



CEDEPLAR

MICROECONOMETRIA

1º RELATÓRIO

Luca Klein

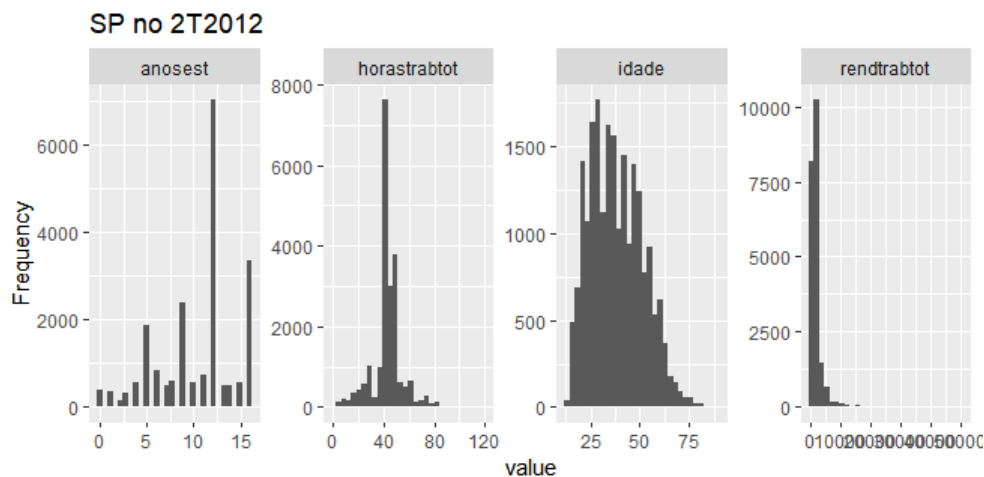
# 1.Dados e Modelo

A base de dados selecionada para o relatório são os microdados da PNAD referentes ao 2T2012 e 2T2022. A motivação por trás do trabalho irá se pautar na comparação entre as UFs com a menor e a maior renda auferida de todos os trabalhos, na média, com base no primeiro período a ser analisado para que, assim, seja possível avaliar a mudança de renda após 10 anos. Para tal, no que tange a UF com a menor renda, o Estado a ser tratado será o Maranhão (MA) que, por sua vez, ainda apresenta o posto de menor rendimento médio de todos os trabalhos no segundo período analisado. Pelo lado dos maiores rendimentos, a UF a ser tratada será São Paulo (SP). Porém, cabe ressaltar que tal UF registra o 2º maior rendimento médio de todos os trabalhos, atrás do Distrito Federal (DF), no 2T12. No entanto, a exclusão do DF se dá pelas especificidades, uma vez que boa parte dos rendimentos do Estado são oriundos do funcionalismo público, principalmente no que se refere a esfera federal. Tendo isso em vista, a comparação do DF com qualquer outra UF traria um viés de seleção que não esboça o cenário brasileiro, em grande medida, dado suas particularidades

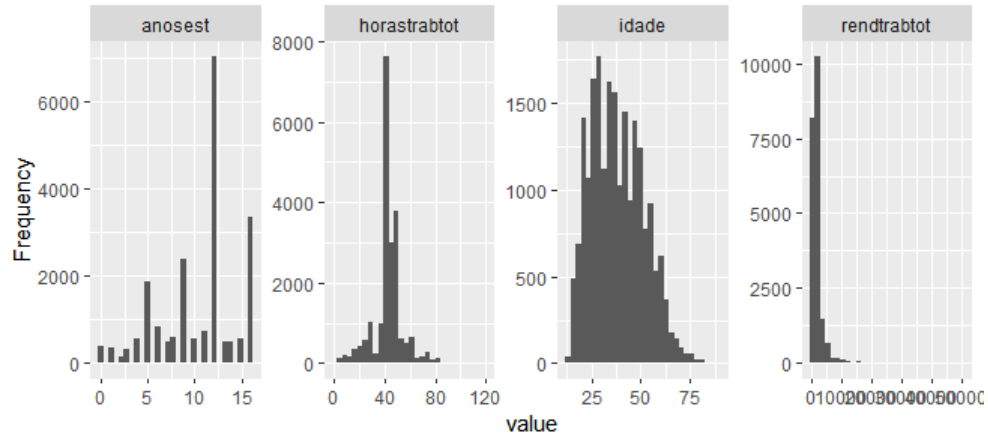
O modelo a ser trabalhado irá considerar como variável explicada os rendimentos totais do trabalho (*rendtrabtot*), uma vez que a informalidade e a volatilidade do mercado de trabalho aumentaram nos últimos tempos e tal variável é capaz de capturar melhor a dinâmica da renda tanto do trabalho principal como dos demais. As variáveis explicativas adotadas são anos de estudos (*anosest*), idade e número de horas totais despendidas nos trabalhos (*horastrabtot*). Por fim, será aplicado um recorte na renda para valores menores e iguais R\$ 5000,00, dado que renda média nacional era R\$ 2304 no 2T2012, sendo que algumas observações ficam acima do valor destacado e podem ser consideradas *outliers*, logo a estimação contendo tais valores podem comprometer os parâmetros e as inferências a serem feitas.

