



Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas
Públicas
Departamento de Economia

Pedro Henrique Medeiros dos Santos
Orientadora: Prof^a Dr^a Ana Carolina Pereira Zoghbi

DIFERENCIAL SALARIAL ENTRE FORMAIS E INFORMAIS NO BRASIL: Uma análise de impacto da pandemia de 2020

Brasília - DF

2024

Resumo

O objetivo deste trabalho é investigar o impacto da pandemia de 2020 no diferencial de rendimentos de trabalhadores formais e informais. Para tanto, considerando a literatura existente sobre o tema, optou-se pela abordagem metodológica da combinação de um balanceamento por *propensity score matching* com o procedimento de diferenças em diferenças para identificar os efeitos médios de tratamento sobre os tratados que indicam alterações no diferencial entre ambos os grupos. Os resultados apontam para uma heterogeneidade do impacto no setor informal, com consequências notadamente negativas para os trabalhadores domésticos e do setor privado. Não foram encontrados indícios de alteração significativa para indivíduos posicionados como empregadores e conta própria.

Palavras-chave: informalidade, *propensity score matching*, diferenças em diferenças, diferencial de rendimentos, mercado de trabalho.

Abstract

The objective of this study is to investigate the impact of the 2020 pandemic on the earnings differential between formal and informal workers. To this end, considering the existing literature on the subject, the methodological approach chosen was the combination of propensity score matching with the differences-in-differences procedure to identify the average treatment effects on the treated, indicating changes in the differential between both groups. The results point to a heterogeneity of impact in the informal sector, with notably negative consequences for domestic and private sector workers. No significant changes were found for individuals positioned as employers and self-employed.

Keywords: informality, propensity score matching, differences-in-differences, earnings differential, labor market.

Sumário

Introdução	5
1. Revisão de Literatura	6
1.1. O conceito de informalidade	6
1.2. O diferencial de rendimentos entre formais e informais	7
1.3 A Informalidade no Brasil.....	9
2. Dados e Metodologia	11
2.1. O Propensity Score Matching.....	12
2.2. A técnica de Diferenças em Diferenças (DiD)	15
3. Estimação e Resultados	16
3.1. O modelo de probabilidade	16
3.2. Resultados	20
3.2. Testes de heterogeneidade.....	22
3.2.1 Trabalhadores do Setor Privado	22
3.2.2 Empregadores e Conta Própria.....	23
3.2.3. Trabalhadores Domésticos	24
4. Conclusão	25
Referências Bibliográficas	26

Introdução

Caracterizado pela ausência de vínculos formais de emprego e de cobertura pela seguridade social, o trabalho informal é um fenômeno ubíquo no mundo, afetando praticamente todos os países independentemente do seu grau de desenvolvimento. A economia informal é especialmente prevalente nas economias emergentes, onde o Brasil figura com destaque (ILO, 2008). Efetivamente, o mercado de trabalho brasileiro é caracterizado por um acentuado nível de informalidade. Entre 1992 e 1999, a proporção de trabalhadores nessa situação em relação a força de trabalho variou de 55,7% para 57,1%, um aumento em partes explicado por uma mudança estrutural da economia brasileira com a retração da indústria, intensiva em postos formais, e a expansão do setor de serviços, mais propenso a informalidade (RAMOS, 2002). A partir dos anos 2000 iniciou-se uma trajetória de queda, chegando a uma taxa de 43,5% em 2014. No entanto, com a crise nesse período, a informalidade retomou uma tendência crescente, atingindo 47,5% da força de trabalho em 2019 (BARBOSA FILHO e VELOSO, 2022)

Com a pandemia em 2020 e o seu brutal impacto na economia, a queda dos empregos formais caiu cerca de 4,1%, ao passo que no setor informal, no qual as condições de emprego não permitiram em muitos casos contornar as medidas de distanciamento social, a retração foi ainda maior, com uma queda de aproximadamente 12,6% da empregabilidade (BARBOSA FILHO e VELOSO, 2022). Não obstante muitas atividades tenham voltado aos níveis de ocupação pré-pandemia desde então, a informalidade voltou a crescer significativamente após a crise sanitária, atingindo um número recorde de 12,9 milhões de pessoas em 2022.

As estimativas supracitadas podem variar em razão das dificuldades de estudar o setor informal, que compreendem não somente os desafios de mensuração inerentes a sua natureza de produção, resultando na subestimação de registros estatísticos e consequente exclusão das contas nacionais, mas também pela própria falta de consenso acerca da definição de atividade econômica informal, cuja literatura é marcada por uma multiplicidade de critérios tanto em trabalhos empíricos quanto teóricos (ULYSSEA, 2006). Nesse contexto, a próxima seção será destinada a situar brevemente a conceituação da informalidade, definir o escopo deste trabalho e realizar uma rápida análise da situação da economia informal no Brasil.

1. Revisão de Literatura

1.1. O conceito de informalidade

Embora a origem do termo “setor informal” tenha surgido no início dos anos 70 com estudos produzidos pela Organização Internacional do Trabalho (OIT) em países africanos, sua incipiente conceituação foi pouco rigorosa, buscando simplesmente caracterizar uma realidade no mundo do trabalho marcadamente diferente dos países centrais. A partir dela se seguiu o desenvolvimento de outros esforços teóricos no âmbito do Programa Regional do Emprego para América Latina e o Caribe (PREALC), que contribuíram para o desenvolvimento de uma concepção de informalidade a partir da noção de excedente estrutural. Tal perspectiva vislumbra a existência da economia informal como uma consequência da insuficiência dos setores mais modernos, cujos vínculos trabalhistas seriam regulados pelo marco legal, em criar empregos formais via demanda por fator trabalho. Nesse contexto, sem mecanismos suficientemente desenvolvidos de amparo e proteção social, o desemprego não poderia ser uma opção em razão das necessidades de subsistência, e a aderência à informalidade resultaria da oferta excedente de trabalho. Esse excedente, por sua vez, compreenderia os indivíduos segregados do processo de desenvolvimento, ou seja, unidades de produção orientadas ao mercado e caracterizadas por alguns aspectos particulares como baixa produtividade e não dissociação do proprietário dos meios de produção em relação ao processo de trabalho. Ademais, além de determinantes estruturais, também é concebível que a ocorrência de volatilidade na magnitude do setor informal seria originada por mudanças conjunturais, seja por transformações na composição setorial de uma região ou por resultado de flutuações ao longo do ciclo econômico. (RAMOS C. 2012)

Num esforço de tentar desenvolver diretrizes para o dimensionamento da informalidade, em 1993 a OIT estabeleceu, por meio da 15ª Conferência Internacional dos Estatísticos do Trabalho, recomendações para avaliar o que seria considerado setor, ou economia informal. Ao adaptar e operacionalizar estatisticamente as recomendações para o Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE), no âmbito da pesquisa Economia Informal e Urbana (ECINF), delimitou como o universo amostral todas as unidades econômicas de propriedade de trabalhadores por conta própria e de empregadores com até cinco empregados, moradores de áreas urbanas, sejam elas a atividade principal ou secundária de seus proprietários.

