

O Atrito nas Pesquisas Longitudinais: O Caso da Pesquisa Mensal de Emprego (PME/IBGE)

▪ Rafael Perez Ribas*

▪ Sergei Suarez Dillon Soares**

Resumo

O objetivo deste artigo é estimar os determinantes da permanência de pessoas no painel da Pesquisa Mensal de Emprego (PME) e, além disso, testar se ignorar a não-aleatoriedade do desgaste no painel causa algum viés em análises de regressão. Os resultados apontam que há uma forte correlação da taxa de atrito com as características geralmente associadas a uma mobilidade geográfica maior. Além disso, tanto características do processo de entrevista quanto fatores socioeconômicos estão significativamente relacionados à probabilidade de atrito no painel. Nos modelos adotados para testar o viés do desgaste no painel, assim como de outros critérios de seleção amostral, constatamos que todos estes são endógenos, apesar de alguns não estarem diretamente correlacionados com a equação de interesse. Por fim, mostramos que a ausência do controle para seleção amostral, o que inclui a retenção no painel, pode incorrer em algumas análises enviesadas.

Palavras-Chave

Pesquisa Mensal de Emprego, desgaste de painel, viés de atrito, correção para viés de seleção

Abstract

This article aims to estimate the determinants of attrition/permanence of people in the panel of the Brazilian Monthly Employment Survey (PME/IBGE) and also to test the bias emerged from ignorability of attrition selection in regression analyses. Results point out a strong correlation between attrition rate and characteristics closely related to geographic mobility. In addition, both survey design aspects and socioeconomic events are significantly related to the probability of attrition. In selection correction models, we find that all selection mechanisms are endogenous, even though they are not directly correlated with the equation of interest. Finally, we show that ignoring the endogeneity of sample selection, such as attrition, may imply to a biased analysis. However, controlling for only some selection mechanism, not controlling for all, may provide estimators even more biased than no controlling.

Keywords

Brazilian Monthly Employment Survey, panel attrition, attrition bias, bias selection corrections

JEL Classification

C33, C81, J60

* University of Illinois at Urbana-Champaign. Endereço para contato: Department of Economics – 419 David Kinley Hall – 1407 W. Gregory Dr. – Urbana – IL 61801 – USA. E-mail: ribas1@illinois.edu.

** Instituto de Pesquisa Econômica Aplicada / Diretoria de Estudos Sociais. E-mail: sergei.soares@ipea.gov.br.

(Recebido em abril de 2008. Aceito para publicação em março de 2009).

1 Introdução

A Pesquisa Mensal de Emprego (PME), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), é uma pesquisa amostral domiciliar de periodicidade mensal e seu levantamento é realizado em seis Regiões Metropolitanas do Brasil. Além da disponibilização mensal de microdados, o painel da PME, que acompanha as unidades amostrais por até oito entrevistas, permite a investigação de fenômenos socioeconômicos no nível individual/domiciliar que ocorrem de forma longitudinal. Contudo, como todo painel domiciliar, a PME está sujeita ao problema de perda das unidades amostrais ao longo do tempo, o chamado desgaste ou atrito do painel.

Para Peracchi e Welch (1995), são duas as principais causas para o desgaste em um painel. A primeira causa de desgaste em uma pesquisa longitudinal está relacionada diretamente à mudança de endereço das pessoas na amostra. Na PME, este problema é ainda maior que em outras pesquisas em painel, como o PSID (*Panel Study of Income Dynamics*) e o BHPS (*British Household Panel Survey*), que se esforçam em encontrar as pessoas em outros endereços. No caso da PME, se os indivíduos deixam seus domicílios, eles são automaticamente excluídos da pesquisa. A segunda causa para o desgaste é a recusa de entrevista. Neste caso, a hipótese é de que, após participar repetidas vezes de uma determinada pesquisa, os participantes podem se tornar desinteressados, recusando-se a ser entrevistados novamente. Existem ainda outras causas de menor proporção para o desgaste, como a morte de indivíduos e o fato de o entrevistador não encontrar nenhum morador no momento em que visita o domicílio.

Independente da perda no número absoluto de observações, um problema maior em qualquer análise de transição ocorre quando o atrito não é aleatório. Por exemplo, indivíduos com elevada mobilidade geográfica tendem a diferir daqueles com maior estabilidade em seu endereço. Com isso, o problema surge quando um determinado grupo de indivíduos torna-se sub-representado na amostra em decorrência do próprio fenômeno socioeconômico analisado.

Neri *et al.* (1997), por exemplo, aponta que, particularmente no caso em que estavam analisando, não havia diferenças significativas nas características demográficas e econômicas entre as amostras atritadas e não atritadas. Por outro lado, o trabalho de Lopes (2002) aponta que algumas características individuais, como a situação de desemprego e a idade, e do domicílio, como o sexo do chefe, estão significativamente relacionadas com o desgaste no painel. Independente disso, Falaris e Peters (1998) salientam que, mesmo quando a sub-amostra de atritados apresenta estatísticas semelhantes às dos demais, a estimação sem os termos de correção do viés pode incorrer em resultados inconsistentes.

O objetivo deste artigo é estimar os determinantes do atrito/permanência no painel da PME e, além disso, testar se ignorar a não-aleatoriedade deste desgaste causa algum viés na análise de regressões. Na seção seguinte a esta introdução, apresentamos uma revisão sobre viés e controle da seleção amostral não aleatória causados pelo atrito em painéis. Em seguida, é apresentada uma descrição sobre a PME e particularmente sobre a estrutura do seu painel. Na quarta seção, são apontados os resultados das regressões sobre a probabilidade de as pessoas permanecerem no painel da PME na entrevista seguinte. A seção posterior apresenta os resultados sobre as implicações de se ignorar um possível viés de atrito em alguns tipos de análises com o painel da PME. Cabe deixar claro que este trabalho não tem a intenção de discutir os temas específicos envolvendo estes resultados. A discussão apresentada está relacionada apenas aos aspectos metodológicos das estimativas. Por fim, algumas conclusões são traçadas com base nos resultados encontrados.

2 Revisão da Literatura

Uma pesquisa longitudinal é aquela na qual uma unidade amostral é entrevistada mais de uma vez, o que quer dizer que se pode montar um painel de dados.¹ Em geral, dados em painel permitem análises mais precisas que dados *cross-section* ou “transversais”, possibilitando, por exemplo, o controle de efeitos fixos e de condições iniciais. Os painéis, por outro lado, apresentam alguns problemas, e o principal deles é o desgaste.

Se a morte, nascimento ou troca de endereço fossem eventos que ocorressem de forma aleatória – i.e., sem nenhuma correlação com variáveis observadas ou não observadas – o atrito seria um problema apenas na medida em que reduzisse o tamanho da amostra, aumentando o intervalo de confiança. Caso contrário, se o atrito é baseado em fatores que são sistematicamente relacionadas à variável resposta, um problema de seleção amostral poderá ocorrer, tornando a análise enviesada.

Hausman e Wise (1979) enfatizam que o atrito não aleatório não torna necessariamente as estimativas de um modelo enviesadas. O atrito relacionado somente a variáveis exógenas não causa problemas, desde que essas variáveis sejam controladas na estatística. Basicamente, o problema de viés de seleção ocorre somente quando o atrito está relacionado aos erros aleatórios da equação de interesse.

1 Exemplos no Brasil são o Censo Escolar, no qual cada escola é entrevistada a cada ano; a Relação Anual de Informações Sociais (RAIS), nas quais as empresas e unidades de produção são entrevistadas repetidas vezes; e a Pesquisa Mensal de Emprego (PME), na qual cada domicílio e, potencialmente, cada morador, é entrevistado até oito vezes. Ademais, o IBGE estuda a transformação da Pesquisa Nacional por Amostragem de Domicílios (PNAD) em uma pesquisa contínua com um painel rotativo de domicílios.

Embora na literatura de amostragem o tema de desgaste de painéis seja tratado há mais tempo (e.g. HORVITZ; THOMPSON, 1952), o trabalho de Hausman e Wise foi um dos primeiros a levantar a questão na literatura econométrica. O estudo em questão era sobre o impacto de um experimento (*Gary Income Maintenance Experiment*) sobre os rendimentos de trabalhadores negros. Com a perda de 35% da amostra no período seguinte, o viés de atrito foi identificado como um problema nas avaliações mais parcimoniosas. Porém, esse viés diminuiu com a incorporação de mais variáveis exógenas no modelo. Para correção do viés, os autores propuseram uma função de verossimilhança muito próxima da utilizada por Heckman (1974).

Apesar da extensa literatura sobre viés de atrito, publicada após os trabalhos de Hausman e Wise (*op. cit.*) e Heckman (*op. cit.*; 1979), alguns trabalhos empíricos continuam não se atendo ao problema.² A prática mais comum utilizada nestes trabalhos é descartar as unidades com observações ausentes nos períodos subsequentes. Dessa forma, algo fortemente relacionado a fatores econômicos e sociais de interesse, como a estabilidade da unidade familiar, acaba sendo imposto como uma exigência de entrada na amostra.

Uma das razões para a frequente ausência de preocupação com o viés de atrito é que, de acordo com Verbeek e Nijman (1992) e Ziliak e Kniesner (1998), a estimação de um modelo de efeitos fixos pode eliminar muita das formas de heterogeneidade não observada. Entre elas, a heterogeneidade na seleção amostral. Contudo, Vella (1998) salienta que outras formas de viés de seleção e heterogeneidade podem não ser eliminadas com o uso de um modelo de efeitos fixos. Este é o caso quando a seleção da amostra está relacionada com os termos idiossincráticos aleatórios da equação de interesse.

Além disso, para estimação de um modelo de efeitos fixos é necessária uma amostra com unidades acompanhadas em três ou mais períodos. Nicoletti e Peracchi (2001), por exemplo, não utilizam um modelo de efeitos fixos para estimar transições na força de trabalho no BHPS (*British Household Panel Survey*). Eles concluem que um choque na probabilidade de permanência na amostra é positivamente correlacionado com um choque na probabilidade de desemprego. A explicação é que o atrito é mais comum entre aqueles que se mudaram por razões relacionadas ao trabalho, sendo impossível estabelecer um novo contato.

Van den Berg e Lindeboom (1998) também estimam a relação entre desgaste do painel e emprego/desemprego, porém utilizando dados do *Labour Supply Panel*

2 Como exemplos de trabalhos que não utilizam técnicas de correção para o viés de atrito podemos citar os de Duryea (1998), Duryea, Lam e Levison (2007), Neri *et al.* (2000), Corseuil e Carneiro (2001), Lemos (2002), Woltermann (2002), Penido e Machado (2003), Gonzaga e Reis (2005), Machado, Ribas e Penido (2007).

Survey da OSA (*Netherlands Organization for Strategic Labour Market Research*). Com uma taxa de atrito elevada, entre 25% e 30%, eles evidenciam uma forte correlação entre os componentes não observados determinantes do tempo de emprego e desemprego e os componentes não observados determinantes do tempo de participação no painel.

