

Desigualdade de Oportunidades no Brasil: Uma Decomposição Quantílica Contrafactual*

Erik Alencar de Figueiredo
Programa de Pós-Graduação em Economia
Universidade Federal da Paraíba, Brasil
Pesquisador do CNPq
E-mail: eafigueiredo@gmail.com

Cleiton Roberto da Fonseca Silva
Programa de Pós-Graduação em Economia
Universidade Federal da Paraíba, Brasil
E-mail: cleiton.roberto@yahoo.com.br

March 16, 2011

Resumo: O objetivo deste estudo é mensurar a desigualdade de oportunidades brasileira. Para tanto, utilizam-se a abordagem teórica desenvolvida por Bourguignon et al. (2007) e uma estratégia empírica baseada na conjugação do método de geração de contrafactuais de Chernozhukov et al. (2009), com as inferências quantílicas sob má especificação, propostas por Angrist et al. (2006). Os resultados indicam que variáveis como educação e ocupação dos pais, raça e região de nascimento, produzem efeitos expressivos (diretos e indiretos) sobre a renda dos indivíduos. Em resumo, a proporção da desigualdade de oportunidades na desigualdade total brasileira pode variar entre 29% e 35%.

Palavras-Chave: Desigualdade de oportunidades; regressões quantílicas; método contrafactual.

Classificação no JEL: C22; E31.

*O primeiro autor gostaria de agradecer o suporte financeiro do CNPq – projeto 475225/2009-0.

1 Introdução

As recentes transformações na economia brasileira, como a abertura comercial, estabilização do nível de preços e a consequente criação de um ambiente propício para a adoção dos programas de transferência de renda, foram decisivas para melhora nos indicadores sociais, tais como concentração de renda, pobreza e bem-estar.¹ Em poucas palavras, a desigualdade de renda, que cresceu continuamente entre 1960 e 1990 (Barros & Mendonça, 1995), passou a declinar de forma expressiva e estatisticamente significativa, a partir da segunda metade da década de 1990 (Azevedo, 2007 e Figueiredo & Ziegelmann, 2009), resultando, conseqüentemente, em uma melhora nos níveis de pobreza e indigência (Ramos & Santana, 2003).

Sob o ponto de vista da igualdade de renda estrita, a análise desse cenário permite concluir que o Brasil vem se tornando uma nação mais justa. Porém, os desenvolvimentos recentes da literatura propõem algumas contestações. Na verdade, nem mesmo os “modern equalitarians”, como Rawls (1971), Dworkin (1981), Arneson (1989), entre outros, consideram o conceito da perfeita igualdade de renda como a norma ideal para a justiça. Inspirados pela tradição rawlsiana, autores como Kolm (1972), Sen (1985) e Roemer (1996, 1998) advogam que as diferenças de renda devem ser tratadas como uma função das diversas necessidades, esforços ou preferências individuais.

Uma forma de captar essa diversidade é considerar o conceito de **desigualdade de oportunidades**.² Sob esse enfoque, os resultados econômicos dos indivíduos passam a ser observados como uma função de fatores de responsabilidade e de não responsabilidade (Roemer, 1998). Em outras palavras, parte da renda é determinada por variáveis de esforço, tais como nível educacional, decisão de migrar, horas trabalhadas por ano, entre outras e; a outra por fatores que fogem do controle dos agentes econômicos, variáveis circunstância, ou seja, background familiar (nível educacional e ocupação dos pais), atributos individuais como raça, gênero, idade ou região de nascimento, entre outras.³ Uma das principais implicações desse novo conceito é que so-

¹Grande parte dessas mudanças se deu a partir da segunda metade da década de 1990. Para detalhes, com atenção específica para a distribuição de renda e pobreza, ver Neri (2006).

²Outra forma é adotar a literatura relacionada à identificação das desigualdades justa e injusta. Para uma introdução ao tema, ver Thomson (2011).

³Alternativamente, a desigualdade de oportunidades pode ser mensurada a partir da comparação de distribuições condicionadas. Nesse caso, o foco é direcionado para o conjunto de oportunidades ao qual os indivíduos estão inseridos. Essa visão foi proposta, inicialmente, por Pattanaik & Xu (1990) e Kranich (1996) e desenvolvida, posteriormente, por Lefranc et al. (2004). A idéia é simples: a igualdade de oportunidades prevalece se e somente se os indivíduos associados a um particular conjunto de circunstâncias, não

mente a desigualdade oriunda das variáveis de circunstância é socialmente indesejável.⁴

Porém, como calcular a desigualdade de oportunidades? Uma alternativa é decompor a iniquidade de uma determinada distribuição de renda, $I(Y^a)$,⁵ a partir da criação de duas distribuições contrafactuais. Uma livre da desigualdade relacionada às circunstâncias, $I(Y^c)$, e a outra onde não há desigualdade resultante do esforço, $I(Y^e)$. Diante disso, a desigualdade de oportunidades pode ser mensurada a partir de dois indicadores: no primeiro se considera a diferença entre as desigualdades da distribuição atual e da distribuição isenta de diferenciais de circunstâncias $I(Y^a) - I(Y^c)$ e; o segundo sumarizando a desigualdade oriunda apenas dos fatores de circunstância, $I(Y^e)$. Para tanto, faz-se necessário classificar o esforço individual como uma variável **observável** ou **não observável**.

Dentro do grupo de estudos que considera o esforço como uma variável observável, destaca-se o paper influente de Bourguignon et al. (2007). Analisando a desigualdade de oportunidades brasileira, os autores propõem uma abordagem paramétrica onde os rendimentos são explicados pelas variáveis de circunstância e de esforço, postulando que as variáveis de circunstância possuem efeito sobre as de esforço. Ou seja, as circunstâncias são endógenas e possuem impactos diretos e indiretos sobre a renda.

Com base nessa estrutura, os autores criam duas distribuições contrafactuais. A primeira livre de efeitos (diretos e indiretos) das circunstâncias e a segunda isenta de somente dos efeitos diretos da circunstâncias. Seus resultados sugerem que as variáveis de circunstância correspondem a cerca de 23% da desigualdade total, onde, 60% são efeitos diretos e 40% indiretos. O background familiar se apresenta como o principal fator de determinação do nível de oportunidades, dado que cerca de 70% do efeito das circunstâncias se deve a escolaridade do pai.

