

TEXTO PARA DISCUSSÃO N° 1347

O ATRITO NAS PESQUISAS LONGITUDINAIS: O CASO DA PESQUISA MENSAL DE EMPREGO (PME) DO IBGE

**Rafael Perez Ribas
Sergei Suarez Dillon Soares**

Rio de Janeiro, agosto de 2008

TEXTO PARA DISCUSSÃO N° 1347

O ATRITO NAS PESQUISAS LONGITUDINAIS: O CASO DA PESQUISA MENSAL DE EMPREGO (PME) DO IBGE*

Rafael Perez Ribas**

Sergei Suarez Dillon Soares***

Rio de Janeiro, agosto de 2008

* Os autores agradecem os comentários de Carlos Henrique Corseuil e Maurício Cortez Reis. Os erros remanescentes são de responsabilidade dos autores. Todas as sintaxes usadas neste trabalho encontram-se à disposição para todos que as queiram utilizar. Enviar pedidos para sergei.soares@ipea.gov.br ou rafael.ribas@undp-povertyncentre.org.

** Do Centro Internacional de Pobreza.

*** Técnico de Planejamento e Pesquisa do Ipea.

Governo Federal

**Ministro de Estado Extraordinário de
Assuntos Estratégicos** – Roberto Mangabeira Unger

**Secretaria de Assuntos Estratégicos
da Presidência da República**

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Marcio Pochmann

Diretor de Administração e Finanças

Fernando Ferreira

Diretor de Estudos Macroeconômicos

João Sicsú

Diretor de Estudos Sociais

Jorge Abrahão de Castro

Diretora de Estudos Regionais e Urbanos

Liana Maria da Frota Carleial

Diretor de Estudos Setoriais

Márcio Wohlers de Almeida

Diretor de Cooperação e Desenvolvimento

Mário Lisboa Theodoro

Chefe de Gabinete

Persio Marco Antonio Davison

Assessor-Chefe de Comunicação

Estanislau Maria de Freitas Júnior

URL: <http://www.ipea.gov.br>

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

ISSN 1415-4765

JEL: C33; C81

TEXTO PARA DISCUSSÃO

Publicação cujo objetivo é divulgar resultados de estudos desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

SINOPSE

ABSTRACT

| | |
|---|----|
| 1 INTRODUÇÃO | 7 |
| 2 COMO CONTROLAR A SELEÇÃO NÃO-ALEATÓRIA – REVISÃO DA LITERATURA | 8 |
| 3 ESTIMATIVA DOS DETERMINANTES DA PERMANÊNCIA DA PERMANÊNCIA NO PAINEL DA PME | 15 |
| 4 COMPARANDO MODELOS COM E SEM CONTROLE PARA SELEÇÃO AMOSTRAL | 25 |
| 5 CONCLUSÃO | 33 |
| REFERÊNCIAS | 34 |

SINOPSE

O objetivo deste artigo é estimar os determinantes do atrito/permanência de pessoas no painel da Pesquisa Mensal de Emprego (PME), do Instituto Brasileiro de Geografia e Estatística (IBGE) e, além disso, testar se ignorar a não-aleatoriedade deste desgaste causa algum viés na análise de regressões. O atrito pode ser de três tipos: *a)* aleatório, que não gera qualquer tipo de viés em estimações de parâmetros; *b)* função de variáveis observadas, cujo viés pode ser corrigido mediante repesagem adequada; e *c)* função de variáveis não-observadas, cujo viés pode ser corrigido usando-se um *probit* bivariado para variáveis discretas ou a correção de Heckman para variáveis contínuas.

Os resultados apontam que a taxa de atrito na pesquisa tende a ser maior nos meses de férias, e que há uma forte correlação dessa taxa com as características geralmente associadas a uma mobilidade geográfica maior. Além disso, tanto características do processo de entrevista quanto fatores socioeconômicos são 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. 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.

ABSTRACT

Our objective is to analyze the gravity of panel attrition in the Brazilian Monthly Employment Survey (PME). There are three types of panel attrition: *a)* random, that does not produce any selection bias; *b)* function of observables, whose bias can be corrected through an appropriate re-weighting scheme; and *c)* function of unobservable variables, whose bias can be corrected through a bivariate probit for discrete variables or a Heckman correction for continuous ones.

To test the gravity of panel attrition bias we estimated a transition to unemployment model both with and without attrition bias controls. Our results are that attrition tends to increase during school vacations and that attrition is strongly correlated to geographical mobility. Our conclusion is that while attrition bias does exist, its magnitude is small and its elimination does not substantially change analysis made using the models. In addition, in some cases, since it is not possible to control simultaneously for observable and non-observable attrition bias, in some cases the best procedure is not use on controls at all.

1 INTRODUÇÃO

A Pesquisa Mensal de Emprego (PME), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), se caracteriza por ser amostral, domiciliar e de periodicidade mensal. Seu levantamento é realizado somente nas regiões metropolitanas (RMs) de Rio de Janeiro, São Paulo, Porto Alegre, Belo Horizonte, Recife e Salvador. Usualmente, os microdados dessa pesquisa são trabalhados na forma *cross-section*, na forma de série de tempo mensal e no formato de painel.

O painel da PME, que acompanha as unidades amostrais por até oito entrevistas, possibilita, sem dúvida alguma, a maneira mais rica de se trabalhar com essa base de dados, pois 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.

Basicamente, são duas as causas para o desgaste do painel. A primeira causa de desgaste em uma pesquisa longitudinal está relacionada diretamente à mobilidade geográfica das pessoas na amostra. Na PME, este problema é ainda maior que em outras pesquisas em painel, como o Panel Study of Income Dynamics (PSID) e o British Household Panel Survey (BHPS), que se esforçam por encontrar as pessoas em outros endereços. No nosso caso, se os indivíduos deixam seus domicílios, eles são automaticamente excluídos da pesquisa. A segunda causa para o desgaste é a recusa de entrevista. Aqui, temos a hipótese de que, após participar repetidas vezes de uma determinada pesquisa, os respondentes podem se tornar desinteressados, recusando-se a ser entrevistados novamente.

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

Alguns trabalhos, como o de Neri *et al.* (1997), ajudaram a disseminar a idéia de que, na PME, o atrito da amostra não introduz vieses nas análises de mobilidade ou transição. A razão para esta afirmação é que não haveria diferença significativa nas características demográficas e econômicas entre as amostras atritadas e não-atritadas. O trabalho de Lopes (2002) mostra, porém, que algumas características individuais, como a situação de desemprego e a idade, e do domicílio, assim como o sexo do chefe, estão significativamente relacionadas com o desgaste no painel. Independentemente disso, Falaris e Peters (1998) salientam que, mesmo quando a subamostra 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 causado pelo atrito em painéis. Em seguida, 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. Por fim, algumas conclusões são traçadas com base nos resultados encontrados.

Em relação aos determinantes do desgaste, verificamos que a taxa de atrito tende a ser maior nos meses de férias, e que há uma forte correlação dessa taxa com as características geralmente associadas a uma mobilidade geográfica maior. Além disso, tanto as características do processo de entrevista quanto os fatores socioeconômicos são significativamente relacionados à probabilidade de atrito no painel. Verificando os modelos adotados para testar o viés do desgaste no painel, assim como de outros critérios de seleção amostral, constatamos que todos eles 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. 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.