Um ponto importante levantado por Falaris e Peters (*op. cit.*) é que, mesmo quando a subamostra de atritados apresenta estatísticas semelhantes às dos demais, a estimação sem os termos de correção pode incorrer em resultados viesados. Zabel (1998), por outro lado, apresenta um caso onde o padrão de atrito é contrário deste último. A primeira evidência foi que atritados apresentavam maior variabilidade em suas horas trabalhadas e em seus salários e menor taxa de participação na força de trabalho que os não atritados. Além disso, as estimativas da equação de oferta de trabalho eram distintas entre as amostras de atritados e não atritados. Contudo, comparando as regressões com e sem o termo de correção, os resultados são de que não há viés de atrito na equação de horas de trabalho ofertadas, mas somente na equação de salários.

Outra forma de seleção amostral ocorre quando pessoas não desaparecem do painel, mas algumas variáveis se tornam não observadas em algum dos períodos. Um exemplo é a estimação de uma equação salarial usando um painel de indivíduos. A população de interesse são pessoas que estão trabalhando no período inicial, porém algumas dessas pessoas poderão ficar desempregadas em períodos subsequentes. Esta situação é diferente de um problema de desgaste, onde pessoas saem completamente da amostra e, normalmente, não reaparecem nos períodos seguintes.

Cappelari e Jenkins (2004a, 2004b) investigam o viés de atrito sobre duas equações um pouco distintas, e ambas envolvem transições entre estados discretos e outros mecanismos de seleção. No primeiro caso (2004b), que investiga as transições entre baixos e altos salários considerando a probabilidade de o adulto estar empregado, a conclusão é que as correções de viés a partir da condição inicial e seletividade do mercado de trabalho são relevantes. A existência de um viés de atrito, porém, não passa no teste de hipótese. No segundo trabalho (2004a), que investiga as probabilidades de permanência na pobreza ou transição para ela, os autores encontram evidências sobre o viés de atrito e de condição inicial. Porém, eles concluem que negligenciar a endogeneidade da condição inicial é mais problemático que negligenciar o controle sobre o desgaste do painel.

Lillard e Panis (1998), que estimaram os efeitos do estado civil sobre a dinâmica da renda familiar e sobre o risco de morte utilizando o PSID, tiveram uma conclusão semelhante à de Cappelari e Jenkins (2004a). Apesar da evidência de significati-

va seletividade no atrito, o viés introduzido ao ignorar o problema é muito baixo. Cabe salientar que a taxa de atrito neste trabalho também era baixa, apenas 2% por ano.

Fitzgerald, Gottschalk e Moffitt (1998) colocam que não há necessariamente uma relação entre tamanho da amostra atritada e magnitude do viés gerado. Elevadas taxas de atrito podem não causar viés se elas forem aleatórias.

2.1 Como Controlar a Seleção Não Aleatória

As formas de não-aleatoriedade do atrito que causam viés na estimação são apresentadas a seguir. Considere um modelo geral onde y_{it} é a variável de interesse e d_{it} é uma *dummy* igual a 1 se a unidade permaneceu na amostra do período, ou seja, y_{it} é observado, e igual a 0 caso contrário, tal que:

$$y_{it}^* = x'_{1,it} \beta_1 + \mu_{1,i} + e_{1,it} \quad (1)$$

$$d_{it}^* = x'_{1,it} \beta_2 + x'_{2,it} \gamma_2 + \mu_{2,i} + e_{2,it} \quad (2)$$

$$d_{it} = \begin{cases} 1 & \text{se } d_{it}^* > 0 \\ 0 & \text{se } d_{it}^* \leq 0 \end{cases} \quad (3)$$

$$y_{it} = y_{it}^* \cdot d_{it} \quad (4)$$

onde i , ($i = 1, \dots, N$), representa o indivíduo e t , ($t = 1, \dots, T$), representa o período. A variável dependente na primeira equação é observada somente para as unidades que satisfazem a regra de seleção ($d_{it}^* > 0$). Para introduzir o viés de seleção é assumido que os erros podem ser divididos em efeitos individuais fixos ($\mu_{1,i}$ e $\mu_{2,i}$) e efeitos aleatórios idiossincráticos ($e_{1,it}$ e $e_{2,it}$), e que cada componente do erro é assumido ser normalmente distribuído e correlacionado com o componente da mesma dimensão na outra equação.³ Para simplificar, definimos que $\varepsilon_{1,it} = \mu_{1,i} + e_{1,it}$ e $\varepsilon_{2,it} = \mu_{2,i} + e_{2,it}$.

Uma forma de atrito não aleatório é quando a seleção ocorre sobre os componentes específicos e invariantes dos indivíduos:

3 Para simplificar, os efeitos aleatórios agregados de períodos estão junto das variáveis explicativas do modelo.

$$e_{2,it} \perp e_{1,it} \text{ e } x_{2,it} \perp \varepsilon_{1,it}, \text{ mas } \varepsilon_{2,it} \not\perp \varepsilon_{1,it} \text{ porque } \mu_{2,i} \not\perp \mu_{1,i} \quad (5)$$

Neste caso, não é necessário utilizar técnicas de correção de viés, desde que seja utilizado um modelo de estimação com efeitos fixos. Porém, cabe salientar que, caso não seja possível estimar os efeitos fixos, não é possível isolar os termos $\mu_{1,i}$ de $\varepsilon_{1,it}$ e $\mu_{2,i}$ de $\varepsilon_{2,it}$. Dessa forma, o comportamento de atrito seria tratado como dentro de um segundo grupo de estimativas, juntamente com aquelas onde a seleção ocorre sobre os componentes não observados, tal que:

$$x_{2,it} \perp \varepsilon_{1,it}, \text{ mas } \varepsilon_{2,it} \not\perp \varepsilon_{1,it} \text{ porque } e_{2,it} \not\perp e_{1,it} \quad (6)$$

Neste caso, a esperança condicional de y_{it} na amostra não atritada será:

$$\begin{aligned} E[y_{it} | x'_{1,it}, x'_{2,it}, d_{it} = 1] &= x'_{1,it}\beta_1 + E[\varepsilon_{1,it} | x_{1,it}, x_{2,it}, \varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2] \\ &= x'_{1,it}\beta_1 + h(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2) \end{aligned} \quad (7)$$

onde h é uma função com parâmetros desconhecidos que pode ser escrita como:

$$h(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2) = \rho_{12}\sigma_1 E[\varepsilon_{2,it} | \varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2] \quad (8)$$

sendo ρ_{12} a correlação entre $\varepsilon_{1,it}$ e $\varepsilon_{2,it}$ e σ_1 o desvio padrão de $\varepsilon_{1,it}$. Para se chegar à função (7), a distribuição conjunta de $\varepsilon_{1,it}$ e $\varepsilon_{2,it}$ deve ser independente de $x_{1,it}$ e $x_{2,it}$.

Aplicações deste modelo geralmente assumem uma distribuição bivariada para $\varepsilon_{1,it}$ e $\varepsilon_{2,it}$. Com as estimativas dos parâmetros da equação de retenção (ou atrito) ou com a probabilidade predita de retenção (ou atrito), a equação (7) se torna uma função em que os parâmetros podem ser consistentemente estimados.

Basicamente, existem duas formas de estimação neste caso, maximizando uma única função de verossimilhança com componentes de distribuição multivariada ou utilizando a estimação em dois estágios proposta inicialmente por Heckman (1979). Mesmo atualmente, processos de estimação baseados na maximização de uma função de verossimilhança única são difíceis de se trabalhar, particularmente quando são mais de um os mecanismos de seleção (ARENDDT; HOLM, 2006). Por isso,

alguns trabalhos ainda preferem a estimação em dois estágios.⁴ Segundo Nicoletti e Perracchi (*op. cit.*) e Arendt e Holm (*op. cit.*), este tipo de estimação também gera resultados consistentes quando a variável de interesse é binária. Cabe salientar que, de acordo com Bourguignon *et al.* (2007), quando o mecanismo de seleção não é uma variável binária, o que não é geralmente o caso de atrito/permanência no painel, a estimação em dois estágios pode gerar resultados enviesados.

Independente da forma de estimação, a identificação de β_1 implica uma restrição de exclusão, que exista $x_{2,it}$ satisfazendo a propriedade de independência de $\varepsilon_{1,it}$ e na qual γ_2 não seja igual a zero. Segundo Fitzgerald *et al.* (*op. cit.*), encontrar tal variável instrumental é mais difícil para os casos de não-resposta que em outras aplicações, pois são poucas as variáveis que afetam não-resposta que podem ser confiavelmente excluídas da equação de interesse. Para os autores, características individuais que são potenciais fontes de instrumentos geralmente estão relacionadas à variável comportamental y_{it} . As variáveis instrumentais mais promissoras são aquelas externas ao indivíduo, tais como características do entrevistador ou do processo de entrevista.

Outra forma de atrito não aleatório, com seleção amostral sobre observáveis, ocorre quando:

$$\varepsilon_{2,it} \perp \varepsilon_{1,it}, \text{ mas } x_{2,it} \not\perp \varepsilon_{1,it} \quad (9)$$

Neste caso, a variável crítica é $x_{2,it}$, que afeta a probabilidade de atrito, mas está, ao mesmo tempo, relacionada à densidade de y_{it} condicionada à $x_{1,it}$. Ou seja, $x_{2,it}$ é endógena à y_{it} . Dessa forma, uma estimação por mínimos quadrados da equação (1) sobre uma amostra de não atritados geraria estimadores inconsistentes de β_1 e uma densidade $g(y_{it} | x_{1,it}, d_{it} = 1)$ que não corresponde à densidade $f(y_{it} | x_{1,it})$ da população, pois a permanência na amostra está relacionada à y_{it} através de $x_{2,it}$. Esta situação em particular ocorre quando há um interesse em investigar somente o efeito de $x_{1,it}$ sobre y_{it} , sem condicioná-lo a $x_{2,it}$, pois a inclusão de $x_{2,it}$ na equação de interesse iria enviesar a estimativa do efeito desejado. Fitzgerald *et al.* (*ibidem*) afirmam

4 Sobre estimação em dois estágios quando há mais de um mecanismo de seleção da amostra, ver Fische *et al.* (1981). Sobre métodos que corrigem viés de seleção sobre não observáveis, ver Vella (*op. cit.*).

que, em um painel, os valores defasados da variável de interesse, y_{it-1} , y_{it-2} , etc., podem cumprir o papel de variável endógena auxiliar $x_{2,it}$.

