

Fatores da Produtividade das Firms no Brasil: Uma Análise Microeconômica com Dados da RAIS 2021

Lucas Scheremetta, Pedro Henrique
UFMG

Setembro 2025

Sumário

1	Introdução	3
2	Metodologia	3
2.1	Fonte de Dados	3
2.2	Variáveis	3
2.3	Estratégia Metodológica	5
2.3.1	Análise Descritiva	5
2.3.2	Modelos de Regressão	5
2.3.3	Decomposição de Blinder-Oaxaca	7
3	Análise Exploratória dos Dados	8
4	Análise dos Modelos Econométricos	11
4.1	Modelos de Mínimos Quadrados Ordinários (MQO)	11
4.2	Regressão Quantílica	13
4.3	Decomposição de Blinder-Oaxaca	14
5	Conclusão	16

1 Introdução

Como revisado intensamente na literatura, a produtividade é um dos principais motores do crescimento econômico e do desenvolvimento de uma nação. No entanto, a produtividade agregada é um reflexo de uma complexa e heterogênea distribuição de eficiências no nível microeconômico, ou seja, no nível das firmas. O Brasil, em particular, é caracterizado por uma persistente baixa produtividade agregada, um fenômeno que esconde uma vasta dispersão no desempenho entre as empresas.

Compreender quais fatores determinam essa heterogeneidade é fundamental para a formulação de políticas públicas eficazes. Este trabalho busca contribuir para essa discussão ao investigar os fatores da produtividade das firmas no setor privado brasileiro. Utilizando uma rica base de dados da Relação Anual de Informações Sociais (RAIS) para o ano de 2021, que combina informações das empresas com características agregadas de seus trabalhadores, buscou-se responder à seguinte pergunta: quais são as características das firmas e de sua força de trabalho que explicam as diferenças de produtividade observadas?

Para tanto, este estudo emprega uma estratégia econométrica em múltiplos estágios. Primeiramente, uma análise exploratória visualiza as distribuições de produtividade através de diferentes recortes setoriais e demográficos. Em seguida, houve a utilização de técnicas não-paramétricas. Posteriormente, foram utilizados modelos de regressão por Mínimos Quadrados Ordinários (MQO) e Regressão Quantílica que são estimados para quantificar o impacto médio e ao longo da distribuição dos fatores da produtividade. Finalmente, a técnica de decomposição de Oaxaca-Blinder é aplicada para investigar especificamente o hiato de produtividade entre firmas com maioria de funcionários homens e mulheres.

Este trabalho está estruturado da seguinte forma: a Seção 2 detalha a metodologia, descrevendo a fonte de dados, as variáveis e os modelos econométricos utilizados. A Seção 3 apresenta a análise exploratória dos dados. A Seção 4 discute os resultados dos modelos de regressão e da decomposição. Por fim, a Seção 5 sumariza as principais conclusões, discute as limitações do estudo e aponta caminhos para futuras pesquisas.

2 Metodologia

2.1 Fonte de Dados

Os dados utilizados neste estudo são provenientes da Relação Anual de Informações Sociais (RAIS) de 2021, disponibilizada pelo Ministério do Trabalho e Emprego do Brasil BRASIL (2021). A unidade de análise é a firma. A base de dados foi previamente tratada para agregar características dos trabalhadores (como escolaridade média, idade média, gênero e raça/cor) no nível da empresa, além de conter informações estruturais da própria firma, como idade, tamanho e setor de atividade (CNAE), que foi utilizado para derivar o nível tecnológico. A análise foca-se nas empresas do setor privado. Todos os dados foram tratados na linguagem R, por meio da IDE RStudio.

2.2 Variáveis

A variável dependente em todas as análises é a produtividade da firma. Portanto, a variável explicada e explicativas são, respectivamente:

Tabela 1: Variável dependente utilizada nos modelos econométricos

Variável	Descrição
<i>produtividade</i>	Massa salarial (soma das remunerações médias reais) dividida pela massa horária (soma das quantidades horárias)

Fonte: Elaboração própria com base nos dados da RAIS.

Tabela 2: Variáveis explicativas utilizadas nos modelos econométricos

Variável	Descrição
<i>qtd_vinc_atv</i>	Quantidade de vínculos ativos na empresa.
<i>tamanho</i>	Variável discreta entre 1 e 5 que demonstra o tamanho da empresa.
<i>nivel_tec1</i>	<i>Dummy</i> que assume valor igual a 1 em empresas que têm o nível de tecnologia 1 (em uma escala discreta de 0 a 5) e 0 caso contrário.
<i>nivel_tec2</i>	<i>Dummy</i> que assume valor igual a 1 em empresas que têm o nível de tecnologia 2 (em uma escala discreta de 0 a 5) e 0 caso contrário.
<i>nivel_tec3</i>	<i>Dummy</i> que assume valor igual a 1 em empresas que têm o nível de tecnologia 3 (em uma escala discreta de 0 a 5) e 0 caso contrário.
<i>nivel_tec4</i>	<i>Dummy</i> que assume valor igual a 1 em empresas que têm o nível de tecnologia 4 (em uma escala discreta de 0 a 5) e 0 caso contrário.
<i>nivel_tec5</i>	<i>Dummy</i> que assume valor igual a 1 em empresas que têm o nível de tecnologia 5 (em uma escala discreta de 0 a 5) e 0 caso contrário.
<i>idade</i>	Variável discreta que mostra a idade da firma em anos.
<i>sexo_med</i>	Variável que está entre 0 e 1. 0 significa que a firma é composta totalmente por pessoas do sexo masculino e 1 o inverso (a firma composta totalmente de pessoas do sexo feminino).
<i>raca_cor_med</i>	Variável que mostra a proporção de negros (pretos e pardos) e indígenas na firma. 0 significa que a firma é composta totalmente por pessoas sem alguma dessas características e 1 o inverso (a firma é composta totalmente de pessoas negras (pretas e pardas) ou indígenas).
<i>idade_med</i>	Variável que mostra a média de idade dos trabalhadores da firma.
<i>tempo_emprego_med</i>	Variável que mostra a média (em anos) de tempo empregado dos trabalhadores da firma.
<i>hb_med</i>	Variável que mostra a proporção de trabalhadores brancos na firma (entre 0 e 1).
<i>hn_med</i>	Variável que mostra a proporção de trabalhadores negros na firma (entre 0 e 1).
<i>ate_fundamental_med</i>	Variável que mostra a proporção de trabalhadores com graus de escolaridade mais baixos que "médio completo".
<i>superior_med</i>	Variável que mostra a proporção de trabalhadores com graus de escolaridade iguais ou superiores a "superior completo".