## 1.1 Análise Descritiva

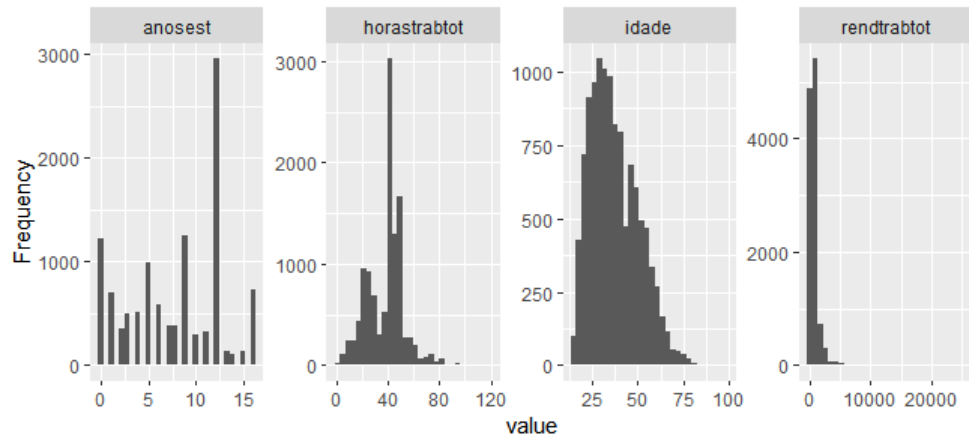
Para o primeiro momento, o corte sobre a renda ainda não será aplicado, tendo em vista a necessidade de evidenciar os *outliers* da amostra total dos microdados. Logo, para SP e MA, para ambos os períodos, verifica-se comportamentos semelhantes em relação as distribuições das variáveis explicativas frente a variável explicada. Então, no que tange aos anos de estudos, há evidências de uma relação positiva entre eles, ou seja, quanto maior a escolaridade maior o rendimento. Já pelo lado das horas de trabalho e da idade, a dinâmica das variáveis parecem sinalizar rendimentos marginais decrescentes, com isso o aumento nas variáveis geram um incremento cada vez menor na renda. Para idade, esse comportamento vem em linha com algumas teorias de consumo, à exemplo Friedman (1957) e Modigliani (1986). Enquanto para horas trabalhadas, o aumento da informalidade conjuntamente com a falta de escolaridade e a necessidade de mais de um trabalho são vetores que favorecem esse tipo de dinâmica.



