

TEXTO PARA DISCUSSÃO Nº 384

**COMPARAÇÃO ENTRE METODOLOGIAS DE IDADE-PERÍODO-COORTE PARA O ESTUDO
DE PROBABILIDADES DE PROGRESSÃO POR SÉRIE
SELECIONADAS NO BRASIL**

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

Junho de 2010

Ficha catalográfica

| | |
|----------|---|
| 304.6028 | Guimarães, Raquel Rangel de Meireles |
| G963c | Comparação entre metodologias de idade-período- |
| 2010 | coorte para o estudo de probabilidades de progressão por série selecionadas no Brasil / Raquel Rangel de Meireles Guimarães ; Eduardo Luiz Gonçalves Rios-Neto. - Belo Horizonte: UFMG/Cedeplar, 2010. |
| | 57p. (Texto para discussão ; 384) |
| | 1. Demografia - Técnica. 2. Projeção demográfica. I. Rios Neto, Eduardo Luiz Gonçalves. II. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. III. Título. IV. Série. |
| | CDD |

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL**

**COMPARAÇÃO ENTRE METODOLOGIAS DE IDADE-PERÍODO-COORTE PARA O ESTUDO
DE PROBABILIDADES DE PROGRESSÃO POR SÉRIE
SELECIONADAS NO BRASIL**

Raquel Rangel de Meireles Guimarães

Doutoranda em Demografia/CEDEPLAR/UFMG - raquel@cedeplar.ufmg.br

Eduardo Luiz Gonçalves Rios-Neto

Professor Titular do Departamento de Demografia/CEDEPLAR/UFMG - eduardo@cedeplar.ufmg.br

**CEDEPLAR/FACE/UFMG
BELO HORIZONTE
2010**

SUMÁRIO

| | |
|---|----|
| 1. INTRODUÇÃO..... | 6 |
| 2. A QUESTÃO DA IDENTIFICAÇÃO DOS MODELOS DE IDADE-PERÍODO-CORTE: UMA REVISÃO | 7 |
| 3. O ESTIMADOR INTRÍNSECO (EI) | 10 |
| 4. DADOS E MÉTODOS..... | 13 |
| 5. RESULTADOS | 20 |
| 6. CONSIDERAÇÕES FINAIS | 26 |
| REFERÊNCIAS | 56 |

RESUMO

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 e da eficiência do modelo. O objeto de interesse são as contribuições dos efeitos de idade, período e coorte nas mudanças temporais nas probabilidades de progressão por série (PPS) no Brasil. A modelagem IPC das probabilidades de progressão por série se justifica pois os efeitos de idade, período e coorte podem impactar de forma significativa as transições escolares: 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. Os resultados deste artigo revelam a potencialidade da solução para o modelo idade-período-coorte baseada no estimador intrínseco, o qual apresenta excelentes propriedades estatísticas, quais sejam: melhor ajuste aos valores preditos, baixa variância, maior significância dos parâmetros. Desta forma, projeções das PPS baseadas neste estimador podem ser bastante promissoras.

ABSTRACT

The aim of this article is to realize a comparative methodological exercise of two APC estimators: the conventional restrict estimator obtained by the generalized linear models (MLGR) and the so-called intrinsec estimator (EI). Both instruments are contrasted by means of the parameter estimates and of the model's efficiency. The object of our interest are the contributions of age, period and cohort effects to temporal changes in grade progression probabilities in Brazil. The APC modelling of grade progression probabilities is justified as age, period and cohort effects may significantly affect school transitions: age effects reflect both the minimal age of school entry as the trade-off between study and work, which becomes strong along the educational carrier; period effects are associated with different economical and political conjunctures, as well as with the state of the educational policies; finally, cohort effects reflect social attributes proper of some group of individuals. This article shows the potentiality of the solution to the age-period-cohort model based on intrinsec estimator, which presents excelent statistical properties, namely: best adjust to predicted values, small variance, and significance of the parameters. Therefore, projections of grade progression probabilities based on this estimator may be very promissing.

Palavras-chave: Modelos Idade-Período-Coorte; Estimador Intrínseco; Probabilidade de Progressão por Série (PPS)

Keywords: Age-Period-Cohort Models; Intrinsic Estimator; Grade Progression Probability

JEL: Y80

1. 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, por variações na idade, ou por variações interativas entre período e idade (coorte). Pressupõe-se que cada um destes efeitos tenha uma interpretação e atuação substantiva sobre o objeto de estudo. Efeitos de idade estão normalmente 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 nos dão uma medida das condições ambientais, econômicas e sociais vigentes. Por sua vez, 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 e Rao 1992, Yang, Fu e Land 2004).

A principal controvérsia na aplicação dos modelos IPC é como lidar com o chamado *problema da identificação*. Uma vez que 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 a identificação do modelo IPC consiste na imposição de uma igualdade entre os parâmetros do modelo (Fienberg e Mason 1985). Todavia, esta solução recebeu muitas críticas, as quais de uma maneira geral 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 invariantes às restrições nos parâmetros dentro do arcabouço IPC. Dentre elas, surgiu o estimador intrínseco (EI), descrito por Yang, Fu e Land (2004). Este estimador é baseado em funções estimáveis da decomposição do valor singular de matrizes, e provê soluções únicas dos estimadores de idade, período e coorte. Mais ainda, o estimador intrínseco 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 e 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 e da eficiência do modelo. O objeto de interesse são as contribuições dos efeitos de idade, período e coorte nas mudanças temporais nas probabilidades de progressão por série (PPS) no Brasil. A modelagem IPC das probabilidades de progressão por série se justifica pois os efeitos de idade, período e coorte podem impactar de forma significativa as transições escolares: *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.

Além de ser de grande importância substantiva, a análise IPC das probabilidades de progressão por série 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. Neste sentido, a comparação entre metodologias adquire uma importância fundamental para que se obtenham projeções educacionais de qualidade e acuradas.

Este artigo está organizado em quatro 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 terceira seção trata exclusivamente do estimador intrínseco descrito por Yang, Fu e Land (2004). A quarta seção descreve os dados e os passos metodológicos. A quinta seção reporta os procedimentos e os resultados da comparação entre o estimador restrito convencional e o estimador intrínseco para a análise de quatro progressões por série selecionadas, quais sejam: conclusão da 1^a série do Ensino Fundamental (e_0); conclusão da 5^a série do Ensino Fundamental, dado que concluiu a 4^a série (e_4); conclusão da 1^a série do Ensino Médio, dado que concluiu a 8^a série do Ensino Fundamental (e_8) e a probabilidade de conclusão do primeiro ano do Ensino Superior, dado que concluiu a 3^a série do Ensino Médio (e_{11}). A sexta seção sumariza as evidências deste estudo e propõe uma agenda futura de pesquisas.

2. 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, $i=1,...,a$ é o indexador do grupo etário, $j=1,...,p$ indexa o período e $k=a+p-1$ denota a coorte. Ainda, ε_{ij} é o termo de erro aleatório 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. Em termos formais, um modelo linear generalizado apresenta três componentes: um *componente aleatório* implícito na variável dependente, cuja distribuição é estipulada *a priori*; um *componente sistemático*, expresso por um preditor linear construído a partir das variáveis explicativas; e uma *função de ligação* que relaciona os componentes aleatório e sistemático.

Neste artigo, tratamos a probabilidade de progressão por série como uma variável dicotômica que apresenta uma distribuição binomial. Ou seja, para cada combinação de idade-período temos neste estudo 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 à não progredir. A ligação canônica é, então, realizada pela função logística, resultando em um modelo logito da seguinte forma:

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

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 através da 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:

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

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

$$b=(X^T X)^{-1}(X^T Y) \quad (4)$$

A questão fundamental é que não existe *apenas um* vetor de coeficientes estimados para o modelo 3. Isto ocorre pois $X^T X$ não é invertível (ou seja, é uma matriz singular), devido a uma perfeita relação linear entre os efeitos de idade, período e coorte. Em outras palavras, algumas colunas da matriz X podem ser linearmente combinadas ou somadas de forma a produzir uma coluna idêntica às outras colunas na mesma matriz. Denomina-se na literatura este impasse como *problema de identificação* na análise IPC. Implícito no problema da identificação está, portanto, o fato de que, para a equação 4, há um infinito número de soluções possíveis. 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, i.e., 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 . Desta forma, bastaria que se admitisse que os coeficientes do primeiro e segundo período, ou da primeira e segunda coorte, ou do primeiro e segundo grupo etário, 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 em se 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 utilizar-se 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 podem 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 é 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. Robertson, Gandini e Boyle (1999) sistematizaram estas contribuições e avaliaram cada uma delas. A primeira classe de contribuições se baseia no uso da *minimização de uma função de perda* como forma de se atingir a identificação, a qual, em geral, leva em consideração as medidas de qualidade do ajuste e os parâmetros do modelo de três fatores (idade-período, período-coorte, idade-coorte) e do modelo completo (idade-período-coorte) osmond,decarlirobertson. A segunda classe utiliza *registros individuais* como forma de se obter coortes de nascimento que não são justapostas e, desta forma, resolve o problema da exata dependência linear entre os parâmetros. A terceira classe de estudos, por sua vez, utiliza *funções estimáveis* que não necessitam de restrições para que ocorra a identificação. Neste caso, um modelo IPC arbitrário é estimado e as suas estimativas são transformadas em curvaturas e desvios: curvaturas representam a segunda diferença das estimativas dos parâmetros, enquanto os desvios são os resíduos obtidos por meio de regressões lineares para os efeitos de idade, período e coorte estimados com base em restrições arbitrárias. Utilizando simulações, os autores concluíram que todos os métodos apresentam algum tipo de viés e que, em alguns casos, este viés acarretar em erro de interpretação do fenômeno.

Na terceira linha das estratégias descritas por Robertson, Gandini e Boyle (1999), a que utiliza funções estimáveis, insere-se a abordagem baseada num novo estimador: o chamado estimador intrínseco. Ele foi introduzido por Fu; Knight e Fu; Fu, Hall e Rohan (apud YANG; FU; LAND, 2004) e se assenta na decomposição do valor singular de matrizes, e provê estimadores robustos dos efeitos de idade, período e coorte. Uma vez que a aplicação deste método é de interesse central neste artigo, a formalização deste método 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 se basearam 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 afetam a variável resposta são de interesse-chave. O teorema proposto pelos autores é que, para a identificação dos coeficientes no modelo IPC, é apenas necessário que sejam especificados todos 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 sobre-identificado e, neste caso, deve-se testar o ajuste do modelo global e de seus componentes.

3. O ESTIMADOR INTRÍNSECO (EI)

Nesta seção formalizamos a construção do estimador intrínseco Yang, Fu e Land (2004), Yang (2008), Yang et al. (2008), bem como as suas propriedades estatísticas. Considere a equação linear geral do modelo IPC (Equação 3). A partir da mesma, a dependência linear entre os efeitos de idade, período e coorte pode ser representada na forma matricial da seguinte forma, a partir de um vetor B_0 não-nulo:

$$XB_0=0 \quad (5)$$

A equação 5 é resultado do fato de que a matriz X é singular, i.e., existe alguma combinação linear das colunas da matriz de design X que resulta em um vetor nulo. Na Álgebra Linear, diz-se que o posto (ou *rank*) de uma matriz X é o número máximo de colunas linearmente independentes de X . Se uma matriz possui o máximo de colunas linearmente independentes, então dizemos que ela possui *posto completo*. Este, todavia, não é o caso da matriz X , dada a relação linear prevista entre idade, período e coorte.

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:

$$P=N\oplus\tau \quad (6)$$

Onde \oplus representa a soma direta dos dois subespaços lineares N e τ , 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 τ é o 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:

$$\hat{b}=B+sB_0 \quad (7)$$

Onde s é um escalar que corresponde a uma solução específica para o problema da identificação e B_0 é 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:

$$B_0 = \frac{\bar{B}_0}{\|\bar{B}_0\|} \quad (8)$$

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

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

Onde:

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

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

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

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 e é 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:

1. Obtenha os autovalores e os autovetores (componentes principais) da matriz $X^T X$;
2. Normalize os autovetores de forma que eles tenham tamanho 1;
3. Identifique o autovetor B_0 que corresponde ao único autovalor 0;
4. 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;

5. 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, apesar do fato de que 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), Yang, Fu e Land (2004) enunciam e demonstram matematicamente algumas propriedades estatísticas do estimador intrínseco. Posto que o propósito deste artigo é tão somente um exercício de aplicação metodológica, deixamos ao leitor interessado nas formulações matemáticas estas referências. De uma maneira geral, podemos verificar pelos estudos até então realizados que o estimador intrínseco apresenta excelentes propriedades estatísticas, as quais serão descritas nos próximos parágrafos.

A primeira vantagem estatística do EI 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 Kupperman (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 a 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 e Land 2004).

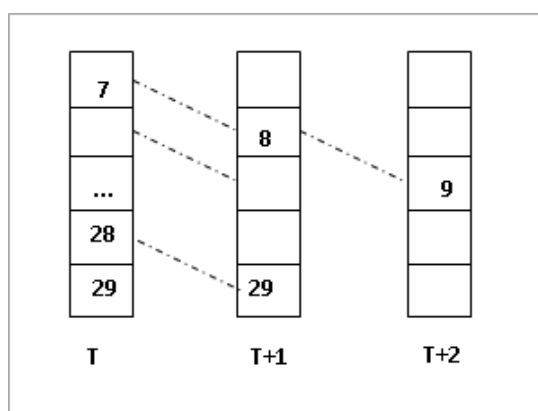
Em segundo lugar, foi demonstrado que o EI é mais eficiente do ponto de vista estatístico - i.e., possui menor variância - do que qualquer estimador MLGR. Em outras palavras, Para qualquer número finito p de períodos de tempo, o estimador intrínseco B tem uma variância menor do que qualquer estimador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

Uma limitação, contudo, 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). O estimador intrínseco, embora resulte em coeficientes distintos do arcabouço MLGR - i.e., livres de qualquer viés - ele apresentará as 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 (Yang et al. 2008).

4. DADOS E MÉTODOS

Para este exercício utilizamos os microdados da PNAD-IBGE, no período compreendido entre 1981 e 2008. Embora a PNAD não se constitua por um painel verdadeiro, i.e., que acompanha os indivíduos ao longo do tempo, sua análise no contexto do modelo IPC pode ser viável. Isto 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). Desta forma, por exemplo, os indivíduos que tinham 7 anos em 1981 teriam 8 anos em 1982, e assim por diante. A Figura 1 ilustra a estratégia de análise, com base em Fienberg e Mason (1985) e Oliveira (2002).

FIGURA 1
Esquema analítico da análise IPC mediante pesquisas transversais repetidas,
conforme ilustrado em Fienberg e Mason, (1985)



A faixa etária de estudo neste artigo foi dos 7 aos 29 anos, pois se requer que, a cada transição, os indivíduos analisados já pudessem ter concluído determinada série. Nossa análise também foi estratificada por sexo. Isto pois há evidências de que homens e mulheres apresentam diferenciais na progressão escolar, bem como a estratificação também é útil na construção de projeções educacionais. Desta forma, as únicas informações extraídas dos microdados foram: idade, sexo, anos de estudo e o período de referência. Construímos também as amostras referentes a cada transição escolar: conclusão da 1^a série do Ensino Fundamental (e_0); conclusão da 5^a série do Ensino Fundamental, dado que concluiu a 4^a série (e_4); conclusão da 1^a série do Ensino Médio, dado que concluiu a 8^a série do Ensino Fundamental (e_8) e a probabilidade de conclusão do primeiro ano do Ensino Superior, dado que concluiu a 3^a série do Ensino Médio (e_{11}).

O primeiro passo para a construção do banco de dados foi verificar quantos indivíduos, segundo o sexo, idade, coorte e período, progrediram na transição de interesse e quantos estavam em risco de realizar aquela progressão. Para a primeira transição, e_0 , todos os indivíduos com pelo menos 7 anos encontravam-se expostos em risco de serem promovidos, mas somente aqueles que tinham pelo menos 1 ano de estudo foram de fato promovidos. No caso da segunda transição, e_4 , apenas os indivíduos com pelo menos 10 anos que

completaram a 4^a série com sucesso estavam em risco de realizar a transição para a 5^a série do Ensino Fundamental, portanto considerou-se os aqueles que possuíam pelo menos 5 anos de estudo como os promovidos. Esta lógica se repetiu para a construção dos bancos de dados referentes à conclusão da 1^a série do Ensino Médio e à conclusão do primeiro ano do Ensino Superior. Como a série histórica da PNAD possui três descontinuidades (1991, 1994 e 2000), utilizamos uma interpolação linear do número de promovidos e indivíduos em risco nos anos adjacentes de forma a completar a série.

As Tabelas 1, 2, 3 e 4 ilustram o diagrama IPC para cada transição escolar investigada neste artigo. Do ponto de vista operacional, da primeira à última transição, o que muda em cada uma das análises IPC é sempre o intervalo etário inferior: no mínimo 7 anos para e_0 , 10 anos para e_4 , 15 anos para e_8 e 18 anos para e_{11} , para garantir que a cada análise se considere apenas aqueles indivíduos expostos ao risco de progredir. Desta forma, na análise de e_0 , temos 23 intervalos unitários de idade \times 28 períodos; para e_4 , temos 20 intervalos unitários de idade \times 28 períodos; para e_8 , temos 15 intervalos unitários de idade \times 28 períodos; e finalmente, para e_{11} , temos 12 intervalos unitários de idade \times 28 períodos.

Neste trabalho, consideramos a probabilidade de progressão por série como sendo uma variável dicotômica que apresenta uma distribuição binomial. Ou seja, para cada combinação de idade-período-coorte-sexo temos neste estudo a frequência absoluta dos que progrediram e dos que não progrediram em determinada transição escolar. A partir destas quantidades foram obtidas as chances de progredir Em relação à não progredir em cada uma das células da tabela idade-período. Para a estimação do modelo, a ligação canônica foi determinada função logística, resultando em um modelo logito da seguinte forma:

$$e_{ij} = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \delta_{ij} \quad (13)$$

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 ij . 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 que sempre ocorre quando a opção é a função de ligação logística. O método de estimação utilizado foi o da máxima verossimilhança.

Em relação à especificação do modelo, optamos 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. Temos em mente, contudo, que o procedimento ideal seria testar a importância 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, através da estatística *deviance* e do R^2 .

Outro procedimento importante na análise IPC que não efetuamos 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 a estimação de um modelo IPC em sua forma completa.

Para a estimação do modelo com base no estimador intrínseco utilizamos o algoritmo disponibilizado no STATA por Schulhofer-Wohl e Yang (2006). De forma sintética, o algoritmo dos autores estima o vetor de parâmetros restrito que corresponderia à projeção dos parâmetros do modelo no espaço não-nulo da matriz X . Isto é efetivado mediante a aplicação de uma transformação ortonormal da matriz $X^T X$, a qual produz os autovalores não-nulos e seus correspondentes autovetores. Em seguida, é estimada uma regressão por componentes principais que utiliza estes autovetores como variáveis explicativas. Por fim, a matriz de transformação ortonormal é novamente empregada, mas desta vez para retornar os coeficientes aos efeitos diretamente interpretáveis de idade, período e coorte.

Schulhofer-Wohl e Yang (2006) 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, tem-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, utilizamos o algoritmo *glm* também disponível no STATA (StataCorp 2007). Nossa 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 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, consideramos que os períodos e coortes mais recentes podem ter diferenças significativas em virtude das políticas educacionais recentemente adotadas no Brasil. Não se cogitou aqui em nenhum momento restringir os parâmetros de idade, pois assumimos que o comportamento da progressão por idade é singular e de interesse substantivo, pois a variação nas probabilidades de progressão por idade numa determinada série reflete o crônico padrão brasileiro de distorção idade-série.

TABELA 1

Diagrama analítico do modelo idade-período-coorte aplicado para a probabilidade de progressão para a 1a série

| | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 7 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 | C47 | C48 | C49 | C50 |
| 8 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 | C47 | C48 | C49 |
| 9 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 | C47 | C48 |
| 10 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 | C47 |
| 11 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 |
| 12 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 |
| 13 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 |
| 14 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 |
| 15 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 |
| 16 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 |
| 17 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 |
| 18 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 |
| 19 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 |
| 20 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 |
| 21 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 |
| 22 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 |
| 23 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 |
| 24 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 |
| 25 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 |
| 26 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 |
| 27 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 |
| 28 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 |
| 29 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 |

Fonte: Elaboração própria

TABELA 2

Diagrama analítico do modelo idade-período-coorte aplicado para a probabilidade de progressão para a 5a série

| | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 11 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 | C46 |
| 12 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 | C45 |
| 13 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 | C44 |
| 14 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 | C43 |
| 15 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 |
| 16 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 |
| 17 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 |
| 18 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 |
| 19 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 |
| 20 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 |
| 21 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 |
| 22 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 |
| 23 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 |
| 24 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 |
| 25 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 |
| 26 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 |
| 27 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 |
| 28 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 |
| 29 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 |

Fonte: Elaboração própria

TABELA 3

Diagrama analítico do modelo idade-período-coorte aplicado para a probabilidade de progressão para a 1ª série do Ensino Médio

| | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 15 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 | C42 |
| 16 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 | C41 |
| 17 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 | C40 |
| 18 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 |
| 19 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 |
| 20 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 |
| 21 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 |
| 22 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 |
| 23 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 |
| 24 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 |
| 25 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 |
| 26 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 |
| 27 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 |
| 28 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 |
| 29 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 |

Fonte: Elaboração própria

TABELA 4

Diagrama analítico do modelo idade-período-coorte aplicado para a probabilidade de progressão para o Ensino Superior

| | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 18 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 | C39 |
| 19 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 | C38 |
| 20 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 | C37 |
| 21 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 | C36 |
| 22 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 | C35 |
| 23 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 | C34 |
| 24 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 | C33 |
| 25 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 | C32 |
| 26 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 | C31 |
| 27 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 | C30 |
| 28 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 | C29 |
| 29 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 | C17 | C18 | C19 | C20 | C21 | C22 | C23 | C24 | C25 | C26 | C27 | C28 |

Fonte: Elaboração própria

5. RESULTADOS

As Figuras 2 e 3 mostram o comportamento das probabilidades de progressão selecionadas em quatro pontos no tempo (1981, 1992, 2001 e 2008): conclusão da 1^a série do Ensino Fundamental (e_0); conclusão da 5^a série do Ensino Fundamental, dado que concluiu a 4^a série (e_4); conclusão da 1^a série do Ensino Médio, dado que concluiu a 8^a série do Ensino Fundamental (e_8) e a probabilidade de conclusão do primeiro ano do Ensino Superior, dado que concluiu a 3^a série do Ensino Médio (e_{11}). Podemos observar que o padrão por idade de e_0 e e_4 é bastante semelhante entre homens e mulheres. Ainda em relação ao padrão das curvas de progressão, pode-se observar que a probabilidade de progressão na idade correta (7 anos para e_0 , 10 anos para e_4 , 15 anos para e_8 e 18 anos para e_{11}) é bastante baixa. Isto indica que poucos conseguem ser promovidos na idade correta. A elevação das probabilidades de progressão nas idades posteriores é condizente com o padrão brasileiro de distorção idade-série. A análise de nível das curvas revela, de uma maneira geral, uma melhoria ao longo do tempo para e_0 e e_4 e no caso das mulheres também em e_8 .

O padrão por idade das probabilidades de progressão dos homens nas duas últimas transições e_8 e e_{11} e das mulheres na probabilidade de progressão para o Ensino Médio (e_8) merece destaque. Podemos observar que estas curvas apresentam bastante oscilações, e que, de uma maneira geral, apontam no sentido de uma piora ao longo do tempo nas probabilidades de progressão. Este fato, contudo, não é surpreendente. Uma vez que as probabilidades de progressão por série são *condicionais* à conclusão da série anterior, o aumento do número de expostos ao risco nas transições tardias é reflexo das políticas educacionais universalistas adotadas a partir da década de 90, e que atingiram sobretudo os primeiros estágios da carreira escolar. Desta forma, com o passar dos anos, um maior número de indivíduos a cada idade passou a estar em risco de transitar para o Ensino Médio ou Superior. Contudo, uma vez que estes níveis de ensino não se expandiram de tal forma que o Ensino Fundamental, uma redução na probabilidade de progressão por série pode ser esperada. Interessante é o padrão discrepante das probabilidades de progressão para o Ensino Médio dos homens a partir de 1992. De certa forma, estas oscilações poderiam ser explicadas por flutuações no mercado de trabalho, uma vez que os homens enfrentam de forma mais sistêmica este *trade-off* entre trabalho e estudo.

A Figura 4 mostra o comportamento das probabilidades de progressão por série conforme a coorte de nascimento e o sexo dos indivíduos ao final de sua carreira escolar, ou seja, aos 29 anos. Nosso objetivo é verificar em que medida as coortes se diferenciaram em termos das probabilidades de progressão e, para tanto, seria necessário garantir que a coorte já tivesse encerrado sua carreira escolar. Por conseguinte, conseguimos acompanhar apenas as coortes nascidas entre 1952 e 1978. Apesar de se tratar de informação passada, este diagnóstico pode revelar a importância da coorte de nascimento para a probabilidade de progressão, e, portanto, nos fornecer uma evidência preliminar da importância dos efeitos de coorte para a trajetória educacional. Podemos observar que, para e_0 , a curva apresenta-se praticamente horizontal, com uma ligeira tendência de aumento entre coortes para homens e mulheres. Para e_4 , a curva tende a ser mais ascendente ao longo das coortes para ambos os sexos. Para e_8 , contudo, a tendência entre coortes já não se

mostra muito clara para os homens, e mostra muitas flutuações. No caso das mulheres, entretanto, as flutuações são mais suaves e evidencia-se tendência de crescimento de e_8 . As evidências entre coortes para e_{11} revelam também ligeira flutuação entre coortes, porém tende a ser mais estável e em um patamar baixo (próximo a 0,4).

Passamos neste momento para a comparação dos resultados dos modelos idade-período-coorte estimados para as transições e_0, e_4, e_8, e_{11} segundo o estimador intrínseco e os modelos lineares generalizados restritos (com a suposição de que as duas coortes mais antigas fossem iguais) segundo o sexo. Vale lembrar que, tal como foi explicitado na seção 3, cada par de modelos comparados (EI versus MLGR) segundo transição escolar e sexo possuem as mesmas medidas de qualidade do ajuste (deviance, AIC, BIC e log-verossimilhança), sendo, portanto, a utilização destes critérios na seleção do melhor modelo inviável.

Resumindo as evidências das regressões em relação aos efeitos mais importantes sobre as probabilidades de progressão selecionadas, verificamos que, de uma maneira geral, os efeitos de idade, período e coorte tem a sua significância alterada conforme o método de estimação (EI ou MLGR). No que tange à probabilidades de progressão para a 1^a série do Ensino Fundamental (e_0), há casos em que os efeitos estimados para o período e coorte pelo estimador intrínseco revelam-se significativos a um nível de 5%, porém são não significativos quando estimados pelo MLGR. É importante salientar que esta discordância possui implicações substantivas: se apenas considerássemos o MLGR, diríamos que as variações de período não foram importantes para a variação em e_0 . Todavia, quando dispomos das estimativas do EI, verificamos que houve um crescimento vertiginoso deste efeito, o que seria condizente com a evolução histórica das políticas educacionais expansionistas no Brasil.