Fonte: Elaboração própria com base nos dados da RAIS.

2.3 Estratégia Metodológica

A análise do presente estudo foi conduzida em 4 etapas: (i) Análise descritiva; (ii) Modelos de regressão; (iii) Decomposição de Blinder-Oaxaca.

2.3.1 Análise Descritiva

Utilizou-se a técnica de Densidade de Kernel para visualizar e comparar as distribuições de produtividade entre diferentes grupos de firmas (público vs. privado, por tamanho, MEI vs. outros, nível tecnológico e sexo majoritário).

2.3.2 Modelos de Regressão

Em um passo posterior, utilizou-se o ferramental econométrico (paramétrico e não-paramétrico) para quantificar os fatores da produtividade. Dentre os modelos paramétricos utilizados, têm-se os modelos de Mínimos Quadrados Ordinários (MQO) e Regressão Quantílica. Enquanto o MQO estima o efeito médio das variáveis explicativas sobre a produtividade, a Regressão Quantílica permite analisar se esses efeitos variam para firmas em diferentes pontos da distribuição de produtividade (i.e., para as menos e mais produtivas). Em uma análise não-paramétrica, foi utilizada a Regressão Kernel, a qual foi usada para explorar a relação não-paramétrica entre produtividade e o número de vínculos empregatícios.

Especificamente, o método de Mínimos Quadrados Ordinários (MQO) é a técnica econométrica mais amplamente utilizada para estimação de parâmetros em modelos de regressão linear. Desenvolvido inicialmente por Legendre (1805) e Gauss (1809), o MQO busca encontrar os valores dos parâmetros que minimizam a soma dos quadrados dos resíduos WOOLDRIDGE (2019).

A forma geral de um modelo de regressão linear múltipla é dada por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

onde:

- Y_i é a variável dependente para a observação i
- X_{ji} são as variáveis explicativas
- β_j são os parâmetros a serem estimados
- ε_i é o termo de erro aleatório

O estimador de MQO é obtido minimizando a função:

$$\min_{\beta} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \min_{\beta} \sum_{i=1}^n \varepsilon_i^2 \quad (2)$$

A solução analítica para este problema de otimização é dada por:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (3)$$

sob as seguintes premissas clássicas GUJARATI; PORTER (2011):

1. Linearidade nos parâmetros

2. Amostragem aleatória
3. Média condicional zero: $E(\varepsilon_i|X) = 0$
4. Não existência de multicolinearidade perfeita
5. Homocedasticidade: $Var(\varepsilon_i|X) = \sigma^2$
6. Não autocorrelação dos erros: $Cov(\varepsilon_i, \varepsilon_j|X) = 0$ para $i \neq j$

Sob essas condições, o teorema de Gauss-Markov estabelece que os estimadores de MQO são os Melhores Estimadores Lineares Não Tendenciosos (BLUE) GREENE (2018). A inferência estatística pode ser realizada através de testes t, F e estatísticas de significância, assumindo normalidade dos resíduos para amostras pequenas.

Apesar de sua ampla utilização, o MQO apresenta limitações quando as premissas do modelo são violadas, particularmente na presença de valores atípicos (outliers), heterocedasticidade ou quando o interesse centra-se em partes específicas da distribuição da variável dependente KOENKER (2005).

Em um espectro mais específico na distribuição, a regressão quantílica, introduzida por KOENKER; BASSETT (1978), representa uma extensão robusta dos modelos de regressão tradicional que permite estimar efeitos em diferentes pontos da distribuição condicional da variável dependente. Enquanto o MQO foca na média condicional, a regressão quantílica modela quantis condicionais específicos (τ), como a mediana ($\tau = 0,5$) ou outros percentis.

O modelo de regressão quantílica para o quantil τ pode ser expresso como:

$$Q_{Y_i}(\tau|X_i) = X_i'\beta(\tau) \quad (4)$$

onde $Q_{Y_i}(\tau|X_i)$ representa o τ -ésimo quantil condicional de Y_i dado X_i .

Os coeficientes $\beta(\tau)$ são estimados resolvendo o seguinte problema de minimização:

$$\min_{\beta} \sum_{i=1}^n \rho_{\tau}(Y_i - X_i'\beta(\tau)) \quad (5)$$

onde $\rho_{\tau}(u)$ é a função de perda assimétrica definida como:

$$\rho_{\tau}(u) = \begin{cases} \tau u & \text{se } u \geq 0 \\ (\tau - 1)u & \text{se } u < 0 \end{cases} = u(\tau - I(u < 0)) \quad (6)$$

Esta função de perda atribui pesos diferentes aos resíduos positivos e negativos, permitindo que diferentes quantis sejam estimados KOENKER; MACHADO (1999).