Ao considerar a variável de esforço como um fator não observável, a construção das distribuições contrafactuais necessita de uma suposição adicional, qual seja: duas pessoas com diferentes níveis de oportunidade, mas no mesmo percentil da sua distribuição condicionada ao seu tipo, empreendem o mesmo nível de esforço. Essa condição é reconhecida como Roemer's Identification Axiom – RIA (Roemer, 1996).

Dentro desse segundo enfoque de investigação, destaca-se o estudo de Checchi & Peragine (2010). Em sua estratégia empírica, os autores dividem

apresentarem vantagens em relação a um outro grupo de indivíduos pertencentes a um outro conjunto de circunstâncias.

⁴Nesse ponto há um importante avanço em relação as teorias tradicionais como, por exemplo, a de Atkinson (1970), onde toda a desigualdade é socialmente indesejável.

⁵Sendo I um índice de desigualdade como o Gini ou Theil, por exemplo.

a população em n tipos e m percentis (ou tranches), a partir da criação de uma matriz M^a de dimensão $n \times m$, representando a distribuição de renda vigente. Essa estrutura faz com que cada combinação do tipo com o tranche, corresponda a um nível de renda. Como supõe-se que em cada tranche o nível de esforço é o mesmo, a distribuição contrafactual é obtida a partir da substituição das colunas da matriz original por colunas onde vigora um valor médio para a renda do tranche. Dessa forma, a nova matriz, M^c , não conterá diferenças de renda resultante das circunstâncias. Por conseguinte, o indicador para a desigualdade de oportunidades é calculado a partir de: $I(M^a) - I(M^c)$. Ou seja, novamente equalizam-se os efeitos das circunstâncias.

A aplicação dessa estratégia aos dados italianos no período de 1993 a 2000, divididos em cinco tipos e vinte tranches, indicaram que a Região Sul do país apresenta o dobro da desigualdade de oportunidades da Região Centro-Norte. A comparação desses resultados com os obtidos por Bourguignon et al. (2007) reforçam a magnitude da desigualdade de oportunidades brasileira. Como já destacado, no Brasil a proporção da desigualdade de oportunidades na desigualdade total é de cerca de 23%. Na Itália essa proporção é de cerca de 10%.⁶

Recentemente, Pistolesi (2009) propôs um método alternativo para a captação da desigualdade de oportunidades. Sua abordagem incondicional é similar às estratégias de Bourguignon et al. (2007) e Checchi & Peragine (2010), avançando em, pelo menos, dois sentidos: a) ao captar os efeitos das variáveis explicativas em diferentes pontos da distribuição da variável dependente e; b) ao sugerir duas medidas de desigualdade de oportunidades, a primeira captando a desigualdade referente ao esforço (a partir da equalização das circunstâncias) e a segunda mensurando a desigualdade referente às circunstância (eliminando o efeito dos fatores de esforço).

Entretanto, apesar do seu procedimento apresentar vantagens em relação às abordagens anteriores, ele não está isento de críticas, principalmente no que se refere ao processo de estimação das distribuições contrafactuais.⁷ Em resumo, o autor utiliza um modelo semi-paramétrico onde se considera que os coeficientes associados às covariáveis se associam de forma discreta com a variável dependente. De acordo com Machado & Mata (2005), essa relação constitui uma restrição importante para o processo de inferência.

⁶Outros indicadores revelam o elevado nível de desigualdade de oportunidades brasileiro. Ver Barros et al. (2009) e Figueiredo & Ziegelmann (2010).

⁷É evidente que todas as abordagens apresentadas estão sujeitas a importantes críticas e vieses resultantes, principalmente da existência de endogeneidade, da má especificação das formas funcionais da equação de rendimento ou da adoção do RIA. Para uma leitura mais detalhada desses métodos, ver Ramos & Van de Gaer (2009).

Diante das estratégias empíricas progressas, em especial Bourguignon et al. (2007) e Pistolesi (2009), um avanço natural para o método de captação da desigualdade de oportunidades seria a utilização de uma ferramenta quantílica onde o problema da má especificação, resultante da omissão de variáveis relevantes e da endogeneidade das variáveis de circunstância, fosse abordado.

Ciente disso, este estudo propõe uma alternativa empírica para a mensuração da desigualdade de oportunidades. Para tanto, serão conjugados o método de inferência de distribuições contrafactuais proposto por Chernozhukov et al. (2009), com a estimação de regressões quantílicas sob má especificação e a consequente construção de bounds para as inferências pontuais contida em Angrist et al. (2006). Esse instrumental será aplicado ao conjunto de dados brasileiro, fornecendo mais uma evidência sobre o nível de desigualdade de oportunidades desse país.

Isso posto, além dessa introdução o artigo conta com quatro seções. Na seção 2 será apresentada a estratégia empírica subdividida em modelo teórico, estimação dos contrafactuais e banco de dados. A seção três é reservada aos resultados e a quarta às considerações finais.

2 Estratégia Empírica

2.1 Mensurando a Desigualdade de Oportunidades

O modelo teórico adotado neste estudo assume que os rendimentos do indivíduo i , Y_i , podem ser explicados por variáveis de circunstância (C_i) e de esforço (E_i). Seguindo a orientação de Bourguignon et al. (2007), postula-se uma relação de dependência entre os dois grupos de covariáveis, de tal forma que:

$$Y_i = f(C_i, E_i[C_i, v_i], u_i), \quad [2.1]$$

onde u_i e v_i são fatores aleatórios, inclusive a sorte.

Sob essa estrutura, as variáveis de circunstância podem afetar os rendimentos de forma “direta” e “indireta”. Para deixar essa relação mais clara, considere uma especificação log-linear:

$$\ln Y_i = \alpha C_i + \lambda E_i + u_i, \quad [2.2]$$

e, devido a endogeneidade das circunstâncias:

$$E_i = \Gamma C_i + v_i. \quad [2.3]$$

Onde α e λ são vetores de parâmetros, Γ representa uma matriz de coeficientes associados as variáveis de circunstância que influenciam o esforço. O sistema (2.2)-(2.3) informa que o efeito direto das circunstâncias é mensurado a partir do vetor α , enquanto que o indireto é captado por $\lambda\Gamma$.

