

# DURAÇÃO DO DESEMPREGO NO BRASIL METROPOLITANO: UMA ANÁLISE ATRAVÉS DE REGRESSÃO QUANTÍLICA CENSURADA

Ana Katarina Campêlo (UFPE)  
Wilton Bernardino da Silva (UFPE)  
Rafael Moura Azevedo (UFPE)

## Resumo

Conforme enfatizado por Chaudhuri, Doksum e Samarov (1997), o vetor  $\beta(\tau)$  de regressão quantílica é “um conceito unificador que representa os vetores de coeficientes nos modelos linear, de Cox, de razões proporcionais, de tempo de falha acelerado e assim por diante”. Ademais, considerar-se-á as vantagens até o presente da técnica de regressão quantílica através da estimação destes modelos para fins comparativos com o efeito quantílico de tratamento derivado do modelo de riscos proporcionais de Cox. Estas modelagens são ilustradas numa primeira instância neste trabalho através de aplicação empírica a dados de duração de desemprego no Brasil, fornecendo evidências atuais e no âmbito nacional e possibilitando um estudo mais detalhado e amplo da relação entre as variáveis do modelo, na medida em que esta abordagem permite que as covariáveis impactem de forma bastante flexível no que se refere não só a locação, mas a escala e forma da distribuição condicional. Os resultados econométricos revelam que o fator idade é irrelevante para a duração do desemprego, enquanto que esta é menor em todos os quantis para pessoas que já trabalharam antes, bem como para os indivíduos que possuem qualquer diploma, destacadamente aqueles que têm pós-graduação. Esta última conclusão difere de trabalhos anteriores, que sugerem um efeito médio positivo para anos de educação. Além disso, o tempo na situação de desemprego tende a ser maior em todos os setores de atividades considerados, relativamente a outros de menor representatividade no mercado de trabalho e serviços domésticos. Em adição, outro resultado interessante é o fato de a duração tender a ser mais curta para os indivíduos que residem nas regiões metropolitanas de Recife, Belo Horizonte e Porto Alegre, em comparação com aqueles que moram em Salvador, Rio de Janeiro e São Paulo. Finalmente, é também digno de nota que as estimativas dos efeitos do indicador de chefe de família e de outras *dummies* relativas ao emprego anterior, como as que indicam se a pessoa foi demitida ou trabalhava no setor formal, além de anos de experiência, são estatisticamente não significativos ao longo de todos os quantis da distribuição condicional da duração do desemprego, de forma distinta ao encontrado em estudos prévios.

**Palavras-chave:** duração do desemprego no Brasil, regressão quantílica pra análise de sobrevivência, efeito quantílico de tratamento no modelo de Cox.

**Códigos de classificação do JEL:** J64, R23, C41, C24, C14, C13.

## Abstract

Chaudhuri, Doksum and Samarov (1997) emphasizes that the quantile regression vector  $\beta(\tau)$  is “a unifying concept that represents the coefficient vectors in the linear, Cox, proportional odds, accelerated failure time model and so on”. In addition, the advantages of the quantile regression technique is considered by estimating these models for comparative analysis with the quantile treatment effect derived from the Cox proportional hazards model. The first are illustrated for the first time nationally in this work throughout an empirical application to unemployment duration data in Brazil, providing a more recent evidence on the topic and allowing for a more detailed and ample study of the relationship between the variables in the model. This approach allows the covariates to impact not only in terms of the location, but

also the scale and form of the conditional distribution. The results reveal that the age factor is irrelevant to the unemployment duration, along with this duration is lower across all quantiles for people who have previously worked, as well as for individuals who hold any degree, notably those who have postgraduate degrees. This last conclusion differs from earlier works, which suggest an average positive effect for years of education. Furthermore, unemployment spells tend to be longer in all sectors of activities considered, relative to those less representative ones in the labor market and domestic services. Finally, another very interesting result is that the duration tends to be shorter for individuals residing in the metropolitan areas of Recife, Belo Horizonte and Porto Alegre, compared to those living in Salvador, Rio de Janeiro and São Paulo. This latter evidence also differs from previous results reported in the literature.

**Keywords:** unemployment duration in Brazil, quantile regression for survival analysis, quantile treatment effect in the Cox model.

**JEL Code:** J64, R1, C41, C24, C14, C13.

Os autores agradecem os comentários de Sergio Firpo (INSPER-SP) e Roger Koenker (UIUC, EUA) e financiamento do INTC/CNPq.

## 1 INTRODUÇÃO

Pesquisadores e especialistas concordam com o fato de que um dos desafios centrais para a recuperação das economias mundiais reside no desemprego, em especial quando a duração deste se prolonga. Este indicador não sinaliza apenas um mercado de trabalho frágil, mas outro cenário igualmente importante: a realidade de que o crescimento econômico está bem aquém de seu potencial. No Brasil, a ocorrência de elevadas taxas de desemprego nos últimos anos e em décadas passadas tem sido apontada como uma preocupação primária para economistas e gestores de políticas públicas. Esta estatística reflete a incidência e também a duração do desemprego, necessitando ser complementada por esta última. Assim, existe tanto o cenário no qual o aumento da taxa de desemprego se dá pela elevação da incidência com duração constante, bem como a situação em que há uma extensão da duração e manutenção da incidência, ou ainda crescimento de ambas. Ademais, é digno de nota que o bem-estar do indivíduo desempregado encontra-se mais fortemente relacionado ao tempo passado sem emprego do que com o desemprego em si (Kiefer, 1988). Neste sentido, a taxa de desemprego seria de menor importância relativa que a média da duração do desemprego. Por conseguinte, tal variável implica em sérias questões que são recorrentes em debates econômicos, nos quais se discute sobre tomadas de decisão com fins de estabelecer soluções eficazes no sentido de promover a empregabilidade destas pessoas, bem como encontrar meios de suporte para as mesmas enquanto permanecem sem emprego. Para tal, é de crucial importância buscar uma compreensão aprofundada e abrangente da duração do desemprego, variável modelada neste trabalho, e sua relação com fatores socioeconômicos, educacionais, ocupacionais e regionais, dentre outros.

No que concerne a metodologias, conforme sumarizado em Heckman & Vytlacil (2007) e Imbens & Wooldridge (2009), a literatura de avaliação de políticas públicas tem se concentrado primariamente nas situações em que a variável de interesse é completamente observada para os grupos de controle e tratamento. No entanto, quando a variável resposta está sujeita a censura, tais procedimentos de inferência podem apresentar resultados espúrios e conclusões incorretas relativamente aos efeitos de tais políticas. Desta forma, neste trabalho é utilizado o modelo de regressão quantílica censurada, o qual leva em conta observações que apresentam censura à direita, que é comum nas aplicações típicas de análise de duração (ver Powell, 1986; Fitzenberger, 1997; Portnoy, 2003; Peng & Huang, 2008). Esta técnica representa um importante complemento aos métodos tradicionais, como o modelo de riscos proporcionais de Cox e suas generalizações, na medida em que a mesma é uma abordagem unificadora para modelos de transformação de forma mais generalizada<sup>1</sup>. A regressão quantílica censurada mantém todas as propriedades da técnica padrão, como a robustez de um modelo semi-paramétrico (*distribution-free*) que fornece estimadores mais eficientes quando os erros não seguem uma distribuição normal, bem como o fato de os coeficientes estimados não serem sensíveis a observações extremas na variável dependente (dado que a função objetiva é uma soma ponderada de desvios absolutos). Em especial, esta modelagem fornece soluções (ou seja, estimativas para os parâmetros) diferentes para quantis distintos, que podem ser interpretadas como diferenças na resposta da variável dependente em vários pontos da distribuição condicional desta e em função das mudanças nas covariáveis. Assim, as funções quantis condicionais caracterizam toda a distribuição da variável resposta dado um conjunto de regressores. Em particular, o método estima o  $\tau$ -ésimo quantil da variável dependente ( $T$ ), dado que  $X = x$  (regressores), demonstrando a natureza condicional desta modelagem. E na formulação dos modelos de duração, Portnoy (2003) salienta que a

regressão quantílica possibilita que sinal do efeito de uma variável explicativa possa mudar com a magnitude da

1 Modelos nos quais a variável resposta é dada por  $g(T)$ , onde  $T$  denota o tempo de duração ou sobrevivência. A variável resposta (ao longo dos quantis), bem como permite que o efeito destas seja significativamente positivo em determinados casos (de maior severidade, por exemplo e que corresponderia à cauda direita) e não em outros. Além disso, como enfatiza Koenker & Gelig (2000), as estimativas por regressão quantílica são consideravelmente mais flexíveis que aquelas relativas aos métodos tradicionais em análise de sobrevivência. Nestes últimos tem-se que os efeitos quantílicos de tratamento para as várias covariáveis são forçosamente proporcionais uns aos outros. Isto implica ainda que as funções de sobrevivência não podem se cruzar, enquanto isto é possível no modelo de regressão quantílica (*cross-over effect*). Finalmente, a análise empírica neste trabalho aplica o estimador do processo dos quantis de regressão ( $\beta(\tau)$ ) proposto por Portnoy (2003). Este estimador do quantil condicional para dados censurados é uma generalização do estimador de Kaplan-Meier para o contexto de modelos de regressão. E, para fins comparativos, também são estimados os efeitos quantílicos de tratamento para o modelo de Cox.

Em He (2017), o precursor da regressão quantílica (ver Koenker & Bassett, 1978), Roger Koenker, salienta com propriedade que as mais significantes questões de políticas públicas em Economia requerem alguma forma de análise empírica: estas não podem ser resolvidas apenas por raciocínio teórico. Na aplicação deste artigo, o banco de dados provém da pesquisa mensal do emprego (PME), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2015<sup>2</sup>. A duração do desemprego (em meses) é obtida ao acompanhar a situação do indivíduo relativamente ao desemprego no período em que o mesmo se encontra nos quatro meses da pesquisa, que compreendem o período amostral. E o período total do desemprego é obtido adicionando ainda a contagem do tempo durante o qual permaneceu desempregado nos últimos doze meses, anteriormente à sua entrada no primeiro mês da pesquisa. Este acompanhamento dos indivíduos no ano de 2015 é realizado através da construção de painéis, os quais seguiram as diretrizes contidas em Ribas & Soares (2008). Finalmente, os resultados apontam para vantagens de saída do desemprego para o emprego se a pessoa já trabalhou antes e para quem possui qualquer diploma educacional, notadamente o de pós-graduação. Esta inclinação de retorno ao emprego também é evidenciada nas regiões metropolitanas de Recife, Belo Horizonte e Porto Alegre, nesta ordem, em comparação às demais regiões (Salvador, Rio de Janeiro e São Paulo). Por sua vez, as estimativas também apontam para desvantagem em direção a um aumento da duração do desemprego em todos os setores de atividades incluídos como controles, relativamente a atividades menos representativas no mercado de trabalho e serviços domésticos.

