Parecer: **A relação de fatores individuais, familiares e escolares com a distorção idade-serie no ensino público brasileiro**.

[Parecerista B] O artigo aborda um tema bastante discutido na literatura empírica brasileira que é a defasagem idade série e suas inter-relações com as variáveis individuais, familiares e escolares. Apesar disto, acho que o artigo pode ter uma contribuição à literatura no que tange à metodologia, trazendo alguns resultados interessantes. Neste sentido, considero a possibilidade do artigo ser aprovado desde que se façam mudanças substanciais que listo abaixo.

Revisão português: muitas palavras repetidas no mesmo parágrafo, problemas de concordância verbal.

Com relação à definição de defasagem idade-série, destacam na introdução que se for igual a zero é o ideal. E se for menor que zero? Quando o estudante estiver avançado, como poderemos classificá-lo?

*[Autores] Em nossa opinião, e tendo em vista o objetivo do artigo, que é de analisar a defasagem idade-escolaridade como um problema passível de políticas públicas, deve-se apenas levar em conta os casos em que a defasagem idade-série é zero ou positiva. Isso porque não faz sentido focar nos estudantes que estariam adiantados, pois estes não são caracterizados como um problema de política pública. Desta forma, sua defasagem foi transformada em zero. Realmente, em uma comparação entre adiantados e aqueles com distorção zero é difícil identificar qual seria a situação “ideal”. Em vista disto, substituiu-se a palavra “ideal” por “recomendada” no texto.*

[P] Na introdução tem uma vasta revisão bibliográfica sobre estudos empíricos feitos para o Brasil. Apesar de vasta, acredito que esta revisão está pouco explorada. Seria interessante que esta revisão destacasse melhor as bases e metodologias utilizadas e que fosse construída a partir de alguma sequência lógica, seja temporalmente ou metodologicamente. Os artigos estão agrupados em dois grupos, mas são analisados separadamente. Seria também interessante trazer nesta revisão empírica alguma análise do que tem sido feito internacionalmente.

*[A] Esse comentário foi incorporado e foi dado destaque para questões metodológicas e para as bases utilizadas nos diversos estudos mencionados. Embora se mantenha a divisão em grupos, a sequência em que os artigos são apresentados em seus respectivos grupos foi alterada para atender a um critério temporal, para o primeiro grupo, e um critério relacionado aos resultados encontrados, no caso do segundo grupo.*

*Quanto a literatura internacional, é importante notar que quando a variável dependente é a proficiência, o tema perpassa os mais diversos países. Entretanto, quando a variável é a distorção idade-série, o problema é muito mais focalizado em alguns países, não sendo muito preponderante nos países desenvolvidos, por exemplo, e desta forma não há muitos artigos sobre este tema, especialmente em se tratando de artigos seminais. Desta forma, a revisão focou-se apenas na literatura brasileira, que já é extensa e relevante.*

[P] Acredito que na introdução o autor poderia apenas destacar o que traz de diferente em relação à bibliografia pesquisada e como se insere nesta literatura. Toda a análise bibliográfica deveria compor uma nova seção de revisão da literatura.

*[A] Essa sugestão do parecerista foi incorporada ao texto, na forma de uma nova seção e algumas alterações na introdução.*

[P] Na página 6, segundo parágrafo, o autor fala que pretende contribuir para a literatura mostrando a influência das características individuais e familiares e do ambiente escolar sobre a defasagem escolar...como será esta contribuição? Pela revisão da literatura, há uma grande quantidade de artigos empíricos que estudam esta relação. Porque a metodologia utilizada é inovadora e de que forma os resultados anteriores podem ser comparados com seus resultados? Qual a intuição para existir diferenças entre os resultados?

*[A] Os autores concordam que não havia ficado clara a contribuição e neste sentido, alterações foram feitas neste trecho do texto, agora presente na introdução do artigo nas páginas 2 e 3.*

[P] Ao longo da década de 90 e anos 2000, existiu um amplo debate sobre políticas de promoção automática. Estas políticas foram aplicadas de forma diferenciada entre municípios/estados e ao longo dos anos. Algumas foram continuadas e outras não. Este debate surgiu a partir da preocupação de que altos índices de reprovação desincentivariam a permanência na escola. Em nenhum momento, o autor cita este debate e, está no cerne do aumento dos estudos sobre defasagem. Desta forma, acho que esta discussão justifica a análise do problema e deve ser inserida na introdução.

