

Comparação entre metodologias de idade-período-coorte para o estudo de uma medida da progressão escolar no Brasil

Raquel Rangel de Meireles Guimarães*
Eduardo Luiz Gonçalves Rios-Neto**

Este artigo procura realizar um exercício de comparação metodológica de dois estimadores dos modelos idade-período-coorte: o estimador convencional obtido pelos modelos lineares generalizados restritos (MLGR); e o estimador intrínseco (EI). O objeto de interesse são as contribuições dos efeitos de idade, período e coorte nas mudanças temporais na probabilidade de progressão (PPS) para a 5ª série do ensino fundamental das mulheres brasileiras. A modelagem IPC se justifica uma vez que os efeitos de idade, período e coorte podem impactar de forma significativa a probabilidade de progressão escolar: efeitos de idade refletem tanto a idade mínima de entrada no sistema de ensino como também o dilema entre trabalho e estudo que surge ao longo da carreira educacional; efeitos de período estão associados às diferentes conjunturas econômica e política, bem como ao estado das políticas educacionais; por fim, efeitos de coorte refletem características sociais peculiares a determinados grupos de indivíduos. Os dois instrumentais foram contrapostos em termos da eficiência, significância e estimativa dos parâmetros. Os resultados revelam a potencialidade da solução para o modelo IPC baseada no estimador intrínseco, o qual apresenta parâmetros com variância inferior aos estimados pelo modelo linear generalizado restrito. Desta forma, projeções das PPS baseadas na extrapolação do erro-padrão dos parâmetros obtidos pelo estimador intrínseco podem ser promissoras.

Palavras-chave: Modelos idade-período-coorte. Estimador intrínseco. Probabilidade de progressão por série.

Introdução

A aplicação dos modelos idade-período-coorte (IPC) tem sido objeto de intenso debate na Demografia desde 1970, especialmente a partir do trabalho de Mason et al. (1973). De forma sintética, estes modelos buscam avaliar em que medida um fenômeno de interesse estaria sendo determinado por variações de período, variações na idade, ou variações interativas entre período

e idade (coorte). Efeitos de idade estão associados ao diferencial nos riscos conforme a idade. Ademais, a idade pode refletir a evolução de processos biológicos, psicológicos e a mudança nos papéis sociais a cada faixa etária. Os efeitos de período, por sua vez, refletem as mudanças no fenômeno de interesse que atingem todos os grupos etários. Em geral, variações de período fornecem uma medida das condições ambientais, econômicas e sociais vigentes. Já

* Doutoranda em Demografia no Cedeplar/UFMG e mestranda em Educação Internacional Comparada pela Stanford University.

** Professor titular do Departamento de Demografia/Cedeplar/UFMG.

os efeitos de coorte podem estar refletindo mudanças entre um grupo de indivíduos com a mesma idade. Esses efeitos podem ser representados por mudanças genéticas ou sociais (RODGERS, 1982; HALLI; RAO, 1992; YANG; FU; LAND, 2004).

A principal controvérsia na aplicação dos modelos IPC é como lidar com o chamado *problema da identificação*. Como há uma dependência linear entre idade, período e coorte (período = idade + coorte), a matriz de design $X^T X$ é singular, ou seja, a inversa de $X^T X$ não existe. Por conseguinte, a solução para as equações estimáveis não é única e não é possível estimar os três efeitos separados sem que sejam impostas restrições de identificação. A usual estratégia para identificação do modelo IPC consiste na imposição de uma igualdade entre os parâmetros do modelo (FIENBERG; MASON, 1985). Todavia, esta solução recebeu muitas críticas, as quais tendem a reforçar o caráter *ad hoc* ou mesmo ateuórico da escolha das restrições de identificação (SMITH, 2004).

A partir de avanços na bioestatística, foram testadas funções que seriam invariáveis às restrições nos parâmetros dentro do arcabouço IPC. Entre estas, surgiu o estimador intrínseco (EI), descrito por Yang, Fu e Land (2004), baseado em funções estimáveis da decomposição do valor singular de matrizes. Este estimador, além de prover soluções únicas dos estimadores de idade, período e coorte, permite que o modelo seja identificado mediante mínimas suposições ou informações *a priori*. Ademais, os autores demonstraram que o EI apresenta propriedades estatísticas desejáveis nas análises IPC com períodos de tempo fixo (YANG; FU; LAND, 2004; YANG, 2008).

Nesse contexto, o objetivo deste artigo é realizar um exercício de comparação metodológica de dois estimadores IPC: o estimador restrito convencional obtido pelos modelos lineares generalizados (MLGR); e o estimador intrínseco (EI). Os dois instrumentais serão contrapostos em termos da estimativa dos parâmetros, do ajuste aos valores observados e da eficiência do modelo. O objeto de interesse são as contribuições dos efeitos de idade, período e coorte nas mudanças temporais na probabilidade de

progressão para a 5ª série do ensino fundamental entre as mulheres. A modelagem IPC das probabilidades de progressão por série possui uma clara interpretação substantiva: *efeitos de idade* refletem tanto a idade mínima de entrada no sistema de ensino como também o dilema entre trabalho e estudo que surge ao longo da carreira educacional; *efeitos de período* estão associados às diferentes conjunturas econômica e política, bem como ao estado das políticas educacionais; finalmente, *efeitos de coorte* refletem características sociais peculiares a determinados grupos de indivíduos.

A opção pelo estudo da probabilidade de progressão para a 5ª série do ensino fundamental deveu-se ao fato de que as chances de progressão nesta série se elevaram de forma considerável a partir da década de 1980 – em cerca de 20 pontos percentuais no período (GUIMARÃES, 2010). Ademais, verificou-se que não há um diferencial no comportamento ao longo do tempo nas probabilidades de progressão entre homens e mulheres e, desta forma, optou-se pela análise das probabilidades de progressão das mulheres.

Além de ser de grande importância substantiva, a análise IPC das probabilidades de progressão por série (PPS) possui aplicação para a elaboração de projeções educacionais. Uma vez que é demonstrada a existência de uma relação formal entre a PPS e os anos médios de estudo (RIOS-NETO, 2004), uma modelagem consistente das probabilidades de progressão por série permite, então, a construção de cenários factíveis das variações futuras nos componentes de idade, período e coorte. A partir destas probabilidades projetadas, Rios-Neto (2004) demonstrou que é possível obter os anos médios de estudo da população no futuro. Nesse sentido, a comparação entre metodologias adquire uma importância fundamental para que se obtenham projeções educacionais de qualidade e acuradas.

Este artigo está organizado em seis seções, incluindo esta introdução. Na segunda seção é apresentada uma revisão da literatura dos modelos IPC, em especial em relação às estratégias e soluções para contornar o problema da identificação. A ter-

ceira trata exclusivamente do estimador intrínseco descrito por Yang, Fu e Land (2004). A quarta descreve os dados e os passos metodológicos, enquanto a quinta seção reporta os procedimentos e os resultados da comparação entre o estimador restrito convencional e o estimador intrínseco para a análise da probabilidade de progressão das mulheres brasileiras para a 5ª série do ensino fundamental, ou seja, a probabilidade de conclusão da 5ª série, dado que concluiu a 4ª série (e_4). Finalmente, a sexta seção sumariza as evidências deste estudo e propõe uma agenda futura de pesquisas.