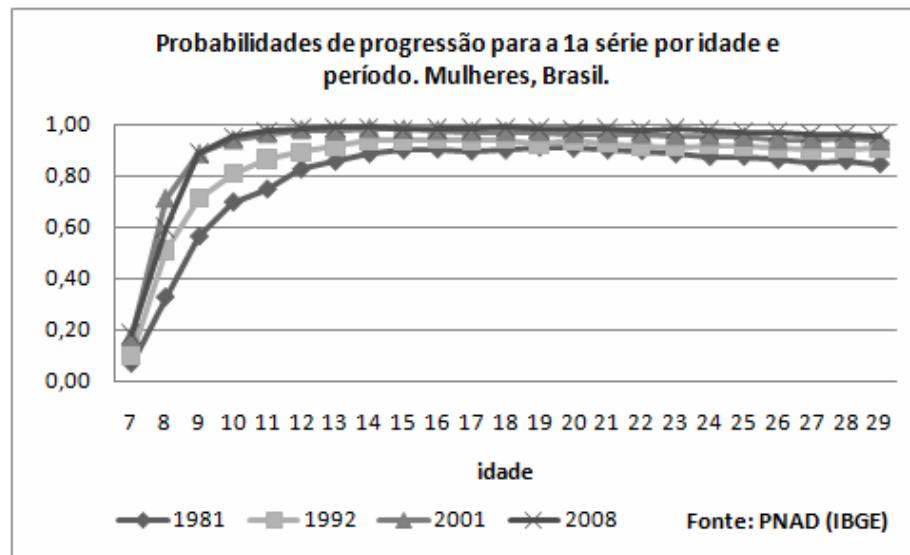
Os coeficientes dos modelos IPC estimados para a probabilidade de progressão para a 5^a série do ensino fundamental também mostram a divergência na significância dos parâmetros entre os dois métodos, tanto para homens quanto para as mulheres. No MLGR, todos os efeitos de período e coorte não são significativos a um nível de 5%, enquanto que o EI revela que uma boa parte dos parâmetros de período e coorte são significativos. Os efeitos de período obtidos pelo estimador intrínseco apontam para uma elevação em e_4 . Já os efeitos de coorte possuem comportamento parabólico para as coortes nascidas entre 1962 e 1992, com um pico para aquelas coortes nascidas na década de 80.

No que diz respeito à probabilidade de progressão para o Ensino Médio (e_8), praticamente todos os efeitos mostram-se não significativos a um nível de 5% quando o estimador adotado é o do MLGR, para ambos os sexos. O estimador intrínseco, por sua vez, revela que alguns efeitos de período e coorte são significativos: os efeitos de período apontam de maneira geral para uma elevação em e_8 , e os efeitos de coorte apresentam grande flutuação, em especial no caso dos homens.

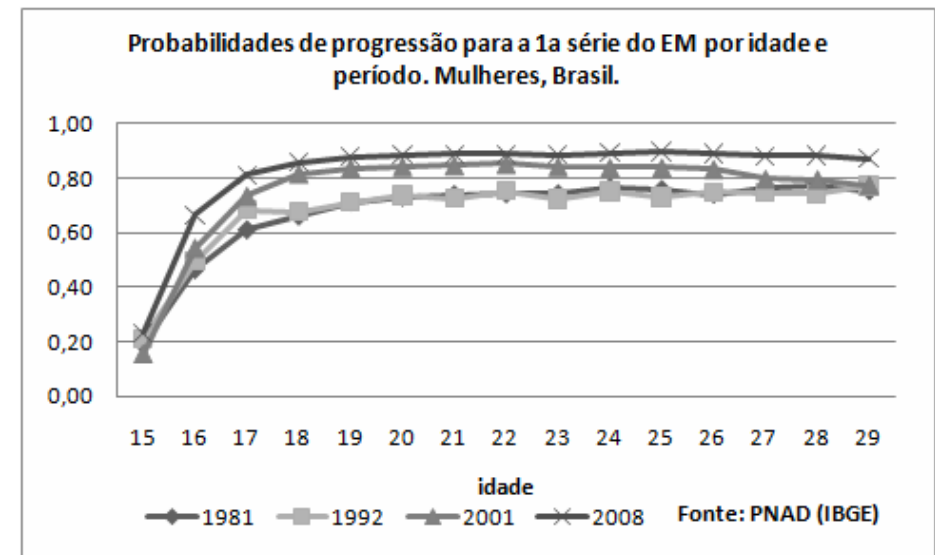
O padrão de discordância na significância dos coeficientes se repete quando a transição analisada é a probabilidade de conclusão do primeiro ano do Ensino Superior (e_{11}). No arcabouço dos MLGR, somente os efeitos referentes às primeiras idades são significativos a um nível de 5%, sendo todos os demais insignificantes tanto para os homens quanto para as mulheres. Todavia, as estimativas com base no estimador intrínseco revelam que todos os efeitos de idade são positivos e crescentes; os efeitos de período, quando significativos, apontam para a redução na probabilidade de progressão ao longo do tempo; e, finalmente, os efeitos de coorte tendem a apresentar bastante flutuações, quando são significativos.

FIGURA 2

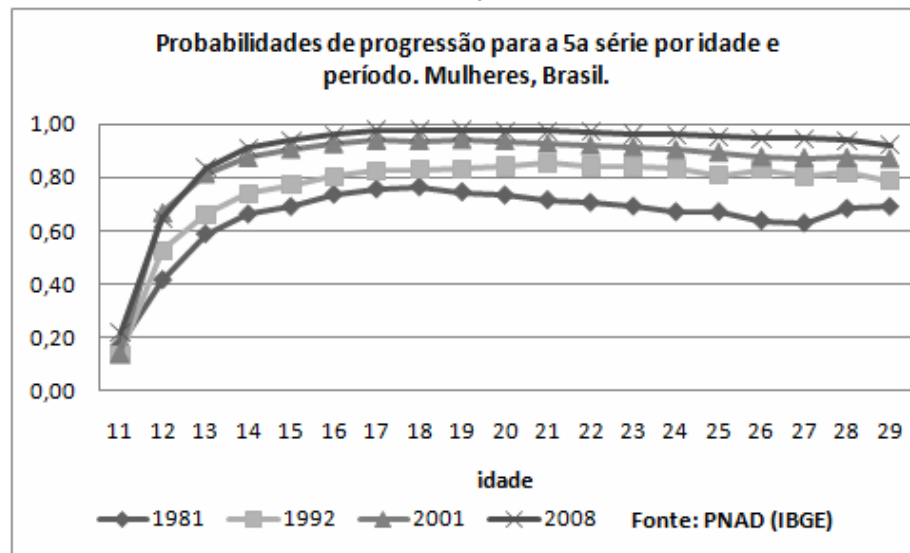
Probabilidade de progressão em séries selecionadas conforme idade e período. Brasil, Mulheres



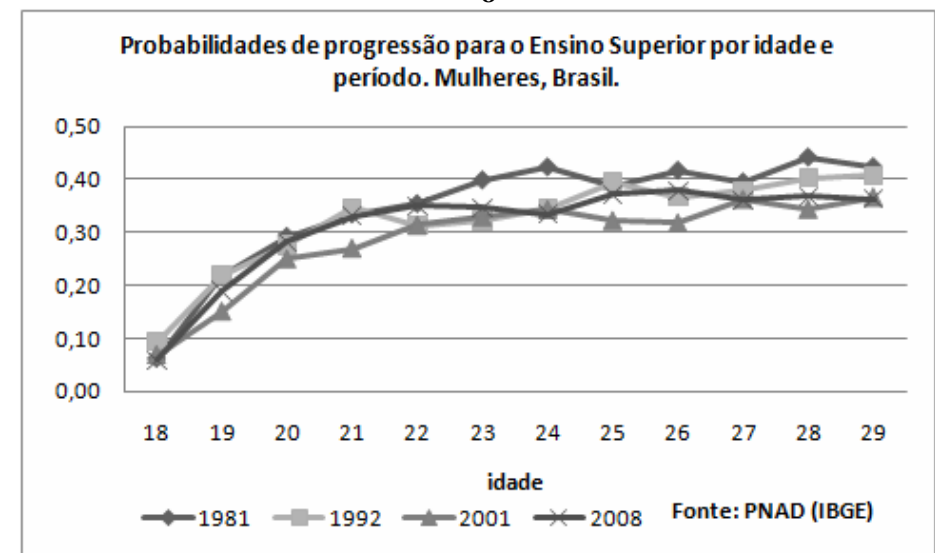
(a) e_0



(c) e_8



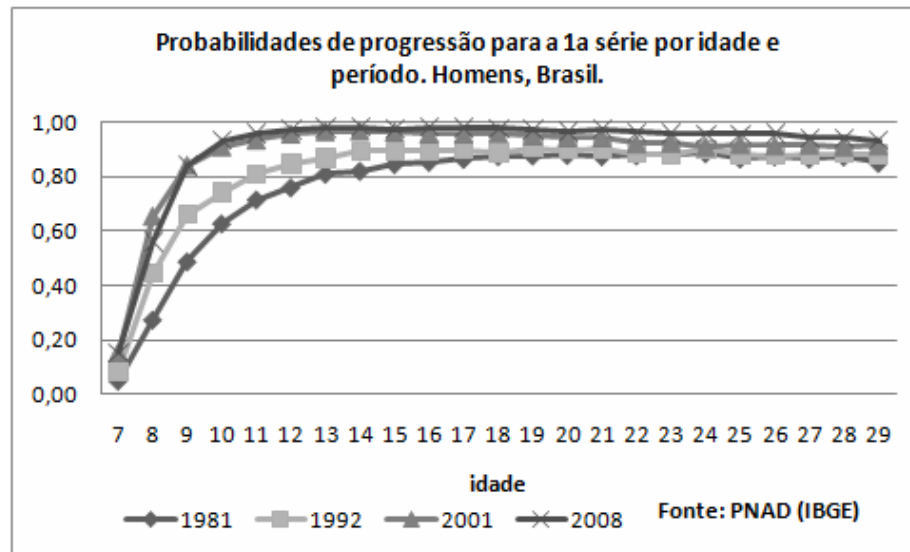
(b) e_4



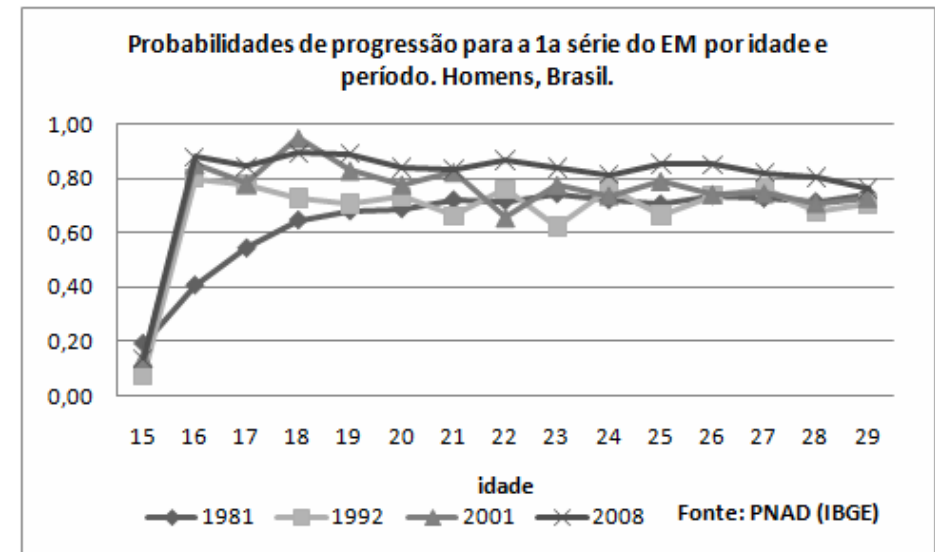
(d) e_{11}

FIGURA 3

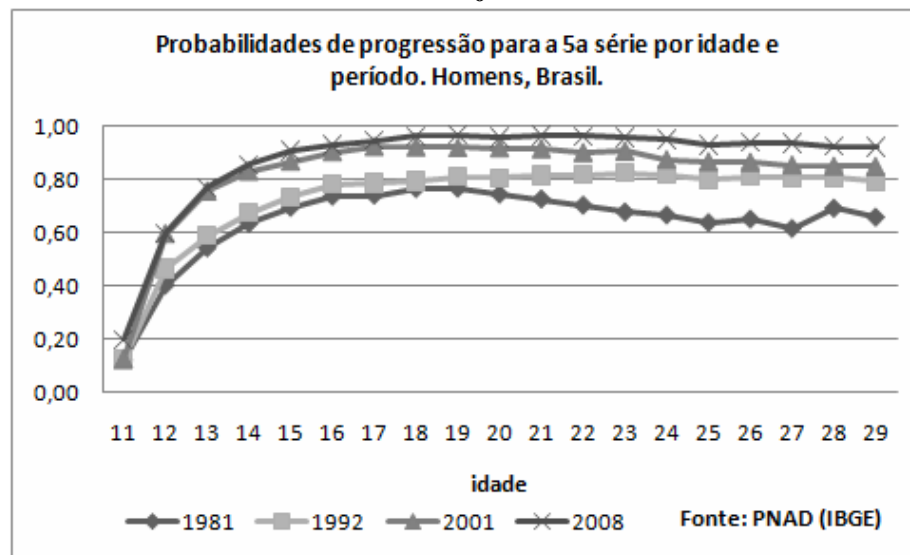
Probabilidade de progressão em séries selecionadas conforme idade e período. Brasil, Homens



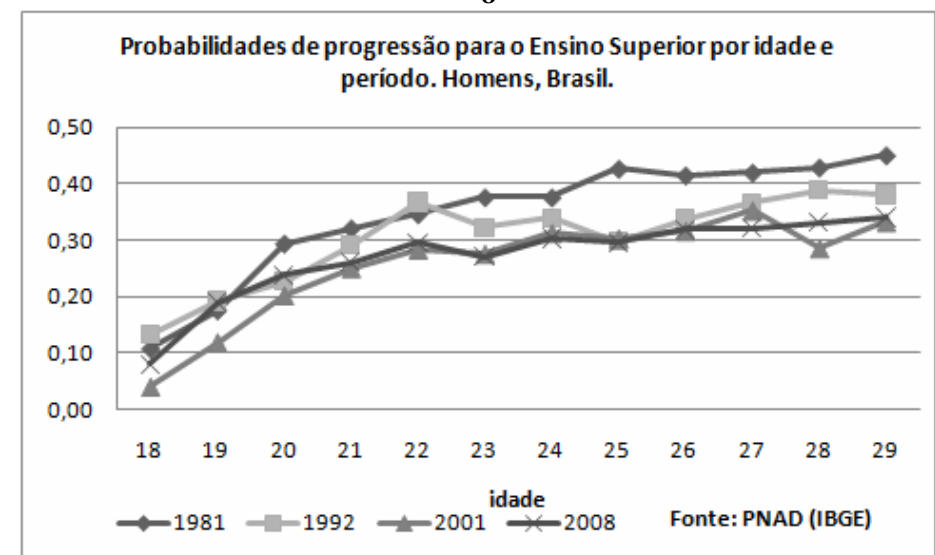
(a) e_0



(c) e_8



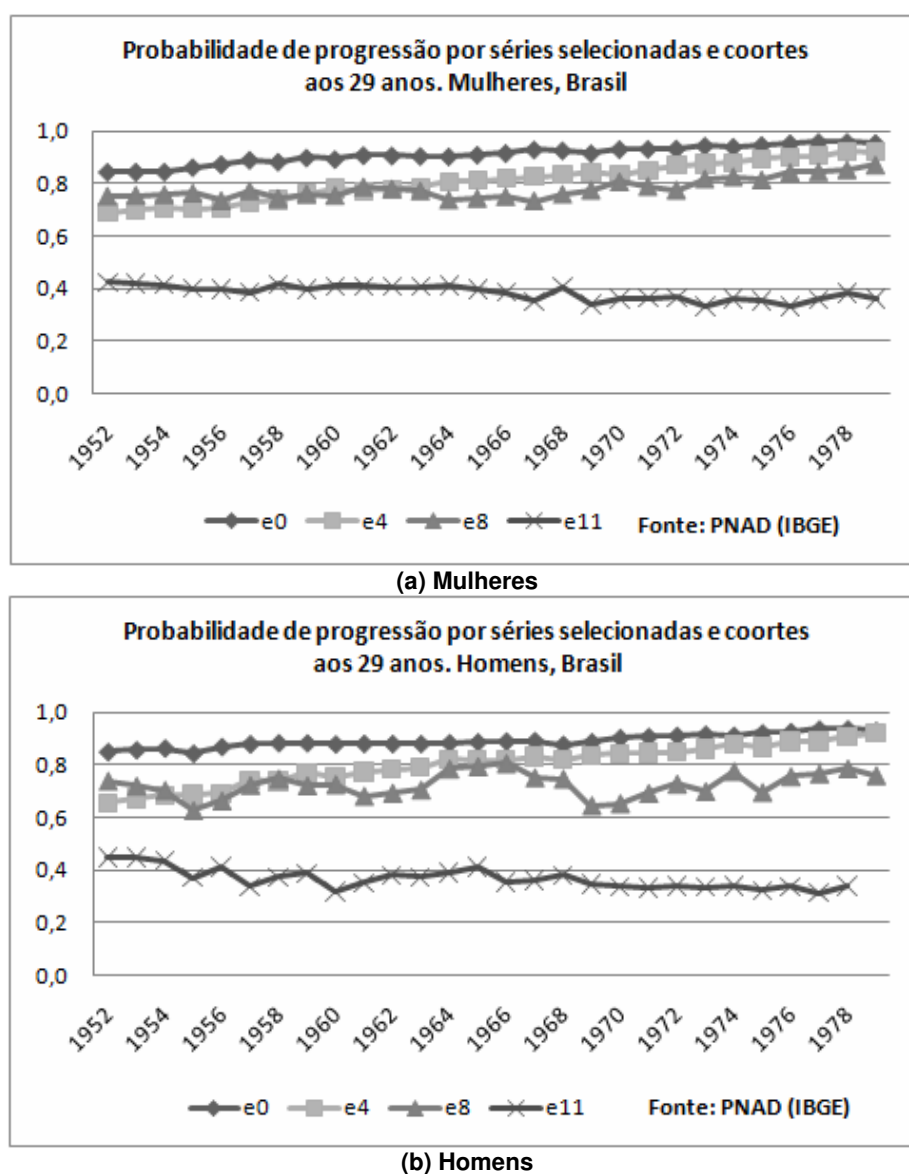
(b) e_4



(d) e_{11}

FIGURA 4

Probabilidade de progressão em séries selecionadas conforme coorte e sexo ao final da carreira escolar. Brasil



Realizamos também uma análise dos coeficientes estimados pelos modelos EI e MLGR em cada uma das transições escolares e conforme o sexo. Com este exercício buscamos verificar em que medida as estimativas dos parâmetros derivada de cada método são discrepantes. Antes contudo, de prosseguirmos com a 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 e Lemeshow 2000).

Vamos analisar, em primeiro lugar, a comparação entre os coeficientes estimados para a probabilidade de progressão para a 1^a série (e_0). Para as mulheres, os efeitos de idade são parecidos nas duas

modelagens, contudo, os efeitos de período e coorte tendem a ser mais discrepantes, embora as figuras mostrem que o comportamento dos coeficientes ao longo dos períodos e coortes tende a ser semelhante. Para os homens, contudo, além da divergência entre os modelos entre os efeitos de idade e coorte, os efeitos de idade também são distintos entre os métodos (Figuras 5 e 6).

Comparando-se os coeficientes estimados para a probabilidade de progressão para a 5^a série (e_4), podemos verificar que, no caso das mulheres, os efeitos de idade apresentam ligeira diferença entre os métodos, enquanto os efeitos de período e coorte tendem a apresentar magnitudes bastante distintas. Este padrão se repete para os homens, e, da mesma forma que foi visto para e_0 , os efeitos de idade se diferem mais do que o observado para as mulheres (Figuras 7 e 8).

Nossos resultados apontam, ainda, que a diferença entre os coeficientes estimados pelos dois métodos pode, em certos casos, levar a diferentes interpretações a respeito dos efeitos. Por exemplo, os efeitos de coorte sobre a probabilidade de progressão para o Ensino Médio, dado que concluiu a oitava série do Ensino Fundamental (e_8) para as mulheres mostra que, entre 1974 e 1988, estes efeitos eram positivos segundo o EI e negativos segundo o MLGR (Figura 9). No caso dos homens, a Figura 10 mostra que os efeitos de período possuem um comportamento inverso conforme o método IPC escolhido. Esta evidência é importante, e corrobora os estudos que revelam o quanto a escolha do método IPC pode influenciar nos resultados e nas interpretações.

Em relação à probabilidade de progressão para o Ensino Superior, dado que concluiu a terceira série do Ensino Médio (e_{11}), as estimativas EI e MLGR tendem a ser mais semelhantes em termos do comportamento ao longo das idades, períodos e coortes no caso das mulheres (Figura 11). Todavia, no caso dos homens mais uma vez os efeitos de período são opostos entre os dois métodos, e as estimativas apresentam ligeira divergência em relação ao padrão observado para as mulheres (Figura 12).

Outra comparação importante entre as metodologias EI e MLGR diz respeito aos valores preditos pelos dois modelos das probabilidades de progressão por série por idade. Procedemos com esta análise em quatro pontos no tempo: 1981, 1992, 2001 e 2008. Nosso objetivo foi verificar em que medida os dois estimadores divergem em relação ao valor verdadeiro da probabilidade de progressão por série selecionada, e em que medida esta divergência sofre influência do período analisado. Cumpre ressaltar que, embora os modelos apresentem as mesmas medida de qualidade do ajuste, não necessariamente os modelos devem prever os mesmos valores da PPS. Considerando-se a definição da *deviance*, pode ser que o resultado da soma dos quadrados das diferenças entre os valores preditos por cada um dos modelos seja a mesma, enquanto os valores preditos se diferem. Neste caso, haverá uma compensação dentro da fórmula que resultará na mesma *deviance*.

No caso de e_0 , observa-se que tanto para homens e mulheres os dois métodos, EI e MLGR, convergem para os valores verdadeiros das PPSs, e este resultado é robusto nos quatro períodos analisados (Figuras 13 e 14). Para e_4 , também os dois métodos convergem para o valor verdadeiro da PPS para ambos os sexos, a exceção em 1981, quando as PPSs preditas pelo MLGR estão sobreestimadas (Figuras 15 e 16).

Interessante é a análise dos resultados dos valores preditos para a probabilidade de progressão para o Ensino Médio, e_8 , especialmente devido a um padrão divergente na magnitude dos estimadores segundo o sexo. No caso dos homens, o estimador MLGR e o EI mostraram-se muito divergente dos valores verdadeiros

em alguns casos, especialmente em 1981 (Figura 17). Já para as mulheres, os dois estimadores tendem a convergir para os valores verdadeiros em todos os períodos (Figura 18).

Por fim, contrastamos as probabilidades observadas de progressão para o Ensino Superior (e_{11}) por idade com aquelas estimadas pelo EI e por MLGR. Fica clara a mudança na escala dos gráficos, uma vez que esta probabilidade de progressão assume valores bastante baixos em relação às demais. Verificamos que ambos os métodos divergem ligeiramente em relação aos valores verdadeiros, para ambos os sexos, e tendem a assumir valores preditos semelhantes. Na realidade, as estimativas dos modelos tendem a suavizar a tendência por idade nestas probabilidades de progressão (Figuras 19 e 20).

Agora passamos à análise da eficiência dos dois métodos. Vimos na seção 3 que se demonstrou matematicamente que o EI possui menor variância do que qualquer estimador MLGR, i.e., qualquer estimador MLGR obtido por quaisquer restrições de identificação. Para investigar esta assertiva no estudo das probabilidades de progressão por série, 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 que, de fato, o estimador intrínseco é mais eficiente - i.e, possui menor variância - do que o MLGR, tanto para os efeitos de idade, período e coorte, para ambos os sexos e em todas as transições selecionadas. Quando analisamos os gráficos para os efeitos de idade, verificamos que o estimador intrínseco possui uma excelente eficiência em relação ao MLGR (Figuras 21, 22, 23, 24). A variância do EI, contudo, aumenta quando os coeficientes para período ou coorte não são significativos, mas ainda assim este estimador possui de fato uma menor variância (Figuras 25, 26, 27, 28, 29, 30, 31, 32). Este resultado é muito importante e desejável para projeções educacionais baseadas na PPS, de forma que uma extrapolação do erro-padrão das estimativas baseadas no estimador intrínseco tenderá a ser bastante eficiente.

6. CONSIDERAÇÕES FINAIS

Os modelos idade-período-coorte possuem uma forte tradição na Demografia. Neste artigo, buscou-se com a análise IPC a investigação sobre as diferentes fontes de variação nas probabilidades de progressão por série. Este arcabouço é adequado, pois fornece evidências substantivas sobre os componentes específicos da dinâmica da progressão escolar no recente contexto histórico.

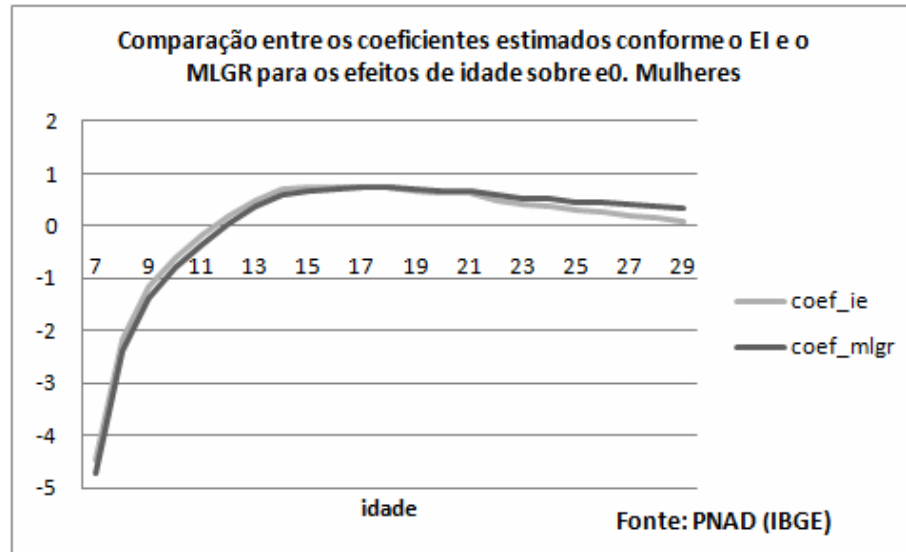
Este artigo possui, contudo, um forte caráter metodológico. Buscou-se comparar duas metodologias, sendo a primeira delas usual na literatura demográfica, baseada nos trabalhos de Fienberg e Mason (1985), a qual denominamos de solução por *modelos lineares generalizados restritos* (MLGR), e uma metodologia 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, através 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. A grande virtude e singularidade deste estimador é que ele não requer a adoção de pressupostos de igualdade sobre os parâmetros de idade, período ou coorte. A única restrição necessária se baseia na orientação do estimador no espaço paramétrico, a qual depende fundamentalmente de uma matriz de design X fixa, i.e., do número de períodos e grupos de idade.

Vimos 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 significância estatística condizente 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 os estimadores 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, com variância maior do que os estimadores EI, e incondizentes com a análise substantiva da evolução dos indicadores de progressão escolar por série no Brasil.