Conforme exposto em Machado e Antigo (2008), na literatura de mercado de trabalho brasileira existe uma confusão entre atividade, relação de trabalho e setor. A delimitação mais

frequentemente utilizada é a da regulação do trabalho, que considera como informais os trabalhadores assalariados sem carteira assinada, ou seja, aqueles cujas atividades produtivas e eventuais relações de trabalho geralmente não são abrangidas por disposições formais legais, ou se encontram na margem da legislação vigente. Essa fronteira de definição pode também se estender significativamente com a inclusão dos trabalhadores por conta própria e firmas sem CNPJ, de modo que a informalidade pode, ao considerá-los, compreender aproximadamente metade de toda a força de trabalho do país. Neste trabalho, optou-se por definir indivíduos informais de forma semelhante ao critério utilizado pelo IBGE para calcular a taxa de informalidade atualmente, isto é, trabalhadores sem carteira assinada e empregadores e conta própria sem CNPJ.

1.2. O diferencial de rendimentos entre formais e informais

Como evidenciado pelos números apresentados na introdução, o impacto da pandemia no que diz respeito às diferenças entre os postos de trabalho formais e informais se mostrou significativamente heterogêneo, ressaltando disparidades existentes entre os trabalhadores de ambos os setores. Com efeito, uma temática estudada com certa frequência por especialistas em mercado de trabalho é a do diferencial de rendimentos existente entre trabalhadores formais e informais. Ao que consta na literatura, as disparidades salariais podem ser explicadas de várias formas, como capital humano, discriminação, segmentação, salário eficiência ou mesmo poder de barganha (Machado e Antigo, 2008). No Brasil, destacam-se estudos que investigaram o diferencial de rendimentos testando a hipótese de segmentação, que de maneira semelhante à concepção de excedente estrutural, explica a disparidade existente na renda de trabalhadores igualmente produtivos pelas características do seu setor de atuação. Tais características podem estar associadas a disparidades tecnológicas (razão capital-trabalho), estruturas de mercado imperfeitas, legislação trabalhista e existência de sindicatos [(Barros e Reis, 1990) (Barros, 1993)]. Além disso, a diferença salarial entre trabalhadores com e sem carteira assinada também pode ser explicada por características atinentes a estrutura e ao desempenho produtivo das empresas, que por sua vez estariam relacionadas à opção de contratar trabalhadores na modalidade formal e, paralelamente, com a qualidade dos empregos ofertados (Ulysea, 2006).

Nesse sentido, a evidência para a segmentação seria a constatação de um diferencial de rendimentos que persiste, ainda que se controle a outros fatores, como o nível educacional. Com efeito, existe significativa produção bibliográfica indicando que trabalhadores informais

seriam, em média, menos educados [Barros et al. (1993), Cacciamali e Fernandes (1993), Fernandes (1996), Neri (2002), Tannuri-Pianto e Pianto (2002) e Soares (2004)]. Esse resultado é razoavelmente amparado no âmbito da teoria microeconômica que diz respeito à decisão de acumular capital humano com os resultados de Carneiro e Henley (2012), que indicam um menor retorno em termos de rendimentos para os trabalhadores informais nos níveis médio e superior, e parcialmente corroborados em Tannuri-Pianto e Pianto (2002) para os quantis inferiores de rendimento. Para além do nível educacional, verifica-se também uma heterogeneidade no retorno da experiência para formais e informais. Kassouf (1998) mostra que o retorno para homens é maior no setor formal, ao passo que para mulheres ocorre o inverso, sendo os anos de experiência mais recompensados para as trabalhadoras informais.

Por outro lado, estudos como Carneiro e Henley (2001) e Menezes Filho, Mendes e Almeida (2001) testemunham contra a hipótese de segmentação, sugerindo que o diferencial é explicado essencialmente por outras características, observáveis ou não, dos trabalhadores de cada setor. Em outras palavras, concluiu-se nesses estudos a existência de um viés de seleção derivado de características não observadas, que explicariam tanto o fato de os indivíduos serem informais (escolha não aleatória), quanto seu nível de rendimento. Tannuri-Pianto e Pianto (2002) sugerem uma sofisticação para essa explicação, propondo uma segmentação parcial do mercado de trabalho. Constatou-se no estudo por meio de regressões quantílicas corrigidas para viés de seleção que o efeito da segmentação seria mais prevalente nos quantis inferiores da distribuição de renda, ao passo que para os quantis superiores a diferença seria causada principalmente por atributos individuais dos trabalhadores. Nessa perspectiva, apenas uma parte do diferencial seria explicada pela segmentação.

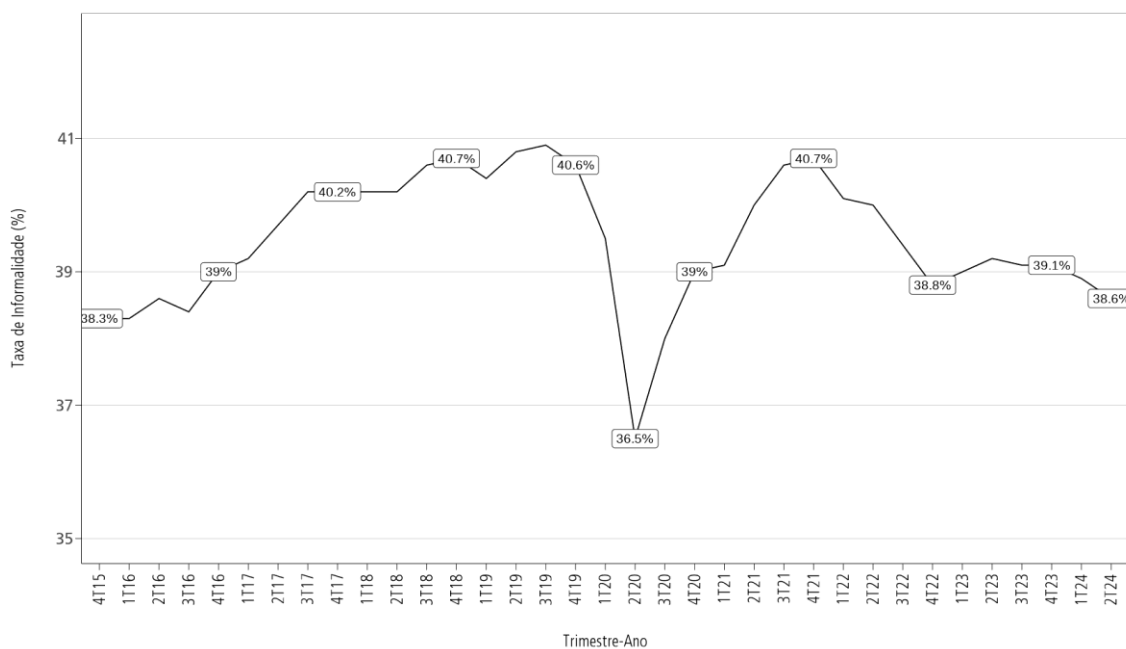
No curso de seu recrudescimento, a análise do diferencial salarial entre trabalhadores com e sem carteira assinada tem sido objeto de significativos avanços metodológicos na literatura, com especial destaque ao tratamento de viés de seleção. O reconhecimento de que a distribuição dos trabalhadores entre os setores não se dá de forma aleatória resultou no uso de métodos de correção como o procedimento de Heckman (Carneiro e Henley 2001; Tannuri-Pianto e Pianto, 2002), modelos de escolha (Kassouf, 1998; Scandiuzzi, 1999) e até mesmo variáveis instrumentais (Menezes Filho et al. 2004). Os resultados conflitantes encontrados sedimentam a ausência de consenso acerca dos determinantes do diferencial de salários entre os setores formal e informal, e salientam a necessidade de realização de mais estudos concernentes a essa temática.

