

Relatório Provisório da Pesquisa: Impactos da Pandemia na Ocupação dos Jovens no Mercado de Trabalho brasileiro

1. Introdução

Este relatório documenta, passo a passo, o desenvolvimento da presente pesquisa, que objetiva, em linhas gerais, mensurar o impacto do choque econômico no contexto da pandemia de COVID-19 sobre a ocupação juvenil no mercado de trabalho. Mais propriamente, dentro do referencial teórico do Modelo de Search & Matching de Diamond-Mortensen-Pissarides (DMP), estimar as probabilidades condicionadas (Cadeias de Markov) de transição entre emprego e desemprego, com foco na integração dos jovens no setor laboral, e complementação com a mensuração de heterogeneidades no impacto dado diferentes características individuais (como sexo, raça, escolaridade, localidade, experiência), mas sobretudo, aplicação do referido modelo para regressão da eficiência de matching para o público-alvo.

Para este fim, utilizou-se microdados da PNAD Contínua Trimestral (2019-2022) e dados de admissões formais disponibilizados pelo CAGED como proxy para o estoque de vagas.

2. Construção do Painel Longitudinal da PNADC

2.1. Obtenção dos Dados

Foram utilizados microdados da PNAD Contínua Trimestral, para os anos de 2019 a 2022, coletados a partir do módulo *PNAD_SOCIAL* do STATA, desenvolvido pelo projeto Data Zoom da PUC Rio, que permite o emparelhamento dos indivíduos consultados pelo IBGE a partir da técnica proposta por Ribas e Soares (2008). Assim, às variáveis padrão são adicionados novos elementos de identificação *hous_id* e *ind_id* para identificação do domicílio e do indivíduo, respectivamente, caso sejam sorteados novamente na nova amostra. Estas informações serão úteis para a observação do mesmo indivíduo em dois ou mais trimestres consecutivos e registro de sua evolução no mercado de trabalho, para posterior geração das matrizes de transição e regressões contra as suas características e os períodos de análise, especialmente o que antecede e o em que se vive a pandemia.

O módulo aplicado às planilhas trimestrais da PNADC, baixadas diretamente do portal do IBGE (<https://www.ibge.gov.br/estatisticas/sociais/trabalho/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html?=&t=microdados>), retornou quatro arquivos de

dados .dta para cada ano de análise. Estes foram, em seguida, importados para o R, onde foram unidos em um mesmo banco de dados, via empilhamento, e feito o filtro das variáveis relevantes, conforme o script *importação_filtro.R*, tratado a seguir.

2.2. Filtragem dos Dados

No processo de filtro dos dados da PNADC já limpos e emparelhados pelo módulo anterior, prosseguimos para a seleção apenas das variáveis relevantes para o método desta pesquisa. Seguindo o convencionado pelo IBGE e estudos da OIT sobre mercado de trabalho, consideramos como Jovens apenas indivíduos na faixa etária de 14 a 29 anos de idade, e filtramos, restringindo às observações cuja variável de código *V2009* (que indica a idade relatada na semana de referência) se situassem neste intervalo de idades.

Para identificação do indivíduo, além das variáveis de emparelhamento geradas pela ferramenta do Data Zoom (*ind_id* foi renomeada para *id_pessoa*), e da idade, selecionamos variáveis idiossincráticas que podem ser úteis para análise dos impactos heterogêneos entre os mesmos jovens, como:

- *V2007*: Sexo (se homem, 1; se mulher, 2)
- *V1022*: Situação do domicílio (se urbana, 1; se rural, 2)
- *V2010*: Cor ou raça (branca, 1; preta, 2; parda, 4; e outras)
- *V3009A*: Escolaridade (fundamental ou menos, 1-3; médio, 4-5; superior, 6)

Seguindo a literatura que compõem o referencial teórico da pesquisa, consideramos a questão da acumulação de experiência, além da qualificação, na inclusão e manutenção no mercado de trabalho. Como não existe nenhuma forma direta de mensurar esta dimensão do capital humano, consideramos como proxy do que denominamos “experiência” a idade do indivíduo, mas propriamente construímos três faixas de idade no intervalo etário de estudo: Adolescente (14-17), Jovem (18-24) e Jovem adulto (25-29).

Para análise do estado ocupacional, dentro do arcabouço do modelo de Search & Matching, é conveniente trabalhar apenas com jovens dentro da força de trabalho, o que nos levou à seleção das seguintes variáveis ocupacionais:

- *VD4001*: Condição em relação a força de trabalho (na força de trabalho, 1; fora da força de trabalho, 2)
- *VD4002*: Condição de ocupação (ocupada, 1; desocupada, 2)

Assim, criamos uma variável indicadora, *estado_ocupacional*, que assume valor 1 para as observações em que se está dentro da força de trabalho e ocupado, ou 0, para aquelas em que se está desocupado, mas ainda assim está dentro da força de trabalho. Ou seja, observar apenas jovens que estão ativamente no mercado de trabalho, seja trabalhando, seja buscando emprego.