2 COMO CONTROLAR A SELEÇÃO NÃO-ALEATÓRIA – REVISÃO DA LITERATURA

Uma pesquisa longitudinal é aquela na qual uma unidade de pesquisa é entrevistada mais de uma vez, o que quer dizer que se pode montar um painel de dados. Exemplos no Brasil são o Censo Escolar, no qual cada escola é entrevistada a cada ano; a Relação Anual de Informações Sociais (Rais), na qual as empresas e unidades de produção são entrevistadas repetidas vezes; e a PME, na qual cada domicílio e, potencialmente, cada morador é entrevistado até oito vezes. Ademais, há uma tendência em direção à realização, cada vez mais, de pesquisas longitudinais. Além de contarmos com registros administrativos cada vez melhores, como o censo escolar e a Rais, que permitem que se monte um painel, as pesquisas amostrais também se movem na direção de permitir a montagem de um painel. Por exemplo, o IBGE estuda a transformação da Pesquisa Nacional por Amostra de Domicílios (Pnad) em uma pesquisa contínua com um painel rotativo de domicílios.

Os dados em painel possibilitam uma riqueza de análise que os dados transversais não permitem, e as pesquisas longitudinais abrem todo um mundo novo de possibilidades de análise. Há, no entanto, um preço a ser pago. Os painéis sofrem de desgaste e isso passa a ser algo que deve ser levado em conta. Se a morte, nascimento ou troca de endereço fossem eventos que ocorressem de forma aleatória – isto é, sem qualquer correlação com qualquer variável presente na pesquisa nem com qualquer variável ausente, mais relevante para as relações que estão sendo identificadas – o atrito seria um problema para um painel de pessoas apenas na medida em que reduzisse a 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 inconsistente e 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 vies de seleção ocorre somente quando o atrito está relacionado aos erros aleatórios da equação de interesse.

Embora na literatura de amostragem, lida por estatísticos, o tema de desgaste de painéis seja tratado há algum tempo, o trabalho de Hausman e Wise foi dos primeiros a levantar a questão na literatura econométrica, lida por economistas. O estudo em questão era sobre o impacto de um experimento (Gary Income Maintenance Experiment) sobre os rendimentos de trabalhadores negros. Neste caso, a hipótese era de que a correlação entre o elevado rendimento pré-experimento, o conseqüente não-recebimento do benefício e a falta de incentivo em permanecer na pesquisa no período seguinte levaria a um vies na estimação do efeito do experimento. Com a perda de 35% da amostra no período seguinte, o vies de atrito foi identificado como um problema nas avaliações mais parcimoniosas. Porém, esse vies diminuiu com a incorporação de mais variáveis exógenas no modelo.

Para correção do vies, os autores propuseram uma função de verossimilhança muito próxima da utilizada por Heckman (1974). A proposta de Heckman (1974) é de extensão do modelo *tobit* para um modelo de equações simultâneas, em que os parâmetros determinantes da probabilidade de estar na amostra diferem dos parâmetros da equação de interesse. Porém, ambas as equações são estimadas conjuntamente. Apesar da extensa literatura sobre vies de atrito, publicada após os trabalhos de Hausman e Wise (1979) e de Heckman (1974, 1979), alguns trabalhos continuam não se atendo ao problema.¹ A prática mais comum utilizada nesses trabalhos é descartar as unidades com observações ausentes nos períodos subseqüentes. 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 freqüente ausência de preocupação com o vies de atrito é que, de acordo com Verbeek e Nijman (1992), a estimação de um modelo de efeitos fixos pode eliminar muita das formas de heterogeneidade não-observada. Entre elas, a heterogeneidade na seleção amostral. O trabalho de Ziliak e Kniesner (1998) chega a conclusões semelhantes às de Verbeek e Nijman. Esses autores estimam um modelo de ciclo de vida da oferta de trabalho, utilizando dados do PSID. Seus resultados mostram que não é necessário utilizar um termo de correção para o vies de atrito quando utilizado um modelo de efeitos fixos. Isso porque, em seu modelo, a probabilidade de atrito está relacionada principalmente a características específicas e invariantes dos indivíduos.

Contudo, Vella (1998) salienta que outras formas de vies 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

1. Como exemplos de trabalhos que não utilizam técnicas de correção para o vies de atrito, nem ao menos testam a aleatoriedade do desgaste, podemos citar os de Duryea (1998); Duryea, Lam e Levison (2007); Neri *et al.* (2000); Camargo e Neri (1999); Corseuil e Carneiro (2001); Lemos (2002); Woltermann (2002); Penido e Machado (2003); Gonzaga e Reis (2005); e Machado, Ribas e Penido (2007).

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. Eles concluem que um choque na probabilidade de permanência na amostra está positivamente correlacionado a um choque na probabilidade de desemprego. Ou seja, retirar da análise pessoas que saem da amostra pode sobreestimar o risco 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 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 (1998) 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 enviesados. Os autores analisaram a escolaridade de homens e mulheres em três coortes do National Longitudinal Surveys of Labor Market Experience (NLS) e do PSID. Apesar das regressões sobre as amostras de atritados e não-atritados não se diferenciarem significativamente, os resultados da regressão com correção do viés eram muito diferentes das regressões nas subamostras, indicando que o atrito não é aleatório e gera estimativas enviesadas. Zabel (1998), por outro lado, apresenta um caso onde o padrão de atrito é o contrário deste último. O autor também utilizou o PSID, além do Survey of Income and Program Participation (SIPP), para estimar a equação de salários e a oferta de trabalho de chefes de família. As taxas de atrito não eram tão elevadas, 12% no PSID e entre 6% e 8% no SIPP. A primeira evidência foi que atritados apresentavam uma variabilidade maior em suas horas trabalhadas e em seus salários, ou seja, eram mais susceptíveis a mudanças ao longo do tempo, 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-se 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.

Uma 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 delas poderão ficar desempregadas em períodos subsequentes. Esta situação é diferente de um problema de desgaste, em que pessoas saem completamente da amostra e, normalmente, não reaparecem nos períodos seguintes.

Cappellari e Jenkins (2004a, 2004b) investigam o viés de atrito sobre duas equações um pouco distintas, sendo que ambas envolvem transições entre estados discretos e outros mecanismos de seleção. Os dados são do BHPS, com uma taxa de atrito baixa, de apenas 8% ao ano. No primeiro trabalho (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 significativa seletividade no atrito, o viés introduzido ao ignorar o problema é muito baixo. Cabe salientar que a taxa de atrito nesse 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. As formas de não-aleatoriedade do atrito que causam viés na estimação são apresentadas a seguir. Considere-se 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 assume-se 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 assumindo-se que cada componente do erro é normalmente distribuído e correlacionado com o componente da mesma dimensão na outra equação.² 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:

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

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

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

$\mu_{2,i}$ de $\varepsilon_{2,it}$. Dessa forma, o comportamento de atrito seria tratado como dentro de um segundo grupo de estimativas, junto com aquelas onde a seleção ocorre sobre os componentes não-observados, tal que:

$$x_{2,it} \perp \varepsilon_{1,it}, \text{ mas } \varepsilon_{2,it} \perp \varepsilon_{1,it} \text{ porque } e_{2,it} \perp e_{1,it} \quad (6)$$

Nesse 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 (ARENDET; HOLM, 2006). Por isso, talvez, alguns trabalhos ainda prefiram a estimação em dois estágios. De acordo com Heckman (1979), no caso de uma variável de interesse contínua e de apenas um mecanismo de seleção, a esperança condicional de $\varepsilon_{2,it}$, na equação (8), pode ser calculada através da razão inversa de Mills, λ :

$$E[\varepsilon_{2,it} | \varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2] = \lambda(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2) \equiv \frac{\phi(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2)}{\Phi(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2)} \quad (9)$$

onde ϕ é uma função densidade de probabilidade normal univariada e Φ é uma função de probabilidade normal univariada acumulada.