Definindo $\Phi(Y)$ como a distribuição marginal dos rendimentos obtida a partir de (2.1), os autores sugerem a construção de dois contrafactuais: a) no primeiro, apagam-se os efeitos totais das circunstâncias, i.e., $\tilde{Y}_i = f(\bar{C}_i, E_i[\bar{C}_i, v_i], u_i)$, e; b) no segundo apagam-se apenas os efeitos diretos das circunstâncias, $\tilde{Y}_i^d = f(\bar{C}_i, E_i[C_i, v_i], u_i)$.

Note que no caso “a”, o vetor de renda resultante conterà apenas a desigualdade fruto das variáveis de esforço. Sendo assim, o share da desigualdade de oportunidades total será:

$$\Theta_I := \frac{I(\Phi(Y)) - I(\Phi(\tilde{Y}))}{I(\Phi(Y))}, \quad [2.4]$$

sendo I um índice de desigualdade como o Gini ou Theil, por exemplo. Já o share da desigualdade oriundo dos efeitos diretos, caso “b”, é representado por:

$$\Theta_I^d := \frac{I(\Phi(Y)) - I(\Phi(\tilde{Y}^d))}{I(\Phi(Y))}. \quad [2.5]$$

Dessa forma, o efeito indireto da desigualdade de oportunidades será $\Theta_I^i = \Theta_I - \Theta_I^d$.

O cálculo de (2.4) e (2.5) envolve as estimações do modelo (2.2), visando isolar os efeitos diretos da desigualdade e, da forma reduzida

$$\ln Y_i = C_i(\alpha + \Gamma\lambda) + (v_i\lambda + u_i), \quad [2.6]$$

representando os efeitos totais das circunstâncias.

É evidente que os problemas relacionados a variáveis omitidas e a endogeneidade das circunstâncias, tornam os parâmetros de (2.2) e (2.6) e, por conseguinte, a geração dos contrafactuais, viesados.

Cientes disso, e na ausência de variáveis instrumentais apropriadas, Bourguignon et al. (2007) promoveram uma simulação de Monte Carlo, obtendo intervalos para variação para os coeficientes estimados por OLS. O procedimento baseou-se no “monotone treatment response” fornecido por Manski & Pepper (2000).

Na subseção seguinte serão propostos métodos alternativos aos de Bourguignon et al. (2007). Em primeiro lugar, a geração dos contrafactuais se baseará em uma estrutura quantílica e não apenas nos valores médios da distribuição condicionada. Nesse sentido, as inferências alinham-se aos esforços

de Pistoletti (2009), porém, com um método mais robusto para a geração das distribuições simuladas. A questão da endogeneidade será tratada a partir da construção de bounds para a inferência de regressões quantílicas.

2.2 Método Contrafactual

A escolha do ferramental desenvolvido por Chernozhukov et al. (2009) é justificável por constituir um importante avanço em relação aos métodos regression-based pgressos. O principal deles é o desenvolvimento das teorias de distribuição e inferência para os estimadores. No mais, os estimadores podem se basear em uma classe ampla de métodos de regressão para as distribuições condicionais, tais como as regressões clássica e quantílica, os modelos de duração e as distribution regressions. Dessa forma, os autores incluem instrumentais anteriores, como os de Gosling et al. (2000), Donald et al. (2000) e Machado & Mata (2005), como casos especiais da sua teoria.⁸

Nessa abordagem, a população investigada é dividida em dois grupos indexados por $j \in \{0, 1\}$. O grupo de referência é representado por $j = 0$, enquanto que o grupo $j = 1$ sumariza os resultados contrafactuais. Para estabelecer a análise de regressão, é importante definir a distribuição condicional (u -quantile) de Y dado X no grupo j , $Q_{Y_j}(u|x)$, e a distribuição marginal para o p -vetor das covariadas X no grupo k , F_{X_k} , para $j, k \in \{0, 1\}$. Em seguida, define-se a variável de interesse Y_j^j como uma função das covariáveis e de um distúrbio não-aditivo U_j^j . Considerando a representação de Skorohod, essa relação será:

$$Y_j^j = Q_{Y_j}(U_j^j|X_j),$$

onde $U_j^j \sim U(0, 1)$ independente de $X_j \sim F_{X_j}$, para $j \in \{0, 1\}$.⁹

O experimento contrafactual adotado nesse estudo, consiste em gerar o resultado da conditional quantile function para o grupo $j = 1$, $Q_{Y_1}(u|x)$, mantendo as distribuições marginais das covariáveis do grupo $j = 0$, ou seja, $X_0 \sim F_{X_0}$. Sendo assim, define-se o resultado contrafactual Y_1^0 por:

$$Y_1^0 := Q_{Y_1}(U_1^0|X_0), \tag{2.7}$$

⁸É importante destacar que esta análise se concentra apenas nos métodos regression-based, em detrimento a literatura relacionada aos métodos propensity score (ver DiNardo et al. (1996) e Firpo et al. (2009), entre outros). Uma comparação entre os métodos de regressão e de propensity score pode ser encontrada no capítulo 3 de Angrist & Pischke (2008).

⁹Diferente dos modelos de regressão clássicos, a representação de Skorohod não necessita da hipótese da relação aditiva entre os erros e as covariáveis. Para detalhes, ver van der Vaart & Wellner (1996).

onde $U_1^0 \sim U(0, 1)$ independente de $X_0 \sim F_{X_0}$.

Como já destacado, (2.7) pode ser inferida a partir de uma série de métodos conhecidos na literatura, inclusive vários modelos de tratamento exógenos e endógenos (cf. Giustinelli, 2010). Porém, na ausência de instrumentos apropriados, se propõe que as distribuições condicionadas sejam inferidas via regressões quantílicas (QR) sob a hipótese da má especificação da Função Quantílica Condicionada (CQF). Para tanto, adota-se o referencial teórico proposto por Angrist et al. (2006), onde a má especificação da CQF pode ser entendida, inclusive, como resultante da omissão de uma variável relevante. Por fim, propõe-se a construção de bounds para as estimativas pontuais.