Concluída esta etapa, de definição do conjunto de dados sobre os quais queremos proceder nossas análises dos impactos do choque econômico desencadeado pela pandemia de COVID-19 sobre a transição dos jovens no mercado de trabalho, passamos para a estimação dessas transições.

3. Construção da Matriz de Markov

Para a confecção das matrizes de transição a lá Markov, principiamos o emparelhamento de observações consecutivas do estado ocupacional de um mesmo indivíduo. Para isso, inicialmente ordenamos os dados por código de identificação individual, ano e trimestre, a fim de organizarmos observações do mesmo indivíduo uma abaixo da outra. Em seguida, agrupamos todas as observações pelo mesmo id, criamos variável *estado_t1* e armazenamos nela o estado ocupacional da observação posterior. Analiticamente isso significa,

$$E_{t1_t} = E_{t+1}$$

Depois, aplicamos um filtro para garantir que estejamos observando apenas registros consecutivos, ou seja, trimestre após trimestre, para que seja possível capturar as transições no espaço de um trimestre.

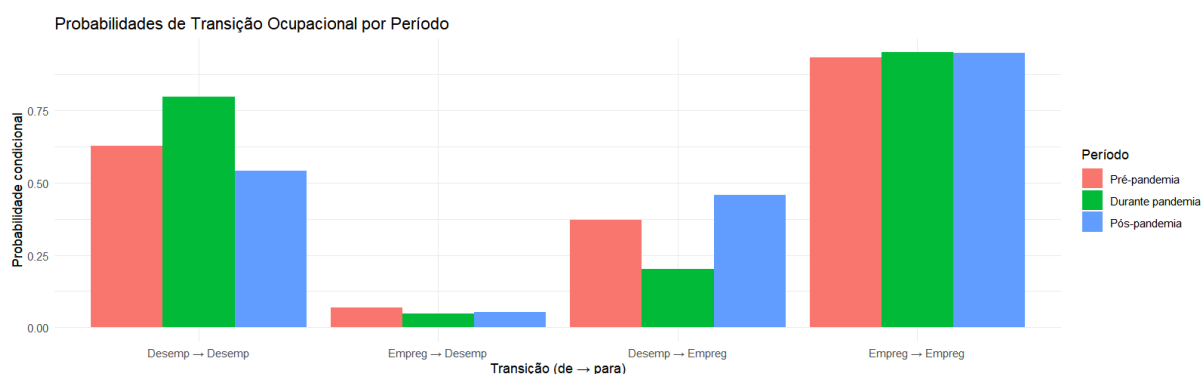
Com as informações contidas nestas duas variáveis, *estado_ocupacional* e *estado_t1*, conseguimos criar uma matriz que os unem como pares ordenados e contabilizam quantas vezes cada uma dessas combinações aparecem no conjunto de dados. Dividindo cada medida pelo total de observações, obtivemos as proporções, ou probabilidades condicionais de transição entre os estados ocupacionais 0 e 1, ou seja, entre o desemprego e emprego.

Como o que nos interessa são as matrizes de cada período de análise para comparação posterior, geramos uma função que estima matrizes de Markov para cada categoria da variável *período* especificada. Por exemplo, a seguinte tabela com matrizes para cada espaço temporal foi confeccionada:

Pré-pandemia			Durante pandemia		Pós-pandemia	
	Desemprego	Emprego	Desemprego	Emprego	Desemprego	Emprego
Desemprego	0,628	0,372	0,798	0,202	0,542	0,458
Emprego	0,068	0,932	0,049	0,951	0,052	0,948

A partir desses dados, já foram possíveis verificações de muitas afirmações da literatura consultada e adotadas como hipóteses de pesquisa, como a histórica dificuldade de inclusão juvenil no mercado de trabalho. Antes mesmo da pandemia, embora só esteja sendo observada as proporções de transição do ano de 2019, pode-se afirmar que, as chances de ingresso de um jovem no mercado de trabalho entre um trimestre e outro, controlada por todas as variáveis pessoais que possam impactar isso, são baixíssimas ($\approx 37,2\%$). Em média, a pandemia realmente teve um impacto negativo, reduzindo-as para cerca de 20,2%; o que pode ser explicado pela redução da atividade econômica, mas não pela rotatividade ou desemprego, já que a probabilidade de transição para o desemprego caiu, e a de permanência no emprego, aumentou (na verdade, nos três períodos, é muito alta). Já a explicação dos dados de retomada é um desafio a hipótese de trabalho, que supunha uma deterioração da participação dos jovens no mercado de trabalho; no entanto, as especulações não ficam de todas equivocadas, dado que a expansão dos serviços de delivery, especialmente por conta das restrições de aglomeração e trânsito durante a pandemia, tem uma fisionomia juvenil. Em outras palavras, o jovem encontrou nesse tipo de prestação de serviços um mercado laboral alternativo, dado a dificuldade de integração ao mercado formalizado.