Segundo Nicoletti e Peracchi (2001), o uso da razão inversa de Mills como fator de correção de viés também gera resultados consistentes na estimação de variáveis de interesse binárias. Contudo, quando há mais de um mecanismo de seleção da amostra, este termo de correção não é apropriado. Supondo-se que s_{it} representa um segundo critério de participação na amostra, onde:

$$s_{it}^* = x'_{1,it}\beta_3 + x'_{3,it}\gamma_3 + \varepsilon_{3,it} \quad (10)$$

e:

$$s_{it} = \begin{cases} 1 & \text{se } s_{it}^* > 0 \\ 0 & \text{se } s_{it}^* \leq 0 \end{cases} \quad (11)$$

reescrevemos a equação (4) da seguinte forma:

$$y_{it} = y_{it}^* \cdot d_{it} \cdot s_{it} \quad (12)$$

Dessa forma, a esperança de $\varepsilon_{1,it}$ na equação (7) é condicionada não só a $\varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2$ mas também a $\varepsilon_{3,it} > -x'_{1,it}\beta_3 - x'_{3,it}\gamma_3$. Considerando y_{it} uma variável contínua, Fische, Trost e Lurie (1981) propõem a seguinte esperança condicional do resíduo da equação de interesse:

$$E[\varepsilon_{1,it} | x_{1,it}, x_{2,it}, x_{3,it}, \varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2, \varepsilon_{3,it} > -x'_{1,it}\beta_3 - x'_{3,it}\gamma_3] = \sigma_{12}\lambda_2 + \sigma_{13}\lambda_3 \quad (13)$$

onde σ_{1m} é a covariância entre os resíduos, $\varepsilon_{1,it}$ e $\varepsilon_{m,it}$, e λ_m representa os termos para correção do viés de seleção, definidos como:

$$\lambda_m = (1 - \sigma_{23}^2)^{-1} (h_m - \sigma_{23}h_n) \quad (14)$$

sendo

$$\begin{aligned} h_m &= E[\varepsilon_{m,it} | \varepsilon_{2,it} > -x'_{1,it}\beta_2 - x'_{2,it}\gamma_2, \varepsilon_{3,it} > -x'_{1,it}\beta_3 - x'_{3,it}\gamma_3] \\ &= \left\{ \phi(-x'_{1,it}\beta_m - x'_{m,it}\gamma_m) \Phi\left(\frac{x'_{1,it}\beta_n + x'_{n,it}\gamma_n - \rho_{23}(x'_{1,it}\beta_m + x'_{m,it}\gamma_m)}{\sqrt{1 - \rho_{23}^2}}\right) \right. \\ &\quad \left. + \rho_{23}\phi(-x'_{1,it}\beta_n - x'_{n,it}\gamma_n) \Phi\left(\frac{x'_{1,it}\beta_m + x'_{m,it}\gamma_m - \rho_{23}(x'_{1,it}\beta_n + x'_{n,it}\gamma_n)}{\sqrt{1 - \rho_{23}^2}}\right) \right\} \\ &\quad \cdot (\Phi_2(x'_{1,it}\beta_2 + x'_{2,it}\gamma_2, x'_{1,it}\beta_3 + x'_{3,it}\gamma_3; \rho_{23}))^{-1}, \end{aligned} \quad (15)$$

$$m, n = 2, 3, m \neq n \text{ e } \rho_{23} = \sigma_{23}(\sigma_2\sigma_3)^{-1} \text{.}^3$$

É fácil ver que, se $\rho_{23} = 0$, ou seja, não há relação entre os dois mecanismos de seleção, o termo de correção proposto por Fische, Trost e Lurie, na equação (14), é idêntico ao proposto por Heckman (1979), na equação (9). Assim como Nicoletti e Peracchi (2001), que analisaram regressões com um mecanismo de seleção, Arendt e Holm (2006) concluíram que a utilização dos termos de correção também gera resultados consistentes no caso de variáveis de interesse binárias, quando são dois os mecanismos de seleção. Esses autores evidenciaram ainda que a técnica de dois estágios gera resultados mais consistentes, em alguns casos, que a maximização de uma única função de verossimilhança multivariada.⁴

3. Para calcular os termos de correção no primeiro estágio, estimam-se as equações (2) e (10) e a correlação entre seus resíduos por meio de um modelo *probit* bivariado. No segundo estágio, a equação (1) é estimada incorporando-se os termos de correção como variáveis explicativas.

4. Para mais exemplos de métodos que corrigem viés de seleção sobre não-observáveis, ver Vella (1998).

Independentemente da forma de estimação, a identificação de β_1 implica uma restrição de exclusão, que exista $x_{2,it}$ (e $x_{3,it}$) satisfazendo a propriedade de independência de $\varepsilon_{1,it}$ e na qual γ_2 (e γ_3) não seja igual a zero. Segundo Fitzgerald, Gottschalk e Moffitt (1998), 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:

$$\varepsilon_{2,it} \perp \varepsilon_{1,it}, \text{ mas } x_{2,it} \not\perp \varepsilon_{1,it} \quad (16)$$

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 a $x_{1,it}$. Ou seja, $x_{2,it}$ é endógena a 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 a 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, Gottschalk e Moffitt (1998) colocam que, em um painel, os valores defasados da variável de interesse, y_{it-1} , y_{it-2} etc., podem cumprir o papel da 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:

$$f(y_{it} | x_{1,it}) = \int_{x_{2,it}} g(y_{it}, x_{2,it} | x_{1,it}, d_{it} = 1) w(x_{1,it}, x_{2,it}) dx_{2,it} \quad (17)$$

onde:

$$w(x_{1,it}, x_{2,it}) = \left[\frac{\Pr(d_{it} = 1 | x_{1,it}, x_{2,it})}{\Pr(d_{it} = 1 | x_{1,it})} \right]^{-1} \quad (18)$$

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

Wooldridge (2001) registra 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 ESTIMATIVA DOS DETERMINANTES DA PERMANÊNCIA NO PAINEL DA PME

Peracchi e Welch (1995) analisam o problema do desgaste na Current Population Survey (CPS). 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 indicam que a principal causa do atrito nesta pesquisa é o fracasso em seguir os jovens em idade de freqüentar a universidade em domicílios emparelhados e de acompanhar famílias jovens que mudam de endereço. A conclusão de seu trabalho é que grande parte do desgaste da CPS resulta da mobilidade das pessoas ligadas às decisões sobre educação, formação de novas famílias e procura de emprego.

Lopes (2002) realiza uma análise semelhante à desses autores, porém na PME. Particularmente, Lopes utiliza dados de março de 1996, 1997, 1998 e 1999 para a RM do Rio de Janeiro para estimar a probabilidade de as pessoas permanecerem na amostra após um ano. Os resultados de seu trabalho indicam 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 a pessoa 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 tanto da antiga pesquisa quanto da nova.⁵ Em uma das estimativas, as informações são referentes aos anos de 1997, 1998 e 1999, para as seis RMs. Na outra estimativa, as informações são novamente referentes às seis RMs, porém utilizando os grupos rotacionais totalmente compreendidos entre os anos de 2002 e 2006. Portanto, as duas estimativas fazem uma comparação entre os resultados utilizando a antiga e a nova metodologia da PME. As pessoas incluídas em ambas as amostras são aquelas entre a primeira e a sétima entrevista. Cabe salientar ainda que todos os painéis

5. 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).

utilizados neste artigo foram reconstituídos através do algoritmo proposto por Ribas e Soares (2008).

A tabela 1 apresenta nossas estimativas dos efeitos marginais sobre a probabilidade de permanência na amostra, *vis-à-vis* o atrito, utilizando dados da antiga PME. Primeiramente, podemos constatar que, sendo constantes as demais características observáveis, a probabilidade de desgaste é maior na RM de Recife (categoria de referência), seguida pelas RMs de Rio de Janeiro, São Paulo e Salvador. Para indivíduos de características iguais, aqueles com maiores probabilidades de permanência na amostra ao longo do painel residiam nas RMs de Belo Horizonte e Porto Alegre.