1.3. A Informalidade no Brasil

Esta seção tem como objetivo apresentar um retrato da informalidade no Brasil de 2015 a 2024. Para tanto, são utilizados dados oficiais divulgados pelo IBGE no âmbito da plataforma SIDRA e, para tabulações não disponíveis nessa plataforma, utilizou-se os microdados da PNAD Contínua Trimestral aplicando-se os respectivos pesos amostrais.

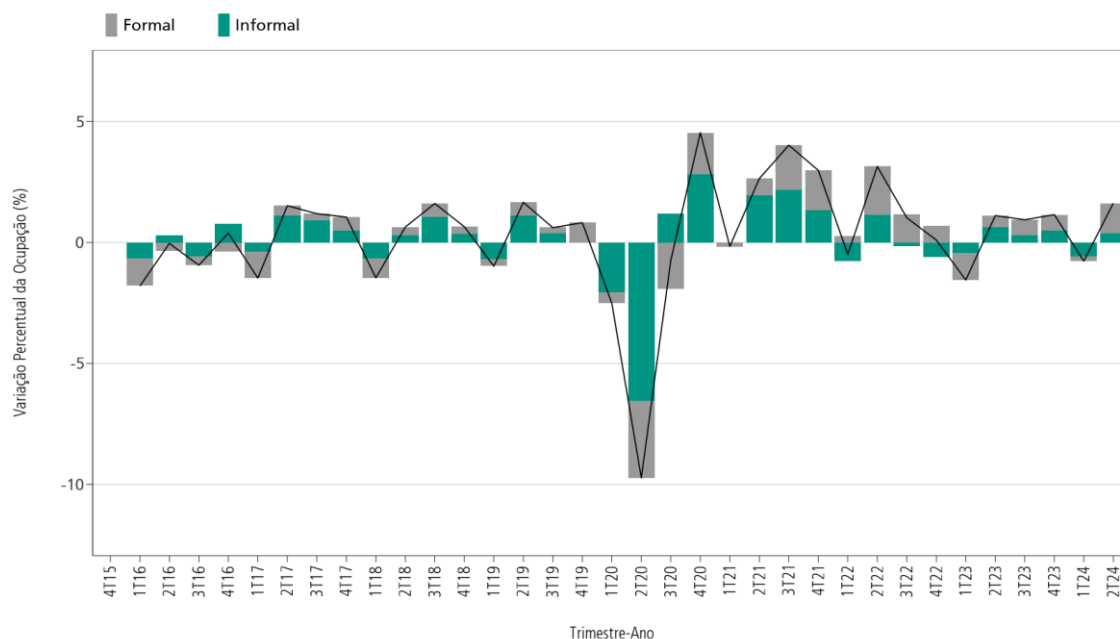
A informalidade no Brasil de 2015 até 2024 variou, em geral, entre 38% e 41% da população ocupada considerando o critério utilizado pelo IBGE — trabalhadores sem carteira assinada e conta própria sem CNPJ. O maior valor do período foi registrado no terceiro trimestre de 2019, e o menor no segundo trimestre de 2020, isto é, logo após a conflagração da pandemia. Esse resultado é parcialmente explicado pela queda mais que proporcional no número de indivíduos informais ocupados. Nesse contexto, é importante mencionar que, devido às medidas de distanciamento social, as entrevistas realizadas pelo IBGE passaram a ser conduzidas por telefone durante a pandemia. Isso gerou dificuldades em obter os números de telefone dos domicílios selecionados, especialmente daqueles que seriam entrevistados pela primeira vez. Como resultado, houve um aumento substancial na taxa de não resposta a partir dos primeiros trimestres de 2020, o que levou o instituto a investigar se a perda de observações poderia ter introduzido vieses na pesquisa. Os estudos, no entanto, não encontraram indícios de vieses de não resposta para os estratos geográficos mais agregados.

Figura 1 – Taxa de Informalidade (2015-2024)



No que diz respeito a variação da taxa de ocupação ilustrada na figura 2, a composição entre formais e informais se mostrou equilibrada no período que se inicia no quarto trimestre de 2015 e se estende até o último trimestre de 2019.

Figura 2 – Composição da variação da taxa de ocupação (%)



No início da pandemia, a queda da população ocupada foi predominantemente composta pela retração no número de indivíduos em situação de informalidade, indicando um impacto mais significativo sobre esse grupo. A recuperação do choque verificado nos trimestres iniciais de 2020 ocorreu logo em seguida, e o novo contingente de ocupados se distribuiu de forma relativamente equilibrada entre ambos os grupos até 2022. A partir desse ano podemos observar uma criação de postos de trabalho formais maior, embora ainda bem distribuída.

No que se refere a atividade econômica, analisando os microdados do quarto trimestre de 2019, a informalidade se mostra mais distribuída do que a formalidade. Para ambos os grupos, o grupamento de atividades que compreende o maior número de trabalhadores é o de comércio, reparação de veículos automotores e motocicletas. Aproximadamente metade da formalidade é composta por três grandes grupamentos, a saber, a indústria geral; educação, saúde humana e serviços sociais; e informação, comunicação e atividades financeiras, imobiliárias, profissionais e administrativas. A informalidade por sua vez concentra cerca de 35% do seu contingente nos seguintes grupamentos: agricultura, pecuária, produção florestal, pesca e aquicultura; construção e serviços domésticos. Os dados corroboram a baixa produtividade do setor informal no país, onde o maior número de trabalhadores dessa categoria se encontra em setores caracterizados por baixo valor agregado.

Tabela 1 – Distribuição entre os grupamentos da atividade principal (2019-4)

Atividade	Formal	Informal
Comércio, reparação de veículos automotores e motocicletas	20.47%	17.55%
Agricultura, pecuária, produção florestal, pesca e aquicultura	3.4%	16.49%
Construção	3.75%	12.59%
Serviços domésticos	3.04%	11.57%
Indústria geral	15.68%	9.23%
Outros Serviços	3.52%	8.41%
Alojamento e alimentação	4.86%	7.73%
Informação, comunicação e atividades financeiras, imobiliárias, profissionais e administrativas	14.5%	6.51%
Transporte, armazenagem e correio	4.55%	6.21%
Educação, saúde humana e serviços sociais	17.44%	3.6%
Administração pública, defesa e seguridade social	8.79%	0.07%
Atividades mal definidas	0.01%	0.05%
Total	100,00%	100,00%

2. Dados e Metodologia

No intento de apresentar uma contribuição para a literatura de diferenciais salariais entre os setores formal e informal, o presente trabalho propõe como forma de analisar o impacto da pandemia no diferencial salarial a combinação dos procedimentos de propensity score matching (PSM) e diferenças em diferenças (DiD). Os dados utilizados serão extraídos dos microdados da PNAD Contínua dos anos de 2019, 2020 e 2021, que são divulgados pelo IBGE. O PSM será a estratégia adotada para tentar mitigar o viés de seleção amostral, e o método de diferenças em diferenças será utilizado para captar o diferencial por meio do cálculo do efeito médio do tratamento sobre os tratados (ATT). A ideia é utilizar o pareamento para construir grupos de controle e tratamento comparáveis a partir de variáveis observáveis.