As vantagens da regressão quantílica incluem:

1. **Robustez:** Menos sensível a valores extremos comparado ao MQO
2. **Flexibilidade:** Permite analisar diferentes pontos da distribuição
3. **Eficiência:** Em algumas distribuições, estimadores de regressão quantílica podem ser mais eficientes que MQO
4. **Invariância a transformações monotônicas:** $Q_{h(Y)}(\tau|X) = h(Q_Y(\tau|X))$ para funções monotônicas h

A inferência em regressão quantílica pode ser realizada usando métodos de bootstrap, uma vez que a distribuição assintótica dos estimadores depende da função densidade desconhecida dos erros BUCHINSKY (1998).

A regressão quantílica tem aplicações particularmente úteis em casos onde a relação entre variáveis difere ao longo da distribuição, como em estudos sobre desigualdade salarial, onde o retorno da educação pode ser diferente para indivíduos nos quantis superiores comparados aos inferiores da distribuição de renda.

Portanto, enquanto o MQO fornece uma estimativa da mudança na média condicional da variável dependente dada uma mudança unitária na variável explicativa, a regressão quantílica oferece um panorama mais completo ao mostrar como essas relações variam em diferentes pontos da distribuição. Esta característica é particularmente valiosa quando há heterogeneidade nos efeitos, permitindo identificar diferenças que seriam mascaradas em uma análise apenas da média ANGRIST; PISCHKE (2009).

Ambas as técnicas serão implementadas utilizando o software estatístico R, com o pacote `quantreg` para as estimativas de regressão quantílica e procedimentos de bootstrap para inferência KOENKER (2022).

2.3.3 Decomposição de Blinder-Oaxaca

Para investigar o hiato de produtividade entre firmas de maioria feminina e masculina, aplicou-se a decomposição de Blinder-Oaxaca OAXACA (1973) e ?. Em suma, esta técnica divide a diferença média na produtividade em duas partes: uma componente *explicada*, atribuível a diferenças nas características médias observáveis (dotações) entre os grupos, e uma componente *não explicada*, atribuível a diferenças nos retornos (coeficientes) a essas características.

A decomposição de Oaxaca-Blinder, desenvolvida independentemente por OAXACA (1973) e ?, é uma técnica econométrica amplamente utilizada para decompor diferenças médias entre grupos em dois componentes: um explicado por características observáveis e outro não explicado, frequentemente atribuído a discriminação ou fatores não observados.

Este método é particularmente útil em estudos de desigualdade salarial, discriminação no mercado de trabalho e diferenças educacionais, permitindo quantificar a parcela das disparidades que pode ser atribuída a diferenças nas características produtivas versus diferenças nos retornos a essas características JANN (2008).

Considere dois grupos, A e B (por exemplo, homens e mulheres), com modelos de regressão linear separados:

$$Y_g = X_g\beta_g + \varepsilon_g, \quad g = \{A, B\} \quad (7)$$

onde:

- Y_g é o vetor de outcomes para o grupo g
- X_g é a matriz de características observáveis
- β_g é o vetor de coeficientes
- ε_g é o termo de erro com média zero

A diferença média entre os grupos pode ser expressa como:

$$\Delta = \bar{Y}_A - \bar{Y}_B = \bar{X}_A\hat{\beta}_A - \bar{X}_B\hat{\beta}_B \quad (8)$$

A decomposição de Oaxaca-Blinder reorganiza esta diferença adicionando e subtraindo o termo $\bar{X}_B\hat{\beta}_A$:

$$\Delta = (\bar{X}_A - \bar{X}_B)\hat{\beta}_A + \bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) \quad (9)$$

Alternativamente, pode-se usar $\bar{X}_A\hat{\beta}_B$ como referência:

$$\Delta = (\bar{X}_A - \bar{X}_B)\hat{\beta}_B + \bar{X}_A(\hat{\beta}_A - \hat{\beta}_B) \quad (10)$$

A equação (9) separa a diferença total em dois componentes:

1. **Componente Explicado** (efeito das características):

$$E = (\bar{X}_A - \bar{X}_B)\hat{\beta}_A \quad (11)$$

Representa a parte da diferença atribuível às disparidades nas características observáveis entre os grupos.

2. **Componente Não Explicado** (efeito dos coeficientes):

$$U = \bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) \quad (12)$$

Representa a parte da diferença atribuível a disparidades nos retornos às características, frequentemente interpretado como discriminação ou efeitos de variáveis não observadas.

float

3 Análise Exploratória dos Dados

A análise visual revela padrões marcantes. A Figura 1 mostra que o setor privado concentra a vasta maioria das firmas em um baixo nível de produtividade, mas possui uma cauda longa de firmas altamente produtivas. As Figuras 2 e 3 detalham a forte relação entre porte e produtividade, com os MEIs formando um grupo distinto de baixa eficiência, enquanto empresas maiores deslocam a distribuição para a direita.