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 de Salvador | 0.02071*** | 0.02009*** | 0.02010*** | 0.02012*** | 0.02079*** | 0.02072*** |
| RM de Belo Horizonte | 0.03213*** | 0.01922*** | 0.01920*** | 0.01926*** | 0.03231*** | 0.03193*** |
| RM do Rio de Janeiro | 0.01780*** | 0.01791*** | 0.01788*** | 0.01789*** | 0.01789*** | 0.01764*** |
| RM de São Paulo | 0.01847*** | 0.01870*** | 0.01880*** | 0.01882*** | 0.01866*** | 0.01783*** |
| RM de 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 |
| Março | 0.00582*** | 0.00562*** | 0.00591*** | 0.00547*** | 0.00587*** | 0.00588*** |
| Abril | 0.00902*** | 0.00846*** | 0.00866*** | 0.00806*** | 0.00903*** | 0.00909*** |
| Mai | 0.00672*** | 0.00685*** | 0.00690*** | 0.00611*** | 0.00677*** | 0.00674*** |
| Junho | 0.00720*** | 0.00726*** | 0.00754*** | 0.00586*** | 0.00723*** | 0.00724*** |
| Julho | 0.00979*** | 0.00990*** | 0.01014*** | 0.00774*** | 0.00981*** | 0.00986*** |
| Agosto | 0.01304*** | 0.01232*** | 0.01252*** | 0.00943*** | 0.01304*** | 0.01310*** |
| Setembro | 0.01153*** | 0.01155*** | 0.01164*** | 0.00781*** | 0.01153*** | 0.01155*** |
| Outubro | 0.00760*** | 0.00793*** | 0.00820*** | 0.00537*** | 0.00762*** | 0.00763*** |
| Novembro | -0.00020 | 0.00016 | 0.00047 | -0.00149 | -0.00021 | -0.00019 |
| 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 | -0.01775*** | -0.01673*** | -0.01668*** | -0.01667*** | -0.01341*** | -0.01766*** |
| Entre 25 e 29 | -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 | -0.00671*** | -0.00641*** | -0.00640*** | -0.00641*** | -0.00233 | -0.00673*** |

(continua)

(continuação)

| Variável | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Entre 40 e 44 | -0.00330* | -0.00332* | -0.00332* | -0.00333* | 0.00082 | -0.00325* |
| Entre 45 e 49 | -0.00112 | -0.00128 | -0.00130 | -0.00131 | 0.00263 | -0.00102 |
| Entre 50 e 54 | 0.00128 | 0.00118 | 0.00112 | 0.00109 | 0.00431* | 0.00136 |
| Entre 55 e 59 | 0.00694*** | 0.00643*** | 0.00638*** | 0.00636*** | 0.00919*** | 0.00697*** |
| Entre 60 e 64 | 0.00878*** | 0.00815*** | 0.00809*** | 0.00806*** | 0.01025*** | 0.00881*** |
| Com 65 anos ou + | 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 | 0.00199 | 0.00156 | 0.00153 | 0.00153 | 0.00242* | 0.00181 |
| De 8 a 10 anos | 0.00531*** | 0.00452*** | 0.00444*** | 0.00442*** | 0.00594*** | 0.00505*** |
| 11 anos ou + | 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 | 0.01811 | 0.02028* | 0.02044* | 0.02056* | 0.01742 | 0.01821 |
| Entre 30 e 34 | 0.02368* | 0.02501** | 0.02524** | 0.02536** | 0.02288* | 0.02382* |
| Entre 35 e 39 | 0.02944** | 0.03038*** | 0.03052*** | 0.03065*** | 0.02867** | 0.02952** |
| Entre 40 e 44 | 0.03313*** | 0.03386*** | 0.03401*** | 0.03414*** | 0.03229*** | 0.03311*** |
| Entre 45 e 49 | 0.03535*** | 0.03582*** | 0.03598*** | 0.03609*** | 0.03446*** | 0.03534*** |
| Entre 50 e 54 | 0.03882*** | 0.03869*** | 0.03880*** | 0.03890*** | 0.03798*** | 0.03872*** |
| Entre 55 e 59 | 0.03924*** | 0.03917*** | 0.03927*** | 0.03937*** | 0.03841*** | 0.03905*** |
| Entre 60 e 64 | 0.04098*** | 0.04056*** | 0.04068*** | 0.04077*** | 0.04016*** | 0.04062*** |
| Com 65 ou + | 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 | 0.00158 | 0.00132 | 0.00128 | 0.00126 | 0.00120 | 0.00124 |
| De 8 a 10 anos | 0.00169 | 0.00142 | 0.00130 | 0.00128 | 0.00112 | 0.00121 |
| 11 ou + | -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*** | | |

(continua)

(continuação)

| Variável | (1) | (2) | (3) | (4) | (5) | (6) |
|-------------------------------|----------|-----------|-----------|------------|-------------|-------------|
| 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*** |
| 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 |
| Número de observações | | 1716958 | 1716958 | 1716958 | 1716958 | 1716375 |
| Log <i>likelihood</i> | | -417026.6 | -403527.2 | -403391.6 | -403366.6 | 416701.3 |
| Pseudo R^2 | | 0.0202 | 0.0519 | 0.0523 | 0.0523 | 0.0206 |

(continua)

(continuação)

| Variável | (1) | (2) | (3) | (4) | (5) | (6) | |
|-----------------------------------|----------|--------|-----|---------------------------------------|----------|--------|-----|
| Wald <i>test</i> | chi d.f. | Modelo | | Wald <i>test</i> | 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: PMEs de 1997, 1998 e 1999.

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

* , ** e *** significativos a 5%, 1% e 0,1%, respectivamente.

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 dos meses de maio e junho, e crescente a partir de setembro. Dezembro é o mês em que a probabilidade de desgaste é maior, seguido de novembro, janeiro e fevereiro. Uma das causas desse resultado pode ser a ausência de pessoas nos domicílios para responder à entrevista nesses meses, considerados como dentro de um período de férias.

O aumento de 1% no número de membros dividindo o mesmo domicílio aumenta em 1,4 ponto percentual (p.p.) a probabilidade de as pessoas permanecerem na amostra. Além disso, a probabilidade de atrito é menor nos domicílios com maior proporção de jovens e adolescentes. 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 aquelas que representam menor probabilidade de atrito. A justificativa para todos esses efeitos encontrados é a menor mobilidade geográfica das pessoas com tais características. Acreditamos que homens, jovens, com pouca escolaridade e residentes em domicílios menores, geralmente unipessoais, compõem o grupo de pessoas mais propenso à migração. Ainda nesse sentido, um resultado esperado é de que as pessoas do núcleo familiar (chefe, cônjuge e filho), principalmente o cônjuge, são aqueles 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.

Nas regressões apresentadas na tabela 1, incluímos paulatinamente 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 decorre do elevado desgaste entre a quarta e a quinta entrevista, que são feitas com um intervalo de oito meses entre uma e outra.

Os domicílios na quarta entrevista possuem uma probabilidade de permanência no painel 8 p.p. menor do que os domicílios na primeira entrevista. Com 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-se que, ao longo do painel, o percentual de pessoas impacientes diminua e o de pessoas pacientes aumente de modo conseqüente, essa mudança na composição da amostra seria a responsável por esse 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. Tal resultado pode ser explicado por fatores motivacionais tanto por parte dos entrevistados como por parte do entrevistador. Alguns conjuntos da amostra, chamados “painéis”, possuem ainda maior probabilidade de desgaste que outros. Essa diferença possivelmente decorre 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 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.

