

Casamentos Seletivos e Desigualdade de Renda no Brasil

LUCIENE PEREIRA*

CEZAR SANTOS†

Sumário: 1. Introdução; 2. Crescimento dos casamentos seletivos; 3. Casamentos seletivos e desigualdade de renda; 4. Conclusão; A. Dados; B. Curva de Lorenz e Índice de Gini; C. Algoritmo Sinkhorn-Knopp (1967).

Palavras-chave: Desigualdade de renda; Casamentos; Pareamento.

Códigos JEL: D13, D31, J12, O11.

O objetivo deste artigo é investigar a evolução do mercado de casamento e o seu impacto na distribuição de renda no Brasil. A partir da análise de dados dos censos demográficos de 1970 a 2010, encontramos evidências de que as pessoas estão cada vez mais se casando com parceiros de características semelhantes, ao longo dos anos. O aumento do número de casamentos seletivos, a princípio, não é capaz de afetar negativamente a desigualdade de renda. No entanto, quando realizamos alguns exercícios contrafactuais, é possível perceber que a melhora na distribuição de renda, ocorrida nesse período, poderia ter sido ainda maior caso essa tendência no mercado de casamento não tivesse ocorrido.

This paper aims to study the evolution of marriage markets and its impact on income inequality in Brazil. Using data from Brazilian censuses from 1970 to 2010, we find evidence that the degree of assortative mating has increased in Brazil. This higher degree of sorting does not seem at first to influence income inequality, since this has decreased over time. However, by computing counterfactual analyses, we find that income inequality would have decreased even further were it not for the higher degree of assortative mating in the marriage market.

1. INTRODUÇÃO

Decisões individuais (tais como casamento, divórcio, fecundidade e educação) afetam diretamente a dinâmica demográfica de um país e estão relacionadas a variáveis macroeconômicas (como renda *per capita* e desigualdade social). O interesse em entender como mudanças demográficas e desigualdade de renda se relacionam tem sido crescente na literatura. Em especial para o Brasil, Barros et alii (2000) apresenta evidências do impacto de fatores demográficos (o tamanho e a composição etária da população) sobre a distribuição de renda entre os anos de 1976 e 1996. Ele chega à conclusão, a partir de microssimulações com dados da PNAD, que o efeito da mudança demográfica nesse período sobre a redução da pobreza corresponde a aproximadamente 15% do efeito que o crescimento da renda teve sobre tal redução.¹ Estudar a relação entre mudanças demográficas e distribuição de renda se mostra particularmente interessante para o Brasil, pois o país possui elevados índices de desigualdade. Segundo dados do Banco Mundial, cerca de 90% das economias apresentam um grau de desigualdade de renda inferior à brasileira. Além disso, o país apresenta um fato notório, o seu grau de desigualdade

*Bahia Asset Management.E-mail: lucienetmp@gmail.com

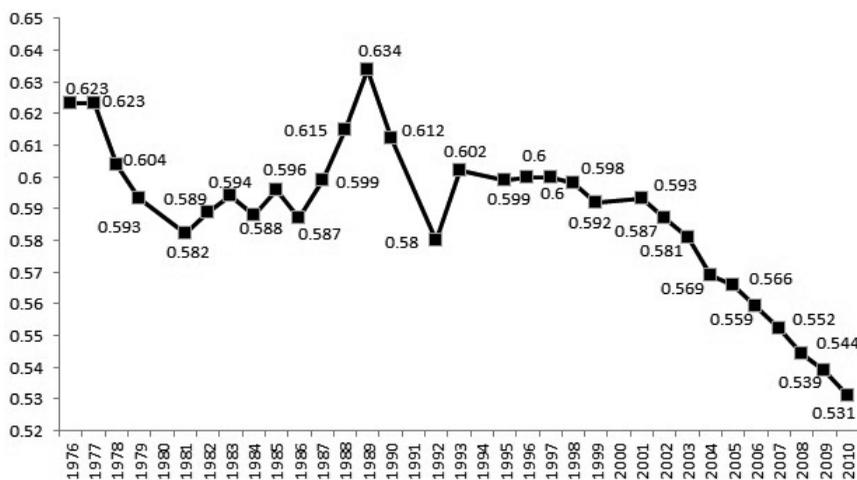
†FGV/EPGE - Escola Brasileira de Economia e Finanças.E-mail: cezar.santos@fgv.br

¹Entre os anos de 1976 a 1996, a renda per capita cresceu em média 3% ao ano.



vem apresentando uma queda contínua e acentuada desde 2001, como podemos observar na Figura 1 abaixo.²

Figura 1: Evolução do Índice de Gini, de 1976 a 2010



Nota: Dados do Banco Mundial.

Neste artigo, estudamos a evolução do mercado de casamento no Brasil nos últimos quarenta anos e como isso tem afetado a distribuição de renda da economia. A partir da análise de dados dos censos demográficos de 1970 a 2010, verificamos que as pessoas estão cada vez mais propensas a se casar com outras que possuem características socioeconômicas semelhantes (aqui representadas por 4 níveis educacionais distintos). Esse fenômeno social será chamado neste artigo de crescimento do casamento seletivo (ou também de queda na diversidade conjugal).

O crescente fenômeno de casamentos entre classes socioeconômicas semelhantes já vem sendo largamente documentado em países desenvolvidos por sociólogos e economistas.³ Para o Brasil, Neri (2005) parece ser o único a se debruçar sobre essa questão. O estudo analisa de forma geral as relações entre casamento e economia, distinguindo os padrões de mulheres e homens. E, mais especificamente, mensura o quanto das relações conjugais acontecem entre pessoas do mesmo nível educacional e como isso evoluiu entre 1970 e 2000. E a partir da análise do percentual de casamentos ocorridos entre os mesmos grupos educacionais (56,7% em 1970 e 49,6% em 2000), conclui que houve um aumento na diversidade conjugal (ou queda no casamento seletivo). O nosso artigo apresenta conclusões opostas a Neri (2005) e reafirma, para o Brasil, a tendência apontada na literatura internacional. Isso ocorre pois Neri (2005) não leva em consideração o fato de que as distribuições marginais de homens e mulheres entre os diferentes níveis educacionais mudaram ao longo dos anos. Quando esse fato é incorporado às análises, somos capazes de concluir que houve um aumento nos casamentos seletivos no Brasil.

Além de estudar essa tendência, também analisamos como ela é capaz de afetar a distribuição de renda do país. A literatura mostra que, para países desenvolvidos, o crescente número de casamentos seletivos tem contribuído para o aumento da desigualdade de renda entre as famílias. No entanto, para o Brasil verificamos em um primeiro momento o oposto, isto é, a desigualdade de renda melhorou entre 1970 e 2010, apesar de os brasileiros estarem cada vez mais se casando com parceiros de características semelhantes. Contudo, a partir de exercícios contrafactuais, mostramos que a distribuição de renda

²Barros et alii (2006) analisa as causas e consequências dessa queda entre 2001 e 2004.