## 2 ESTUDOS EMPÍRICOS NA LITERATURA INTERNACIONAL E NACIONAL

Nos trabalhos pioneiros na área, os métodos para a análise da duração do desemprego resultam de formas reduzidas de teorias de procura por trabalho (*job search*). As referidas teorias advogam um efeito positivo do benefício do seguro desemprego tanto na duração esperada do desemprego quanto no salário esperado pós-desemprego. Os resultados em Ehrenberg e Oaxaca (1976) indicam que um aumento no benefício do seguro desemprego implica em uma elevação de maior magnitude na duração do desemprego e no salário pós-desemprego para homens mais velhos relativamente às mulheres. Em contraste, para homens e mulheres mais novas, o efeito é o de apenas aumentar a duração do desemprego e não se encontra impacto algum no salário pós-desemprego para este grupo populacional. Em

estudo subsequente, Katz (1986) leva em conta a perspectiva de o indivíduo voltar para o antigo

2 Esta pesquisa foi descontinuada em fevereiro de 2016. Então, o último ano da mesma é 2015, período este que é considerado neste estudo.

emprego como um caminho alternativo para a saída do desemprego e propõe um modelo de procura de trabalho que resulta naturalmente numa especificação de riscos competitivos. Este autor conclui que a função de risco de retorno é crescente com a estabilidade no emprego anterior (*job tenure*) e a função de risco de um novo emprego, por sua vez, é decrescente em função desta variável. Ademais, salienta que a teoria proposta serve ainda para a avaliação de situações de escolha entre o trabalhador permanecer em sua área de atuação na perspectiva de melhoras nas condições de mercado relativamente a arcar com o custo de busca em ocupações alternativas. Por sua vez, Ham e LaLonde (1991) utilizam dados de um experimento social para estimarem o impacto de um treinamento na duração do emprego como também dos períodos de desemprego para os beneficiários do *Aid to families with dependent children (AFDC)*. No que tange os resultados, estes sugerem efeitos iniciais potencialmente positivos de tal programa, que persistiriam e até mesmo viriam a ser mais significativos ao longo do tempo. Em contrapartida, o treinamento não acarretou em períodos mais curtos de desemprego. O trabalho aponta que o programa de emprego apoiado (*supported work*) elevou as taxas de emprego porque este auxiliou mulheres que encontraram trabalho a permanecerem empregadas por um período superior ao que teriam sem tal suporte. E que tal resultado seria encorajador, dado que quanto mais longo o período no emprego, maior a acumulação de capital humano. Outra questão na literatura que é verificada no trabalho de Abraham e Shimer (2001) é o aumento da duração do desemprego nos EUA relativamente à taxa de desemprego, em especial a de curto-prazo, que se manteve num patamar baixo durante o período de 3 décadas anterior ao estudo. Neste mesmo ano, o artigo de Koenker & Biliias (2001) ilustram a abordagem de regressão quantílica pra análise de sobrevivência proposta em Koenker & Gelig (2001) ao reanalisar os dados americanos de experimentos conduzidos no período de 1988 a 1989 no Estado da Pensilvânia (*Pennsylvania Reemployment Bonus Experiments*), os quais foram estabelecidos no intuito de se testar a eficácia de pagamentos em dinheiro para aqueles que se reempregassem de forma mais imediata, visando encurtar a duração da permanência no seguro-desemprego. Já o estudo mais recente de Kroft *et al.* (2014) foca o período da Grande Recessão nos EUA, no qual tanto a taxa do desemprego no curto prazo quanto a de longo prazo elevam-se de forma aguda nos anos de 2008 e 2009. No entanto, enquanto a primeira retorna a seu patamar no ano de 2013, o desemprego de longo prazo (DLP) permanece e segue em nível elevado historicamente.

A literatura nacional na área é ainda incipiente e um dos primeiros artigos que aborda o tema de duração do desemprego é o de Bivar (1993). Neste estudo a autora estima a duração esperada das ocorrências completas do desemprego para a região metropolitana de São Paulo com os dados da PME no período de 1983 a 1990. Os resultados indicam que a duração média estimada pelo modelo markoviano (1,63 mês) é inferior à transcorrida (3,43 meses), enquanto que o modelo semi-markoviano aponta para um valor três vezes superior (6,20) meses, sugerindo que o desemprego no Brasil tem curta duração. Em artigo subsequente, que mantém igualmente a base de dados da PME e foca novamente na região metropolitana de São Paulo, Menezes-Filho & Picchetti (2000) salienta que em Bivar (1993) não houve o controle por características dos desempregados nos modelos utilizados e que a incorporação destas seria uma importante contribuição do trabalho dos mesmos. Estes autores colocam que, se o grupo de desempregados for heterogêneo, uma queda na probabilidade de saída do desemprego poderia ser atribuída à sua duração de forma indevida, quando na realidade a variação nestas probabilidades se deve às características individuais. Através do

uso do modelo Cox associado à distribuição lognormal, este estudo aponta para as seguintes conclusões: a duração esperada é menor para os chefes de família e aqueles que já tiveram algum trabalho anterior, enquanto que é maior para os indivíduos de nível educacional mais elevado, bem como para os mais idosos e para as pessoas que, no emprego anterior, foram demitidas, trabalhavam por mais tempo e no setor formal. Ademais, o estimador de Kaplan-Meier sinaliza que a duração média do desemprego é de 6,64 meses. Por fim, a probabilidade condicional de encontrar um novo emprego é crescente do primeiro até o sexto mês de desemprego e decai a partir deste ponto. Penido e Machado (2002) e Avelino (2000) ampliam o artigo de Menezes-Filho & Picchetti (2000) para as seis regiões metropolitanas pesquisadas pela PME e para o período de 1984-1997, respectivamente, e encontram resultados semelhantes. E mais recentemente, na mesma linha, Cacciamali *et al.* (2015) atualizam a análise para o período de 2002 a 2010, porém focalizando nos subperíodos das crises econômicas de 1992, 1998 e 2009. Estes últimos autores destacam que, dentre os resultados novos, verifica-se que a pessoa desempregada apresenta maior chance de sair dessa condição em 2009 do que nos demais anos, entretanto, o indivíduo que não consegue uma nova ocupação nos primeiros meses de busca em 2009 permanece mais tempo desempregado. Adicionalmente, observa-se que variáveis como escolaridade, posição na família e experiência apresentam relevância nas diversas crises, enquanto para as demais não foi possível identificar padrão comum.

### 3 METODOLOGIA

Os modelos de duração têm recebido bastante atenção nas últimas décadas por parte de pesquisadores em estatística, econometria e áreas afins. Chaudhuri, Doksum e Samarov (1997) ressaltaram a vantagem da formulação da regressão quantílica (Koenker & Bassett, 1978) para a análise de sobrevivência, argumentando que a mesma provê uma abordagem unificadora para modelos de transformação de forma mais generalizada. Segundo estes autores, o vetor dos quantis de regressão,  $\beta(\tau)$ , é um conceito unificador que representa os vetores no modelo linear padrão, no modelo de Cox, nos modelos de razões proporcionais, no modelo de tempo de falha acelerado e assim por diante. Neste contexto, Koenker e Gelig (2001) propõem analisar o tempo de sobrevivência simplesmente pela estimação dos quantis de regressão para uma transformação deste, dada por  $\log(T)$ . Isto é possível em decorrência da propriedade de equivariância a transformações monotônicas da função quantil condicional. Esta transformação em  $\log$  possibilita a modelagem por regressão quantílica do modelo usual do tempo de falha acelerado, permitindo passar de um modelo de locação da média condicional para toda a função de sobrevivência. Adicionalmente, estes autores colocam que os métodos de regressão quantílica consistem numa abordagem natural e de grande potencial para análise da variabilidade estatística causada pela heterocedasticidade presente nos dados e não homogeneidade na população. Por sua vez, Powell (1986) realizou uma importante contribuição ao estender o método de regressão quantílica para modelos de regressão cuja variável dependente apresenta observações censuradas. Neste artigo seminal, a abordagem proposta pelo autor é equivalente ao modelo linear para a variável resposta, mas assumindo que as observações não podem ser observadas acima de um valor fixo  $C$ . Esta abordagem foi estendida para o caso em que o tempo de censura,  $C_i$ , é observado para cada observação (mesmo aquelas não-censuradas), mas esta não pode ser estendida para o caso mais geral de censura. Em sequência, Portnoy (2003) propõe um estimador para os quantis condicionais para dados censurados que é uma generalização do estimador de Kaplan-Meier. Esta última abordagem representa um avanço substancial na direção de identificar e modelar fontes de heterogeneidade que não podem ser modeladas através dos modelos de riscos proporcionais.

Especificamente, qualquer fonte de heterogeneidade que afeta a variável dependente tende a ser refletida em efeitos quantílicos que são substancialmente distintos ao longo da distribuição condicional desta. Por fim, Fitzenberger & Wilker (2006) sumariza várias abordagens de estimação usando regressão quantílica para dados que apresentam censura à direita.

### 3.1 CONCEITOS PRELIMINARES: FUNÇÃO DE SOBREVIVÊNCIA, TAXA DE RISCO E CENSURA

Os modelos de sobrevivência ou de duração representam importantes formulações em determinadas áreas, como em economia do trabalho. Estes modelos, originários de aplicações biomédicas, seguros e controle de qualidade, estão sendo mais amplamente aplicados ao desemprego, aposentadoria, finanças e a uma gama de outras questões.