A questão da identificação dos modelos de idade-período-coorte: uma revisão

O problema da identificação dentro do arcabouço IPC pode ser descrito da seguinte forma. Considere um modelo linear geral cuja variável dependente seja uma taxa demográfica T_{ij} , expressa em termos da razão entre o número de ocorrências O_{ij} e expostos ao risco E_{ij} a cada grupo etário i e período j :

$$T_{ij} = \frac{O_{ij}}{E_{ij}} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij} \quad (1)$$

Neste modelo, μ é a constante do modelo, α corresponde aos parâmetros para os efeitos de idade e $i = 1, \dots, a$ é o indexador para o grupo etário, β compreende os parâmetros referentes aos efeitos de período, onde $j = 1, \dots, p$ indexa o período e γ representa os parâmetros para os efeitos de coorte e o indexador $k = a + p - 1$ denota a coorte. Ainda, ε_{ij} é o termo de erro aleatório para a idade i e período j cuja esperança é zero.

O modelo 1 se insere na classe dos chamados Modelos Lineares Generalizados (MLG). De acordo com Rios-Neto e Oliveira (1999), um MLG é construído a partir da escolha de uma função de ligação apropriada ao fenômeno de interesse e de uma distribuição de probabilidade para a variável resposta.

Neste artigo, a probabilidade de progressão por série é tratada como uma variável dicotômica que apresenta uma distribuição binomial. Ou seja, para cada combinação de idade-período tem-se a frequência absoluta

dos que progrediram em determinada transição escolar e daqueles que não progrediram. A partir destas quantidades, são obtidas as chances de progredir em relação a não progredir. A ligação canônica é, então, realizada pela função logística, resultando em um modelo logito da seguinte forma:

$$e_{ij} = \log \left(\frac{P_{ij}}{1 - P_{ij}} \right) = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij} \quad (2)$$

Onde e_{ij} expressa o logaritmo da razão das chances de progredir numa transição escolar para um determinado grupo etário e período e p_{ij} é a probabilidade de progressão na célula i, j . De acordo com Yang, Fu e Land (2004), o modelo 2 pode apenas ser operacionalizado pela centralização dos parâmetros ou por meio da imposição de identificação de uma de cada covariável de idade, período e coorte como sendo a categoria de referência.

Escolhida uma das operacionalizações para o modelo, pode-se reescrever o modelo 2 na forma de uma equação linear geral, onde Y representa o logaritmo da chance de progressão em cada uma das células, X é a matriz constituída por variáveis indicadoras com dimensão $m = 1 + (a-1) + (p-1) + (a+p-2)$ e ε é o vetor de erro aleatório:

$$Y = X\beta + \varepsilon \quad (3)$$

Onde β corresponde aos parâmetros populacionais, cujos estimadores amostrais são:

$$b = (\mu, \alpha_1, \dots, \alpha_{a-1}, \beta_1, \dots, \beta_{p-1}, \gamma_1, \dots, \gamma_{a+p-2})^T \quad (4)$$

A solução numérica para obtenção dos parâmetros amostrais para β por máxima verossimilhança do modelo 3 pode ser descrita por:

$$b = (X^T X)^{-1} (X^T Y) \quad (5)$$

A questão fundamental é que não existe apenas um vetor de coeficientes estimados para o modelo 3. Isto ocorre porque $X^T X$ não é invertível (ou seja, é uma matriz singular), devido a uma perfeita relação linear entre os efeitos de idade, período e coorte. Na literatura, denomina-se este impasse como *problema de identificação* na análise IPC. Por conseguinte, torna-se impossível separar as estimativas dos efeitos de coorte, idade e período sem que sejam impostas restrições adicionais aos coeficientes do

modelo, isto é, além da centralização ou da adoção de categorias de referência.

A partir desta limitação, iniciou-se nas ciências sociais e na epidemiologia um intenso debate sobre qual seria a melhor restrição de identificação a ser adotada nos modelos IPC. Mason et al. (1973) e Fienberg e Mason (1985) foram os primeiros a propor uma solução a este problema. De acordo com os autores, uma forma de tornar a matriz $X^T X$ invertível e, conseqüentemente, obter uma solução única para os parâmetros seria impor uma restrição de igualdade no vetor de coeficientes β . Dessa forma, bastaria admitir que os coeficientes do primeiro e segundo períodos, ou da primeira e segunda coortes, ou do primeiro e segundo grupos etários, por exemplo, fossem iguais.

Esta estratégia, contudo, mostrou-se sujeita a diversas fragilidades, as quais foram apontadas pela literatura. A primeira delas se pautou na dificuldade de separar os efeitos de idade, período e coorte por meio dos métodos estatísticos. Glenn (1976), por exemplo, argumentou pela necessidade de que, na aplicação do IPC, houvesse um esforço adicional para averiguar em que medida os efeitos de idade, período e coorte seriam aditivos. Dito de outra forma, seria necessário que o pesquisador se certificasse de que, por exemplo, os efeitos de idade sejam constantes a cada período ou coorte. Esta questão é relevante na medida em que os eventos históricos (período) podem afetar os indivíduos de forma distinta ao longo do ciclo de vida (idade). De forma análoga, pode-se suspeitar que os efeitos de idade podem não ser os mesmos entre várias coortes.

Outras críticas posteriores reforçaram o caráter *ad hoc* na restrição dos coeficientes do modelo, a possibilidade de que sejam obtidos efeitos falsos sobre o fenômeno ao se utilizar a proposta de Mason et al. (1973) e Fienberg e Mason (1985), ou mesmo a necessidade de um conhecimento profundo substantivo que forneça subsídios à correta identificação. Utilizando exercícios de simulação, Rodgers (1982) mostrou que a

adoção de diferentes estratégias de identificação pode não resultar em estimativas verdadeiras dos efeitos de idade, período e coorte. Sua principal contribuição foi argumentar, portanto, que a solução proposta por Mason et al. (1973) pode impactar de forma significativa as estimativas do modelo. Heckman e Robb (1985), por outro lado, argumentam que informações adicionais devem ser utilizadas para superar o problema de identificação, ou seja, os cientistas sociais deveriam aprimorar seu conhecimento de variáveis intervenientes ou *proxies* dos efeitos de idade, período e coorte. De uma perspectiva não tão crítica, Mason e Smith (1985) propõem que uma regra plausível para a seleção das restrições seria utilizar conhecimento *a priori* ou informações externas. Finalmente, a crítica feita por Rodgers (1982) e Yang et al. (2008) é que todo e qualquer modelo exatamente identificado a um conjunto de dados produzirá as mesmas medidas de qualidade de ajuste do modelo e, portanto, este fato inviabiliza o uso de qualquer critério para selecionar o modelo restrito com melhor ajuste.

A literatura recente que discute o problema da identificação recebeu contribuições significativas da bioestatística e da epidemiologia.¹ Em uma das linhas deste ramo de estudo que utiliza funções estimáveis, insere-se a abordagem baseada num novo estimador: o chamado estimador intrínseco, introduzido por Fu, Knight e Fu; Fu, Hall e Rohan (*apud* YANG; FU; LAND, 2004). Esse instrumento se assenta na decomposição do valor singular de matrizes e provê estimadores robustos dos efeitos de idade, período e coorte. Como a aplicação deste método é de interesse central neste artigo, sua formalização será apresentada na próxima seção.