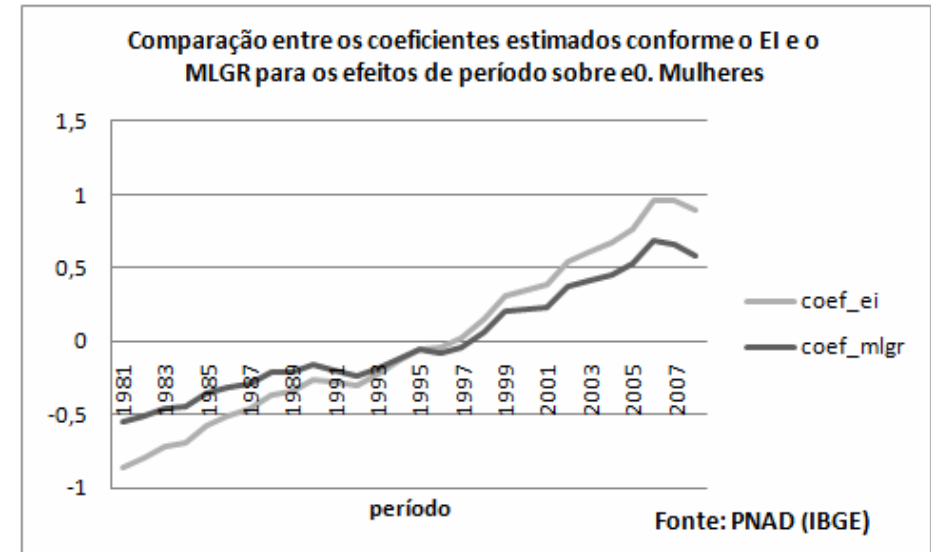
Diante disso tudo, argumentamos que o estimador intrínseco apresenta-se de fato como uma ferramenta poderosa na análise IPC. Por conseguinte, a construção de projeções probabilísticas das PPSs a partir deste modelo se mostra bastante promissora. Este é o próximo passo que pretendemos implementar em artigos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

FIGURA 5

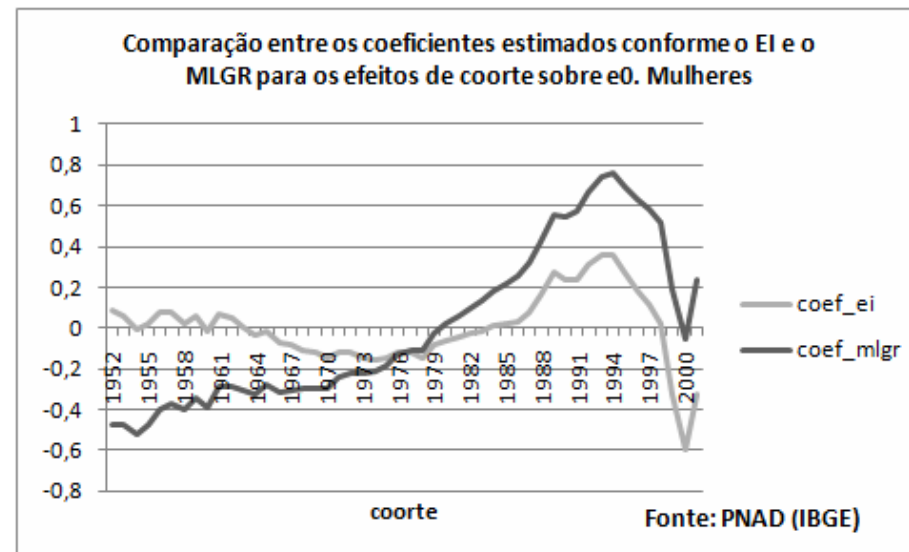
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_0 Brasil, Mulheres



(a) Idade



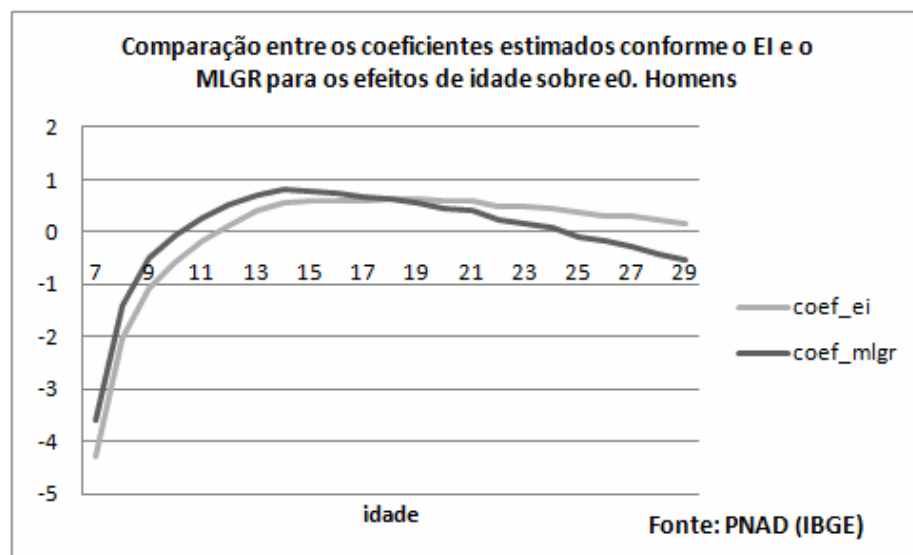
(b) Período



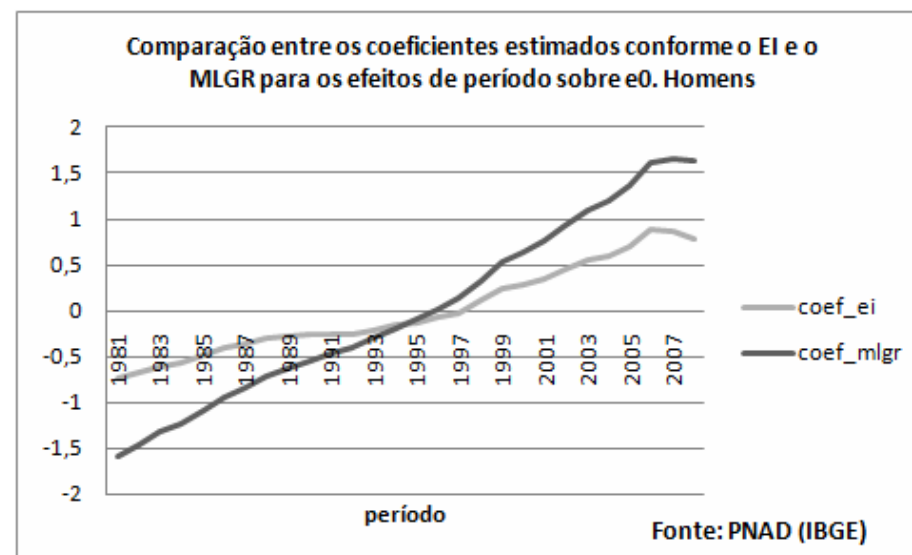
(c) Coorte

FIGURA 6

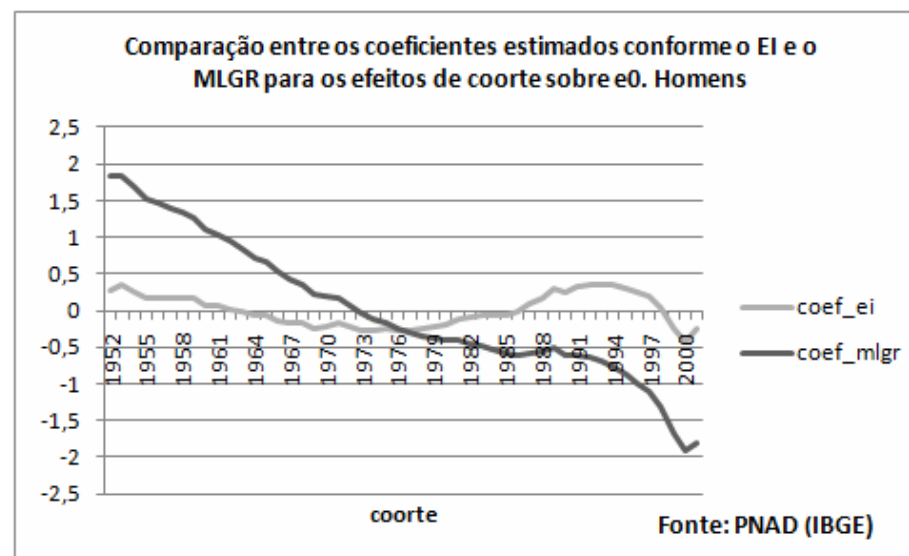
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_0 Brasil, Homens



(a) Idade



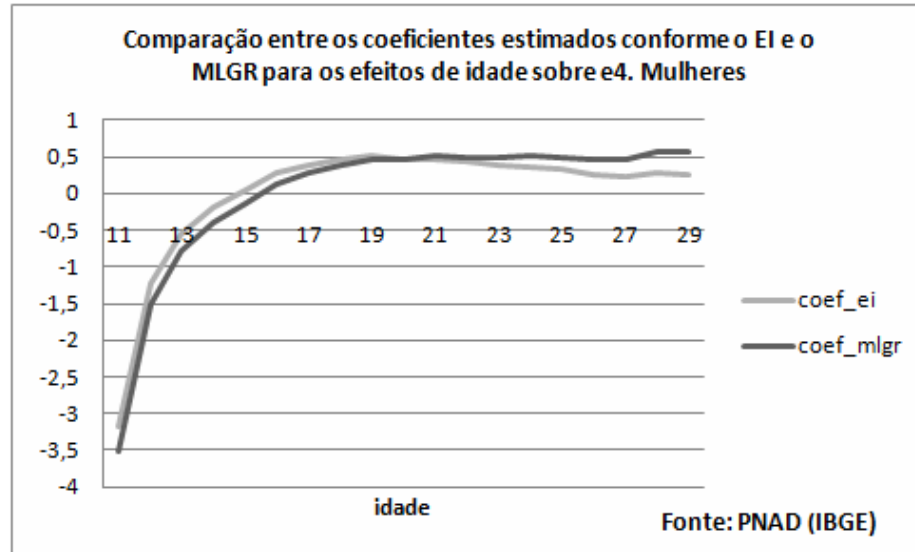
(b) Período



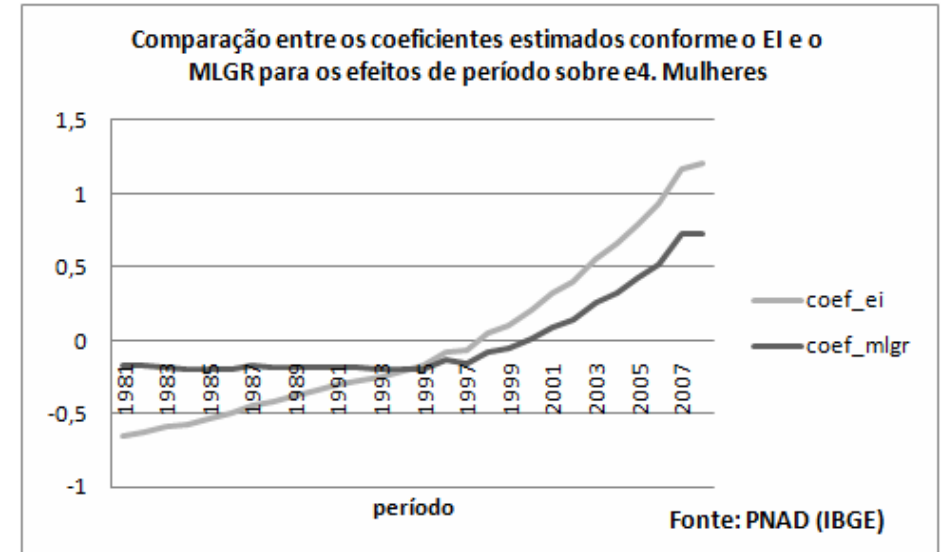
(c) Coorte

FIGURA 7

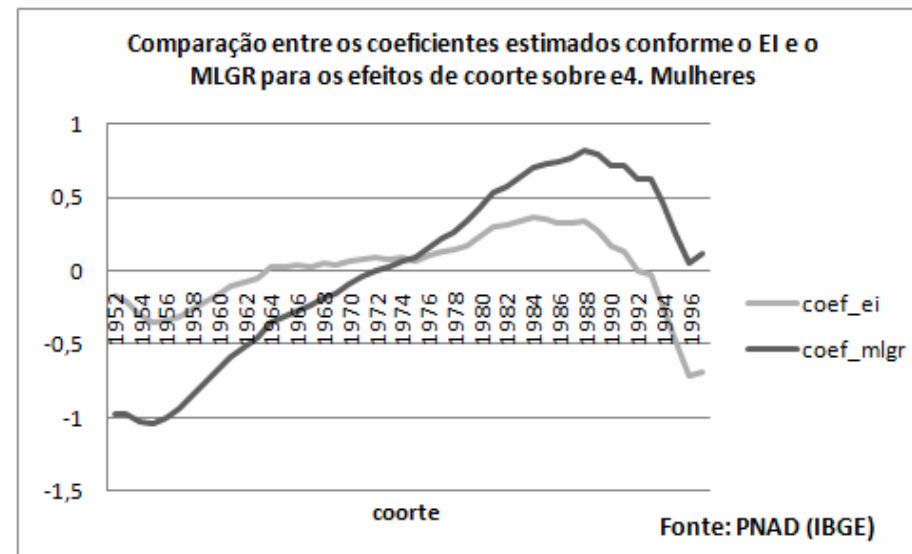
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_4 Brasil, Mulheres



(a) Idade



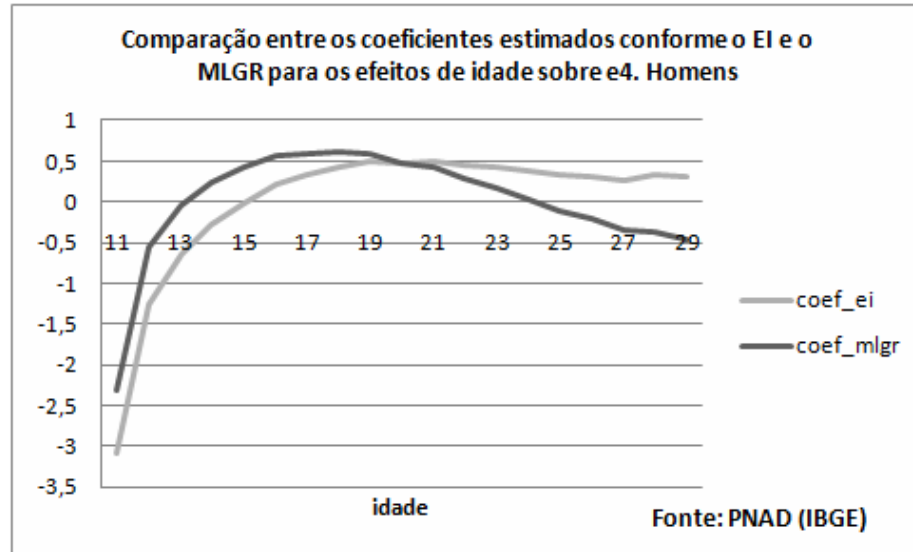
(b) Período



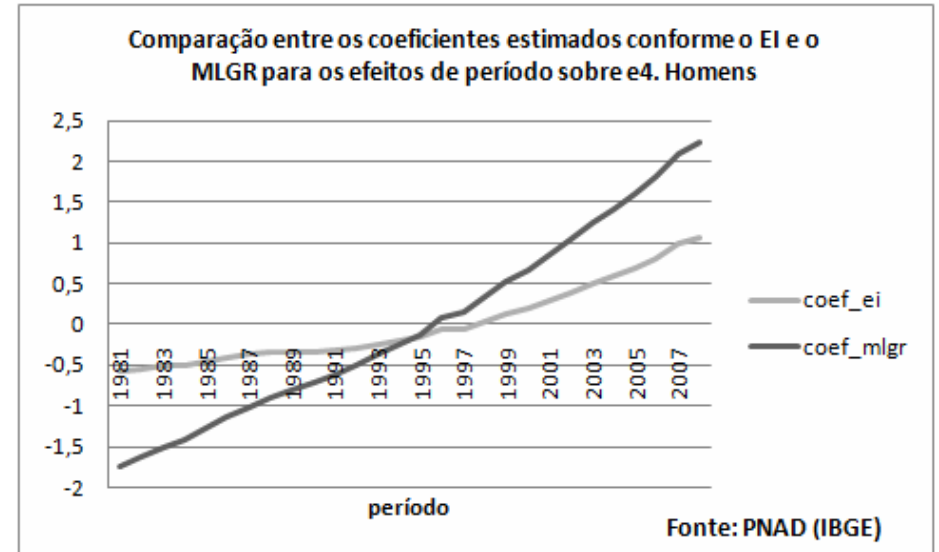
(c) Coorte

FIGURA 8

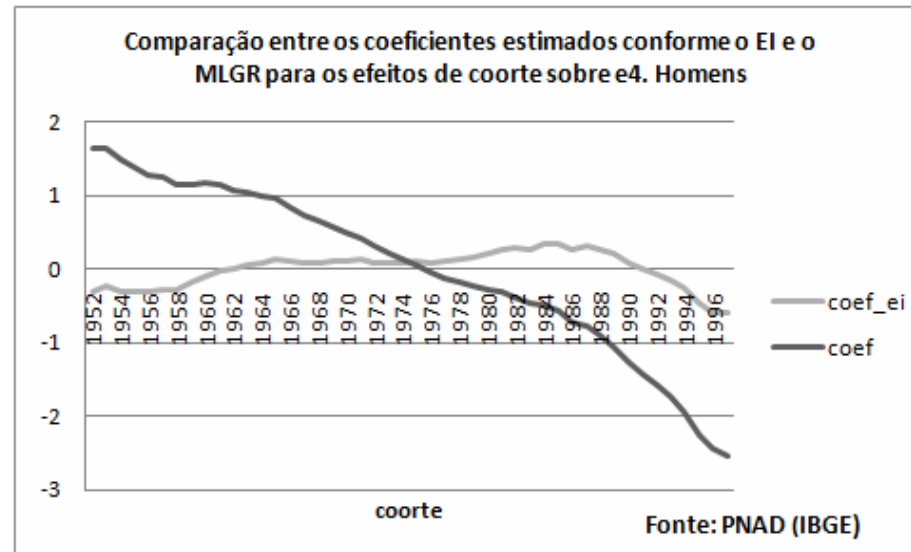
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_4 Brasil, Homens



(a) Idade



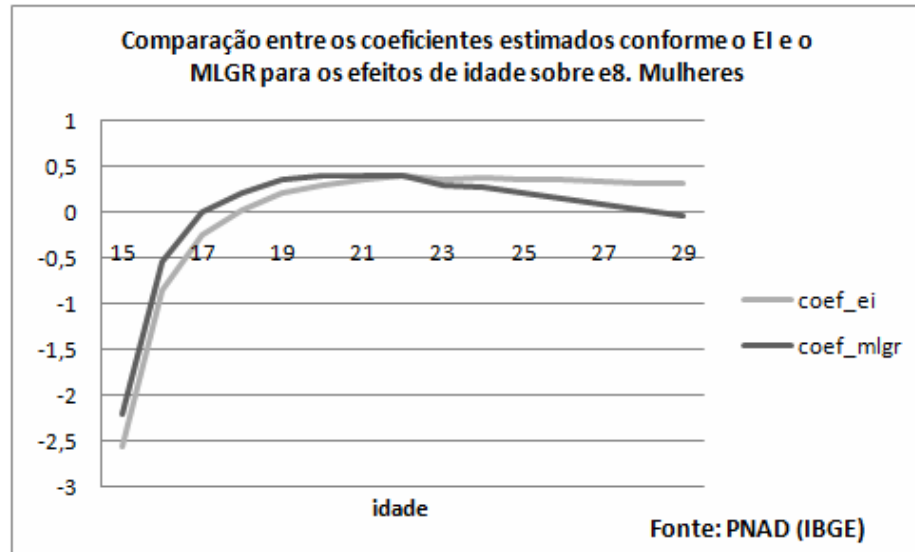
(b) Período



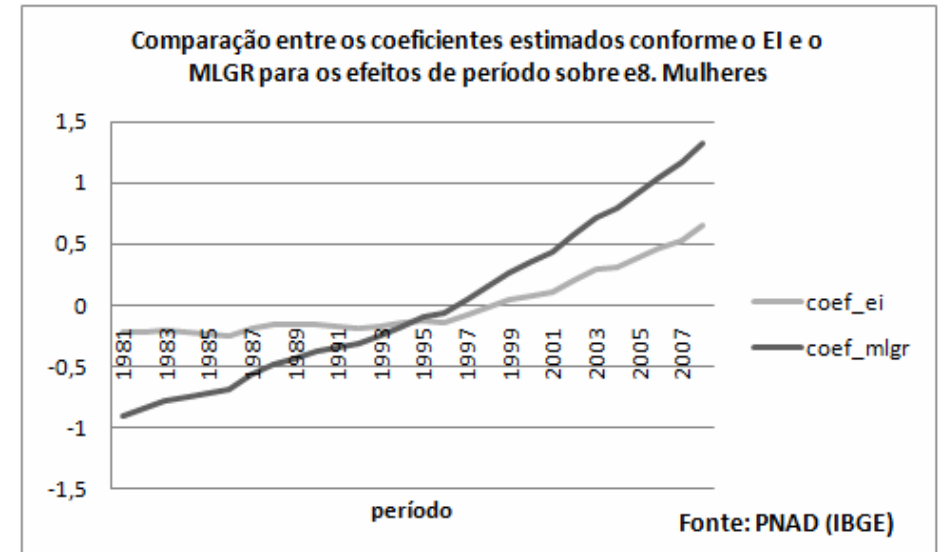
(c) Coorte

FIGURA 9

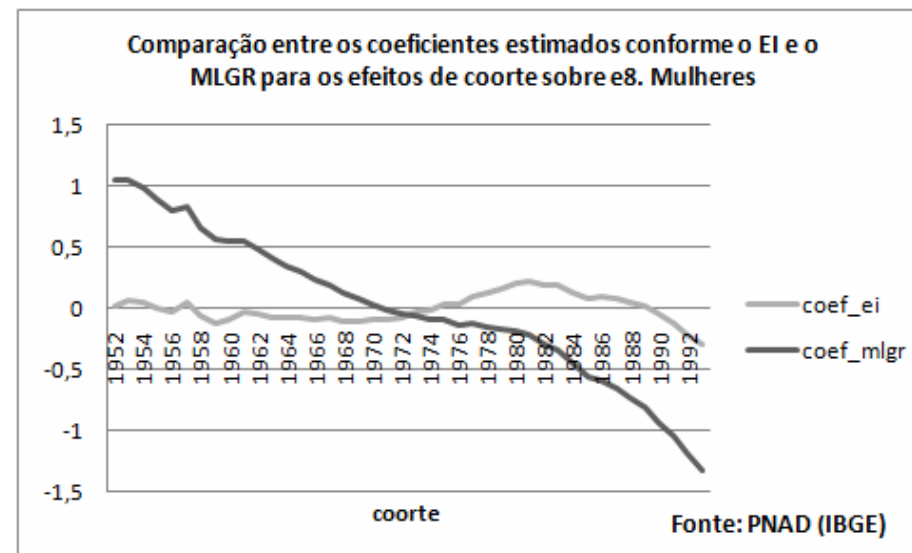
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_8 Brasil, Mulheres



(a) Idade



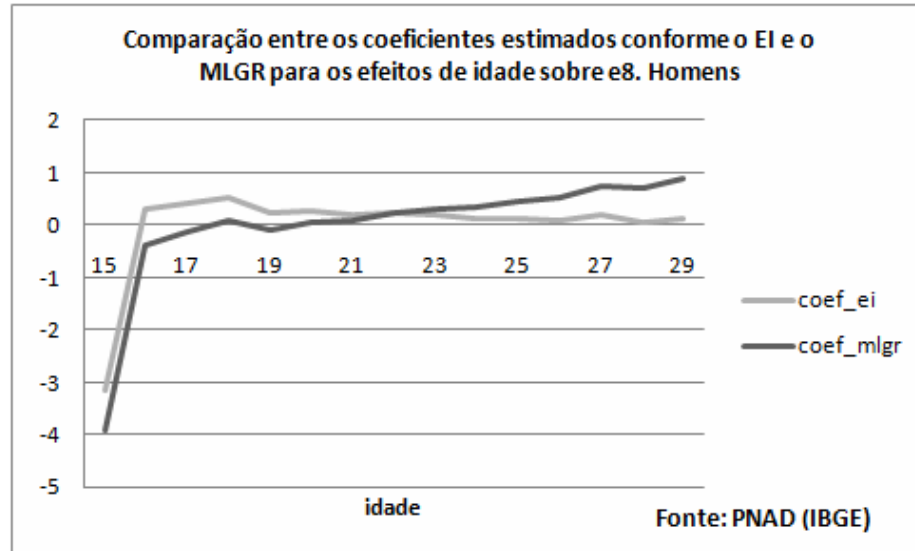
(b) Período



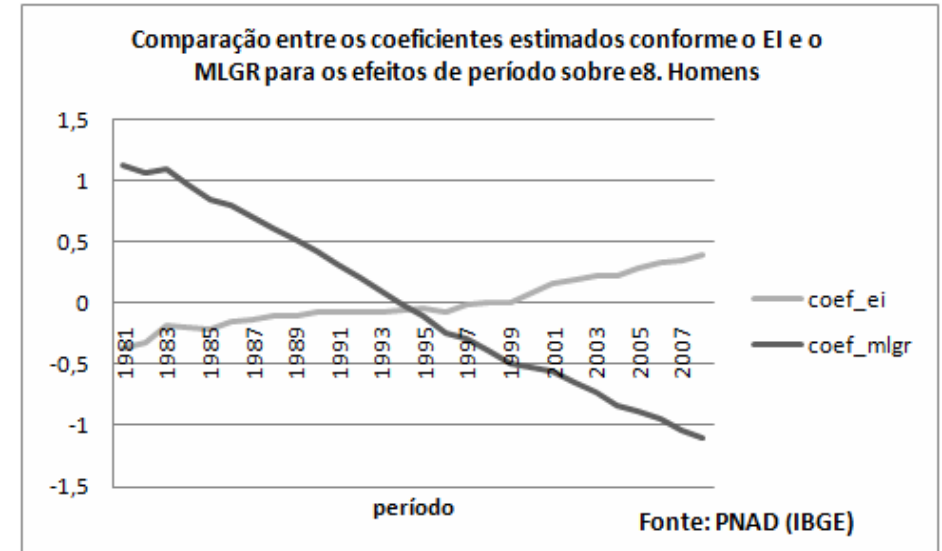
(c) Coorte

Figura 10

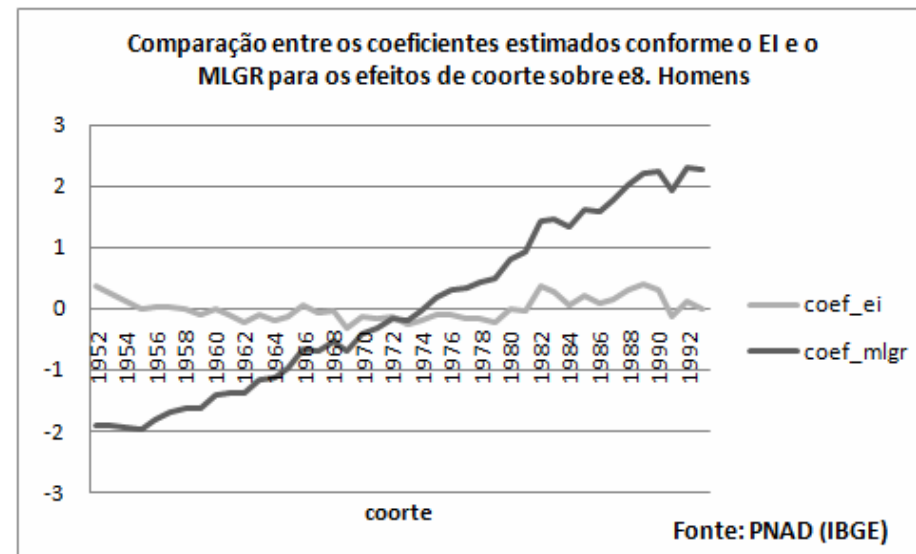
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_8 Brasil, Homens