A solução para eliminar o viés de atrito sobre observáveis é a utilização do método de Ponderação pela Probabilidade Inversa (IPW). Sendo $f(y_{it}, x_{2,it} | x_{1,it})$ a densidade conjunta de y_{it} e $x_{2,it}$ da população e $g(y_{it}, x_{2,it} | x_{1,it}, d_{it} = 1)$ a densidade da amostra de não atritados, então:

$$\begin{aligned} g(y_{it}, x_{2,it} | x_{1,it}, d_{it} = 1) &= \frac{g(y_{it}, x_{2,it}, d_{it} = 1 | x_{1,it})}{\Pr(d_{it} = 1 | x_{1,it})} \\ &= \frac{\Pr(d_{it} = 1 | y_{it}, x_{1,it}, x_{2,it}) f(y_{it}, x_{2,it} | x_{1,it})}{\Pr(d_{it} = 1 | x_{1,it})} \\ &= \frac{\Pr(d_{it} = 1 | x_{1,it}, x_{2,it}) f(y_{it}, x_{2,it} | x_{1,it})}{\Pr(d_{it} = 1 | x_{1,it})}. \end{aligned}$$

Logo, $f(y_{it}, x_{2,it} | x_{1,it}) = g(y_{it}, x_{2,it} | x_{1,it}, d_{it} = 1) \cdot w(x_{1,it}, x_{2,it})$ que, integrando dos dois lados em relação a $x_{2,it}$, é reescrito como:

$$f(y_{it} | x_{1,it}) = \int_{x_{2,it}} g(y_{it}, x_{2,it} | x_{1,it}, d_{it} = 1) w(x_{1,it}, x_{2,it}) dx_{2,it} \quad (10)$$

onde

$$w(x_{1,it}, x_{2,it}) = \left[\frac{\Pr(d_{it} = 1 | x_{1,it}, x_{2,it})}{\Pr(d_{it} = 1 | x_{1,it})} \right]^{-1} \quad (11)$$

é o peso normalizado a ser utilizado na regressão sobre a amostra de não atritados da equação (1) para estimar $f(y_{it} | x_{1,it})$.⁵

5 Wooldridge (2001) aponta que o método IPW pode ser aplicado a qualquer estimador M , o que inclui modelos não lineares. As propriedades assintóticas do método IPW são demonstradas por este autor.

3 O Painel da PME

A Pesquisa Mensal de Emprego (PME) do Instituto Brasileiro de Geografia e Estatística (IBGE) é uma pesquisa amostral domiciliar de periodicidade mensal que abrange seis Regiões Metropolitanas (RMs) brasileiras – Rio de Janeiro, São Paulo, Porto Alegre, Belo Horizonte, Recife e Salvador.

De janeiro de 1980 a janeiro de 1982, a PME era parte integrante da Pesquisa Nacional por Amostra de Domicílios (PNAD), utilizando-se do mesmo desenho amostral. Em fevereiro de 1982, deu início à implantação gradativa de reformulações na amostra e, em maio daquele mesmo ano, introduziu-se também um novo e mais amplo questionário. Este questionário deveria ser respondido por pessoas com 10 ou mais anos de idade residentes no domicílio selecionado para a amostra. Para obter resultados para cada RM separadamente, a PME incorporou aspectos de um plano amostral autoponderado. Dentro de cada RM, a probabilidade de um domicílio ou de uma pessoa qualquer pertencer à amostra é constante e igual à fração amostral.

Para atender às mudanças na estrutura produtiva, na alocação da mão de obra e nas relações de trabalho, juntamente com as recomendações da Organização Internacional do Trabalho (OIT) para a investigação da força de trabalho, o IBGE realizou, no início da década de 2000, uma revisão da PME em todos os seus aspectos, metodológicos e processuais. Assim, em março de 2002, foi a campo uma nova PME.⁶ Em relação ao desenho amostral, aumentou-se em quase 500 o número de setores selecionados e calculou-se uma nova fração amostral, que diminuiu a probabilidade de um domicílio qualquer pertencer à amostra. Por consequência, a nova pesquisa reduziu em quase 3 mil casos o número de unidades domiciliares selecionadas. Em relação ao questionário, houve uma ampliação e algumas seções passaram a ser respondidas também por pessoas com menos de 10 anos de idade.

Tanto na antiga PME, que foi realizada até dezembro de 2002, como na nova PME, levantada a partir de março de 2002, existe um esquema de rotação de painéis que entrevista o mesmo grupo de domicílio durante certo período. A principal razão para este acompanhamento é a maior segurança nas comparações mensais dos resultados sobre o mercado de trabalho, garantindo que as variações verificadas não estão sendo provocadas pela troca de informantes. Mesmo assim, para minimizar o cansaço imposto aos informantes, os mesmos domicílios não são entrevistados

6 Os principais objetivos da revisão foram: implementação de algumas mudanças conceituais no tema trabalho; ampliação da investigação para se ter melhor conhecimento da População Economicamente Ativa (PEA) e da População em Idade Ativa (PIA); e melhor operacionalização dos quesitos para captação das informações de forma a aprimorar a mensuração dos fenômenos (IBGE, 2002).

durante todos os meses de pesquisa, mas eles entram e saem da amostra de acordo com um padrão pré-definido (IBGE, 2002).

Na antiga PME, um painel equivale a um conjunto de domicílios selecionados que é dividido em quatro grupos rotacionais correspondentes cada qual a uma remessa de setores entrevistados em uma semana específica do mês. Os grupos rotacionais são indicados por uma letra, que identifica o painel, acompanhada de um subscrito correspondente à semana do mês. Se no mês t , por exemplo, for aplicado o painel B (B1, B2, B3, B4), no mês $t + 1$ será aplicado apenas 75% do seu todo (B1, B2, B3), entrando um quarto do painel seguinte C (C4), e assim sucessivamente. Assim, há a garantia de que 75% dos domicílios são comuns em dois meses consecutivos.

O esquema de rotação, chamado 4-8-4, determinava que, de outubro de um ano ímpar a setembro do ano seguinte, todo mês um grupo de domicílios entrava na pesquisa e era entrevistado por quatro meses consecutivos. Do quinto ao décimo segundo mês, este grupo saía da amostra, retornando no décimo terceiro mês e sendo entrevistado por mais quatro vezes. Os domicílios saíam da amostra definitivamente dezesseis meses depois de sua entrada. Importante salientar que, de outubro de um ano par a setembro do ano seguinte, nenhum grupo novo de domicílios entrava na amostra. Neste período, a cada mês, apenas retornavam os grupos entrevistados que estavam no intervalo de oito meses sem entrevista. Por consequência, a cada par de anos, 100% da amostra se repetia.

No caso da nova PME, o padrão 4-8-4 foi mantido, mas houve um ajustamento no processo de rotação para dar mais condições de acompanhamento longitudinal dos resultados. Os grupos rotacionais foram sincronizados de tal forma que não aconteça o chamado *blackout* a cada dois anos, quando toda a amostra era trocada. Dessa forma, aumentou-se o número de grupos rotacionais de quatro para oito, rodando dois grupos por mês ao invés de apenas um, como era anteriormente.

Conhecer o esquema de rotação da PME e as variáveis que identificam o mesmo domicílio na pesquisa é o primeiro passo para reconstituir o painel de indivíduos ou famílias a partir dos microdados. Ademais, Ribas e Soares (2008) apontam outros problemas que dificultam esta reconstituição. Em decorrência dos domicílios que, em períodos subsequentes, passam a ser inexistentes, estão fechados ou cujos habitantes se recusaram a responder à pesquisa, o emparelhamento dos domicílios não é perfeito. Como a identificação do domicílio é precisa, nada se pode fazer com relação a esta forma de desgaste no painel. A segunda forma de desgaste – pessoas não emparelhadas porque seu domicílio não foi entrevistado ou porque elas não faziam parte do domicílio no momento da entrevista – é obviamente maior.

De acordo com Ribas e Soares (*ibidem*), em ambas as PMEs, as taxas de atrito no intervalo de um mês são em torno de 4% para domicílios e de 13% para indivíduos, sendo crescentes com o aumento no intervalo dos meses. No intervalo de 12 meses, a perda proporcional de domicílio chega a perto de 10%. Porém, as perdas relativas de indivíduos são muito mais altas na PME nova a partir dos nove meses de intervalo, com mais de 50% de perda, chegando a quase 70% no 16º mês. Na antiga PME, a taxa de atrito de indivíduos varia de 20% a 30% a partir dos nove meses de intervalo.

No entanto, Ribas e Soares (*ibidem*) apontam, ainda, que estas taxas de atrito para indivíduos estão possivelmente contaminadas com o chamado “falso atrito”. Isso ocorre porque a informação utilizada para o emparelhamento de indivíduos pode não ser tão precisa quanto a informação utilizada no emparelhamento de domicílios. Assim, a sobreposição de pessoas pode ser subestimada, desde que haja algum erro na informação individual reportada em algum dos meses. Para minimizar este problema, os autores propõem um algoritmo para reconstituição do painel que utiliza critérios não só de exatidão nas variáveis de identificação, mas também de proximidade nas respostas. Após a utilização deste algoritmo, é possível recuperar, em ambas PMEs, entre 7 e 10 pontos percentuais da amostra nos primeiros meses de intervalo. Nos intervalos maiores, recuperaram-se entre 5 e 6 pontos percentuais da amostra na antiga PME e entre 15 e 25 pontos percentuais da amostra na nova PME.

Para evitar o problema de ‘falso atrito’ causado por erro de informação, todos os painéis utilizados neste artigo foram reconstituídos através do algoritmo proposto por Ribas e Soares (*ibidem*). Para maiores informações sobre o painel da PME, consultar IBGE (1998, 2002), Lopes (*op. cit.*) e Ribas e Soares (*op. cit.*).

4 Estimativa dos Determinantes da Permanência no Painel da PME

Peracchi e Welch (1995) analisam o problema do desgaste na CPS (*Current Population Survey*). A CPS é uma pesquisa mensal semelhante à PME, possuindo oito grupos rotativos de domicílios e um esquema de rotação 4-8-4. Seus resultados apontam que a principal causa do atrito nesta pesquisa é o fracasso em seguir os jovens em idade de frequentar a universidade em domicílios emparelhados e de acompanhar famílias de jovens que mudam de endereço. A conclusão de seu trabalho é que grande parte do desgaste da CPS resulta da mobilidade das pessoas ligada às decisões sobre educação, formação de novas famílias e procura de emprego.

Lopes (*op. cit.*) realiza uma análise semelhante à destes autores, porém na PME. Particularmente, Lopes utiliza dados de março de 1996, 1997, 1998 e 1999 para a Região Metropolitana do Rio de Janeiro para estimar a probabilidade de as pessoas permanecerem na amostra após um ano. Os resultados de seu trabalho apontam que a probabilidade de atrito é crescente até os 29 anos de idade da pessoa, depois ela se torna decrescente, maior entre os desempregados, e crescente com a renda do indivíduo. Além disso, cinco ou mais anos de estudo e residir em domicílio chefiado por mulher aumentam a probabilidade de permanecer na amostra. A condição de ocupação do chefe não é um fator determinante do atrito do domicílio.

Nesta seção, apresentamos os resultados sobre a probabilidade de as pessoas permanecerem na entrevista seguinte da PME. Os microdados utilizados são da antiga pesquisa⁷, referentes aos anos de 1997, 1998 e 1999 para as seis Regiões Metropolitanas (RMs). A amostra é composta por todas as pessoas com 10 anos de idade ou mais inquiridas entre a primeira e a sétima entrevista. Além disso, considerando que existe uma correlação entre pessoas vivendo em um mesmo domicílio, as seguintes estimativas foram obtidas delimitando cada família como um cluster.⁸

A Tabela 1 apresenta as estimativas dos efeitos marginais sobre a probabilidade de permanência na amostra, *vis-à-vis* o atrito. Podemos constatar que, dadas constantes as demais características observáveis, a probabilidade de desgaste é maior na RM de Recife (categoria de referência), seguida pelas RMs do Rio de Janeiro, São Paulo e Salvador. Para indivíduos de mesmas características, os com maiores probabilidades de permanência na amostra ao longo do painel residiam nas RMs de Belo Horizonte e Porto Alegre.