Por fim, uma contribuição importante para as estratégias de identificação no arcabouço IPC foi dada por Winship e Harding (2008). Os autores basearam-se na abordagem dos modelos causais ou estruturais proposta por Pearl (2000) e desenvolveram uma metodologia na qual os mecanismos pelos quais a idade, o período e a coorte

¹ Para uma revisão das contribuições da bioestatística para os modelos IPC, ver Robertson, Gandini e Boyle (1999).

afetam a variável resposta são de interesse central. O teorema proposto pelos autores é que, para a identificação dos coeficientes no modelo IPC, é apenas necessário que sejam especificadas todas as variáveis intervenientes dos efeitos de idade, período e coorte, respectivamente. Por se tratar de uma abordagem estrutural, os autores consideram que o modelo poderá eventualmente estar sobreidentificado e, neste caso, deve-se testar o ajuste do modelo global e de seus componentes.

O estimador intrínseco (EI)

Nesta seção formaliza-se a construção do estimador intrínseco, bem como suas propriedades estatísticas, com base nos trabalhos de Yang, Fu e Land (2004), Yang (2008) e Yang et al. (2008). Considerando-se a equação 3, que expressa o modelo linear geral do modelo IPC, a dependência linear entre os efeitos de idade, período e coorte pode ser representada na forma matricial da seguinte maneira, a partir de um vetor B_0 não nulo:

$$XB_0 = 0 \tag{5}$$

A equação 5 é resultado do fato de que a matriz X é singular, isto é, existe alguma combinação linear das colunas da matriz de *design* X que resulta em um vetor nulo. Nos termos da Álgebra Linear, diz-se que a matriz X não possui *posto completo*.

Kupper et al. (*apud* YANG; FU; LAND, 2004) demonstraram que, se uma matriz possui posto menor que completo, seu espaço paramétrico pode ser decomposto na soma direta de dois subespaços lineares que são perpendiculares entre si:

$$P = N \oplus T \tag{6}$$

Onde: \oplus representa a soma direta dos dois subespaços lineares N e T , que são perpendiculares entre si; N é o espaço nulo com uma dimensão de X medido pelo vetor SB_0 com um número real S ; e T corresponde ao subespaço complementar ortogonal a N . Devido a esta decomposição ortogonal do espaço paramétrico, cada uma das infinitas soluções do modelo IPC irrestrito pode ser escrita como:

$$\hat{b} = B + SB_0 \tag{7}$$

Onde: S é um escalar que corresponde a uma solução específica para o problema da identificação e B_0 compreende um autovetor de norma euclidiana ou de tamanho 1. Yang, Fu e Land (2004) argumentam que este autovetor B_0 independe das taxas observadas Y e, portanto, é completamente determinado pelo número de grupos etários e períodos, ou seja, B_0 possui uma forma específica que é função tão somente da matriz de *design* X .

Kupper et al. (*apud* YANG; FU; LAND, 2004) mostraram que B_0 tem a seguinte forma:

$$B_0 = \frac{\tilde{B}_0}{\|\tilde{B}_0\|} \tag{8}$$

A implicação direta da equação 8 é que B_0 é o vetor normalizado de \tilde{B}_0 , o qual corresponde a:

$$\tilde{B}_0 = (0, I, P, C)^T \tag{9}$$

Onde:

$$I = \left(1 - \frac{i+1}{2}, \dots, [a-1] - \frac{i+1}{2} \right),$$

$$P = \left(\frac{p+1}{2} - 1, \dots, \frac{p+1}{2} - [p-1] \right),$$

$$C = \left(1 - \frac{i+p}{2} - 1, \dots, [a+p-2] - \frac{i+p}{2} \right)$$

e i , p e c denotam, respectivamente, os grupos etários, períodos e coortes.

Yang, Fu e Land (2004) ressaltam que a grande importância da equação 9 é que o vetor B_0 é fixo, ou seja, é independente da variável-resposta Y e, portanto, não possui nenhum papel na determinação dos coeficientes do modelo. Todavia, quando alguma restrição é imposta ao vetor de coeficientes, tal como propõem Fienberg e Mason (1985), então este princípio é violado na medida em que s , na equação 6, assume um valor diferente de zero.

Portanto, é possível demonstrar que qualquer estimador do modelo IPC pode ser descrito por uma restrição de identificação na matriz de *design* X , conforme a equação 6, e B é o denominado *estimador intrínseco* que é ortogonal ao espaço nulo, sendo determinado pela inversa generalizada de Moore-Penrose.

Yang et al. (2008) sugerem o seguinte algoritmo computacional para que se obtenha o estimador intrínseco, com base na regressão por componentes principais:

- obtenha os autovalores e os autovetores (componentes principais) da matriz $X^T X$;
- normalize os autovetores de forma que eles tenham tamanho 1;
- identifique o autovetor B_0 que corresponde ao único autovalor 0;
- estime uma regressão por componentes principais com variável-resposta Y e a matriz de *design* U , sendo que esta última contém os vetores-coluna que são os componentes principais determinados pelos autovalores não nulos;
- utilize uma matriz ortonormal de todos os autovetores para transformar os coeficientes da regressão por componentes principais em coeficientes da regressão do estimador intrínseco B .

Yang et al. (2008) alertam que, embora o estimador intrínseco possa ser derivado de uma regressão por componentes principais, os coeficientes estimados por este modelo não serão interpretáveis em termos da idade, período e coorte. Portanto, torna-se necessária a transformação ortogonal destas estimativas.

Fu, Hall e Rohan (2004) e Yang, Fu e Land (2004) enunciam e demonstram matematicamente algumas propriedades estatísticas do estimador intrínseco.² A primeira vantagem estatística do EI, segundo os autores, é que ele satisfaz a condição para a estimação de funções lineares do vetor de parâmetros b . Este é um dos fatores positivos das abordagens baseadas em funções estimáveis: são invariantes em relação à qual solução para as equações normais for obtida. Ademais, estas funções são desejáveis como estimadores estatísticos, na medida em que são funções lineares do vetor de parâmetros não identificado que pode ser estimado sem

viés – em outras palavras, o EI fornece estimadores não viesados dos efeitos de idade, período e coorte. Cabe mencionar que esta condição, devidamente formalizada por Kupper et al. (*apud* YANG; FU; LAND, 2004), implica que qualquer estimador restrito, ou seja, aquele que é obtido mediante a imposição de restrições de igualdade sobre o vetor de parâmetros, sempre produzirá estimativas viesadas dos efeitos de idade, período e coorte. Em resumo, a primeira propriedade estatística do EI é que ele produz estimativas não viesadas dos coeficientes no arcabouço IPC para análise de taxas populacionais, considerando-se número finito fixo de períodos p . A propriedade assintótica do estimador intrínseco sugere que, à medida que o número de períodos aumenta, $p \rightarrow \infty$, a arbitrariedade dos múltiplos possíveis estimadores por MLGR é removida e estes estimadores convergem para o estimador intrínseco B (YANG, FU; LAND, 2004).