A tabela 2 apresenta o mesmo tipo de estimativa, porém utilizando dados coletados pela nova PME. Neste caso, os erros-padrão das regressões foram calculados considerando-se o novo desenho amostral da pesquisa, que é estratificado em dois estágios. De acordo com essa tabela, Recife continua sendo uma das RMs com maior probabilidade de desgaste. Porém, na nova pesquisa, ela está acompanhada de Porto Alegre. 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 aquelas com maiores taxas de permanência no painel. Entre os meses da pesquisa, identificamos novamente que agosto é o mês com menor taxa de desgaste, seguido de julho, maio e setembro. No período de férias, que inclui os meses de dezembro, janeiro e fevereiro, as taxas de desgaste tendem a aumentar.

TABELA 2

Efeitos marginais sobre a probabilidade de permanecer na amostra na entrevista seguinte para a nova PME

| Variável | (1) | (2) | (3) | (4) | (5) | (6) |
|---------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| RM de Salvador | 0.01197*** | 0.01525*** | 0.01530*** | 0.01532*** | 0.01217*** | 0.01200*** |
| RM de Belo Horizonte | 0.01436*** | 0.01400*** | 0.01405*** | 0.01403*** | 0.01452*** | 0.01426*** |
| RM do Rio de Janeiro | 0.03617*** | 0.03746*** | 0.03745*** | 0.03784*** | 0.03635*** | 0.03620*** |
| RM de São Paulo | 0.01890*** | 0.02049*** | 0.02053*** | 0.02055*** | 0.01904*** | 0.01883*** |
| RM de Porto Alegre | -0.00060 | -0.00126 | -0.00119 | -0.00113 | -0.00071 | -0.00102 |
| Mês | | | | | | |
| Fevereiro | -0.00082 | 0.00265 | 0.00290 | 0.00278 | -0.00082 | -0.00081 |
| Março | -0.00290 | 0.00579* | 0.00653** | 0.00632** | -0.00287 | -0.00285 |
| Abril | -0.01207** | 0.00210 | 0.00298 | 0.00277 | -0.01204** | -0.01204** |
| Maio | 0.00767** | 0.00741** | 0.00739** | 0.00724** | 0.00771** | 0.00771** |
| Junho | 0.00066 | -0.00381 | -0.00409 | -0.00325 | 0.00068 | 0.00072 |
| Julho | 0.01408*** | 0.00638** | 0.00619** | 0.00775*** | 0.01408*** | 0.01412*** |
| Agosto | 0.01974*** | 0.00896*** | 0.00914*** | 0.01130*** | 0.01974*** | 0.01978*** |
| Setembro | 0.00698** | 0.00695** | 0.00693** | 0.00980*** | 0.00699** | 0.00699** |
| Outubro | 0.00458 | 0.00548* | 0.00544* | 0.00761** | 0.00458 | 0.00456 |
| Novembro | 0.00297 | 0.00387 | 0.00416 | 0.00575* | 0.00297 | 0.00296 |
| Dezembro | -0.00710 | -0.00787*** | -0.00750** | -0.00644** | -0.00709 | -0.00711 |
| Log do número de membros do domicílio | 0.02753*** | 0.02687*** | 0.02691*** | 0.02703*** | 0.02759*** | 0.02763*** |
| Proporção de | | | | | | |
| Crianças com menos de 10 anos | -0.00954** | -0.00488 | -0.00498 | -0.00466 | -0.00953** | -0.00932** |
| Adolescente entre 10 e 15 anos | 0.01181*** | 0.01272*** | 0.01277*** | 0.01294*** | 0.01221*** | 0.01222*** |
| Jovens entre 16 e 18 anos | 0.01161** | 0.01318*** | 0.01317*** | 0.01350*** | 0.01208** | 0.01208** |
| Idade da pessoa | | | | | | |
| Entre 16 e 18 anos | -0.00592*** | -0.00540*** | -0.00540*** | -0.00533*** | -0.00497*** | -0.00581*** |
| Entre 19 e 24 | -0.01455*** | -0.01381*** | -0.01382*** | -0.01362*** | -0.01291*** | -0.01441*** |
| Entre 25 e 29 | -0.01470*** | -0.01451*** | -0.01451*** | -0.01445*** | -0.01309*** | -0.01459*** |
| Entre 30 e 34 | -0.01018*** | -0.00985*** | -0.00982*** | -0.00983*** | -0.00860*** | -0.01019*** |
| Entre 35 e 39 | -0.00537*** | -0.00531*** | -0.00530*** | -0.00527*** | -0.00382** | -0.00540*** |
| Entre 40 e 44 | -0.00296** | -0.00323** | -0.00318** | -0.00320** | -0.00149 | -0.00291** |
| Entre 45 e 49 | 0.00259* | 0.00212 | 0.00214 | 0.00202 | 0.00397** | 0.00272* |
| Entre 50 e 54 | 0.00608*** | 0.00577*** | 0.00583*** | 0.00574*** | 0.00736*** | 0.00627*** |
| Entre 55 e 59 | 0.00946*** | 0.00864*** | 0.00868*** | 0.00844*** | 0.01047*** | 0.00966*** |
| Entre 60 e 64 | 0.00930*** | 0.00894*** | 0.00901*** | 0.00878*** | 0.01007*** | 0.00955*** |
| Com 65 anos ou + | 0.01123*** | 0.01084*** | 0.01091*** | 0.01077*** | 0.01148*** | 0.01141*** |
| Mulher | 0.00427*** | 0.00433*** | 0.00433*** | 0.00438*** | 0.00433*** | 0.00421*** |

(continua)

(continuação)