A variável categórica que indica se um indivíduo se situa ou não na informalidade foi construída considerando os trabalhadores domésticos e do setor privado sem carteira assinada, assim como os empregadores e conta própria sem CNPJ. Optou-se pela exclusão de indivíduos que trabalhavam sem receber rendimentos, grupo majoritariamente composto por trabalhadores familiares auxiliares.

Considerando que o pareamento deve ser realizado pré-tratamento e como o IBGE não divulga chaves que permitam a identificação longitudinal dos indivíduos em cada um dos

arquivos de microdados publicados (apenas de domicílios), optou-se pela utilização de um algoritmo para a construção do painel identificado a nível de indivíduo. A metodologia, proposta por Ribas e Soares (2008) para a Pesquisa Mensal de Emprego (PME), consiste em identificar cada indivíduo a partir de um conjunto de características consideradas fixas no tempo, tomando como base o código do domicílio divulgado pelo instituto. O procedimento também incorpora estratégias avançadas para reduzir o atrito associado a causas diversas, como mobilidade geográfica das pessoas, recusa de entrevistas e imprecisão nas respostas. Desenvolvido no âmbito do projeto Data Zoom pela iniciativa de membros do Departamento de Economia da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO), o algoritmo de identificação foi operacionalizado por meio da extensão “datazoom_social”, no software estatístico STATA 15, utilizando-se os arquivos de microdados trimestrais dos anos supracitados. No tratamento dos dados e realização das regressões foi usada a linguagem R, onde para fazer o pareamento foi usado o pacote “MatchIt”. A coorte selecionada foi a de indivíduos pertencentes a domicílios entrevistados em algum trimestre igual ou anterior ao primeiro trimestre de 2020 e que foram identificados em pelo menos um período subsequente até o momento esperado para sua saída da amostra da pesquisa. Em virtude da possibilidade de que um indivíduo salte de um grupo para o outro ao longo do tempo, isto é, da formalidade para a informalidade ou vice-versa, foi aplicada uma segunda filtragem que censura os indivíduos na amostra a partir do momento em que mudam sua condição em relação a situação inicial registrada. Reconhece-se a possibilidade de que tal filtragem pode ter acrescentado um viés nas estimativas.

Adicionalmente, o regressando (log do salário hora) considerado na estimação do ATT foi transformado para contornar problemáticas associadas a aplicação do logaritmo natural em variáveis que podem assumir valores muito próximos ou mesmo iguais a zero. Aqui vale ressaltar que, embora seja adotada neste trabalho, é necessário cautela na interpretação em termos de percentuais para modelos log-nível onde a variável dependente pode assumir tais valores extremos ou indefinidos. Ademais, como há sensibilidade dos coeficientes obtidos em relação a transformação realizada, foi utilizada uma abordagem recorrente na literatura, qual seja, a de somar uma unidade na variável antes de aplicar o logaritmo. Destarte, seja *salh* o salário hora, a variável dependente *lnsalh* foi calculada como $\ln(1 + salh)$.

2.1. O Propensity Score Matching

Uma dificuldade central da pesquisa em ciências sociais é a usual impossibilidade de obtenção de dados experimentais. Nesse contexto, o consequente uso de dados observacionais disponíveis exige uma série de cuidados ao extrair conclusões, especialmente pela existência de vieses capazes de levar o analista incauto a resultados equivocados. Um dos vieses mais significativos é o de seleção, que dentre outras causas pode acontecer quando a probabilidade de uma determinada unidade observada ser atribuída ao tratamento é dependente da variável de resultado em análise. Seja T uma *dummy* de tratamento e Y a variável de resultado analisada, o efeito médio do tratamento sobre os tratados (ATT) é definido por:

$$ATT = E(Y(1) - Y(0) | T = 1) = E[Y(1) | T = 1] - E[Y(0) | T = 1]$$

Onde o contrafactual não observado por definição dentre ambos os resultados potenciais é: $E[Y(0) | T = 1]$. Ou seja, a esperança da variável de resultado para a mesma unidade tratada em um cenário onde ela não recebeu o tratamento. Por não ser possível observar ao mesmo tempo ambos os resultados, o enquadramento do modelo causal de Rubin sugere a necessidade de se utilizar um substituto apropriado para o contrafactual no cálculo do efeito causal médio. Em estudos com dados não experimentais (atribuição não aleatória), o resultado médio dos não tratados, definido como $E[Y(0) | T = 0]$, frequentemente não é um bom substituto em virtude da existência de variáveis confundidoras que afetam concomitantemente a variável de resultado e a probabilidade de tratamento. Isso significa que mesmo na ausência do tratamento existiriam diferenças significativas. Essa diferença seria exatamente o viés seleção, ou mais precisamente, o viés de auto seleção. Somando e subtraindo $E[Y(0) | T = 0]$ na equação, obtemos:

$$ATT + E[Y(0) | T = 1] - E[Y(0) | T = 0] = E[Y(1) | T = 1] - E[Y(0) | T = 0]$$

Onde o termo que soma o ATT no lado direito da equação é o viés associado, e o lado esquerdo da equação corresponde ao efeito observado extraído da diferença entre a média da variável de resultado para ambos os grupos. Se a atribuição ao tratamento fosse feita de forma aleatória, garantindo a independência da variável de resultado em relação a probabilidade de tratamento, o viés de seleção seria igual a zero pois o efeito médio seria o mesmo para os indivíduos tratados e não tratados.

Com o fito de lidar com essa problemática, Rosenbaum e Rubin (1983) introduziram o propensity score matching. Esse procedimento quasi-experimental é utilizado para emular grupos comparáveis a fim de lidar com o viés de seleção e obter estimativas mais acuradas sobre o efeito médio do tratamento. O procedimento consiste em balancear a distribuição dos grupos por meio de um escore calculado a partir de um vetor de características observáveis mensuradas pré-tratamento. O escore de propensão utilizado indica a probabilidade de uma unidade ser tratada condicional ao vetor de covariáveis supramencionado. O balanceamento é

então realizado por meio de algum método de pareamento que busca correspondências aproximadas de escore entre os integrantes de cada um dos grupos. O êxito desse procedimento em eliminar eventuais vieses depende da satisfação da hipótese de independência condicional (HIC). Seja R a variável de resultado, p a probabilidade de tratamento e X o vetor de covariáveis; A HIC é definida da seguinte forma:

$$Y \perp p|X$$

Que significa a independência (ou ortogonalidade) da variável de resultado em relação à probabilidade condicional de ser tratado. Neste trabalho, a variável de resultado é o rendimento efetivo de todos os trabalhos, a probabilidade condicional de tratamento é calculada por meio de uma regressão logística (logit) que tem como variável dependente a *dummy* de informalidade segundo os critérios adotados, e como regressores um conjunto de variáveis captadas pelo questionário da PNAD no qual a escolha será discutida adiante. Assume-se que a hipótese forte de que a independência condicional é satisfeita para as variáveis observadas escolhidas.