Em segundo lugar, foi demonstrado que o EI é mais eficiente do ponto de vista estatístico – isto é, possui menor variância – do que qualquer estimador MLGR. Em outras palavras, para qualquer número finito p de períodos de tempo, o estimador intrínseco B tem uma variância menor do que qualquer estimador restrito de MLGR. Ou seja, $\text{var}(\hat{b}) - \text{var}(B)$ é uma função definida e positiva para uma não trivial restrição de identificação no contexto do MLGR (YANG, FU; LAND 2004). Outra importante propriedade do estimador intrínseco que foi provada é que ele é assintoticamente consistente, isto é, quando $p \rightarrow \infty$ ele converge para os parâmetros verdadeiros que geram a sequência de taxas do IPC.

Uma limitação, contudo, presente nos estimadores por modelos lineares generalizados restritos ainda persiste na solução EI. Foi dito na seção anterior que todo e qualquer modelo exatamente identificado a um conjunto de dados produzirá as mesmas medidas de qualidade de ajuste do modelo (RODGERS 1982, YANG et al., 2008). Dessa forma, o estimador intrínseco apresentará as

² Posto que o propósito deste artigo é tão somente um exercício de aplicação metodológica, deixamos ao leitor interessado nas formulações matemáticas estas referências.

mesmas medidas de qualidade de ajuste, tais como a log-verossimilhança e a *deviance*. Portanto, estas medidas não devem ser utilizadas para selecionar o modelo correto (YANG et al., 2008).

Dados e métodos

Para este exercício, utilizaram-se os microdados da PNAD-IBGE, no período 1981 a 2008. Embora a PNAD não se constitua num painel verdadeiro, que acompanha os indivíduos ao longo do tempo, sua análise no contexto do modelo IPC pode ser viável, pois uma forma de se acompanhar cada coorte ao longo de repetidas pesquisas é olhar para os membros dessa coorte que foram selecionados de forma aleatória em cada ano (OLIVEIRA, 2002). Dessa forma, por exemplo, os indivíduos que tinham sete anos em 1981 teriam oito anos em 1982, e assim por diante.

A faixa etária de estudo, neste artigo, corresponde dos 10 aos 29 anos, pois se requer que, na análise da transição escolar de nosso interesse, os indivíduos analisados estivessem expostos ao risco de realizar a progressão. Dessa forma, têm-se 20 intervalos unitários de idade \times 28 períodos. Como a série histórica da PNAD possui três descontinuidades (1991, 1994 e 2000), utilizou-se uma interpolação linear do número de promovidos e de indivíduos em risco nos anos adjacentes, de forma a completar a série.

Para a estimação do modelo, a ligação canônica empregada foi a função logística. A opção por esta forma funcional do modelo teve o objetivo de garantir que as probabilidades de progressão por série preditas estivessem inseridas no intervalo (0,1). O método de estimação utilizado foi o da máxima verossimilhança.

Em relação à especificação do modelo, optou-se pela estimação do modelo idade-período-coorte em sua forma completa, para verificar as diferenças entre as metodologias EI e MLGR e avaliar a importância substantiva dos efeitos de idade, período e coorte para a probabilidade de transição em cada uma das séries. Sabe-se, contudo, que o procedimento ideal seria testar a im-

portância de cada uma das variáveis a partir de um modelo nulo e, então, incorporar uma a uma as variáveis de idade, período e coorte e avaliar sua significância, por meio da estatística *deviance* e do R^2 .

Outro procedimento importante na análise IPC não efetuado neste trabalho foi testar a presença de efeitos interativos entre idade, período e coorte, ou mesmo a necessidade de inclusão de termos quadráticos. Contudo, por ser este um artigo direcionado para a comparação de metodologias, não se buscou fundamentalmente um ajuste perfeito aos dados, mas sim verificar o potencial de cada arcabouço para estimação de um modelo IPC em sua forma completa.

Para estimação do modelo com base no estimador intrínseco, foi utilizado o algoritmo disponibilizado no STATA por Schulhofer-Wohl e Yang (2006). Estes autores ressaltam que no algoritmo, para computar o estimador intrínseco, é adotada a restrição de que a soma dos coeficientes seja igual a zero. Para fins computacionais, são incluídas variáveis indicadoras para cada um dos valores das variáveis idade, período e coorte na matriz de variáveis explicativas, mas uma das categorias de cada uma delas é omitida. Após a regressão por componentes principais, contudo, a restrição de que os parâmetros devem ter soma zero permite que sejam obtidas as estimativas para as categorias omitidas e, portanto, têm-se as estimativas para todas as categorias de idade, período e coorte, o que não ocorre com o MLGR.

Para estimar o modelo linear generalizado restrito, empregou-se o algoritmo *glm* também disponível no STATA (StataCorp 2007). A estratégia de identificação consistiu em adotar a imposição de que as duas coortes mais antigas tivessem os mesmos coeficientes em todas as transições escolares analisadas (ou seja, as coortes de 1952 e 1953 seriam iguais). Considera-se que esta alternativa é plausível, uma vez que pode-se admitir que as duas coortes mais antigas não tenham passado por um processo de mudança social tão significativo. Cabe mencionar, ainda, que outras estratégias poderiam ser utilizadas, quais sejam: igualar os dois últimos períodos, igualar as duas coortes mais recentes ou os dois períodos

mais recentes. Contudo, considerou-se que os períodos e coortes mais recentes podem ter diferenças significativas em virtude das políticas educacionais recentemente adotadas no Brasil. Não se cogitou aqui restringir os parâmetros de idade, pois assumiu-se que o comportamento da progressão por idade é singular e de interesse substantivo, uma vez que a variação nas probabilidades de progressão por idade numa determinada série reflete o crônico padrão brasileiro de distorção idade-série.

Resultados

Passa-se, neste momento, para a comparação dos resultados dos modelos idade-período-coorte estimados para a probabilidade de progressão das mulheres brasileiras para a 5ª série do ensino fundamental, segundo o estimador intrínseco e os modelos lineares generalizados restritos (com a suposição de que as duas coortes mais antigas fossem iguais). Antes de realizar a análise, cabe ressaltar que a interpretação aqui pode ser feita sem qualquer prejuízo, tomando-se como base os coeficientes estimados como medida para o comportamento das probabilidades de progressão por série. Isto porque as probabilidades de progressão preditas pelo modelo podem ser obtidas diretamente pelo exponencial dos coeficientes, tanto no modelo linear generalizado restrito quanto naquele baseado no estimador intrínseco. Dessa forma, a interpretação do comportamento dos coeficientes estimados é similar à interpretação do comportamento das probabilidades de progressão por série, já que a segunda é uma transformação uma-a-uma da primeira.

Vale lembrar ainda que, tal como explicitado anteriormente, os dois modelos possuem as mesmas medidas de qualidade do ajuste (deviance, AIC, BIC e log-verossimilhança), sendo, portanto, inviável a utilização destes critérios na seleção do melhor modelo. Os valores dos coeficientes estimados, bem como as estatísticas de qualidade do ajuste dos dois modelos, estão reportados na Tabela 1 do Anexo.