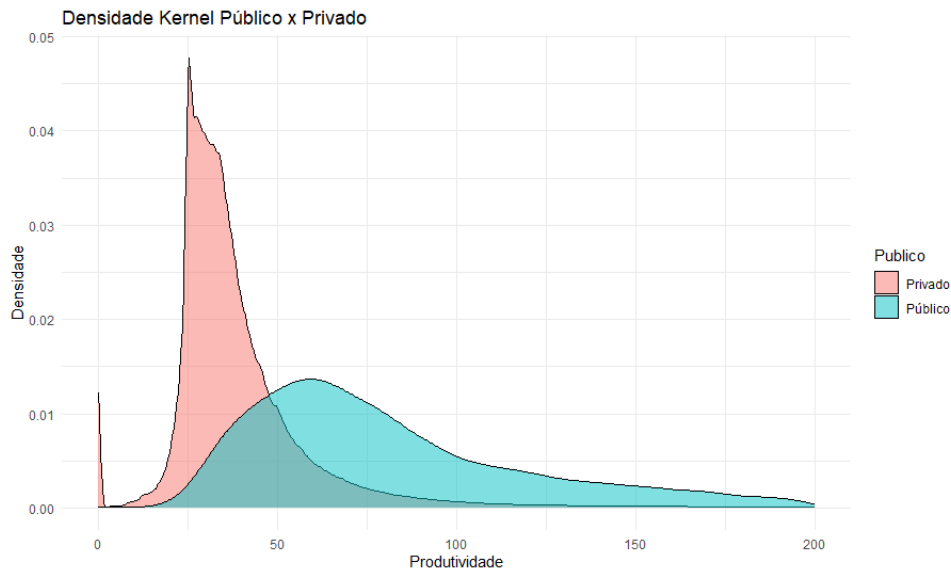


Figura 1: Densidade Kernel da Produtividade: Público vs. Privado

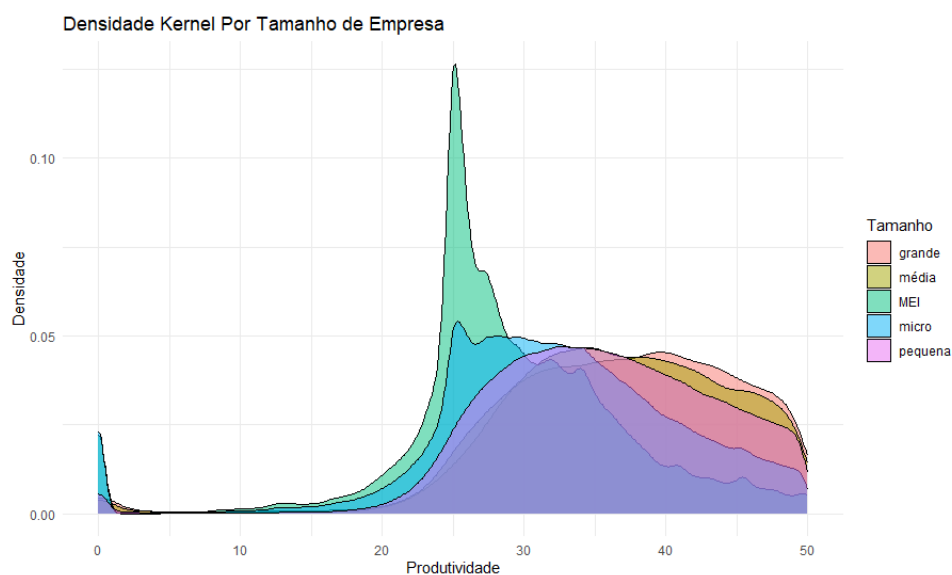


Figura 2: Densidade Kernel da Produtividade por Tamanho de Empresa

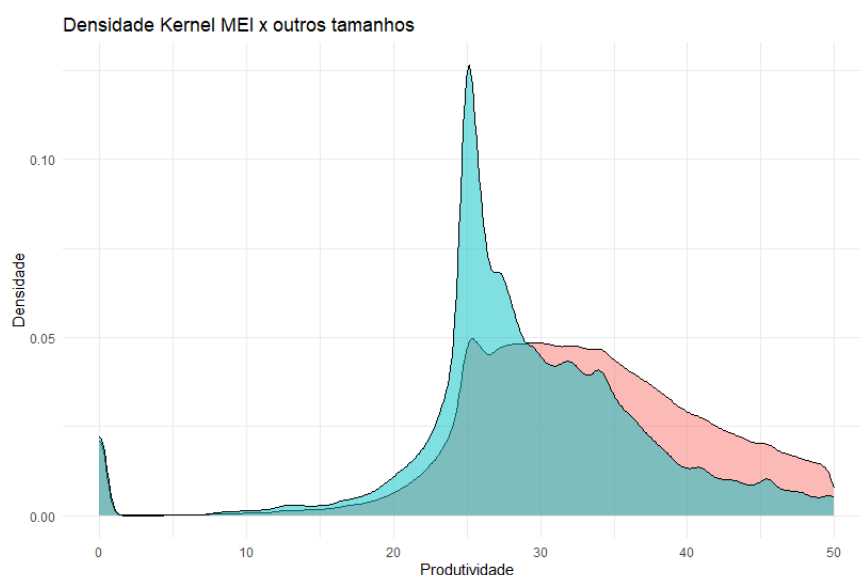


Figura 3: Densidade Kernel da Produtividade: MEI vs. Outros Portes

Além disso, as Figuras 4 e 5 indicam que firmas de maior nível tecnológico e aquelas com maioria masculina tendem a ter produtividade média superior. Por fim, a Figura 6 ilustra a extrema concentração de firmas com poucos funcionários e a presença de outliers, enquanto a Figura 7 sugere uma relação positiva, porém ruidosa, entre o número de vínculos e a produtividade após a remoção dos valores extremos.