³Ver Schwartz e Mare (2005), Blossfeld e Timm (2003), Greenwood et alii (2014), Greenwood et alii (2016), Eika et alii (2014).

poderia ter melhorado ainda mais caso essa tendência no mercado de casamentos não estivesse presente. No primeiro experimento, medimos o efeito sobre o Índice de Gini caso os casamentos tivessem ocorrido de forma aleatória e observamos que, tanto para 1970 quanto para 2010, há uma melhora nesse índice. No segundo, trocamos o percentual de solteiros entre os anos de 1970 e 2010 e mais uma vez verificamos uma melhora na desigualdade de renda, embora o efeito se mostre pequeno. Esses exercícios contrafactuais só fazem sentido e produzem algum efeito na distribuição de renda caso a participação de mulheres no mercado de trabalho seja positiva e crescente, o que se mostra verdadeiro quando analisamos os dados. A partir disso, realizamos o exercício de assumir que todos os casamentos ocorreram de forma aleatória em 1970 e 2010 mas trocando a participação das mulheres casadas no mercado de trabalho entre os anos de 1970 e 2010. E comprovamos que a inserção de mulheres casadas afeta a distribuição de renda.

Investigar o impacto de um maior número de casamentos seletivos sobre a distribuição de renda requer um certo cuidado. Durante os últimos quarenta anos, observamos, pela Figura 3, que o nível educacional da população brasileira como um todo tem aumentado e que as mulheres a partir de 1990 passaram a ter em média mais anos de estudo que os homens. Portanto a distribuição marginal de homens e mulheres entre os diferentes níveis educacionais mudou nesse período. Assim, precisamos distinguir os efeitos sobre a distribuição de renda provocados pelas mudanças dos níveis de educação de homens e mulheres e pela maior tendência de casamentos seletivos. Para isso, no exercício contrafactual de trocar a distribuição de casamentos entre os anos de 1970 e 2010, utilizamos a tabela de contingência padronizada.⁴ E mais uma vez observamos uma melhora na desigualdade de renda.

Deve-se notar também que há interações entre algumas das margens acima que não são exploradas no presente artigo. Por exemplo, há vários estudos que relacionam interações entre casamento e participação feminina no mercado de trabalho: ver, por exemplo, o artigo clássico de Lundberg (1985). Goldin (2006) relaciona mudanças na família, educação e oferta de trabalho feminina por um período de mais de 100 anos. Há ainda estudos sobre a relação entre os retornos da educação no mercado de casamento: ver, por exemplo, Chiappori et alii (2009) e Greenwood et alii (2016).

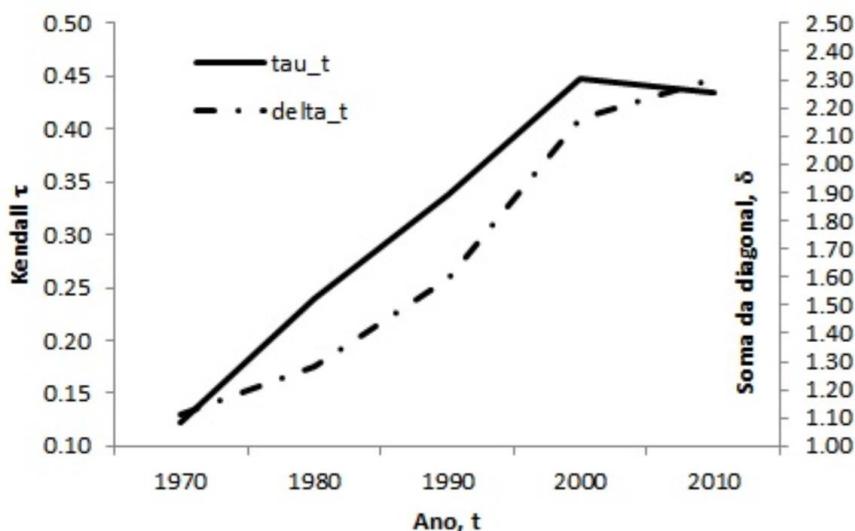
O artigo está organizado da seguinte maneira. Na próxima seção, analisamos a evolução da diversidade conjugal no mercado de casamentos brasileiro. A partir disso, verificamos como essa tendência pode afetar a distribuição de renda e realizamos alguns experimentos contrafactuais na seção 3. Por fim, concluímos na seção 4.

2. CRESCIMENTO DOS CASAMENTOS SELETIVOS

Para analisar o crescente fenômeno de casamentos entre pessoas de características semelhantes, dividimos as classes socioeconômicas brasileiras em quatro níveis educacionais: educação primária incompleta (PC-), primária (PC), secundária (SC) e universidade (C). Primeiramente, estimamos o coeficiente de correlação de Kendall τ entre os níveis educacionais do marido e da esposa para cada ano da nossa amostra (1970, 1980, 1991, 2000 e 2010). A evolução desse coeficiente ao longo do tempo pode ser observada na Figura 2 a seguir. Note que, apesar da série não ser monotonicamente crescente, o coeficiente τ é claramente maior em 2010 (0.43516) do que em 1970 (0.12160).

Em segundo, computamos a tabela de contingência dos casamentos (os anos de 1970 e 2010 podem ser vistos na Tabela 1). Cada célula representa o percentual de casamentos com aquele específico pareamento educacional entre marido e esposa. Vale mencionar que a moda entre todas as combinações educacionais de marido e esposa em 1970 era entre pessoas de baixa escolaridade (ensino primário incompleto) e passou para casais com ensino secundário em 2010, ou seja, houve uma melhora dos níveis

⁴Método proposto por Mosteller (1968) e descrito com maiores detalhes na seção 3.

Figura 2: Crescimento de casamentos seletivos

Nota: As variáveis τ_t and δ_t são medidas da tendência de casamentos seletivos para os anos $t = 1970, \dots, 2010$. Quanto maior o valor de uma dessas variáveis, maior é o grau de pareamento seletivo entre os casais. Veja o texto para maiores detalhes e o Apêndice para a descrição dos dados usados.

educacionais na população brasileira (que pode ser facilmente verificada ao observarmos a evolução do número médio de anos de estudo para mulheres e homens na Figura 3 abaixo).⁵

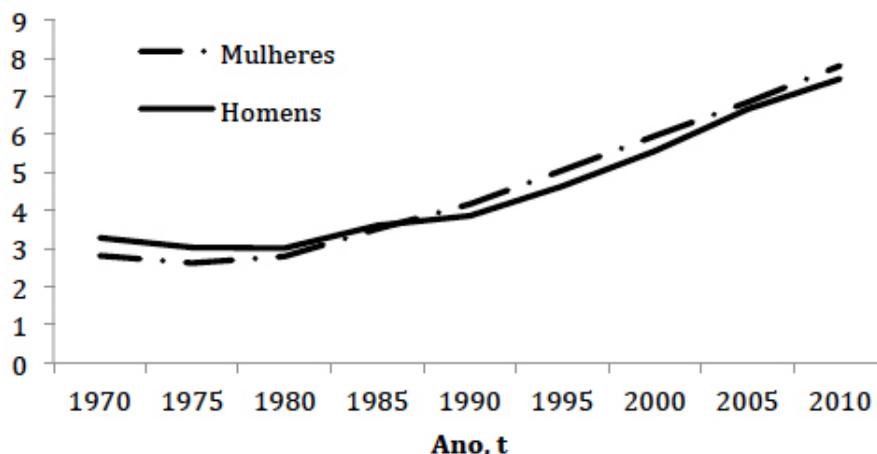
Computamos também as tabelas de contingência quando a hipótese de casamento aleatório é assumida (em itálico na Tabela 1). A partir dessas duas categorias de tabela de contingência, podemos apresentar outra estatística que indica a queda na diversidade conjugal ao longo dos anos. Essa estatística é a razão entre a soma dos traços (diagonal principal) das tabelas de contingência real e aleatória. Note que os elementos da diagonal principal representam os casamentos entre pessoas de mesmo nível educacional, assim quanto mais alto forem esses números, maior é a fração na população de casamentos entre indivíduos com características semelhantes. Como δ é a razão da soma da diagonal principal das tabelas de contingência real e aleatória, quanto mais alto for essa razão, maior é o grau de casamento seletivo. Observando ainda a Figura 2, podemos notar que essa estatística é sempre maior do que um (o que indica que o pareamento entre marido e mulher de mesmo nível educacional é sempre superior ao caso aleatório) e que ela cresceu mais do que 100% entre 1970 e 2010. Portanto, o grau de casamento seletivo cresceu de forma considerável na sociedade brasileira.