O Gráfico 1 mostra os coeficientes estimados pelos modelos EI e MLGR para o

modelo IPC da probabilidade de progressão para a 5ª série do ensino fundamental das mulheres. Mediante a análise gráfica, busca-se verificar em que medida as estimativas dos parâmetros derivadas de cada método são discrepantes. Antes, contudo, de dar prosseguimento à análise dos gráficos, cabe mencionar que, na estimação do modelo, o EI utiliza a restrição de que a soma dos coeficientes de idade, período e coorte seja igual a zero. Por sua vez, o MLGR utiliza a restrição de omitir uma categoria de referência, no caso a primeira categoria de idade, período e coorte. Diante disso, para manter a comparabilidade entre os coeficientes dos dois modelos, os parâmetros do modelo MLGR foram centralizados em torno da média dos coeficientes de idade, período e coorte. Este procedimento é conhecido como *effect coding* e, a partir dele, o intercepto é igual à média global e o intercepto para cada variável expressa a diferença entre o grupo e a média global (HOSMER; LEMESHOW, 2000).

Comparando-se os coeficientes estimados para a probabilidade de progressão para a 5ª série, é possível verificar que, no caso das mulheres, os efeitos de idade apresentaram pouca diferença entre os métodos, enquanto os efeitos de período e coorte mostraram magnitudes bastante distintas (Gráfico 1). A análise substantiva do comportamento dos coeficientes de idade, período e coorte é capaz de revelar características interessantes sobre as variações nas taxas observadas. Verifica-se, num primeiro momento, que a probabilidade de progressão para a 5ª série é muito baixa aos 11 anos, elevando-se, a partir daí, de forma acelerada até aproximadamente os 19 anos de idade, quando então se estabiliza (Gráfico 1). Este padrão por idade das probabilidades de progressão é coerente com o elevado padrão de defasagem idade-série no Brasil, decorrente do fato de que os indivíduos tendem a não concluir a 5ª série do ensino fundamental na idade adequada (RIOS-NETO et al., 2010). Os efeitos de período obtidos pelo estimador intrínseco e modelo linear generalizado restrito apontam para uma elevação na probabilidade de progressão das mulheres para a 5ª série do

ensino fundamental, comportamento este coerente com a expansão desse nível de ensino ocorrida no Brasil (Gráfico 1). Já os efeitos de coorte possuem comportamento parabólico para as coortes nascidas entre 1952 e 1997, com um pico para a importância deste efeito para as coortes nascidas na década de 1980 (Gráfico 1).

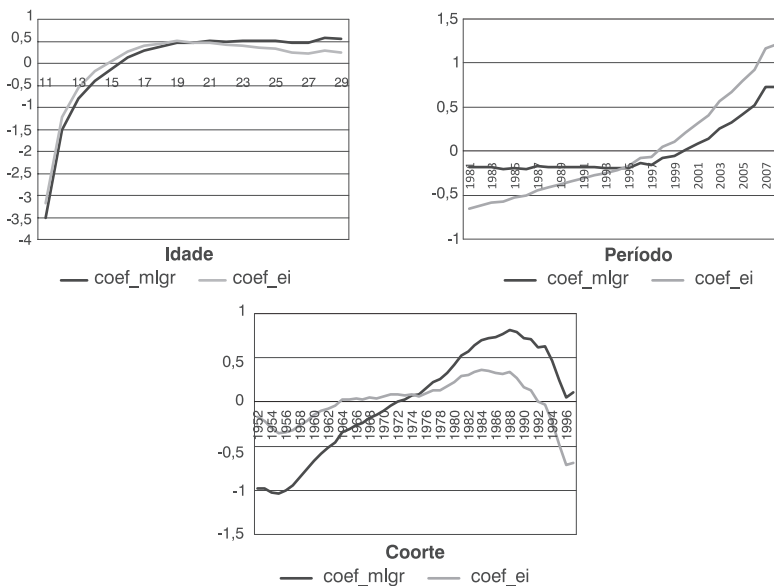
Uma explicação plausível para o comportamento parabólico dos efeitos de coorte sobre a probabilidade de progressão para a 5ª série é o aumento expressivo das taxas de matrícula para as coortes assinaladas. Oliveira (2007) mostra que, antes da década de 1960, as taxas de matrícula no ensino fundamental como um todo eram bastante baixas. Contudo, no período da ditadura militar, o número de matriculados cresceu a uma taxa média de 3,9% ao ano, e depois desacelerou entre 1985 e 1999. Dessa forma, embora o aumento de número de matrículas possa estar associado a efeitos de período (políticas educacionais), pode-se especular que tal crescimento vertiginoso na década de

1980 esteja também relacionado às características sociais peculiares destas coortes.

Agora passa-se à análise da eficiência dos dois métodos. Como visto anteriormente, foi demonstrado matematicamente que o EI possui menor variância do que qualquer estimador MLGR, isto é, qualquer estimador MLGR obtido por quaisquer restrições de identificação. Para investigar esta assertiva no estudo das probabilidades de progressão das mulheres para a 5ª série do ensino fundamental, construímos gráficos que mostram o comportamento dos coeficientes estimados por EI e pelo MLGR com seus respectivos intervalos de confiança a um nível de 95%.

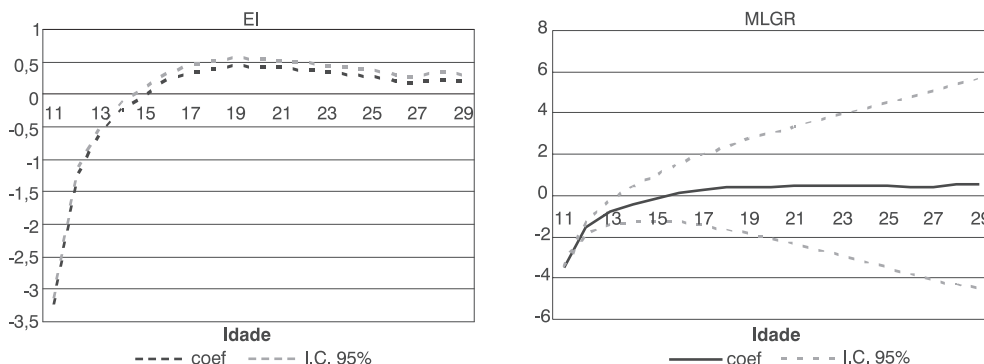
É possível depreender pelos Gráficos 2, 3 e 4 que, de fato, o estimador intrínseco é mais eficiente – isto é, possui menor variância – do que o MLGR, para os efeitos de idade, período e coorte. Analisando os gráficos para os efeitos de idade, verifica-se que o estimador intrínseco possui uma excelente eficiência em relação ao MLGR. A variância do EI, contudo, aumenta quando

GRÁFICO 1
Coefficientes estimados para os efeitos de idade, período e coorte para mulheres,
segundo os estimadores EI e MLGR
Brasil – 1981-2008



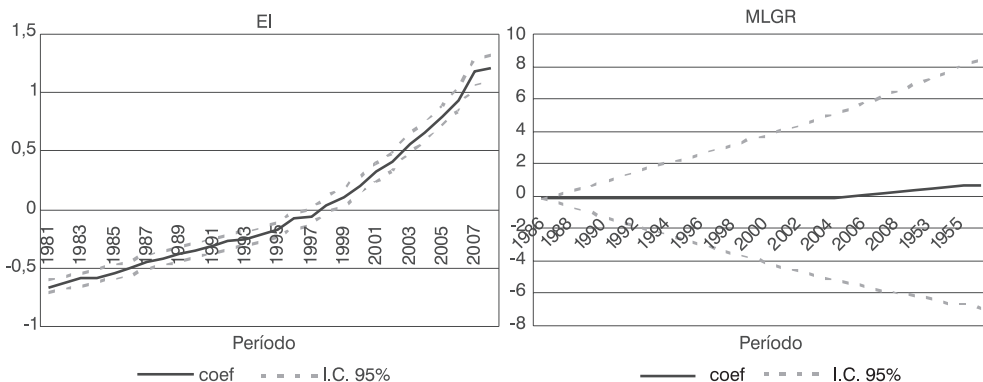
Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 1981 a 2008.