Graficamente ficará melhor a visualização dessas chances de transição nos diferentes cenários estudados:



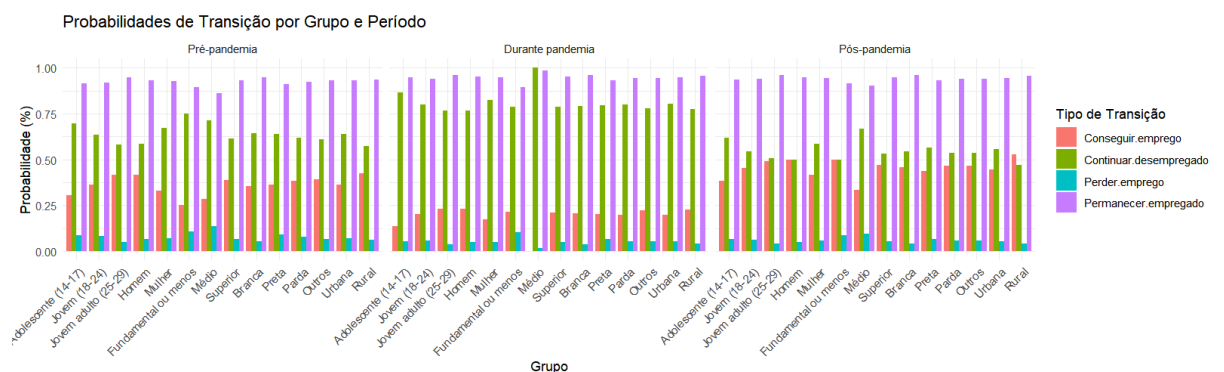
É visível que as chances de conseguir uma ocupação em um intervalo de três meses, para os jovens de 14 a 29 anos, é baixíssima, e, portanto, o tempo no desemprego é desalentador,

especialmente no período pandêmico. Mas a diferença entre as chances de ficar desempregado e de conseguir emprego se reduzem no pós-pandemia. Que mudanças na distribuição dos postos de trabalho, por exemplo, podem explicar esse fenômeno? Em pouco tempo as características do jovem que o tornam atrativo para o mercado de trabalho, e, portanto, dentro da visão do modelo DMP, aumentam a eficiência do matching, não devem ter mudado muito; mas, como se dá a distribuição dessas especificidades, e como ela impacta nas chances de o trabalhador conseguir uma ocupação nessa “nova economia”? Prosseguimos então com a geração de matrizes por grupo de características individuais.

3.1. Matrizes por Grupo e Período

Para geração dessas matrizes segmentadas tornamos explícitas as variáveis selecionadas no processo de filtragem, e criamos uma função a partir da qual possam ser geradas todas as matrizes de subgrupos.

O seguinte gráfico com as probabilidades de transição pode ser gerado:

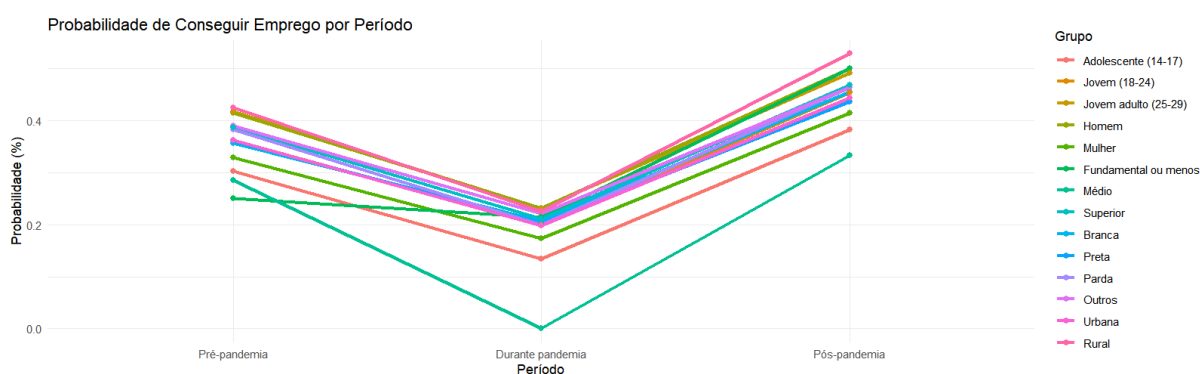


Quanto aos padrões de inclusão, segue-se o já observado na literatura, a pandemia apenas reduz as chances em escala, sem alterar as relações intragrupo. Quanto maior a idade, e na nossa abordagem, maior a experiência, maior as chances de conseguir um emprego; mulheres estão em desvantagem em relação aos homens; maior escolaridade acompanha maior integração, apesar de a diferença ser marginal entre quem tem Ensino Médio e quem tem apenas o Fundamental ou menos; jovens de áreas urbanas tem mais facilidade de transição para o emprego que jovens de regiões rurais, devido a um mercado de trabalho mais espesso.