3. CASAMENTOS SELETIVOS E DESIGUALDADE DE RENDA

A literatura tem apontado que um crescimento na tendência de casamentos seletivos vem se traduzindo em uma maior desigualdade de renda para países desenvolvidos (em especial os Estados Unidos). O interesse nessa relação causal pode ser observado tanto em termos teóricos quanto em termos pu-

⁵Repare que as mulheres apresentaram um crescimento mais acelerado do que os homens de acumulação de anos de estudo até os anos 1990. A partir desse ano, ambos apresentaram tendência semelhante, embora as mulheres com um maior nível de educação.

Figura 3: Evolução dos anos médios de estudo



Nota: Dados de Barro e Lee (2010), população brasileira de 25 ou mais anos de idade.

ramente empíricos.⁶ Seguindo a metodologia de Greenwood et alii (2014) de um arcabouço empírico que independe de um modelo teórico, estudaremos essa relação para o Brasil. Em um primeiro momento, quando apenas olharmos para a melhora do Índice de Gini entre 1970 e 2010 e o crescimento dos casamentos seletivos nesse mesmo período, chegaremos à conclusão de que o Brasil se diferencia dos demais desenvolvidos. No entanto, quando realizamos exercícios contrafactuais, podemos perceber que a melhora na desigualdade de renda brasileira poderia ter sido ainda maior caso o aumento na tendência de casamentos seletivos não tivesse ocorrido.

A primeira análise consiste em olhar para as estatísticas de renda de pessoas casadas nos diferentes níveis educacionais. Na Tabela 1, cada célula nos fornece a renda daquele determinado tipo de casal em relação à renda média da população inteira, casados e solteiros. Quando analisamos de forma geral todas as células, podemos perceber que, com exceção dos casamentos entre pessoas com escolaridade primária incompleta, em todos os tipos de pareamento, a renda do casal se aproximou à renda média da população entre 1970 e 2010, em outras palavras, a renda da sociedade como um todo se tornou mais homogênea. Além disso, quando olhamos para a diagonal principal das matrizes (ou seja, para os casamentos em que os parceiros têm o mesmo nível educacional), as rendas dos casais de níveis educacionais secundário e universidade foram as que mais se aproximaram da renda média.⁷ E vale lembrar que o casamento entre pessoas de nível secundário se tornou o mais comum na sociedade em 2010. Assim, o tipo de casamento de maior frequência foi um dos que mais teve a sua renda se aproximando à renda média da economia. Portanto, apenas por essa análise inicial, temos indícios de que a distribuição da renda brasileira tenha melhorado no período em questão. Mas para confirmar, precisamos computar a Curva de Lorenz e o Índice de Gini.

Para isso, defina $f_{i,j}$ como a fração das famílias que são do tipo i e estão no percentil j de renda e $r_{i,j}$ como a renda desse tipo de família em relação à renda média. O índice percentual j é expresso em termos de fração (ou seja, 0.10 no lugar de 10). Os tipos são classificados da seguinte maneira: casados ou solteiros. No caso dos casados, cada membro da família é identificado também pelo nível

⁶Para análises com modelos teóricos, ver Fernández e Rogerson (2001) e Greenwood et alii (2016). Para trabalhos empíricos, ver Lam (1997), Cancian e Reed (1998, 1999), Schwartz (2010) e Greenwood et alii (2014).

⁷Deve-se notar que o retorno à educação no Brasil deve ter um papel importante também. Para este assunto, ver por exemplo Barbosa Filho e Pessôa (2008).

**Tabela 1:** Tabela de Contingência: Casamento Seletivo e Renda por Nível Educacional

Casamento Seletivo por Nível Educacional								
1970								
Marido	Esposa							
	PC-	PC	SC	C	PC-	PC	SC	C
PC-	0.850 ¹	0.790 ²	0.018	0.046	0.009	0.035	0.000	0.006
PC	0.036	0.058	0.018	0.003	0.009	0.003	0.001	0.000
SC	0.011	0.032	0.011	0.002	0.012	0.001	0.001	0.000
C	0.003	0.022	0.007	0.001	0.011	0.001	0.004	0.000
Marginal	0.900		0.054		0.041		0.006	
2010								
PC-	0.166	0.068	0.078	0.081	0.039	0.095	0.007	0.047
PC	0.049	0.068	0.148	0.081	0.077	0.095	0.017	0.047
SC	0.016	0.069	0.047	0.082	0.180	0.097	0.052	0.048
C	0.001	0.029	0.006	0.035	0.031	0.041	0.086	0.020
Marginal	0.232		0.279		0.327		0.162	
Renda Familiar por Nível Educacional								
1970								
PC-	0.565 ³	0.029 ⁴	1.583	0.118	2.424	0.326	4.172	0.425
PC	1.81	0.037	2.785	0.075	3.469	0.228	5.653	0.349
SC	2.852	0.033	3.928	0.053	4.580	0.174	6.158	0.301
C	5.160	0.025	6.324	0.037	7.042	0.091	8.690	0.177
2010								
PC-	0.357	0.348	0.428	0.329	0.610	0.364	1.062	0.536
PC	0.459	0.296	0.603	0.309	0.729	0.334	1.408	0.481
SC	0.629	0.262	0.724	0.256	1.034	0.318	1.976	0.457
C	1.420	0.188	1.684	0.203	2.120	0.227	4.102	0.362

Note: Cada célula na tabela de contingência possui duas entradas. No painel superior, elas se referem: 1) ao pareamento observado entre marido e esposa e 2) ao que aconteceria se os casamentos fossem aleatórios. No painel inferior, elas denotam: 3) a renda da família relativa à renda média total e 4) a fração da renda feminina na renda total da família. A linha Marginal nos mostra a fração de mulheres em cada nível educacional, ou seja, a distribuição marginal de educação para as mulheres.

educacional. A esposa recebe uma classificação de acordo com a participação no mercado de trabalho (trabalha ou não). Os solteiros são apenas identificados pelo seu nível educacional. E ainda há dez percentis (decis) de renda, de maneira que $j \in \{0.1, 0.2, \dots, 1.0\}$. São gerados, assim, 360 (i, j) possíveis combinações de família para cada ano. A fração da renda agregada que cabe ao percentil j é dada por $s_j = \sum_i f_{i,j} r_{i,j}$. A fração de renda acumulada no percentil p é $l_p \equiv \sum_j^p s_j = \sum_j^p \sum_i f_{i,j} r_{i,j}$. A Curva de Lorenz pode ser representada pelo gráfico de l_p contra $p = \sum_j^p \sum_i f_{i,j}$. E o Índice de Gini, g , é duas vezes a área entre a Curva de Lorenz e a reta de 45°. ⁸ Se p se mover continuamente, então o coeficiente de Gini passa a ser definido por $g = 2 \int_0^1 |l_p - p| dp$, onde $0 \leq g \leq 1$. Quanto maior for g , maior será

⁸Veja o Apêndice para maiores detalhes de como essas curvas são construídas.

a desigualdade de renda. A Curva de Lorenz e o Índice de Gini são funções de $f_{i,j}$ e $r_{i,j}$, para todo i e j . Assim, podemos escrever $l_p = \text{LORENZ}_p(\{f'_{ij}\}, \{r_{ij}\})$ e $g = \text{GINI}(\{f'_{ij}\}, \{r_{ij}\})$.