| Variável | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Escolaridade da pessoa | | | | | | |
| De 1 a 3 anos de estudo | 0.00363** | 0.00338** | 0.00341** | 0.00318** | 0.00383** | 0.00375** |
| De 4 a 7 | 0.00418*** | 0.00384** | 0.00388*** | 0.00362** | 0.00434*** | 0.00424*** |
| De 8 a 10 | 0.00678*** | 0.00606*** | 0.00611*** | 0.00573*** | 0.00666*** | 0.00674*** |
| 11 ou + | 0.01020*** | 0.00923*** | 0.00928*** | 0.00873*** | 0.00966*** | 0.01020*** |
| Condição no domicílio | | | | | | |
| Cônjuge | 0.00243*** | 0.00234*** | 0.00234*** | 0.00226*** | 0.00208*** | 0.00242*** |
| Filho | -0.00517*** | -0.00526*** | -0.00522*** | -0.00539*** | -0.00559*** | -0.00510*** |
| Agregado | -0.04581*** | -0.04401*** | -0.04401*** | -0.04403*** | -0.04602*** | -0.04588*** |
| Empregado ou pensionista | -0.07778*** | -0.07557*** | -0.07564*** | -0.07529*** | -0.07287*** | -0.07720*** |
| Idade do chefe | | | | | | |
| Entre 16 e 18 anos | 0.14480*** | 0.13105*** | 0.13129*** | 0.13074*** | 0.14427*** | 0.14414*** |
| Entre 19 e 24 anos | 0.16990*** | 0.15542*** | 0.15571*** | 0.15527*** | 0.16907*** | 0.16861*** |
| Entre 25 e 29 | 0.18351*** | 0.16894*** | 0.16921*** | 0.16866*** | 0.18265*** | 0.18174*** |
| Entre 30 e 34 | 0.18611*** | 0.17132*** | 0.17161*** | 0.17104*** | 0.18528*** | 0.18445*** |
| Entre 35 e 39 | 0.19433*** | 0.17893*** | 0.17919*** | 0.17869*** | 0.19349*** | 0.19268*** |
| Entre 40 e 44 | 0.19867*** | 0.18295*** | 0.18319*** | 0.18253*** | 0.19787*** | 0.19699*** |
| Entre 45 e 49 | 0.20094*** | 0.18513*** | 0.18540*** | 0.18482*** | 0.20014*** | 0.19927*** |
| Entre 50 e 54 | 0.20350*** | 0.18751*** | 0.18776*** | 0.18702*** | 0.20268*** | 0.20208*** |
| Entre 55 e 59 | 0.20452*** | 0.18829*** | 0.18855*** | 0.18785*** | 0.20365*** | 0.20318*** |
| Entre 60 e 64 | 0.20561*** | 0.18964*** | 0.18997*** | 0.18931*** | 0.20468*** | 0.20425*** |
| Com 65 ou + | 0.20953*** | 0.19325*** | 0.19349*** | 0.19272*** | 0.20855*** | 0.20801*** |
| Chefe mulher | -0.00611*** | -0.00553*** | -0.00548*** | -0.00569*** | -0.00598*** | -0.00497*** |
| Escolaridade do chefe | | | | | | |
| De 1 a 3 anos de estudo | 0.01052*** | 0.01062*** | 0.01060*** | 0.01039*** | 0.01047*** | 0.01057*** |
| De 4 a 7 anos | 0.01251*** | 0.01163*** | 0.01163*** | 0.01127*** | 0.01232*** | 0.01233*** |
| De 8 a 10 anos | 0.01242*** | 0.01163*** | 0.01162*** | 0.01142*** | 0.01216*** | 0.01183*** |
| 11 anos de estudo ou + | 0.00924*** | 0.00810*** | 0.00805*** | 0.00774*** | 0.00893*** | 0.00805*** |
| Domicílio na | | | | | | |
| 2ª entrevista | (i) | 0.00800*** | 0.00798*** | 0.00804*** | | |
| 3ª entrevista | (ii) | 0.00917*** | 0.00909*** | 0.00928*** | | |
| 4ª entrevista | (iii) | -0.14455*** | -0.14460*** | -0.14442*** | | |
| 5ª entrevista | (iv) | 0.00926*** | 0.00926*** | 0.00932*** | | |
| 6ª entrevista | (v) | 0.01553*** | 0.01555*** | 0.01558*** | | |
| 7ª entrevista | (vi) | 0.01889*** | 0.01885*** | 0.01896*** | | |
| Entrevista ocorrida na | | | | | | |
| 2ª semana | (vii) | | -0.00151 | -0.00147 | | |
| 3ª semana | (viii) | | -0.00487 | -0.00475 | | |

(continua)

(continuação)

| Variável | | (1) | (2) | (3) | (4) | (5) | (6) |
|-------------------------------|-----------|-----------|-----------|-----------|------------|--------------|-------------|
| 4ª semana | (ix) | | | -0.00647* | -0.00627* | | |
| Painel E | (x) | | | | 0.00644*** | | |
| Painel F | (xi) | | | | 0.01530*** | | |
| Painel G | (xii) | | | | 0.01720*** | | |
| Pessoa ocupada | (xiii) | | | | | -0.00517 *** | |
| Pessoa desempregada | (xiv) | | | | | -0.00176 | |
| Setor de ocupação | | | | | | | |
| Indústria | (xv) | | | | | 0.00551 *** | |
| Construção civil | (xvi) | | | | | -0.00289 | |
| Serviços | (xvii) | | | | | 0.00068 | |
| Público | (xviii) | | | | | 0.00411 *** | |
| Doméstico | (xix) | | | | | -0.00530 *** | |
| outro | (xx) | | | | | 0.00079 | |
| Condição na ocupação | | | | | | | |
| Conta-própria | (xxi) | | | | | 0.00502 *** | |
| Empregador | (xxii) | | | | | -0.00631 ** | |
| Não-remunerado | (xxiii) | | | | | 0.00717 | |
| Tem carteira assinada | (xxiv) | | | | | 0.00306 ** | |
| Chefe ocupado | (xxv) | | | | | | -0.00479 * |
| Chefe desempregado | (xxvi) | | | | | | -0.00368 |
| Setor de ocupação do chefe | | | | | | | |
| Indústria | (xxvii) | | | | | | 0.00546 *** |
| Construção civil | (xxviii) | | | | | | -0.00023 |
| Serviços | (xxix) | | | | | | 0.00011 |
| Público | (xxx) | | | | | | 0.00382 * |
| Doméstico | (xxxi) | | | | | | -0.00486 |
| Outro | (xxxii) | | | | | | 0.00123 |
| Condição na ocupação do chefe | | | | | | | |
| Conta-própria | (xxxiii) | | | | | | 0.00475 ** |
| Empregador | (xxxiv) | | | | | | -0.00364 |
| Não-remunerado | (xxxv) | | | | | | 0.00030 |
| Tem carteira assinada | (xxxvi) | | | | | | 0.00645 *** |
| Probabilidade média predita | 0.9152622 | 0.9152293 | 0.9152306 | 0.9152402 | 0.9152616 | 0.9152622 | |
| Número de observações | 2362683 | 2362683 | 2362683 | 2362683 | 2362681 | 2362681 | |
| Tamanho da população | 1.058E+09 | 1.06E+09 | 1.06E+09 | 1.06E+09 | 1.06E+09 | 1.06E+09 | |
| Número de estratos | 131 | 131 | 131 | 131 | 131 | 131 | |

(continua)

(continuação)

| Variável | (1) | (2) | (3) | (4) | (5) | (6) | |
|-----------------------------------|------------|----------------|--|------------------|----------------|------------|--------|
| Número de UPAs | 2146 | 2146 | 2146 | 2146 | 2146 | 2146 | |
| Estatística-F | 140.97 | 437.78 | 421.53 | 417.25 | 119.54 | 119.86 | |
| d.f. | (56, 1960) | (62, 1954) | (65, 1951) | (68, 1948) | (68, 1948) | (68, 1948) | |
| Wald <i>test</i> | F | d.f. | Modelo | Wald <i>test</i> | chi | d.f. | Modelo |
| (i) = (ii) = ... = (vi) = 0 | 3148.25 | (6, 2010) *** | (2) (vii) = (viii) = (ix) = 0 | 2.79 | (3, 2013) * | (3) | |
| (i) = (ii) = 0 | 13.54 | (2, 2014) *** | (2) (i) = (ii) = ... = (ix) = 0 | 2111.94 | (9, 2007) *** | (3) | |
| (iii) = 0 | 6966.97 | (1, 2015) *** | (2) (x) = (xi) = (xii) = 0 | 50.83 | (3, 2013) *** | (4) | |
| (iv) = (v) = (vi) | 13.80 | (2, 2014) *** | (2) (i) = (ii) = ... = (xii) = 0 | 1655.88 | (12, 2004) *** | (4) | |
| (xiii) = (xiv) = 0 | 9.85 | (2, 2014) *** | (5) (xxv) = (xxvi) = 0 | 4.19 | (2, 2014) * | (6) | |
| (xv) = (xvi) = ... = (xx) = 0 | 10.74 | (6, 2010) *** | (5) (xxvii) = (xxviii) = ... = (xxxii) = 0 | 4.83 | (6, 2010) *** | (6) | |
| (xxi) = (xxii) = (xxiii) = 0 | 16.08 | (3, 2013) *** | (5) (xxxiii) = (xxxiv) = (xxxv) = 0 | 6.89 | (3, 2013) *** | (6) | |
| (xxiv) = 0 | 9.96 | (1, 2015) ** | (5) (xxxvi) = 0 | 19.92 | (1, 2015) *** | (6) | |
| (xiii) = (xiv) = ... = (xxiv) = 0 | 13.44 | (12, 2004) *** | (5) (xxv) = (xxvi) = ... = (xxxvi) = 0 | 8.84 | (12, 2004) *** | (6) | |

Fonte: PMEs de 2002, 2003, 2004, 200 e 2006.

