) são estimativas por MQO enquanto modelos (3) e (4) são estimados por 2SLS. Instrumentos usados são markup defasados $(t-2)$ e impostos com uma defasagem. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[5] Berry, Steven T., Martin Gaynor, e Fiona Scott Morton. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives,* 33 (3), 2019.

[6] Boar, Corina e Virgiliu Midrigan. "Markups and Inequality." *Review of Economic Studies,* a ser publicado, 2024.

[7] Christensen, Laurits R., Dale W. Jorgenson, e Lawrence J. Lau. "Transcendental Logarithmic Production Frontiers." *Review of Economics and Statistics,* 55 (1), 1973.

[8] Collard-Wexler, Allan, e Jan De Loecker. "Productivity and Capital Measurement Error." Unpublished manuscript, 2021.

[9] Coury, Michael, Toru Kitagawa, Allison Shertzer, e Matthew A. Turner. "The Value of Piped Water and Sewers: Evidence from 19th Century Chicago." *The Review of Economics and Statistics,* 2024.

[10] De Loecker, Jan. "Detecting Learning by Exporting." *American Economic Journal: Microeconomics,* 5 (3), 2013.

[11] De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, e Nina Pavcnik. "Prices, Markups, and Trade Reform." *Econometrica,* 84, 2016.

[12] De Loecker, Jan e Chad Syverson, "An Industrial Organization Perspective on Productivity." *Handbook of Industrial Organization*, Volume 4. Kate Ho, Ali Hortaçsu, e Alessandro Lizzeri (eds.). Amsterdam, Elsevier, 2021.

[13] Dearing, Adam. "Estimating Structural Demand and Supply Models Using Tax Rates as Instruments." *Journal of Public Economics,* 205, 2022.

[14] Doraszelski, Uli e Jordi Jaumandreu, "Reexamining the De Loecker & Warzynski (2012) method for estimating markups." 2021.

[15] Evans, David S. e James J. Heckman. "Multiproduct Cost Function Estimates and Natural Monopoly Tests for the Bell System." In: David S. Evans (ed.). *Breaking Up Bell.* Elsevier, Amsterdam, 1983.

[16] Galvão Junior, Alceu de Castro e Wanderley da Silva Paganini. "Aspectos Conceituais da Regulação dos Serviços de Água e Esgoto no Brasil." *Engenharia Sanitária e Ambiental*, 14 (1), 2009.

[17] Joskow, Paul L. "Regulation of Natural Monopoly." In: A. Mitchell Polinsky e Steven Shavell (eds). *Handbook of Law and Economics,* Volume 2. Amsterdam, North Holland, 2007.

[18] Kahn, Alfred. *The Economics of Regulation.* Volume II. New York, John Wiley, 1971.

[19] Kim, H. Youn e Robert M. Clark. "Economies of scale and scope in water supply." *Regional Science and Urban Economics,* 18 (4), 1988.

[20] Levinsohn, James, e Amil Petrin. "Estimating Production Functions using Inputs to Control for Unobservables." *Review of Economic Studies,* 70(2) 2003.

[21] Lucinda, Claudio R. de e Francisco Anuatti. "Economies of Scale and Scope in the Sanitation Sector." *Brazilian Review of Econometrics,* 37 (2), 2017.

[22] Olley, G. Steven, e Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica,* 64(6), 1996.

[23] Raval, Devesh. "Testing the Production Approach to Markup Estimation." *Review of Economic Studies,* 90, 2023.

[24] Secretaria Nacional de Saneamento Ambiental (SNSA), Ministério das Cidades. *Diagnóstico Temático Serviços de Água e Esgoto.* Visão Geral. Ano de referência 2022. Brasília, Dezembro, 2023.

[25] Schmalensee, Richard. "A Note on Economies of Scale and Natural Monopoly in the Distribution of Public Utility Services." *The Bell Journal of Economics,* 9 (1), 1978.

[26] Seroa da Motta, R. Ajax Moreira. "Efficiency and Regulation in the Sanitation Sector in Brazil." *Utilities Policy,* 14, 2006.

[27] Sutton, John. *Sunk Costs and Market Structure:* Price Competition, Advertising, and the Evolution of Concentration. Cambridge, MIT Press, 1991.

[28] Syverson, Chad. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives,* 33 (3), 2019.

[29] Syverson, Chad. "Markups and Markdowns." Becker-Friedman Insitute Working Paper, University of Chicago, 2024.

[30] Timmins, Chris. "Measuring the Dynamic Efficiency Costs of Regulators' Preferences: Municipal Water Utilities in the Arid West." *Econometrica,* 70 (2), 2002.

Apêndice

A Dados

O SNIS esgoto reporta 3717 prestadores entre administração direta, autarquia, sociedade de economia mista, organização social, empresa pública e privada. Prestadores são diferentes de firmas, para o SNIS uma operada local é uma prestadora. Neste estudo, a firma se aproxima de um grupo econômico que pode operar controlando mais de uma prestadora.

De acordo com "Diagnóstico Temático Serviços de Água e Esgoto" (SNSA, 2023), 96,3% do total de investimentos são realizados pelo prestador de serviço. O restante é despesa enderaçada a estados e municípios. 63,3% são investimentos realizados com recursos prórpios, 31,7% utilizam algum mecanismo de financiamento e 5% são chamados de não-onerosos.

Medida de capital. O capital é mensurado seguindo o método do estoque perpétuo com depreciação de 10%. Para obter o estoque de capital foi acumulado o investimento deflacionado por IPCA a partir de 2003. Para o ponto inicial do estoque de capital, K_0, foi utilizada a média do investimento nos primeios seis anos. O total de despesa de investimento no setor contempla fortemente despesas de manutenção e reparação das redes, mas isto faz parte do estoque de capital produtivo. Especificamente, o investimento realizado pela CEDAE no município do Rio de Janeiro foi repartido entre as novas prestadoras no ano que estas assumiram o controle da concessão.

De acordo com a lei 2020, art. 4, buscar modicidade tarifária. Nos contratos de concessão a tarifa base é estabelecida pela tarifa de referência dos contratos de concessão. A revisão tarifária trata da recomposição inflacionária da tarifa definida nos contratos.[33] Revisão tarifária deve incluir repasse de ganhos de produtividade (inciso 8), mas a tarifação busca assegurar "o equilíbrio econômico-financeiro dos contratos." Neste sentido, a revisão tarifária visa permitir investimentos.

B Estimação

[33] A tarifa deverá ser reajustada anualmente de acordo com a metodologia de correção monetária prevista no contrato. Resolução ANA 183, 5 de fevereiro de 2024.

Tabela 5: Parâmetros da Função de Produção

Parâmetros	Cobb-Douglas	Translog
β_k	0,0644	0,8064
β_l	0,4065	0,7940
β_m	0,6054	0,2550
β_{k^2}		0,0241
β_{l^2}		-0,1070
β_{m^2}		0,0035
β_{kl}		-0,0787
β_{km}		-0,0956
β_{lm}		0,1115
β_{klm}		0,0048
N	782	782

Nota: N é o número de observações.