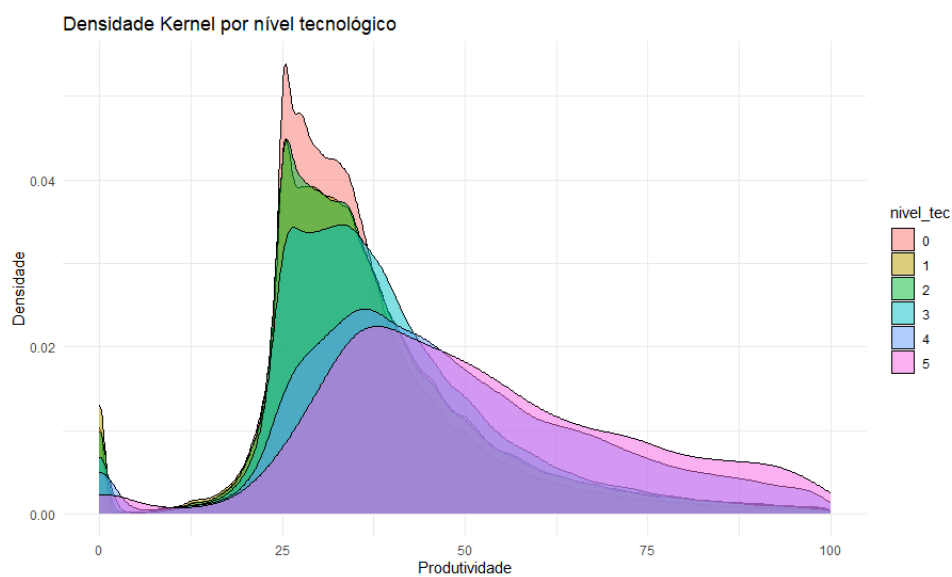


Figura 4: Densidade Kernel da Produtividade por Nível Tecnológico

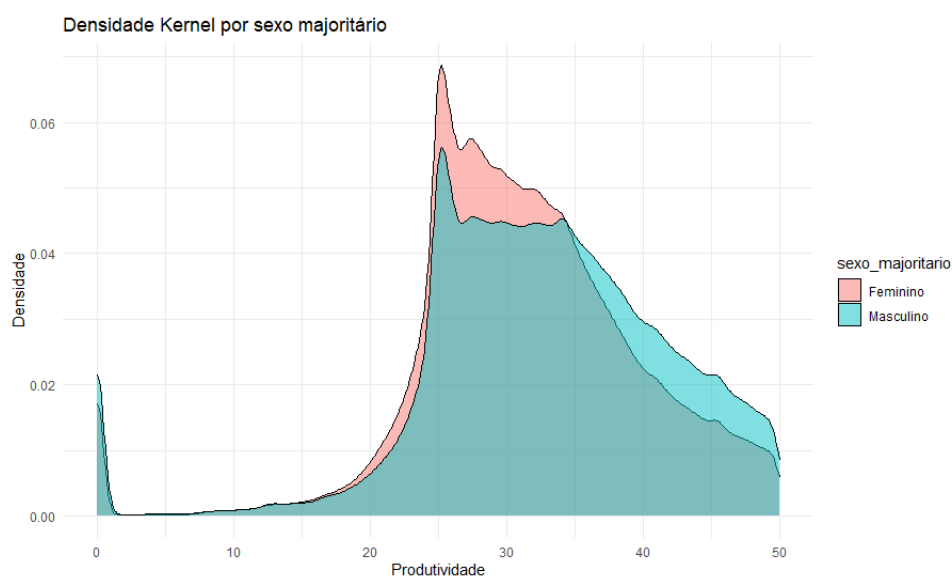


Figura 5: Densidade Kernel da Produtividade por Sexo Majoritário na Firma

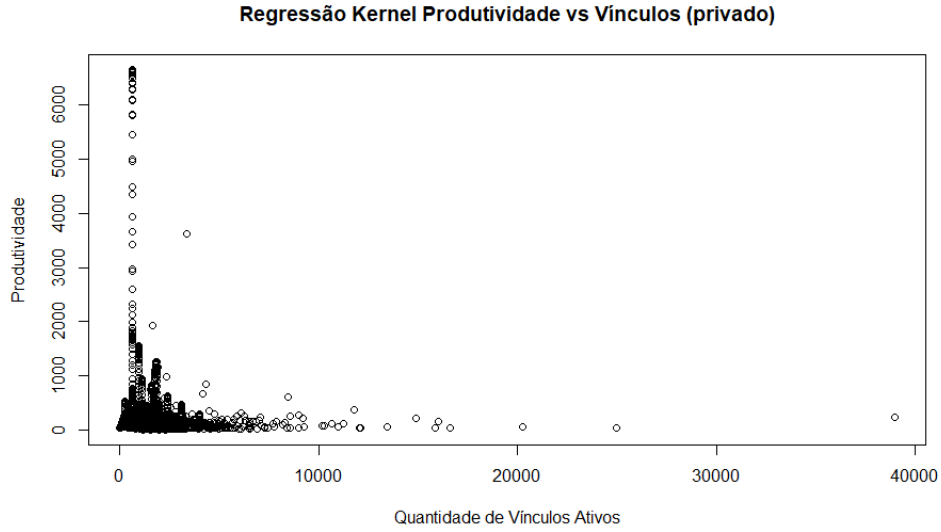


Figura 6: Relação entre Produtividade e Vínculos (com outliers)