Sob o modelo de QR, a CQF pode ser obtida a partir da solução do problema de otimização:

$$Q_Y(u|X) \equiv \underset{q(X)}{\operatorname{argmin}} E[\rho_u(Y - q(X))].$$

Onde $\rho_u(w) = (u - \mathbb{I}(w \leq 0))w$ é uma check function, com \mathbb{I} representando uma função indicadora que assume valor 1 se o argumento é correto e 0 caso contrário. Caso $q(X)$ seja uma função linear bem especificada, o estimador de QR para vetor de parâmetros $\beta(u)$ resolvem o problema populacional

$$\beta(u) \equiv \underset{\beta \in R^d}{\operatorname{argmin}} E[\rho_u(Y - X'\beta)].$$

Segundo Powell (1986) esse vetor de parâmetros pode ser interpretado como um estimador GMM que satisfaz a condição de momento:

$$E[u - \mathbb{I}\{Y \leq X'\beta(u)\}|X] = 0. \quad [2.8]$$

Caso a CQF seja mal especificada devido a omissão de uma variável relevante, a condição de ortogonalidade em (2.8) não prevalecerá. Por exemplo, suponha o interesse em uma regressão com as seguintes covariáveis, $X = [X_1', X_2']'$, Porém, X_2 não encontra-se disponível, tal como a habilidade na equação de salários ou as variáveis associadas a fatores genéticos geralmente omitidas em estudos relacionados a desigualdade de oportunidades. Sendo assim, o problema de otimização quantílico será:

$$\gamma_1(u) \equiv \underset{\gamma_1}{\operatorname{argmin}} E[\rho_u(Y - X_1'\gamma_1)].$$

Contudo, a especificação correta para a CQF seria:

$$(\beta_1(u)', \beta_2(u)')' \equiv \underset{\beta_1, \beta_2}{\operatorname{argmin}} E[\rho_u(Y - X_1'\gamma_1 + X_2'\gamma_2)].$$

Diante disso, pode-se mensurar o erro de especificação associado a omissão da variável X_2 por:

$$R_u(X) := Q_Y(u|X) - X_1'\beta_1.$$

Note que se a CQF é linear, então $R_u(X) = X_2'\beta_2$. Sob algumas condições de regularidade, Angrist et al. (2006) demonstra que a relação entre $\beta_1(u)$ e $\gamma_1(u)$ é dada por: $\gamma_1(u) = \beta_1(u) + B_1(u)$, com

$$B_1(u) = E[w_u(X)X_1X_1']^{-1}E[w_u(X)X_1R_u(X)]. \quad [2.9]$$

Onde a função de ponderação (“importance weights”), é definida por:

$$w_u(X) = \frac{1}{2} \int_0^1 f_{\epsilon_u}(t\Delta_u(X, \gamma_1)|X) dt, \quad [2.10]$$

$\Delta_u := X'\beta - Q_Y(u|X)$ é o erro de especificação relacionado à estrutura linear da regressão, $\epsilon_u := Y - Q_Y(u|X)$ é o resíduo quantílico, definido como o desvio da variável dependente em relação ao quantil condicional de interesse e $f_{\epsilon_u}(\cdot)$ é a densidade condicionada.

A expressão (2.10) é calculada a partir da seguinte aproximação discreta:

$$\hat{w}_u(X) = \frac{1}{101} \sum_{t=1}^{101} \frac{1}{2} \hat{f}_Y \left(\frac{t-1}{101} x' \hat{\beta}(u) + \left(1 - \frac{t-1}{101}\right) \hat{Q}_Y(u|X=x)|X=x \right).$$

Com as densidades condicionais inferidas a partir de um kernel gaussiano.¹⁰ Em suma, esse procedimento permite calcular o viés dos estimadores de QR e, por conseguinte, obter estimativas não viciadas para os coeficientes do modelo. Nesse sentido, pode-se entender que os resultados estabelecidos nesse estudo constituem uma generalização não viciada das estimativas OLS contidas em Bourguignon et al. (2007).

2.3 Dados

Os dados utilizados neste estudo são similares aos de Bourguignon et al. (2007). Tratam-se de informações captadas na Pesquisa Nacional por Amostragem de Domicílios (PNAD) para o ano de 1996. Este ano foi escolhido por conta do suplemento de mobilidade social, onde são sumarizadas as informações relacionadas à educação, escolaridade e ocupação dos pais dos indivíduos selecionados. Consideraram-se somente os homens chefes de família,

¹⁰Detalhes sobre o processo de estimação podem ser encontrados no suplemento computacional de Angrist et al. (2006) disponível em: <http://econ-www.mit.edu/files/385>.

com idade entre 25 e 65 anos, residindo em zonas urbanas. Os rendimentos são representados pelo salário por hora¹¹, expresso em Reais de setembro de 2007.

O vetor de variáveis de circunstâncias inclui: a) uma *dummie* para raça, que assume valor 1 para os não-brancos (i.e. negros, pardos e indígenas) e 0 para os brancos (brancos e asiáticos); b) uma *dummie* para as regiões, que assume valor 1 se o indivíduo reside nas Regiões mais dinâmicas do país, Sul, Sudeste e Centro-Oeste e 0 caso contrário (Regiões Norte e Nordeste); c) educação do pai e da mãe, expressas em anos de estudo,¹² e; d) a posição ocupacional do pai, dividida nas seis categorias propostas por Pastore & Silva (1999).¹³

As variáveis de esforço são representadas por: a) os anos de estudo do indivíduo; b) uma *dummie* para migração¹⁴ e; c) o seu status no mercado de trabalho (trabalhador formal, sem carteira de trabalho assinada ou trabalhando por conta própria). Consideram-se ainda *dummies* para os coortes, onde, o primeiro coorte representa os indivíduos nascidos entre 1936 e 1940 (variável *dummie* C_1); o segundo os que nasceram entre 1941 e 1945 e assim por diante até o último coorte, contendo aqueles que nasceram entre 1966 e 1970 (variável *dummie* C_7).

A amostra conta com 24.293 indivíduos. Ao considerar o desenho complexo, ela é expandida para 19.565.783 indivíduos. Algumas estatísticas descritivas são apresentadas na Tabela 1, onde as informações estão divididas em variáveis de esforço e circunstâncias. Destaca-se o maior nível educacional médio dos filhos em relação aos pais. A comparação entre as escolaridades médias dos pais e das mães denuncia que elas não são significativamente diferentes. Nota-se também que há uma maior incidência de migrantes e de trabalhadores com carteira assinada.