De acordo com Caliendo e Kopeinig (2008), a aplicação do PSM implica a realização de diversas escolhas. Para algumas, como o as que dizem respeito ao modelo de probabilidade a ser utilizado na estimação dos escores, existe pouca ou nenhuma recomendação na literatura. Já para outras, como a escolha do algoritmo de pareamento, a forma de checagem das hipóteses fundamentais e a avaliação da qualidade do pareamento é possível encontrar algumas orientações.

No que diz respeito ao algoritmo de pareamento, sua escolha pode ser de relevância quando existem poucas observações, ou seja, métodos que tendem a podar uma quantidade maior de dados podem não ser tão adequados nesse cenário. O algoritmo escolhido para a realização deste estudo foi o de vizinho próximo (*nearest neighbour*) sem possibilidade de reposição e caliper de 0,1. Ele consiste em parear os indivíduos com o escore de propensão mais próximo em termos absolutos. A impossibilidade de reposição significa que um determinado indivíduo não tratado não pode ser o controle de mais do que um indivíduo tratado. O *caliper* de 0.1 que delimita a distância máxima dos escores de dois indivíduos pareados foi escolhido arbitrariamente. A abordagem sem reposição foi escolhida em razão da relativamente grande quantidade de dados disponíveis.

Para que a aplicação do PSM na construção de um contrafactual apropriado à estimação do ATT seja válida, é necessário outra hipótese principal além da HIC seja satisfeita, qual seja, a hipótese de suporte comum (HSC). Ela consiste na constatação da existência de unidades de

observação com a mesma probabilidade de receber o tratamento em ambos os grupos. Uma forma de verificar essa hipótese é a construção de um gráfico de densidade com a frequência relativa de cada escore observado. Essa hipótese será testada na seção 3, onde será apresentado o modelo de probabilidade utilizado para auferir os escores de propensão.

2.2. A técnica de Diferenças em Diferenças (DiD)

Ao investigar relações de causalidade em determinado experimento buscamos comparar mudanças na variável dependente em função de um tratamento. Uma técnica utilizada para realizar inferências a respeito do efeito causal é a de Diferenças em Diferenças (DiD). Ela consiste em comparar a diferença na variável de resultado intragrupo ao longo do tempo e, então, tomar a diferença da diferença entre ambos os grupos. Ao tomar a diferença temporal das variáveis de resultado dentro de um mesmo grupo no tempo, o procedimento capta a mudança global na variável de resultado, causada tanto de fatores observados (como o tratamento, no caso do grupo de tratados) quanto fatores não observados. No grupo de controle, a diferença obtida corresponde ao efeito devido a outros fatores ao longo do tempo. Se a hipótese de tendências paralelas for satisfeita, então espera-se que o resultado das unidades tratadas também tenha sofrido o mesmo efeito, de modo que ao tomar a diferença da diferença podemos isolar do efeito global sobre os tratados a parcela correspondente ao tratamento. Esse procedimento é amplamente utilizado na análise de inferência causal.

Neste trabalho, a combinação do propensity score matching com a técnica de diferenças em diferenças é utilizada visando propiciar ainda mais robustez nos resultados obtidos. Para isso, considerando que o objetivo é investigar a dinâmica do diferencial em múltiplos períodos, foi utilizada uma abordagem diferente em relação ao modelo canônico de diferenças em diferenças restrito a dois períodos. A metodologia utilizada é apresentada em Callaway e Sant’Anna (2021) e foi operacionalizada por meio do pacote “did” no software R Studio. Em suma, esse método flexibiliza o modelo de diferenças em diferenças convencional para os casos em que existem múltiplos períodos e é possível que as unidades sejam tratadas em diferentes momentos. Nesse contexto, os autores propõem o *group-time average treatment effect*, que corresponde ao efeito médio de tratamento sobre os tratados no período t para o grupo de unidades tratadas no período g . Dentre as diversas alternativas de identificação elencadas no artigo, optou-se pela utilização do $ATT_{or}^{nev}(g, t)$, onde o subscrito *or* significa que o procedimento utilizado foi o de “outcome regression” e o sobrescrito *nev* indica que as unidades

do grupo de controle se manterão nesse grupo ao longo do tempo. Como foi realizado o pareamento por propensity score matching na etapa anterior, nenhuma covariável adicional foi acrescida ao modelo. Por fim, note que os indivíduos tratados no caso deste estudo, que são os trabalhadores informais que atravessaram a pandemia nessa condição, constituem apenas um único grupo (g) de momento do tratamento, que é o das unidades tratadas no primeiro semestre de 2020.

3. Estimação e Resultados

3.1. O modelo de probabilidade

Entendido como a probabilidade de um indivíduo ser tratado condicional a um vetor de covariáveis, o escore que será utilizado para o pareamento foi calculado a partir de um modelo de regressão logística. As variáveis explicativas foram escolhidas por exercerem influência, por hipótese, tanto na variável de rendimento médio efetivo de todos os trabalhos quanto na probabilidade de tratamento. O objetivo desta seção é explicar a motivação da escolha de cada uma e, por fim, apresentar a equação estimada.

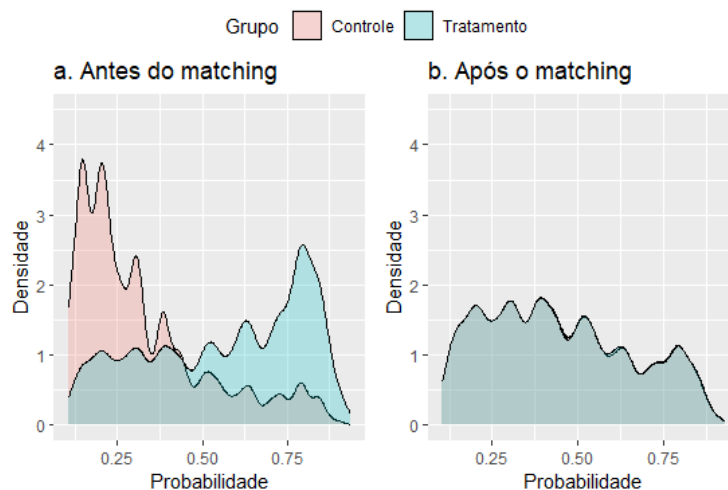
A partir da consideração da literatura de diferenciais salariais brevemente sumarizada na seção 1, onde foi constatado que os trabalhadores informais são, em média, menos escolarizados, optou-se pela inclusão de variáveis binárias de nível educacional tomando como base os indivíduos com ensino fundamental incompleto. Além disso, como o fenômeno da informalidade é mais prevalente em determinados setores da economia, também foram incluídas variáveis categóricas de grupamento da atividade econômica do trabalho principal para cada setor compreendido no questionário tomando como grupo base o grupamento de comércio, reparação de veículos automotores e motocicletas que, como constatado, é o setor no qual se concentram a maioria dos trabalhadores informais. Com o acréscimo dessas *dummies*, espera-se que o nível médio de escolaridade e a participação relativa em cada setor seja próxima para cada um dos grupos. Adicionalmente, outras variáveis como região de residência (urbana ou rural) e sexo foram incluídas. O modelo estimado para as variáveis mensuradas pré-tratamento (pt) portanto, foi:

$$\begin{aligned} inform = & urb_{pt} + mulher_{pt} + ppi_{pt} + seminst_{pt} + fundcomp_{pt} + medioincomp_{pt} + supincomp_{pt} + supcomp_{pt} \\ & + agro_{pt} + ind_{pt} + cons_{pt} + sd_t + tac_{pt} + icfipa_{pt} + apds_{pt} + edss_{pt} + outro_{pt} + amd_{pt} \\ & + sd_{pt} \end{aligned}$$