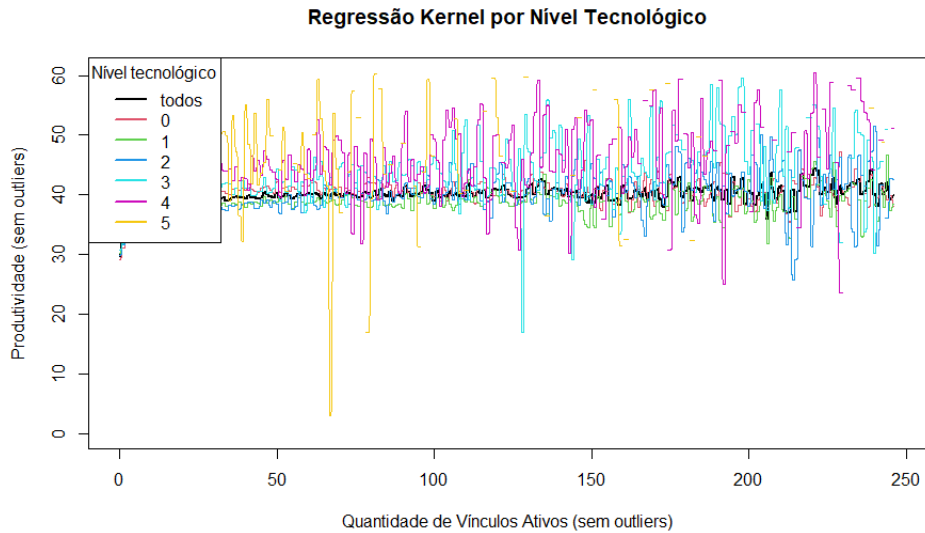


Figura 7: Regressão Kernel entre Produtividade e Vínculos (sem outliers)

4 Análise dos Modelos Econométricos

Nesta seção, interpretamos os resultados das tabelas de regressão e da decomposição.

4.1 Modelos de Mínimos Quadrados Ordinários (MQO)

A Tabela 3 apresenta os resultados de quatro especificações do modelo MQO geral. De forma consistente, a variável com maior impacto positivo é a proporção de trabalhadores com ensino superior (**superior_med**), destacando o papel central do capital humano. Tamanho e nível tecnológico também se mostram positivamente correlacionados com a produtividade. Por outro lado, a maior proporção de mulheres (**sexo_med**) e de não-brancos (**raca_cor_med**) está associada a uma menor produtividade. O R^2 baixo (em

torno de 4-5%) é esperado para dados em corte transversal dessa natureza e indica que uma grande parte da variação da produtividade é devida a fatores não observados.

Tabela 3: Resultados dos Modelos de Regressão Gerais para a Produtividade

Variável	(1) mqo_geral1	(2) mqo_geral2	(3) mqo_geral_dummies	(4) mqo_geral3
qtd_vinc_atv	0.033*** (0.001)	0.033*** (0.001)		0.014*** (0.001)
tamanho			9.086*** (0.085)	8.549*** (0.088)
nivel_tec1	4.223*** (0.128)	4.227*** (0.128)	4.085*** (0.127)	4.077*** (0.127)
nivel_tec2	0.859*** (0.189)	0.865*** (0.189)	0.226 (0.189)	0.194 (0.189)
nivel_tec3	6.020*** (0.370)	6.019*** (0.370)	4.687*** (0.370)	4.671*** (0.370)
nivel_tec4	18.330*** (0.437)	18.276*** (0.437)	15.890*** (0.437)	15.780*** (0.437)
nivel_tec5	23.787*** (1.801)	23.725*** (1.801)	19.959*** (1.798)	19.490*** (1.798)
idade	0.127*** (0.006)	0.128*** (0.006)	0.024*** (0.006)	0.022*** (0.006)
sexo_med	-11.291*** (0.152)	-8.725*** (0.613)	-10.333*** (0.152)	-10.350*** (0.152)
raca_cor_med	-6.122*** (0.138)	-4.737*** (0.213)	-6.863*** (0.137)	-6.878*** (0.137)
idade_med	0.218*** (0.007)	0.218*** (0.007)	0.233*** (0.007)	0.233*** (0.007)
tempo_emprego_med	0.070*** (0.002)	0.070*** (0.002)	0.080*** (0.002)	0.080*** (0.002)
hb_med		3.644*** (0.610)		
hn_med		0.996 (0.623)		
ate_fundamental_med	-7.531*** (0.196)	-7.517*** (0.196)	-7.225*** (0.196)	-7.245*** (0.196)
superior_med	60.195*** (0.250)	60.211*** (0.250)	58.042*** (0.251)	58.040*** (0.251)
(Intercept)	31.935*** (0.277)	28.839*** (0.655)	13.946*** (0.324)	14.930*** (0.328)
Observações	2,096,380	2,096,378	2,096,380	2,096,379
R^2	0.043	0.043	0.047	0.047
R^2 Ajustado	0.043	0.043	0.047	0.047

Nota: Erros-padrão entre parênteses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

No Modelo (2), duas variáveis foram omitidas devido a singularidades (multicolinearidade).

A Tabela 4 detalha os resultados por porte da empresa, revelando que os retornos ao capital humano (*superior_med*) e as penalidades associadas a gênero e raça se intensificam com o aumento do tamanho da firma.