As Curvas de Lorenz de 1970 e 2010 nos mostram uma queda na desigualdade de renda, com o Índice de Gini indo de 0.609 para 0.572.⁹ A Tabela 2 nos mostra a renda relativa de cada percentil. Note que, entre 1970 e 2010, os percentis mais baixos (0.1 e 0.2) apresentaram uma perda acentuada em sua renda relativa, em especial o percentil 0.1 com uma queda de 82% (indo de 0.03258 para 0.00572). No entanto, todos os percentis entre 0.3 e 0.7 viram as suas rendas relativas crescerem, com destaque para os percentis 0.4 e 0.5 que tiveram um crescimento de 13% e 10%, respectivamente. Já os percentis 0.8 e 0.9 apresentaram uma pequena queda, na ordem de 4% e 2% respectivamente. E por fim, o percentil 1 verificou um aumento na renda relativa de apenas 1,7%. A partir dessa análise, podemos perceber que, em linhas gerais, a renda relativa de praticamente todos os percentis (com exceção para os dois mais baixos) tem se aproximado da renda média, o que corrobora a melhora da distribuição de renda brasileira. Quando comparamos essa distribuição de renda brasileira com a norte-americana entre 1960 e 2005, podemos perceber o quanto a evolução foi diferente entre essas duas economias.¹⁰ Nos Estados Unidos, houve um crescimento da renda relativa dos percentis mais altos (0.8, 0.9 e 1), enquanto os mais baixos (0.1 até 0.7) viram a sua renda cair. E essa tendência, naturalmente, se traduziu em uma piora da desigualdade (e aumento do Índice de Gini).

Tabela 2: Desigualdade de Renda

Percentil de renda	Renda familiar relativa à média	
	1970	2010
0.1	0.03258	0.00572
0.2	0.14659	0.11823
0.3	0.22434	0.23933
0.4	0.29699	0.33641
0.5	0.38578	0.42601
0.6	0.50822	0.54831
0.7	0.70218	0.71148
0.8	1.00548	0.98516
0.9	1.66101	1.59004
1	5.14919	5.23769

Embora para o Brasil não observemos uma piora na desigualdade de renda, apesar do aumento de casamentos seletivos, contrariando assim o que a literatura vem apontando nos países desenvolvidos, surge uma pergunta: será que a distribuição de renda brasileira teria melhorado ainda mais caso o crescimento na tendência de casamentos seletivos não tivesse ocorrido? Para respondê-la, realizamos o seguinte experimento: substituímos o pareamento observado na tabela de contingência presente na Tabela 1 pelo padrão da tabela de contingência aleatória (representada em itálico também na Tabela 1). Vejamos como isso é realizado.

Primeiro, defina \mathcal{M} como o conjunto de índice para os casados e \mathcal{S} , o conjunto para os solteiros. O conjunto \mathcal{M} é definido particularmente da seguinte forma. Considere uma distribuição qualquer $\{f_{ij}\}$ de famílias. Vale lembrar que as famílias compostas por casais são indexadas pelo nível educacional do marido e da esposa, a participação de cada um no mercado de trabalho e o número de filhos. Defina

⁹Ver Tabela 2.

¹⁰Ver Figure 2 em Greenwood et alii (2014).



\mathcal{M}_{E_H} como os conjuntos que contêm os índices de todas as famílias de casais cujo marido possui nível educacional $E_H \in \{PC-, PC, SC, C\}$. De forma similar, os conjuntos \mathcal{M}_{E_W} contêm todas as famílias de casais cuja esposa possui nível educacional E_W . Além disso, \mathcal{M}_{LFP_H} (\mathcal{M}_{LFP_W}) contêm todas as famílias de casados cujo (a) marido (esposa) possui a participação no mercado de trabalho $LFP_{H(W)} \in \{TRABALHA_{H(W)}, \sim TRABALHA_{H(W)}\}$. E por fim, \mathcal{M}_{KIDS} contêm todas as famílias de casados com um particular número de filhos $KIDS \in \{0, 1, 2, 2+\}$. O conjunto de todas as famílias de casados com um particular pareamento de nível educacional \mathcal{M}_{E_H, E_W} é dado por

$$\mathcal{M}_{E_H, E_W} = \mathcal{M}_{E_H} \cap \mathcal{M}_{E_W}.$$

Assim, o conjunto que contém todas as diferentes famílias é:

$$\mathcal{M} = \bigcup_{E_H, E_W, LFP_H, LFP_W, KIDS} (\mathcal{M}_{E_H} \cap \mathcal{M}_{E_W} \cap \mathcal{M}_{LFP_H} \cap \mathcal{M}_{LFP_W} \cap \mathcal{M}_{KIDS}),$$

onde o termo dentro do parênteses representa todos os casados do tipo $(E_H, E_W, LFP_H, LFP_W, KIDS)$.

Com os conjuntos \mathcal{M} e \mathcal{S} devidamente definidos, agora mostraremos como de fato o experimento de casamento aleatório é realizado. Ele se resume a substituir os $\{f_{i,j}\}$'s observados para $(i,j) \in \mathcal{M}$ pelo conjunto obtido caso os casamentos tivessem ocorrido de forma aleatória, aqui representado por $\{\tilde{f}_{i,j}\}$ para $(i,j) \in \mathcal{M}$. O Índice de Gini é então dado por $GINI(\{f'_{i,j}\}, \{r_{i,j}\})$, onde $\{f'_{i,j}\} \equiv \{\tilde{f}_{i,j}\}_{\mathcal{M}} \cup \{f_{i,j}\}_{\mathcal{S}}$. Assim, considere o primeiro elemento da tabela de contingência de 1970 presente na Tabela 1, que representa os casamentos entre pessoas com nível primário de educação incompleto. Em 1970, essa fração era de 0.85 e pode ser representada da seguinte forma

$$\frac{\sum_{i \in \mathcal{M}_{PC-, PC-}} \sum_{j=0.1}^1 f_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 f_{ij}^{1970}} = 0.85.$$

Agora, impondo que os casamentos sejam aleatórios, a fração desse tipo de casamento passa a ser 0.79. Chame essa nova distribuição de \tilde{f}_{ij}^{1970} e de forma similar à expressão acima, temos

$$\frac{\sum_{i \in \mathcal{M}_{PC-, PC-}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 f_{ij}^{1970}} = 0.79.$$