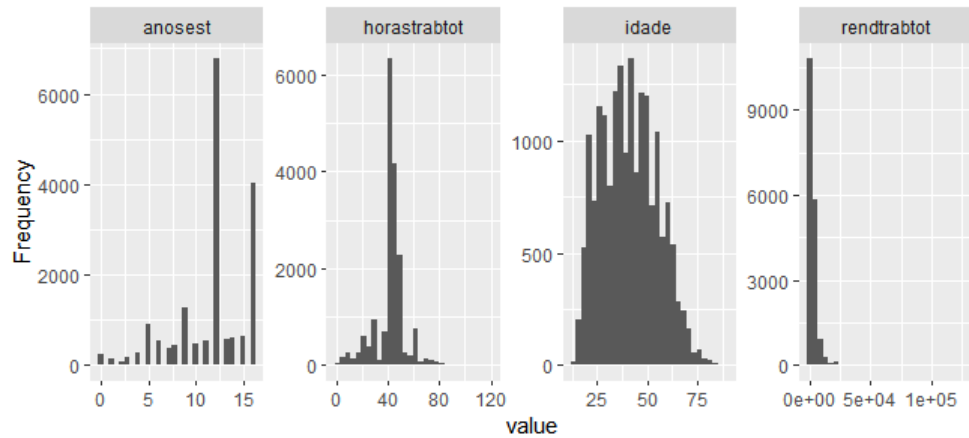
SP no 2T2012



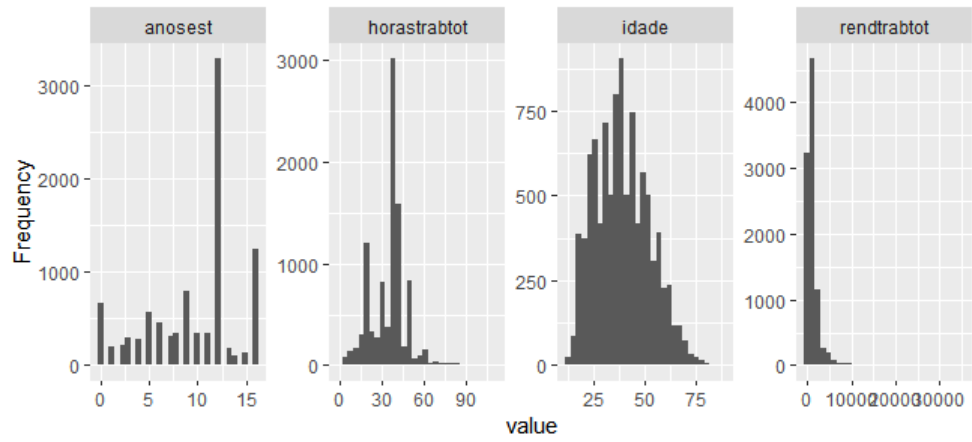
MA no 2T2012



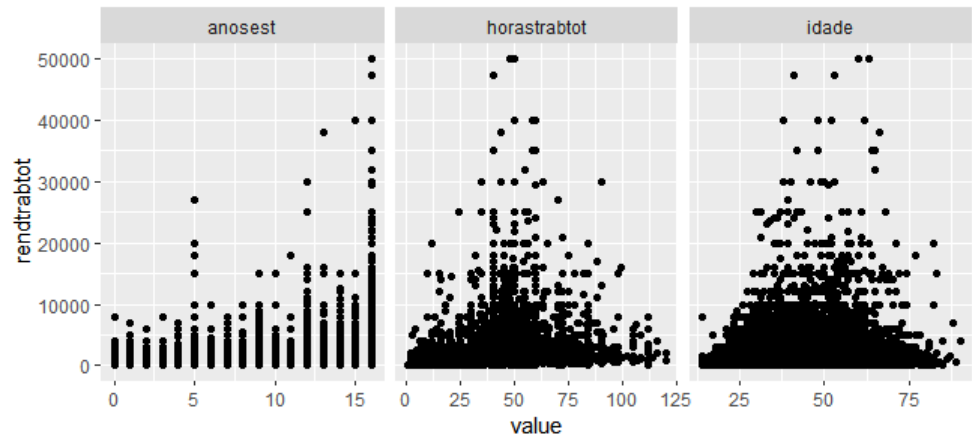
SP no 2T2022



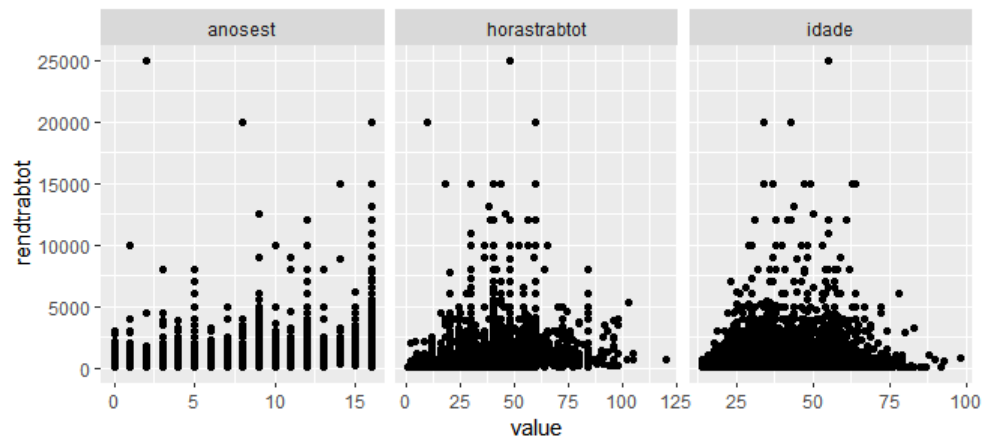
### MA no 2T2022



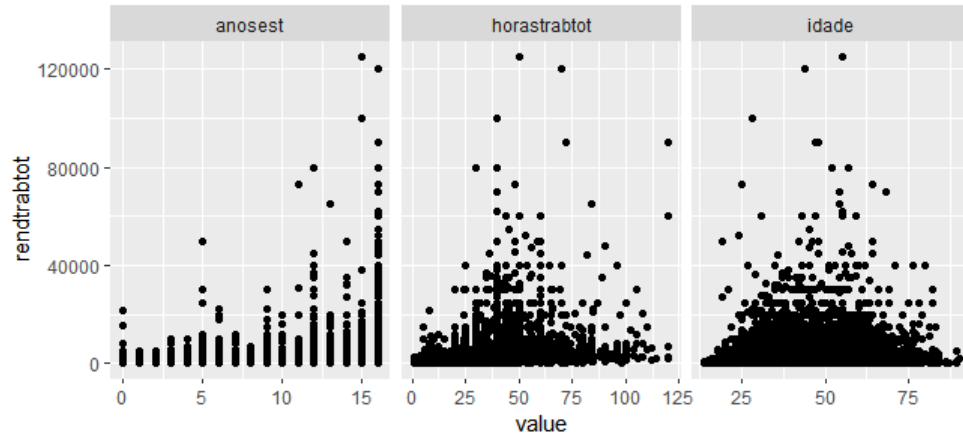
### SP no 2T2012



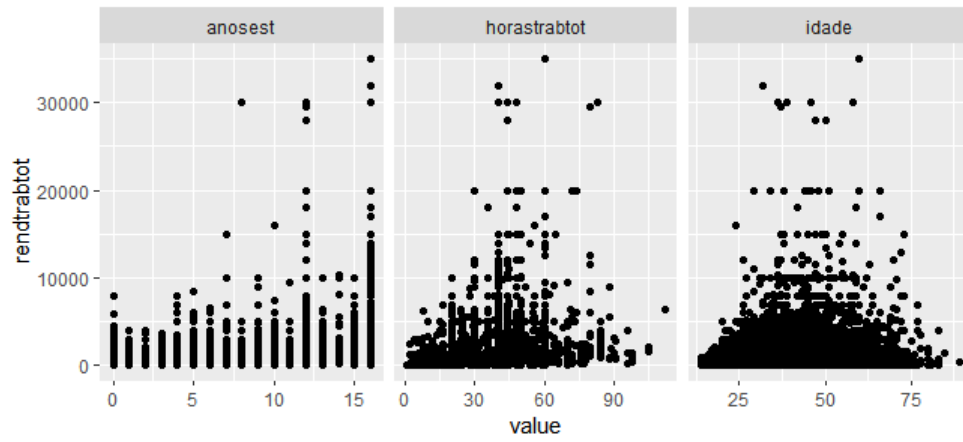
### MA no 2T2012



SP no 2T2022



MA no 2T2022



## 2. Métodos não Paramétricos

### 2.1 Introdução

Em linhas gerais, pode-se considerar a estimação não paramétrica como uma extensão de problemas de estimadores com parâmetros. Nesse sentido, se tomarmos uma sequência de variáveis aleatórias independentes e identicamente distribuídas (*i.i.d*) qualquer  $(X_n)$ , em que sua distribuição é dada por uma função de densidade de probabilidade  $F(., \theta)$ , temos que sua estimação é dependente do parâmetro  $\theta$ . Por outro lado, quando a mesma sequência depende somente de uma função  $F(.)$  não teremos conhecimento da sua distribuição. No entanto, quando  $X_n$  é *i.i.d*, seja ele contínuo ou discreto, pode existir uma função  $g(F) \in R$  não viesada para estimação do parâmetro, tal que para um grau  $m \geq 1$  do parâmetro que queremos estimar de  $g(.)$  teremos o menor  $n$  possível para um estimador não viciado de  $X_n$ . Isto é, estimamos a distribuição por meio de uma função baseada no tamanho mínimo da amostra, tal método é conhecido como Kernel.

A função de Kernel pode ser descrita como:

$$\hat{f}(x) = 1/Nh_n \sum_{i=1}^n K((x - X_i)/h_n) \quad (1)$$

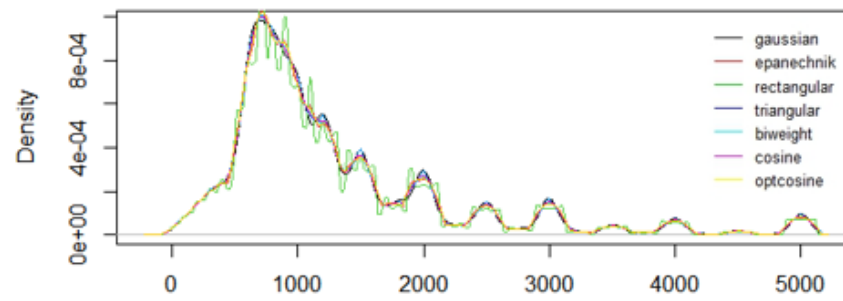
A partir de (1) temos que  $K(.)$  é a função de Kernel, em que assumimos ela como não negativa, simétrica em torno de zero e sua integral tende a 1, dado que ela representa uma função de densidade de probabilidade. Ademais, temos que  $h_n$  é a largura da banda a ser escolhida de maneira arbitrária, de tal modo a suavizar a estimação. Finalmente, a estimação da densidade será feita via avaliação em torno de  $f(x_0)$  frente ao comprimento de  $X_n$ .