Em relação ao período de entrevista, os coeficientes mostram que o desgaste no painel tende a ser decrescente até o mês de agosto, com exceção para os meses de maio e junho, e crescente a partir de setembro. Dezembro é o mês no qual a probabilidade de desgaste é maior, seguido de novembro, janeiro e fevereiro. Uma possível razão para este resultado pode ser a ausência de pessoas nos domicílios para responder à entrevista nestes meses, considerados como dentro de um período de férias para algumas pessoas e de aumento de ocupações sazonais para outras.

O aumento de 1% no número de membros dividindo o mesmo domicílio aumenta em 1,4 pontos percentuais a probabilidade de pessoas permanecerem na amostra. Além disso, a probabilidade de atrito é menor nos domicílios com maior proporção de jovens e adolescentes. Claro que este resultado pode ser decorrente da pró-

7 A PME com a antiga metodologia foi a campo até dezembro de 2002. A partir de março de 2002, outra PME, com uma nova metodologia, passou a ser levantada (IBGE, 2002).

8 Uma alternativa seria estimar cada modelo utilizando efeitos fixos por família.

pria correlação entre pessoas de uma mesma família. Corroborando o resultado de Lopes, identificamos que o desgaste é maior para os indivíduos entre 25 e 29 anos, sendo decrescente a partir desta idade. Podemos identificar também que as pessoas com ensino fundamental completo (oito anos de estudo ou mais), assim como as mulheres, são as com menor probabilidade de atrito.

De acordo com Justo e Silveira Neto (2008), homens, jovens, com pouca escolaridade e residentes em domicílios menores, geralmente unipessoais, compõem o grupo de pessoas mais propenso à migração. Ou seja, parte das características que estão relacionadas ao atrito está também relacionada à mobilidade geográfica. Contudo, salientamos que a mobilidade geográfica não é necessariamente a principal razão para o atrito. Outras razões incluem, por exemplo, o falecimento do indivíduo e a mudança no estado conjugal. Ainda nesse sentido, um resultado esperado é de que as pessoas do núcleo familiar (chefe, cônjuge e filho), principalmente o cônjuge, são as com maiores chances de permanência no painel.

Assim como para a idade da pessoa, a probabilidade de permanência no painel é crescente com a idade do chefe do domicílio a partir dos 30 anos e maior caso ele seja homem. Por outro lado, controlados os demais fatores, a escolaridade do chefe do domicílio não é um fator significativamente determinante do desgaste no painel.

Tabela 1 – Efeitos Marginais sobre a Probabilidade de Permanecer na Amostra na Entrevista Seguinte para a Antiga PME

Variável	(1)	(2)	(3)	(4)	(5)	(6)
RM Salvador	0.02071 ***	0.02009 ***	0.02010 ***	0.02012 ***	0.02079 ***	0.02072 ***
RM Belo Horizonte	0.03213 ***	0.01922 ***	0.01920 ***	0.01926 ***	0.03231 ***	0.03193 ***
RM Rio de Janeiro	0.01780 ***	0.01791 ***	0.01788 ***	0.01789 ***	0.01789 ***	0.01764 ***
RM São Paulo	0.01847 ***	0.01870 ***	0.01880 ***	0.01882 ***	0.01866 ***	0.01783 ***
RM Porto Alegre	0.02770 ***	0.02723 ***	0.02723 ***	0.02723 ***	0.02785 ***	0.02738 ***
mês fevereiro	0.00194	0.00207 *	0.00235 *	0.00211 *	0.00190	0.00196
mês março	0.00582 ***	0.00562 ***	0.00591 ***	0.00547 ***	0.00587 ***	0.00588 ***
mês abril	0.00902 ***	0.00846 ***	0.00866 ***	0.00806 ***	0.00903 ***	0.00909 ***
mês maio	0.00672 ***	0.00685 ***	0.00690 ***	0.00611 ***	0.00677 ***	0.00674 ***
mês junho	0.00720 ***	0.00726 ***	0.00754 ***	0.00586 ***	0.00723 ***	0.00724 ***
mês julho	0.00979 ***	0.00990 ***	0.01014 ***	0.00774 ***	0.00981 ***	0.00986 ***
mês agosto	0.01304 ***	0.01232 ***	0.01252 ***	0.00943 ***	0.01304 ***	0.01310 ***
mês setembro	0.01153 ***	0.01155 ***	0.01164 ***	0.00781 ***	0.01153 ***	0.01155 ***
mês outubro	0.00760 ***	0.00793 ***	0.00820 ***	0.00537 ***	0.00762 ***	0.00763 ***

(continua...)

Variável	(1)	(2)	(3)	(4)	(5)	(6)
mês novembro	-0.00020	0.00016	0.00047	-0.00149	-0.00021	-0.00019
mês dezembro	-0.00553 ***	-0.00572 ***	-0.00539 ***	-0.00632 ***	-0.00556 ***	-0.00555 ***
log do número de membros do domicílio	0.01446 ***	0.01423 ***	0.01416 ***	0.01416 ***	0.01451 ***	0.01445 ***
Proporção de						
crianças com menos de 10 anos	-0.00316	-0.00224	-0.00211	-0.00208	-0.00360	-0.00275
adolescente entre 10 e 15 anos	0.01222 ***	0.01264 ***	0.01284 ***	0.01286 ***	0.01234 ***	0.01232 ***
jovens entre 16 e 18 anos	0.01065 ***	0.01086 ***	0.01092 ***	0.01086 ***	0.01095 ***	0.01097 ***
Idade da pessoa						
entre 16 e 18 anos	-0.00861 ***	-0.00796 ***	-0.00793 ***	-0.00792 ***	-0.00668 ***	-0.00853 ***
entre 19 e 24 anos	-0.01775 ***	-0.01673 ***	-0.01668 ***	-0.01667 ***	-0.01341 ***	-0.01766 ***
entre 25 e 29 anos	-0.01802 ***	-0.01704 ***	-0.01702 ***	-0.01702 ***	-0.01319 ***	-0.01800 ***
entre 30 e 34 anos	-0.01032 ***	-0.00936 ***	-0.00938 ***	-0.00938 ***	-0.00568 ***	-0.01035 ***
entre 35 e 39 anos	-0.00671 ***	-0.00641 ***	-0.00640 ***	-0.00641 ***	-0.00233	-0.00673 ***
entre 40 e 44 anos	-0.00330 *	-0.00332 *	-0.00332 *	-0.00333 *	0.00082	-0.00325 *
entre 45 e 49 anos	-0.00112	-0.00128	-0.00130	-0.00131	0.00263	-0.00102
entre 50 e 54 anos	0.00128	0.00118	0.00112	0.00109	0.00431 *	0.00136
entre 55 e 59 anos	0.00694 ***	0.00643 ***	0.00638 ***	0.00636 ***	0.00919 ***	0.00697 ***
entre 60 e 64 anos	0.00878 ***	0.00815 ***	0.00809 ***	0.00806 ***	0.01025 ***	0.00881 ***
com 65 anos ou mais	0.01058 ***	0.01017 ***	0.01013 ***	0.01012 ***	0.01109 ***	0.01053 ***
Mulher	0.00172 **	0.00177 **	0.00175 **	0.00175 **	0.00061	0.00174 **
Escolaridade da pessoa						
de 1 a 3 anos de estudo	-0.00240	-0.00242 *	-0.00241 *	-0.00241 *	-0.00209	-0.00246 *
de 4 a 7 anos de estudo	0.00199	0.00156	0.00153	0.00153	0.00242 *	0.00181
de 8 a 10 anos de estudo	0.00531 ***	0.00452 ***	0.00444 ***	0.00442 ***	0.00594 ***	0.00505 ***
11 ou mais anos de estudo	0.00573 ***	0.00487 ***	0.00478 ***	0.00476 ***	0.00715 ***	0.00556 ***
Condição no domicílio						
cônjuge	0.00432 ***	0.00398 ***	0.00399 ***	0.00399 ***	0.00264 **	0.00425 ***
filho	-0.00139	-0.00163	-0.00166	-0.00168	-0.00176	-0.00137
agregado	-0.03294 ***	-0.03175 ***	-0.03178 ***	-0.03178 ***	-0.03331 ***	-0.03297 ***
empregado ou pensionista	-0.07496 ***	-0.07320 ***	-0.07340 ***	-0.07335 ***	-0.07294 ***	-0.07438 ***
Idade do chefe						
entre 16 e 18 anos	-0.01993	-0.01475	-0.01492	-0.01461	-0.01965	-0.01911
entre 19 e 24 anos	0.00492	0.00815	0.00835	0.00850	0.00447	0.00529
entre 25 e 29 anos	0.01811	0.02028 *	0.02044 *	0.02056 *	0.01742	0.01821
entre 30 e 34 anos	0.02368 *	0.02501 **	0.02524 **	0.02536 **	0.02288 *	0.02382 *
entre 35 e 39 anos	0.02944 **	0.03038 ***	0.03052 ***	0.03065 ***	0.02867 **	0.02952 **
entre 40 e 44 anos	0.03313 ***	0.03386 ***	0.03401 ***	0.03414 ***	0.03229 ***	0.03311 ***
entre 45 e 49 anos	0.03535 ***	0.03582 ***	0.03598 ***	0.03609 ***	0.03446 ***	0.03534 ***

(continua...)

Variável	(1)	(2)	(3)	(4)	(5)	(6)
entre 50 e 54 anos	0.03882 ***	0.03869 ***	0.03880 ***	0.03890 ***	0.03798 ***	0.03872 ***
entre 55 e 59 anos	0.03924 ***	0.03917 ***	0.03927 ***	0.03937 ***	0.03841 ***	0.03905 ***
entre 60 e 64 anos	0.04098 ***	0.04056 ***	0.04068 ***	0.04077 ***	0.04016 ***	0.04062 ***
com 65 anos ou mais	0.04707 ***	0.04664 ***	0.04670 ***	0.04678 ***	0.04618 ***	0.04639 ***
Chefe mulher	-0.00435 ***	-0.00433 ***	-0.00432 ***	-0.00430 ***	-0.00424 ***	-0.00481 ***
Escolaridade do chefe						
de 1 a 3 anos de estudo	0.00309 **	0.00279 **	0.00271 **	0.00268 **	0.00284 **	0.00292 **
de 4 a 7 anos de estudo	0.00158	0.00132	0.00128	0.00126	0.00120	0.00124
de 8 a 10 anos de estudo	0.00169	0.00142	0.00130	0.00128	0.00112	0.00121
11 ou mais anos de estudo	-0.00216 *	-0.00233 *	-0.00249 *	-0.00252 *	-0.00306 **	-0.00238 *
Domicílio na						
2ª entrevista	(i)	0.00559 ***	0.00563 ***	0.00561 ***		
3ª entrevista	(ii)	0.01019 ***	0.01023 ***	0.01017 ***		
4ª entrevista	(iii)	-0.08294 ***	-0.08297 ***	-0.08314 ***		
5ª entrevista	(iv)	0.01149 ***	0.01147 ***	0.01137 ***		
6ª entrevista	(v)	0.01437 ***	0.01433 ***	0.01426 ***		
7ª entrevista	(vi)	0.01413 ***	0.01406 ***	0.01396 ***		
Entrevista ocorrida na						
2ª semana	(vii)		-0.00136 *	-0.00132 *		
3ª semana	(viii)		-0.00315 ***	-0.00308 ***		
4ª semana	(ix)		-0.00799 ***	-0.00792 ***		
Painel H	(x)			0.00084		
Painel I	(xi)			0.00424 ***		
Pessoa ocupada	(xii)				-0.00760 ***	
Pessoa desempregada	(xiii)				-0.01022 ***	
Setor de ocupação						
indústria	(xiv)				0.00165	
construção civil	(xv)				-0.00486 ***	
comércio	(xvi)				-0.00258 **	
outro	(xvii)				-0.00422 ***	
Condição na ocupação						
conta própria	(xviii)				0.00479 ***	
empregador	(xix)				0.00120	
não remunerado	(xx)				0.01693 ***	
tem carteira assinada	(xxi)				0.00258 **	
chefe ocupado	(xxii)					-0.00365 ***
chefe desempregado	(xxiii)					-0.00978 ***
Setor de ocupação do chefe						
indústria	(xxiv)					0.00472 ***

(continua...)