Assim, a razão entre o número total de casamentos aleatórios entre pessoas com nível primário incompleto e o total de casamentos desse tipo observados nos dados é

$$\frac{\sum_{i \in \mathcal{M}_{HS-, HS-}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}_{HS-, HS-}} \sum_{j=0.1}^1 f_{ij}^{1970}} = \frac{0.79}{0.85} = 0.93.$$

Portanto, sob a hipótese de casamentos aleatórios, o número de casamentos do tipo $(PC-, PC-)$ é reduzido por um fator de $0.93 = 0.79/0.85$. Assuma que essa redução possa ser generalizada para todos os percentis de renda, isto é, para todos os j 's. Assim, quando realizarmos o exercício de casamentos aleatórios, a fração \tilde{f}_{ij}^{1970} deve ser construída da seguinte forma

$$\tilde{f}_{ij}^{1970} = \frac{0.79}{0.85} f_{ij}^{1970}, \text{ para } i \in \mathcal{M}_{PC-, PC-} \text{ e todo } j.$$

Esse procedimento é realizado para todos os demais 15 tipos de casamentos. E para os indivíduos solteiros e divorciados, mantemos as frações originais (provenientes dos dados), ou seja, $\tilde{f}_{ij}^{1970} = f_{ij}^{1970}$.

Na Tabela 3, podemos observar os resultados desse exercício contrafactual. Note que, tanto para 1970 quanto para 2010, há uma queda no Índice de Gini se comparamos com os dados (indo de 0.609 para 0.587 em 1970 e de 0.572 para 0.543 em 2010). A partir desses números, podemos concluir que o crescimento na tendência de casamentos seletivos impacta a distribuição de renda em uma economia e que a queda ligeiramente maior para o ano de 2010 mostra um nível de casamento seletivo mais elevado. Assim, a resposta à pergunta anterior fica evidente: a desigualdade de renda brasileira teria melhorado ainda mais caso a diversidade conjugal não tivesse caído ao longo dos anos.

Tabela 3: Índice de Gini, Dados e Experimentos

Base para o Índice de Gini	1970	2010
Dados	0.609	0.572
Pareamento Aleatório	0.587	0.543
Aleatório + 2010 PFFT	0.562	
Aleatório + 1970 PFFT		0.585
Tabela Padronizada	0.603	0.561
Tabela Padronizada + 2010 PFFT	0.591	
Tabela Padronizada + 1970 PFFT		0.600
Mais solteiras (nível 2010)	0.614	
Menos solteiras (nível 1970)		0.568
Mais mulheres trabalhando (nível 2010)	0.596	
Menos mulheres trabalhando (nível 1970)		0.603

Outro experimento interessante a ser feito é verificar o que ocorreria com a distribuição de renda em 1970 e 2010, caso trocássemos o percentual de solteiros entre os anos, ou seja, caso assumíssemos que o nível de solteiros de 1970 fosse o mesmo de 2010 e o de 2010 fosse igual ao de 1970. Note pela Tabela 3 que o Índice de Gini subiria de 0.609 para 0.614 (diferença de 0.005) em 1970 e cairia de 0.572 para 0.568 (diferença de 0.004) em 2010. Esses resultados nos levam a conclusão de que um maior número de casamentos ajuda a melhorar a distribuição de renda brasileira, embora esse efeito não seja muito expressivo.

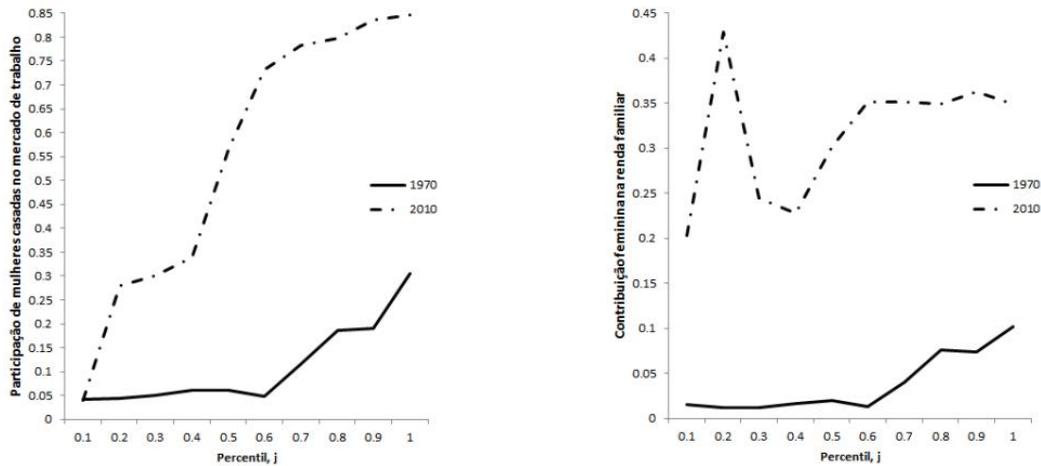
Os dois experimentos acima só causam algum impacto na distribuição de renda se pelo menos algumas mulheres casadas participarem do mercado de trabalho. Olhando para a Figura 4, que nos mostra a participação de mulheres casadas nesse mercado por percentil de renda, podemos perceber que essa participação é muito maior em 2010 do que em 1970, para todos os percentis acima de 0.1. Além disso, houve um crescimento bastante elevado dessa participação a partir do (incluindo) percentil 0.5, com destaque para o 0.6 (o percentual de mulheres casadas trabalhando cresceu de 5% para 73%). Ainda na Figura 4, também podemos observar a contribuição da renda da mulher na renda da família por percentil. Note que, novamente, esse tipo de participação é muito superior em 2010 se comparada a 1970. Além disso, para todos os percentis houve um crescimento da renda da mulher na renda total da família da ordem de 20 pontos percentuais (p.p.).¹¹ Assim, temos fortes indícios de que a oferta de mão de obra das mulheres casadas no mercado de trabalho impacta a distribuição de renda.

Para comprovar o ponto anterior, realizaremos o seguinte exercício contrafactual: primeiro, assumamos que todos os casamentos ocorreram de forma aleatória em 1970 e 2010 e depois, que as mulheres casadas em 1970 participaram no mercado de trabalho com o mesmo nível de 2010 e que em 2010 elas trabalharam ao nível de 1970. O algoritmo para obter as tabelas de contingência assumindo casamentos aleatórios já foi descrito anteriormente. Suponha agora que, além disso, impomos que as mulheres

¹¹O percentil 0.2 viu sua participação crescer quase 42 p.p., o que pode ser uma anomalia dos dados.