*[A] Um parágrafo tratando da discussão sobre as políticas de promoção automática foi incorporado ao texto, na seção de revisão da literatura, página 8.*

[P] Com relação à metodologia, a amostra escolhida é formada por crianças entre 8 e 17 anos de idade. Primeiro, porque o recorte final foi em 17 anos de idade? Existem muitos jovens no Brasil com mais de 17 anos que cursam o nível médio. Outra observação, uma criança de 6 ou 7 anos de idade que não está frequentando a escola já não pode ser considerada defasada? Retirando da amostra estas duas faixas etárias, acredito que o fenômeno de defasagem idade série está subestimado, sobretudo se há um grande problema na idade de entrada na escola. Como também está restringindo o universo aos estudantes (os que estão matriculados), a defasagem idade série é subestimada, sobretudo para as idades acima de 14 anos...Acredito que o maior problema para as faixas etárias maiores seja o abandono e não somente a repetência. A repetência na verdade estimula o abandono. Parte do problema está sendo descartado nesse caso. Porque retiram os estudantes que não frequentam? Se for pelo fato de olhar estudantes na rede pública, aponto outro tipo de problema que me remete ao título do trabalho. Como a PNAD capta informação da rede de ensino da escola que o aluno frequentou ou frequenta, não capta bem o histórico do aluno. O aluno, sobretudo, nos anos iniciais do ensino fundamental, pode frequentar a escola pública... contudo ao longo dos anos pode ter mudado para escola particular e também retornado a escola pública. Neste sentido, o seu objeto de estudo não é a distorção no idade série no ensino público. Alunos que evadiram do ensino público e evadiram do sistema escolar, por exemplo, não estão sendo captados pela sua amostra de dados. Ademais, um aluno que estudou na escola particular, repetiu na escola particular e apenas em 2013 foi para rede pública, está defasado não por conta da escola pública... mas sua trajetória foi de escola particular. Enfim, só dei alguns exemplos para ponderar o título com a amostra escolhida e pontuar que o problema tratado não está sendo a existência de defasagem escolar no ensino público brasileiro, mas de uma parcela de estudantes que frequenta ensino público no ano de 2013.

*[A] Conforme discutido no texto, foi feito o recorte aos 17 anos pois a partir desta idade e dada o método de cálculo da distorção (Idade – 7 anos – anos de escolaridade), todos os alunos de 18 anos matriculados estariam defasados, o que não é interessante metodologicamente. Isso ocorre pois, como se está trabalhando apenas com alunos matriculados, o nível mais alto de matricula é o terceiro ano do ensino médio e, desta forma, seu maior nível de escolaridade seria aquele obtido quando completa o segundo ano do ensino médio, totalizando 10 anos de estudo. Assim, todos os alunos acima desta idade teriam necessariamente ao menos 1 ano de distorção, o que causaria viés na estimação.*

*Com relação às crianças de 6 e 7 anos não matriculados, a observação do parecerista é pertinente, mas duas considerações devem ser feitas. Em primeiro lugar, a não matricula desses alunos não chega a representar um grande problema, visto que apenas 1,37% das observações de 7 anos de idade não estão matriculadas em algum nível de ensino e menos de 6% delas estão matriculadas no jardim de infância ou classe de alfabetização, sendo a grande maioria já matriculada no ensino fundamental. Para alunos de seis anos esses mesmo números são de 4,19% para não matriculados e cerca de 35% para jardim de infância ou classe de alfabetização. Em segundo lugar, mesmo para os não matriculados não seria interessante considera-los defasados neste momento, uma vez que esta defasagem teria que ser interpretada como futura ou antecipada, que se realizará apenas quando esta criança tiver 8 anos de idade, mas nenhum ano de escolaridade completo, resultando assim, em um ano de defasagem.*

*Quanto a utilizar uma amostra que contém apenas crianças e jovens matriculados, a decisão derivou de uma opção metodológica. Primeiramente, porém, é preciso considerar que não faz muito sentido falar de distorção para crianças e adolescentes que evadiram o sistema de ensino. Nestes casos, o problema não é o aluno estar atrasado, mas ter deixado de estudar permanentemente. Esse ponto foi acrescentado ao texto, em nota de rodapé na página 9. De fato, crianças e jovens que evadiram estão acumulando anos de atraso, o que levaria a uma subestimação do problema como um todo se os considerássemos no computo da distorção. Além disso, dado que o objetivo do artigo é analisar como variáveis escolares podem afetar a distorção, então o uso exclusivo de alunos matriculados não pode ser evitado para que se possa introduzir as variáveis obtidas a partir dos dados do INEP. Mais ainda, se o objetivo é avaliar a possibilidade de intervenções públicas, então também se faz necessário o recorte pelo ensino básico público. Finalmente, dado a natureza dos dados e da variável dependente, faz-se impossível incorporar ao modelo mudanças de escolas, embora isso certamente ocorra. Para o estudo deste aspecto da distorção, seria necessário dados dos alunos na forma de painel.*

[P] Com relação às variáveis incluídas no modelo, listo algumas dúvidas:

1. Como calcularam a escolaridade do pai? Na PNAD, identificamos a mãe. E se a criança não tem pai nem mãe? Qual o valor da escolaridade que é imputado?