No presente trabalho considera-se a densidade condicional do período de saída do desemprego, dado que o indivíduo estava desempregado até o tempo  $t$ . Tomemos uma variável aleatória positiva *ex ante*,  $T$ , com função de densidade  $f(t)$  e função de distribuição  $F(t)$ . A partir destas, seguem uma série de terminologias subjacentes à esta modelagem. A função de sobrevivência é definida por  $S(t) = 1 - F(t) = P(T > t)$ . Dado a sobrevivência até  $t$ , a probabilidade de morte antes de  $t + dt$  é dada por:

$$P(T < t + dt | T > t) = \frac{P(t < T < t + dt)}{P(T > t)} = \frac{F(t+dt) - F(t)}{1 - F(t)}. \quad (1)$$

Assim, para se obter uma taxa de morte (ou de falha), por unidade de tempo, entre  $t$  e  $t + dt$ , toma-se o seguinte limite:

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(T < t + dt | T > t)}{dt} = \lim_{dt \rightarrow 0} \frac{\frac{F(t+dt) - F(t)}{1 - F(t)}}{dt} = \frac{f(t)}{1 - F(t)}, \quad (2)$$

a qual denomina-se de função taxa de risco<sup>3</sup>. Tem-se então que  $\lambda(t)dt = P(t < T < t + dt | T > t)$  é similar à probabilidade condicional de se deixar um estado no intervalo ínfimo de  $(t, t + dt)$ , condicional a este ter durado até o tempo  $t$  (dado por  $S(t)$ ). E, nesta aplicação,  $\lambda(t)dt = P[\text{sair do desemprego no tempo } t + 1 | \text{desempregado no tempo } t]$ . Ademais, é digno de nota que a função de risco não é uma probabilidade, dado que esta refere-se ao tempo  $t$  exato e não ao pequeno intervalo  $dt$  em diante. Em consequência:

$$\int_0^t \lambda(u)du = -\log[1 - F(u)]|_0^t = -\log[1 - F(t)] = -\log[S(t)]. \quad (3)$$

Ou seja,  $\lambda(t) = \partial\{-\log[S(t)]\}/\partial t$ . Além de que:  $S(t) = \exp\left\{-\int_0^t \lambda(s)ds\right\}$ . Estas são denominadas na literatura como integrais produto e dispõem de uma rica teoria<sup>4</sup>.

Finalmente, um problema comum a estes tipos de dados é o fato de observarmos  $T$  para apenas algumas observações, enquanto que para outras observa-se apenas que  $T_i$  é maior que um tempo de censura,  $C_i$ . Por exemplo, no presente estudo, alguns indivíduos podem estar

ainda desempregados ao final do período amostral, então estas observações são ditas censuradas. Então observa-se  $\tilde{T}_i$  tal que:

3 A recíproca da taxa de risco é chamada razão de Mills.

4 Ver Gill & Johanson (1990).

$$\tilde{T}_i = \begin{cases} T_i, & \text{se } T_i < C_i \\ C_i, & \text{se } T_i \geq C_i. \end{cases} \quad (4)$$

### 3.2 MODELOS SEMI-PARAMÉTRICOS

#### 3.2.1 Modelo de Risco Proporcional de Cox e Efeito Quantílico de Tratamento

Os modelos de sobrevivência, mais comumente chamados de modelos de duração em economia, analisam a duração esperada de tempo até o acontecimento de um ou mais eventos. Em particular, estes relacionam o tempo que decorre até a ocorrência de um evento a covariáveis que estejam associadas a este período de tempo. Muitos dos modelos paramétricos e semi-paramétricos de análise de duração, podem ser expressados na forma de alguma transformação monotônica do tempo de duração observado,  $g(T_i)$ , representado como um preditor linear mais um termo de erro independente e identicamente distribuído (i.i.d.):

$$g(T_i) = x_i' \beta + u_i, \quad (5)$$

onde  $x$  representa um vetor de covariáveis, enquanto que  $\beta$  é um vetor de parâmetros desconhecidos e  $u$  é um termo de erro com distribuição  $F$ . Comumente, na literatura de análise de sobrevivência toma-se  $g(T_i)$  igual a  $\log(T_i)$ , resultando no modelo denominado de modelo de tempo de falha acelerado.

No modelo de Cox (Cox, 1972), que é um dos seminais na área, tem-se (retornando a (2))  $\lambda(t|x) = \frac{f(t|x)}{1-F(t|x)}$ . E a hipótese crucial do modelo de Cox é a de que:

$$\lambda(t|x) = e^{-x'\beta} \lambda_0(t), \quad (6)$$

onde  $\lambda_0(t)$  é o risco de linha de base ou função de risco basal. Com a integração desta obtém-se a função de risco basal acumulada  $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ . Note que a forma  $h(x) = e^{-x'\beta}$  implica em separabilidade multiplicativa da função em  $t$  e  $x$ . E se esta hipótese é válida, os coeficientes podem ser estimados sem a necessidade de se determinar  $\lambda_0(t)$ . Em termos de interpretação, num modelo de riscos proporcionais o efeito único do aumento em 1 unidade nas covariáveis é multiplicativo com respeito ao risco de linha de base<sup>5</sup>. Assim, diversas formas de heterogeneidade são excluídas, a menos que diferentes conjuntos da população possam ser associados a funções de risco basal diferenciadas. Especificamente, o modelo de riscos proporcionais não permite uma situação na qual o sinal do efeito de uma variável explicativa possa mudar com a magnitude da variável resposta. Esta modelagem também não possibilita haver um efeito que seja significativamente positivo em determinados casos (de maior severidade, por exemplo) e não em outros. No entanto, o modelo de Cox é uma formulação natural quando há um interesse primário na função taxa de risco. Em particular,



numa situação em que a intensidade da transição é de especial importância para o processo estocástico subjacente através da presença de covariáveis que variam no tempo, por exemplo. Ao integrar (6) tem-se a função de risco acumulada  $\Lambda(t|x) = \Lambda_0(t)e^{-x'\beta}$ . Esta última, usando (3), é equivalente a:

5 Por exemplo, a taxa de falha de um produto pode dobrar ao se mudar um material de um componente manufaturado deste.

$$(7) \quad \log\{-\log[S(t|x)]\} = \log\{-\log[1 - F(t|x)]\} = \log[\Lambda_0(t)] - x'\beta,$$

onde  $S(t|x)$  representa a função de sobrevivência condicional e tem-se assim uma formulação que pode ser expressa na forma de (5). Em particular, quando a função de risco acumulada toma a forma da Weibull, tal que  $\log[\Lambda_0(t)] = \theta \log(t) - \alpha$ , e incorporando-se as constantes no preditor linear, obtém-se especificamente o modelo de tempo de falha acelerado, conforme visto na formulação em (5) igualmente.

A definição a seguir é útil na interpretação do papel exercido pela hipótese de Cox:

*Definição:* uma família de distribuições  $\mathcal{F}$  constitui uma família de  $\mathcal{L}$ -alternativas de Lehmann se existe  $F_0 \in \mathcal{F}$ ,  $1 - F(t) = (1 - F_0)^\gamma$  para algum  $\gamma > 0$  e pra todo  $t$ .

Assim, o modelo de riscos proporcionais de Cox implica uma família de  $\mathcal{L}$ -alternativas por (5):

$$(8) \quad S(t|x) = \exp\left\{-e^{-x'\beta} \int_0^t \lambda_0(u) du\right\} = S_0(t)^{\gamma(x)},$$

onde  $\gamma(x) = e^{-x'\beta}$ . No caso especial de duas amostras,  $x'\beta$  toma os valores 1 ou 0. Assim  $S_1(t|x) = S_0(t)^\gamma$ , para alguma constante  $\gamma$ . Sendo, portanto, bastante restritivo. Em particular, tal formulação impede que haja efeitos de covariáveis que sejam favoráveis por um período e desfavoráveis em outro e vice-versa.

Ademais, Koenker & Gelig (2001) mostram que as funções quantis do tempo de duração,  $T$ , neste modelo são:

$$(9) \quad Q_T(\tau|x) = S_0^{-1}[(1 - \tau)^{1/\gamma(x)}].$$

Estes autores sugerem que uma medida local do efeito do coeficiente no modelo de riscos proporcionais nos quantis condicionais em  $\tau$ , que denominar-se-á de efeito quantílico de tratamento no modelo de Cox, é dado pela derivada da função quantil em (9) avaliada no valor médio das variáveis explicativas:

$$(10) \quad b(\tau) = \frac{\partial Q_T(\tau|x)}{\partial x_i} \Big|_{\bar{x}} = \frac{(1-\tau)\log(1-\tau)\gamma(\bar{x})}{S_0'[Q_T(\tau|x)]} \beta_i.$$

Estes serão comparados, na aplicação empírica neste estudo, aos quantis de regressão  $\beta(\tau)$ , obtidos por regressão quantílica censurada.

Finalmente, outros tipos de modelagens em análise de sobrevivência, como o modelo de tempo de falha acelerado já mencionado, não exibem riscos proporcionais. Neste as

covariáveis impactam de forma a acelerar ou desacelerar o tempo de duração. Esta modelagem é a mais usada de forma generalizada atualmente e a mesma consiste no modelo subjacente à técnica de regressão quantílica para análise de sobrevivência conforme segue. E, para fins comparativos entre estas modelagens, considerar-se-á o  $\log(T)$  no efeito quantílico de tratamento também<sup>6</sup>.

### 3.2.2 Modelo de Regressão Quantílica Censurada

O modelo de Cox (e suas numerosas variações) representou a base de análise do efeito de covariáveis em dados de sobrevivência durante as últimas décadas. No entanto, o próprio Sir David Cox reconheceu uma possível artificialidade adjacente a tal técnica que modela a função taxa de falha (a qual não apresenta uma interpretação física substantiva, o que enfraquece a atratividade desta modelagem) ao invés do tempo de duração (ou de sobrevida) diretamente (Reid 1994). A regressão quantílica oferece uma abordagem substancialmente mais flexível no contexto de fácil interpretação do modelo linear<sup>7</sup>. Particularmente, os quantis de regressão,  $\beta(\tau)$ , podem ser interpretados como efeitos de regressão diretos nos tempos de sobrevida (ou no logaritmo destes no modelo de tempo de falha acelerado). Ademais, primeiramente, Portnoy (2003) evidencia que a regressão quantílica é mais promissora em identificar e modelar fontes de heterogeneidade que não são passíveis de serem modeladas pelos métodos de riscos proporcionais. Em segundo, modelos tradicionais, como o de riscos proporcionais, estabelecem que qualquer efeito na função quantil condicional seja uma função monotônica da correspondente função de risco basal. Porém, ressalva que o modelo de Cox é especialmente natural quando a função taxa de risco (ou de falha) é de interesse primário. E quando as covariáveis variarem no tempo, este modelo se aplica, mas não há quantil condicional para o tempo de duração (ou sobrevida) neste caso. Por conseguinte, este autor advoga que estas abordagens são fundamentalmente complementares e recomenda a utilização de ambas simultaneamente, pois tem-se assim, comumente, uma análise mais completa e confiável.