### 2.2 Aplicação

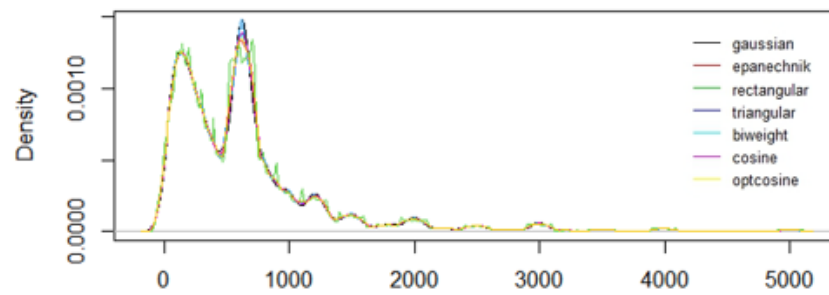
Feito o corte na amostra da renda como já comentado, para ambos os períodos foi adotado algumas estimações via função de Kernel. Logo, para o período do 2T12, o MA acusou, para todas as métricas adotadas, uma alta concentração de rendimentos em torno do salário mínimo da época (R\$ 622,00) e quantis tímidos para rendimentos maiores. Ao mesmo tempo, SP registrou uma concentração mais alta perto de rendimentos ligeiramente maiores que o salário mínimo, porém com quantis mais robustos ao longo de rendimentos mais elevados.

Para o período do 2T2022, MA apresentou uma dinâmica similar ao período passado, ou seja, em torno do salário mínimo (R\$ 1212,00), contudo, houve uma melhora para rendimentos mais elevados com quantis mais robustos. SP, por sua vez, registrou comportamento semelhante ao do MA, concentração alta próxima do salário mínimo e uma melhora para quantis de renda mais elevada.

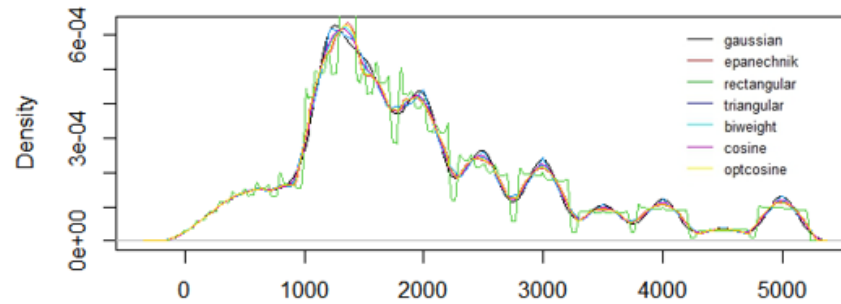
**Kernels da Renda em SP no 2T12**



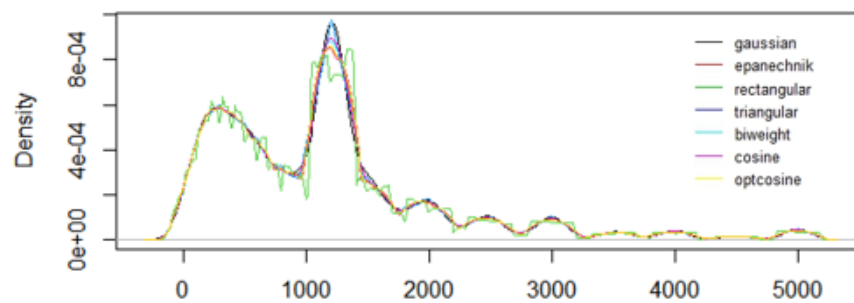
**Kernels da Renda no MA no 2T12**



**Kernels da Renda em SP no 2T22**



**Kernels da Renda no MA no 2T22**



### 3. Regressão Quantílica

#### 3.1 Introdução

O método de Regressões Quantílicas (RQ) se mostra mais eficiente do que o Mínimo Quadrados Ordinários (MQO) quando buscamos mais informações de localização dentro de uma distribuição, seja na mediana, nas caudas ou em qualquer quantil desejado. Em outras palavras, ao passo que o MQO produz estimativas de uma esperança condicional de uma variável, ou seja, uma média, a RQ, por sua vez, é capaz de realizar o mesmo método, porém em determinados pontos da distribuição. Desta maneira, a RQ tende a produzir estimativas mais precisas ao longo de toda distribuição de uma variável à ponto de demonstrar as nuances das variáveis explicativas na variável resposta.

Ademais, quando se trata de RQ, o método leva algumas vantagens sobre o MQO. Dado que a RQ, segundo Koenker & Bassett, 1978, é uma soma ponderada dos desvios absolutos da média, temos que a regressão lida melhor com a presença de *outliers* e com a falta de gaussianidade dos erros. Logo, em grande medida, os estimadores da RQ tendem a serem mais eficientes e robustos quando comparados aos do MQO.

#### 3.2 Aplicação

Primeiramente, vale destacar que todos os parâmetros de cada modelo passaram no teste  $f$  e, respectivamente, no teste  $t$  ao nível de significância de 5%. Os modelos, por sua vez, representam estimações para os quantis 10%, 25%, 50%, 75%, 90% e, por fim, uma regressão pelo MQO.

Para o 2T2012, os coeficientes de SP demonstram um impacto maior das variáveis explicativas sobre a renda em comparação aos betas do MA. Ademais, cada variável registrou um efeito maior na medida que os quantis crescem, por exemplo, o impacto da idade para o quantil que representa os 50% mais ricos é maior do que para os quantis 10% e 25%, de maneira generalizada. Além disso, os anos de estudos se demonstraram a variável de maior impacto sobre a renda, sendo essa a variável explicativa de maior dispersão entre os quantis. Por fim, ficou evidente que os parâmetros estimados via MQO ficaram próximos da média dos coeficientes estimados para os quantis.