Variável	(1)	(2)	(3)	(4)	(5)	(6)	
construção civil	(xxv)					-0.00404 ***	
comércio	(xxvi)					-0.00177 *	
outro	(xxvii)					-0.00497 ***	
Condição na ocupação do chefe							
conta própria	(xxviii)					0.00289 ***	
empregador	(xxix)					-0.00024	
não remunerado	(xxx)					0.01730 *	
tem carteira assinada	(xxxi)					0.00347 ***	
probabilidade média predita	0.93556	0.94004	0.94010	0.94012	0.93562	0.93561	
Número de observações	1716958	1716958	1716958	1716958	1716375	1716485	
Log likelihood	-417026.6	-403527.2	-403391.6	-403366.6	-416701.3	-416777.8	
Pseudo R2	0.0202	0.0519	0.0523	0.0523	0.0206	0.0206	
Wald test	chi	d.f.	modelo	Wald test	chi	d.f.	modelo
(i) = (ii) = ... = (vi) = 0	20123.04	6	*** (2)	(vii) = (viii) = (ix) = 0	195.82	3	*** (3)
(i) = (ii) = 0	163.16	2	*** (2)	(i) = (ii) = ... = (ix) = 0	20333.01	9	*** (3)
(iii) = 0	7174.66	1	*** (2)	(x) = (xi) = 0	36.09	2	*** (4)
(iv) = (v) = (vi)	19.66	2	*** (2)	(i) = (ii) = ... = (xi) = 0	20372.23	11	*** (4)
(xii) = (xiii) = 0	124.68	2	*** (5)	(xxii) = (xxiii) = 0	53.85	2	*** (6)
(xiv) = (xv) = (xvi) = (xvii) = 0	35.72	4	*** (5)	(xxiv) = (xxv) = (xxvi) = (xxvii) = 0	102.52	4	*** (6)
(xviii) = (xix) = (xx) = 0	55.04	3	*** (5)	(xxviii) = (xxix) = (xxx) = 0	21.00	3	*** (6)
(xxi) = 0	10.69	1	** (5)	(xxxi) = 0	22.26	1	*** (6)
(xii) = (xiii) = ... = (xxi) = 0	212.75	10	*** (5)	(xxii) = (xxiii) = ... = (xxxi) = 0	221.96	10	*** (6)

Fonte: PME 1997, 1998, 1999.

Nota: * significativa a 5%; ** significativa a 1%; *** significativa a 0.1%.

Obs: Regressões realizadas utilizando os painéis G, H e I.⁹

Nas regressões apresentadas na Tabela 1, incluímos paulatinamente ainda conjuntos de variáveis relacionadas a características das entrevistas, potenciais instrumentos em modelos de seleção sobre não observáveis, e relacionadas a fatores socioeconômicos, de potencial uso em modelos de seleção sobre observáveis. De acordo com os testes de significância conjunta, o número da entrevista no domicílio é o fator, entre todos os demais incluídos, que mais explica a probabilidade de permanência no painel. Contudo, isso é decorrente do elevado desgaste entre a quarta e a quinta entrevista, que possuem um intervalo de oito meses entre si.

Os domicílios na quarta entrevista possuem uma probabilidade de permanência no painel 8 pontos percentuais menor que os domicílios na primeira entrevista. Com

⁹ A numeração do painel segue a numeração oficial da pesquisa estabelecida em IBGE (1998).

exceção da quarta entrevista, a probabilidade de permanência é crescente com o tempo de participação no painel. Isso, a princípio, refuta a hipótese de que as famílias se cansam e se recusam a responder à entrevista com o passar do tempo, ao menos na PME. Contudo, este resultado pode ser, na verdade, decorrente da heterogeneidade entre os indivíduos. Supondo que, ao longo do painel, o percentual de pessoas impacientes diminua e o de pessoas pacientes aumente conseqüentemente, essa mudança na composição da amostra seria a responsável por este resultado.

Em relação à época da entrevista, identificamos que as chances de desgaste são menores na primeira semana e maiores na última semana do mês. Este resultado pode ser explicado por fatores motivacionais tanto por parte dos entrevistados como por parte do entrevistador. Alguns conjuntos da amostra, chamados de “painéis”, possuem ainda maior probabilidade de desgaste que outros. Essa diferença é possivelmente decorrente de mudanças, muitas vezes para melhor, no processo de entrevistas e na orientação repassada aos entrevistadores.

Entre as características socioeconômicas incluídas nas regressões, as relacionadas à ocupação do chefe do domicílio apresentam efeitos semelhantes às relacionadas à ocupação da pessoa em si. As pessoas inativas, assim como as residentes em domicílios chefiados por um inativo, possuem a maior probabilidade de permanência no painel. O desgaste, como esperado, está mais relacionado ao desemprego da pessoa ou do chefe. Entre os trabalhadores ocupados, os de maiores chances de permanecer no painel são os que estão nos setores de serviços (categoria de referência) e indústria. Os ocupados no setor de construção civil ou em outros setores diversos (exclusive comércio), considerados como os mais vulneráveis no mercado de trabalho, são os com maiores chances de atrito. Ainda nesse sentido, os empregados sem carteira assinada possuem maior probabilidade de desgaste que os demais trabalhadores ocupados.¹⁰

5 Comparando Modelos com e sem Controle para Seleção Amostral

Nesta seção, apresentamos os resultados de modelos estimados com e sem controle para seleção amostral, o que inclui, entre outras coisas, a restrição de permanecer no painel. O primeiro exercício é a estimação da probabilidade de transitar do emprego para o desemprego no intervalo de um mês. Para executá-lo, utilizamos a amostra da antiga PME, de 1995 a 2002, para a RM de São Paulo. No segundo

10 De maneira geral, os resultados utilizando a nova PME corroboram aqueles encontrados com a antiga PME. A única mudança significativa nas regressões foi em relação aos coeficientes das *dummies* de RMs. As RMs do Rio de Janeiro e de São Paulo que, na antiga pesquisa, estavam entre as com maiores taxas de atrito, agora se apresentam como as com maiores taxas de permanência no painel.

exercício, utilizando dados da nova PME, de 2002 a 2006, incluindo todas as RMs, realizamos a estimação da probabilidade de progressão escolar para crianças entre 10 e 15 anos de idade. Esta estimação é inspirada nos trabalhos de Duryea (1998) e Duryea, Lam e Levison (2007), que buscam testar o efeito de choques transitórios no domicílio sobre o atraso escolar das crianças.

Ambos os exercícios iniciam com um modelo mais simples, sem controle para seleção amostral, que posteriormente é comparado com modelos de seleção sobre observáveis e sobre não observáveis. Como seleção amostral, consideramos ainda duas formas de ocorrência: possuir as características para pertencer à amostra, o chamado problema de condição inicial, e permanecer no painel, a chamada retenção. Com exceção dos modelos com seleção sobre observáveis, que foram estimados em dois-estágios, todos os outros foram estimados através da maximização de uma única função de verossimilhança; os modelos multivariados com três ou mais componentes foram estimados por máxima verossimilhança simulada.¹¹

5.1 Transição para o desemprego

O modelo de transição para o desemprego é definido como:

$$y_{it}^* | y_{it-1}^* \leq 0 = x'_{1,t-1} \alpha_1 + x'_{1,i} \beta_1 + \varepsilon_{1,it} \quad (12)$$

onde $x_{1,t-1}$ é o vetor que representa as condições no mercado de trabalho no mês anterior, $x_{1,i}$ representa as características idiossincráticas da pessoa i , α_1 e β_1 são os respectivos vetores de coeficientes, $\varepsilon_{1,it}$ é o termo aleatório não explicado da equação, e y_{it}^* e y_{it-1}^* são as variáveis latentes que representam a propensão ao desemprego da pessoa i no mês t e no mês anterior ($t-1$), respectivamente, tal que:

se $y_{it}^* > 0$, $y_{it} = 1$ e a pessoa é observada como desempregada,

se $y_{it}^* \leq 0$, $y_{it} = 0$ e a pessoa é observada como empregada.

Podemos notar que a transição para o desemprego está condicionada a um estado inicial. Ou seja, só há transição quando a pessoa i estiver empregada no período

11 Modelos multivariados com mais de dois componentes endógenos só podem ser estimados utilizando técnicas de simulação ou por cálculo numérico de integrais. No nosso caso, utilizamos o estimador GHK de máxima verossimilhança simulada, proposto por Geweke (1991), Hajivassiliou (1990) e Keane (1994). Este estimador pode ser calculado em Stata através dos algoritmos apresentados por Cappelari e Jenkins (2006).

anterior. Essa condição inicial também pode ser definida através de uma equação, descrita da seguinte forma:

$$y_{it-1}^* = x'_{2,t-1}\alpha_2 + x'_{2,i}\beta_2 + \varepsilon_{2,it-1} \quad (13)$$

onde $x_{2,t-1}$ e $x_{2,i}$ são os vetores que representam, respectivamente, as condições no mercado de trabalho e as características idiossincráticas, α_2 e β_2 são os respectivos vetores de coeficientes, e $\varepsilon_{2,it-1}$ é o termo aleatório não explicado da equação.

Outra restrição inicial imposta à equação (12) é que a pessoa i seja observada em dois pontos no tempo, $t-1$ e t . Essa restrição, chamada de retenção na amostra ou no painel, pode ser definida como:

$$d_{it}^* = x'_{3,t-1}\alpha_3 + x'_{3,i}\beta_3 + \varepsilon_{3,it} \quad (14)$$

onde $x_{3,t-1}$ e $x_{3,i}$ são os vetores que representam, respectivamente, as condições no mercado de trabalho e as características idiossincráticas, α_3 e β_3 são os respectivos vetores de coeficientes, $\varepsilon_{3,it}$ é o termo aleatório não explicado da equação, e d_{it}^* é a variável latente que representa a propensão da pessoa i em permanecer na amostra no tempo t , tal que:

se $d_{it}^* > 0$, $d_{it} = 1$ e a pessoa é observada no tempo t ,

se $d_{it}^* \leq 0$, $d_{it} = 0$ e a pessoa não é observada no tempo t .