Figura 4: Mulheres casadas no mercado de trabalho, 1970 e 2010



Nota: O lado direito mostra a participação de mulheres casadas no mercado de trabalho por percentil de renda para 1970 e 2010. O lado esquerdo ilustra a fração da renda feminina na renda total da família.

casadas participem do mercado de trabalho em 1970 com os níveis de 2010.¹² A participação de mulheres casadas na força de trabalho em 1970, quando assumimos a hipótese de casamentos aleatórios, é

$$\frac{\sum_{i \in \mathcal{M}_{TRABALHA_W}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}} = 0.13,$$

enquanto essa participação em 2010 é

$$\frac{\sum_{i \in \mathcal{M}_{TRABALHA_W}} \sum_{j=0.1}^1 f_{ij}^{2010}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 f_{ij}^{2010}} = 0.64.$$

Chamemos essa nova distribuição contrafactual de famílias de casados em 1970 de \hat{f}_{ij}^{1970} , para $i \in \mathcal{M}$ e para todo j . Essa distribuição para 1970 deve nos dar a participação feminina no mercado de trabalho ao nível de 2010, ou seja,

$$\frac{\sum_{i \in \mathcal{M}_{TRABALHA_W}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970}} = 0.64.$$

É válido ter em mente que a fração de pessoas casadas em 1970 não muda nesses experimentos contrafactuals, isto é,

$$\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 f_{ij}^{1970} = \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970} = \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970}.$$

Portanto,

$$\frac{\sum_{i \in \mathcal{M}_{TRABALHA_W}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970}} = \frac{0.64}{0.13} \frac{\sum_{i \in \mathcal{M}_{TRABALHA_W}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}} = 0.64.$$

¹²O procedimento é o mesmo para o caso em que as mulheres em 2010 trabalharam ao nível de 1970.

Impor a taxa de participação feminina no mercado de trabalho de 2010 em 1970 implica em escalonar para cima todas as combinações (i, j) de famílias de casados em que a esposa trabalha. Por outro lado, as famílias de casados em que a esposa não trabalha devem ser escalonadas para baixo de tal maneira que a fração total de famílias de casados não se altere.

Portanto, a distribuição contrafactual $\{\hat{f}_{ij}^{1970}\}$ deve ser construída da seguinte maneira:

$$\hat{f}_{ij}^{1970} = \frac{0.64}{0.13} \tilde{f}_{ij}^{1970}, \text{ para } i \in \mathcal{M}_{TRABALHAW} \text{ e para todo } j,$$

e

$$\hat{f}_{ij}^{1970} = \frac{1 - 0.64}{1 - 0.13} \tilde{f}_{ij}^{1970}, \text{ para } i \in \mathcal{M}^{-TRABALHAW} \text{ e para todo } j.$$

Logo, a fração total de famílias de casados permanece constante,

$$\begin{aligned} \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970} &= \sum_{i \in \mathcal{M}_{TRABALHAW}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970} + \sum_{i \in \mathcal{M}^{-TRABALHAW}} \sum_{j=0.1}^1 \hat{f}_{ij}^{1970} \\ &= \frac{0.64}{0.13} \sum_{i \in \mathcal{M}_{TRABALHAW}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970} + \frac{1 - 0.64}{1 - 0.13} \sum_{i \in \mathcal{M}^{-TRABALHAW}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970} \\ &= 0.64 \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970} + (1 - 0.64) \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970} \\ &= \sum_{i \in \mathcal{M}} \sum_{j=0.1}^1 \tilde{f}_{ij}^{1970}. \end{aligned}$$

Assim como a distribuição contrafactual do experimento anterior, a distribuição para solteiros e divorciados permanece inalterada, $\hat{f}_{ij}^{1970} = f_{ij}^{1970}$.

Nesse experimento, os Índices de Gini passam a ser 0.562 para 1970 e 0.585 para 2010. A hipótese de casamentos aleatórios faz com que as rendas fiquem mais equalizadas entre as famílias, pois acaba diversificando a renda entre maridos e esposas de diferentes níveis educacionais. Mas esse efeito só se torna real à medida em que as mulheres casadas participarem do mercado de trabalho. Assim, comparando o Índice de Gini do caso de casamentos aleatórios com o caso de casamentos aleatórios mais a troca entre os anos da participação das mulheres casadas no mercado de trabalho, observamos que essa estatística caiu de 0.587 para 0.562 em 1970 e cresceu de 0.543 para 0.585 em 2010. Portanto, o fato das mulheres casadas se inserirem mais ou menos no mercado afeta a desigualdade de renda brasileira.

Outro importante experimento a se pensar seria medir a distribuição de renda caso os casamentos em 2010 tivessem ocorrido da mesma maneira que em 1970, ou seja, caso a tabela de contingência de 2010 fosse substituída pela de 1970. No entanto, esse exercício contrafactual apresenta um grande problema. Em 2010, as pessoas (homens e principalmente mulheres) tinham níveis educacionais mais altos do que em 1970. Com isso, a distribuição marginal de maridos e esposas em cada nível educacional mudou bastante entre 1970 e 2010, inviabilizando em um primeiro momento esse novo experimento. Assim, caso realizássemos esse experimento sem qualquer correção, não seria possível distinguir os efeitos dos casamentos seletivos dos efeitos do aumento do nível educacional da população sobre a distribuição de renda.¹³ Para contornar essa questão, podemos utilizar o procedimento proposto por Mosteller (1968) e construir a chamada tabela de contingência padronizada para 1970, usando as distribuições marginais de níveis educacionais para homens e mulheres de 2010, e para 2010 usando as distribuições de 1970.

Mosteller (1968) sugere que, quando comparamos duas tabelas de contingência, elas devem primeiro ser padronizadas de maneira que ambas possuam as mesmas distribuições marginais associadas às colunas e às linhas. Considere uma tabela 4×4 . Ela pode ser padronizada tal que cada elemento das duas

¹³Isto é, a mudança das distribuições marginais de homens e mulheres entre os diferentes níveis educacionais.



distribuições marginais seja $1/4$. Isso pode ser feito usando o algoritmo de Sinkhorn e Knopp (1967), que escalona cada linha e cada coluna de forma iterativa.¹⁴ A padronização preserva o padrão central de associação em uma tabela de contingência. Por exemplo, Tan et alii (2004) notam que tal padronização não altera as razões de possibilidade (medida típica usada para avaliar o padrão de associação entre variáveis) em uma tabela de contingência.

As tabelas padronizadas para 1970 e 2010 estão ilustradas na Tabela 4. A soma dos elementos da diagonal principal em 2010 é maior do que em 1970, o que nos leva a concluir que houve um crescimento na tendência de casamentos seletivos.

Tabela 4: Tabela de Contingência Padronizada: Casamento Seletivo por Nível Educacional

Casamento Seletivo por Nível Educacional				
1970				
Distribuição Marginal = $(1/4, \dots, 1/4)$				
Marido	Esposa			
	PC-	PC	SC	C
PC-	0.201	0.028	0.017	0.004
PC	0.034	0.114	0.067	0.035
SC	0.012	0.074	0.101	0.063
C	0.003	0.034	0.065	0.148
Marginal, Esposas	$1/4$	$1/4$	$1/4$	$1/4$
2010				
Distribuição Marginal = $(1/4, \dots, 1/4)$				
PC-	0.166	0.058	0.021	0.005
PC	0.058	0.130	0.048	0.014
SC	0.022	0.048	0.130	0.049
C	0.004	0.013	0.050	0.182
Marginal, Esposas	$1/4$	$1/4$	$1/4$	$1/4$

Nota: O painel superior mostra a tabela de contingência para 1970 quando ela é normalizada usando o algoritmo de Sinkhorn-Knopp tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual $1/4$. O painel inferior mostra o mesmo para 2010.

Note que cada distribuição marginal das tabelas padronizadas não precisa necessariamente ser $1/4$. A padronização da tabela de 1970 pode ser tal que cada distribuição marginal coincida com aquelas provenientes dos dados de 2010, ou vice-versa. Esses resultados são mostrados na Tabela 5.