Tabela 1 AQUI

Com relação ao restante das informações, observa-se que há uma predominância de indivíduos brancos e de residentes nas áreas mais dinâmicas do país. No que se refere à ocupação dos pais, cerca de 86% trabalham em setores intensivos em baixo estoque de capital humano.

¹¹Renda de todos os trabalhos dividida pelo número de horas trabalhadas.

¹²Seguindo a mesma classificação de Bourguignon et al. (2007).

¹³Categorias: 1) Baixo Inferior: Lavrador, pescador, lenhador, etc; 2) Baixo Superior: braçal, servente, vigias, etc.; 3) Médio Inferior: pedreiro, eletricista, marceneiro, etc.; 4) Médio Médio: praticistas, viajantes, chefes de nível baixo, etc.; 5) Médio Superior: assessores, gerentes de alto nível, diretores, etc. e; 6) Alto: grandes proprietários, magistrados, nível superior, etc.

¹⁴Se o indivíduo reside em um cidade diferente do seu local de nascimento.

3 Resultados

A captação do nível de desigualdade de oportunidades requer a comparação entre as distribuições para a renda observada e simulada. Para tanto, serão adotados dois procedimentos. No primeiro, calculam-se os efeitos totais das circunstâncias sobre os rendimentos. Para isso, os contrafactuais serão estimados a partir da especificação reduzida (2.6). Em seguida observam-se os efeitos diretos das circunstâncias, com os contrafactuais calculados a partir da especificação (2.2). Em ambos os casos, podem-se fixar todas as variáveis de circunstâncias ou uma a uma. As inferências serão baseadas no método de regressões quantílicas, com bounds construídos a partir do ferramental apresentado na subseção 2.2.

Os detalhes relativos aos procedimentos iniciais, em especial a especificações das funções de regressão e o cálculos das distribuições condicionadas, serão abordados na subseção 3.1. Em seguida, discutem-se os resultados relacionados à captação da desigualdade de oportunidades brasileira (subseção 3.2).

3.1 Análise Preliminar

As inferências contrafactuais considerarão os quantis (0.10, 0.20, 0.50, 0.80, 0.90). No mais, os bounds com 90% de confiança serão computados para cada coeficiente, usando o cálculo para os vieses do vetor de parâmetros especificados em (2.9). As Tabela 2 e 3 sumarizam as informações das regressões (2.2) e (2.6) considerando a mediana. Convém destacar que a função de “importance weights” é estimada a partir da aproximação discreta (2.10), tornando possível a mensuração dos vieses das estimativas de QR.

Tabela 2 AQUI

A Tabela 2 apresenta os coeficientes de QR, os erros padrões robustos e os bounds (inferior e superior) para a estimativas pontuais. Grosso modo, os resultados indicam: a) um retorno da educação próximo a 11%, resultado similar aos disponíveis na literatura brasileira (cf. Sachsida et al., 2004); b) a educação dos pais como um fator relevante para determinação dos salários; c) os não brancos apresentam um nível de renda mais baixo em relação aos brancos (brancos e asiáticos); d) as pessoas que residem nas regiões mais dinâmicas (Sul, Sudeste e Centro-oeste) têm um nível de renda mais alto do que as que residem nas regiões Norte e Nordeste; e) a intensividade em capital humano do setor de ocupação do pai possui uma relação direta com o nível de renda do filho e; f) as variáveis de coorte não se mostraram significativas. As estimativas para a regressão reduzida (2.6), estão dispostas na Tabela

3. Note que nessa estrutura não há variáveis de esforço, portanto, o modelo perde em poder explicativo.

Tabela 3 AQUI

A estimação das funções condicionadas constitui o primeiro passo da análise. Em seguida, inferem-se as distribuições contrafactuais a partir da fixação de um grupo de referência. No caso das variáveis não dicotômicas, como educação dos pais e das mães, foi criada uma variável dummie que indicará o grupo de referência. Para essas variáveis, o grupo de referência será o dos pais com ensino elementar completo. Já a referência para a ocupação do pai será a categoria Médio Inferior. Os demais grupos serão os migrantes, os brancos e a região mais dinâmica. Nessa etapa será possível isolar o efeito de cada variável de circunstância sobre o indicador de desigualdade. Por conta disso, serão calculados seis índices de desigualdade, considerando as equalizações: Total (todas as circunstâncias fixadas); por Raça; por Região; pela Educação do Pai; pela Educação da Mãe e pela Ocupação do Pai.

Um exemplo da geração de um contrafactual é apresentado na Tabela 4. Nela têm-se os resultados para as distribuições do logaritmo da renda observada (“Obs.”) e simuladas, com efeitos total (“Tot.”) e direto (“Dir.”), em quantis selecionados, considerando todas as circunstâncias apagadas. Assim como esperado, as simulações diminuíram a dispersão na distribuição da renda. Isto é, houve uma redução nos desvios padrões das distribuições (“SE”), indicando que as variáveis de circunstância possuem um efeito expressivo sobre a distribuição dos rendimentos brasileiros.

Tabela 4 AQUI

A observação dos quantis selecionados indica que a equalização das circunstâncias promove um impacto mais expressivo na parte de baixo da distribuição, i.e., considerando os quantis 0,10 e 0,20, o valor relativo às distribuições simuladas é sempre superior ao da distribuição observada. O efeito inverso é observado na parte de cima da distribuição.

3.2 Índices de Desigualdade de Oportunidades

Como já destacado, a simulação contrafactual tornou a distribuição dos rendimentos mais compacta. Resta saber a magnitude desse movimento. Diante disso, propõe-se o cálculo dos indicadores de desigualdade brasileiros observando os efeitos “Total” e “Direto” das variáveis de circunstâncias.¹⁵

¹⁵Os resultados dessa seção baseiam-se na aplicação do exponencial sobre os contrafactuais, dado que na especificação da regressão, utilizou-se o log da renda.

A Tabela 5 sintetiza a desigualdade de oportunidades, mensurada pelo índice Theil.¹⁶ No painel A têm-se as informações relacionadas ao efeito Total das variáveis de circunstâncias, destacando-se as estimativas pontuais para as desigualdades com seus respectivos bounds e a participação desses indicadores na desigualdade de renda (Eq. 2.4). Já no painel B têm-se as informações relativas ao efeito Direto das variáveis de circunstâncias (Eq. 2.5).