Nota: Regressões realizadas utilizando os painéis D, E, F e G.

*, ** e *** significativos a 5%, 1% e 0,1%, respectivamente.

Em relação às características individuais e domiciliares, os resultados da tabela 2 são, de maneira geral, próximos dos resultados da tabela 1. Entre as características do processo de entrevista, confirmamos que pessoas na quarta entrevista são aquelas com maiores chances de atrito e que, à exceção dessa quarta, a probabilidade de permanência aumenta com o número de entrevista. Outro resultado que se confirma é que, quanto mais recente o painel, menor é a probabilidade de desgaste. Comparando-se os efeitos da semana de entrevista, verifica-se que, na nova PME, eles não são tão significativos quanto eram na antiga PME.

Nas características socioeconômicas incluídas na tabela 2, os coeficientes relacionados ao desemprego não são significativos. Porém, todas as outras combinações de variáveis possuem significância conjunta. Com uma classificação mais desagregada de setores de atividade, identificamos que, na nova PME, as menores taxas de atrito estão relacionadas à indústria e ao serviço público. O maior desgaste, por sua vez, está entre os trabalhadores domésticos. Empregados por conta própria ou com carteira assinada possuem maior probabilidade de permanência no painel que os demais trabalhadores.

Em suma, podemos concluir nesta seção que, controladas as demais características, a taxa de atrito é maior na RM de Recife e que, nas RMs de São Paulo e Rio de Janeiro, foi onde a taxa de retenção mais aumentou relativamente com a passagem da antiga para a nova PME. O desgaste possui ainda certa sazonalidade, com maior taxa nos meses de férias (dezembro, janeiro e fevereiro). Características individuais fortemente relacionadas ao atrito são aquelas geralmente associadas a uma mobilidade geográfica maior. De fato, pessoas do sexo masculino, jovens, com pouca escolaridade, residentes em domicílios menores, geralmente unipessoais, ou que não fazem parte do núcleo familiar são as mais propensas ao atrito no painel. 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 as características do processo de entrevista quanto os fatores socioeconômicos são significativamente relacionados à probabilidade de retenção no painel. Nas características da entrevista, podemos destacar que, com exceção da quarta entrevista, a probabilidade de permanência é crescente com o tempo de participação no painel e que, devido a melhoras no processo, a taxa de atrito tende a diminuir ao longo do tempo.

4 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 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. Essa 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, sendo que os modelos multivariados com três ou mais componentes foram estimados por máxima verossimilhança simulada.⁶

4.1 TRANSIÇÃO PARA O DESEMPREGO

O modelo de transição para o desemprego é definido como:

$$y_{it}^* | y_{it-1}^* \leq 0 = x'_{1,t-1} \alpha_1 + x'_{1,i} \beta_1 + \varepsilon_{1,it} \quad (19)$$

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:

6. Modelos multivariados com mais de dois componentes endógenos só podem ser estimados utilizando-se 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).

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

$$y_{it-1}^* = x'_{2,t-1}\alpha_2 + x'_{2,i}\beta_2 + \varepsilon_{2,it-1} \quad (20)$$

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 (19) 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:

$$d_{it}^* = x'_{3,t-1}\alpha_3 + x'_{3,i}\beta_3 + \varepsilon_{3,it} \quad (21)$$

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 junto com as equações (19), (20) e (21).

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

a) 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);

b) 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);

c) 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);

d) 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); e

e) o modelo (5) é um *probit* quase idêntico ao modelo (1), porém utilizando um peso amostral corrigido pela probabilidade bivariada de estar inicialmente no emprego e permanecer na amostra, tal como descrito na equação (18) (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 essas 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 se mostraram 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. 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 muito bem a condição inicial.

A tabela 3 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 (19) e (21), 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 significa 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.

TABELA 3
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 ** |
| + de 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: PMEs de 1995, 1996, 1997, 1998, 1999, 2000, 2001 e 2002.

* e ** significativos a 5% e 1%, respectivamente.

De acordo com os coeficientes estimados, podemos notar que o efeito de variação na taxa de desemprego é sobreestimado 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-se 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 fosse incluído nenhum mecanismo.

4.2 O EFEITO DE CHOQUES TRANSITÓRIOS SOBRE A PROGRESSÃO ESCOLAR

O modelo de progressão escolar, apresentado por Duryea (1998), pode ser representado da seguinte forma:

$$y_{it}^* = x'_{1,it} \beta_1 + T'_{it} \gamma_1 + \varepsilon_{1,it} \quad (22)$$

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 (22), é um indicador do choque transitório sobre a renda domiciliar. De acordo com Duryea, a identificação de choques exógenos na renda permanente é extremamente difícil empiricamente, principalmente utilizando-se um painel curto como o da PME. Portanto, a melhor opção para medir o efeito desses 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 freqüente a escola no período t . Seja f_{it}^* a variável latente que representa a propensão à freqüência à escola da criança i no período t , podemos definir a seguinte equação:

$$f_{it}^* = x'_{2,it} \beta_2 + \varepsilon_{2,it} \quad (23)$$

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 (22) é de que esse chefe esteja inicialmente empregado. Assim, estimamos a probabilidade desse evento através da seguinte equação:

$$e_{it}^* = x'_{3,it} \beta_3 + \varepsilon_{3,it} \quad (24)$$

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 (22), 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 (23) e (24) definem a condição do chefe do domicílio e da criança justamente na primeira entrevista. Nas segunda, terceira e quarta entrevistas, é que pode ocorrer o possível choque transitório, o qual afetaria a progressão escolar. Na quinta entrevista é que se confirma se a criança foi aprovada ou não no curso que freqüentava.

Dessa forma, a última restrição imposta à equação (22) 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:

$$d_{it}^* = x'_{4,it} \beta_4 + \varepsilon_{4,it} \quad (25)$$

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 da 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:

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

b) o modelo (2) é um *probit* bivariado que estima simultaneamente a probabilidade de a criança progredir na escola e a probabilidade de ela estar freqüentando 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);

c) 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 freqüentavam a escola na primeira entrevista e permaneceram no painel até a quinta entrevista (modelo de seleção amostral sobre não-observáveis);

d) o modelo (4) é um *probit* trivariado que estima simultaneamente a probabilidade de progressão, a probabilidade de freqüentar 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 à quinta entrevista (modelo de seleção amostral sobre não-observáveis);

e) o modelo (5) é um *probit* quase idêntico ao modelo (1), porém utilizando-se um peso amostral corrigido pela probabilidade bivariada de a criança freqüentar a escola e de o chefe estar empregado na primeira entrevista (modelo de seleção amostral sobre observáveis);

f) 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 freqüentavam a escola na primeira entrevista, ao mesmo tempo em que seus chefes estavam empregados (modelo de seleção amostral sobre não-observáveis); e

g) o modelo (7) é um *probit* tetravariado que estima simultaneamente a probabilidade de progressão, a probabilidade de freqüentar 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-se o desenho amostral complexo da nova PME.

Para instrumentalizar a probabilidade de freqüência à escola nos modelos (2), (4) e (7), utilizamos a proporção de demais crianças do setor censitário freqüentando 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.

Todas estas variáveis instrumentais, com exceção da proporção de demais chefes de domicílio ocupados no setor censitário, possuem um alto poder de explicação nas equações de seleção e um insignificante poder de explicação na equação de interesse. 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 4, verificamos que há correlações significativamente positivas entre os resíduos das equações de seleção. Dessa forma, uma probabilidade maior não-explicada de o chefe do domicílio estar empregado está relacionada a uma probabilidade maior de a criança freqüentar a escola e permanecer no painel. Em termos da correlação direta com a equação de interesse, no modelo (6), identificamos que uma probabilidade maior não-explicada de progressão na escola está relacionada a uma probabilidade maior 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 (22) e (23). Ou seja, quanto maiores as chances não-explicadas de a criança freqüentar a escola, menores são as chances de aprovação na série que freqüenta.