A partir dessas tabelas de contingência padronizadas, podemos computar a desigualdade de renda para o experimento que estávamos pensando anteriormente. Os Índices de Gini associados são 0.603 para 1970 e 0.561 para 2010. Assim, caso as pessoas em 2010 se casassem como em 1970 (seguindo a tabela de contingência padronizada), o coeficiente de Gini teria caído de 0.572 para 0.561, ou seja, teríamos uma melhora na distribuição de renda brasileira.

Além do experimento com essas tabelas de contingência padronizadas, também podemos realizar outro exercício adicionando a hipótese de troca da participação das mulheres casadas no mercado de trabalho entre os anos. Assim, assumindo para 1970 a distribuição proveniente da tabela de contingência padronizada e a participação das mulheres casadas ao nível de 2010, o Índice de Gini vai de 0.609

¹⁴Ver Apêndice para maiores detalhes sobre esse algoritmo.

Tabela 5: Tabela de Contingência Padronizada: Casamento Seletivo por Nível Educacional

Casamento Seletivo por Nível Educacional				
1970				
Usando as Distribuições Marginais de 2010				
Marido	Esposa			
	PC-	PC	SC	C
PC-	0.201	0.047	0.036	0.005
PC	0.024	0.131	0.101	0.034
SC	0.008	0.082	0.146	0.059
C	0.001	0.017	0.043	0.063
Marginal, Esposas	0.234	0.277	0.326	0.161
2010				
Usando as Distribuições Marginais de 1970				
PC-	0.830	0.034	0.013	0.000
PC	0.047	0.012	0.005	0.000
SC	0.018	0.004	0.012	0.001
C	0.007	0.003	0.010	0.005
Marginal, Esposas	0.902	0.053	0.040	0.006

Nota: O painel superior mostra a tabela de contingência para 1970 quando ela foi normalizada usando o algoritmo de Sinkhorn-Knopp tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual aos dados de 2010. O painel inferior mostra a tabela de contingência para 2010 quando ela foi normalizada tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual aos dados de 1970.

(dados) para 0.591. Fazendo o mesmo para 2010 e diminuindo a participação das mulheres casadas ao nível de 1970, essa estatística cresce de 0.572 (dados) para 0.6, aumentando assim a desigualdade de renda. Mais uma vez mostramos que a participação das mulheres no mercado de trabalho é crucial para entender a distribuição de renda.

4. CONCLUSÃO

Neste artigo, a partir de análises dos censos demográficos brasileiros e de exercícios puramente contábeis, mostramos que ao longo do tempo os brasileiros estão se casando cada vez mais com parceiros que possuem características socioeconômicas semelhantes. Esse fenômeno já vem sendo observado para alguns países desenvolvidos e está largamente documentado por sociólogos e economistas. Além disso, a literatura tem mostrado que a tendência crescente de casamentos seletivos impacta de forma negativa a desigualdade de renda. Quando nos debruçamos sobre essa questão para o Brasil, na verdade observamos, em um primeiro momento, o contrário: a desigualdade de renda melhorou de 1970 para 2010, embora a frequência de casamentos seletivos tenha aumentado. No entanto, quando realizamos alguns exercícios contrafactuais, podemos observar que a distribuição de renda brasileira teria melhorado ainda mais, caso a tendência de casamentos seletivos não tivesse ocorrido.



BIBLIOGRAFIA

- Barbosa Filho, F. d. H. & Pessôa, S. (2008). Retorno da Educação no Brasil. *Pesquisa e Planejamento Econômico*, 38(1):97–126.
- Barro, R. & Lee, J.-W. (2010). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104:184–198.
- Barros, R. P. d., de Carvalho, M., Franco, S., & Mendonça, R. (2006). Consequências e Causas Imediatas da Queda Recente da Desigualdade de Renda Brasileira. Technical report, Discussion Papers 1201, IPEA.
- Barros, R. P. d., Firpo, S., Guedes, R., & Leite, P. (2000). Demographic Changes and Poverty in Brazil. Technical report, Discussion Papers 0096, IPEA.
- Blossfeld, H.-P. & Timm, A. (2003). *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*. Kluwer Academic, Dordrecht.
- Cancian, M. & Reed, D. (1998). Assessing the Effects of Wives' Earnings on Family Income Inequality. *The Review of Economics and Statistics*, 80(1):73–79.
- Cancian, M. & Reed, D. (1999). The Impact of Wives' Earnings on Income Inequality: Issues and Estimates. *Demography*, 36(2):173–84.
- Chiappori, P.-A., Iyigun, M., & Weiss, Y. (2009). Investment in Schooling and the Marriage Market. *American Economic Review*, 99(5):1689–1713.
- Eika, L., Mogstad, M., & Zafar, B. (2014). Educational Assortative Mating and Household Income Inequality. NBER Working Paper 20271.
- Fernández, R. & Rogerson, R. (2001). Sorting and Long-Run Income Inequality. *Quarterly Journal of Economics*, 116(4):1305–41.
- Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. NBER Working Paper 11953.
- Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2014). Marry your Like: Assortative Mating and Income Inequality. *American Economic Review Papers & Proceedings*, 104(5):348–353.
- Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2016). Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment and Married Female Labor-Force Participation. *American Economic Journal: Macroeconomics*, 8(1):1–41.
- Lam, D. (1997). Demographic Variables and Income Inequality. In Rosenzweig, M. R. & Stark, O., editors, *Handbook of Population and Family Economics*, pages 1015–1059. Elsevier North Holland.
- Lundberg, S. (1985). The Added Worker Effect. *Journal of Labor Economics*, 3(1):11–37.
- Mosteller, F. (1968). Association and Estimation in Contingency Tables. *Journal of the American Statistical Association*, 63(321):1–28.
- Neri, M. (2005). *Sexo, Casamento e Economia*. FGV/IBRE, CPS, Rio de Janeiro.
- Rao, V. M. (1969). Two Decompositions of Concentration Ratio. *Journal of the Royal Statistical Society, Series A (General)*, 132(3):418–425.
- Schwartz, C. R. (2010). Earnings Inequality and the Changing Association between Spouses' Earnings. *American Journal of Sociology*, 115(5):1524–1557.

- Schwartz, C. R. & Mare, R. D. (2005). Trends in Educational Assortative Marriage from 1940 to 2003. *Demography*, 42(4):621–646.
- Sinkhorn, R. & Knopp, P. (1967). Concerning Nonnegative Matrices and Doubly Stochastic Matrices. *Pacific Journal of Mathematics*, 21(2):343–348.
- Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. *Information Systems*, 29(4):293–313.

A. DADOS

A base de dados usada neste estudo é o censo demográfico brasileiro dos anos de 1970, 1980, 1991, 2000 e 2010 e pode ser facilmente obtida através da página do IPUMS (Integrated Public Use Microdata Series) na internet. As variáveis incluídas na nossa base de dados são: ano do censo (nome da variável: year), código identificador da esposa (sploc), número de membros em uma família (famsize), número de filhos (nchild), idade (age), sexo (sex), estado civil (marst), nível educacional (educ), relação de emprego (empstat) e renda individual (inc). No nosso estudo só consideramos casais e solteiros entre 25 e 54 anos. Os adultos vivem por conta própria ou com seus filhos, que têm idade inferior a 19 anos. Famílias em que existem outros membros tais como avós, tios/tias, ou quaisquer outros indivíduos estão excluídas da base de dados. Assim como famílias com subfamílias de qualquer tipo. Além disso, viúvos/viúvas e indivíduos casados cujas esposas estão ausentes também são excluídos. As rendas são estritamente positivas.