Onde *inform* é uma dummy de tratamento, *urb*, *mulher* e *ppi* correspondem a variáveis de região de residência, sexo e raça, respectivamente. As variáveis *seminst*, *fundcomp*, *medioincomp*, *supincomp* e *supcomp* são dummies de nível de instrução, e os demais regressores como *agro*, *ind*, *cons*, *sd*, *tac*, *icfipa*, *apds*, *edss*, *amd* e *outro* são variáveis categóricas de grupamento de atividade principal do empreendimento do trabalho principal.

O pareamento foi aplicado considerando a entrevista realizada no trimestre imediatamente anterior ao momento do tratamento, que para fins de balanceamento foi considerado como o segundo trimestre de 2020. Essa escolha se deu por duas razões, a primeira é o fato de que podemos delimitar sem muita controvérsia o início da pandemia no Brasil em meados de março de 2020, de modo que podemos afirmar de maneira razoável que muitas entrevistas nesse período foram realizadas pré-tratamento. A segunda é o ganho em observar mais um trimestre na análise ao considerar esse período como o momento do tratamento para fins de balanceamento. Em outras palavras, conseguimos observar o impacto até o início de 2021 incidente sobre a coorte de indivíduos que entraram na pesquisa em janeiro e fevereiro de 2020. Ao aplicar o modelo no pareamento, podemos verificar que a hipótese de suporte comum é satisfeita:

Figura 3 - Hipótese de suporte comum



Como podemos observar no gráfico a da figura 1, antes da realização do matching havia uma região de suporte comum, isto é, um intervalo no qual existem indivíduos com a mesma probabilidade de tratamento em ambos os grupos. Essa constatação nos permite dar continuidade a aplicação do balanceamento a partir do escore de propensão obtido com o modelo de probabilidade acima exposto. A realização do matching com os parâmetros supramencionados na seção 2.1 nos retorna uma amostra cujo gráfico equivalente de densidade

se encontra à direita na figura (b). O aparente êxito se deve, principalmente, a numerosa quantidade de dados na amostra. Podemos também avaliar a qualidade do procedimento a partir da média das variáveis explicativas utilizadas. As tabelas abaixo demonstram algumas informações sobre o pareamento.

Tabela 2 – Estatísticas do pareamento

	Controle	Tratamento
Total	69.744	44.166
Pareados	26.173	26.173
Não Pareados	42.631	17.993
Descartados	940	0

Foram pareados 26.173 indivíduos de um total de 44.166 unidades tratadas. Adicionalmente, podemos avaliar o pareamento na tabela 3, onde constam as respectivas médias antes e após o pareamento para os regressores mensurados no trimestre disponível imediatamente anterior ao tratamento. Como esperado, podemos verificar disparidades significativas nas médias, especialmente em variáveis como grupamento da atividade principal e nível educacional. Para um pareamento sem reposição, podemos considerar que o matching teve êxito em balancear ambos os grupos.

Tabela 3 – Médias antes e depois do pareamento

Variável	Média				
	Antes do Matching		Após o Matching		
	Tratados	Controle	Tratados	Controle	Dif. de Média
distance	0.574	0.270	0.463	0.463	*
urb_pt	0.657	0.871	0.781	0.803	
mulher_pt	0.354	0.452	0.392	0.392	*
ppi_pt	0.641	0.512	0.600	0.589	
seminst_pt	0.049	0.008	0.020	0.018	
fundcomp_pt	0.099	0.062	0.099	0.096	*
medioincomp_pt	0.076	0.047	0.077	0.076	*
supincomp_pt	0.034	0.075	0.051	0.055	*

supcomp_pt	0.069	0.314	0.110	0.122	
agro_pt	0.273	0.051	0.131	0.130	*
ind_pt	0.078	0.143	0.126	0.119	
cons_pt	0.116	0.029	0.066	0.072	
tac_pt	0.057	0.040	0.082	0.089	
icfipa_pt	0.052	0.128	0.084	0.085	*
apds_pt	0.0003	0.115	0.0004	0.0004	*
edss_pt	0.025	0.209	0.041	0.041	*
outro_pt	0.070	0.028	0.063	0.071	
amd_pt	0.001	0.0002	0.0003	0.0004	*
sd_pt	0.112	0.028	0.067	0.071	*

* Não há diferença estatisticamente significativa na média (5%)

3.2. Resultados

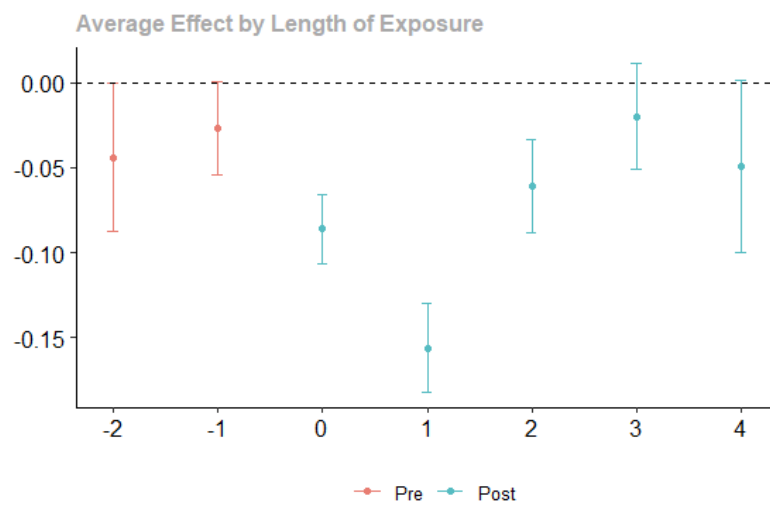
Estimando um modelo de diferenças em diferenças incluindo cada trimestre dos anos escolhidos, obtemos resultados que indicam que o gap salarial existente entre trabalhadores formais e informais foi ampliado com a pandemia, embora esse efeito tenha apresentado uma leve reversão ao final do ano de 2020 e início de 2021. Os efeitos médios de tratamento foram estimados tomando como base sempre o período imediatamente anterior. No primeiro trimestre de 2020, o efeito médio de tratamento sobre os tratados indicativo de alterações no diferencial salarial foi de aproximadamente -8,8% para os dados pareados, uma estimativa estatisticamente significativa a 5% que denota um aumento na diferença da remuneração por hora trabalhada em desfavor dos informais. Esse efeito foi provavelmente limitado pelo fato de que as medidas de distanciamento social foram iniciadas no final desse trimestre, em meados de março. De fato, com o degrading da crise sanitária, os resultados apontam que o gap apresentou maior ampliação em pontos percentuais já no segundo trimestre de 2020, onde o diferencial na remuneração foi majorado 15,7 p.p em desfavor dos trabalhadores informais. No entanto, essa tendência de alta no diferencial em desfavor da informalidade apresentou desaceleração já no terceiro trimestre de 2020, onde foi de cerca de 6.00 p.p, até se tornar estatisticamente não significativa no quarto trimestre desse ano. No início de 2021 o coeficiente estimado indica uma última ampliação no diferencial salarial de 4,96%, e a indisponibilidade dos dados em períodos posteriores em virtude do esquema de rotatividade amostral da pesquisa nos impede de analisar como o diferencial evoluiu para a mesma coorte de indivíduos analisada. Além disso, os coeficientes encontrados são relativamente semelhantes tanto para os dados básicos (sem pareamento) quanto para o modelo estimado após o balanceamento com matching (PSM), sendo os ATT's dos dados pareados maior em termos absolutos para todos os trimestres. Por fim, não obstante a significância estatística dos resultados encontrados, o teste da hipótese de tendências paralelas indica que não é possível atestar sua validade. Com efeito, o teste indica para uma rejeição da hipótese mesmo para níveis de significância tão baixos quanto 1%. É possível que tal resultado decorra da heterogeneidade nos rendimentos existentes entre os diferentes tipos de trabalhadores considerados. Por exemplo, a prevalência de trabalhadores por conta própria na informalidade e a existência de reajustes pontuais para o grupo de controle como decorrência de reajustes no salário-mínimo podem ter contribuído para comprometer a similaridade da tendência de rendimentos.