#### 3.2.2.1 Censura Aleatória: o Estimador de Portnoy

A censura aleatória acontece em situações nas quais os valores censurados,  $C_i$ , são apenas observados para as observações censuradas (ou, de forma equivalente, não são observados para observações não-censuradas). Neste contexto, são observados apenas os tempos dos eventos,  $T_i$ , e um indicador de censura,  $\delta_i$ , que toma o valor 1 se a observação é não-censurada e 0 se a observação é censurada. O modelo básico postula variáveis aleatórias  $\{T_i: i = 1, \dots, n\}$  que podem ser censuradas e que dependem de covariáveis  $\{x_i \in \mathfrak{R}^k: i = 1, \dots, n\}$ . É natural considerar os pares  $\{T_i, x_i\}$  como uma amostra multivariada, mas a teoria em Portnoy (2003) é desenvolvida condicionando a  $\{x_i\}$ . Ou seja,  $\{x_i\}$  são tomadas como fixas e, se necessário,  $P_{x_i}$  é usada para denotar a probabilidade condicional como uma função de  $x_i$ . Em aplicações  $T_i$  pode ser o tempo de duração ou alguma função deste, tal qual a log-transformação que provê o usual modelo de tempo de falha acelerado. O modelo quantílico especifica a dependência linear dos quantis condicionais de  $T_i$  em  $x_i$ . Especificamente, tome  $\beta(\tau)$  como mapeando  $0 \leq \tau \leq 1$  em  $\mathfrak{R}^k$  como um vetor de coeficientes de regressão tal que  $x_i' \beta(\tau)$  seja o  $\tau$ -ésimo quantil condicional de  $T_i$  (dado  $x_i$ ) tal que  $P_{x_i}\{T_i \leq x_i' \beta(\tau)\} = \tau$  com  $i = 1, \dots, n$ . E seja  $f_i$  a densidade (condicional) e  $F$  a função de distribuição acumulada de  $T_i$ . Em adição, há tempos de censura,  $\{C_i: i = 1, \dots, n\}$ . A distribuição de  $C_i$  pode depender de  $x_i$ , mas  $T_i$  é independente de  $C_i$  (condicional a  $x_i$ )<sup>7</sup>. Assim, tem-se o modelo geral de censura aleatória. E as variáveis aleatórias censuradas (observadas) e o indicador de censura são, novamente, dados por  $\tilde{T}_i = \min\{T_i, C_i\}$ ,  $\delta_i =$

$I\{T_i \leq C_i\}$ . E a tripla  $\{T_i, C_i, \delta_i\}$ , é comumente considerada como uma amostra aleatória. De forma usual,  $\beta(\tau)$  deve ser expresso em termos dos valores observados censurados  $\{\tilde{T}_i\}$  e  $\{\delta_i\}$ .

Portnoy (2003) propõe um método que estima recursivamente as funções quantis condicionais lineares a partir de dados de sobrevivência censurados, bem como prova a

6 Tem-se que  $\frac{\partial \log[Q_T(\tau|x)]}{\partial x_i} = \frac{\partial Q_T(\tau|x)}{\partial x_i} / Q_T(\tau|x)$ .

7 Sob censura,  $E[T|X]$  não é identificável. Em contrapartida, os quantis condicionais  $Q_T(\tau|x)$  são identificáveis para algum  $\tau$ . Por exemplo, se houver censura à direita de 40% dos dados, a região identificável dos quantis é  $\tau \in (0, 0,6)$ .

consistência e  $\sqrt{n}$ -convergência dos estimadores propostos<sup>8</sup>. Este método pode ser visto como uma generalização do estimador de Kaplan-Meier<sup>9</sup> para o contexto de um modelo de regressão. Este autor observou que os quantis da função de distribuição de Kaplan-Meier,  $\hat{F}(t) = 1 - \hat{S}(t)$  podem ser expressos como um problema de otimização de quantis ponderados, no qual o peso associado a observações censuradas é separado em dois termos. Uma parte desta é mantida na sua posição inicial no tempo de censura e o restante é deslocado para a direita (em efeito, para  $+\infty$ ). A ideia fundamental reside no fato de que a função quantil empírica em qualquer ponto depende apenas do sinal dos resíduos. Assim, a massa em  $C_i$ , pode ser redistribuída para qualquer ponto acima de todos os  $\{x'_i\beta: i \in L\}$ <sup>10</sup>, (por exemplo,  $+\infty$ ) e não necessariamente para observações específicas acima de  $C_i$ , conforme previamente sugerido na literatura afim<sup>11</sup>. A vantagem desta percepção é a de que tal reponderação se aplica a contextos mais gerais de regressão, nos quais não se sabe quais observações estão acima de  $C_i$ . Para visualizar isto, considere o contexto de apenas uma amostra sem censura, para a qual os quantis ordinários amostrais podem ser expressos como:

$$\hat{\xi} = \operatorname{argmin}_{\xi} \sum_{i=1}^n \rho_{\tau}(T_i - \xi) \quad (11)$$

Para obter a função escada:

$$\hat{\xi}(\tau) = t_{(i)} \quad \text{para } \tau \in [(i-1)/n, i/n], \quad (12)$$

onde  $\rho_{\tau}(u) = u[\tau - I(u < 0)]$  para  $\tau \in [0,1]$ . É de auxílio tomar o problema acima como paramétrico em  $\tau$ : na medida em que  $\tau$  aumenta a partir de 0,  $t_{(1)}$  é a solução até atingir  $\tau = 1/n$ , quando  $t_{(2)}$  passa a ser também uma solução e assim por diante. E quando há observações censuradas, procede-se de forma similar, à exceção de quando se encontra  $\tilde{t}_i$  tal que  $\hat{\xi}(\tilde{t}_i) = t_{(i)}$  e  $\delta_i = 0$ , e separa-se esta observação censurada em duas partes, uma destas permanece na sua posição original,  $t_{(i)}$ , recebendo o peso  $w_i(\tau) = \frac{\tau - \tilde{t}_i}{1 - \tilde{t}_i}$ , em todo o  $\tau$  subsequente e a outra é deslocada para  $t_{\infty} = +\infty$ , com peso complementar  $1 - w_i(\tau)$ . Note que o peso  $w_i(\tau)$  estima a probabilidade condicional de  $T_i \leq F^{-1}(\tau)$  dado que  $T_i > C_i$ . Isto é:

$$E\{I(T_i - x'_i\beta(\tau) \leq 0 | T_i > C_i, C_i, x_i)\} = \frac{\tau - \tilde{t}_i}{1 - \tilde{t}_i}. \quad (13)$$

Esta reponderação garante que  $\hat{\xi}(\tau)$  é constante numa vizinhança aberta de qualquer  $\tilde{t}_i$  e a massa restante correspondente à parte  $1 - w_i(\tau)$  de cada observação censurada seja distribuída apropriadamente. A percepção crucial subjacente é simplesmente a de que os quantis

8 O estimador de Kaplan-Meier da função de sobrevivência é uma técnica eficiente de estimação e pode ser interpretado como um estimador de máxima verossimilhança não-paramétrico (Andersen, Borgan, Gill, & Keiding, 1991).

9 O estimador de Kaplan-Meier da função de sobrevivência,  $S(t)$ , pode ser escrito como  $\hat{S}(t) = \prod_{i:t_{(i)} \leq t} [1 - 1/(n - i + 1)]^{\delta_i}$ , onde os  $t_{(i)}$ 's representam os tempos ordenados dos eventos e  $\delta_{(i)}$ 's são os respectivos indicadores de censura. Este representa uma técnica eficiente de estimação e pode ser interpretado como um estimador de máxima verossimilhança não-paramétrico (Andersen, Borgan, Gill & Keiding, 1991).

10  $L$  denota o índice que define as observações censuradas.

11 Uma abordagem alternativa foi proposta por Ying *et al.* (1995). No entanto, estes autores assumiram independência (não-condicional) entre  $T_i$  e  $C_i$ , que é consideravelmente mais forte que a hipótese de independência condicional (dado  $x_i$ ) em Portnoy (2003).

12 Ver Efron (1967).

dependem apenas da proporção de massa inferior e superior aos mesmos e o deslocamento de parte da massa censurada para  $+\infty$  assegura que todas as observações não censuradas subsequentes recebam a contribuição adequada devido às observações censuradas. Em sequência, denotando por  $L(\tau)$  o conjunto dos índices das observações censuradas até  $\tau$ , os quantis da distribuição de Kaplan-Meier,  $\hat{F}$ , podem ser expressados como a solução do seguinte problema:

$$\text{Min } \sum_{i \notin L(\tau)} \rho_{\tau}(T_i - \xi) + \sum_{i \in L(\tau)} [w_i(\tau) \rho_{\tau}(T_i - \xi) + (1 - w_i(\tau)) \rho_{\tau}(t_{\infty} - \xi)]. \quad (14)$$

A vantagem desta formulação é que esta pode ser generalizada para o contexto de regressão substituindo  $\xi$  por  $x_i' \beta(\tau)$ . Em particular, na ausência de censura,  $\tilde{T}_i = T_i$  e  $\beta(\tau)$  pode ser estimado minimizando  $S_n(\beta) = n^{-1} \sum_{i=1}^n \rho_{\tau}(T_i - x_i' \beta)$ .

O valor minimizador de  $S_n(\beta)$  é uma raiz da equação estimada:

$$D_n(\beta) = n^{-1} \sum_{i=1}^n x_i \{ \tau - I(T_i - x_i' \beta(\tau) \leq 0) \} \approx 0, \quad (15)$$

onde  $D_n(\beta)$  é a função gradiente. Conforme colocado anteriormente, o gradiente depende apenas dos sinais de  $T_i - x_i' \beta(\tau)$ . E na presença de censura, o algoritmo descrito em Portnoy (2003) define  $\{\hat{\tau}_i: i = 1, \dots, L\}$  indutivamente para as  $L$  observações censuradas tal que estes satisfazem  $x_i' \beta(\hat{\tau}_i) = C_i$  e  $\delta_i = 0$ . E seja  $i(j)$  tal que  $\hat{\tau}_{i(j)}$  estejam ordenados  $\epsilon \leq \hat{\tau}_{i(1)} < \dots < \hat{\tau}_{i(L)} \leq 1 - \epsilon$ . Assim posto, o estimador de regressão quantílica censurada é então  $\hat{\beta}(\tau, w)$ , a estimativa do quantil de regressão ponderado para  $\hat{\tau}_{i(j)} < \tau < \hat{\tau}_{i(j+1)}$  com pesos definidos por:

$$w_i(\tau) = \begin{cases} 1, & \tau \leq \hat{\tau}_i \text{ ou } \delta_i = 1 \\ \frac{\tau - \hat{\tau}_i}{1 - \hat{\tau}_i}, & \tau > \hat{\tau}_i \end{cases} \quad (16)$$

ou seja, tem-se: i) para observações não-censuradas, tome  $w_i = 1$ ; ii) para observações censuradas e ainda não cruzadas (acima do  $\tau$ -ésimo quantil tal que  $\delta_i = 0$  e  $T_i = C_i > x_i' \beta(\tau)$ ), trate-as como não-censuradas e tome  $w_i = 1$ ; iii) para observações censuradas e cruzadas  $\delta_i = 0$  e  $T_i = C_i \leq x_i' \beta(\tau)$  (ou seja,  $\tilde{\tau}_i = F(C_i | x_i) < \tau$ ), tome  $w_i(\tau) = \frac{\tau - \tilde{\tau}_i}{1 - \tilde{\tau}_i}$ . Desta forma, iniciando por  $\tau$  próximo de zero, a cada passo é possível avaliar o tamanho do intervalo dos  $\tau$ 's para o qual a solução corrente do seguinte problema de regressão quantílica ponderada:

$$\begin{aligned}
& \text{Min } \sum_{i \in L(\tau)} \rho_{\tau}(T_i - x_i' \beta(\tau)) + \sum_{i \in L(\tau)} [w_i(\tau) \rho_{\tau}(T_i - x_i' \beta(\tau)) \\
& \quad (17) \\
& \quad + (1 - w_i(\tau)) \rho_{\tau}(t_{\infty} - x_i' \beta(\tau))]
\end{aligned}$$

permanece ótima. Em seguida, retoma-se o problema atualizando-o e este é resolvido para o limite superior do intervalo, e a iteração prossegue até atingir  $\tau = 1$ .