Na nossa análise, há 384 tipos diferentes de famílias e elas são classificadas da seguinte maneira:

1. Estado civil: casados, homens nunca casados, mulheres nunca casadas, homens divorciados e mulheres divorciadas;
2. Educação: primária incompleta, primária, secundária e universidade;
3. Mercado de trabalho: trabalha, não trabalha;
4. Filhos: nenhum filho, um, dois, ou mais de dois filhos.

Por fim, as famílias são divididas em dez percentis. Assim, para cada ano, temos 3840 (i, j) combinações de famílias para todos os tipos e decis.

B. CURVA DE LORENZ E ÍNDICE DE GINI

Suponha uma amostra de diferentes tipos de famílias, $i \in \{1, 2, \dots, m\}$, situadas em diferentes percentis, $j \in \mathcal{J}$, da distribuição de renda. Defina f_{ij} como a fração de famílias que são do tipo i e estão no percentil j e r_{ij} como a renda y_{ij} relativa à renda média y . A renda de cada família é ajustada para a base equivalente de um adulto de acordo com a metodologia da OCDE, que considera o primeiro adulto como 1, o segundo como 0.5 e cada criança como 0.3. As rendas equalizadas das famílias são então divididas pela renda média da população.

A fração de renda do percentil j é

$$s_j = \sum_i f_{ij} r_{ij}.$$

A Curva de Lorenz é derivada plotando as frações acumuladas da população indexadas por p ,

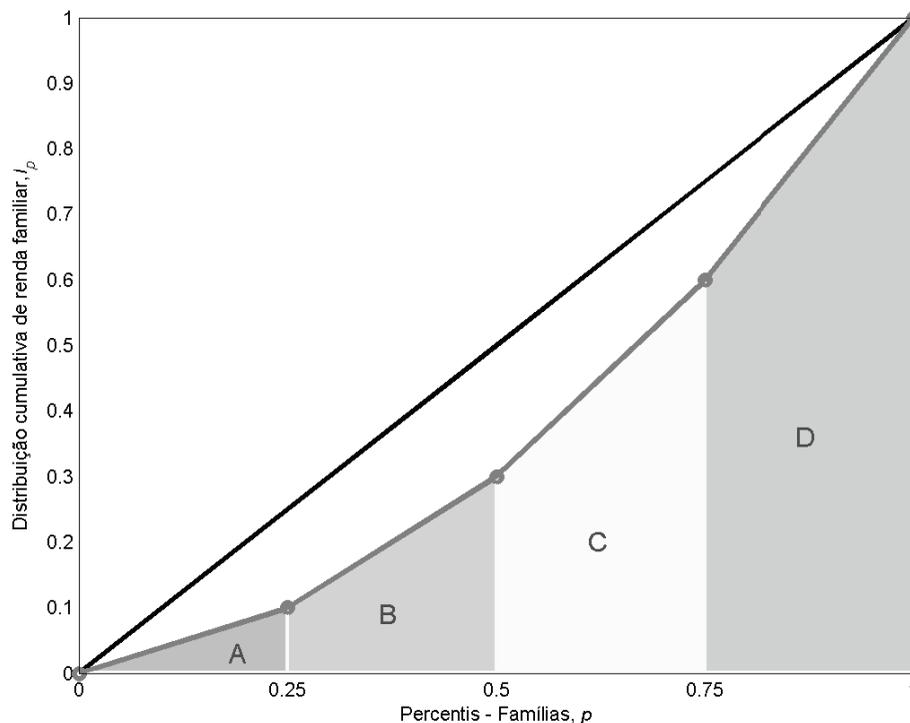
$$p = \sum_j \sum_i f_{ij}$$

no eixo x contra a fração de renda acumulada pelo percentil p ,

$$l_p = \sum_j^p s_j,$$

no eixo y . Suponha que o intervalo unitário seja dividido em n segmentos iguais. Assim, $j \in \mathcal{J} = \{1/n, \dots, 1 - 1/n, 1\}$.

Imagine o exemplo em que $n = 4$ (quartil). A Curva de Lorenz descrita acima está plotada na Figura 1. O Índice de Gini associado a essa Curva de Lorenz é duas vezes a área entre a curva e a linha de 45 graus. O índice também pode ser calculado como $1 - 2\Delta$, onde Δ é a área abaixo da Curva de Lorenz. No caso em que $n = 4$, o índice é a soma das áreas do triângulo A e dos trapezóides B , C , e D . As coordenadas no eixo x são dadas por $0, p_1 = 0.25, p_2 = 0.5, p_3 = 0.75$, e 1.0 . As coordenadas no eixo y são $0, l_{0.25}, l_{0.5}, l_{0.75}$, e 1.0 .



Nota: A figura mostra a construção de uma Curva de Lorenz quando há quatro percentis (quartis). O Índice de Gini é duas vezes a área entre a linha de 45 graus e a Curva de Lorenz.

Assim, utilizando as fórmulas para as áreas geométricas A , B , C , e D , o Índice de Gini, g , pode ser derivado como

$$g = 1 - 2 \left(\underbrace{\frac{p_1 l_1}{2}}_{\text{Área A}} + \underbrace{\frac{(l_1 + l_2)(p_2 - p_1)}{2}}_{\text{Área B}} + \underbrace{\frac{(l_2 + l_3)(p_3 - p_2)}{2}}_{\text{Área C}} + \underbrace{\frac{(l_3 + 1)(1 - p_3)}{2}}_{\text{Área D}} \right).$$

Após rearranjar a expressão acima, o Índice de Gini é dado por

$$g = (p_1 l_2 - p_2 l_1) + (p_2 l_3 - p_3 l_2) + (p_3 - l_3).$$

As frações acumuladas da população, os p 's, são baseadas nos quartis; i.e., $p_1 = 1/4, p_2 = 2/4, \dots$. Portanto, a expressão anterior pode ser escrita como

$$g = \frac{1}{4} [(l_2 - 2l_1) + (2l_3 - 3l_2) + (3 - 4l_3)].$$

No caso geral de n percentis, o Índice de Gini é

$$g = \sum_{p=1/n}^{1-1/n} [pl_{p+1} - (p + 1/n)l_p].$$

A versão dessa fórmula para um número arbitrário de grupos de renda e de sub-populações (tipos) é apresentada em Rao (1969).

C. ALGORITMO SINKHORN-KNOPP (1967)

1. Comece uma iteração com uma tabela de contingência.
2. Essa tabela de contingência possui uma distribuição marginal associada a cada linha (marido) obtida a partir da soma de cada linha ao longo de suas colunas para obter o total daquela linha. Divida cada linha por 4 vezes o seu total. A distribuição marginal associada às linhas passa a ser $(1/4, 1/4, 1/4, 1/4)$.
3. Compute a distribuição marginal associada a cada coluna (esposa) somando cada coluna ao longo de suas linhas para obter o total daquela coluna. Divida cada coluna por 4 vezes o seu total.
4. Recompute a distribuição marginal associada a cada linha. Verifique se ela mudou após as duas etapas anteriores, checando a distância até a distribuição marginal desejada $(1/4, 1/4, 1/4, 1/4)$. Se atingiu o nível de proximidade desejado, então pare. Caso contrário, volte ao passo 1.