(a) Idade



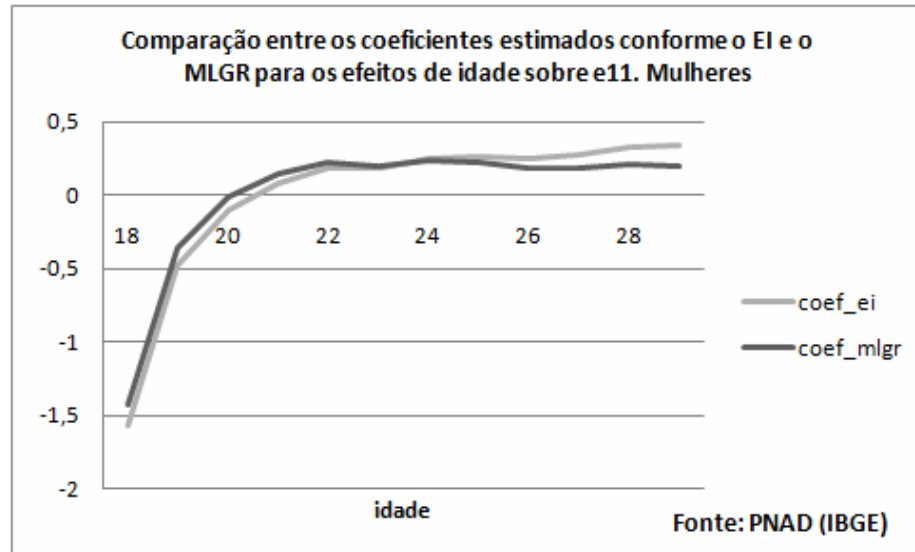
(b) Período



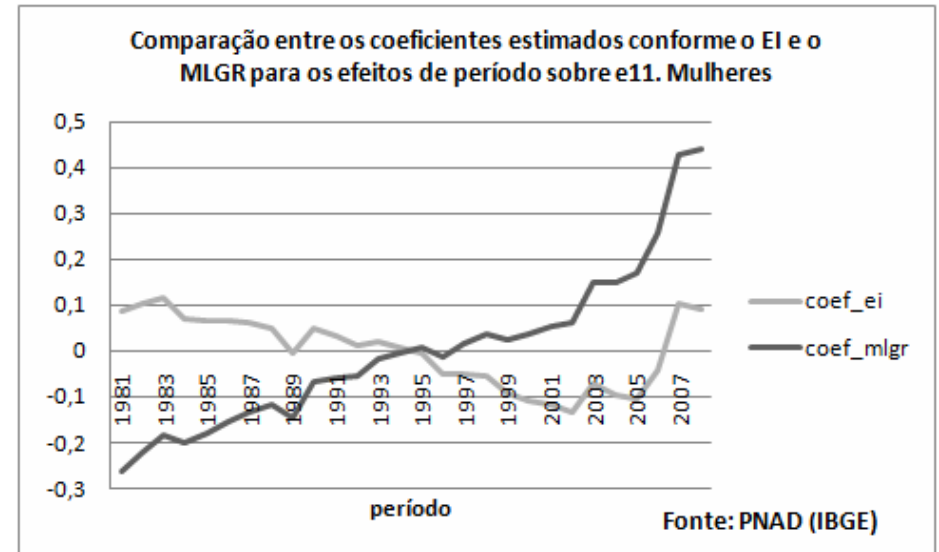
(c) Coorte

FIGURA 11

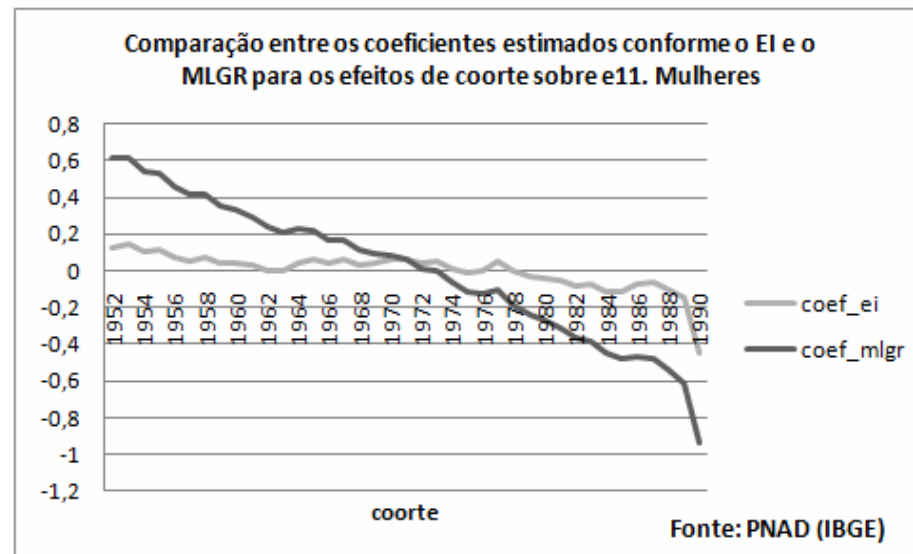
Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_{11} Brasil, Mulheres



(a) Idade



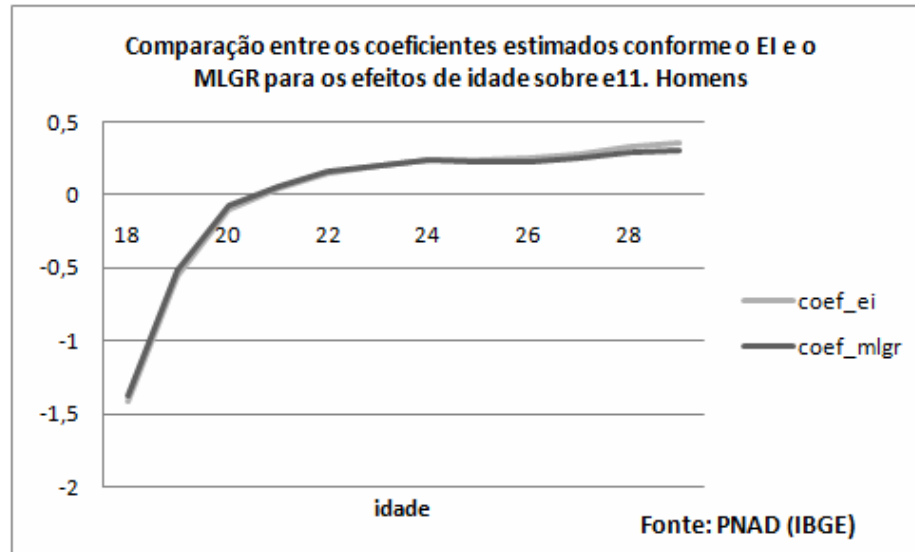
(b) Período



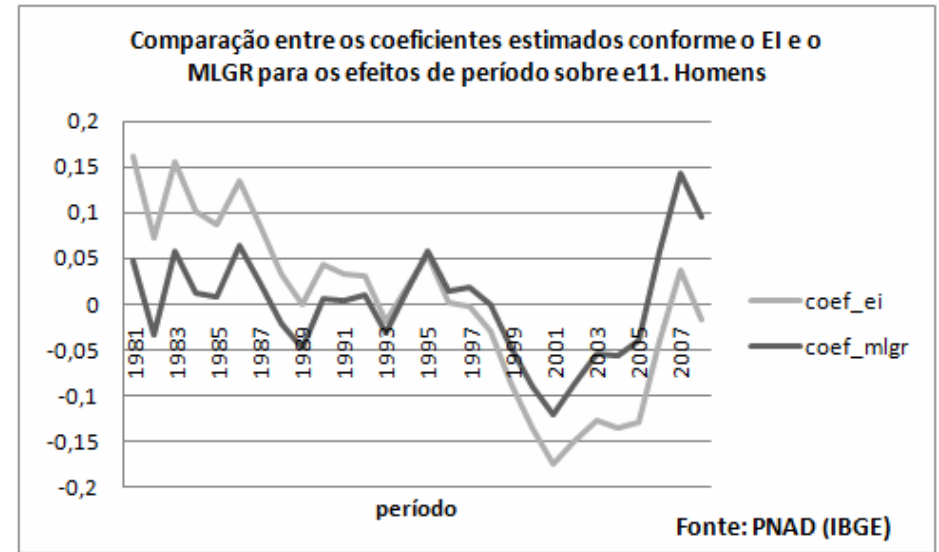
(c) Coorte

FIGURA 12

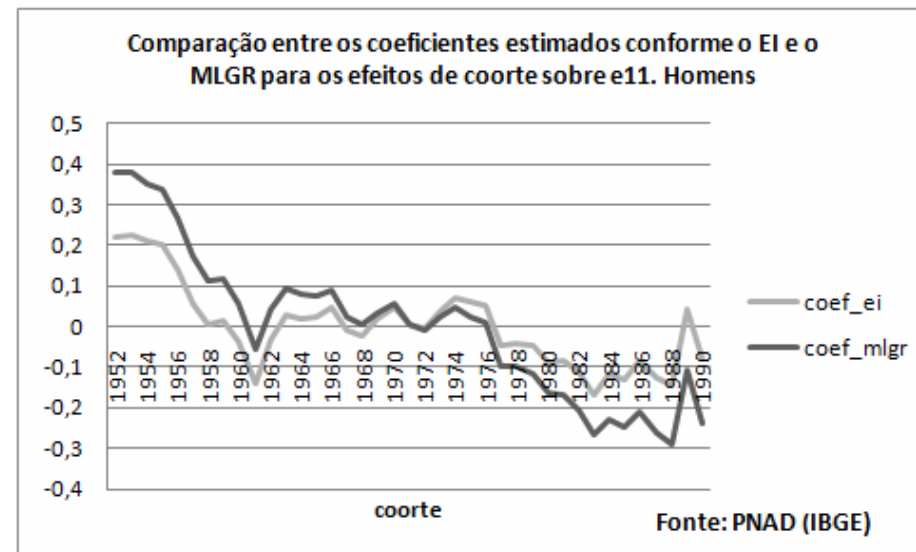
: Comparação entre os coeficientes estimados para os efeitos de idade, período e coorte para e_{11} Brasil, Homens



(a) Idade



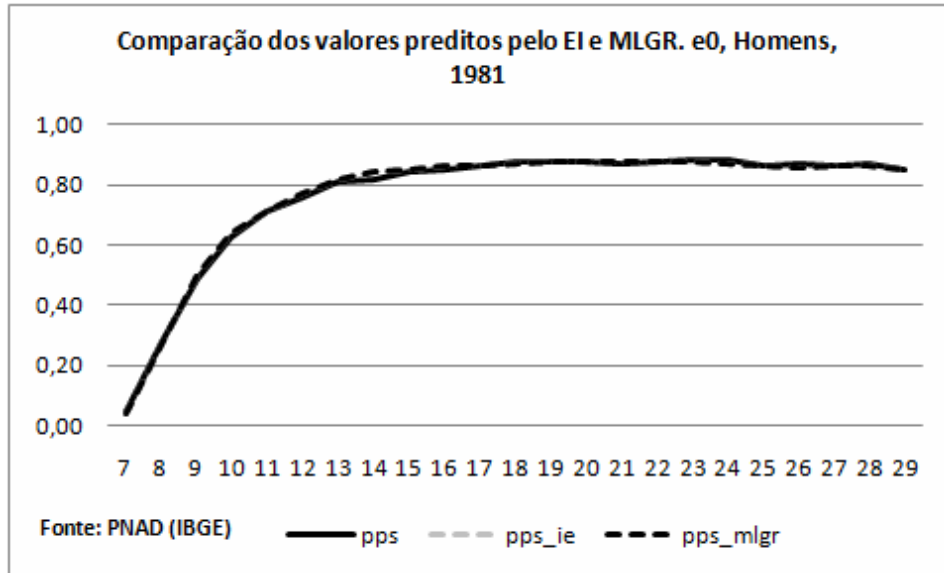
(b) Período



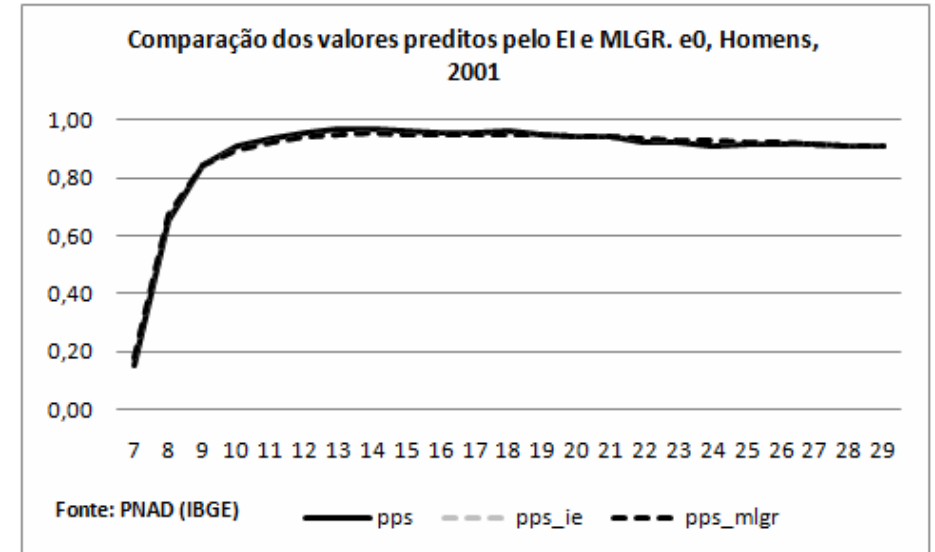
(c) Coorte

FIGURA 13

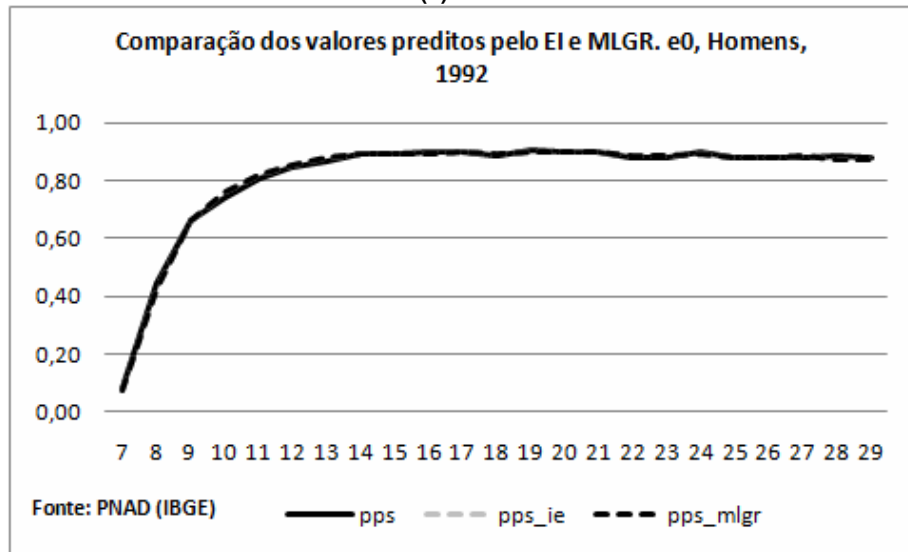
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a primeira série, e_0 , Homens



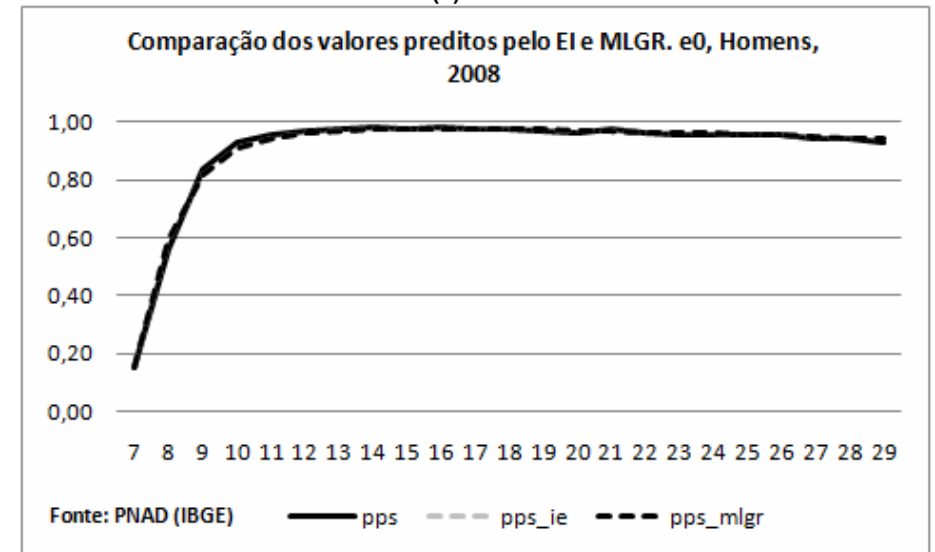
(a) 1981



(c) 2001



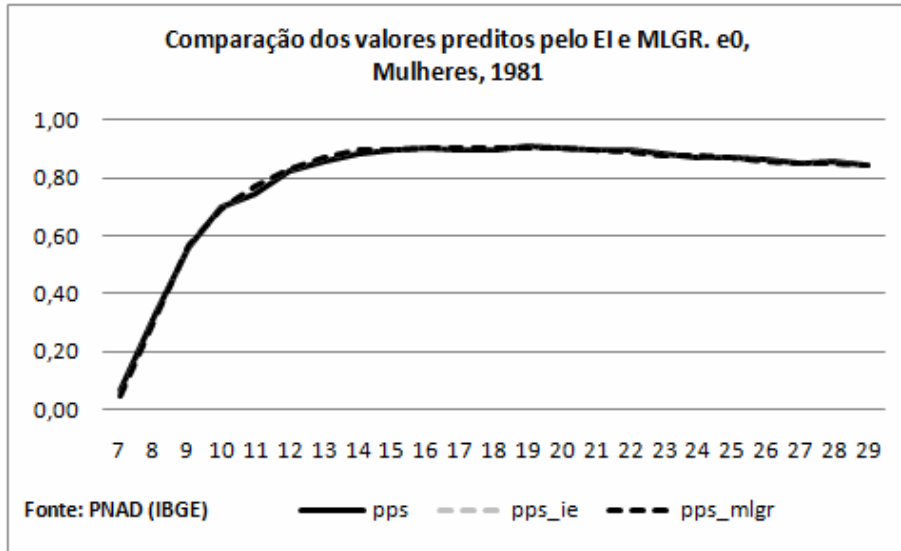
(b) 1992



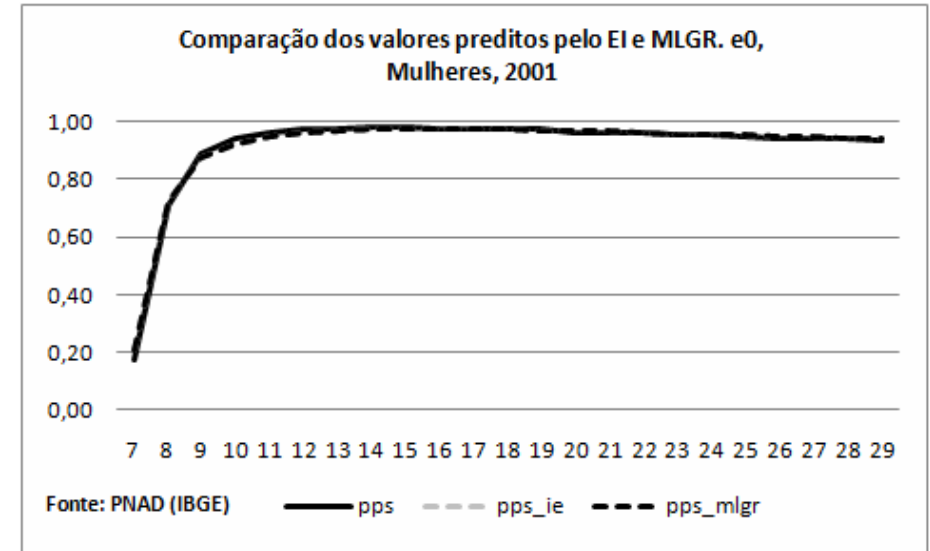
(d) 2008

FIGURA 14

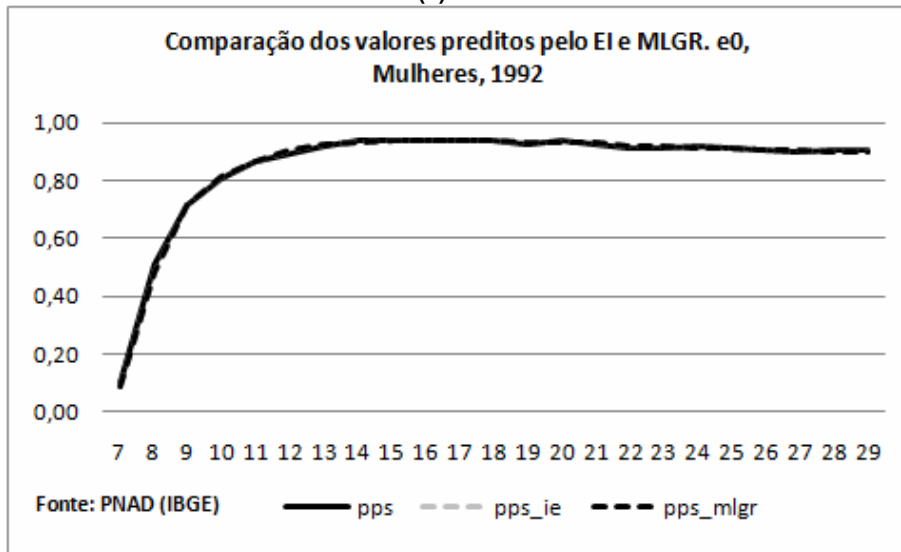
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a primeira série, e_0 , Mulheres



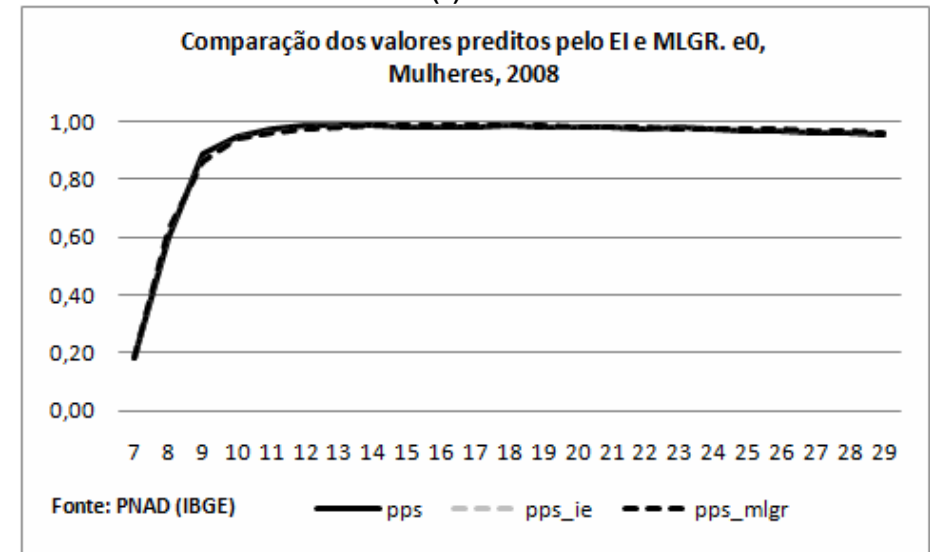
(a) 1981



(c) 2001



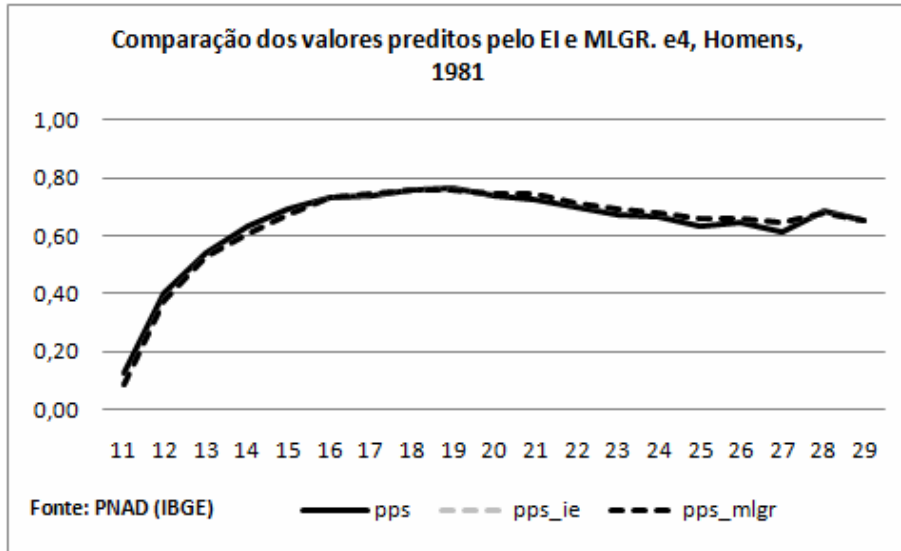
(b) 1992



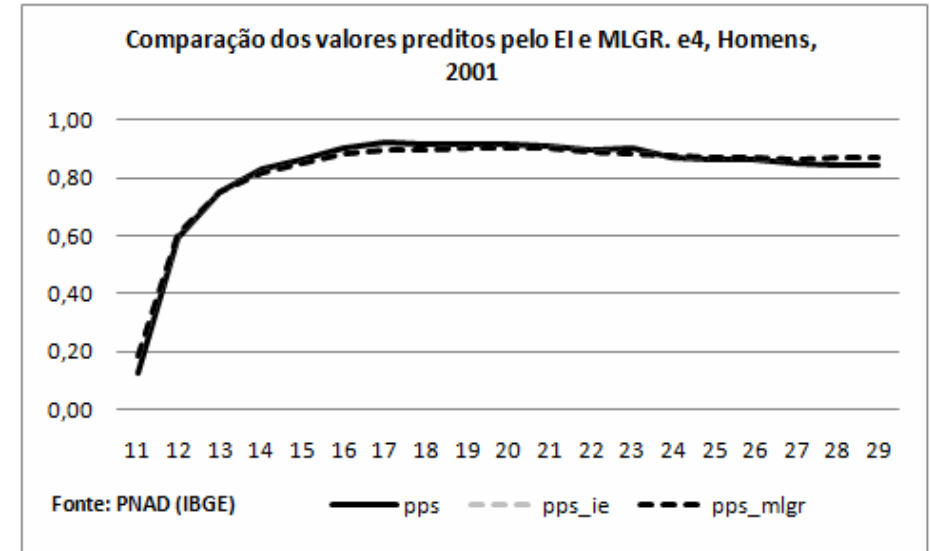
(d) 2008

FIGURA 15

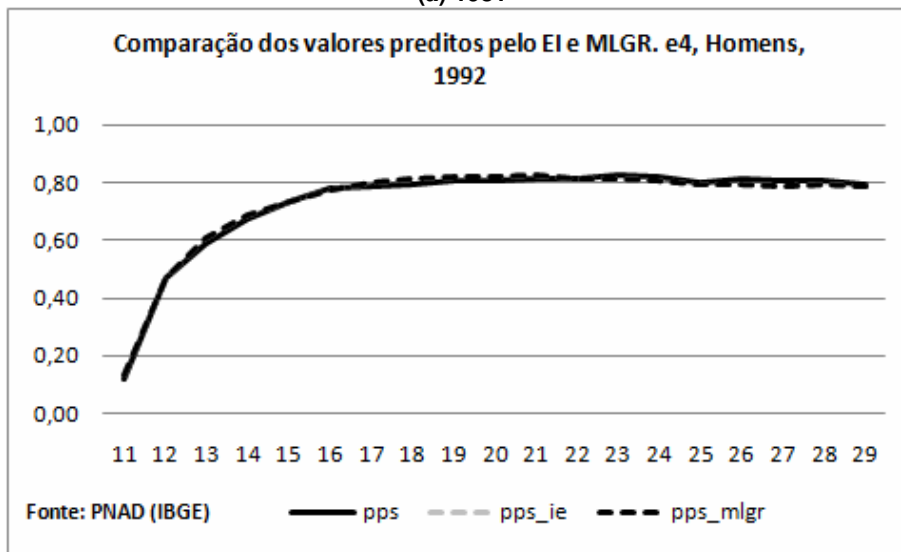
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a quinta série, e_4 , Homens