No entanto, uma configuração nos parece interessante à investigação da pesquisa, pessoas de outras cores ou raças têm vantagens em relação às brancas, no processo de integração ao mercado de trabalho. Preliminarmente, pode-se afirmar que isto está associado a composição dos setores e postos de trabalho, a uma economia de empregos de baixa qualidade e amplo setor

informal, dado às desigualdades raciais consagradas na literatura de ciências sociais. Mas não se deve apegar a esta hipótese, pois os dados da pandemia podem esclarecer que não passa de um grande preconceito, e quem vai discordar dos números?

A pandemia, como já constatado, reduziu igualmente todas as probabilidades de inserção, mas significativamente achatou todas as chances associadas a categorias de raça e origem, elevou a adesão de quem tem fundamental ou menos, e significativamente zerou os de quem tem o ensino médio. Por outro lado, elevou todas as chances de ficar desempregado, e, portanto, o tempo no desemprego.



A retomada que se torna um objeto interessante da pesquisa, mais propriamente, as mudanças no durante pandemia, e o rearranjo da economia na recuperação, que explicam a melhora das chances de emprego, quando o que se vê na literatura é desmantelamento das cadeias de produção, crescente substituição da mão de obra humana por capital tecnológico e otimização de processos.

4. Modelagem Econométrica da Inclusão Juvenil no Mercado de Trabalho

Seguimos então para a modelagem econométrica dessas relações. Como o papel metodológico da pesquisa é aplicar o modelo Diamond-Mortensen-Pissarides ao estudo dos efeitos da pandemia, e, particularmente, o nosso interesse é mensurar essa reverberação no segmento juvenil da força de trabalho, empreenderemos uma estimação da eficiência do matching, que depende, além da disponibilidade de vagas e número de desempregados, de características dos trabalhadores que determinem sua eficiência e seu poder de barganha. Por isso, antes rodamos regressões Logit para as características, visando ver as que afetam a

probabilidade de transição do emprego com alguma significância estatística, inclusive analisando suas interações com o tempo.

Neste objetivo, inicialmente criamos a variável *transição_emprego* que foi nossa variável resposta nas especificações dos modelos. Ela assume valor 1 se a pessoa saiu do desemprego para o emprego, ou seja, se observações cuja variável *estado_ocupacional* é 0 têm como 1 o valor de *estado_t1*.

Após isso, como as especificações dos modelos foram tais que os coeficientes das variáveis explicativas fossem estimados em relação a um grupo base (homem, branca, fundamental ou menos, adolescente (14-17), urbano, pré-pandemia), criamos uma variável binária para mulher (que assume valor 1 se for; 0, caso contrário) e transformamos as outras e fator, tal que permita ao R criar dummies para cada grupo automaticamente.

A primeira especificação do modelo não considerou a interação com o tempo, e assumiu a seguinte forma:

$$P(Y_{it+1} = 1|X_{it}) = \text{logit}^{-1}(X)_{it}'$$

Onde:

- Y_{it+1} : variável dependente *transição_emprego* = 1 se ocorreu transição de desempregado para empregado.
- X_{it} : vetor de características individuais (sexo, raça, escolaridade, etc.) no tempo t.

Os coeficientes de cada variável assumem a forma de log-odds, de forma que se pode dizer que, o coeficiente da variável binária mulher é o efeito de ser mulher comparado a homem, que o coeficiente de *periodoDurante pandemia* é o efeito da pandemia comparado ao pré-pandemia, e assim por diante.

Em código R, a regressão tem a seguinte fórmula:

```
1. modelo_logit <- glm(transicao_emprego ~ mulher + raca + escolaridade + localidade +  
2. experiencia + periodo, data = microdados_transicao, family = binomial)
```

Para obter os coeficientes Odds-ratios (o quanto a elasticidade de uma variável representa da elasticidade da variável base), aplicou-se o exponencial aos coeficientes em log-odds obtidos do sumário da regressão. Ademais, considerou-se significativo apenas estimativas a um grau de 0.05 de significância. Abaixo, o sumário obtido:

Sumário do Modelo Logit Inicial

Variável Explicativa	Coef. Estimado	Odd-ratio	Variação odd	Erro padrão	Significância
Intercepto	-3,241771	0,039094597	-96,09%	0,418956	1,01E-14
Mulher	0,142067	1,152653874	15,27%	0,014793	<2,2e-16
Raça Preta	0,414842	1,514131489	51,41%	0,025315	<2,2e-16
Raça Parda	0,352104	1,42205641	42,21%	0,016389	<2,2e-16
Raça Outras	0,324429	1,38324059	38,32%	0,074964	1,51E-05
Ens. Médio	-0,316711	0,728541273	-27,15%	0,717351	0,65885
Ens. Superior	0,432454	1,541034586	54,10%	0,415477	0,29794
Urbana	0,23352	1,263038088	26,30%	0,018342	<2,2e-16
Jovem (18-24)	-0,167972	0,845377505	-15,46%	0,053886	0,001826
Jovem adulto (25-29)	0,618994	1,857058901	85,71%	0,054413	<2,2e-16
Durante pandemia	-0,459	0,631915245	-36,81%	0,016709	<2,2e-16
Pós-pandemia	-0,189145	0,827666486	-17,23%	0,018754	2,05E-06
AIC	154742,1				
Pseudo-R²	0,0164				
N (observações)	380099				

Estes resultados nos permitem fazer algumas inferências:

- Raças não-brancas tem consistentemente mais chances de conseguir emprego.
- Mulheres, que historicamente tinham menos chances, estão com 15% a mais que os homens.
- O maior dinamismo urbano permite aos jovens dessas regiões uma vantagem de 26% sobre os das áreas rurais.
- Jovens adultos (25-29) têm grande dificuldade de inserção, e por outro lado, quanto mais novo menor a desvantagem em relação aos adolescentes.
- Período da pandemia teve efeito fortemente negativo (-37%), com recuperação parcial no pós (-9%);
- Ter ensino superior ou ao menos o Médio, nesse período, não altera significativamente as chances de transição comparado ao fundamental.

Aqui se confirma, para o período em análise, que, as desigualdades raciais têm pouco efeito sobre o processo de inserção no mercado de trabalho. Como também, que, houve uma melhora significativa na presença das mulheres; o que pode ser parcialmente explicado pela expansão de setores em que elas compõem uma porção significativa da força de trabalho, como o setor de saúde e de cuidados.

No sentido contrário do observado na análise das matrizes de transição por experiência, os resultados da regressão mostram uma maior dificuldade dos mais velhos em aos adolescentes para ingressar no mercado de trabalho. No período pandêmico, especificamente, podemos supor que os mais velhos, devido ao maior risco de vida, foram preteridos aos mais novos, e isto resulte em um maior tempo de desemprego; ou, maior informalidade da economia, que

implique, agora seguindo o modelo de Search do McCall, em maior rejeição de ofertas pelos mais velhos, que são mais exigentes em termos de salário e estabilidade.

Claramente as áreas rurais foram as mais afetadas pela redução da atividade econômica, especialmente devido às medidas de restrição.

De fato, mais uma confirmação é feita do choque negativo da pandemia sobre a ocupação dos jovens no mercado de trabalho. E, de acordo com as estimativas, escolaridade não determinou maior ou menor grau de vulnerabilidade ao choque.

Ainda no esforço de entender o papel da heterogeneidade dos agentes no seu desempenho no mecanismo de matching, avançamos em uma especificação que considere as interações das características com os períodos. Ou seja, caberá a pergunta se, estatisticamente, ser mais jovem, e, portanto, ter menos experiência, ou ter um nível de escolaridade menor e - se considerarmos ter alguma relação causal - ser menos produtivo reduz as chances de admissão, em um cenário de redução da atividade econômica e muitas incertezas.

Assim adotamos um modelo DIF-IN-DIF, ou seja, que isole Diferenças em Diferenças com Heterogeneidade. Sua especificação foi a seguinte:

```
1. modelo_3fases <- glm(transicao_emprego ~ pandemia_fase * (mulher + raca + escolaridade +
2. localidade + experiencia), data = microdados_transicao, family = binomial)
```

Na próxima página, o sumário da regressão:

Modelo Dif-in-Dif					
Variável Explicativa	Coef. Estimado	Odd-ratio	Variação odd	Erro padrão	Significância
Intercepto	-4,194151	0,0150835	-98,49%	-4,141	3,45E-05
Mulher	0,156468	1,1693738	16,94%	6,8036	1,02E-11
Raça Preta	0,488409	1,6297217	62,97%	12,256	<2,2e-16
Raça Parda	0,440727	1,5538369	55,38%	17,100	<2,2e-16
Raça Outras	0,241820	1,2735656	27,36%	1,9302	5,36E-02
Ens. Médio	1,282137	3,6043364	260,43%	1,0296	3,03E-01
Ens. Superior	1,298107	3,6623583	266,24%	1,2861	0,198422