Tabela 4: Resultados dos Modelos de Regressão por Tamanho da Empresa

Variável	(1) MEI	(2) Micro	(3) Pequena	(4) Média	(5) Grande
qtd_vinc_atv	2.029*** (0.166)	1.038*** (0.021)	0.186*** (0.022)	0.170* (0.083)	0.007** (0.002)
nivel_tec1	0.862*** (0.159)	3.879*** (0.114)	5.420*** (0.498)	11.830*** (2.925)	7.887 (4.080)
nivel_tec2	0.090 (0.250)	0.500** (0.171)	1.093 (0.679)	4.455 (3.747)	12.860** (4.826)
nivel_tec3	2.416*** (0.510)	4.869*** (0.351)	4.302*** (1.159)	5.931 (5.994)	9.779 (7.804)
nivel_tec4	5.306*** (1.009)	11.636*** (0.412)	10.196*** (1.310)	12.021* (5.803)	-4.978 (6.576)
nivel_tec5	3.839 (8.092)	14.058*** (1.884)	7.125 (4.796)	1.181 (18.750)	23.736 (15.878)
idade	-0.032** (0.011)	-0.019*** (0.005)	-0.098*** (0.019)	-0.070 (0.092)	-0.389*** (0.102)
sexo_med	-3.401*** (0.154)	-9.428*** (0.135)	-35.762*** (0.827)	-72.333*** (5.278)	-76.829*** (7.236)
raca_cor_med	-2.882*** (0.151)	-6.037*** (0.122)	-13.313*** (0.622)	-26.590*** (3.864)	-32.481*** (5.467)
idade_med	0.130*** (0.007)	0.287*** (0.007)	0.371*** (0.048)	-0.049 (0.295)	0.954* (0.419)
tempo_emprego_med	0.001 (0.002)	0.071*** (0.001)	0.135*** (0.008)	0.071 (0.044)	0.104 (0.054)
ate_fundamental_med	-1.625*** (0.188)	-6.656*** (0.177)	-16.763*** (1.053)	-27.424*** (6.223)	-14.628 (8.537)
superior_med	14.374*** (0.305)	48.259*** (0.220)	131.443*** (1.146)	201.667*** (6.120)	278.640*** (8.093)
(Intercept)	28.927*** (0.302)	27.241*** (0.258)	37.654*** (1.719)	63.290*** (12.034)	40.903** (14.720)
Observações	291,502	1,418,761	322,163	35,481	28,417
R^2	0.013	0.053	0.061	0.050	0.072
R^2 Ajustado	0.013	0.053	0.061	0.050	0.071

Nota: Erros-padrão entre parênteses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, · $p < 0.1$.

4.2 Regressão Quantílica

A Tabela 5 aprofunda a análise, mostrando como os efeitos variam ao longo da distribuição de produtividade. O resultado mais notável é o coeficiente da variável `superior_med`, que salta de 4.6 no quantil 0.10 para 119.8 no quantil 0.90. Isso indica que o capital humano qualificado tem um impacto muito maior em firmas que já são altamente produtivas, atuando como um fator de diferenciação e potencializando a desigualdade de desempenho entre as empresas.

Tabela 5: Resultados das Regressões Quantílicas para a Produtividade

Variável	(1) Quantil 0.10	(2) Quantil 0.50	(3) Quantil 0.90
qtd_vinc_atv	0.0003 (0.0003)	0.0126*** (0.0007)	0.0466*** (0.0026)
tamanho	2.7601*** (0.0163)	4.0860*** (0.0187)	4.3942*** (0.0646)
nivel_tec1	-0.3577*** (0.0204)	0.6985*** (0.0182)	3.8323*** (0.0546)
nivel_tec2	-0.7046*** (0.0315)	-0.7020*** (0.0244)	-0.0912 (0.0838)
nivel_tec3	0.0747 (0.0499)	1.1878*** (0.0608)	3.5045*** (0.1715)
nivel_tec4	2.2218*** (0.1070)	8.0061*** (0.1356)	16.7227*** (0.3062)
nivel_tec5	5.0361*** (0.7150)	11.4301*** (0.6024)	20.0675*** (2.7509)
idade	0.0635*** (0.0008)	0.0439*** (0.0009)	-0.0180*** (0.0024)
sexo_med	-1.5177*** (0.0242)	-4.1961*** (0.0201)	-8.6138*** (0.0605)
raca_cor_med	-2.2256*** (0.0213)	-4.4542*** (0.0175)	-6.0386*** (0.0541)
idade_med	0.0135*** (0.0013)	0.1217*** (0.0010)	0.3200*** (0.0032)
tempo_emprego_med	-0.0024*** (0.0005)	0.0577*** (0.0004)	0.1806*** (0.0011)
ate_fundamental_med	-1.5011*** (0.0380)	-2.0548*** (0.0220)	-5.0631*** (0.0665)
superior_med	4.6386*** (0.0538)	26.3886*** (0.1122)	119.8460*** (0.5298)
(Intercept)	19.9821*** (0.0527)	22.6994*** (0.0482)	29.0164*** (0.1560)

Nota: Erros-padrão entre parênteses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.3 Decomposição de Blinder-Oaxaca

A Tabela 6 apresenta o resultado agregado da decomposição da diferença de produtividade entre firmas de maioria feminina e masculina. A diferença total é de 5.52 em favor das firmas de maioria masculina. Surpreendentemente, a componente explicada (dotações) é negativa (-2.027), indicando que as firmas de maioria feminina possuem, em média, características observáveis mais favoráveis. A totalidade do hiato de produtividade é, portanto, atribuída à componente não explicada (8.936), que reflete retornos inferiores a essas características.

Tabela 6: Decomposição de Blinder-Oaxaca da Diferença de Produtividade

Componente	Coeficiente	Erro Padrão	% do Total
Média do Grupo A (Feminino)	46.182		
Média do Grupo B (Masculino)	40.661		
Diferença Total	5.521		100.0%
Componentes da Diferença:			
Parte Explicada (Dotações)	-2.027	(0.051)	-36.7%
Parte Não Explicada (Coeficientes)	8.936	(0.170)	161.8%
Interação	-1.388	(0.108)	-25.1%

A Tabela 7 mostra que as variáveis `idade_med` e `superior_med` são as que mais contribuem para essa componente não explicada, sugerindo que a experiência e a qualificação dos trabalhadores são menos "recompensadas" em termos de produtividade em firmas de maioria feminina.