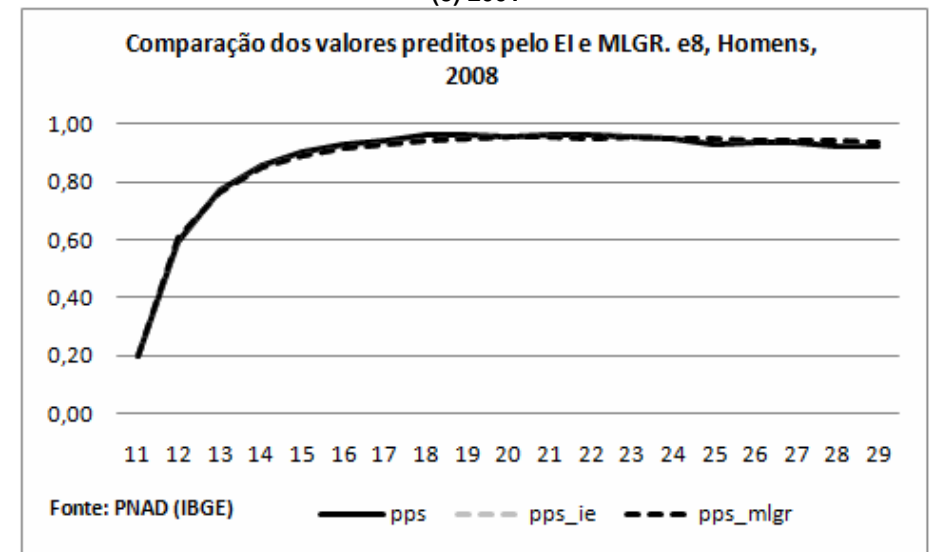
(a) 1981



(c) 2001



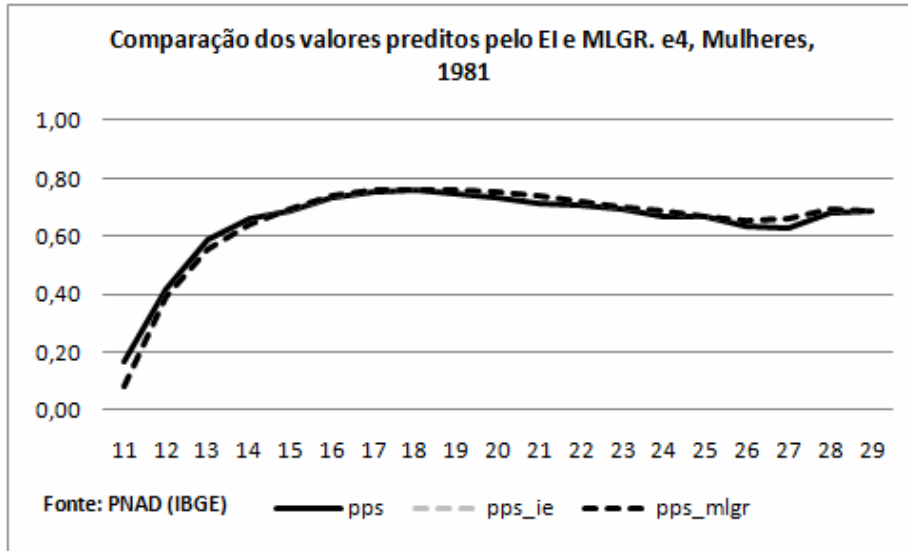
(b) 1992



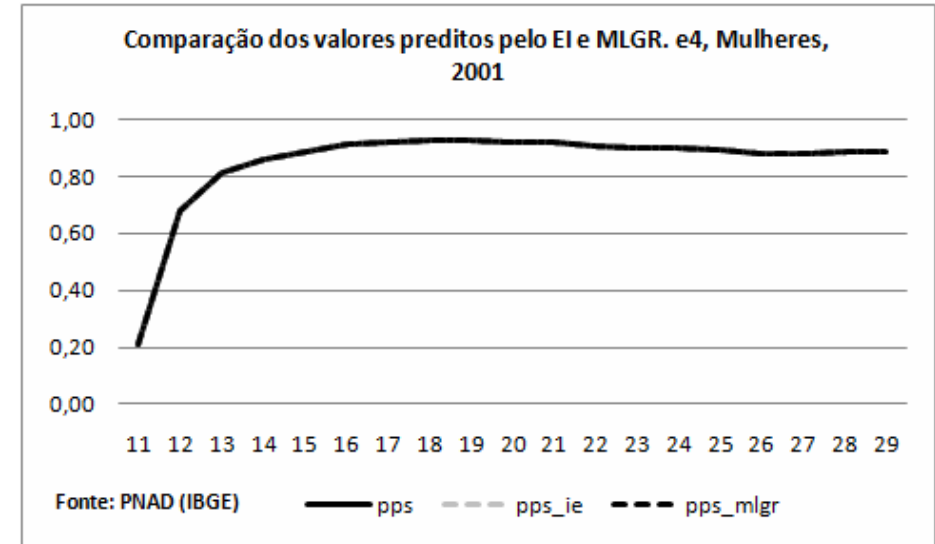
(d) 2008

FIGURA 16

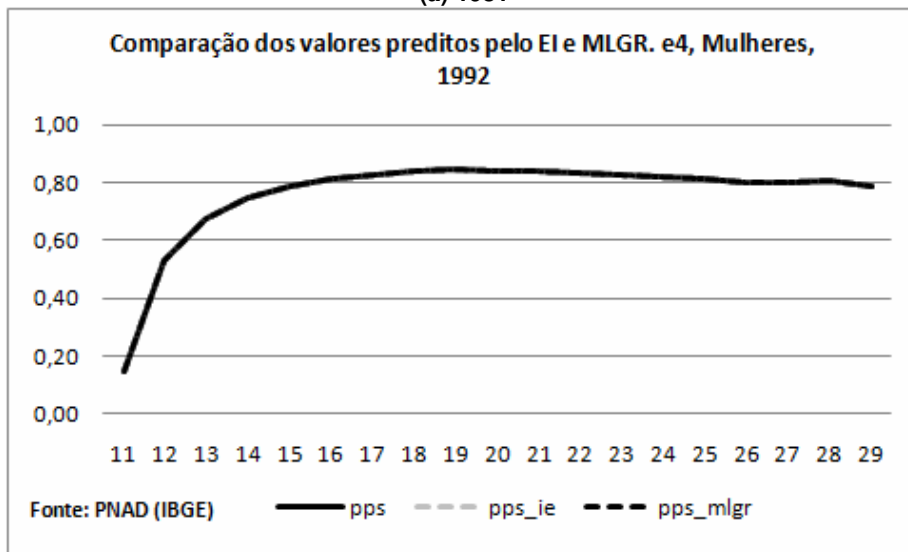
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a quinta série, e_4 , Mulheres



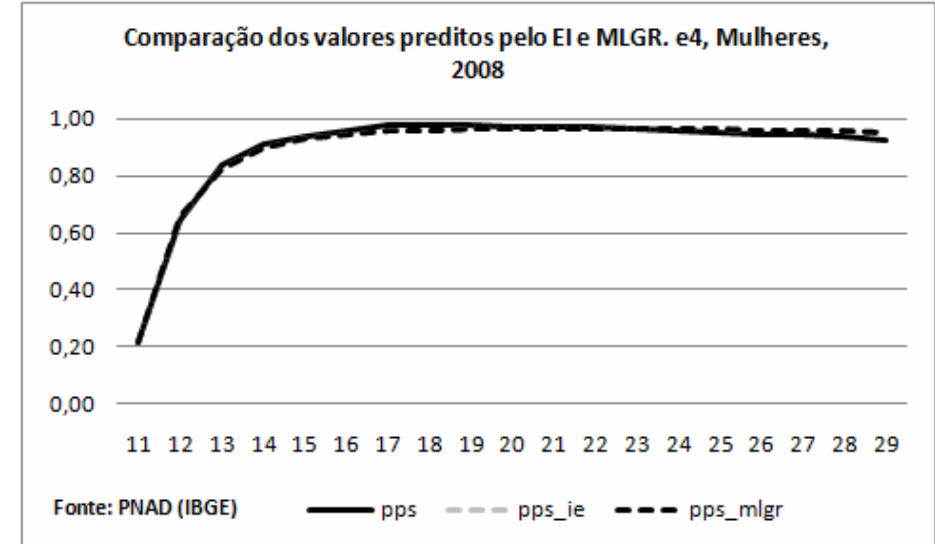
(a) 1981



(c) 2001



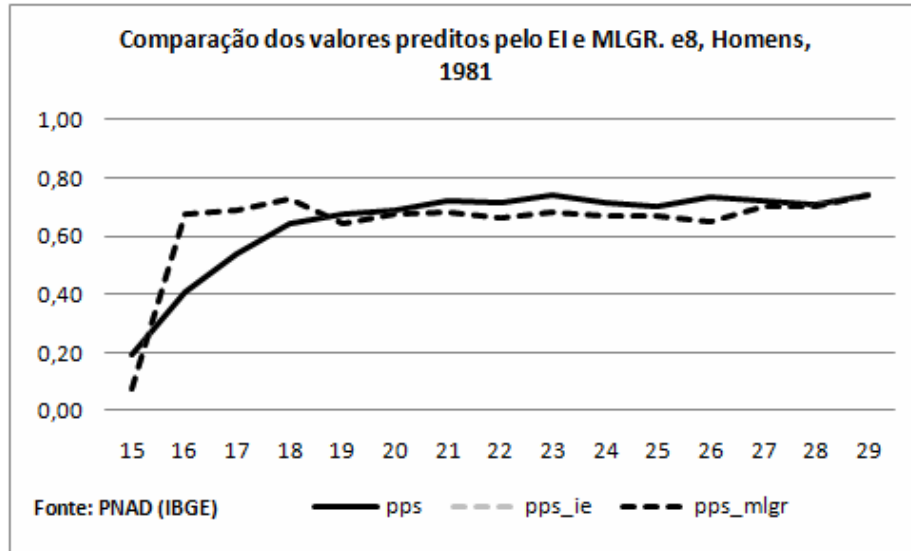
(b) 1992



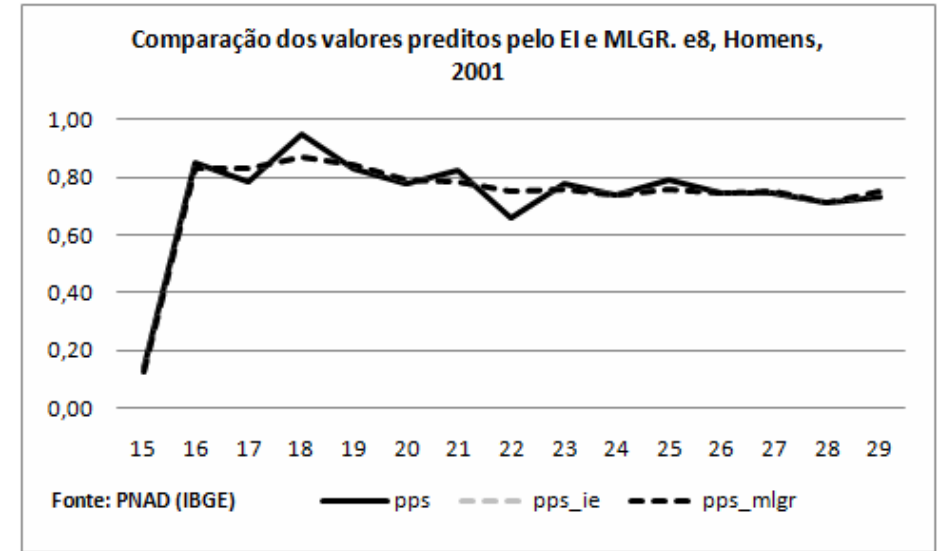
(d) 2008

FIGURA 17

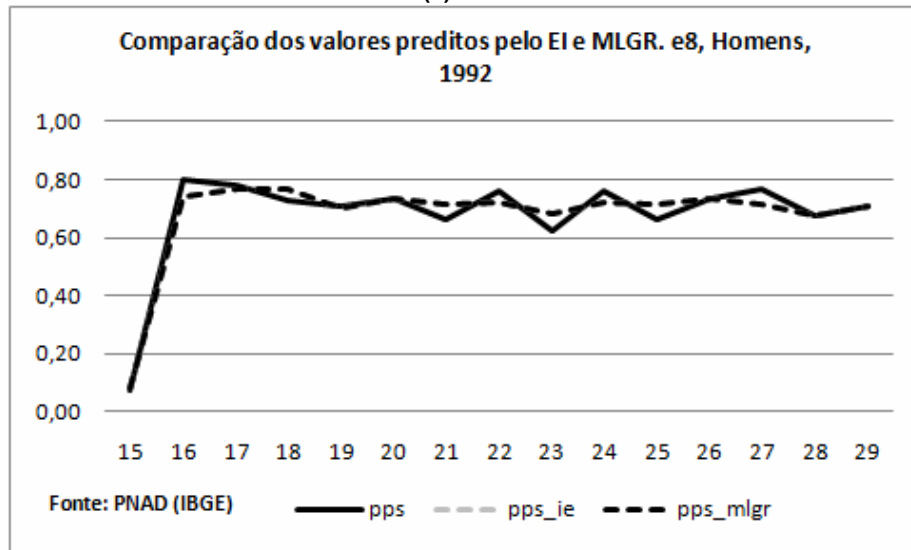
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a 1a série do Ensino Médio, e8, Homens



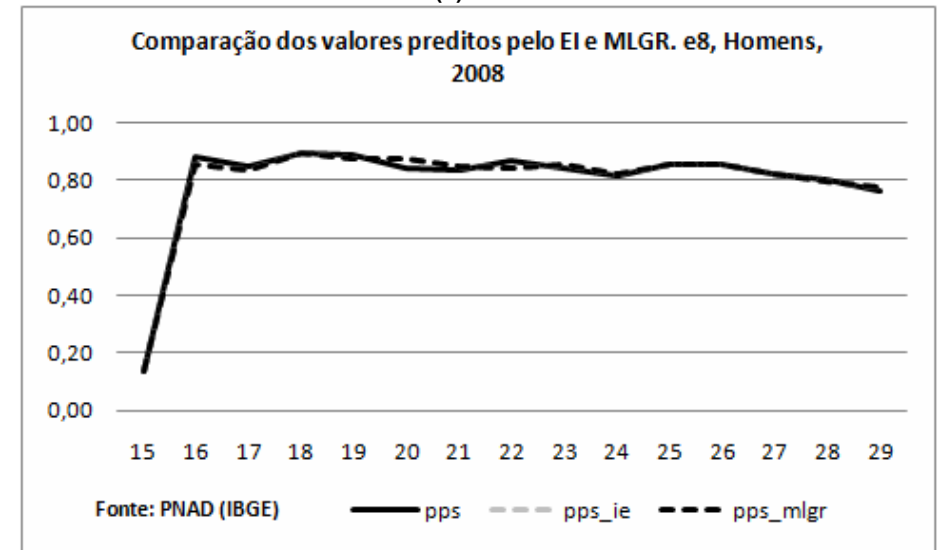
(a) 1981



(c) 2001



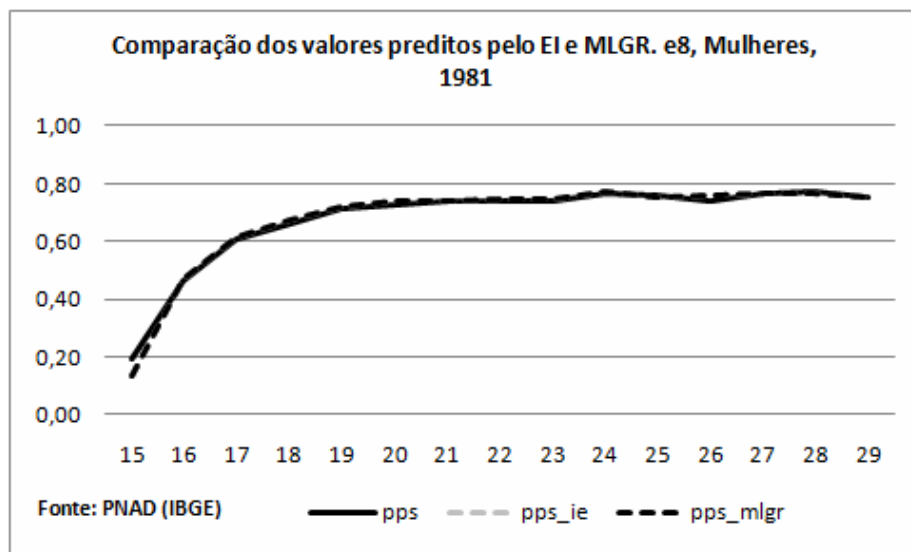
(b) 1992



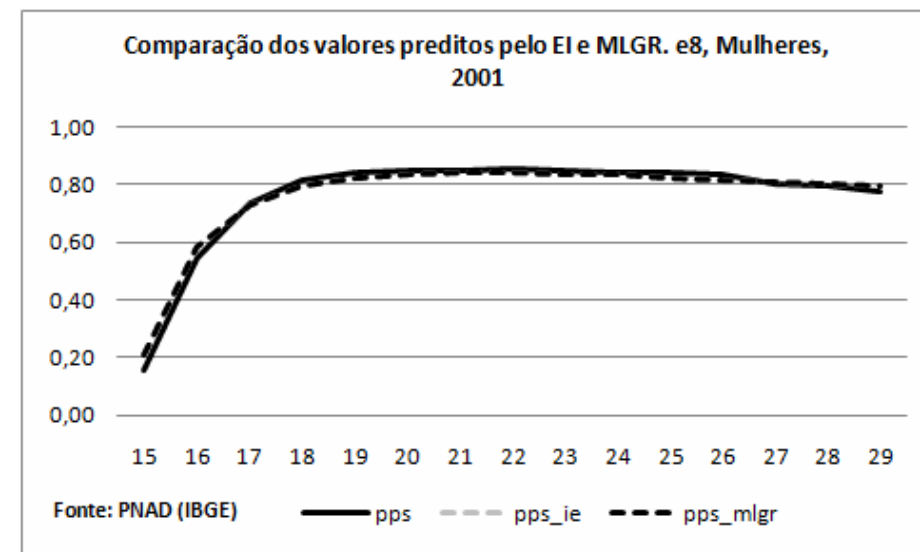
(d) 2008

FIGURA 18

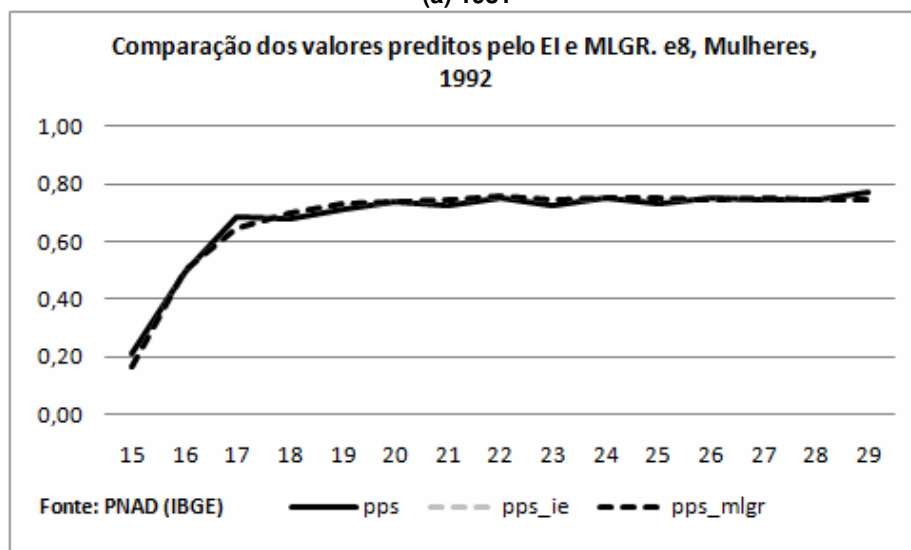
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para a 1a série do Ensino Médio, e8, Mulheres



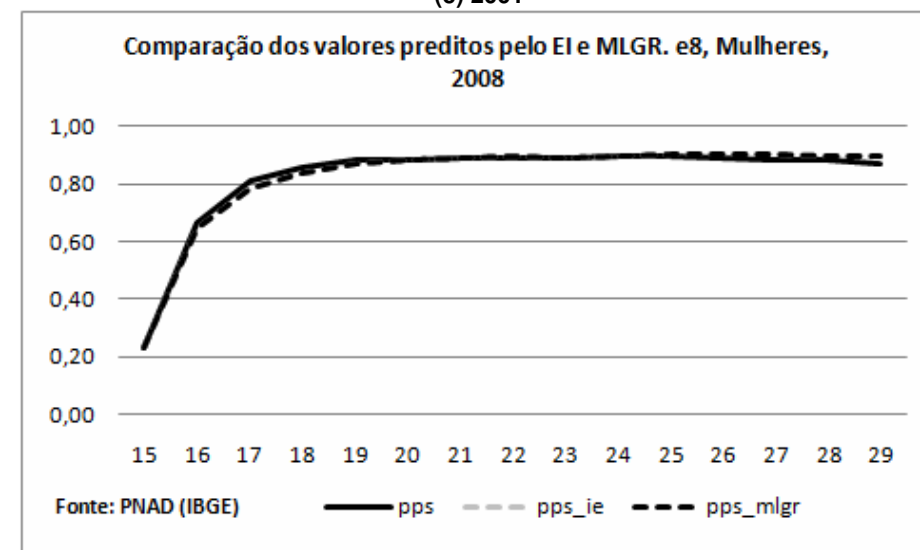
(a) 1981



(c) 2001



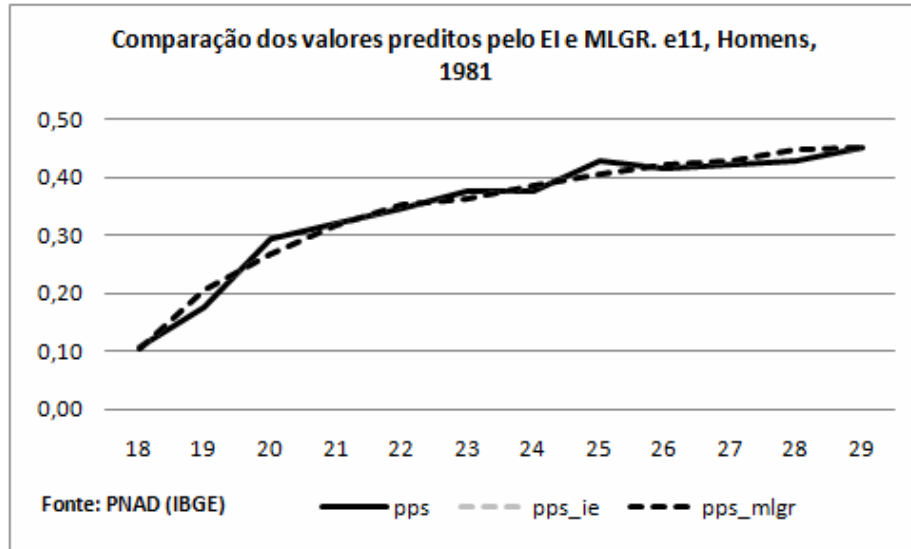
(b) 1992



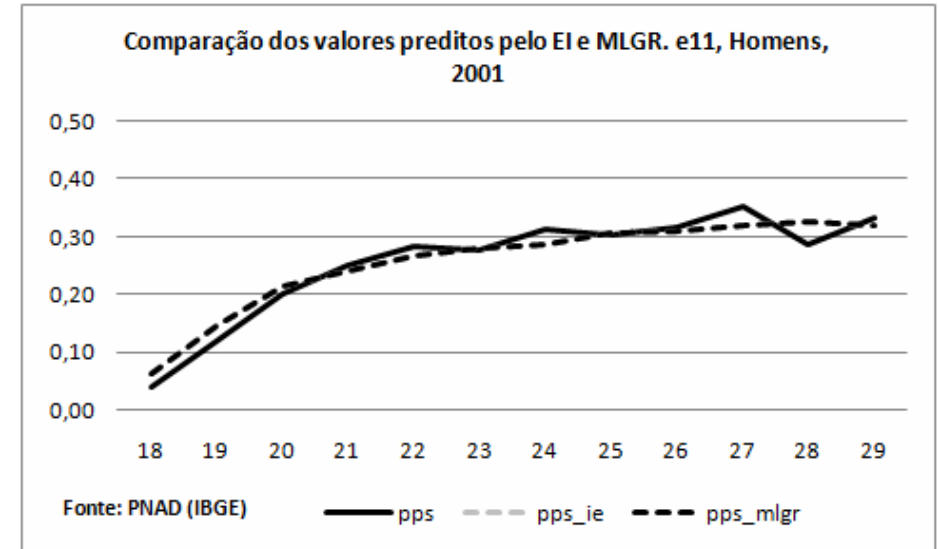
(d) 2008

FIGURA 19

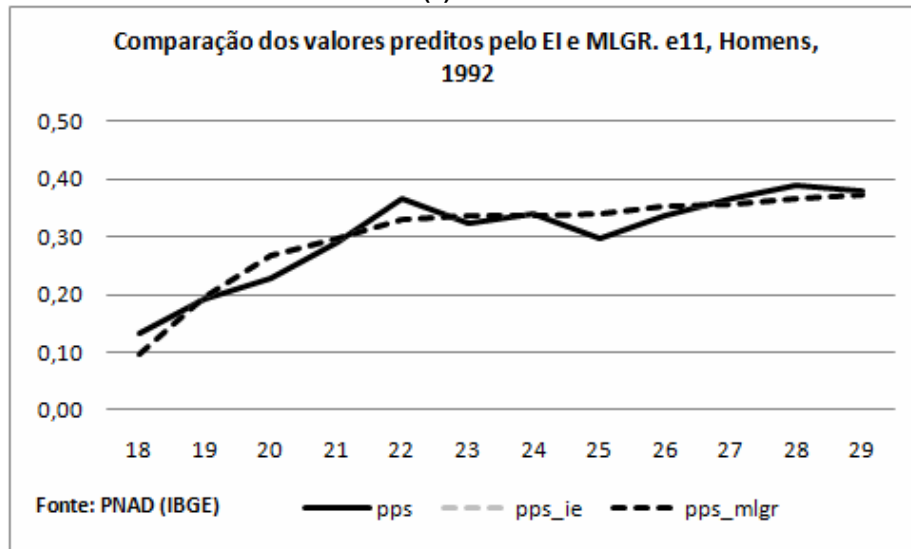
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para o Ensino Superior, e_{11} , Homens