Urbana	0,256428	1,2923066	29,23%	9,0577	<2,2e-16
Jovem (18-24)	-0,158676	0,8532723	-14,67%	-1,989	0,303195
Jovem adulto (25-29)	-0,625733	0,5348691	-46,51%	-7,756	0,198422
Durante pandemia	0,772659	0,5216063	-47,84%	0,658	<2,2e-16
Pós-pandemia	1,360137	1,0361756	3,62%	1,0866	0,046691
Mulher x Durante pandemia	-0,650842	0,8856247	-11,44%	-0,914	8,71E-15
Mulher x Pós-pandemia	0,035536	0,8837556	-11,62%	0,9316	5,11E-01
Raça Preta x Durante pandemia	-0,121462	0,8856247	-11,44%	-2,069	0,277216
Raça Preta x Pós-pandemia	-0,123574	0,8837556	-11,62%	-1,919	0,055524
Raça Parda x Durante pandemia	-0,168657	0,8447984	-15,52%	-4,483	0,351535
Raça Parda x Pós-pandemia	-0,121130	0,8859185	-11,41%	-2,856	0,038474
Raça Outras x Durante pandemia	0,088755	1,0928129	9,28%	0,4953	0,054893
Raça Outras x Pós-pandemia	0,177558	1,1942979	19,43%	0,9563	7,34E-06
Ens. Médio x Durante pandemia	-10,35933	3,16955E-5	-100,00%	-0,277	0,004278
Ens. Médio x Pós-pandemia	-1,659437	0,1902460	-80,98%	-0,938	0,620398
Ens. Superior x Durante pandemia	-1,082614	0,3387087	-66,13%	-0,927	0,338896
Ens. Superior x Pós-pandemia	-1,308413	0,2702485	-72,98%	-1,051	0,781539
Urbana x Durante pandemia	-0,025127	0,9751857	-2,48%	-0,595	0,348191
Urbana x Pós-pandemia	-0,060047	0,9417202	-5,83%	-1,272	0,353821
Jovem (18-24) x Durante pandemia	-0,020860	0,9793552	-2,06%	-0,165	0,293078
Jovem (18-24) x Pós-pandemia	-0,007340	0,9926862	-0,73%	-0,053	0,5514992
Jovem adulto (19-24) x Durante pandemia	0,0574106	1,0590905	5,91%	0,4518	0,203091
Jovem adulto (19-24) x Pós-pandemia	-0,056524	0,9450437	-5,50%	-0,41	0,868419
AIC	169160,2				
Pseudo-R² (McFadden)	0,0167				
N (observações)	380099				

Os resultados desta estimação corroboram os das anteriores. As generalizações possíveis estão organizadas na tabela abaixo:

Grupo	Durante Pandemia	Pós-pandemia
Mulheres	Penalizadas marginalmente (~6%)	Recuperadas (sem efeito no pós)
Pardos	Penalizados significativamente (~15%)	Penalização persiste (~11%)
Pretos	Penalizados (~11%)	Recuperação no pós

Outras Raças	Sem efeito claro	Sem efeito claro
Escolaridade	Sem interação significativa	—
Capital	Benefício mantido, sem mudança	—
Jovens (18–29)	Desvantagem consistente, mas sem mudança ao longo dos períodos	

Podemos concluir então que a pandemia penalizou especialmente os jovens pardos e, em menor grau, pretos e mulheres; e que, apenas as mulheres se recuperaram plenamente no pós-pandemia.

5. Estimação do Modelo de Search & Matching (DMP)

A nossa pretensão inicial de estimação do modelo DMP se refere especificamente à mensuração do tightness θ , ou tensão do mercado de trabalho, que é uma medida da eficiência de matching, para os jovens. Isso nos forneceria uma estimativa do grau de inclusão dos jovens no mercado de trabalho, e internalizado no modelo estrutural, possibilitaria observarmos o seu comportamento diante de choques simulados.

Iniciamos construindo um painel trimestral agregado. Adicionamos uma variável de transição para do desemprego para o emprego, tal qual a que criamos anteriormente, só que simétrica. Depois, agregamos os dados por período e calculamos as seguintes variáveis:

- Estoques:
 - U : total de jovens desocupados, mas buscando trabalho
 - E : total de jovens empregados
 - $N = U + E$: total de jovens na força de trabalho
- De Transições:
 - T_{UE} : Número de transições desempregado-empregado
 - T_{EU} : Número de transições empregado-desempregado
- Indicadoras:
 - Taxas de transição: $f_t = T_{UE}/U$, $s_t = T_{EU}/E$
 - Taxa de desemprego: $u_t = U/(U + E)$
 - Vagas V_t : proxy (não coletada diretamente pelo IBGE), deve ser aproximada a partir da tightness.