Se, de alguma forma, $\varepsilon_{1,it}$ estiver correlacionado a y_{it-1}^* , a condição inicial de emprego será endógena no modelo. Da mesma forma, se $\varepsilon_{1,it}$ estiver correlacionado a d_{it}^* , a permanência no painel será endógena. Se alguma dessas correlações for significativa, pertencer à amostra não é um fenômeno aleatório do ponto de vista da transição para o desemprego. No caso de $\varepsilon_{2,it-1} \perp \varepsilon_{1,it}$, a correlação entre os resíduos é representada por ρ_{12} , e, quando $\varepsilon_{3,it} \perp \varepsilon_{1,it}$, a correlação é representada por ρ_{13} . Ambos os parâmetros, assim como o que representa a correlação entre $\varepsilon_{2,it-1}$ e $\varepsilon_{3,it}$ (ρ_{23}), podem ser estimados conjuntamente com as equações (12), (13) e (14).

Para estimar os determinantes da transição para o desemprego, utilizamos uma amostra de pessoas, entre 18 e 65 anos, da RM de São Paulo, entrevistadas na an-

tiga PME, entre 1995 e 2002. Foram cinco os modelos estimados utilizando esta amostra:

- 1) O modelo (1) é um probit univariado que estima a probabilidade de desemprego a partir de uma amostra de ocupados que permaneceram no painel durante, pelo menos, dois meses seguidos (modelo sem controle de seleção amostral);
- 2) O modelo (2) é um probit bivariado que estima simultaneamente a probabilidade de desemprego e a probabilidade de permanecer no painel, utilizando uma amostra de ocupados no mês anterior (modelo de seleção amostral sobre não observáveis);
- 3) O modelo (3) é um probit bivariado que estima simultaneamente a probabilidade de transição para desemprego e a probabilidade de estar inicialmente empregado, utilizando uma amostra de pessoas que permaneceram no painel durante, pelo menos, dois meses seguidos (modelo de seleção amostral sobre não observáveis);
- 4) O modelo (4) é um probit trivariado que estima simultaneamente a probabilidade de transição para desemprego, a probabilidade de estar inicialmente empregado e a probabilidade de permanecer no painel, tomando como amostra todas as pessoas economicamente ativas (modelo de seleção amostral sobre não observáveis);
- 5) O modelo (5) é um probit quase idêntico ao modelo (1), porém utilizando um peso amostral multiplicado pela probabilidade bivariada de estar inicialmente empregado e permanecer na amostra, tal como descrito na equação (11) (modelo de seleção amostral sobre observáveis).

Como variáveis instrumentais no modelo (2), utilizaram-se *dummies* identificando o número da entrevista no domicílio. No modelo (3), utilizou-se como variável instrumental a proporção de demais pessoas em idade ativa ocupadas no setor censitário. Todas estas variáveis também foram utilizadas no modelo (4). No modelo (5), as variáveis auxiliares foram os números de demais pessoas ocupadas e desempregadas no domicílio e, novamente, as *dummies* identificando o número da entrevista.

Tanto o número da entrevista quanto a proporção de demais pessoas ocupadas no setor censitário mostraram-se como excelentes variáveis para instrumentalizar, respectivamente, a probabilidade de retenção no painel e a probabilidade de estar inicialmente empregado. Ambas possuem um alto poder de explicação nas equações de seleção e um insignificante poder de explicação na equação de interesse,

dada a inclusão das demais covariáveis. Para o modelo (5) de seleção sobre observáveis, as variáveis auxiliares mais adequadas que encontramos foram os números de demais pessoas ocupadas e desempregadas no domicílio. Essas características são claramente endógenas no processo de transição para o desemprego e explicam significativamente a condição inicial.

A Tabela 2 apresenta os resultados das estimativas dos modelos acima descritos. Primeiramente, podemos notar que, de acordo com a significância dos parâmetros de correlação entre os resíduos, ambos os critérios de seleção amostral são endógenos. No modelo (4), apesar de não encontrarmos uma correlação direta entre os resíduos das equações (12) e (14), a retenção no painel também é endógena, pois os resíduos das equações de seleção são significativamente correlacionados.

De maneira geral, os parâmetros de correlação apontam que os fatores não observados que aumentam a probabilidade de permanecer no painel diminuem a probabilidade de estar inicialmente desempregado. Essa redução na probabilidade inicial de desocupação aumenta, por sua vez, a probabilidade de transição para o desemprego. Isso pode significar que um trabalhador altamente condicionado ao desemprego, de acordo com suas características observáveis, que recebe uma oportunidade de ocupação, por exemplo, aumenta suas chances de permanecer no painel, porém possui poucas chances de permanecer ocupado.

De acordo com os coeficientes estimados, podemos notar que o efeito de variação na taxa de desemprego é sobre-estimado no caso de ausência do controle para a condição inicial. A ausência do controle para retenção (ou atrito) subestima, por sua vez, tanto o efeito da variação na taxa de desemprego quanto o efeito de variação no salário médio dos empregados da indústria. Comparando principalmente os modelos (1), (4) e (5), verifica-se que os demais coeficientes apresentam poucas diferenças entre as especificações. Contudo, cabe salientar que, neste caso, a inclusão do controle para somente um dos mecanismos de seleção incorreu em um viés sobre os coeficientes ainda maior do que se não fosse incluído nenhum mecanismo.

Tabela 2 - Coeficientes dos Modelos de Transição para o Desemprego

	(1)	(2)	(3)	(4)	(5)
taxa de desemprego	2.3729 **	2.7438 **	1.9882 *	2.2285 **	2.4393 **
log do salário médio - indústria	0.2965	0.3554 *	0.2939	0.3110 *	0.3146 *
log do salário médio - construção	-0.1135	-0.2080 **	-0.1157	-0.1349	-0.1081
log do salário médio - comércio	0.2088	0.1425	0.2088	0.1955	0.2220 *
log do salário médio - serviços	-0.2353	-0.5103 *	-0.2349	-0.2904	-0.2639
log do salário médio - outros	0.1697 *	0.1226	0.1708 *	0.1592	0.1762 *
mulher	-0.2016 **	-0.1988 **	-0.1889 **	-0.1960 **	-0.2036 **
idade	-0.0280 **	-0.0268 **	-0.0239 **	-0.0258 **	-0.0277 **
idade ao quadrado	0.0002 **	0.0002 **	0.0001 **	0.0002 **	0.0002 **
chefe do domicílio	-0.1120 **	-0.1180 **	-0.1024 **	-0.1088 **	-0.1084 **
chefe mulher	0.1697 **	0.1703 **	0.1571 **	0.1646 **	0.1656 **
casado ou unido	0.0157	0.0244	0.0185	0.0192	0.0112
1-3 anos de estudo concluídos	-0.0450	-0.0439	-0.0433	-0.0440	-0.0447
4 anos de estudo concluídos	-0.0687 **	-0.0657 **	-0.0681 **	-0.0680 **	-0.0645 *
5-7 anos de estudo concluídos	-0.0138	-0.0123	-0.0186	-0.0158	-0.0099
8 anos de estudo concluídos	-0.0977 **	-0.0940 **	-0.0976 **	-0.0970 **	-0.0911 **
9-10 anos de estudo concluídos	-0.1114 **	-0.1030 **	-0.1153 **	-0.1115 **	-0.1031 **
11 anos de estudo concluídos	-0.2041 **	-0.1937 **	-0.2003 **	-0.2001 **	-0.1987 **
12-14 anos de estudo concluídos	-0.3608 **	-0.3479 **	-0.3445 **	-0.3500 **	-0.3548 **
+15 anos de estudo concluídos	-0.4638 **	-0.4533 **	-0.4489 **	-0.4532 **	-0.4593 **
número de crianças (<10)	0.0117 *	0.0114 *	0.0113 *	0.0115 *	0.0123 *
número de adolescentes (10-14)	0.0068	0.0085	0.0058	0.0068	0.0084
número de jovens (15-17)	0.0066	0.0088	0.0044	0.0061	0.0075
número de adultos (18-65)	0.0132 **	0.0133 **	0.0123 **	0.0128 **	0.0082 *
número de idosos (>65)	0.0531 **	0.0566 **	0.0474 **	0.0512 **	0.0502 **
Variáveis de controle					
mês	x	x	x	x	x
ano	x	x	x	x	x
Coeficientes de correlação entre resíduos					
ρ_{12}	0	0	-0.4897 **	-0.1812 **	-
ρ_{13}	0	0.9645 **	0	0.0599	-
ρ_{23}	0	0	0	-0.0743 **	-0.0743 **
			teste	F(2, 608863)	
			$\rho_{13} = \rho_{23} = 0$	111.61 **	
			$\rho_{12} = \rho_{23} = 0$	140.15 **	

Fonte: PME 1995, 1996, 1997, 1998, 1999, 2000, 2001 e 2002.

Nota: * significativa a 5%; ** significativa a 1%.

5.2 O Efeito de Choques Transitórios Sobre a Progressão Escolar

O modelo de progressão escolar, apresentado por Duryea (*op. cit.*), pode ser representado da seguinte forma:

$$y_{it}^* = x'_{1,it} \beta_1 + T'_{it} \gamma_1 + \varepsilon_{1,it} \quad (15)$$

onde $x_{1,it}$ representa as características da criança, assim como do domicílio em que vive, β_1 é o respectivo vetor de coeficientes, $\varepsilon_{1,it}$ é o termo aleatório não explicado da equação e y_{it}^* é a variável latente que representa a propensão à progressão escolar da criança i no ano t , tal que:

se $y_{it}^* > 0$, $y_{it} = 1$ e a criança é aprovada na escola,

se $y_{it}^* \leq 0$, $y_{it} = 0$ e a criança é reprovada na escola.

O vetor T_{it} , associado ao vetor de coeficientes γ_1 na equação (15), é um indicador do choque transitório sobre a renda domiciliar. De acordo com Duryea, a melhor opção para medir o efeito destes choques é utilizar uma *proxy* de transição do chefe do domicílio para o desemprego.

Assim como o modelo anterior, o modelo de progressão escolar também depende de uma condição inicial. Essa condição é que a criança frequente a escola no período t . Seja f_{it}^* a variável latente que representa a propensão à frequência à escola da criança i no período t . Podemos definir a seguinte equação:

$$f_{it}^* = x'_{2,it} \beta_2 + \varepsilon_{2,it} \quad (16)$$

onde $x_{2,it}$ é o vetor que representa as características da criança e do lugar onde vive, β_2 é o respectivo vetor de coeficientes e $\varepsilon_{2,it}$ é o termo aleatório não explicado da equação.