Para o período do 2T2022, as estimações apresentaram dinâmica semelhante ao período passado. Dentre as diferenças, cabe ressaltar que para ambas as UFs os coeficientes demonstraram avanços frente aos estimados para 2T2012. Para tal, o comportamento apresentando é favorecido pelo aumento da renda no período. Tal fato fica mais evidente quando olhamos para os Kerneis estimados na seção passada quando para o mesmo período de referências os quantis de renda mais alta ficaram mais robustos.

| Parâmetros de SP no 2T12 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | - 264.77   | 10.14        | 2.75  | 41.33   |
| Quantil 25%              | -373.75    | 11.68        | 6.08  | 41.33   |
| Quantil 50%              | -819.39    | 14.76        | 14.17 | 71.80   |
| Quantil 75%              | -1440.24   | 19.77        | 26.87 | 116.51  |
| Quantil 90%              | -1977.36   | 25.12        | 41.60 | 168.44  |
| MQO                      | -1281.84   | 15.77        | 20.44 | 112.24  |

| Parâmetros do MA no 2T12 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | -214.96    | 3.62         | 2.45  | 22.21   |
| Quantil 25%              | -364.35    | 6.15         | 4.56  | 40.55   |
| Quantil 50%              | -345.56    | 7.86         | 5.76  | 45.87   |
| Quantil 75%              | -525.31    | 12.62        | 10.35 | 60.56   |
| Quantil 90%              | -844.48    | 18.90        | 17.80 | 94.99   |
| MQO                      | -706.52    | 10.60        | 11.60 | 69.46   |



Figura 1: SP no 2T12

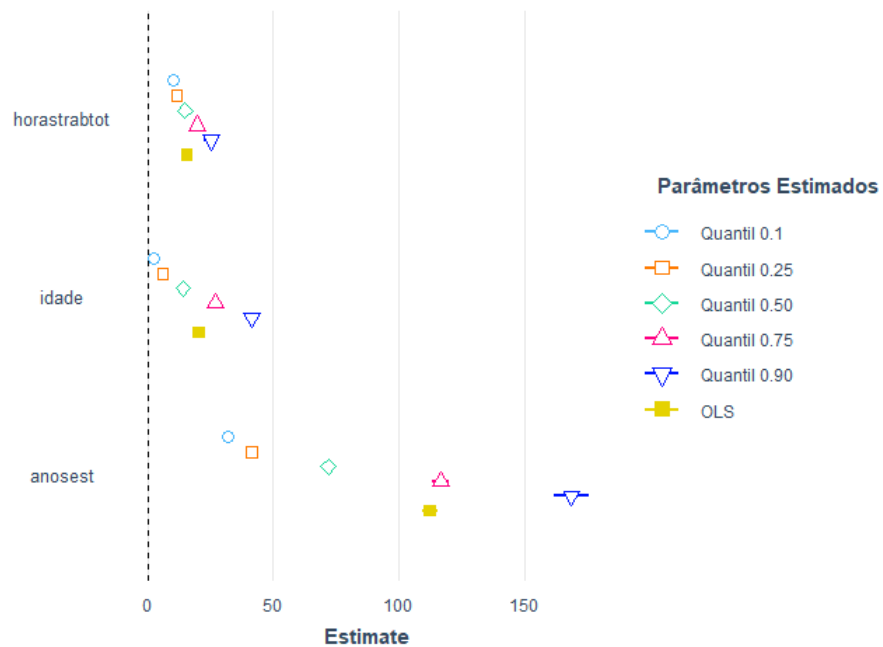
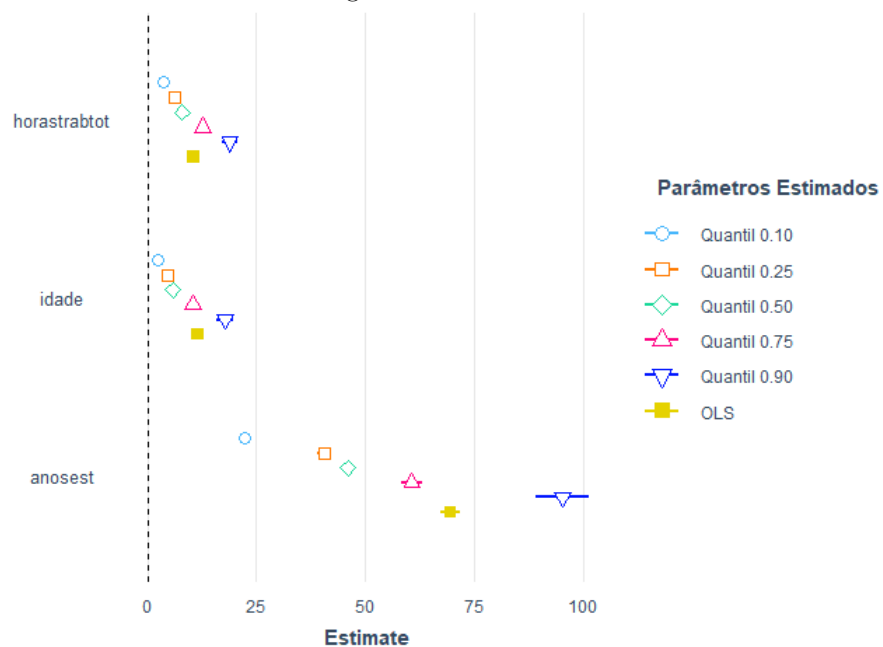


Figura 2: MA no 2T12



| Parâmetros de SP no 2T22 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | -549.34    | 21.52        | 3.81  | 48.13   |
| Quantil 25%              | -654.80    | 24.13        | 8.15  | 59.50   |
| Quantil 50%              | -1198.02   | 28.53        | 17.92 | 99.32   |
| Quantil 75%              | -1734.02   | 33.18        | 29.49 | 151.16  |
| Quantil 90%              | -1629.30   | 34.15        | 36.18 | 191.18  |
| MQO                      | -1139.90   | 26.14        | 18.17 | 120.19  |

| Parâmetros do MA no 2T22 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | -459.60    | 9.03         | 4.41  | 33.70   |
| Quantil 25%              | -889.13    | 16.79        | 9.51  | 64.16   |
| Quantil 50%              | -799.23    | 20.55        | 10.52 | 70.64   |
| Quantil 75%              | -1183.60   | 24.90        | 22.44 | 95.37   |
| Quantil 90%              | -1669.44   | 32.42        | 36.26 | 137.90  |
| MQO                      | -1146.04   | 20.30        | 18.02 | 95.91   |