GRÁFICO 2
Coefficientes estimados para os efeitos de idade na probabilidade de progressão das mulheres para a 5ª série do ensino fundamental, segundo os estimadores EI e MLGR
Brasil – 1981-2008



Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 1981 a 2008.

GRÁFICO 3
Coefficientes estimados para os efeitos de período na probabilidade de progressão das mulheres para a 5ª série do ensino fundamental, segundo os estimadores EI e MLGR
Brasil – 1981-2008



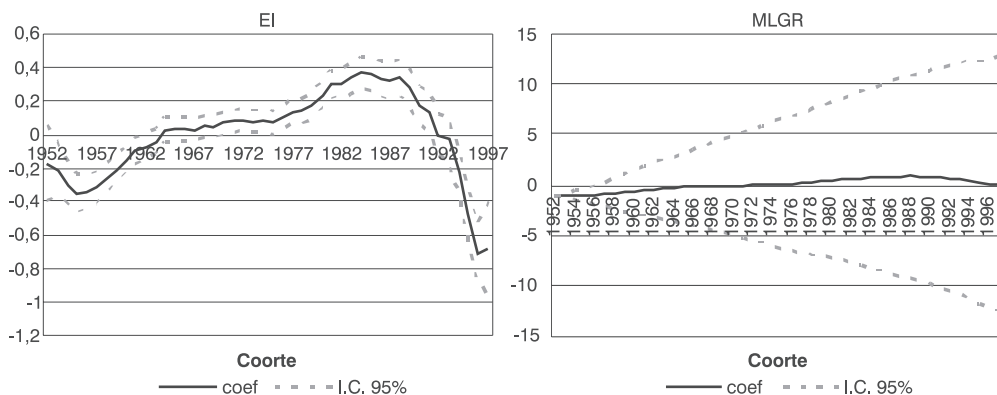
Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 1981 a 2008.

os coeficientes para período ou coorte não são significativos, mas ainda assim este estimador possui de fato uma menor variância que o MLGR.

Cabe ressaltar que a utilização de parâmetros com baixa variância no modelo IPC é muito importante e desejável para a construção de projeções educacionais aleatórias com base na PPS e nos modelos IPC. Isto porque uma forma de se obterem estimativas

futuras das probabilidades de progressão por série, que incluam uma margem de erro (por exemplo, intervalo de confiança), é extrapolar a tendência observada no erro-padrão. Dessa maneira, ter-se-ia uma medida da aleatoriedade nas projeções educacionais. Portanto, se dispomos de estimativas cujo erro-padrão é menor, então a extrapolação desta medida tenderá a ser mais eficiente do que outra cujo erro-padrão é mais elevado.

GRÁFICO 4
Coefficientes estimados para os efeitos de coorte na probabilidade de progressão das mulheres para a 5ª série do ensino fundamental, segundo os estimadores EI e MLGR
 Brasil – 1981-2008



Fonte: IBGE. Pesquisa Nacional por Amostra de Domicílios – PNAD 1981 a 2008.

Considerações finais

Os modelos idade-período-coorte possuem uma forte tradição na Demografia. Neste artigo, com a análise IPC, buscou-se investigar as diferentes fontes de variação na probabilidade de progressão para a 5ª série do ensino fundamental das mulheres brasileiras.

Este artigo possui, contudo, um forte caráter metodológico. Procurou-se comparar duas metodologias, sendo a primeira delas usual na literatura demográfica, baseada nos trabalhos de Fienberg e Mason (1985), a qual denominamos de solução por *modelos lineares generalizados restritos* (MLGR), e uma metodologia recente que emergiu de avanços na epidemiologia, o chamado estimador intrínseco (EI). Argumentou-se que a solução baseada no EI busca solucionar um grande impasse no que concerne ao problema da identificação no arcabouço IPC. A grande inovação deste estimador é que, a partir da decomposição do espaço paramétrico do modelo IPC irrestrito (ou seja, aquele não identificado), este estimador especial B pode ser derivado tanto pelo método de projeção quanto pelo método de regressão por componentes principais.

Ademais, outra singularidade deste estimador é que a única restrição necessária

baseia-se na orientação do estimador no espaço paramétrico, a qual depende fundamentalmente de uma matriz de *design* X fixa, isto é, do número de períodos e grupos de idade.

Foi visto também que o estimador intrínseco apresenta, segundo a literatura, excelentes propriedades estatísticas. Em grande medida, as evidências empíricas deste artigo corroboram com esta assertiva. Além de apresentar parâmetros estimados condizentes com a evolução histórica das políticas educacionais no Brasil e de convergir para os valores verdadeiros dos parâmetros em grandes amostras, este estimador é mais eficiente do que aqueles baseados nos modelos lineares generalizados restritos (MLGR). Por sua vez, os estimadores por MLGR, com a suposição de que os parâmetros das duas coortes mais antigas fossem iguais, se mostraram na grande maioria das vezes não significativos e com variância maior do que os EI.

Diante do exposto, argumenta-se que o estimador intrínseco apresenta-se de fato como uma ferramenta poderosa na análise IPC. Por conseguinte, a construção de projeções probabilísticas das PPSs, a partir deste modelo, se mostra bastante promissora. Este é o próximo passo que pretendemos implementar em artigos

futuros. Para tanto, é necessário que se proceda com a construção passo-a-passo do modelo IPC, com o teste da inclusão de cada uma das variáveis indicadoras de idade, período e coorte, a comparação do

ajuste dos modelos e o teste para inclusão de variáveis interativas ou termos quadráticos. Este procedimento permitirá que se obtenha o melhor modelo IPC para as séries de progressão escolar.

Referências

DECARLI, A.; VECCHIA, C. L. Age, period and cohort models: A review of knowledge and implementation in glim. **Riv. Stat. Applic.**, v. 20, p. 397-410, 1987.

FIENBERG, S. E.; MASON, W. M. Specification and implementation of age, period and cohort models. In: MASON, W. M.; FIENBERG, S. E. (Eds.). **Cohort analysis in social research**. (S.I.): Springer Verlag, 1985.

FU, W. J. Ridge estimator in singular design with application to age-period-cohort analysis of disease rates. **Communications in Statistics: Theory and Method**, v. 29, p. 263-278, 2000.

FU, W. J.; HALL, P.; ROHAN, T. E. Age-period-cohort analysis: structure of estimators, estimability, sensitivity, and asymptotics. **Journal of the American Statistical Association**, 2004.

GLENN, N. D. Cohort analysts' futile quest: statistical attempts to separate age, period and cohort effects. **American Sociological Review**, v. 41, n. 5, p. 900-904, oct. 1976.