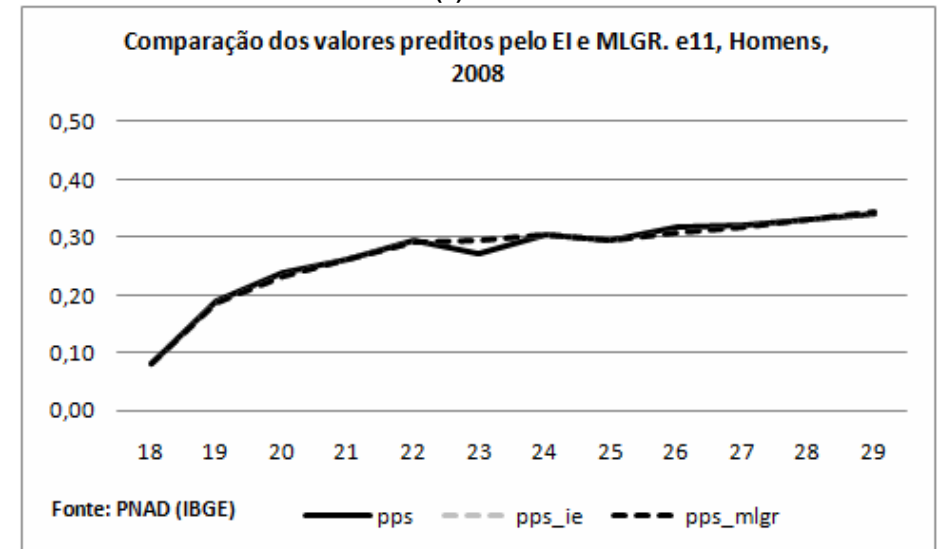
(a) 1981



(c) 2001



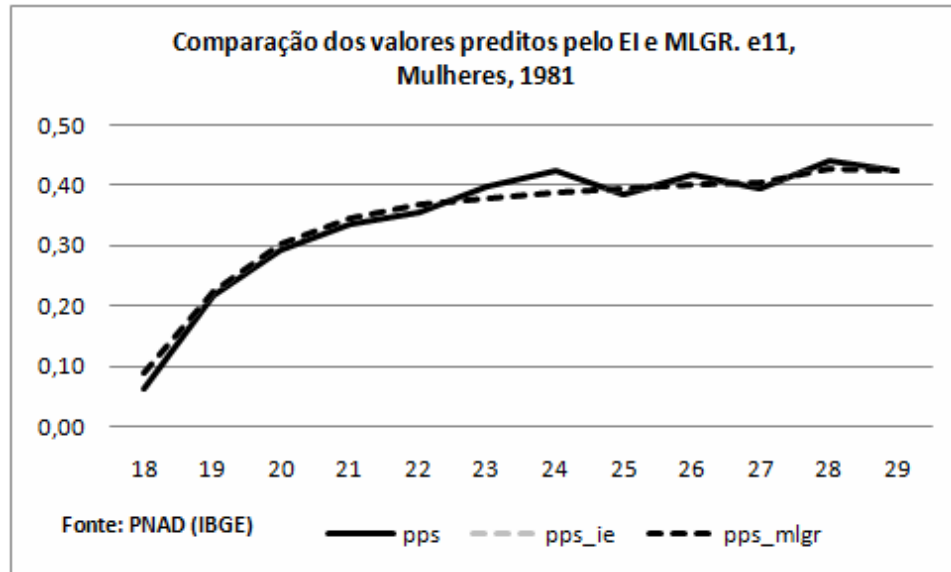
(b) 1992



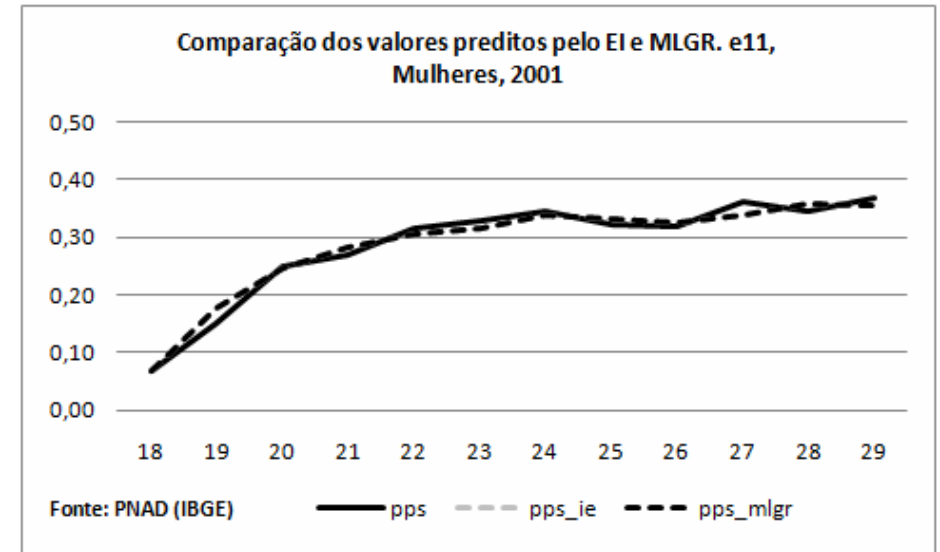
(d) 2008

FIGURA 20

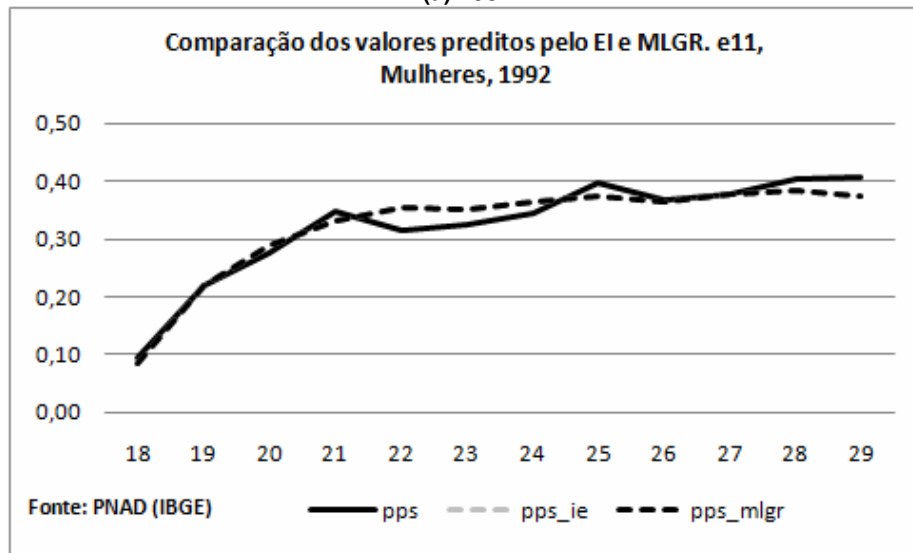
Comparação dos valores preditos dos métodos EI e MLGR para a probabilidade de progressão para o Ensino Superior, e_{11} , Mulheres



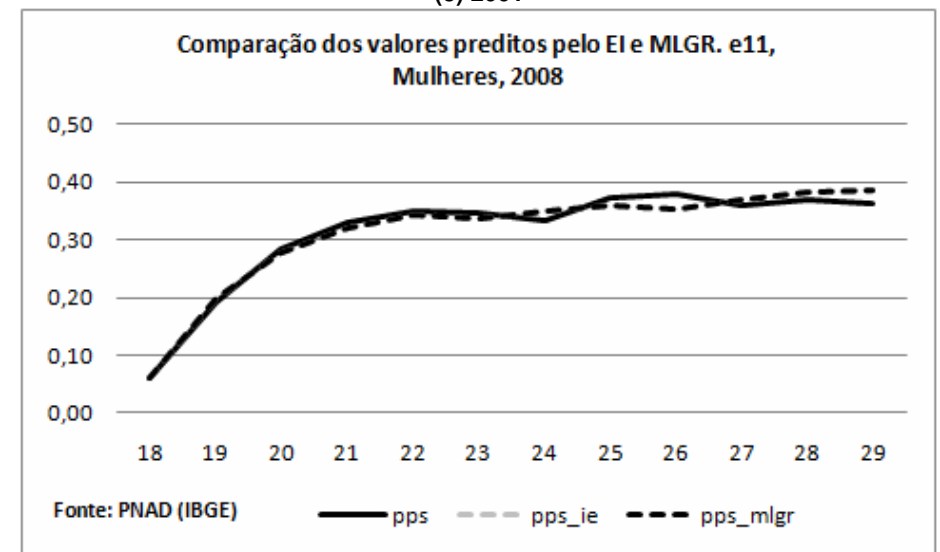
(a) 1981



(c) 2001



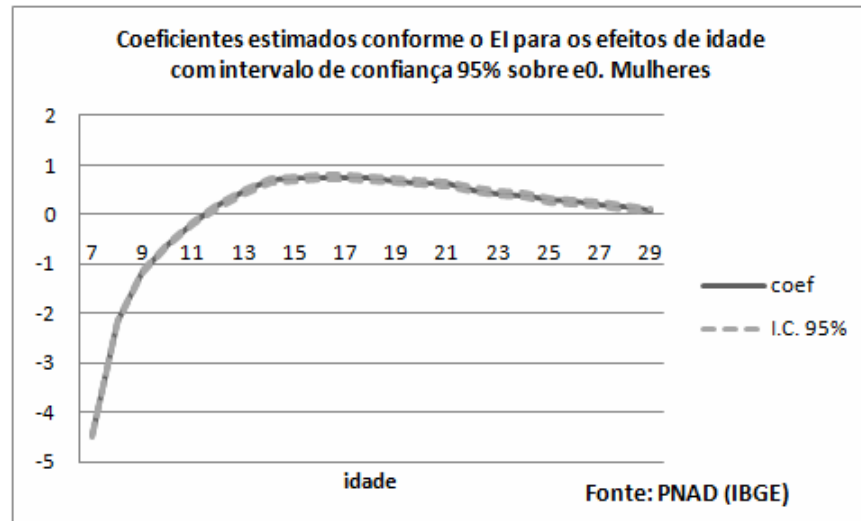
(b) 1992



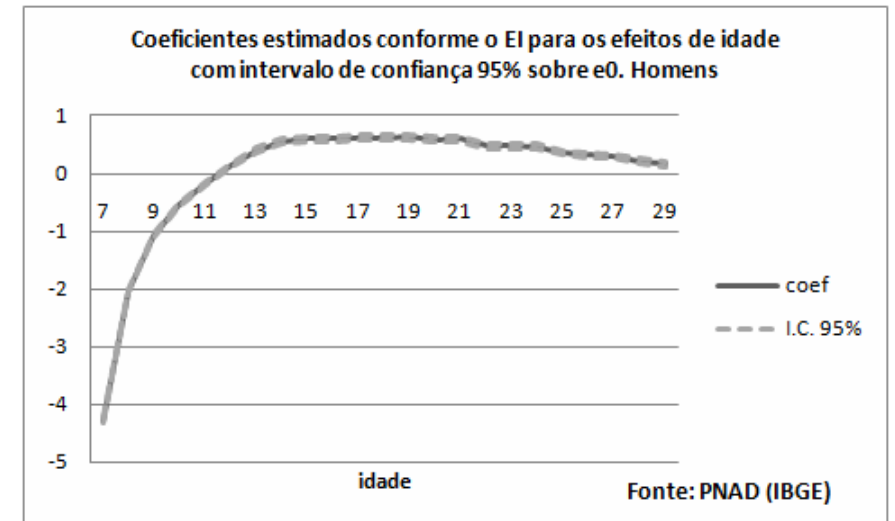
(d) 2008

FIGURA 21

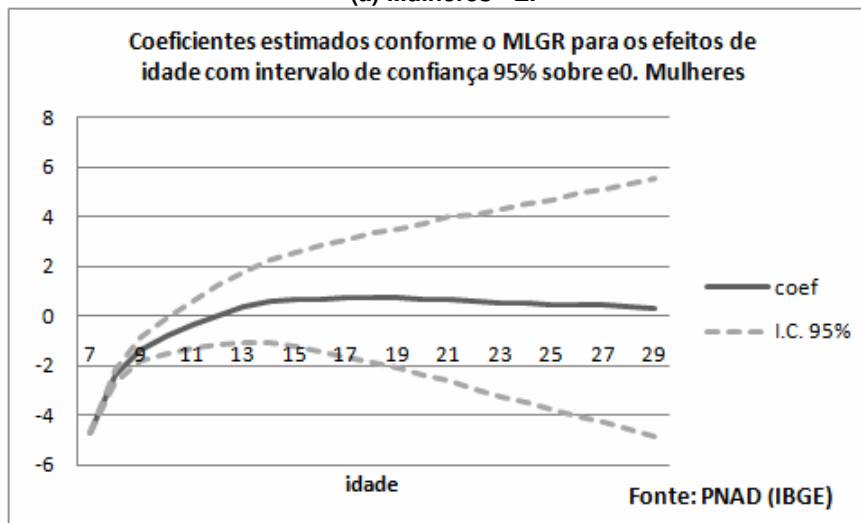
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de idade na probabilidade de progressão para a 1a série



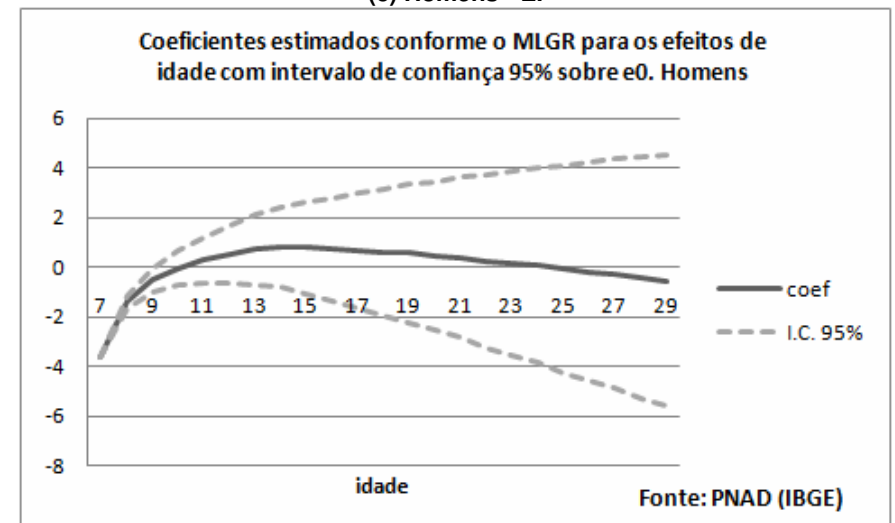
(a) Mulheres - EI



(c) Homens - EI



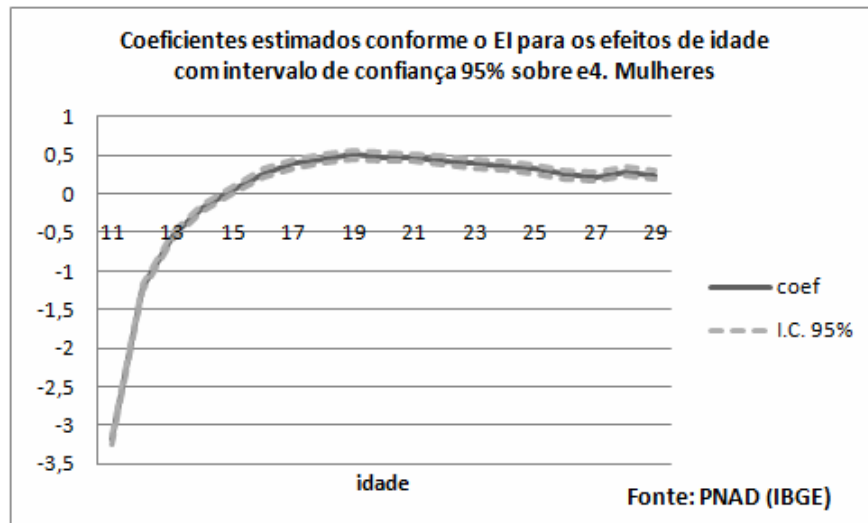
(b) Mulheres - MLGR



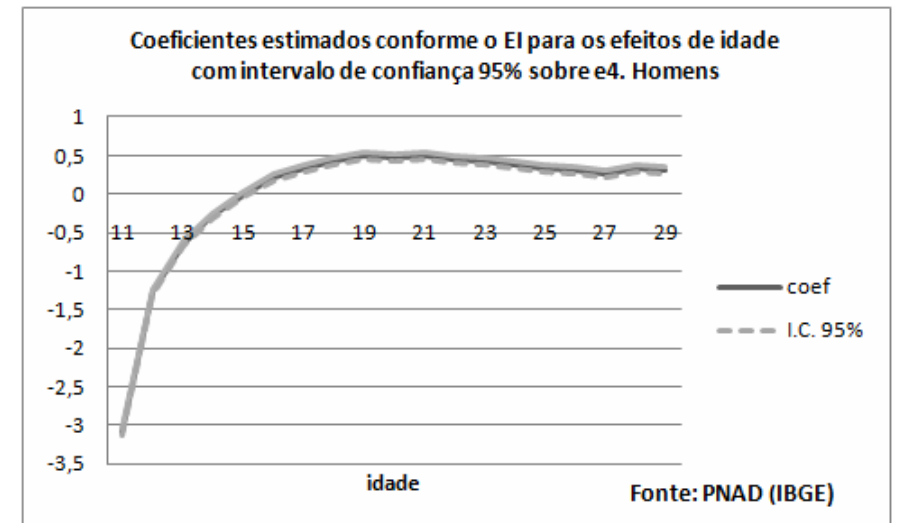
(d) Homens - MLGR

FIGURA 22

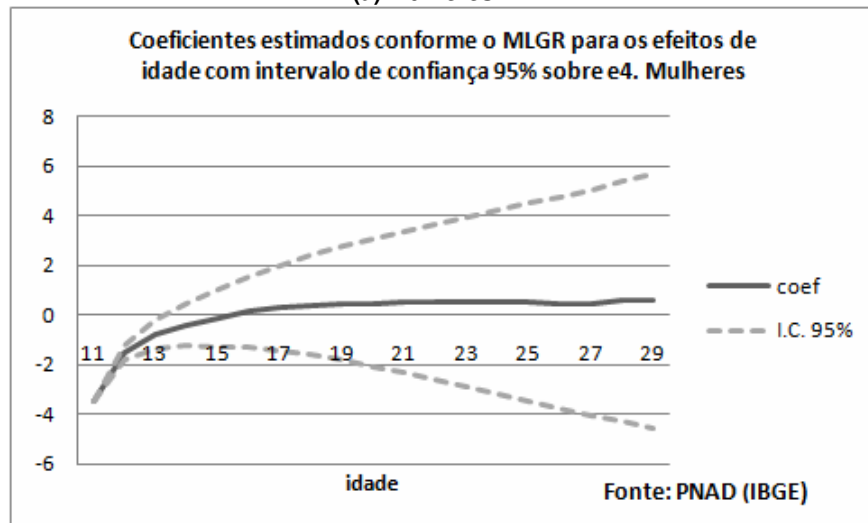
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de idade na probabilidade de progressão para a 5a série



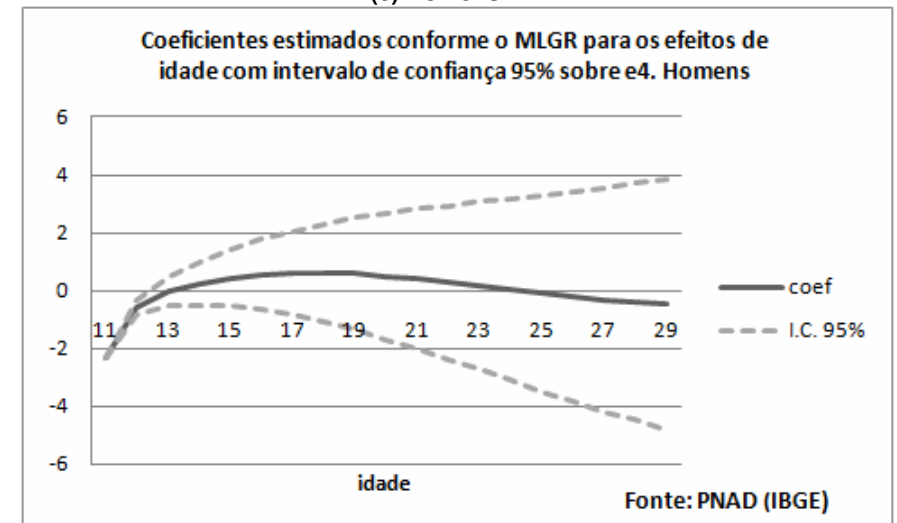
(a) Mulheres - EI



(c) Homens - EI



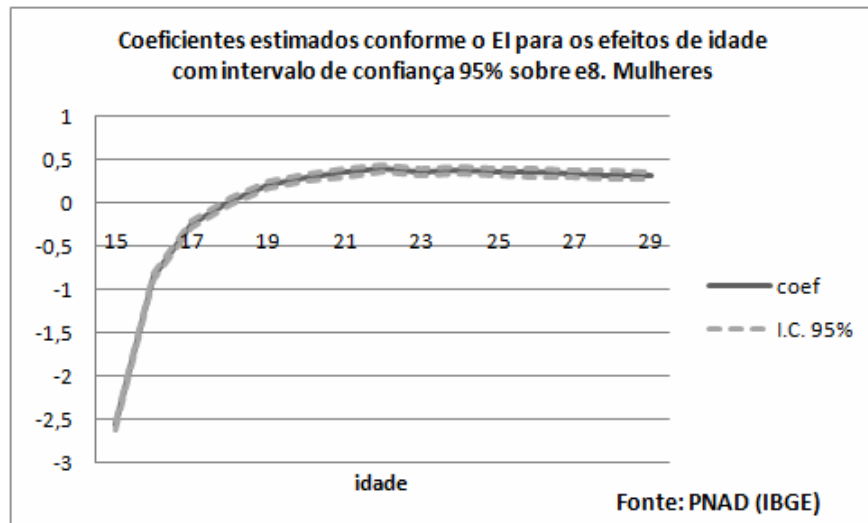
(b) Mulheres - MLGR



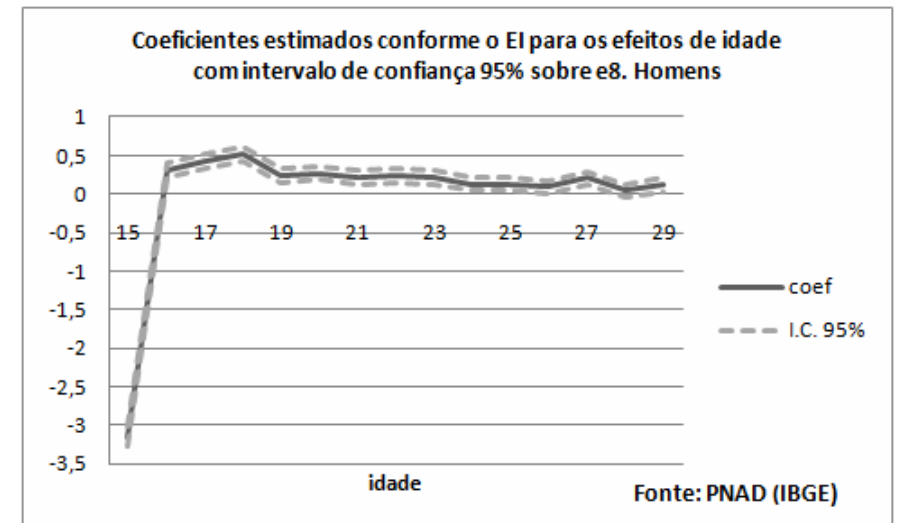
(d) Homens - MLGR

FIGURA 23

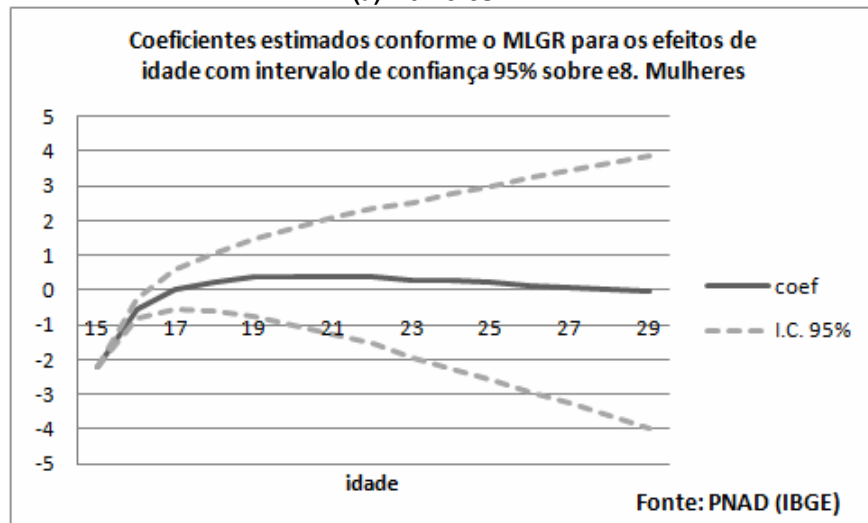
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de idade na probabilidade de progressão para a 1a série do Ensino Médio