#### 4 DADOS

Dentre os dados de duração, certamente os de desemprego são os que mais têm recebido atenção em termos de estudos empíricos. Os economistas estão interessados em descrever a distribuição entre os indivíduos da duração do desemprego. Os dados utilizados neste estudo são provenientes da PME para o ano de 2015, sendo este o período mais recente, pois a mesma foi descontinuada ao final de fevereiro de 2016. Esta pesquisa adota um esquema de rotação de painéis (*rotating panel*). Isto implica que os domicílios não são entrevistados durante todos os meses, mas estes entram e saem da amostra de acordo com um padrão predefinido. O esquema de rotação, denominado de 4-8-4, estabelece que, de outubro de um ano ímpar a setembro do ano seguinte, todo mês um grupo de domicílios entraria na pesquisa e seria entrevistado por quatro meses consecutivos. Do quinto ao décimo segundo mês, este grupo sairia da amostra, retornando no décimo terceiro mês e sendo entrevistado por mais quatro vezes. Na nova PME houve um ajustamento no processo de rotação para dar mais condições de acompanhamento longitudinal (Ribas & Soares, 2008). A disponibilidade de dados em painéis de domicílios permite que se acompanhe a condição de desemprego durante o período dos 4 meses (período amostral) em que o indivíduo é entrevistado. E, para tal, foram utilizados os procedimentos em Ribas & Soares (2008) no intuito de elevar a taxa de identificação do painel de pessoas através de algoritmo de emparelhamento proposto por estes autores. O banco de dados se restringe ao ano de 2015 no intuito de se evitar as ocorrências de durações múltiplas, quando vários períodos de desemprego podem ser observados para cada indivíduo na amostra (ver Kalbeisch & Prentice, 1980, seção 7.3). Nesta última condição, seria necessário o uso de modelagem para dados multivariados, na qual seria considerada uma variedade de noções de dependência, seja na duração e entre períodos (ver Heckman e Borjas, 1980).

A seleção do banco de dados segue Menezes-Filho & Picchetti (2000), com a distinção de que são consideradas todas as seis regiões metropolitanas e não apenas a de São Paulo. A amostra totaliza 1.906 indivíduos ao se manter apenas aqueles que procuraram emprego no mês anterior ao da entrevista. Em sequência, são consideradas as informações relativas à situação do indivíduo no mercado de trabalho nos três meses subsequentes em que este foi entrevistado pelo IBGE. Assim, obtém-se o número de meses em que a pessoa permanece desocupada no período amostral e, para calcular o tempo de duração do desemprego, adiciona-se a este valor os meses nos quais o indivíduo procurou trabalho no período de 365 dias anterior à entrevista. Desta forma, trabalha-se com a possibilidade de a duração do desemprego poder persistir além do curto prazo (período este no horizonte de apenas um ano). No que concerne às covariáveis, foram excluídos vários controles cujas estimativas dos coeficientes não são estatisticamente significativas em todos os quantis estimados no conjunto  $\tau = \{0,1; 0,25; 0,5; 0,75; 0,9\}$ . Dentre os regressores excluídos estão as dummies de condição na família (pessoa de referência, cônjuge e filho), idade ao quadrado, gênero e variáveis relativas ao emprego anterior: se foi demitido, experiência em anos e se o trabalho era no setor formal com carteira assinada. Por este critério, as seguintes variáveis explicativas foram

mantidas: idade, etnia oriental, dummies de diplomas (fundamental, médio, superior e mestrado/doutorado), indicadores das seis regiões metropolitanas, variável indicadora de o indivíduo já ter trabalhado anteriormente e dummies de ocupação. As dummies de ocupação foram construídas com base na tabela referente ao Código Nacional de Atividade Econômica (CNAE) e conforme a elaboração destas como variáveis derivadas na PME.

A Tabela 1 fornece as estatísticas descritivas dos regressores presentes na análise. Primeiramente, estes números indicam que os indivíduos se encontram na situação de desemprego por quase cinco meses em média e o máximo valor observado de inatividade é 15 meses. A faixa etária dos indivíduos que estão à procura de emprego varia bastante, sendo a média de 34,2 anos, e o primeiro e terceiro quartis são 25 e 42 anos, respectivamente. Ademais, tem-se que 0,3% dos desempregados são orientais. Em termos educacionais, a maioria das pessoas sem emprego (58,80%) têm apenas o nível médio, seguido pelo fundamental (23,71%), superior (15,70%) e pós-graduação (0,2%). Ou seja, na amostra, a quase totalidade das pessoas tem baixa escolaridade. Verifica-se ainda que,

Tabela 1: Descrição das Variáveis

	Mínimo	1º Quar.	Mediana	Média	Desv. Pad.	3º Quar.	Máximo
Meses de desemprego	1	2	3	4,549	3,60	7	15
Idade	16	25	33	34,198	11,453	42	84
Oriental	0	0	0	0,003	-	0	1
Fundamen-tal	0	0	0	0,2371	-	0	1
Médio	0	0	1	0,588	-	1	1
Superior	0	0	0	0,157	-	0	1
Mest./Dout.	0	0	0	0,002	-	0	1
Recife	0	0	0	0,117	-	0	1
Salvador	0	0	0	0,185	-	0	1
Belo Horizonte	0	0	0	0,163	-	0	1
Rio de Janeiro	0	0	0	0,143	-	0	1
Porto Alegre	0	0	0	0,147	-	0	1
Já trabalhou	0	0	1	0,512	-	1	1
Indústria	0	0	0	0,052	-	0	1
Comércio	0	0	0	0,063	-	0	1
Finan./Serv.	0	0	0	0,046	-	0	1
Construção	0	0	0	0,018	-	0	1
Pub./Saú./Soc.	0	0	0	0,018	-	0	1
Outros Serv.	0	0	0	0,047	-	0	1

Fonte: elaboração própria com dados da PME 2015 (IBGE).

em termos de percentual de desempregados por região metropolitana, este varia no intervalo de 11,7% em Recife a 18,5% em Salvador, com um valor intermediário em torno de 14% em ambos o Rio de Janeiro e Porto Alegre, e um pouco mais elevado em Belo Horizonte (16,3%). Adicionalmente, é interessante examinar também a situação dos desempregados em termos de experiência profissional anterior. Tem-se que metade da amostra, em torno de 51,2%, é composta de pessoas que já trabalharam anteriormente. Os setores de atividades do trabalho anterior de maior representatividade são o comércio (6,3%), seguido da indústria, outros serviços (Outros Serv.) e intermediação financeiro-imobiliária e serviços prestados à empresa

no trabalho principal (Finan./Serv.), nesta ordem. E os de menor representatividade são a construção com 1,8%, além dos que tiveram empregos na administração pública, defesa, seguridade social, educação, saúde e serviços (igualmente com 1,8%). Por fim, vale salientar que 44,1% das observações são censuradas, implicando em um número bastante elevado de pessoas que permanecem desempregadas ao final do período amostral.

## 5 RESULTADOS

Neste trabalho é assumida a formulação da equação de regressão em (5) tomando como variável resposta a transformação logarítmica do tempo do desemprego (em meses) seja para o modelo de regressão quantílica censurada quanto para o modelo de riscos proporcionais de Cox (para o cálculo do efeito quantílico de tratamento nesta modelagem). A escolha da transformação em log resulta primariamente pela facilidade em se ter linearidade na especificação paramétrica e por ser de fácil interpretação. É importante observar que o papel da transformação é atrativo no contexto da regressão quantílica por esta ser invariante a transformações monotônicas,  $g(\cdot)$ , ou seja,  $Q_{g(\tau)}(\tau|x) = x'\beta$  implica que  $Q_{\tau}(\tau|x) = g^{-1}(x'\beta)$ . Em contraste, o papel das transformações quando se modela a média condicional é bastante complicado, dado que a transformação afeta não apenas a locação, como também a escala e a forma da variável dependente.

A atenção nesta análise é centrada nos resultados para o vetor de coeficientes (os quantis de regressão),  $\beta(\tau)$ , e as estimativas destes são apresentadas na Tabela 2, por quantis da distribuição da duração do desemprego do universo metropolitano do país. A partir destes coeficientes estimados, esta tabela fornece as diferenças percentuais na duração do desemprego de todas as cinco regiões metropolitanas (RMs) consideradas em relação à RM de São Paulo (SP), a saber: Recife (RE), Salvador (SA), Rio de Janeiro (RJ), Belo Horizonte (BH) e Porto Alegre (PA). Inicialmente, foram consideradas como potenciais variáveis explicativas para os modelos todas as variáveis relativas ao indivíduo no período em que cada indivíduo entrou no desemprego e estava à procura de trabalho nos últimos 30 dias. Em seguida, foram retiradas todas as covariáveis que não apresentaram coeficientes estatisticamente significativos em nenhum dos quantis, mantendo-se as demais. Como exemplo, a manutenção da variável idade se deve a esta ser estatisticamente significativa apenas na cauda direita da distribuição de duração. Isto reflete que o fator idade contribui na permanência no desemprego quando este é de mais longa duração e, em particular, o efeito é positivo e de 0,003 pontos percentuais no quantil 0,9. No que concerne à educação, dois resultados se mostram interessantes. Primeiramente, percebe-se que o efeito (negativo) é crescente em termos de diminuir a duração do desemprego à medida em que se detém níveis mais elevados de diplomas, com exceção do diploma de graduação, que figura de forma menos impactante do que o de nível médio. Em segundo, ter diploma de pós-graduação (mestrado ou doutorado) implica em menores permanências na situação do desemprego ao longo de todos os quantis e, mais especificamente, a estimativa -0,471 pode ser traduzida num efeito multiplicativo no quantil 0,1 de  $\exp(-0.471) \approx -0,6244$  ou ainda -37,56% e de -65,77% [ $\exp(-1,072) - 1$ ] no quantil 0,9. Os números da Tabela 2 também revelam diferenças importantes na duração do desemprego intra-regionais e entre regiões. Um resultado esperado, devido às semelhanças como grandes regiões metropolitanas, é o de que as estimativas dos parâmetros relativos ao Rio de Janeiro não são estatisticamente significativos em todos os quantis da distribuição do desemprego, revelando que estes não diferem daquele da dummy de base (São Paulo). Outro resultado bastante relevante é o de que os efeitos entre regiões são estatisticamente significativos apenas na cauda esquerda da distribuição, enquanto perdem significância na mediana e na cauda direita (quantis 0,75 e