Como estamos interessados em estimar o impacto da perda do emprego do chefe do domicílio sobre a progressão da criança, outra condição imposta à estimação da equação (15) é que este chefe esteja inicialmente empregado. Dessa forma, estimamos a probabilidade deste evento através da seguinte equação:

$$e_{it}^* = x'_{3,it} \beta_3 + \varepsilon_{3,it} \quad (17)$$

onde $x_{3,it}$ é o vetor de características explicativas, β_3 é o respectivo vetor de coeficientes, $\varepsilon_{3,it}$ é o termo aleatório não explicado da equação, e e_{it}^* é a variável latente que define se o chefe estava ou não inicialmente empregado.

Para estimar a equação (15), utilizamos uma amostra de crianças, entre 10 e 15 anos de idade, entrevistadas na nova PME, entre 2002 e 2007, em todas em RMs. Para captar o choque transitório e, ao mesmo, a progressão escolar, tomamos as informações da primeira à quarta entrevista do domicílio na PME, no ano t , e da quinta entrevista no ano seguinte. Para isso, é necessário restringir a amostra a crianças em domicílios entrevistados pela primeira vez entre os meses de janeiro a agosto, tal que as quatro primeiras entrevistas estejam dentro de um período escolar e a quinta entrevista esteja no período escolar seguinte.

As equações (16) e (17) definem a condição do chefe do domicílio e da criança justamente na primeira entrevista. Nas segunda, terceira e quarta entrevistas, é quando o possível choque transitório, que afetaria a progressão escolar, ocorre. Na quinta entrevista, é quando se confirma se a criança foi aprovada ou não no curso que frequentava.

Dessa forma, a última restrição imposta à equação (15) é que a criança i seja observada durante quatro meses no ano t e na primeira entrevista no ano $t + 1$. Assim, a retenção no painel pode ser especificada como:

$$d_{it}^* = x'_{4,it} \beta_4 + \varepsilon_{4,it} \quad (18)$$

onde $x_{4,it}$ é o vetor de características explicativas, β_4 é o respectivo vetor de coeficientes, $\varepsilon_{4,it}$ é o termo aleatório não explicado da equação e d_{it}^* é a variável latente que representa a propensão de a criança i em permanecer na amostra.

Para estimar os determinantes da progressão escolar, assim como a correlação ρ_{mn} entre os resíduos $\varepsilon_{m,it}$ e $\varepsilon_{n,it}$, onde $m \neq n$, utilizamos sete modelos distintos:

- 1) O modelo (1) é um probit univariado que estima a probabilidade de progressão escolar a partir de uma amostra de crianças que frequentavam a escola na primeira entrevista, ao mesmo tempo em que seus chefes estavam empregados, e que permaneceram no painel da primeira a quinta entrevista (modelo sem controle para seleção amostral);

- 2) O modelo (2) é um probit bivariado que estima simultaneamente a probabilidade de a criança progredir na escola e a probabilidade de estar frequentando a escola na primeira entrevista, utilizando uma amostra de crianças que permaneceram no painel e que seus chefes estavam empregados na primeira entrevista (modelo de seleção amostral sobre não observáveis);
- 3) O modelo (3) é um probit bivariado que estima simultaneamente a probabilidade de a criança progredir na escola e a probabilidade de seu chefe estar empregado na primeira entrevista, utilizando uma amostra de crianças que frequentavam a escola na primeira entrevista e permaneceram no painel até a quinta entrevista (modelo de seleção amostral sobre não observáveis);
- 4) O modelo (4) é um probit trivariado que estima simultaneamente a probabilidade de progressão, a probabilidade de frequentar a escola e a probabilidade de o chefe estar empregado na primeira entrevista, tomando como amostra todas as crianças que permaneceram no painel da primeira a quinta entrevista (modelo de seleção amostral sobre não observáveis);
- 5) O modelo (5) é um probit quase idêntico ao modelo (1), porém utilizando um peso amostral corrigido pela probabilidade bivariada de a criança frequentar a escola e do chefe estar empregado na primeira entrevista (modelo de seleção amostral sobre observáveis);
- 6) O modelo (6) é um probit bivariado que estima simultaneamente a probabilidade de progressão na escola e a probabilidade de permanecer no painel até a quinta entrevista, utilizando uma amostra de crianças que frequentavam a escola na primeira entrevista, ao mesmo tempo em que seus chefes estavam empregados (modelo de seleção amostral sobre não observáveis);
- 7) O modelo (7) é um probit tetravariado que estima simultaneamente a probabilidade de progressão, a probabilidade de frequentar a escola, a probabilidade de o chefe estar empregado e a probabilidade de permanecer no painel, tomando como amostra todas as crianças da pesquisa (modelo de seleção amostral sobre não observáveis).

Cabe salientar que os erros padrão de todos estes modelos foram calculados considerando o desenho amostral complexo da nova PME.

Para instrumentalizar a probabilidade de frequência à escola nos modelos (2), (4) e (7), utilizamos a proporção de demais crianças do setor censitário frequentando a escola. Nos modelos (3), (4) e (7), a probabilidade de o chefe estar empregado foi

instrumentalizada através da proporção de demais chefes de domicílio ocupados no setor censitário. Ambos os instrumentos foram utilizados também como variáveis auxiliares no modelo (5). Nos modelos (6) e (7), a probabilidade de retenção no painel foi determinada, entre outras variáveis, pela semana de entrevista.

A variável de proporção de demais chefes ocupados, apesar de explicar bem a probabilidade de ocupação do chefe da criança, não pode ser considerada um instrumento ideal, pois possui um efeito significativo sobre a probabilidade de progressão escolar. Por isso, estimamos também o modelo (5) de seleção sobre observáveis, utilizando esta variável como auxiliar.

De acordo com os resultados apresentados na Tabela 3, verificamos que há correlações significativamente positivas entre os resíduos das equações de seleção. Dessa forma, uma maior probabilidade não explicada de o chefe do domicílio estar empregado está relacionada a uma maior probabilidade de a criança frequentar a escola e permanecer no painel. Em termos da correlação direta com a equação de interesse, no modelo (6), identificamos que uma maior probabilidade não explicada de progressão na escola está relacionada a uma maior probabilidade de atrito no painel. Contudo, essa correlação não se mantém significativa no modelo (7). Este último modelo apresenta, por outro lado, uma correlação significativamente positiva entre os resíduos das equações (15) e (16). Ou seja, quanto maiores as chances não explicadas de a criança frequentar a escola, menores são as chances de aprovação na série que frequenta.

Todos os testes de significância conjunta dos parâmetros de correlação apontam que os critérios de seleção amostral são endógenos. Contudo, o modelo probit univariado (1), sem o controle de seleção amostral, não apresenta grandes divergências em relação aos demais modelos. Os únicos coeficientes deste modelo que parecem estar um pouco sobreestimados são os da escolaridade do chefe. Comparando o modelo (1) com os modelos (2), (4), (6) e (7), constatamos que estes coeficientes estão sobre-estimados possivelmente devido à ausência do controle de seleção sobre a frequência à escola e sobre a retenção no painel.

Tabela 3 – Coeficientes dos Modelos de Progressão Escolar

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Chefe							
desempregado no ano escolar	-0.1274 **	-0.1241 **	-0.1291 **	-0.1263 **	-0.1309 **	-0.1073 *	-0.1266 **
desempregado após o ano escolar	-0.1343 *	-0.1342 *	-0.1314	-0.1344 *	-0.1376 *	-0.1172	-0.1329 *
que deixou a força de trabalho	-0.0840	-0.0847	-0.0817	-0.0840	-0.0814	-0.0760	-0.0842
que saiu do domicílio - atritado	0.0016	0.0028	0.0007	0.0025	0.0085	-0.0008	0.0020
criança do sexo feminino	0.1690 **	0.1669 **	0.1639 **	0.1683 **	0.1685 **	0.1543 **	0.1684 **
criança não branca	-0.0453	-0.0407	-0.0556	-0.0424	-0.0521	-0.0224	-0.0411
com 11 anos de idade	-0.1558 **	-0.1497 **	-0.1469 **	-0.1529 **	-0.1534 **	-0.1501 **	-0.1522 **
com 12 anos de idade	-0.2014 **	-0.1990 **	-0.1968 **	-0.2002 **	-0.1978 **	-0.1842 **	-0.1994 **
com 13 anos de idade	-0.2387 **	-0.2310 **	-0.2304 **	-0.2348 **	-0.2361 **	-0.2302 **	-0.2360 **
com 14 anos de idade	-0.2873 **	-0.2727 **	-0.2770 **	-0.2802 **	-0.2822 **	-0.2718 **	-0.2810 **
com 15 anos de idade	-0.3485 **	-0.3187 **	-0.3382 **	-0.3331 **	-0.3531 **	-0.3217 **	-0.3327 **
com dois anos ou mais de atraso	-0.0159	0.0131	-0.0258	-0.0016	-0.0181	0.0395	0.0009
filho do chefe	-0.0385	0.0069	-0.1132	-0.0280	-0.0341	0.2238 **	-0.0129
n. de crianças <10 anos	-0.0572 **	-0.0532 **	-0.0622 **	-0.0556 **	-0.0581 **	-0.0395 **	-0.0551 **
n. de demais crianças 10-15 anos	-0.0544 **	-0.0568 **	-0.0532 **	-0.0556 **	-0.0534 **	-0.0511 **	-0.0559 **
chefe mulher	-0.0875 **	-0.0710 *	-0.1809	-0.0818	-0.0825 **	0.0464	-0.0750
idade do chefe	-0.0027	-0.0023	-0.0061	-0.0027	-0.0028	-0.0013	-0.0027
chefe 1-3 anos de estudo	0.0431	0.0339	0.0623	0.0426	0.0364	-0.0095	0.0412
chefe 4 anos de estudo	0.0716	0.0501	0.1025	0.0660	0.0792	-0.0269	0.0608
chefe 5-7 anos de estudo	0.1076	0.0895	0.1417	0.1040	0.1199	0.0169	0.0994
chefe 8 anos de estudo	0.1838 **	0.1626 *	0.2280 **	0.1782 *	0.1900 **	0.0897	0.1744 *
chefe 9-10 anos de estudo	0.1952 *	0.1747 *	0.2339 *	0.1900 *	0.2008 *	0.0760	0.1836
chefe 11 anos de estudo	0.2749 **	0.2490 **	0.3304 **	0.2671 **	0.2766 **	0.1481 *	0.2614 **
chefe 12-14 anos de estudo	0.3044 **	0.2765 **	0.3606 **	0.2941 *	0.3038 **	0.2249 *	0.2959 **
chefe +15 anos de estudo	0.2802 **	0.2528 **	0.3700 **	0.2716 *	0.2839 **	0.1521 *	0.2669 **
RM Salvador	-0.1254 *	-0.1251 *	-0.1142 *	-0.1250 *	-0.1236 *	-0.0820	-0.1219 *
RM Belo Horizonte	0.1065 **	0.1071 **	0.1261 **	0.1087 *	0.1082 **	0.0334	0.1040 *
RM Rio de Janeiro	-0.0292	-0.0336	0.0152	-0.0293	-0.0265	-0.1588 **	-0.0393
RM São Paulo	0.1871 **	0.1884 **	0.2034 **	0.1897 **	0.1838 **	0.1357 **	0.1879 **
RM Porto Alegre	-0.1751 **	-0.1671 **	-0.1502 **	-0.1692 **	-0.1790 **	-0.1917 **	-0.1723 **
Variáveis de controle							
mês	x	x	x	x	x	x	x
ano	x	x	x	x	x	x	x
Coeficientes de correlação entre resíduos							
ρ_{12}		-0.3297		-0.1924			-0.2564 *
ρ_{13}			0.2077	-0.0105			0.0089
ρ_{14}						-0.6878 **	-0.2585
ρ_{23}				0.1511 **	0.0205 **		0.3051 **
ρ_{24}							0.0623 **
ρ_{34}							0.8413 **
			teste	F(2, 1914)		teste	F(3, 1951)
			$\rho_{12} = \rho_{23} = 0$	56.83 **	$\rho_{12} = \rho_{23} = \rho_{24} = 0$		286.18 **
			$\rho_{13} = \rho_{23} = 0$	56.20 **	$\rho_{13} = \rho_{23} = \rho_{34} = 0$		19643.6 **
					$\rho_{14} = \rho_{24} = \rho_{34} = 0$		19556.1 **

Fonte: PME 2002, 2003, 2004, 2005, 2006 e 2007.