Após, foi feita a especificação empírica do modelo de busca ou função de matching, que é:

$$M_t = m \cdot U_t^\alpha \cdot V_t^{1-\alpha}$$

Mas estimamos de forma log-linearizada a partir da seguinte função:

$$\log(f_t) = \beta_0 + \beta_1 \log(\theta_t) + \epsilon_t$$

Com:

$\beta_1 = 1 - \alpha$: estimativa da elasticidade da função de matching

$$\theta_t = \frac{V_t}{u_t}$$

No entanto, como não temos acesso ao estoque de vagas, nossa estratégia empírica inicial consistiu em utilizar como proxy a proporção de admissões, mas, apesar de gerar um coeficiente elevado (elasticidade > 1), trata-se de um caso de endogeneidade estrutural, pois $f_t = T_{UE}/U$ e $\theta_t = E/U$ são construídos com a mesma base.

Uma alternativa foi utilizar dados de ocupação total trimestral do IPEADATA como proxy para o estoque de vagas da economia (sabendo que isso pressupõe plena utilização da capacidade instalada). Apesar de teoricamente mais plausível, a elasticidade estimada foi negativa e não significativa, provavelmente pela rigidez do estoque de ocupados totais diante de variações específicas entre os jovens.

Por fim, utilizamos dados de admissões formais mensais disponibilizados pelo Novo CAGED. Agregamos as observações por trimestre, e construímos uma nova variável tightness, tal que:

$$\theta_{CAGED,t} = \frac{\text{Admissões formais}}{U_t}$$

Reestimamos a equação $\log(f_t) = \beta_0 + \beta_1 \log(\theta_t) + \epsilon_t$ e chegamos a um resultado que reforça a direção estabelecida pela teoria entre as variáveis, mas, a elasticidade continua insignificante.

Estimação MDP			
Variável Explicativa	Coef. Estimado	Erro padrão	Significância
Intercepto	-3,3157	2,1747	0,151
log(theta_caged)	0,3047	0,3274	0,369
AIC	22,59096		
N (observações)	15		

Em resumo, as três alternativas de estimação da tightness resultam em:

- Com proxy interna (E/U): $\beta_1 \approx 1.33$, $R^2 > 0.6$ (mas enviesado)
- Com total de ocupados: $\beta_1 < 0$, não significativo
- Com CAGED: $\beta_1 = 0.30$, sinal correto, mas $p > 0.3$ (baixo poder estatístico)

6. Abordagem alternativa

Antes de avançar na estimação do modelo estrutura, decidimos testar duas novas abordagens para melhorar estatisticamente a estimativa do θ , para isso consideramos um conjunto mais amplo de dados (para todas as idades), e na nova estimativa da tightness introduzimos a idade como variável dependente; no outro caminho, tentamos aumentar o intervalo de tempo (mas o máximo possível foi de 2012-2023).

6.1. Considerando todas as idades

Nesta nova abordagem desconsideramos a variável *experiencia*, utilizando a transformando a em um indicador de idades *faixa_idade*, que assume quatro valores: Adolescente (14-17), Jovem (18-24), Jovem adulto (25-29) e Adulto (>29).

Assim, rodamos novamente a regressão logit, obtendo o seguinte resultado:

Sumário do Modelo Logit Alternativo					
Variável Explicativa	Coef. Estimado	Odd-ratio	Variação odd	Erro padrão	Significância
Intercepto	-2,766041	0,06291057	-93,71%	0,074979	<2,2e-16
Mulher	0,083917	1,08753862	8,75%	0,008930	<2,2e-16
Raça Preta	0,561192	1,75276054	75,28%	0,014893	<2,2e-16
Raça Parda	0,475963	1,60956346	60,96%	0,009895	<2,2e-16
Raça Outras	0,293186	1,34069213	34,07%	0,045154	8,42E-11
Ens. Médio	-0,510502	0,60019420	-39,98%	0,060143	<2,2e-16

Ens. Superior	-0,10304	0,90209088	-9,79%	0,053671	0,0548783
Urbana	0,267021	1,30606787	30,61%	0,011445	<2,2e-16
Jovem (18-24)	-0,18858	0,82813425	-17,19%	0,052264	0,00031
Jovem adulto (25-29)	-0,646726	0,52375775	-47,62%	0,052709	<2,2e-16
Adulto (>29)	-1,215882	0,29644843	-70,36%	0,010051	<2,2e-16
Durante pandemia	-0,401396	0,66938493	-33,06%	0,011429	<2,2e-16
Pós-pandemia	-0,104719	0,90057754	-9,94%	0,011429	<2,2e-16
AIC	469205				
Pseudo-R² (McFadden)	0,0316				
N (observações)	1796369				

Os resultados enfatizam a redução nas chances de ser admitido na pandemia, em relação as chances no período anterior. Especialmente, quanto mais velho o trabalhador maior a sua desvantagem em relação aos adolescentes no período pandêmico que no ano de 2019.