GUIMARÃES, R. R. de M. **Probabilidade de progressão por série no Brasil: evolução, seletividade e aplicação de modelos de idade-período-coorte**. Dissertação (Mestrado) – Centro de Desenvolvimento e Planejamento Regional – Cedeplar/UFMG, 2010.

HALLI, S. S.; RAO, K. V. **Advanced techniques of population analysis**. (S.I.): Plenum Press, 1992.

HECKMAN, J.; ROBB, R. Using longitudinal data to estimate age, period and cohort effects in earnings equation. In: MASON, W. M.; FIENBERG, S. E. (Eds.). **Cohort analysis in social research**. (S.I.): Springer Verlag, 1985.

HOSMER, D.; LEMESHOW, S. **Applied logistic regression**. (S.I.): Wiley, 2000.

KNIGHT, K.; FU, W. J. Asymptotics for lasso-type estimations. **Annals of Statistics**, v. 28, p. 1356-78, 2000.

KUPPER, L. L. et al. Statistical age-period-cohort analysis: a review and critique. **Journal of Chronic Diseases**, v. 38, p. 811-830, 1985.

MASON, K. O. et al. Some methodological issues in cohort analysis of archive data. **American Sociological Review**, v. 38, p. 242-258, 1973.

MASON, W. M.; SMITH, H. L. Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis. In: MASON, W. M.; FIENBERG, S. E. (Eds.). **Cohort analysis in social research**. (S.I.): Springer Verlag, 1985.

OLIVEIRA, A. M. H. C. de. **Acumulando informações e estudando mudanças ao longo do tempo: análises longitudinais do mercado de trabalho brasileiro**. Tese (Doutorado) – Centro de Desenvolvimento e Planejamento Regional – Cedeplar/UFMG, 2002.

OLIVEIRA, R. P. Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica. **Educação e Sociedade**, Campinas, v. 28, n. 100 Especial, p. 661-690, out. 2007.

OSMOND, C.; GARDNER, M. J. Age, period and cohort models applied to cancer mortality. **Statistical Medicine**, v. 1, p. 245-259, 1982.

PEARL, J. **Causality: models, reasoning and inference**. (S.I.): Cambridge University Press, 2000.

RIOS-NETO, E. L. G. O método de probabilidade de progressão por série. In: RIOS-NETO, E. L. G.; RIANI, J. de L. R. (Eds.). **Introdução à demografia da educação**. Campinas: Associação Brasileira de Estudos Populacionais, 2004, cap. 1.

RIOS-NETO, E. L. G.; OLIVEIRA, A. M. H. C. de. Aplicação de um modelo de idade-período-coorte para a atividade econômica no Brasil metropolitano. **Pesquisa e Planejamento Econômico**, v. 29, n. 2, p. 243-272, ago. 1999.

RIOS-NETO, E. L. G.; GUIMARÃES, R. R. de M.; PIMENTA, P. S. F.; MORAIS, T. A. **Análise de indicadores educacionais no Brasil, 1981 a 2008**. Centro de Desenvolvimento e Planejamento Regional – Cedeplar/UFMG Junho 2010 (Texto para discussão, n. 386).

ROBERTSON, C.; BOYLE, P. Age, period and cohort models: the use of individual records. **Statistical Medicine**, v. 5, p. 527-538, 1986.

ROBERTSON, C.; GANDINI, S.; BOYLE, P. Age-period-cohort models: A comparative study of available methodologies. **Journal of Clinical Epidemiology**, v. 52, n. 6, p. 569-583, 1999.

RODGERS, W. L. Estimable functions of age, period, and cohort effects. **American Sociological Review**, v. 47, n. 6, p. 774-787, dez. 1982.

Resumen

Comparación entre metodologías de edad-período-cohorte para el estudio de una medición de la progresión escolar en Brasil

El objetivo de este artículo es realizar un ejercicio de comparación metodológica de dos estimadores de los modelos edad-período-cohorte: el estimador convencional, obtenido por los modelos lineales generalizados restringidos (MLGR), y el estimador intrínseco (EI). El objeto de estudio es la contribución de los efectos de edad, período y cohorte en los cambios temporales dentro de la probabilidad de progresión (PPS) en el 5º curso de la Enseñanza Fundamental de las mujeres brasileñas. El modelo IPC se justifica, ya que los efectos de edad, período y cohorte pueden impactar de forma significativa en la probabilidad de progresión escolar: los efectos de edad se reflejan tanto en la edad mínima de entrada en el sistema de enseñanza, como también en el dilema entre trabajo y estudio que surge a lo largo de la carrera educacional; los efectos del período están asociados a las diferentes coyunturas económicas y políticas, así como al estado de las políticas educacionales; finalmente, los efectos de cohorte reflejan características sociales peculiares de determinados grupos de individuos. Los dos instrumentos

SCHULHOFER-WOHL, S.; YANG, Y. **APC: Stata module for estimating age-period-cohort effects**. Statistical Software Components, Boston College Department of Economics, ago. 2006. Disponible em: <<http://ideas.repec.org/c/boc/bocode/s456754.html>>.

SMITH, H. Response: cohort analysis redux. **Sociological Methodology**, v. 34, p. 111-119, 2004.

STACORP **Stata Statistical Software**: release 10. 2007.

WINSHIP, C.; HARDING, D. J. A mechanism-based approach to the identification of age-period-cohort models. **Sociological Methods Research**, v. 36, p. 362-401, 2008.

YANG, Y. Trends in U.S. adult chronic disease mortality, 1960-1999: age, period and cohort variations. **Demography**, v. 45, n. 2, p. 387-416, maio 2008.

YANG, Y.; FU, W. J.; LAND, K. C. A methodological comparison of age-period-cohort models: the intrinsic estimator and conventional generalized linear models. **Sociological Methodology**, v. 35, p. 75-110, 2004.

YANG, Y. et al. The Intrinsic Estimator for age-period-cohort analysis: What it is and how to use it. **American Journal of Sociology**, v. 113, n. 6, p. 1697-1736, maio 2008.

fueron contrapuestos en términos de eficiencia, significancia y estimativa de los parámetros. Los resultados de este artículo revelan la potencialidad de una solución para el modelo de IPC basado en el estimador intrínseco, el cual presenta parámetros con variancia inferior a los estimados por el modelo lineal generalizado restringido. De esta forma, proyecciones de las PPS, basadas en la extrapolación del error-padrón de los parámetros obtenidos por el estimador intrínseco, pueden ser prometedoras.

Palabras-clave: Modelos edad-período-cohorte. Estimador intrínseco. Probabilidad de progresión por curso.

Abstract

A comparison between age-period-cohort methodologies for an indicator of school progression in Brazil.

The aim of this article is to conduct a comparative methodological essay of two estimators of age-period-cohort models: the conventional estimator obtained by the generalized linear restricted models (MLGR) and the so-called intrinsic estimator (EI). The objects are contributions of age, period and cohort effects to temporal changes in the progression probability to the 5th grade of elementary school for Brazilian women. The APC modeling is justified because age, period and cohort effects may significantly affect the probability of grade progression: age effects reflect both the minimal age of school entry and the trade-off between study and work; period effects are associated with different economical and political conjunctures, as well as with current educational policies; finally, cohort effects reflect social attributes unique to a group. Both are juxtaposed in terms of the efficiency, significance and parameter estimates. The results reveal the potentiality of the solution to the age-period-cohort model based on intrinsic estimator, which presents parameters with smaller variance than those estimated from generalized linear restricted models. Therefore, projections of grade progression probabilities based on the extrapolation of standard errors of the intrinsic estimator can be promising.