Tabela 5 AQUI

Convém ressaltar que a desigualdade do vetor de renda observado é de 0,5639. Diante disso, a primeira coluna de resultados da Tabela 5 informa que após a equalização de todos os fatores de circunstância, o índice de desigualdade recua para 0,3814. Ou seja, o peso Total das variáveis de circunstância sobre a desigualdade de renda é de aproximadamente 32% $[(0,5639 - 0,3814)/0,5639]$. Considerando apenas os efeitos Diretos dos fatores de não responsabilidade, esse peso cai para 16%. Esses percentuais indicam que há um equilíbrio entre a contribuição dos fatores Diretos e Indiretos sobre a desigualdade de oportunidades. Em outras palavras, variáveis como educação e ocupação dos pais, raça e região de nascimento, produzem efeitos expressivos não só sobre a renda do indivíduo, como também sobre o seu nível de esforço.

Ainda na primeira coluna, observam-se as estimativas para os bounds inferior e superior dos índices. Os resultados indicam que o peso Total das circunstâncias sobre a desigualdade de renda pode variar entre 29% e 35%. Já o efeito Direto se situa entre 12% e 19% e, conseqüentemente, o share do efeito Indireto se limita ao intervalo de 16% e 17%.

Nos demais resultados, cada fator de circunstância será fixado um a um. Esse exercício indica que a educação do pai (Ed.P) é o fator que mais contribui para a desigualdade de oportunidades isoladamente. Isso reforça as evidências da baixa mobilidade intergeracional da educação brasileira (Ferreira & Veloso, 2003). Logo em seguida, vem mais duas variáveis relacionadas ao background familiar: ocupação do pai (Ocp.P), o que se coaduna com as evidências da baixa mobilidade social brasileira (Pastore & Silva, 1999); e a educação da mãe (Ed.M). As equalizações por Raça e Região Geográfica são as de menor impacto dentro do conjunto de variáveis de circunstância disponível.

Em linhas gerais, esses resultados mostram algumas diferenças em relação aos obtidos por Bourguignon et al. (2007). A mais expressiva reside na men-

¹⁶Convém ressaltar que há uma discussão relacionada à aplicação de um índice de desigualdade sobre uma variável obtida a partir de um processo contrafactual. Para detalhes, ver Devooght (2008).

suração do impacto da desigualdade de oportunidades sobre a desigualdade total. De acordo com esses autores, o peso da desigualdade de oportunidades brasileira (Eq. 2.4), é de cerca de 23%, valor bem inferior ao share para a estimativa pontual da Tabela 5 e, até mesmo, ao seu limite inferior (29%).

Dado que o banco de dados utilizados por esses estudos é similar, acredita-se que a discrepância nos resultados se deve às diferenças metodológicas. Enquanto o estudo de Bourguignon et al. (2007) se baseou em contrafactuais gerados a partir do método de Mínimos Quadrados Ordinários, observando os valores condicionados médios, os resultados da Tabela 5 consideraram o os efeitos das variáveis explicativas em diferentes pontos da distribuição da variável dependente.

Essa impressão é reforçada pelo estudo de Figueiredo (2011). Usando o modelo teórico proposto por Pistolesi (2009), o autor calcula dois indicadores de desigualdade de oportunidades: i) o primeiro captando a desigualdade referente ao esforço (a partir da equalização das circunstâncias) e; ii) o segundo mensurando a desigualdade referente às circunstância (eliminando o efeito dos fatores de esforço). Seus resultados, também baseados em contrafactuais quantílicos, indicam que as participações da desigualdade de oportunidades na desigualdade de renda são de 35% e 34%, respectivamente.

No mais, empregou-se a estimação de regressões quantílicas sob má especificação da CQF, conforme proposto por Angrist et al. (2006). Admitir a omissão de variáveis relevantes permite uma estimativa robusta para Θ_I . Como a estratégia adotada por Bourguignon et al. (2007) não considerou esse tipo de problema, não houve plena equalização das variáveis de circunstância em (2.4), de modo a gerar um contrafactual possivelmente viesado. Ou seja, os autores superestimaram $I(\Phi(\tilde{Y}))$, o que gerou, via (2.4), uma baixa estimativa (23%) para a desigualdade de oportunidades brasileira em comparação aos resultados aqui apresentados (entre 29% e 35%).

Em resumo, a utilização de um método mais robusto para a geração das distribuições contrafactuais e a consideração explícita da má especificação da CQF demonstrou que o nível de desigualdade de oportunidades brasileiro é ainda maior do que o já computado na literatura. Observou-se ainda uma discrepância no que se refere aos efeitos diretos e indiretos das variáveis de circunstância sobre a desigualdade total. Enquanto Bourguignon et al. (2007) estima que 40% da influência das variáveis de circunstância sobre a desigualdade se dá de forma indireta, as estimativas aqui apresentadas indicam algo em torno de 50%. A explicação para tal fato por ser visualizada a partir de (2.5). No procedimento de “apagar” os efeitos diretos das circunstâncias visando obter $I(\Phi(\tilde{Y}^d))$, os autores não excluíram apenas o efeito direto, mas também o efeito indireto de variáveis omitidas, uma vez que não consideraram

problemas de má especificação na estimativa. Isso posto, “superequalizaram” os supostos efeitos diretos em (2.5), de forma a superestimar Θ_7^d e subestimar Θ_7^i .

Isso deixa claro como problemas relacionados às variáveis omitidas tornam os parâmetros de (2.2), (2.6) e, conseqüentemente, os contrafactuais, viesados. Como a estimativa da desigualdade de oportunidades se baseia nesses contrafactuais, a ocorrência desse problema, como em Bourguignon et al. (2007), remete a uma estimativa viciada para a desigualdade de oportunidades.

4 Considerações Finais

O objetivo desse artigo foi mensurar a desigualdade de oportunidades brasileira. Para tanto, adotou-se o modelo teórico desenvolvido por Bourguignon et al. (2007), onde as variáveis de circunstância possuem efeitos diretos e indiretos sobre os resultados econômicos dos indivíduos. O banco de dados foi fornecido pelo suplemento social da PNAD de 1996.