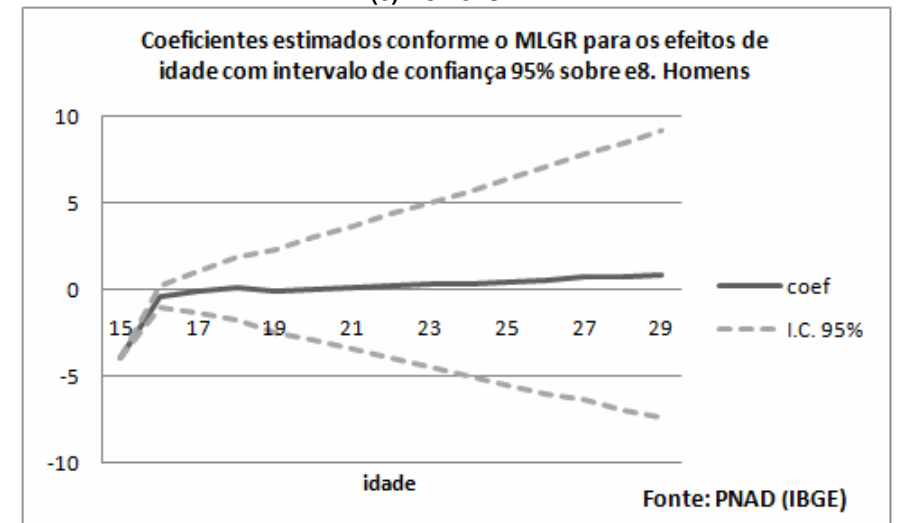
(a) Mulheres - EI



(c) Homens - EI



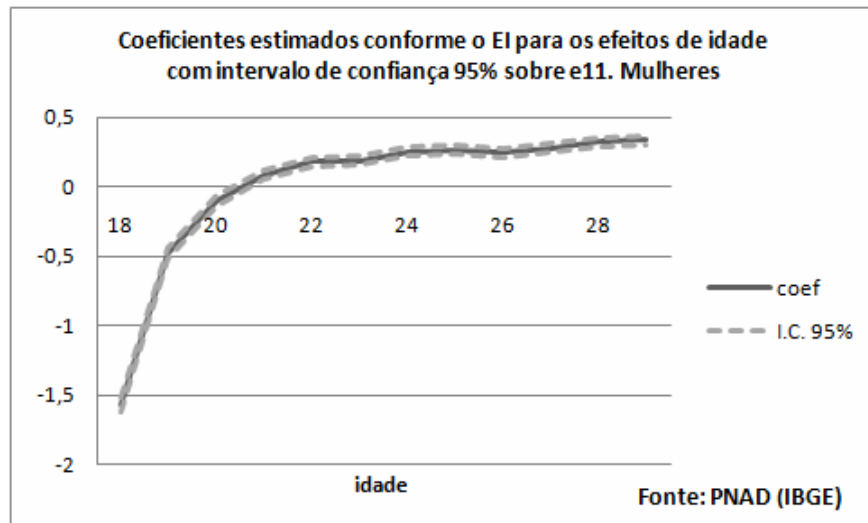
(b) Mulheres - MLGR



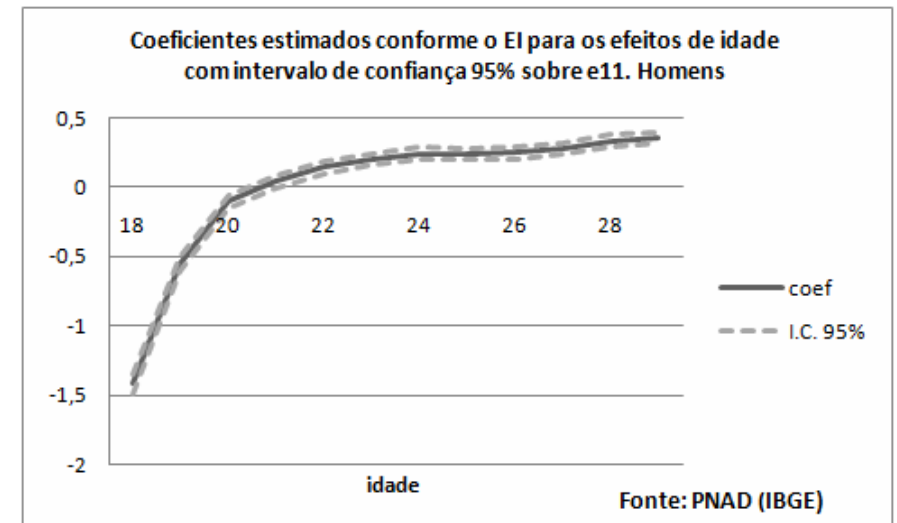
(d) Homens - MLGR

FIGURA 24

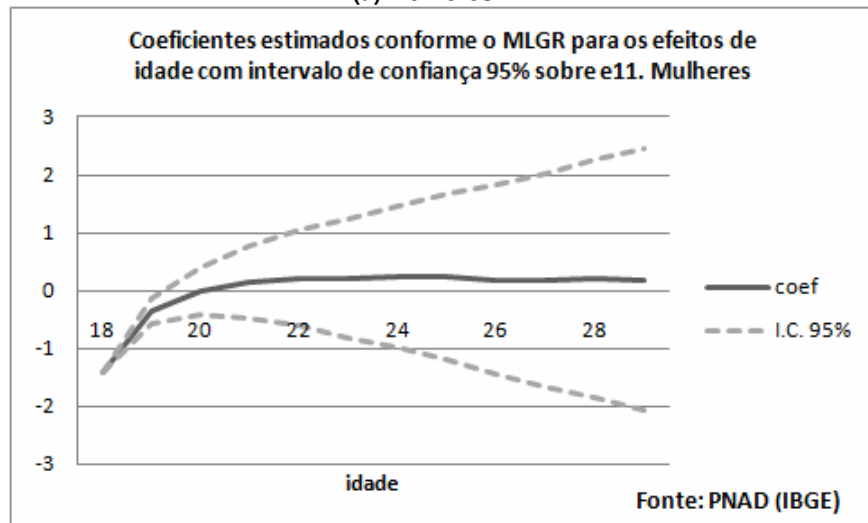
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de idade na probabilidade de progressão para o Ensino Superior



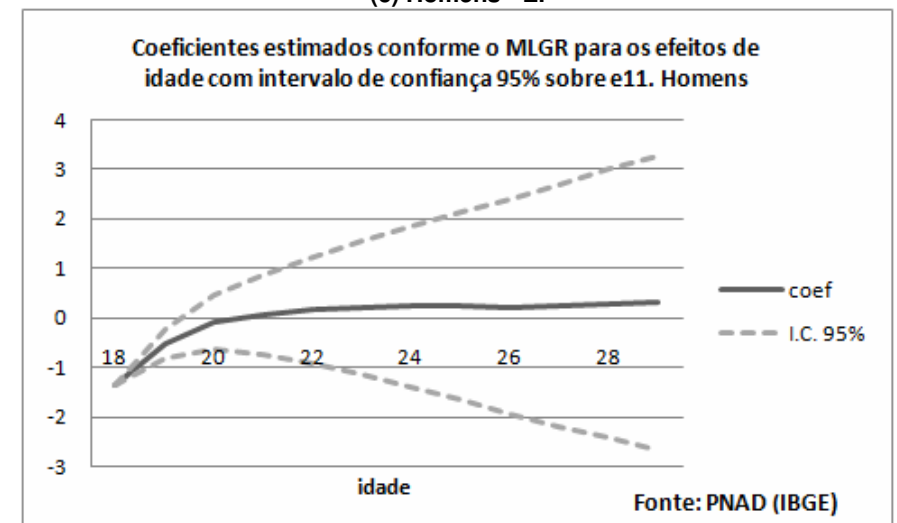
(a) Mulheres - EI



(c) Homens - EI



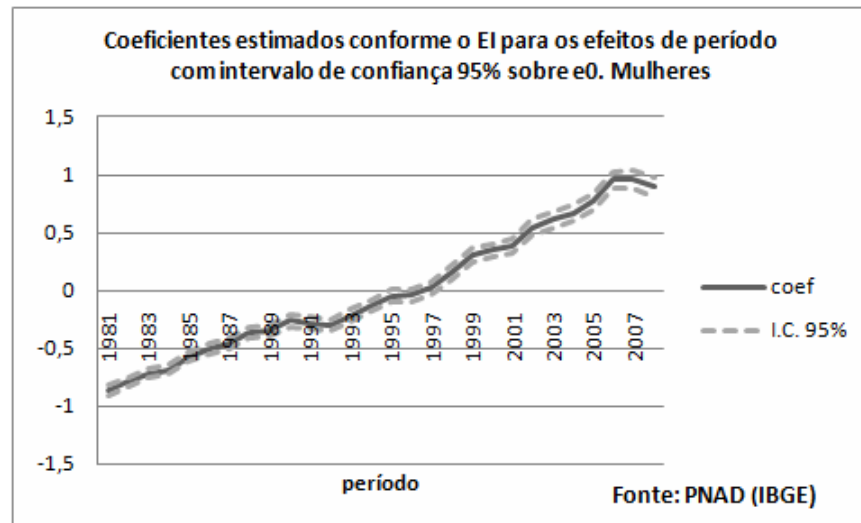
(b) Mulheres - MLGR



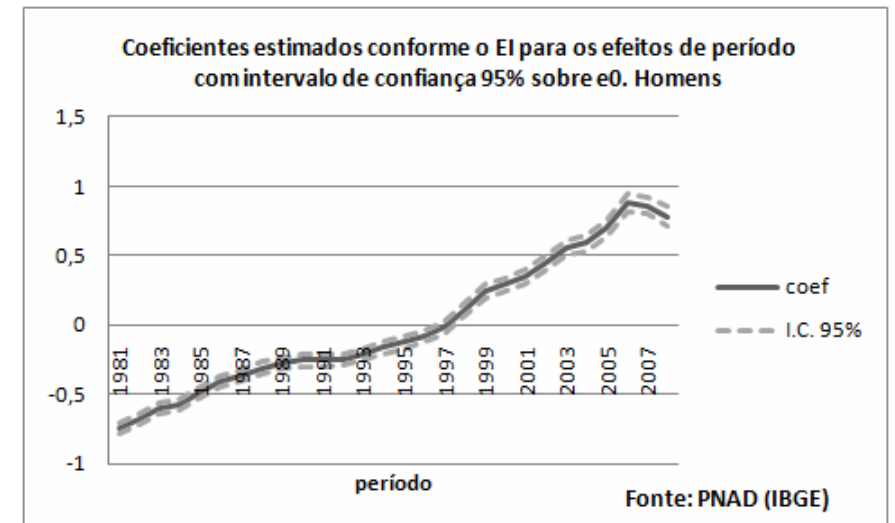
(d) Homens - MLGR

FIGURA 25

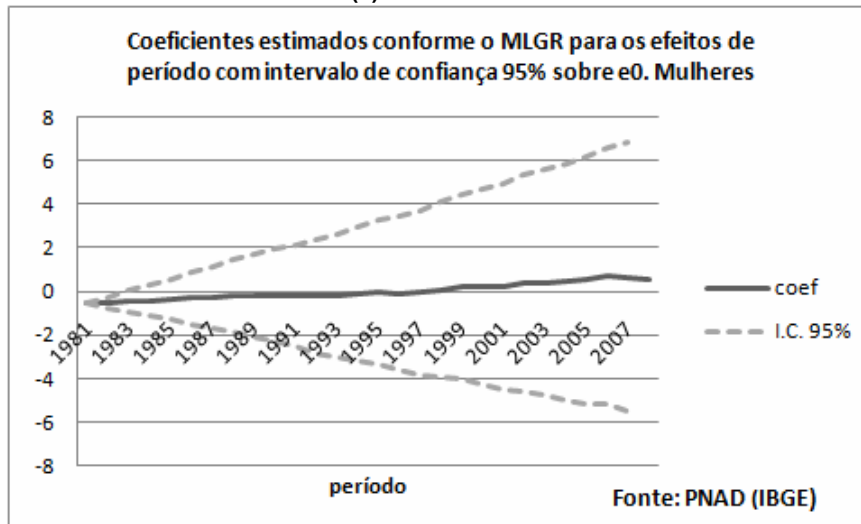
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de período na probabilidade de progressão para a 1a série



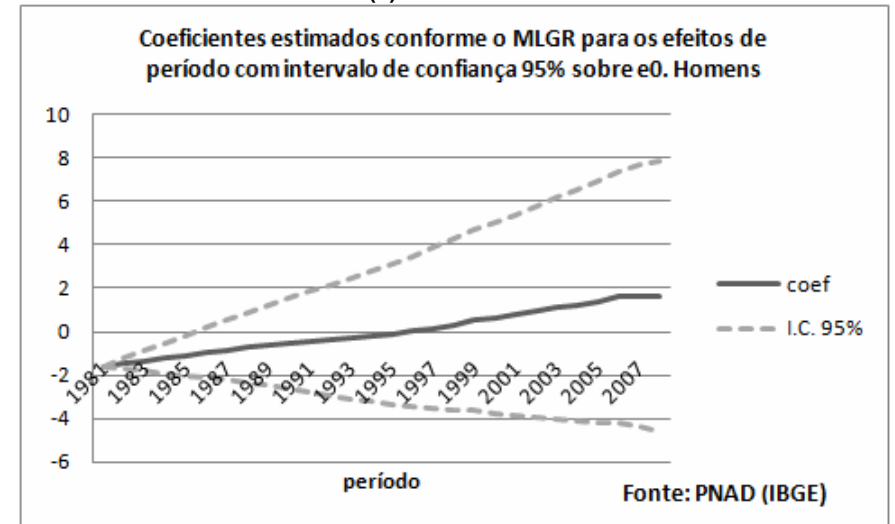
(a) Mulheres - EI



(c) Homens - EI



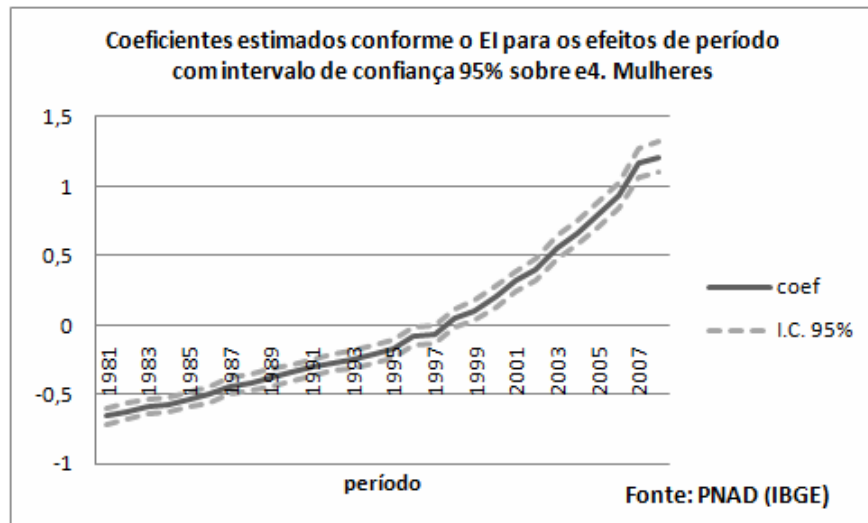
(b) Mulheres - MLGR



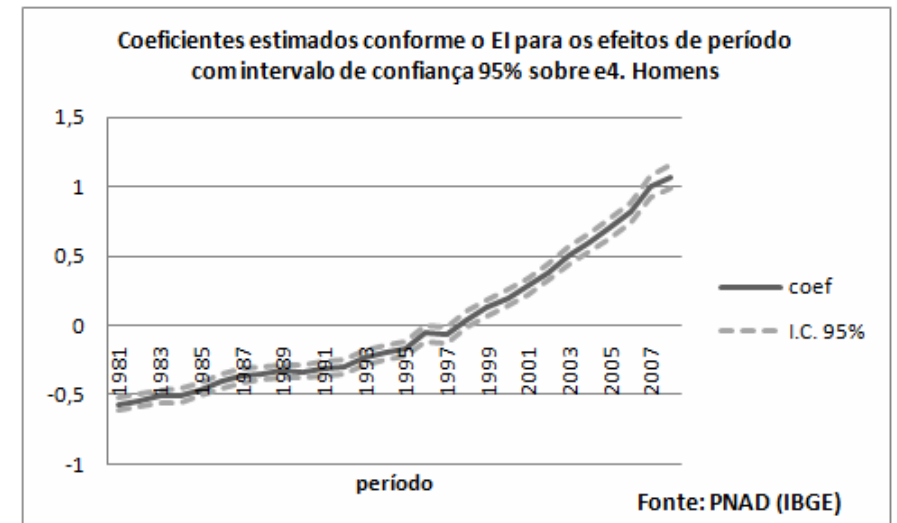
(d) Homens - MLGR

FIGURA 26

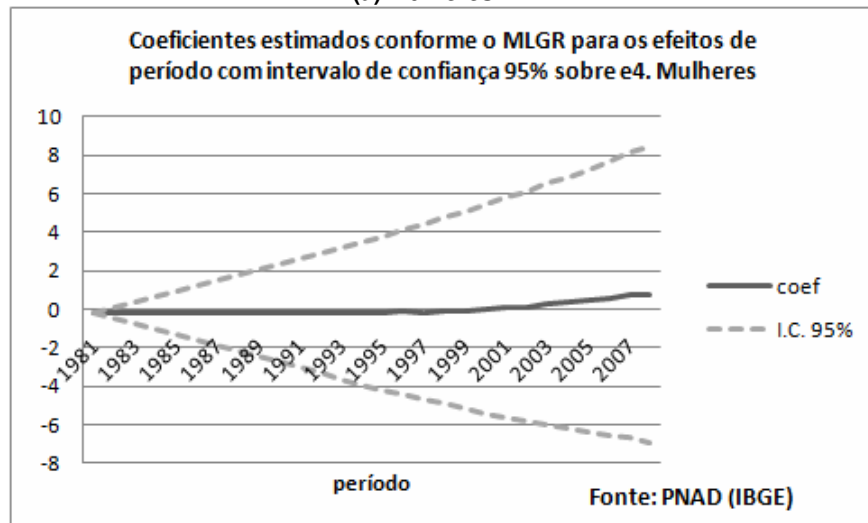
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de período na probabilidade de progressão para a 5a série



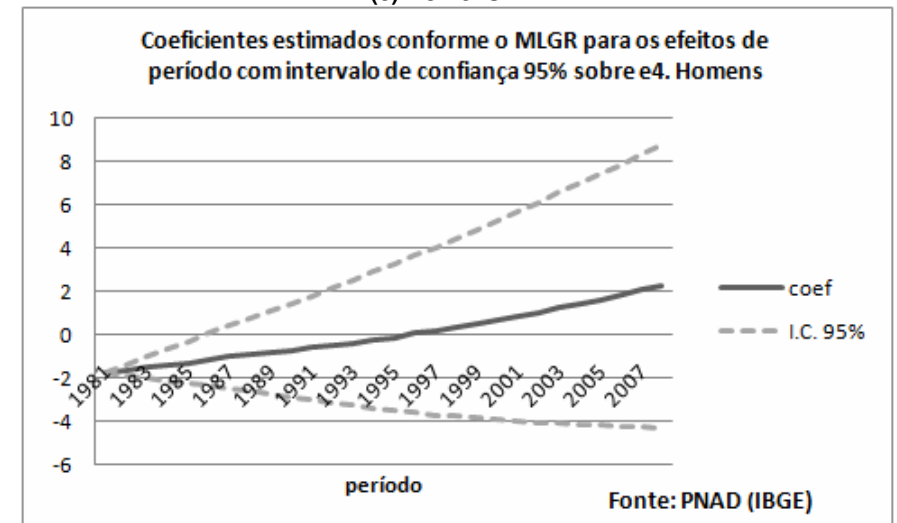
(a) Mulheres - EI



(c) Homens - EI



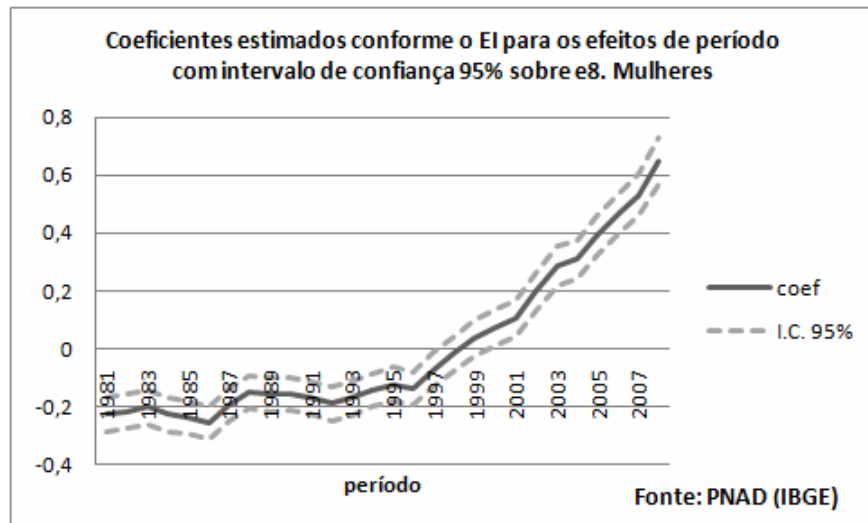
(b) Mulheres - MLGR



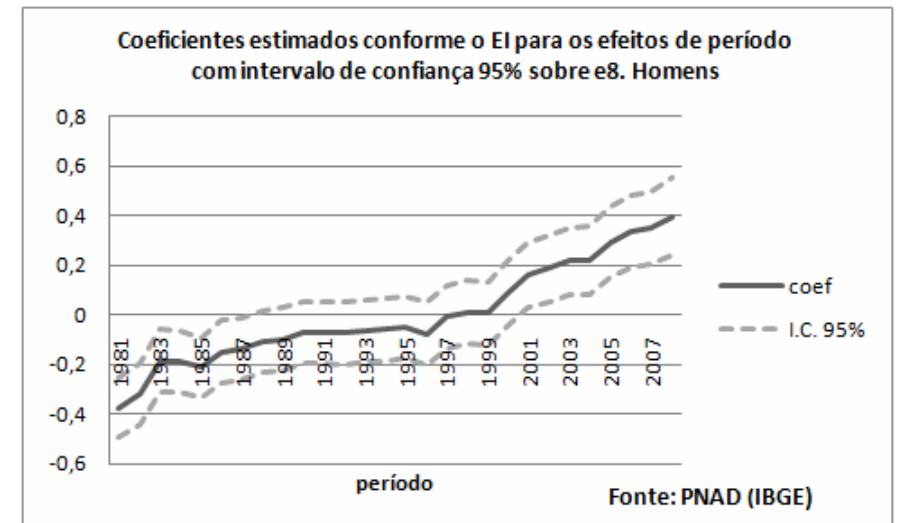
(d) Homens - MLGR

FIGURA 27

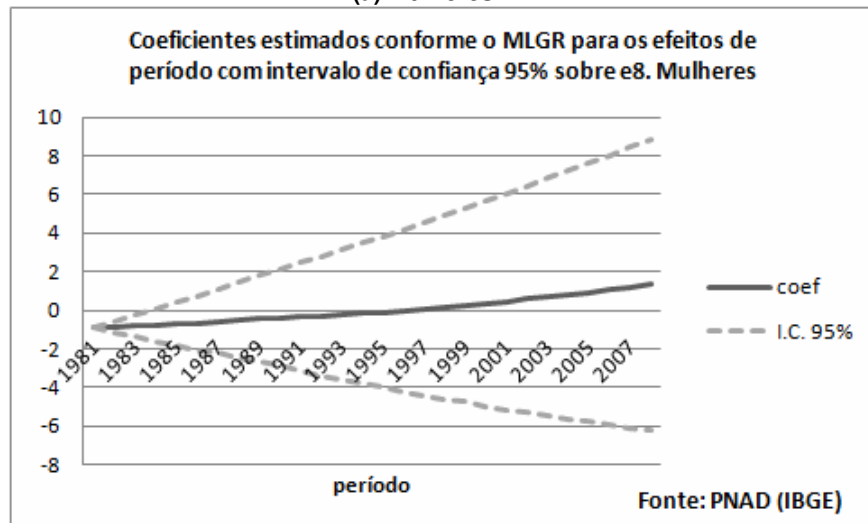
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de período na probabilidade de progressão para a 1a série do Ensino Médio



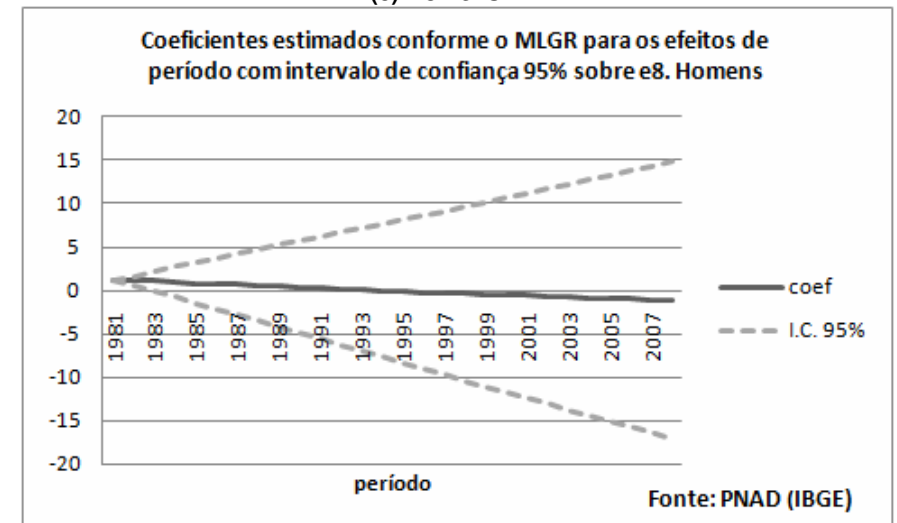
(a) Mulheres - EI



(c) Homens - EI



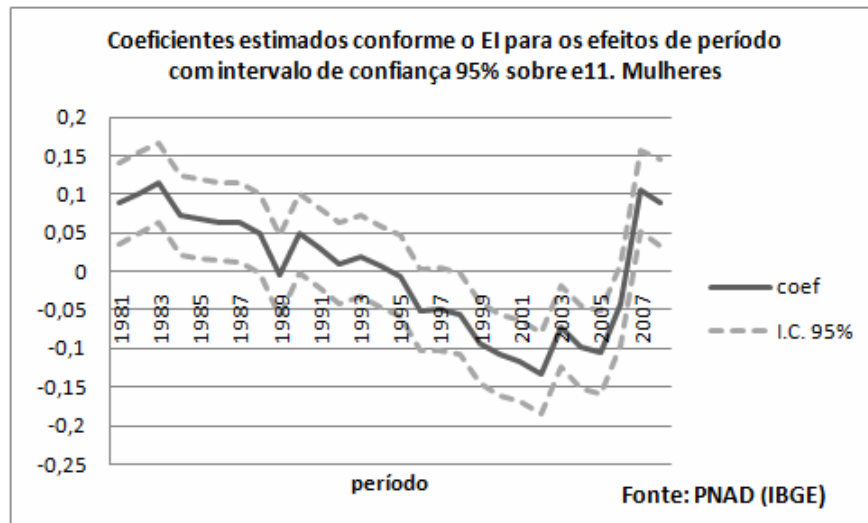
(b) Mulheres - MLGR



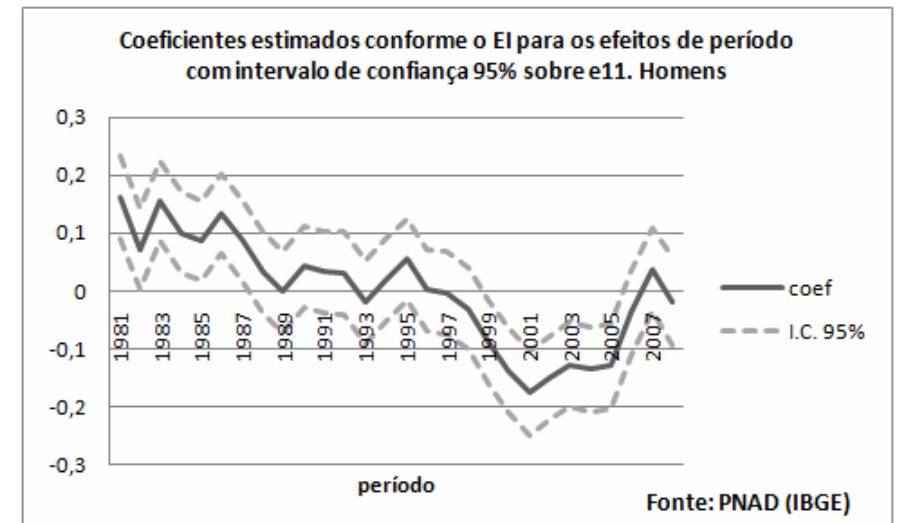
(d) Homens - MLGR

FIGURA 28

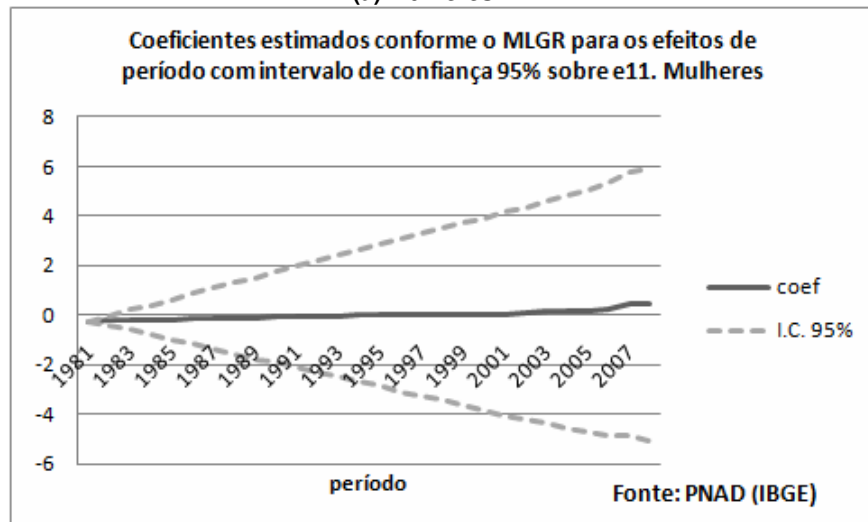
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de período na probabilidade de progressão para o Ensino Superior