Figura 3: SP no 2T22

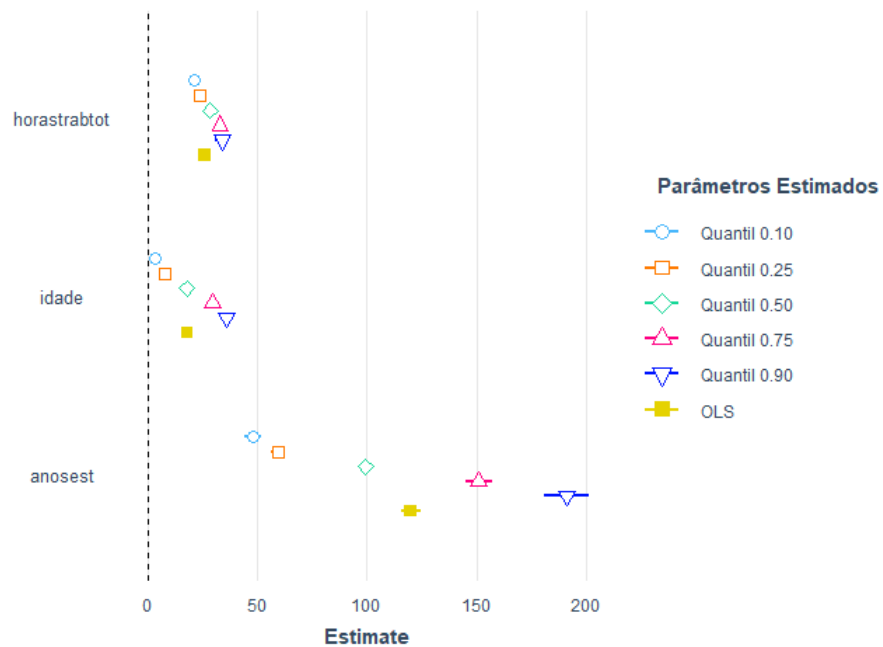
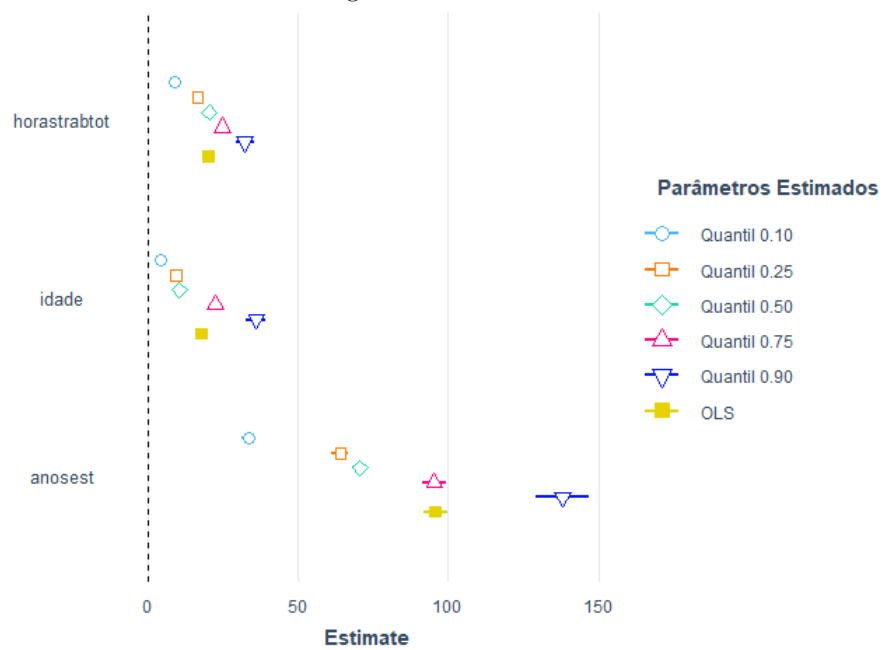


Figura 4: MA no 2T22



## 4. Decomposição Diferencial

### 4.1 Oaxaca-Blinder

#### 4.1.1 Introdução

A partir dos trabalhos de Oaxaca (1973) e Blinder (1973) a metodologia de decomposição salarial entre grupos ganha novos contornos dentro da literatura, sendo que existem alguns tipos de abordagens do método, os mais conhecidos são o *twofold*, que será o foco do trabalho, e *threefold*. De modo geral, a métrica *twofold* é capaz de segmentar efeitos diferenciais observáveis e não observáveis. A primeira fração da diferença é entendida como uma composição do grupo dado pelas diferentes distribuições dos mesmos. Pelo lado não explicado, entende-se que tal diferença reflete o efeito estrutural dos grupos, ou seja, as características exclusivas de cada um. Isto é, pelo lado observável, temos que a diferença entre os grupos será dada pelo hiato entre as matrizes das variáveis explicativas da regressão, ao passo que a parte não observável será dada pelo hiato entre os parâmetros estimados para cada grupo.

#### 4.1.2 Aplicação

A decomposição de Oaxaca-Blinder do presente trabalho considerou os dois períodos já analisados e categorizou por meio de uma variável dummy as UFs em questão, sendo que para SP (Grupo A) foi atribuído o valor 0 e para MA (grupo B) o valor 1. As tabelas das regressões vão indicar na coluna *group.weight* os pesos relativos dos coeficientes de uma regressão em observações dos Grupos A e B, respectivamente, no vetor de coeficiente de referência. Os dois pesos negativos indicam que os coeficientes de referência são provenientes de regressões agrupadas sem (-1) ou com (-2) da variável indicadora de grupo incluída como covariável.

Para o primeiro período, podemos interpretar, a partir da soma da linha 2 dos coeficientes explicados e não explicados, que existe um diferencial de renda de R\$ 976,44, em que, dado os valores da linha 5 das mesmas colunas, R\$ 643,00 pode ser explicado pelas diferenças entre as distribuições e R\$ 333,43 é um efeito não observável. Quando olhamos para o gráfico notamos que os maiores condicionantes para tal diferença se dão pelos anos de estudos, na parte explicada e não explicada, e para idade, na parte não explicada.