Tabela 4 – Efeitos médios de tratamento em cada trimestre

Grupo (g) [momento do tratamento]	Período (t)	ATT(g,t) [PSM]	ATT(g,t) [sem pareamento]
2020.1	2020.1	-0.0882*	-0.0811*
2020.1	2020.2	-0.1577*	-0.1402*
2020.1	2020.3	-0.0632*	-0.0578*
2020.1	2020.4	-0.0247	-0.0187
2020.1	2021.1	-0.0548*	-0.0493*
Observações		n = 215.638	n = 478.734
P-valor da hipótese de tendências paralelas:		0.00	0.00

* Significância a 5%

Figura 4 – Intervalos dos efeitos médios de tratamento



3.2. Testes de heterogeneidade

A fim de investigar o impacto para diferentes categorias de trabalhadores e tentar alcançar grupos de controle e tratamento nos quais a hipótese de tendências paralelas seja satisfeita, procederemos nesta seção a aplicação da metodologia proposta nos dados filtrados de forma a manter apenas determinados grupos de interesse. O filtro aplicado manteve na base apenas os indivíduos que não mudaram de categoria de ocupação ao longo do tempo.

3.2.1 Trabalhadores do Setor Privado

Tabela 5 – Trabalhadores do Setor Privado

Grupo (g) [momento do tratamento]	Período (t)	ATT(g,t) [PSM]	ATT(g,t) [sem pareamento]
2020.1	2020.1	-0.0711*	-0.0759*
2020.1	2020.2	-0.0428*	-0.0354*
2020.1	2020.3	0.0424	-0.0405
2020.1	2020.4	0.0673	0.0691
2020.1	2021.1	0.0695	0.0716
Observações		n = 69.076	n = 195.703
P-valor da hipótese de tendências paralelas:		0.2992	0.03803

* Significância a 5%

Filtrando a base apenas para trabalhadores do setor privado e aplicando o PSM novamente podemos analisar o diferencial nesse grupo utilizando a metodologia proposta. De acordo com os resultados encontrados, o impacto da pandemia no diferencial salarial de formais e informais dentro do grupo de trabalhadores do setor privado se mostrou estatisticamente significativo nos dois primeiros trimestres, sendo que o maior aumento no diferencial em pontos percentuais se deu já no trimestre inicial, de cerca de 7.1 pontos percentuais (p.p.). A tendência de elevação do diferencial em desfavor dos informais se manteve no segundo trimestre, onde o gap foi ampliado em cerca de 4.2 p.p.. Tais dados contrastam com as estimativas inicialmente apresentadas na medida em que o impacto naquele caso se deu no segundo, e não no primeiro trimestre de 2020. É possível criar várias hipóteses que explicam esse fato, como por exemplo sazonalidades ou o efeito do reajuste do salário-mínimo no primeiro trimestre, favorecendo os

trabalhadores formais comparativamente, assim como a maior exposição dos trabalhadores informais a reduções salariais dada a flexibilidade característica de vínculos dessa natureza. Embora as consequências mais graves da pandemia não tenham se concentrado no primeiro trimestre de 2020, é plausível crer que os agentes já antecipassem os riscos e impactos da crise sanitária. De fato, como foi exposto na seção 1, a queda na ocupação posterior à deflagração da pandemia se deu inicialmente de forma mais prevalente sobre a informalidade, evidenciando a maior fragilidade dessas posições de ocupação. Por fim, como demonstrado na tabela, a análise restrita a esse grupo possibilitou a não rejeição da validade da hipótese de tendências paralelas após o balanceamento dos grupos de controle e tratamento, conferindo maior robustez aos resultados encontrados.

3.2.2 Empregadores e Conta Própria

Tabela 6 – Empregadores e Conta Própria

Grupo (g) [momento do tratamento]	Período (t)	ATT(g,t) [PSM]	ATT(g,t) [sem pareamento]
2020.1	2020.1	-0.0066	0.0098
2020.1	2020.2	-0.0011	0.1175*
2020.1	2020.3	0.0352	0.0947*
2020.1	2020.4	0.0435	0.0790*
2020.1	2021.1	0.0396	0.0779
Observações		n = 59.694	n = 133.942
P-valor da hipótese de tendências paralelas		0.3854	0.0442

* Significância a 5%

Filtrando para empregadores e trabalhadores por conta própria obtemos resultados estatisticamente não significantes para todos os trimestres com os dados pareados, indicando que a pandemia não alterou o diferencial salarial para esse grupo. Já para os dados sem pareamento, foram encontrados ATT's estatisticamente significantes e positivos, indicando uma redução do diferencial salarial em favor dos trabalhadores informais classificados como empregadores ou conta própria. Contudo, embora os resultados encontrados para os dados sem pareamento sejam estatisticamente significativos, a hipótese de tendências paralelas é rejeitada aos níveis de significância convencionais. A não rejeição desta hipótese para os dados pareados evidencia o êxito do balanceamento pelo score de propensão em construir grupos de controle e tratamento passíveis de serem comparados.