Tabela 7: Decomposição Detalhada por Variáveis

Variável	Explicado (Dotações)		Não Explicado (Coef.)		Interação	
	Coef.	E.P.	Coef.	E.P.	Coef.	E.P.
(Intercept)	0.000	(0.000)	-6.975	(0.434)	0.000	(0.000)
qtd_vinc_atv	0.052	(0.012)	0.190	(0.064)	0.043	(0.013)
nivel_tec1	0.520	(0.018)	0.832	(0.072)	0.388	(0.034)
nivel_tec2	-0.001	(0.001)	0.502	(0.049)	0.003	(0.002)
nivel_tec3	0.070	(0.020)	0.049	(0.006)	0.202	(0.023)
nivel_tec4	0.183	(0.042)	0.063	(0.019)	0.144	(0.043)
nivel_tec5	0.016	(0.003)	-0.005	(0.004)	-0.004	(0.003)
idade	0.251	(0.013)	-0.667	(0.148)	-0.086	(0.019)
raca_cor_med	-0.226	(0.005)	-0.541	(0.075)	-0.062	(0.009)
idade_med	0.129	(0.012)	9.764	(0.464)	0.554	(0.026)
tempo_emprego_med	0.478	(0.014)	-0.204	(0.090)	-0.041	(0.018)
ate_fundamental_med	-0.732	(0.016)	0.053	(0.029)	0.044	(0.024)
superior_med	-2.766	(0.034)	5.874	(0.195)	-2.572	(0.088)

A Tabela 8 confirma as melhores dotações médias do grupo feminino em diversas variáveis, como tempo de emprego e idade dos funcionários.

Tabela 8: Estatísticas Descritivas das Covariáveis por Grupo

Variável	Grupo A (Feminino)	Grupo B (Masculino)
qtd_vinc_atv	12.0525	9.8295
nivel_tec1	0.4759	0.3246
nivel_tec2	0.1134	0.1128
nivel_tec3	0.0384	0.0076
nivel_tec4	0.0252	0.0077
nivel_tec5	0.0012	0.0007
idade	13.0491	11.5587
raca_cor_med	0.4344	0.3896
idade_med	36.5328	34.5702
tempo_emprego_med	41.3804	34.5277
ate_fundamental_med	0.2581	0.1422
superior_med	0.0815	0.1450
Observações (N)	1,185,692	910,702

5 Conclusão

Este estudo analisou os fatores da produtividade das firmas no Brasil, revelando uma forte heterogeneidade e o papel central de fatores como tamanho, tecnologia e, sobretudo, capital humano.

As principais conclusões são:

1. **A produtividade é um fenômeno de cauda longa:** Uma vasta base de empresas de baixa produtividade coexiste com um número menor de firmas altamente eficientes.
2. **O capital humano é o grande diferenciador:** A proporção de trabalhadores com ensino superior é o preditor mais forte de alta produtividade.
3. **Os retornos são crescentes:** Os benefícios do capital humano e de outras características favoráveis são significativamente maiores para as firmas que já estão no topo da distribuição de produtividade, exacerbando a desigualdade de desempenho.
4. **Existem desvantagens estruturais de gênero:** Firms de maioria feminina apresentam produtividade inferior, não por terem piores características, mas por obterem menores retornos a suas dotações, que são, em média, superiores. Isso aponta para a existência de barreiras não observadas ou vieses sistêmicos.

As limitações deste estudo incluem o baixo poder explicativo dos modelos, o que é comum em análises de microdados, e a natureza correlacional dos resultados, que não permite inferências causais diretas. A análise também se baseia em um único corte transversal, não capturando a dinâmica da produtividade ao longo do tempo.

Pesquisas futuras poderiam se beneficiar do uso de dados em painel para controlar por efeitos fixos das firmas e explorar a causalidade com maior robustez. Além disso, a inclusão de variáveis sobre práticas de gestão, inovação e acesso a crédito poderia ajudar a explicar a grande parcela da variação da produtividade que permanece não observada.

Referências

- ANGRIST, J. D.; PISCHKE, J. S. **Mostly Harmless Econometrics: An Empiricist's Companion**. Princeton: Princeton University Press, 2009.
- BLINDER, A. S. Wage Discrimination: Reduced Form and Structural Estimates. **The Journal of Human Resources**, v. 8, n. 4, p. 436–455, 1973.
- BRASIL. Ministério do Trabalho e Emprego. **Relação Anual de Informações Sociais (RAIS)**. 2021. Disponível em: <<http://rais.gov.br>>. Acesso em: 10 abr. 2024.
- BUCHINSKY, M. Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. **Journal of Human Resources**, v. 33, n. 1, p. 88–126, 1998.
- GREENE, W. H. **Econometric Analysis**. 8^a ed. Nova York: Pearson, 2018.
- GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5^a ed. Porto Alegre: AMGH Editora, 2011.
- HLAVAC, M. *oaxaca: Blinder-Oaxaca Decomposition in R*. Versão 0.1.5. 2022. Disponível em: <<https://CRAN.R-project.org/package=oaxaca>>. Acesso em: 07 set. 2025.
- JANN, B. The Blinder–Oaxaca Decomposition for Linear Regression Models. **The Stata Journal**, v. 8, n. 4, p. 453–479, 2008.
- KOENKER, R. **Quantile Regression**. Cambridge: Cambridge University Press, 2005.
- KOENKER, R. Ranking and Selection from Pairwise Comparisons: Empirical Bayes Methods for Citation Analysis. **Journal of the American Statistical Association**, v. 117, n. 538, p. 833–845, 2022.
- KOENKER, R.; BASSETT, G. Regression Quantiles. **Econometrica**, v. 46, n. 1, p. 33–50, 1978.
- KOENKER, R.; MACHADO, J. A. F. Goodness of Fit and Related Inference Processes for Quantile Regression. **Journal of the American Statistical Association**, v. 94, n. 448, p. 1296–1310, 1999.
- OAXACA, R. Male-Female Wage Differentials in Urban Labor Markets. **International Economic Review**, v. 14, n. 3, p. 693–709, 1973.
- WOOLDRIDGE, J. M. **Introdução à Econometria: Uma Abordagem Moderna**. 6^a ed. São Paulo: Cengage Learning, 2019.