0,90). É também digno de nota que as estimativas para a região metropolitana de Recife são as que apresentam maior magnitude (negativa), revelando que os indivíduos nesta região experimentam menores permanências no desemprego relativamente às demais, bem como é a única que apresenta efeito negativo e estatisticamente significativo na mediana. Esta é seguida pelas regiões metropolitanas de Belo Horizonte e Porto Alegre, que se mostram equiparadas, e Salvador em derradeiro, porém ainda apresentando menores durações de desemprego que as registradas no RJ e SP, mas apenas para o primeiro quartil (os efeitos são nulos nos demais quantis). Os efeitos relativos às regiões metropolitanas acarretam quedas na duração do desemprego que se situam entre -9,52% (Salvador/quartil 0,25) e -29,60% (Recife/quartil 0,1). Em adição, encontra-se variações distintas para o impacto relativo aos setores de atividades na distribuição condicional do desemprego, porém o sinal é sempre positivo, transparecendo, portanto, que a duração do desemprego é mais elevada para pessoas empregadas em atividades classificadas (indústria, construção, comércio, etc) em termos de mercado de trabalho relativamente às demais (classificadas como outras atividades na PME) de menor representatividade e trabalhos domésticos. Vale salientar os seguintes resultados. Os efeitos referentes aos setores de atividades tendem a ser maiores na mediana em particular e na cauda esquerda, com tendência decrescente no sentido desta última. Os setores de atividades nos quais os coeficientes estimados apresentam maior magnitude relativa são a indústria, seguida do comércio e serviços financeiro/imobili-

Tabela 2 Resultados das Regressões Quantílicas Censuradas.

duração desemprego	Quantil 0,1	Quantil 0,25	Quantil 0,5	Quantil 0,75	Quantil 0,9
constante	1,589*	1,803*	2,431*	2,407*	2,405*
idade	0,002	-0,001	-0,001	0,004	0,003*
oriental	0,432*	0,243*	0,136*	-0,687*	-1,043*
fundamental	-0,194*	-0,116	-0,200*	-0,156**	-0,067
medio	-0,297*	-0,224*	-0,256*	-0,163***	-0,041
superior	-	-0,110***	-0,186*	-0,074	0,049
mestdout	0,208**	-0,939*	-0,571*	-0,916*	-1,072*
Rec	-0,471*	-0,322*	-0,161**	-0,049	-0,099
Salv	-0,351*	-0,102*	0,002	-0,014	-0,063
BelHoriz	-0,098	-0,198*	-0,084	0,016	0,001
RJ	-	-0,029	0,040	0,039	0,045
PoA	0,232**	-0,215**	-0,050	0,061	0,079*
jatrabalhou	-0,077	-0,312*	-0,767*	-0,315*	-0,055
industria	-	0,600*	0,746*	0,239*	0,057
comercio	0,178**	0,491*	0,641*	0,167*	0,062
finanserv	-0,342*	0,575*	0,647*	0,140	0,054
construcao	0,500*	0,409*	0,705*	0,113	-0,061
pubsausoc	0,462*	0,324*	0,478*	0,110	0,038
outserv	0,299*	0,473*	0,572*	0,134	0,045
	0,286*				
	0,435*				
	0,4312*				

Nota: \* denota significância ao nível de 1%, \*\* de 5% e \*\*\* de 10%

ários e para atividade fins da empresa, além de outros serviços, nesta ordem. Os valores das

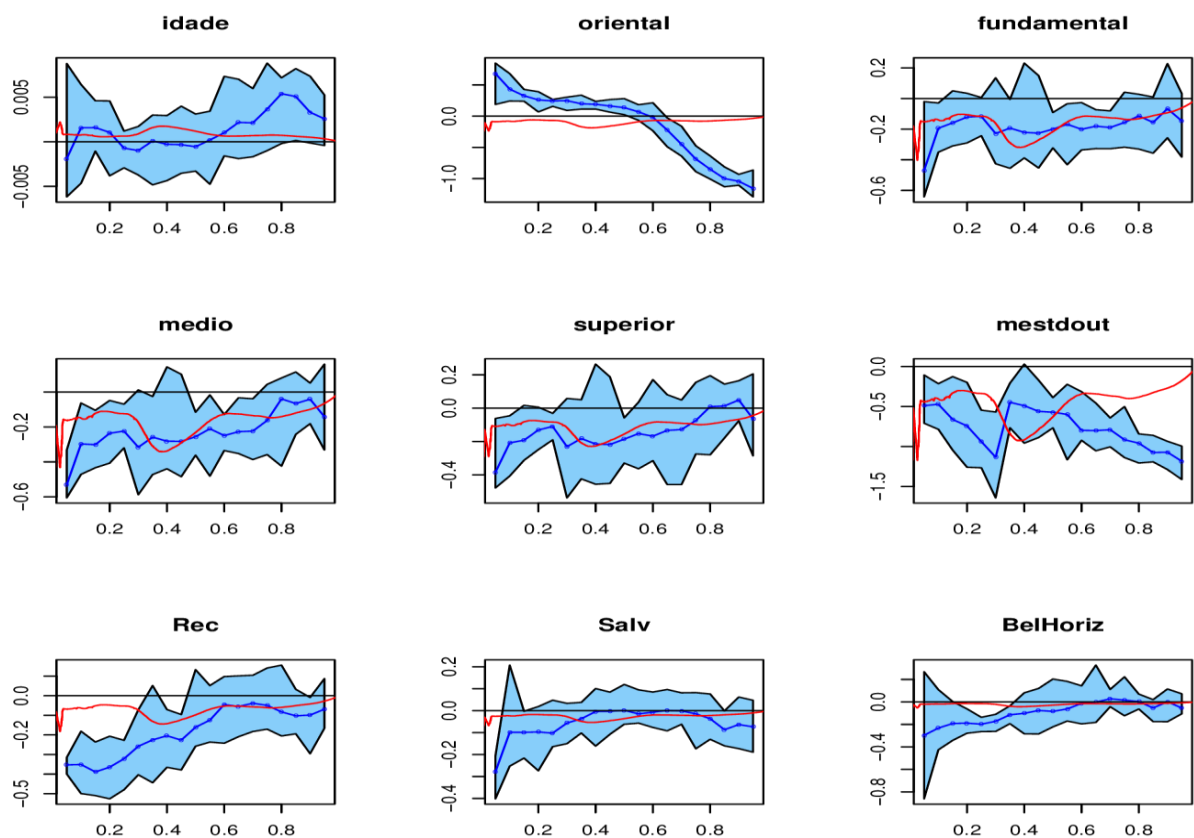


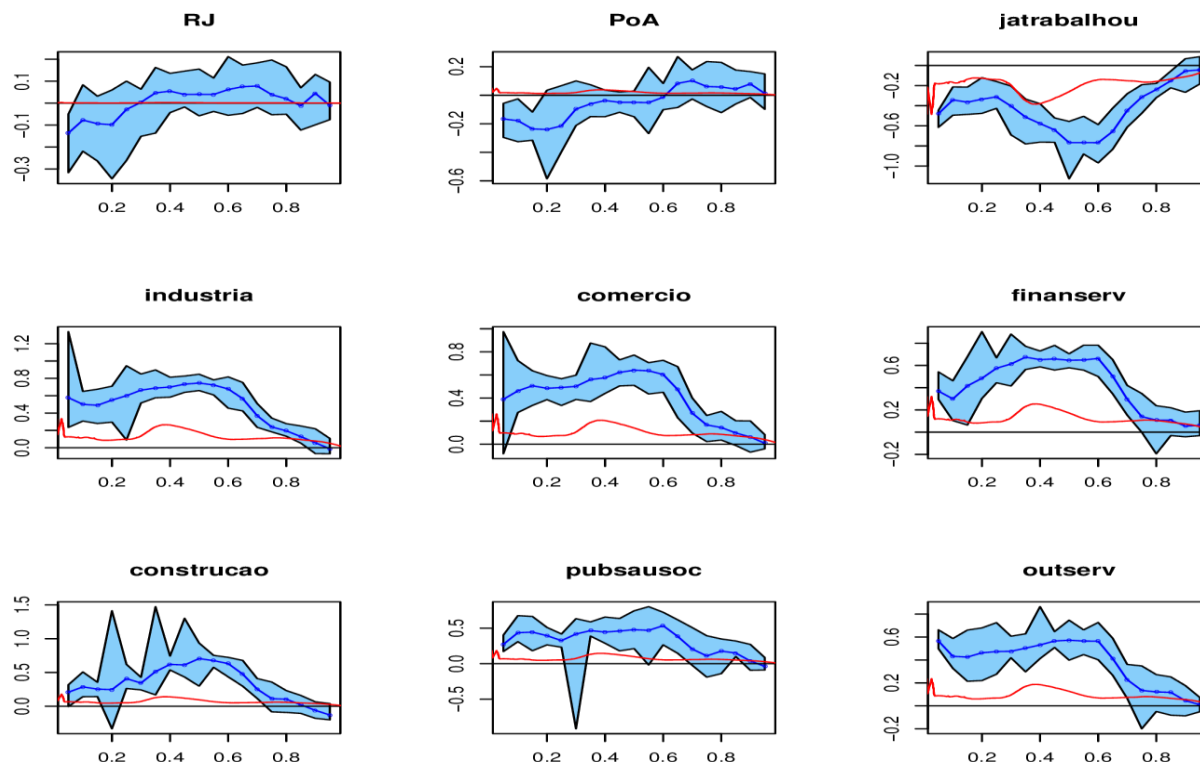
estimativas dos quantis de regressão relativos aos setores de atividade se encontram entre os limites de 18,18% (comércio/quantil 0,75) a 110,86% (indústria/mediana) ao longo da distribuição, enquanto que estes são estatisticamente significativos no quantil 0,75 apenas para a indústria e comércio, e perdem a significância no quantil 0,9. Por fim, outro ponto que merece destaque é a relevância de ter tido alguma experiência anterior de trabalho (variável já trabalhou) que acarreta uma diminuição da duração do desemprego em quase toda a distribuição condicional desta, em especial na mediana (-53,56%) e quantitativamente menor nos demais quantis (em torno de -27% tanto abaixo da mediana quanto para o quantil 0,75), enquanto que perde a significância em durações mais prolongadas (quantil 0,9).