Para o segundo período, temos que os resultados acusaram um diferencial de R\$ 1666,22, em que R\$ 1027,34 são frutos da parte explicada e R\$ 638,88 da parte não explicada. A partir do gráfico fica evidente que, para parte explicada, houve um ganho de significância para todas as variáveis, principalmente para horas de trabalho. Enquanto a parte não explicada também apresentou ganhos de significância, porém a escolaridade predomina o impacto no diferencial de rendimentos em contraste ao período anterior.

|      | group.weight      | coef(explained)     | se(explained)     | coef(unexplained)   |
|------|-------------------|---------------------|-------------------|---------------------|
| [1,] | 0.0000000         | 307.0150            | 9.548345          | 669.4333            |
| [2,] | 1.0000000         | 814.3909            | 23.745400         | 162.0574            |
| [3,] | 0.5000000         | 560.7029            | 13.518537         | 415.7454            |
| [4,] | 0.6477311         | 635.6581            | 11.205324         | 340.7902            |
| [5,] | -1.0000000        | 643.0086            | 15.424667         | 333.4397            |
| [6,] | -2.0000000        | 597.9926            | 14.859060         | 378.4557            |
|      | se(unexplained)   | coef(unexplained A) | se(unexplained A) | coef(unexplained B) |
| [1,] | 23.04250          | 6.694333e+02        | 2.304250e+01      | 0.00000             |
| [2,] | 16.08176          | 0.000000e+00        | 0.000000e+00      | 162.05741           |
| [3,] | 15.81244          | 3.347166e+02        | 1.152125e+01      | 81.02871            |
| [4,] | 17.40409          | 2.358205e+02        | 1.492534e+01      | 104.96962           |
| [5,] | 13.37304          | 1.174604e+02        | 4.775015e+00      | 215.97925           |
| [6,] | 14.84906          | 3.017249e-10        | 4.050586e-10      | 378.45573           |
|      | se(unexplained B) |                     |                   |                     |
| [1,] | 0.000000          |                     |                   |                     |
| [2,] | 16.081756         |                     |                   |                     |
| [3,] | 8.040878          |                     |                   |                     |
| [4,] | 5.665103          |                     |                   |                     |
| [5,] | 8.716331          |                     |                   |                     |
| [6,] | 14.849062         |                     |                   |                     |

Figura 5: Decomposição no 2T212

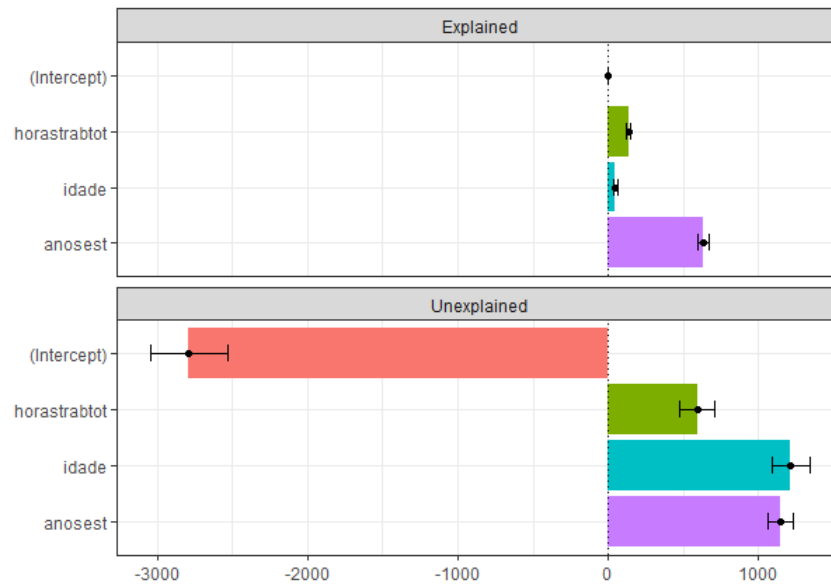
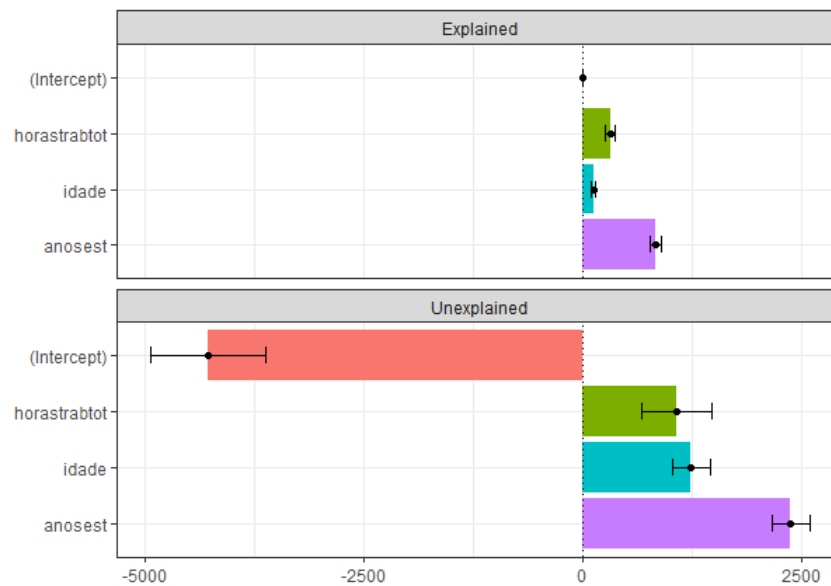


Figura 6: Decomposição no 2T22



## 4.2 Recentered Influence Function (RIF)

### 4.2.1 Introdução

Primeiramente, cabe comentar acerca das funções de influência (IF) que, basicamente, são uma ferramenta para analisar as distribuições de variáveis e verificar a robustez de estimadores. Ademais, a IF é capaz de informar o efeito de uma mudança em uma observação em um estimador e ajudar a calcular matrizes de variância-covariância para certos tipos de estimadores, especialmente quando os métodos mais diretos se tornam difíceis de implementar.