Também se observa a mesma vantagem das raças não-brancas. E, forte desvantagem de quem tem ensino médio em relação as pessoas de menor escolaridade.

O modelo de DIF-IN-DIF também foi reestimado:

Sumário do Modelo Dif-in-Dif Alternativo					
Variável Explicativa	Coef. Estimado	Odd-ratio	Variação odd	Erro padrão	Significância
Intercepto	-2,684251	0,06827232	-93,17%	0,1121923	<2,2e-16
Mulher	0,062702	1,06471046	6,47%	0,014077	0,4272817
Raça Preta	0,589310	1,80274517	80,27%	0,0235913	<2,2e-16
Raça Parda	0,506621	1,659675	65,97%	0,01566	<2,2e-16
Raça Outras	0,319200	1,37602760	37,60%	0,0711294	7,20E-06
Ens. Médio	-0,522898	0,59279994	-40,72%	0,0909792	9,06E-09
Ens. Superior	-0,155111	0,85631986	-14,37%	0,0818141	0,0579734
Urbana	0,222608	1,24933174	24,93%	0,0176582	<2,2e-16
Jovem (18-24)	-0,177455	0,83739826	-16,26%	0,0768269	0,0208987
Jovem adulto (25-29)	-0,647778	0,52320689	-47,68%	0,0775801	<2,2e-16
Adulto (>29)	-1,231868	0,29174691	-70,83%	0,0760978	<2,2e-16
Durante pandemia	-0,619022	0,53847069	-46,15%	0,1742337	8,15E-06

Pós-pandemia	-0,150999	0,85984812	-14,02%	0,1902112	0,0003811
Mulher x Durante	-0,003155	0,99684957	-0,32%	0,0204381	0,8773041
pandemia					
Mulher x Pós-	0,096343	1,10113690	10,11%	0,0232233	3,35E-05
pandemia					
Raça Preta x Durante	-0,064692	0,93735546	-6,26%	0,0344213	0,0601843
pandemia					
Raça Preta x Pós-	-0,018939	0,98123892	-1,88%	0,0381244	0,6193475
pandemia					
Raça Parda x Durante	-0,070527	0,93190239	-6,81%	0,0225825	0,001786
pandemia					
Raça Parda x Pós-	-0,020029	0,98017024	-1,98%	0,0258953	0,4392494
pandemia					
Raça Outras x Durante	-0,050829	0,95044089	-4,96%	0,105062	0,6285257
pandemia					
Raça Outras x Pós-	-0,033552	0,96700414	-3,30%	0,1141838	0,7688752
pandemia					
Ens. Médio x Durante	0,004217	1,00422600	0,42%	0,1386849	0,9757416
pandemia					
Ens. Médio x Pós-	0,047736	1,04889413	4,89%	0,154804	0,7578033
pandemia					
Ens. Superior x					
Durante pandemia	0,128183	1,13676123	13,68%	0,1241748	0,301941
Ens. Superior x Pós-	0,033650	1,03422297	3,42%	0,1369736	0,8059369
pandemia					
Urbana x Durante	0,089511	1,09364034	9,36%	0,0262344	0,0006449
pandemia					
Urbana x Pós-	0,055398	1,05696151	5,70%	0,0296692	0,0618737
pandemia					
Jovem (18-24) x					
Durante pandemia	-0,005987	0,99403058	-0,60%	0,1225604	0,9610372
Jovem (18-24) x Pós-	-0,023075	0,97718831	-2,28%	0,1321801	0,8614102
pandemia					
Jovem adulto (19-24) x					
Durante pandemia	0,062657	1,06466191	6,47%	0,1235213	0,6119734
Jovem adulto (19-24) x					
Pós-pandemia	-0,075579	0,92720610	-7,28%	0,1334359	0,5711151
Adulto (>29) x Durante	0,100311	1,10551501	10,55%	0,1214978	0,4090181
pandemia					
Adulto (>29) x Pós-	-0,070634	0,93180268	-6,82%	0,1230999	0,58997524
pandemia					
AIC	169160,2				
Pseudo-R²					
(McFadden)	0,0317				
N (observações)	1796369				

Nenhuma nova informação a respeito dos jovens pode ser obtida. Assim partimos para a estimação da tightness e ainda obtivemos um resultado coerente com a teoria, mas estatisticamente não significativo.

Estimação MDP - Alternativo			
Variável Explicativa	Coef. Estimado	Erro padrão	Significância
Intercepto	-2,737200	1,92	0,178
log(theta_caged)	0,2682	0,3241	0,423
AIC	469205		
N (observações)	1796369		

6.2. Estimação da tightness com base de dados mais longa