O gráfico 1 mostra os quantis de regressão ( $\beta(\tau)$ ) e o efeito quantílico de tratamento ( $b(\tau)$ ) dado por (10) em complemento com a na nota de rodapé 5. Ambos foram computados usando os pacotes *crq* e *survival* no software livre R (ver Koenker (2008) e Portnoy (2003)). Esta figura provê um resumo visual conciso dos resultados das estimativas dos parâmetros das 18 covariáveis ao longo dos quantis da distribuição condicional da duração de desemprego. Nestes, as estimativas  $\hat{\beta}(\tau)$  por regressão quantílica censurada são dadas pela curva central que estão sombreadas em azul por um intervalo de confiança de 95%, obtido através do método de *bootstrap*. Por sua vez, a curva em vermelho representa os valores estimados do efeito quantílico de tratamento (EQT) para o modelo de Cox, ou seja,  $\hat{b}(\tau)$ . Há ainda uma linha reta no valor zero, que serve de referência pra avaliar a significância estatística das estimativas. Em primeiro plano, percebe-se que para algumas variáveis as estimativas para o QTE de Cox diferem de forma significativa daquelas para os quantis de regressão para alguns regressores. Isto parece mostrar que a natureza global do modelo de Cox dificulta a estimação do verdadeiro efeito quantílico (ver Portnoy, 2003). Na primeira figura, nota-se que a linha reta traçada no valor zero se encontra dentro do intervalo de confiança para as estimativas dos coeficientes da variável explicativa idade, sendo este fator sem influência ao longo de toda a distribuição da duração, a não ser pela parte bem extrema da cauda direita. Na segunda figura é descrito o efeito devido à etnia oriental, que representa a diferença estimada nos quantis do log do tempo da duração para as pessoas orientais relativamente à demais raças. É evidente que este efeito é consideravelmente diferenciado de um deslocamento locacional simples. Ademais, há uma tendência decrescente visível ao longo dos quantis da cauda esquerda para a direita, que demonstra que os orientais têm uma desvantagem de sair do desemprego para o emprego em períodos mais curtos de desemprego até a mediana, e a partir desta e em direção à cauda direita (período mais prolongados), esta etnia apresenta vantagem relativamente às demais raças. É importante salientar que para esta variável dummy em particular, os EQTs de Cox estimados diferem bastante das estimativas dos quantis de regressão, indicando não captar efeito algum ao longo dos quantis. Em sequência, as curvas das estimativas para as dummies educacionais, do diploma fundamental ao nível superior, aparentam ser bastante semelhantes, com tendência crescente a partir do primeiro quartil (quantil 0,25) e situando-se no quadrante negativo, demonstrando que a obtenção de diplomas em qualquer nível educacional é vantajoso no sentido de diminuir o tempo passado na situação de desemprego em todos os quantis. A curva para o efeito relativo à dummy de mestrado e doutorado também se enquadra nas características das demais dummies educacionais, porém esta tem projeção decrescente por volta do quantil 0,4 em diante, implicando que a vantagem de se ter grau de pós-graduação é mais acentuada quanto maior à duração do desemprego (da mediana à cauda direita). Em adição, as estimativas dos EQTs de Cox seguem mais fielmente as curvas dos quantis de regressão relativos às dummies educacionais em comparação aos resultados encontrados para as demais covariáveis. Passando para as dummies de regiões metropolitana, os quantis de regressão estimados para o RJ e Salvador são estatisticamente não significativos na totalidade

da distribuição ordem, que os demais para saírem do desemprego, apresentando estimativas no quadrante inferior e estatisticamente significativas nos quantis abaixo da mediana. Enquanto que os que moram em Porto Alegre apresentam tal padrão num intervalo menor entre o primeiro quartil (0,25) e a mediana. Um dos regressores que mais impactam em termos de retomada da duração do desemprego, com exceção dos efeitos por volta de uma região pequena em torno do quantil 0,25 no que concerne a dummy relativa a Salvador. Nota-se que os residentes das regiões metropolitanas de Recife e Belo Horizonte são mais rápidos, nesta do emprego mais cedo, ao longo de toda a distribuição com exceção da parte extrema da cauda direita, é aquele que indica que o indivíduo tem experiência de trabalho anterior. O formato da curva em U demonstra que a saída do desemprego se dá a taxas decrescentes até por volta da mediana e crescentes a partir desta, bem como ilustra um efeito diferenciado que se reflete no cruzamento das funções de

Gráfico 1: Quantis de regressão  $\beta(\tau)$  no modelo de regressão quantílica censurada (e intervalo de 95% de confiança por *bootstrap*) em azul e o efeito quantílico de tratamento no modelo de Cox (em vermelho).

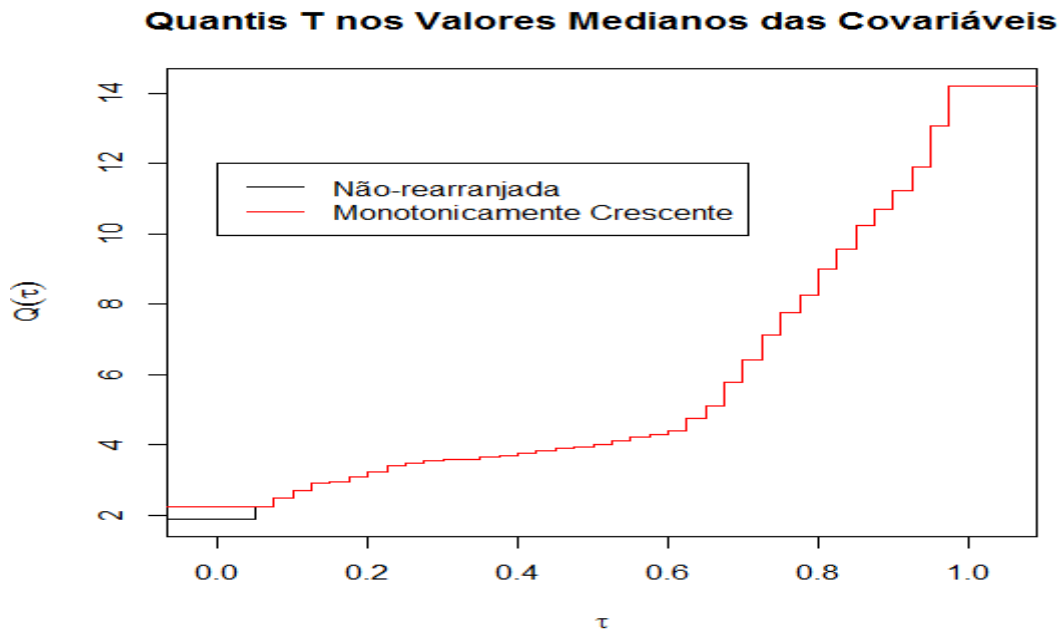




sobrevivência, denominado de *cross-over effect* (ver Koenker & Gelig, 2000). Finalmente, as últimas figuras apresentam as curvas relativas aos efeitos das dummies de setores de atividades. Todas se encontram no quadrante superior, implicando em desvantagem no retorno à situação de emprego em todas as categorias, relativamente às dummies de base que representam outras atividades em conjunto com os serviços domésticos. Estas curvas apresentam ligeiro formato de U. As estimativas são mais expressivas em termos de magnitude para os setores da indústria, comércio, serviços financeiro-imobiliários e outros serviços. Nota-se ainda que estas estimativas perdem a significância estatística na cauda superior, por volta do quantil 0,75 em diante. Por fim, os respectivos efeitos estimados para os QTEs de Cox se mostram distintos e aquém das estimativas dos quantis de regressão, no entanto, conforme ressaltado antes, as estimativas por regressão quantílica se mostram mais confiáveis (Portnoy, 2003).

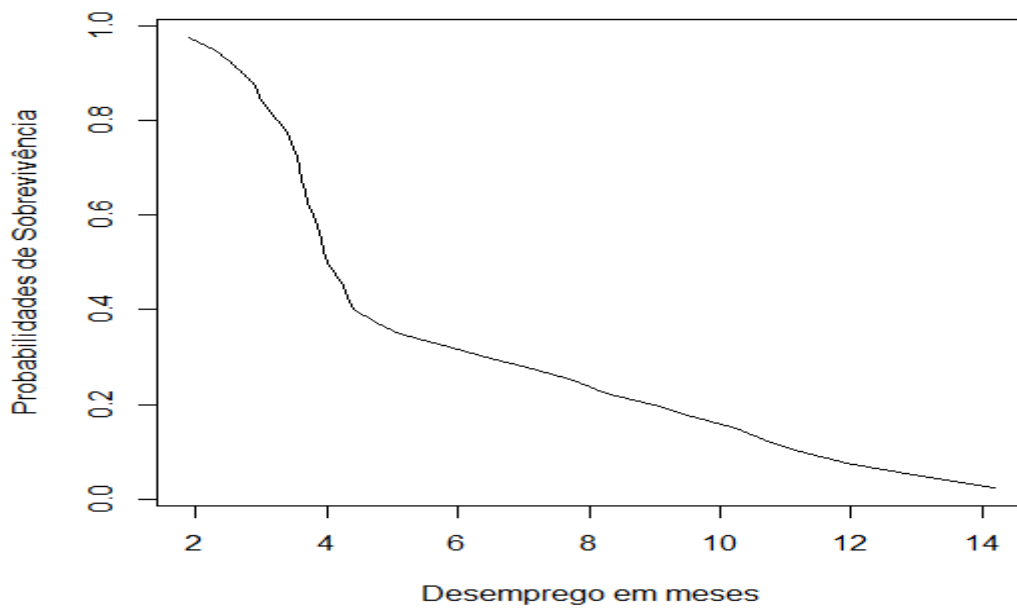
A partir dos modelos estimados, a função quantil condicional pode ser estimada para quaisquer valores das covariáveis. Avaliada nos valores médios das variáveis explicativas,  $\bar{X}$ , a função quantil ajustada é monotonicamente crescente (Koenker, 2005, Theorem 2,5), mas isto não é garantido para outros valores. Uma forma simples e teoricamente atrativa de lidar com tais violações da monotonicidade foi introduzida por Chernozhukov, Fernández-Val & Galichon (2006). No Gráfico 2 tem-se a função quantil condicional ajustada nos valores medianos dos controles e a linha em vermelho é a versão rearranjada, demonstrando que pequenas violações de monotonicidade são observadas apenas no extremo da cauda esquerda da distribuição condicional da duração do desemprego.