Nota: * significativa a 5%; ** significativa a 1%.

Entretanto, deve-se ter cuidado na interpretação dos resultados do modelo (6), no qual somente o controle de retenção no painel é considerado. Os resultados deste modelo são muito divergentes dos apresentados nos demais. Portanto, assim como na estimação da transição para o desemprego na seção anterior, concluímos que a inclusão do controle para somente um dos mecanismos de seleção pode incorrer em um viés sobre os coeficientes ainda maior do que se não fosse incluído nenhum mecanismo.

6 Conclusão

Em relação aos determinantes do desgaste, identificamos que este possui certa sazonalidade, com maior taxa nos meses entre novembro e fevereiro, e forte correlação com as características geralmente associadas a maior mobilidade geográfica. A escolaridade do chefe do domicílio, por sua vez, não é um fator tão determinante do desgaste quanto à escolaridade da própria pessoa. Por fim, tanto características do processo de entrevista quanto fatores socioeconômicos são significativamente relacionados à probabilidade de retenção no painel. Nas características da entrevista, podemos destacar que, devido a melhoras no processo, a taxa de atrito tende a diminuir ao longo do tempo.

Para testar o viés do desgaste no painel, assim como de outros critérios de seleção amostral, sobre algumas análises de regressão, estimamos dois modelos utilizando os microdados da PME. Em ambos os casos, constatamos que todos os critérios de seleção amostral são endógenos, apesar de alguns não estarem diretamente correlacionados com a equação de interesse. No primeiro caso, verificamos que uma maior probabilidade não explicada de permanecer no painel está relacionada a uma maior probabilidade de estar inicialmente empregado. Contudo, esse aumento não explicado na probabilidade inicial de ocupação aumenta, por sua vez, a probabilidade de transição para o desemprego. No segundo caso, uma maior probabilidade não explicada de o chefe do domicílio estar empregado está relacionada a uma maior probabilidade de a criança frequentar a escola e de permanecer no painel. Além disso, quanto maiores as chances não explicadas de a criança frequentar a escola, menores são as chances de aprovação na série que frequenta.

De acordo com os resultados apresentados, a ausência do controle para seleção amostral, o que inclui a retenção no painel, pode incorrer em algumas análises enviesadas. Contudo, a inclusão do controle para somente alguns mecanismos de seleção pode fornecer estatísticas ainda mais inconsistentes do que se não fosse incluído nenhum mecanismo.

Referências

- ARENDET, J. N.; HOLM, A. 2006. *Probit models with binary endogenous regressors*. Department of Business and Economics, University of Southern Denmark, 2006. (Discussion Papers on Business and Economics No. 4/2006).
- BOURGUIGNON, F.; FOURNIER, M.; GURGAND, M. Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. *Journal of Economic Surveys*, v. 21, n. 1, p. 174-205, 2007.
- CAPPELARI, L.; JENKINS, S. P. Modelling low income transitions. *Journal of Applied Econometrics*, v.19, n. 5, p. 593-610, 2004a.
- _____. *Modelling low pay transition probabilities, accounting for panel attrition, non-response and initial conditions*. ISER, University of Essex, 2004b. (ISER Working Paper n. 2004-08)
- _____. Calculation of multivariate normal probabilities by simulation, with application to maximum simulated likelihood estimation. *Stata Journal*, v. 6, n. 2, p. 156-189, 2006.
- CORSEUIL, C. H.; CARNEIRO, F. G. *Os impactos do salário mínimo sobre emprego e salários no Brasil: evidências a partir de dados longitudinais e séries temporais*. Rio de Janeiro: IPEA, 2001. (Texto para Discussão, n. 849).
- DURYEA, S. *Children's advancement through school in Brazil: the role of transitory shocks to household income*. Inter-American Development Bank, 1998. (RES Working Paper 376).
- _____.; LAM, D.; LEVISON, D. Effects of economic shocks on children's employment and schooling in Brazil. *Journal of Development Economics*, v. 84, n. 1, p. 188-214, 2007.
- FALARIS, E. M.; PETERS, H. E. Survey attrition and schooling choices. *Journal of Human Resources*, v. 33, n. 2, p. 531-554, 1998.
- FISHE, R. P. H.; TROST, R. P.; LURIE, P. Labor force earnings and college choice of young women: an examination of selectivity bias and comparative advantage. *Economics of Education Review*, v. 1, n. 2, p. 169-191, 1981.
- FITZGERALD, J.; GOTTSCHALK, P.; MOFFIT, R. An analysis of sample attrition in panel data: the Michigan panel study of income dynamics. *Journal of Human Resources*, v. 33, n. 2, p. 251-299, 1998.
- GEWEKE, J. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In: COMPUTING SCIENCE AND STATISTICS: THE TWENTY-THIRD SYMPOSIUM ON THE INTERFACE, 23th, 1991, Seattle. *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*. Seattle: American Statistical Association, 1991. p. 571-578.
- GONZAGA, G.; REIS, M. C. Os efeitos trabalhador adicional e desalento no Brasil. In: ENCONTRO NACIONAL DE ECONOMIA, 23. *Anais...* João Pessoa: ANPEC, 2005.

- HAIJIVASSILIOU, V. *Smooth simulation estimation of panel data LDV models*. Yale University, 1990. Manuscript.
- HAUSMAN, J. A.; WISE, D. A. Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica*, v. 47, n. 2, p. 455-474, 1979.
- HECKMAN, J. Shadow prices, market wages, and labor supply. *Econometrica*, v. 42, n. 4, p. 679-694, 1974.
- _____. Sample selection bias as a specification error. *Econometrica*, v. 47, n. 1, p. 153-162, 1979.
- HORVITZ, D. G.; THOMPSON, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, v. 47, n. 260, p. 663-685, 1952.
- IBGE – Departamento de Emprego e Rendimento. *Para Compreender a PME: (um texto simplificado)*. 4ª ed. Rio de Janeiro, 1998.
- _____. *Pesquisa Mensal de Emprego*. Rio de Janeiro, 2002. (Relatório Metodológico v. 23).
- JUSTO, W. R.; SILVEIRA NETO, R. M. Quem são e para onde vão os migrantes no Brasil? O Perfil do Migrante Interno Brasileiro. In: XXXVI ENCONTRO NACIONAL DE ECONOMIA. Salvador: ANPEC, 2008.
- KEANE, M. P. A computationally practical simulation estimator for panel data. *Econometrica*, v. 46, n. 6, p. 931-959, 1994.
- LEMONS, S. *The effects of the minimum wage on wages and employment in Brazil: a menu of minimum wage variables*. Department of Economics, University College London, 2002. (Discussion Paper 02-02).
- LILLARD, L. A.; PANIS, C. W. Panel attrition from the panel study of income dynamics: household income, marital status, and mortality. *Journal of Human Resources*, v. 33, n. 2, p. 437-457, 1998.
- LOPES, M. D. *Avaliação de desgaste de painéis em estudos longitudinais: uma aplicação na Pesquisa Mensal de Emprego (PME/IBGE)*. 2002. Dissertação (Mestrado). Orientadora: Denise Britz do Nascimento Silva. ENCE, Rio de Janeiro.
- MACHADO, A. F.; RIBAS, R. P.; PENIDO, M. Mobilidade entre estados de pobreza e inserção no mercado de trabalho: uma análise para o Brasil Metropolitano em 2004. *Economia Aplicada*, v. 11, n. 2, p. 253-279, 2007.
- NERI, M.; COELHO, D.; ANCORA, M.; PINTO, A. Aspectos dinâmicos do desemprego e da posição na ocupação. *Estudos Econômicos*, v. 27, n. especial, p.: 137-159, 1997.
- NERI, M. C., GUSTAFSSON-WRIGHT, E.; SEDLACEK, G.; COSTA, D. R. da; PINTO, A. 2000. Microeconomic instability and children's human capital accumulation: the effects of idiosyncratic shocks to father's income on child

- labor, school drop-outs and repetition rates in Brazil. In: LACEA 2000 ANNUAL MEETING, Rio de Janeiro, 2000.
- NICOLETTI, C.; PERACCHI, F. *Two-step estimation of binary response models with sample selection*. In: BHPS 2001 CONFERENCE, Colchester, 2001.
- PENIDO, M.; MACHADO, A. F. Duração do desemprego no Brasil Metropolitano. In: WAJNMAN, S.; MACHADO, A. F. (Ed.). *Mercado de trabalho: uma análise a partir das pesquisas domiciliares no Brasil*. Editora UFMG, 2003, p. 203-218.
- PERACCHI, F.; WELCH, F. How representative are matched cross-sections? Evidence from the current population survey. *Journal of Econometrics*, v. 68, n. 1, p. 153-179, 1995.
- RIBAS, R. P.; SOARES, S. D. *Sobre o painel da Pesquisa Mensal de Emprego (PME) do IBGE*. Rio de Janeiro: IPEA, 2008. (Texto para Discussão, n. 1348)
- VAN DEN BERG, G. J.; LINDEBOOM M. Attrition in panel survey data and the estimation of multi-state labor market models. *Journal of Human Resources*, v. 33, n. 2, p. 458-478, 1998.
- VELLA, F. Estimating models with sample selection bias: a survey. *Journal of Human Resources*, v. 33, n. 1, p. 127-169, 1998.
- VERBEEK, M.; NIJMAN, T. Testing for selectivity bias in panel data models. *International Economic Review*, v. 33, n. 3, p. 681-703, 1992.
- WOLTERMANN, S. *Job-search methods and labor market transitions in a segmented economy: some empirical evidence from Brazil*. Ibero-America Institute for Economic Research (IAI), Georg-August-Universität Göttingen, 2002. (Discussion Paper 88).
- WOOLDRIDGE, J. M. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, v. 1, n. 2, p. 117-139, 2001.
- ZABEL, J. E. An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *Journal of Human Resources*, v. 33, n. 2, p. 479-506, 1998.
- ZILIAK, J. P.; KNIESNER, T. J. The importance of sample attrition in life cycle labor supply estimation. *Journal of Human Resources*, v. 33, n. 2, p. 507-530, 1998.