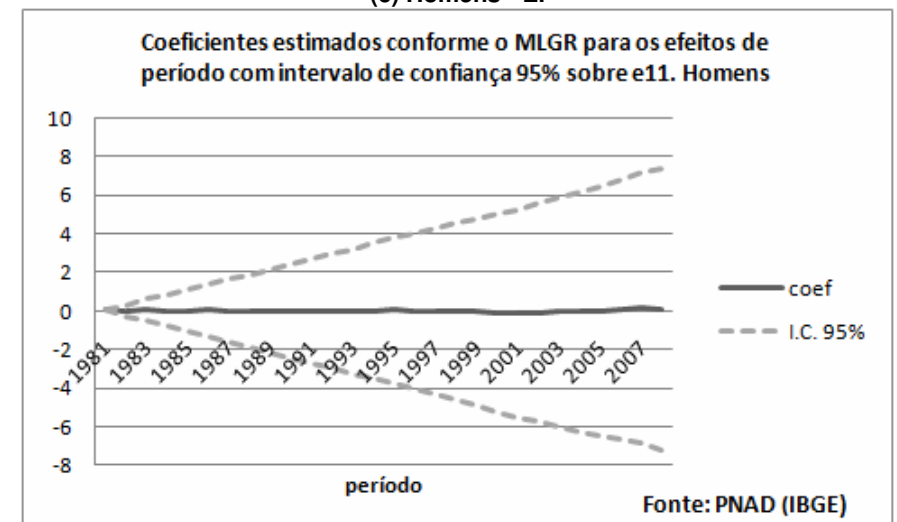
(a) Mulheres - EI



(c) Homens - EI



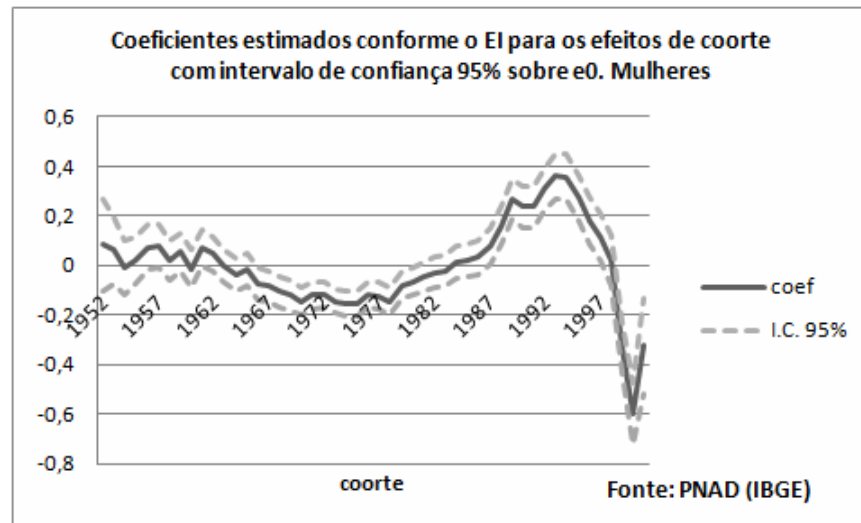
(b) Mulheres - MLGR



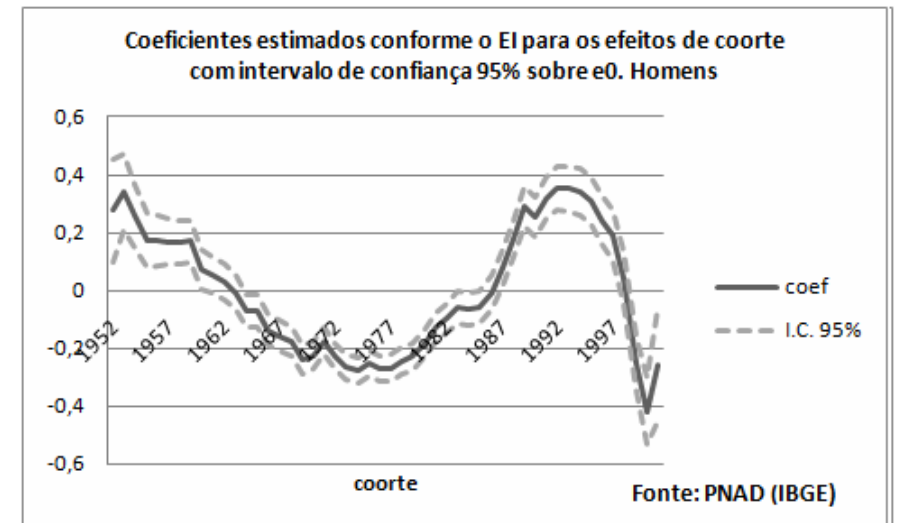
(d) Homens - MLGR

FIGURA 29

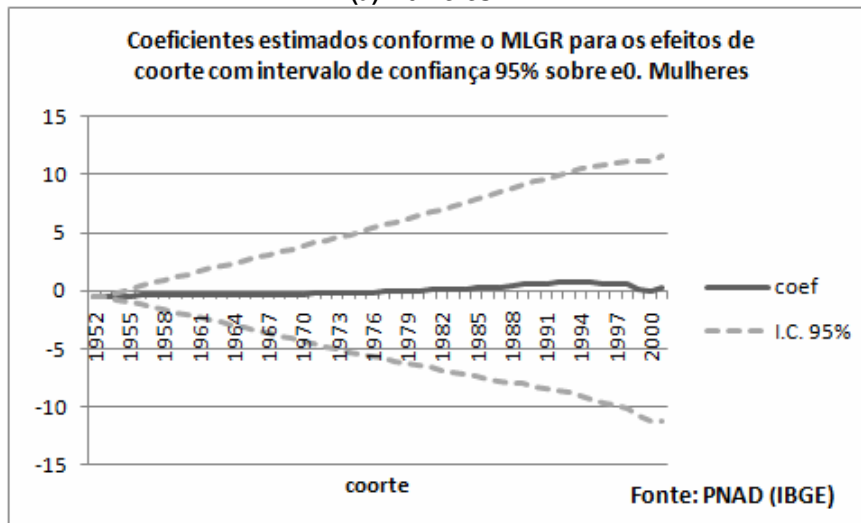
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de coorte na probabilidade de progressão para a 1a série



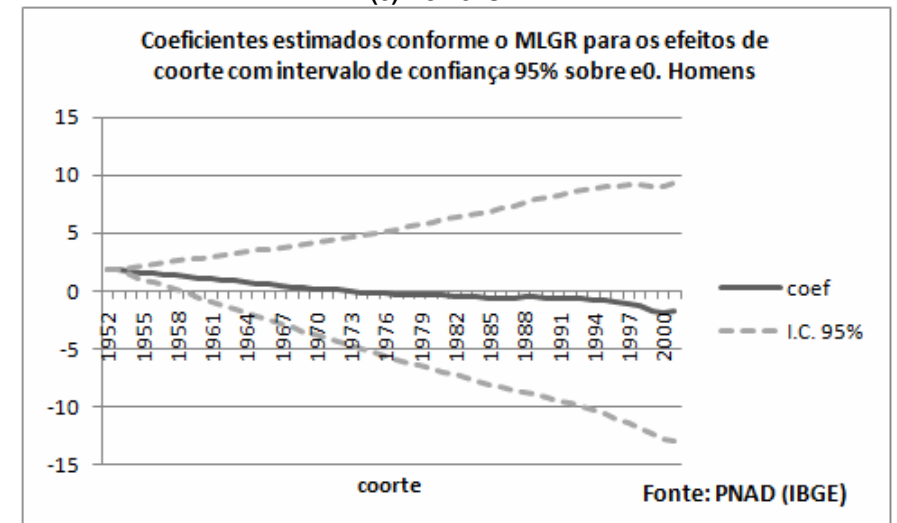
(a) Mulheres - EI



(c) Homens - EI



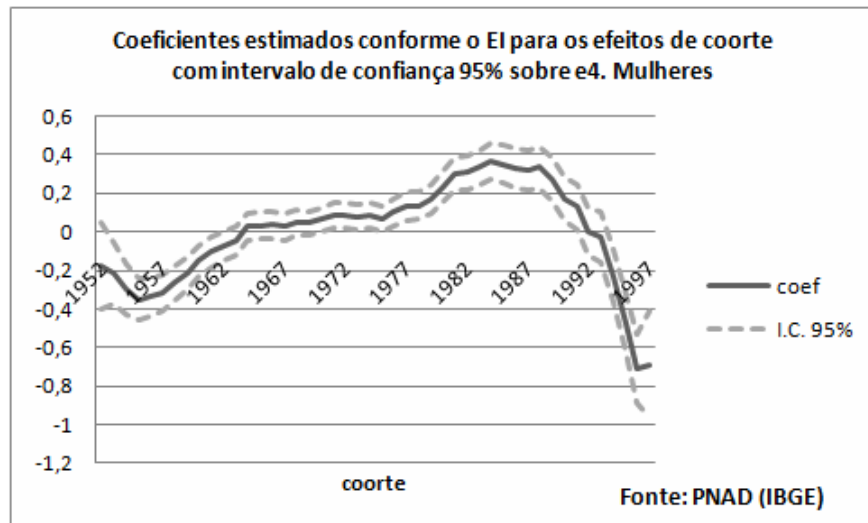
(b) Mulheres - MLGR



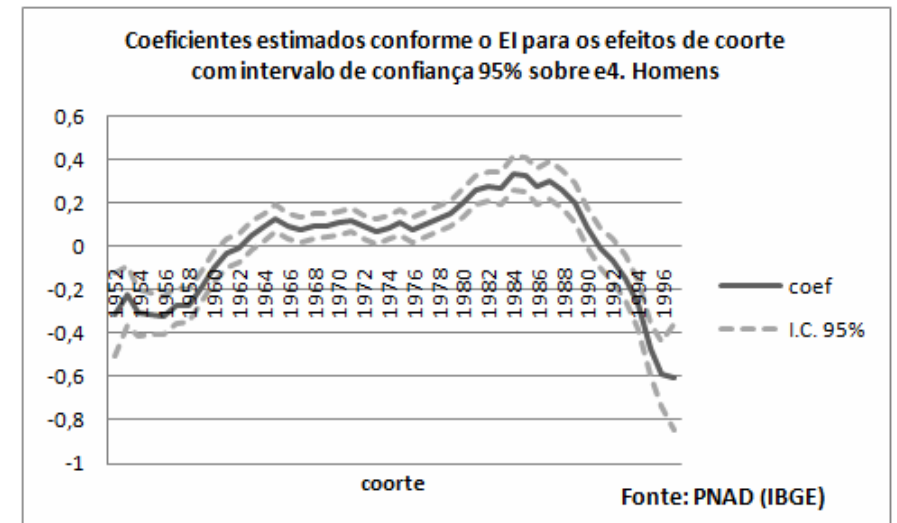
(d) Homens - MLGR

FIGURA 30

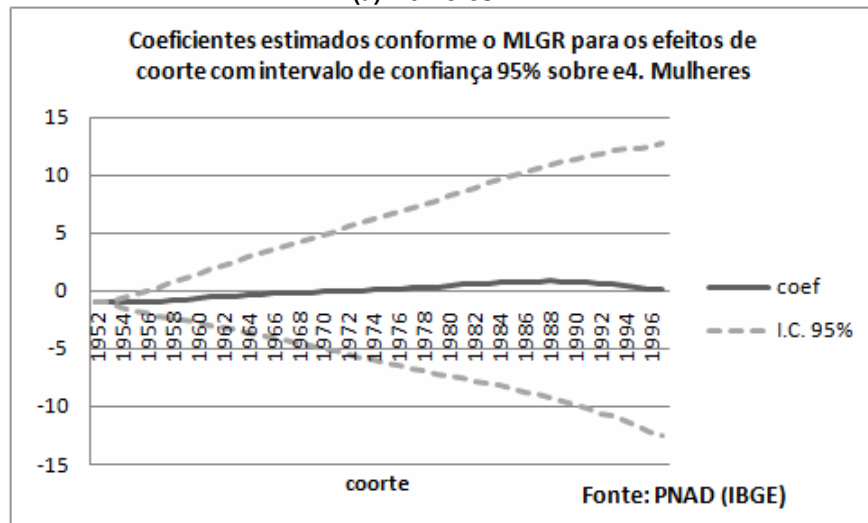
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de coorte na probabilidade de progressão para a 5a série



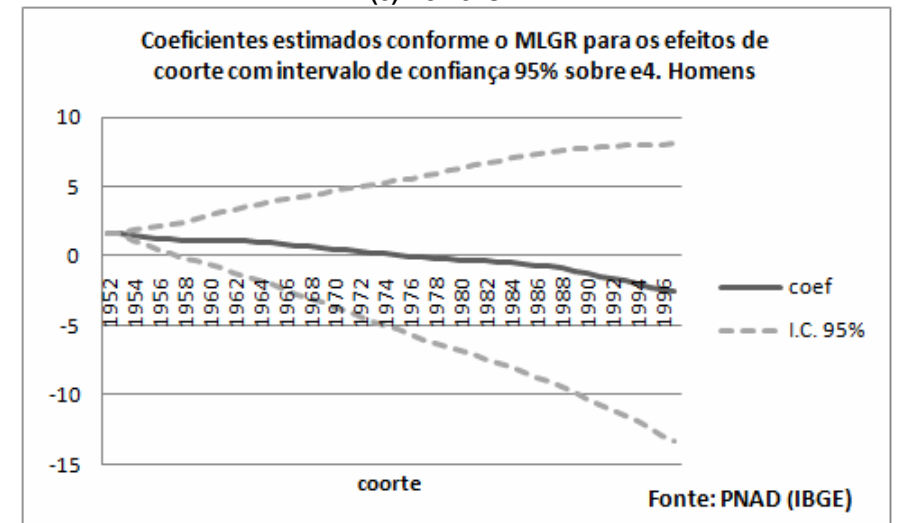
(a) Mulheres - EI



(c) Homens - EI



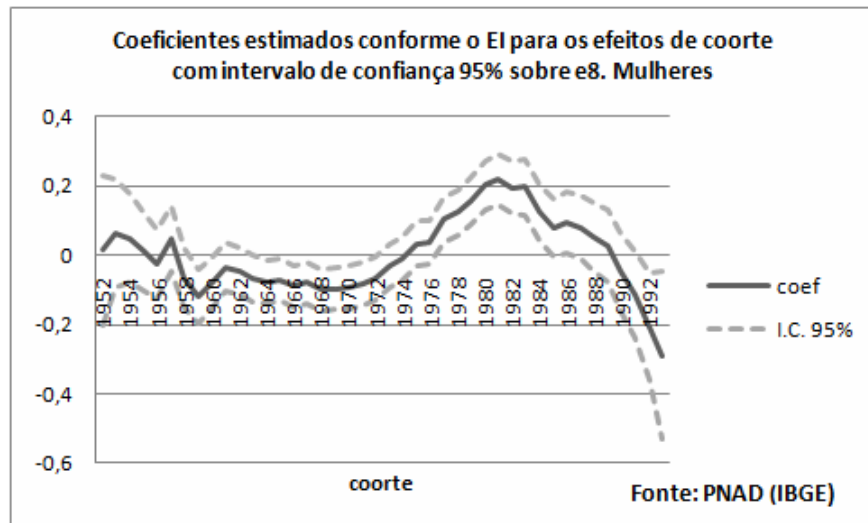
(b) Mulheres - MLGR



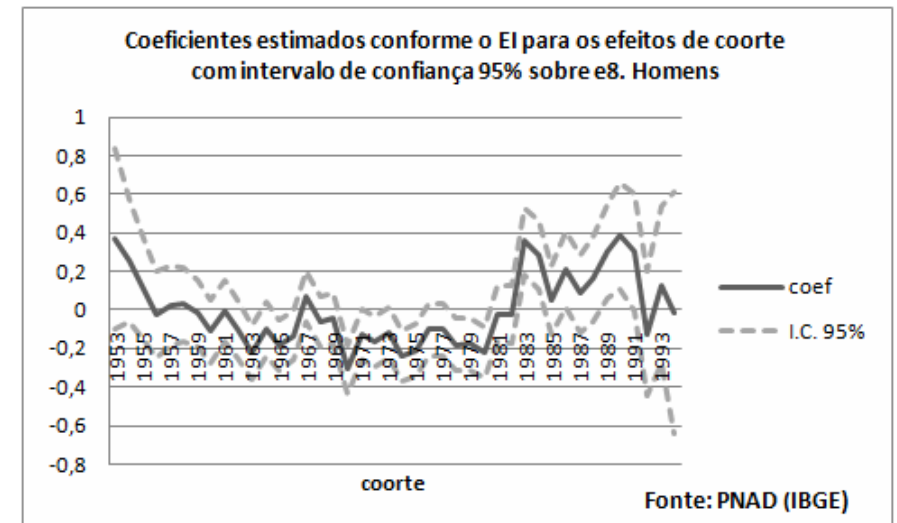
(d) Homens - MLGR

FIGURA 31

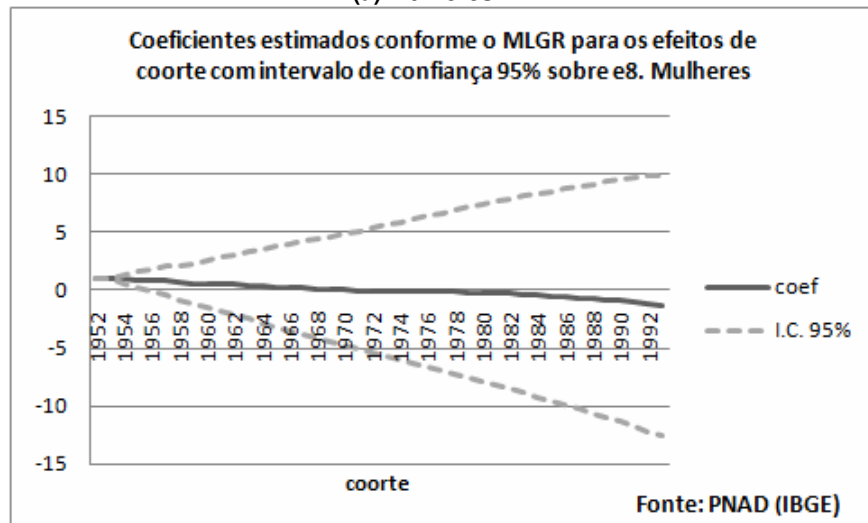
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de coorte na probabilidade de progressão para a 1a série do Ensino Médio



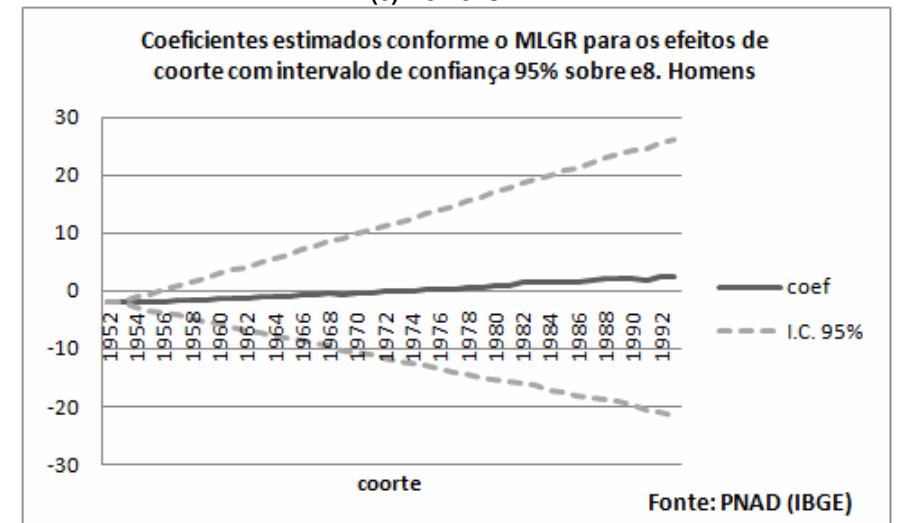
(a) Mulheres - EI



(c) Homens - EI



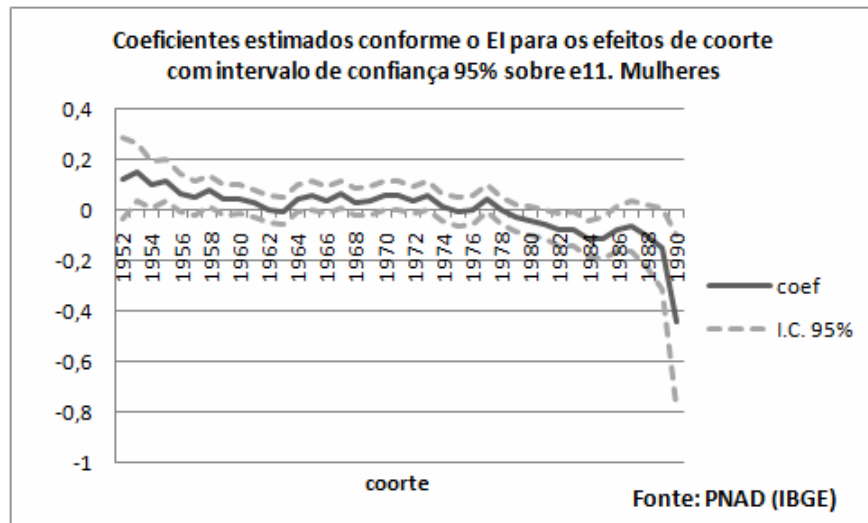
(b) Mulheres - MLGR



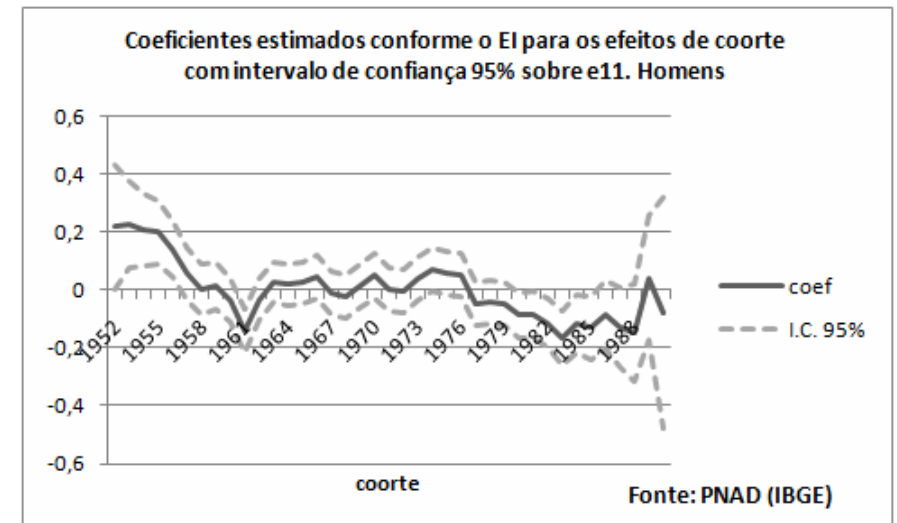
(d) Homens - MLGR

FIGURA 32

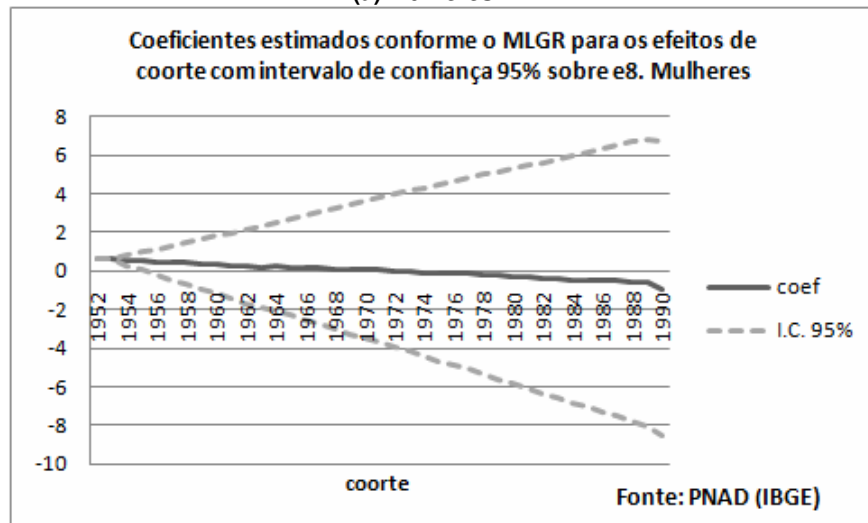
Comparação entre os coeficientes estimados pelo EI e MLGR para os efeitos de coorte na probabilidade de progressão para o Ensino Superior



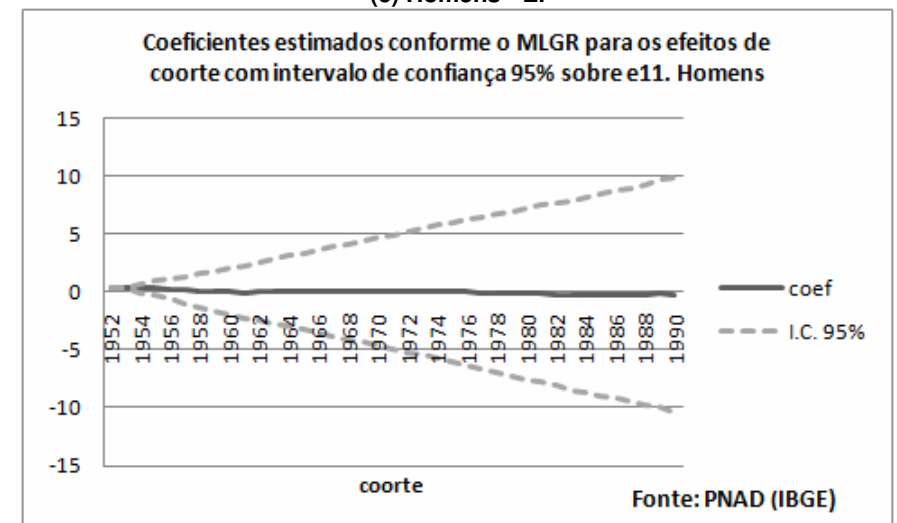
(a) Mulheres - EI



(c) Homens - EI



(b) Mulheres - MLGR



(d) Homens - MLGR

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. (Ed.). *Cohort analysis in social research*. (S.l.): 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, out 1976.
- HALLI, S. S.; RAO, K. V. *Advanced techniques of population analysis*. (S.l.): 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. (Ed.). *Cohort analysis in social research*. (S.l.): Springer Verlag, 1985.
- HOSMER, D.; LEMESHOW, S. *Applied logistic regression*. (S.l.): 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. (Ed.). *Cohort analysis in social research*. (S.l.): 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, 2002.
- 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.l.): 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. (Ed.). *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.
- 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.
- SCHULHOFER-WOHL, S.; YANG, Y. APC: Stata module for estimating age-period-cohort effects. ago. 2006. Statistical Software Components, Boston College Department of Economics. Disponível em: <<http://ideas.repec.org/c/boc/bocode/s456754.html>>.
- SMITH, H. Response: cohort analysis redux. *Sociological Methodology*, v. 34, p. 111–119, 2004.
- STATA CORP. *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.