3.2.3. Trabalhadores Domésticos

Tabela 7 – Trabalhadores Domésticos

Grupo (g) [momento do tratamento]	Período (t)	ATT(g,t) [PSM]	ATT(g,t) [sem pareamento]
2020.1	2020.1	-0.0943*	-0.1202*
2020.1	2020.2	-0.1802*	-0.1838*
2020.1	2020.3	-0.0852*	-0.0915*
2020.1	2020.4	0.0034	-0.0283*
2020.1	2021.1	-0.0396	-0.0773
Observações		n = 13.578	n = 23.744
P-valor da hipótese de tendências paralelas		0.4492	0.2030

* Significância a 5%

Para os trabalhadores domésticos podemos verificar um impacto significativo no diferencial salarial em desfavor dos informais no período que compreende o início da crise sanitária. Os resultados indicam uma dinâmica de três ampliações do diferencial salarial estatisticamente significantes iniciadas já no primeiro trimestre de 2020. O maior efeito médio de tratamento estimado foi no segundo trimestre daquele ano, com uma ampliação da diferença salarial de cerca de 18 p.p. em relação ao grupo de controle. Ademais, a magnitude relativamente maior dos coeficientes indica que os trabalhadores domésticos foram especialmente afetados em termos de renda pela crise sanitária, um resultado razoavelmente esperado se considerarmos que se trata de uma ocupação caracterizada pela proximidade física entre empregador e empregado na medida em que este último presta serviços no próprio domicílio daquele. Os receios aventados pela disseminação do vírus e consequentemente pela necessidade de distanciamento social podem ter resultado em demissões, reduções de jornada ou mesmo diminuições de salário para o grupo de tratamento, dadas suas características mais flexíveis. Ressalta-se que o diferencial em questão, assim como em todos os casos apresentados até aqui, pode ser causado tanto por fatores que afetam o salário hora do grupo de controle quanto de tratamento. Deste modo, assim como existem fatores (como os supracitados) que afetam negativamente o salário dos informais, também é possível que a ampliação do diferencial tenha sido causada por um aumento na variável de resultado dos formais comparativamente ao grupo de controle.

4. Conclusão

Os resultados obtidos a partir da metodologia proposta indicam que a pandemia impactou de forma significativa, relevante e heterogênea o diferencial salarial de trabalhadores formais e informais. Em geral, o impacto da pandemia sobre a variável de resultado se deu no sentido de ampliar a diferença pré-existente no salário hora, expondo um cenário especialmente desfavorável para os trabalhadores em situação de informalidade. Enquanto o efeito foi especialmente deletério para os trabalhadores domésticos informais, não foi possível encontrar coeficientes significativos de efeito médio de tratamento sobre os tratados para os empregadores e conta própria em situação análoga, sugerindo uma maior homogeneidade do impacto da pandemia nessa categoria entre os grupos de controle e tratamento. Em geral, os efeitos se concentraram nos trimestres iniciais e os coeficientes encontrados foram negativos, indicando um aumento da diferença salarial em desfavor dos trabalhadores informais. A metodologia utilizada e a dinâmica da pesquisa não permitiram a análise da dinâmica para períodos posteriores ao primeiro trimestre de 2021. Finalmente, destaca-se que as filtragens aplicadas comprometeram o desenho amostral da pesquisa, de forma que não é possível assegurar a representatividade populacional dos resultados.

Referências Bibliográficas

BARBOSA FILHO, Fernando de Holanda; VELOSO, Fernando. Mercado de Trabalho e Políticas Públicas. Rio de Janeiro: FGV/IBRE, 2022. Disponível em: https://ibre.fgv.br/sites/ibre.fgv.br/files/arquivos/u65/mercado_de_trabalho_e_politicas_publicas_final.pdf. Acesso em: 22 de julho de 2023.

BARROS, R.P.; REIS, J.G.A. “Desigualdade salarial: resultados de pesquisas recentes”. In: CAMARGO, J.M. e GIAMBIAGI, F. (org.). **Distribuição de renda no Brasil**. Paz e Terra, São Paulo, 1991.

BARROS, R.P. “The Informal Labor Market in Brazil”. Mimeo, 1993.

BARROS, R.P., MELLO, R.; PERO, V. “Informal Labor Contracts: a solution or a problem?”. IPEA, Texto para Discussão, n.291, 1993.

CACCIAMALI, M.C.; FERNANDES, R. “Distribuição dos trabalhadores e diferenciais de salários entre mercados de trabalho regulamentado e não regulamentado”, **Pesquisa e Planejamento Econômico**, v. 23, n. 1, pp.135-56, 1993.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. **Journal of Economic Surveys**, v. 22, n. 1, p. 31–72, 2008.

CALLAWAY, B.; SANT’ANNA, P. H. C. Difference-in-Differences with multiple time periods. **Journal of Econometrics**, 2020. Disponível em: <https://doi.org/10.1016/j.jeconom.2020.12.001>. Acesso em: 20 set. 2024.

CARNEIRO, F.G.; HENLEY, A. “Modelling Formal vs. Informal Employment and Earnings: microeconomic evidence for Brazil”, Anais do XXIX Encontro Nacional da Anpec, 2001.

FERNANDES, R. “Mercado de trabalho não-regulamentado: participação relativa e diferenciais de salários”, **Pesquisa e Planejamento Econômico**, v. 26, n. 3, 1996.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Economia Informal Urbana (ECINF). Rio de Janeiro: IBGE, 2003. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv6150.pdf>. Acesso em: 22 de julho de 2023.

KASSOUF, A.L. “Wage gender discrimination and segmentation in the Brazilian labor market”. **Economia Aplicada**, v. 2, 1998.

MACHADO, A. F.; OLIVEIRA, A. M. H. C. DE; ANTIGO, M. Evolução do diferencial de rendimentos entre setor formal e informal no Brasil: o papel das características não observadas. **Revista de Economia Contemporânea**, v. 12, p. 355–388, 1 ago. 2008.

MENEZES-FILHO, N.A.; MENDES, M.; DE ALMEIDA, S.E. “O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção?”. **Revista Brasileira de Economia**, v. 58, n. 2, 2004.

NERI, M.C. “Decent Work and the Informal Sector in Brazil”. EPGE, **Ensaio Econômico**, n. 461, 2002.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO (ILO). Women and men in the informal economy: a statistical picture (third edition). International Labour Office — Geneva. 2018.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO (OIT). Resolução da 15ª Conferência Internacional de Estatísticos do Trabalho. [S.l.: s.n.], 1993. Disponível em: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/meetingdocument/wcms_221686.pdf. Acesso em: 22 de julho de 2023.

_____. Resolução da 15ª Conferência Internacional de Estatísticos do Trabalho. [S.l.: s.n.], 1993. Disponível em: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/meetingdocument/wcms_221686.pdf. Acesso em: 22 de julho de 2023.

RAMOS, Lauro. A evolução da informalidade no Brasil metropolitano: 1991-2001. Brasília: IPEA, 2002. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/5599>. Acesso em: 22 de julho de 2023.

RIBAS, Rafael Perez; SOARES, Sergei Suarez Dillon. Sobre o painel da Pesquisa Mensal de Emprego (PME) do IBGE. IPEA, Texto para discussão, 2008.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41–55, 1983.

SCANDIUZZI, J.C. “Desemprego, informalidade e política salarial no mercado de trabalho brasileiro”. PUC-Rio, tese de doutorado, 1999.

SOARES, F.V. “Some Stylized Facts of the Informal Sector in Brazil”. IPEA, texto para discussão, n. 1020, 2004.

TANNURI-PIANTO, M.E.; PIANTO, D. “Informal Employment in Brazil — A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach”. Anais do XXIV Encontro Brasileiro de Econometria, v. 2, 2002.