Posteriormente, com o trabalho de Firpo, Fortin & Lemieux (2009), introduziu-se o conceito de RIF que incorpora a análise do impacto que mudanças na distribuição de variáveis explicativas tem na distribuição incondicional de Y, ou seja, um impacto marginal. Em outras palavras, o método RIF visa, por meio de uma regressão quantílica incondicional, obter efeitos parciais de variáveis explicativas em qualquer quantil incondicional da variável dependente. Logo, os autores sugerem uma abordagem simples, uma vez que a estratégia para a estimação da regressão quantílica incondicional se baseia na utilização de modelos lineares como o método mais fácil para aproximar esses efeitos parciais.

### 4.2.2 Aplicação

Para o presente trabalho, utilizou-se da regressão RIF com quantis. Em suma, o método funciona fornecendo uma linearidade de aproximação a um funcional não linear da distribuição, de tal forma que a linearização permite aplicar a lei das expectativas iteradas às estatísticas distributivas de interesse e, assim, calcular efeitos parciais aproximados de mudanças na distribuição de cada covariável no funcional de interesse. Sendo assim, as estimações foram feitas para SP e MA nos dois períodos de referência a partir da função de distribuição acumulada feita estimada por meio do método Kernel gaussiano com bootstrap igual a 100 e para os quantis 10%, 25%, 50%, 75% e 90%.

Pelo lado dos resultados, a interpretação é de que a cada 1 unidade adicional de uma variável explicativa teremos um impacto marginal médio na renda. Então, por exemplo, em SP no 2T12 para o quantil 10%, a cada um ano a mais de estudos devemos ter um acréscimo médio de R\$ 321,18 na renda. Ademais, cabe ressaltar que todas as variáveis dos modelos de SP e MA para o primeiro período acusaram ser relevantes para os níveis de significância de 1%, 5% e 10%, à exceção da idade para SP no quantil 10% que não registrou significância alguma e para mesma variável no MA os resultados revelaram que ela foi significativa somente para os níveis de 1% e 10%. No que tange as diferenças entre as UFs, é interessante notar que SP registrou um impacto marginal das variáveis explicativas superior aos do MA quase que de forma generalizada, somente para o cálculo feito no quantil 90% para horas trabalhadas que o MA superou o rendimento adicional paulista. No entanto, para o mesmo quantil, SP apresentou uma diminuição do acréscimo marginal, ao passo que o MA manteve a dinâmica crescimento.

| Parâmetros de SP no 2T12 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | -101.44    | 11.11        | 0.42  | 21.18   |
| Quantil 25%              | -182.19    | 10.51        | 3.23  | 34.04   |
| Quantil 50%              | -728.51    | 12.65        | 13.48 | 71.89   |
| Quantil 75%              | -5604.32   | 33.93        | 64.16 | 331.39  |
| Quantil 90%              | -2259.55   | 22.47        | 50.28 | 246.26  |

| Parâmetros do MA no 2T12 |            |              |       |         |
|--------------------------|------------|--------------|-------|---------|
| Modelo                   | Intercepto | horastrabtot | idade | anosest |
| Quantil 10%              | -121.90    | 3.93         | 0.51  | 9.70    |
| Quantil 25%              | -270.99    | 6.92         | 1.38  | 25.93   |
| Quantil 50%              | -61.93     | 6.37         | 3.83  | 37.62   |
| Quantil 75%              | -691.38    | 10.93        | 13.57 | 77.45   |
| Quantil 90%              | -2960.96   | 24.26        | 46.39 | 215.03  |

Para o 2T22, todas as variáveis apresentaram significância nos níveis de 1%, 5% e 10%, à exceção do intercepto do MA estimado na mediana. No tocante às diferenças entre os Estados, SP registrou efeitos marginais mais elevados quase que para todas as estimações, porém para o quantil 90% o MA se manteve a frente dos paulistas menos para variável de anos de estudos. Por fim, quando

comparamos os efeitos entre os períodos, cada variável registrou um avanço no 2T22 frente ao 2T2012, menos para SP no quantil 75% que registrou uma queda.

| Parâmetros de SP no 2T22 |            |              |        |         |
|--------------------------|------------|--------------|--------|---------|
| Modelo                   | Intercepto | horastrabtot | idade  | anosest |
| Quantil 10%              | -1241.15   | 34.46        | 4.86   | 54.30   |
| Quantil 25%              | -696.95    | 21.03        | 7.84   | 65.81   |
| Quantil 50%              | -1922.10   | 25.95        | 25.47  | 156.21  |
| Quantil 75%              | -3629.56   | 34.16        | 47.53  | 303.38  |
| Quantil 90%              | -8593.21   | 63.13        | 103.07 | 671.32  |

| Parâmetros do MA no 2T22 |            |              |        |         |
|--------------------------|------------|--------------|--------|---------|
| Modelo                   | Intercepto | horastrabtot | idade  | anosest |
| Quantil 10%              | -415.41    | 10.67        | 1.39   | 23.06   |
| Quantil 25%              | -1186.48   | 23.24        | 7.00   | 63.25   |
| Quantil 50%              | 40.46      | 11.64        | 7.35   | 49.36   |
| Quantil 75%              | -1686.42   | 23.70        | 30.89  | 127.19  |
| Quantil 90%              | -10363.88  | 66.75        | 140.00 | 561.38  |

## 5. Referências

- Blinder, Alan S. "Wage discrimination: reduced form and structural estimates." *Journal of Human resources* (1973): 436-455.
- FIRPO, Sergio; FORTIN, Nicole M.; LEMIEUX, Thomas. Unconditional quantile regressions. *Econometrica*, v. 77, n. 3, p. 953-973, 2009.
- Friedman, Milton. "The permanent income hypothesis." *A theory of the consumption function*. Princeton University Press, 1957. 20-37.
- Koenker, Roger, and Gilbert Bassett Jr. "Regression quantiles." *Econometrica: journal of the Econometric Society* (1978): 33-50.
- Koenker, Roger, and Kevin F. Hallock. "Quantile regression." *Journal of economic perspectives* 15.4 (2001): 143-156.
- Modigliani, Franco. "Life cycle, individual thrift, and the wealth of nations." *Science* 234.4777 (1986): 704-712.
- Oaxaca, Ronald. "Male-female wage differentials in urban labor markets." *International economic review* (1973): 693-709.