TABELA 4

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 – azeitado | 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 | -0.2014** | -0.1990** | -0.1968** | -0.2002** | -0.1978** | -0.1842** | -0.1994** |
| Com 13 | -0.2387** | -0.2310** | -0.2304** | -0.2348** | -0.2361** | -0.2302** | -0.2360** |
| Com 14 | -0.2873** | -0.2727** | -0.2770** | -0.2802** | -0.2822** | -0.2718** | -0.2810** |
| Com 15 | -0.3485** | -0.3187** | -0.3382** | -0.3331** | -0.3531** | -0.3217** | -0.3327** |
| Com 2 anos de atraso ou + | -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 |
| Número de crianças <10 anos | -0.0572** | -0.0532** | -0.0622** | -0.0556** | -0.0581** | -0.0395** | -0.0551** |
| Número de demais crianças com 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 com 1-3 anos de estudo | 0.0431 | 0.0339 | 0.0623 | 0.0426 | 0.0364 | -0.0095 | 0.0412 |
| Chefe com 4 anos de estudo | 0.0716 | 0.0501 | 0.1025 | 0.0660 | 0.0792 | -0.0269 | 0.0608 |
| Chefe com 5-7 anos de estudo | 0.1076 | 0.0895 | 0.1417 | 0.1040 | 0.1199 | 0.0169 | 0.0994 |
| Chefe com 8 anos de estudo | 0.1838** | 0.1626* | 0.2280** | 0.1782* | 0.1900** | 0.0897 | 0.1744* |
| Chefe com 9-10 anos de estudo | 0.1952* | 0.1747* | 0.2339* | 0.1900* | 0.2008* | 0.0760 | 0.1836 |
| Chefe com 11 anos de estudo | 0.2749** | 0.2490** | 0.3304** | 0.2671** | 0.2766** | 0.1481* | 0.2614** |
| Chefe com 12-14 anos de estudo | 0.3044** | 0.2765** | 0.3606** | 0.2941* | 0.3038** | 0.2249* | 0.2959** |
| Chefe com 15 anos ou + de estudo | 0.2802** | 0.2528** | 0.3700** | 0.2716* | 0.2839** | 0.1521* | 0.2669** |
| RM de Salvador | -0.1254* | -0.1251* | -0.1142* | -0.1250* | -0.1236* | -0.0820 | -0.1219* |
| RM de Belo Horizonte | 0.1065** | 0.1071** | 0.1261** | 0.1087* | 0.1082** | 0.0334 | 0.1040* |
| RM do Rio de Janeiro | -0.0292 | -0.0336 | 0.0152 | -0.0293 | -0.0265 | -0.1588** | -0.0393 |
| RM de São Paulo | 0.1871** | 0.1884** | 0.2034** | 0.1897** | 0.1838** | 0.1357** | 0.1879** |
| RM de 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: PMEs de 2002, 2003, 2004, 2005, 2006 e 2007.

* e ** significativos a 5% e 1%, respectivamente.

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 desse 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 esses coeficientes estão sobreestimados, possivelmente devido à ausência do controle de seleção sobre a frequência à escola e sobre a retenção no painel.

Entretanto, deve-se ter cuidado na interpretação dos resultados do modelo (6), em que somente o controle de retenção no painel é considerado. Os resultados desse 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.

5 CONCLUSÃO

A finalidade deste artigo foi levantar alguns aspectos sobre como lidar com o desgaste em painéis, usando-se o painel da PME como exemplo.

Em relação aos determinantes do desgaste, concluímos que a taxa de atrito tende a ser maior na RM de Recife e que, nas RMs de São Paulo e Rio de Janeiro, foi onde a taxa de retenção mais cresceu com a passagem da antiga para a nova PME. Identificamos ainda que o desgaste possui certa sazonalidade, com uma taxa maior nos meses de férias, e uma forte correlação com as características geralmente associadas a uma mobilidade geográfica maior. 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. O primeiro modelo era de transição mensal para o desemprego e o segundo modelo era de progressão de crianças na escola. 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 probabilidade maior não-explicada de permanecer no painel está relacionada a uma probabilidade maior 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 probabilidade maior não-explicada de o chefe do domicílio estar empregado está relacionada a uma probabilidade maior 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 alguns resultados apresentados, a ausência do controle para a 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

- ARENDT, J. N.; HOLM, A. *Probit models with binary endogenous regressors*. Department of Business and Economics, University of Southern Denmark, 2006 (Discussion Papers on Business and Economics, n. 4/2006).
- CAMARGO, J. M.; NERI, M. Structural reforms, macroeconomic fluctuations and income distribution in Brazil. *Economic Reforms Series*, n. 39, Cepal, 1999.
- 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, n. 376).
- DURYEA, S.; 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.; MOFFITT, 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. *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*. Alexandria: American Statistical Association, p. 571-578, 1991.
- GONZAGA, G.; REIS, M. C. Os efeitos trabalhador adicional e desalento no Brasil. ENCONTRO NACIONAL DE ECONOMIA, 33., 2005, João Pessoa. *Anais ...* João Pessoa: Anpec, 2005..

- HAIJIVASSILIOU, V. *Smooth simulation estimation of panel data LDV models*. Yale University, 1999. 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.
- IBGE. *Pesquisa Mensal de Emprego*. Rio de Janeiro: Departamento de Emprego e Rendimento, 2002 (Relatório Metodológico, n. 23).
- KEANE, M. P. A computationally practical simulation estimator for panel data. *Econometrica*, v. 46, n. 6, p. 931-959, 1994.
- LEMOS, 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, n. 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). Escola Nacional de Ciências Estatísticas (Ence), Rio de Janeiro, RJ, 2002. Orientadora: Denise Britz do Nascimento Silva.
- 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.; DA COSTA, D. R.; PINTO, A. 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. *LACEA 2000 Annual Meeting*, Rio de Janeiro, 2000.
- NICOLETTI, C.; PERACCHI, F. *Two-step estimation of binary response models with sample selection*. BHPS 2001 Conference, Colchester, 2001.
- PENIDO, M.; MACHADO, A. F. Duração do desemprego no Brasil metropolitano. In: WAJNMAN, S.; MACHADO, A. F. (Eds.). *Mercado de trabalho: uma análise a partir das pesquisas domiciliares no Brasil*. Editora UFMG, p. 203-218. 2003.
- 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*. Brasília: Ipea (Texto para Discussão). Forthcoming.

- 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, n. 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.

EDITORIAL

Coordenação

Iranilde Rego

Supervisão

Andrea Bossle de Abreu

Revisão

Lucia Duarte Moreira

Alejandro Sainz de Vicuña

Eliezer Moreira

Elisabete de Carvalho Soares

Míriam Nunes da Fonseca

Editoração

Roberto das Chagas Campos

Aeromilson Mesquita

Camila Guimarães Simas

Camila Oliveira de Souza

Carlos Henrique Santos Vianna

Brasília

SBS – Quadra 1 – Bloco J – Ed. BNDES,

9ª andar – 70076-900 – Brasília – DF

Fone: (61) 3315-5090

Fax: (61) 3315-5314

Correio eletrônico: editbsb@ipea.gov.br

Rio de Janeiro

Av. Nilo Peçanha, 50/609

20044-900 – Rio de Janeiro – RJ

Fone: (21) 3515-8522 – 3515-8426

Fax (21) 3515-8585

Correio eletrônico: editrj@ipea.gov.br

Tiragem: 135 exemplares