Keywords: Age-period-cohort models. Intrinsic estimator. Grade progression probability.

Recebido para publicação em: 26/08/2010

Aceito para publicação em: 09/11/2010

Anexo

TABELA 1
Valores dos coeficientes estimados segundo os dois modelos IPC para a probabilidade de progressão das
mulheres para a 5ª série do ensino fundamental
Brasil – 1981-2008

	EI	MLGR
I_7	-	-
I_8	-	-
I_9	-	-
I_10	-	-
I_11	-3,188 [0,029]	-
I_12	-1,217 [0,021]	2,006 [0,150]
I_13	-0,546 [0,023]	2,713 [0,293]
I_14	-0,179 [0,025]	3,115 [0,438]
I_15	0,045 [0,027]	3,375 [0,583]
I_16	0,270 [0,028]	3,635 [0,728]
I_17	0,392 [0,029]	3,792 [0,873]
I_18	0,454 [0,030]	3,890 [1,018]
I_19	0,508 [0,030]	3,980 [1,163]
I_20	0,473 [0,030]	3,980 [1,308]
I_21	0,470 [0,029]	4,013 [1,453]
I_22	0,428 [0,029]	4,006 [1,598]
I_23	0,395 [0,028]	4,008 [1,744]
I_24	0,364 [0,028]	4,013 [1,889]
I_25	0,323 [0,027]	4,007 [2,034]
I_26	0,251 [0,027]	3,971 [2,179]
I_27	0,224 [0,027]	3,979 [2,325]

(continua)

(continuação)

	EI	MLGR
I_28	0,289 [0,027]	4,081 [2,472]
I_29	0,245 [0,027]	4,072 [2,612]
P_1981	-0,656 [0,028]	-
P_1982	-0,620 [0,028]	0,000 [0,145]
P_1983	-0,589 [0,028]	-0,005 [0,290]
P_1984	-0,573 [0,028]	-0,023 [0,435]
P_1985	-0,533 [0,029]	-0,019 [0,580]
P_1986	-0,502 [0,029]	-0,023 [0,725]
P_1987	-0,442 [0,030]	0,001 [0,870]
P_1988	-0,416 [0,030]	-0,009 [1,015]
P_1989	-0,378 [0,030]	-0,006 [1,160]
P_1990	-0,343 [0,031]	-0,007 [1,305]
P_1991	-0,306 [0,031]	-0,005 [1,450]
P_1992	-0,271 [0,031]	-0,006 [1,595]
P_1993	-0,250 [0,032]	-0,020 [1,741]
P_1994	-0,212 [0,032]	-0,018 [1,886]
P_1995	-0,176 [0,033]	-0,017 [2,031]
P_1996	-0,078 [0,034]	0,045 [2,176]
P_1997	-0,070 [0,034]	0,018 [2,322]
P_1998	0,049 [0,036]	0,101 [2,467]
P_1999	0,106 [0,036]	0,123 [2,612]
P_2000	0,206 [0,038]	0,187 [2,757]
P_2001	0,317 [0,039]	0,263 [2,903]

(continua)

(continuação)

	EI	MLGR
P_2002	0,407 [0,040]	0,317 [3,048]
P_2003	0,563 [0,042]	0,438 [3,193]
P_2004	0,665 [0,043]	0,504 [3,338]
P_2005	0,795 [0,045]	0,599 [3,484]
P_2006	0,930 [0,047]	0,698 [3,629]
P_2007	1,169 [0,051]	0,902 [3,774]
P_2008	1,209 [0,056]	0,906 [3,920]
C_1952	-0,175 [0,114]	-
C_1953	-0,210 [0,081]	-
C_1954	-0,298 [0,066]	-0,052 [0,215]
C_1955	-0,350 [0,057]	-0,069 [0,349]
C_1956	-0,340 [0,051]	-0,023 [0,490]
C_1957	-0,317 [0,048]	0,035 [0,633]
C_1958	-0,262 [0,045]	0,126 [0,777]
C_1959	-0,210 [0,043]	0,213 [0,922]
C_1960	-0,151 [0,041]	0,308 [1,067]
C_1961	-0,103 [0,040]	0,392 [1,211]
C_1962	-0,073 [0,039]	0,457 [1,356]
C_1963	-0,046 [0,038]	0,519 [1,501]
C_1964	0,026 [0,038]	0,626 [1,646]
C_1965	0,032 [0,036]	0,668 [1,791]
C_1966	0,035 [0,035]	0,706 [1,937]
C_1967	0,026 [0,034]	0,733 [2,082]

(continua)

(continuação)

	EI	MLGR
C_1968	0,048 [0,033]	0,791 [2,227]
C_1969	0,045 [0,032]	0,823 [2,372]
C_1970	0,063 [0,032]	0,877 [2,517]
C_1971	0,082 [0,033]	0,931 [2,662]
C_1972	0,089 [0,034]	0,974 [2,807]
C_1973	0,078 [0,034]	0,998 [2,953]
C_1974	0,086 [0,035]	1,042 [3,098]
C_1975	0,068 [0,035]	1,059 [3,243]
C_1976	0,102 [0,036]	1,129 [3,388]
C_1977	0,132 [0,037]	1,194 [3,534]
C_1978	0,137 [0,038]	1,235 [3,679]
C_1979	0,173 [0,039]	1,306 [3,824]
C_1980	0,229 [0,040]	1,398 [3,969]
C_1981	0,299 [0,042]	1,503 [4,115]
C_1982	0,307 [0,044]	1,547 [4,260]
C_1983	0,336 [0,046]	1,611 [4,405]
C_1984	0,367 [0,048]	1,678 [4,550]
C_1985	0,353 [0,050]	1,699 [4,696]
C_1986	0,329 [0,052]	1,711 [4,841]
C_1987	0,320 [0,053]	1,737 [4,986]
C_1988	0,335 [0,056]	1,788 [5,131]
C_1989	0,274 [0,057]	1,763 [5,277]
C_1990	0,170 [0,059]	1,693 [5,422]

(continua)

(continuação)

	EI	MLGR
C_1991	0,129 [0,061]	1,688 [5,567]
C_1992	-0,001 [0,063]	1,594 [5,712]
C_1993	-0,030 [0,066]	1,600 [5,858]
C_1994	-0,229 [0,070]	1,437 [6,003]
C_1995	-0,477 [0,077]	1,224 [6,148]
C_1996	-0,711 [0,091]	1,026 [6,294]
C_1997	-0,687 [0,143]	1,086 [6,440]
C_1998	-	-
C_1999	-	-
C_2000	-	-
C_2001	-	-
constante	1,383 [0,009]	-3,274 [2,520]
N. Obs.	532	532
L	-162,66	-162,66
AIC	0,95	0,95
BIC	-2773,04	-2773,04
Deviance	1,24	1,24

Fonte: Microdados da PNAD 1981-2008.

Obs.: Erro padrão entre colchetes.