Diante de possíveis vieses decorrentes da má especificação das equações de rendimento e da necessidade de se considerar os efeitos das variáveis explicativas em diferentes pontos da distribuição da variável dependente, optou-se por construir as distribuições contrafactuais a partir de um método quantílico. A alternativa empírica envolveu a inferência de distribuições contrafactuais proposta por Chernozhukov et al. (2009) e a estimação de regressões quantílicas sob má especificação da CQF, desenvolvido por Angrist et al. (2006).

Os resultados indicaram um equilíbrio entre os impactos direto e indireto das circunstâncias sobre os rendimentos. A utilização dessa estratégia robusta apontou para um nível de desigualdade de oportunidades maior do que o já computado na literatura. Em resumo, estimou-se que o impacto da desigualdade de oportunidades na desigualdade total brasileira pode variar entre 29% e 35%.

5 Referências

Angrist, J., Chernozhukov, V. & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica*, 74(2), 539-563.

Angrist, J. & Pischke, J.-S. (2008). *Mostly harmless econometrics: an empiricist’s companion*. Princeton: Princeton University Press.

- Arneson, R. (1989). Equality and equal opportunity for welfare. *Philosophical Studies*, 56, 77-93.
- Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3), 244-263.
- Azevedo, J. (2007). Avaliando a significância estatística da queda na desigualdade no Brasil. In: Barros, R., Foguel, M. & Ullysea, G. (Orgs). *Desigualdade de renda no Brasil: uma análise da queda recente*. Brasília: IPEA.
- Barros, R. & Mendonça, R. (1995). A evolução do bem-estar, pobreza e desigualdade no Brasil ao longo das últimas décadas - 1960/90. *Pesquisa e Planejamento Econômico*, 25(1), 115-164.
- Barros, R., Ferreira, F., Vega, J., Chanduvi, J., Carvalho, M., Franco, S., Freije-Rodríguez, S. & Gignoux, J. (2009). *Measuring inequality of opportunities in Latin America and the Caribbean*. Washington, World Bank.
- Bourguignon, F., Ferreira, F. & Menéndez, M. (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth*, 53, 585-618.
- Checchi, D. & Peragine, V. (2010). Inequality of opportunity in Italy. *Journal of Economic Inequality*, 8(4), 429-450.
- Chernozhukov, V., Fernández-Val, I. & Melly, B. (2009). *Inference on Counterfactual Distributions*. CEMMAP Working Papers, CWP09/09.
- Devooght, K. (2008). To each the same and to each his own: a proposal to measure responsibility-sensitive income inequality. *Economica* 75, 280-295.
- Dworkin, R. (1981). What is Equality? Part 2: Equality of Resources. *Philosophy and Public Affairs*, 10, 185-246.
- DiNardo, J., Fortin, N., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. *Econometrica* 64, 1001-1044.
- Donald, S., Green, A., & Paarsch, H. (2000). Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates, *Review of Economic Studies* 67, 609-633.
- Ferreira, S. & Veloso, F. (2003). Mobilidade intergeracional de educação no Brasil. *Pesquisa e Planejamento Econômico*, 33, 481-583.
- Figueiredo, E. & Ziegelmann, F. (2009). Mudanças na distribuição de renda brasileira: significância estatística e bem-estar social. *Revista de Economia Aplicada*, 13, 257-277.

- Figueiredo, E. & Ziegelmann, F. (2010). Estimation of opportunity inequality in Brazil using nonparametric local logistic regression. *Journal of Development Studies* 46(9), 1593-1606.
- Figueiredo, E. (2011). *Mensurando a desigualdade de oportunidades brasileira*. Texto para discussão, PPGE/UFPB.
- Firpo, S., N. Fortin & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3), 953-973.
- Giustinelli, P. (2010). Non-parametric bounds on quantiles under monotonicity assumptions: with an application to the Italian education returns. *Journal of Applied Econometrics*.
- Gosling, A., S. Machin, & Meghir, C. (2000). The changing distribution of male wages in the U.K.. *Review of Economic Studies*, 67, 635-666.
- Kolm, S-C. (1972). *Justice et équité*. Editions du Centre National de la Recherche Scientifique, Paris.
- Kranich, L. (1996). Equitable opportunities: an axiomatic approach. *Journal of Economic Theory*, 71, 132-147.
- Lefranc, A., Pistolesi, N. & Trannoy, A. (2004). *Harder times for heirs: social background and income in France, 1979-2000*. Mimeo THEMA.
- Machado, J. & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20, 445-465.
- Manski, C. & Pepper, J. (2000). Monotone Instrumental Variables, with an Application to the Returns to Schooling. *Econometrica*, 68(4), 997-1012.
- Neri, M. (2006). *Desigualdade, estabilidade e bem-estar social*. EPGE-FGV: Ensaio Econômico, n. 637.
- Pastore, J. & Silva, N. (1999). *Mobilidade social no Brasil*. São Paulo: Makron Books.
- Pattanaik, P. & Xu, Y. (1990). *On ranking opportunity sets in terms of freedom of choice*. Recherches Économiques de Louvain, 56, 383-390.
- Pistolesi, N. (2009). Inequality of opportunity in the land of opportunities. *Journal of Economic Inequality*, 7, 411-433.
- Powell, J. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica*, 54(6), 1435-1460.
- Ramos, C. & Santana, R. (2003). *Quão pobres são os pobres?* Brasil: 1991-2001. UNB: Texto para Discussão.

- Ramos, X. & Van de gaer, D. (2009). *Empirical evidence on inequality of opportunity*. Working Paper.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.
- Roemer, J. (1998). *Equality of Opportunity*. New York: Harvard University Press.
- Roemer, J. (1996). *Theories of distributive justice*. Harvard University Press, Cambridge.
- Sachsida, A., Loureiro, P. & Mendonça, M. (2004). Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia*, 58(2), 249-265.
- Sen, A. (1985). *Commodities and Capabilities*. Amsterdam, Elsevier.
- Thomson, W. (2011). *Fair allocation rules*. In: Arrow, K., Sen, A. & Suzumura, K. (eds). *Handbooks in Economics: Social choice and welfare*, v. II. Amsterdam: Elsevier.
- van der Vaart, A. & Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.