Gráfico 2: Função Quantil Condicional Prevista.



A partir da função quantil condicional ajustada, Koenker & Gelig (2001) observam que, ao invés de fazer o gráfico de  $\tau \mapsto \hat{Q}_T(\tau|X)$ , pode-se mapear a função de sobrevivência como  $\hat{Q}_T(\tau|X) \mapsto 1 - \tau$ , ilustrada no Gráfico 3. Neste percebe-se que a probabilidade de permanecer no desemprego decai a uma taxa maior até o quarto mês de desemprego, quando esta queda se suaviza deste ponto em diante, demonstrando que as chances de permanecer no desemprego são menores a partir de quatro meses na situação de desocupação.

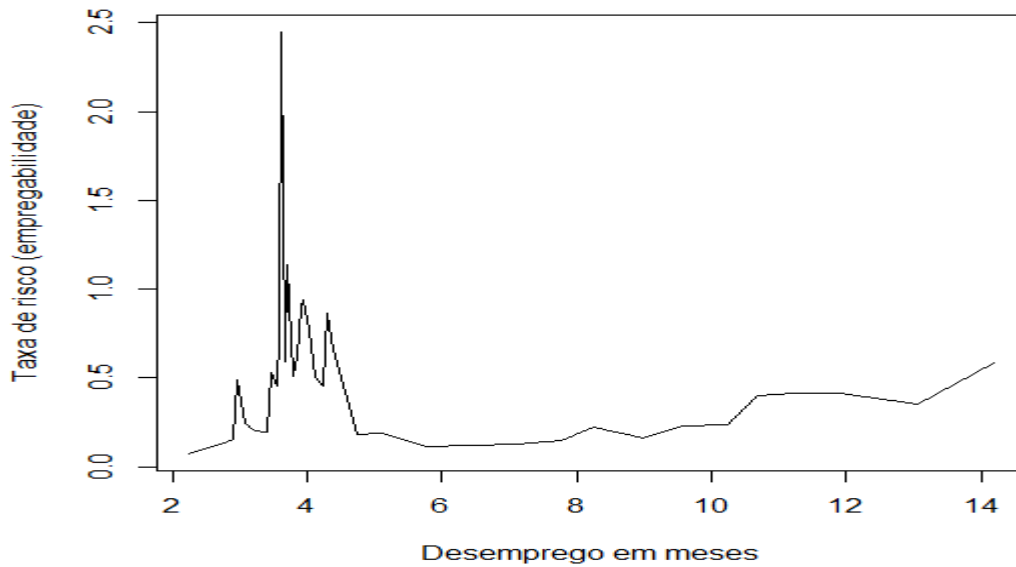
Gráfico 3: Função de Sobrevivência (valores medianos das covariáveis).



Por fim, o Gráfico 4 mostra as taxas de risco condicionais estimadas a partir do

modelo de regressão quantílica censurada. Tradicionalmente, as análises de duração

Gráfico 4: Função de Risco (valores medianos das covariáveis).



focam o impacto de covariáveis na função taxa de risco. A regressão quantílica censurada, por sua vez, estima a distribuição de  $T_i$  condicional às covariáveis, sendo possível estimar as taxas condicionais de risco a partir das estimativas de regressão quantílica. Conforme descrito em Fitzenberger & Wilke (2006), uma abordagem direta consiste em construir uma estimativa da densidade a partir dos quantis condicionais ajustados  $\hat{Q}_T(\tau|X) = g^{-1}(x_i'\beta(\tau))$ . Uma estimativa simples da taxa de risco pode ser construída como uma aproximação da taxa de risco entre dois  $\tau$ -quantis diferentes:

$$\hat{\lambda}_i(t) = \frac{\tau_2 - \tau_1}{(g^{-1}(x_i'\beta(\tau_2)) - g^{-1}(x_i'\beta(\tau_1)))(1 - 0,5(\tau_1 - \tau_2))}, \quad (18)$$

onde  $\hat{\lambda}_i(t)$  aproxima as taxas de risco entre o  $\tau_1$ -quantil e o  $\tau_2$ -quantil estimados. O Gráfico 4 sugere que há mais chances de saída do desemprego por volta do quarto mês na situação de desocupação.

## 6 CONCLUSÕES

O presente estudo inova na aplicação da regressão quantílica censurada para modelar a duração do desemprego no universo das regiões metropolitanas brasileiras, fornecendo um complemento rico em termos de resultados empíricos relativamente aos encontrados em artigos anteriores na literatura nacional e que utilizam os métodos tradicionais. Esta técnica estatístico-econométrica possibilita a estimação do efeito das covariáveis nos distintos quantis da distribuição condicional do desemprego, permitindo que estas impactem de forma mais flexível tanto na locação como na escala e forma desta, além de fornecer uma visão mais ampla da relação entre as variáveis. Os resultados mostram que a idade do indivíduo praticamente não interfere na duração do desemprego, a não ser no quantil 0,9, com efeito positivo. Ademais, há evidências de que a duração seja menor para indivíduos que já

trabalharam anteriormente, bem como aqueles que possuem diplomas educacionais, em especial os de mestrado e doutorado. Vale também salientar que há vantagens de saída do desemprego para o emprego nas regiões metropolitanas de Recife, Belo Horizonte e Porto Alegre, nesta ordem, em comparação com as outras (Salvador, Rio de Janeiro e São Paulo). No que diz respeito à experiência de trabalho anterior, o efeito dos indicadores de todos os setores de atividades considerados como controles são positivos ao longo de todos os quantis, sugerindo um impacto no sentido de maior duração no desemprego relativamente a outras atividades menos representativas e serviços domésticos. É também relevante o impacto da raça oriental na direção de desvantagem em termos de permanência na situação de desemprego nos quantis inferiores (sinal positivo) e vantagem (sinal negativo) relativamente a outras raças da mediana em direção à cauda direita. Por fim, diferentemente de estudos prévios com dados da PME, não foram encontradas evidências de impacto na duração de desemprego em função do indivíduo ser a pessoa de referência na família, bem como as estimativas dos coeficientes da variável anos de experiência no emprego anterior e dos indicadores referentes a emprego formal e se a pessoa foi demitida neste, não são estatisticamente significativos em ponto alguma da distribuição de duração.

## REFERÊNCIAS

- Abraham, K., Shimer, R. (2001). *Changes in unemployment duration and labor force attachment*. NBER Working Paper No. 8513.
- Andersen P. K., Borgan, Gill RD, Keiding N (1991). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Bivar, W. (1993). Estimativas da duração média do desemprego no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, **23**(2), 275-312.
- Cacciamali, M. C., Lima, T. T., Tatei F. (2015). Determinantes da Duração do Desemprego no Brasil em Crises Econômicas. *Revista Venezuelana de Análisis de Coyuntura*, **XXI**(1), 187-209
- Chaudhuri, P., Doksum, K., Samarov, A. (1997). On average derivative quantile regression. *Annals of Statistics*, **25**, 715-44.
- Chernozhukov, V., Fernández-Val, I., Galichon, A. (2006). Quantile and probability curves without crossing. *Econometrica*, **78** (3), 1093-1125
- Cox D.R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

Efron, B. (1967). The Two-Sample Problem With Censored Data. In: *Proc. Fifth Berkeley Symposium in Mathematical Statistics*, IV, eds. L. Le Cam and J. Neyman, New York: Prentice-Hall, 831-853.

Ehrenberg, R., Oaxaca, R. L. (1976). Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Growth. *American Economic Review*, **66**, 754-766.

Fitzenberger, B., Wilke, R. (2006). Using Quantile Regression for Duration Analysis. *Allgemeines Statistisches Archiv*, **90**, 103–118.

Gill, R., D., Johansen, S. (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis. *The Annals of Statistics*, **18**(4), 1501-1555.

Ham, J., LaLonde, R. (1991). *Estimating the Effect of Training on Employment and Unemployment Durations: Evidence from Experimental Data*. NBER Working Paper, No. 3912.

He, X. (2017). A Convergence with Roger Koenker. *International Statistical Review*, **85**, 1, 46–60.

Heckman, J., Borjas, G. (1980). Does Unemployment Causes Future Unemployment? Definitions, Questions and Answers for a Continuous Time Model of Heterogeneity and State Dependence. *Economica*, **47**, 247-285.

Heckman, J., Vytlacil, E. (2007). Econometric Evaluation of Social Programs, part I: Causal Models, Structural Models and Econometric Policy Evaluation, *Handbook of Econometrics*, **6B**, 4779-4874.

Imbens, G., Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47**, 5-86.

Kalbeisch, J., Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Katz, L.F., (1986). *Layoffs, Recall and the Duration of Unemployment*. NBER Working Paper, vol. 1825, January.

Kiefer, N. (1988). Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, **XXVI**, 649-679.

Koenker, R. (2005). *Quantile Regression*. Cambridge U. Press, London.

Koenker, R. (2008). Censored Quantile Regression Redux. *Journal of Statistical Software*, **27**(6), 1-25. URL <http://www.jstatsoft.org/v27/i06/>.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Koenker, R., Biliias, Y. (2001). Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments. *Empirical Economics*, **26**, 199-220.

- Koenker, R., Geling, O. (2001). Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. *Journal of the American Statistical Association*, Vol.96, No.454, 458-468.
- Menezes-Filho, N. A., Picchetti, P. (2000). Os determinantes da duração do desemprego em São Paulo. *Pesquisa e Planejamento Econômico*, **30**(1), 23-47.
- Penido, M., Machado, A. F. (2002). Desemprego: Evidências da Duração no Brasil Metropolitano. *Texto para Discussão No. 176 – UFMG/ Cedeplar*.
- Peng L, Huang, Y. (2008). Survival Analysis with Quantile Regression Models. *Journal of American Statistical Association*, **103**, 637–649.
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**(464):1001–1012.
- Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, **32**, 143-55.
- Avelino, R. R. G. (2000) Os Determinantes do Desemprego em São Paulo. *Textos para Discussão – IPE/USP*.
- Ribas, R. P., Soares, S. D. (2008). Sobre o painel da Pesquisa Mensal de Emprego (PME) do IBGE. *Texto para Discussão, n. 1348*. Rio de Janeiro: IPEA.
- Reid, N. (1994). A Conversation With Sir David Cox. *Statistical Science*, **9**, 439-455.
- Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, **90**:178–184.