Tabela 1: Estatísticas Descritivas

Variáveis	
Média dos Rendimentos	1.655,36
Variáveis de Esforço	
Média da Escolaridade	6,97
Migrantes (%)	64%
Status no Mercado de Trabalho (%)	
Formal	57%
Informal	14%
Conta-própria	29%
Variáveis de Circunstâncias	
Média Esc. do Pai	3,06
Média Esc. da Mãe	2,99
Raça (%)	
Branco	61%
Não-branco	39%
Regiões (%)	
Norte & Nordeste	31%
Sul, Sudeste e Centro-oeste	69%
Ocupação dos Pais (%)	
Baixo Inferior	26%
Baixo Superior	34%
Médio Inferior	26%
Médio Médio	8%
Médio Superior	4%
Alto	2%

Tabela 2: Equação de Redimentos – 1996

Quantil 0.50				
	Coef.	Std. Dev.	L. Bound	U. Bound
Constante	2.2636*	0.2342	1.9656	2.8954
Educação	0.1123*	0.0016	0.1072	0.1243
Migração	0.1011*	0.0118	0.0357	0.1654
Idade1	0.4773*	0.0129	0.3456	0.6521
Idade2	-0.0003**	0.0001	-0.0005	-0.0001
Educ. Pai	0.0242*	0.0023	0.0125	0.0298
Educ. Mãe	0.0137*	0.0021	0.0106	0.0179
Raça	-0.1884*	0.0127	-0.2396	-0.1002
Regiões	0.2140*	0.0132	0.1064	0.3459
Mercado de Trabalho (Ref.: Trabalhador Formal)				
Trabalhador Informal	-0.0629*	0.0130	-0.0749	-0.0334
Conta Própria	-0.3099*	0.0173	-0.3566	-0.2344
Ocupação do Pai (Ref.: Baixo Inferior)				
Baixo Superior	-0.0234	0.0160	-0.0329	-0.0200
Médio Inferior	0.0667*	0.0173	0.0541	0.0756
Médio Médio	0.1404*	0.0262	0.1023	0.2306
Médio Superior	0.1267*	0.0269	0.0999	0.2187
Alto	0.2316*	0.0366	0.1458	0.3218
Coorte (Ref.: Coorte C_7)				
C_1	-0.2470**	0.1176	-0.3183	-0.1834
C_2	-0.1231	0.0824	-0.1954	-0.0542
C_3	-0.0181	0.0561	-0.0587	-0.0091
C_4	0.0129	0.0378	0.0072	0.0136
C_5	0.0053	0.0254	0.0008	0.0093
C_6	0.0023	0.0336	0.0010	0.0086
R^2			0.2886	
Amostra			24,278	

Nota: * $p < 0.01$ and ** $p < 0.05$.

Tabela 3: Equação de Rendimentos – Forma Reduzida

Quantil 0.50				
	Coef.	Std. Dev.	L. Bound	U. Bound
Constante	3.5783*	0.0283	3.0958	3.9831
Educ. Pai	0.0745*	0.0028	0.0439	0.0970
Educ. Mãe	0.0400*	0.0026	0.0096	0.1075
Raça	-0.3175*	0.0156	-0.4076	-0.0934
Regiões	0.2905*	0.0163	0.1548	0.4490
Ocupação do Pai (Ref.: Baixo Inferior)				
Baixo Superior	0.0577*	0.0198	0.0198	0.1290
Médio Inferior	0.3256*	0.0208	0.2556	0.4675
Médio Médio	0.4307*	0.0320	0.1743	0.6879
Médio Superior	0.3994*	0.0330	0.1056	0.5277
Alto	0.4936*	0.0330	0.3209	0.6498
Coorte (Ref.: Coorte C_7)				
C_1	0.1276*	0.0333	0.0568	0.2887
C_2	0.2293*	0.0295	0.1087	0.5482
C_3	0.3522*	0.0260	0.1560	0.6431
C_4	0.3699*	0.0241	0.2400	0.5777
C_5	0.2932*	0.0235	0.0987	0.4309
C_6	0.1562*	0.0233	0.0845	0.2964
R^2		0.1786		
Amostra		24,278		

Nota: * $p < 0.01$ and ** $p < 0.05$.

Tabela 4: Distribuições Observada e Simulada para o Log da Renda

Quantis	Distribuições		
	Obs.	Tot.	Dir.
.10	4.00	5.02	4.34
.20	4.39	5.09	4.75
.50	5.22	5.56	5.02
.80	5.98	5.81	6.01
.90	6.40	6.03	6.36
SE	0.95	0.81	0.90

Tabela 5: Índice e Contribuição da Desigualdade de Oportunidades Brasileira

	Total	Raça	Regiões	Ed.P	Ed.M	Ocp.P
Painel A: Efeito total das Circunstâncias						
(Upper bound)	(0,4032)	(0,4487)	(0,4321)	(0,4098)	(0,4254)	(0,4243)
Média	0,3814	0,4265	0,4125	0,3984	0,4102	0,4076
(Lower bound)	(0,3654)	(0,4097)	(0,3901)	(0,3805)	(0,3989)	(0,3891)
% das circunstâncias na desigualdade de renda						
(Upper bound)	(0,3520)	(0,2734)	(0,3082)	(0,3252)	(0,2926)	(0,3099)
Share da Média	0,3236	0,2436	0,2684	0,2934	0,2725	0,2771
(Lower bound)	(0,2849)	(0,2042)	(0,2337)	(0,2732)	(0,2456)	(0,2475)
Painel B: Efeito direto das Circunstâncias						
(Upper bound)	(0,4985)	(0,5123)	(0,5102)	(0,5013)	(0,5129)	(0,5087)
Média	0,4710	0,4967	0,4903	0,4845	0,4909	0,4856
(Lower bound)	(0,4565)	(0,4787)	(0,4705)	(0,4634)	(0,4723)	(0,4701)
% das circunstâncias na desigualdade de renda						
(Upper bound)	(0,1904)	(0,1510)	(0,1656)	(0,1782)	(0,1624)	(0,1663)
Share da Média	0,1647	0,1191	0,1305	0,1408	0,1294	0,1388
(Lower bound)	(0,1159)	(0,0915)	(0,0952)	(0,1110)	(0,0904)	(0,0978)