*[A] A variável que identifica o nível de escolaridade máxima dos pais e mães e dos alunos foi criada da seguinte maneira: tomou-se o indivíduo de referência da família e seu cônjuge, e aqueles que eram do sexo masculino foram tomados como “pai” e aqueles do sexo feminino como “mãe”. De fato, há casos em que esses indivíduos não são pais ou mães daqueles utilizados no estudo. Porém, na amostra final, mais de 85% das observações são classificados como filhos da pessoa de referência, de modo que o método utilizado é adequado para estes. Além disso, cerca de 13% da amostra é de indivíduos cuja condição familiar é de parentes da pessoa de referência. Embora não se possa dizer que a escolaridade é de seus pais, ela serve como uma boa proxy se assumirmos que a escolaridade de parentes é similar. Do restante das observações, menos de 1% é de pessoas de referência ou cônjuges, que tem a escolaridade do pai ou mãe como a sua própria, e 0,49% são de pessoas classificadas na família como “agregados”. Utilizando o método de identificação da mãe presente nos microdados da PNAD, foi gerada uma variável com a escolaridade da mãe. Essa variável teve uma correlação de 0,95 com a variável criada por nós para identificar a escolaridade da “mãe”, e de 0,80 com a variável empregada no estudo que identifica a educação máxima do “pai” e “mãe”. Desta forma, dada a alta correlação e a incorporação à base de 2174 alunos que moram apenas com o pai e de 6064 alunos que moram com parentes, se preferiu trabalhar com a variável criada. Além disso, as mesmas regressões foram feitas substituindo a variável empregada pela educação da mãe utilizando o identificador da PNAD e não houve alterações qualitativas na interpretação dos coeficientes, havendo apenas um aumento pequeno no coeficiente da educação da mãe e um aumento maior na magnitude do coeficiente da variável relativa a morar com a mãe. Foi introduzida uma nota de rodapé na página 10 explicando o processo de construção da variável contendo as mesmas considerações feitas aqui. Também, o nome da variável foi substituído para educação da pessoa de referência, para incorporar o fato de estar incluídos pessoas além dos filhos.*

1. [P] Com relação às variáveis escolares, há uma hipótese forte de que a média do estado reflete bem as condições encontradas nos sistemas escolares. Contudo, será que isto é verdade? Como grande parte da escolaridade fundamental é municipal, existem muitas diferenças em termos de infraestrutura e qualidade dos docentes entre os municípios. Gostaria que o autor argumentasse melhor a inclusão destas variáveis em comparação com uma especificação que incluísse apenas dummies para os estados.

*[A] Seria possível fazer essa estimação apenas com dummies para o estado, mas nesse caso estaríamos assumindo que variáveis para escolas municipais e estaduais seriam iguais em todos os estados, aumento ainda mais a homogeneidade entre as observações. De fato, o método utilizado não é o ideal, sendo melhor utilizar diretamente variáveis a nível escolar, mas esta opção só seria possível utilizando bases como a do SAEB. Neste sentido, ao optar pelo uso destas bases, não seria interessante seguir usando os dados da PNAD, uma vez que não há representatividade para os municípios ou dentro deles (apenas região metropolitana e estado)*

[P] Com relação à modelagem econométrica, sugiro que a apresentação dos modelos usados (página 9 a 12) seja somente referenciada e/ou resumida. Grande parte do que está contido nesta parte pode ser encontrada na bibliografia citada. Seria interessante apenas que o autor especificar os modelos conforme suas variáveis e tentando construí-los a partir de uma especificação comum.

*[A] A apresentação dos modelos foi reduzida, conforma solicitado. Os dois modelos estimados também foram especificados na página 14.*

[P] As seções devem ser numeradas.

A seção descritiva deve estar separada da seção econométrica.

Nas tabelas, algumas variáveis (tamanho\_fam, etc) estão com nomes que não são citados ao longo do artigo. O nome das variáveis deve constar na metodologia.

Na seção descritiva, sugiro que incluam também o desvio padrão.

Seria interessante vermos a distorção idade série conforme a idade (na página 17 destacam o resultado da idade, logo seria interessante ver a estatística descritiva sobre isto).

*[A] A sugestões foram incorporadas ao texto. Uma tabela contendo o percentual de alunos com e sem defasagem de acordo com sua idade foi incluída na página 17.*

[P] Com relação aos resultados econométricos, o sinal de urbano, a meu ver, não parece tão contraditório. Talvez este sinal esteja sendo influenciado pela escolha da amostra. Como restringe aos que frequentam, na área rural o problema de defasagem idade série pode estar muito mais associado a questão de não frequência, por falta de escolas, por exemplo. Talvez por isso o sinal tenha ficado positivo.A tabela descritiva mostrava um percentual já alto de pessoas com distorção na área urbana (77% aproximadamente).

*[A] Este comentário foi incorporado ao texto, na página 18.*

[P] Sugiro que o autor explore mais os resultados econométricos, sobretudo, os que se diferenciam do que já foi feito (modelos zeros inflados). A apresentação das tabelas 3 e 4 está confusa e poderia ser melhorada. Talvez seja mais interessante começar a seção com estes resultados diferenciados, os demais são resultados de robustez. A tabela 6, que contém os testes poderia ir para um Anexo e referenciada num texto, tendo em vista que considero mais uma análise de robustez.

*[A] As tabelas 3 e 4 foram refeitas e sua apresentação melhorada significativamente. A tabela 6 foi transferida para o anexo, conforme sugestão. Entretanto, preferiu-se manter a discussão dos resultados econométricos na sequência original, uma vez que se parte de modelos mais rudimentares para modelos mais sofisticados (com zeros inflados), e nesse sentindo acrescentando graus de complexidade à análise. Quanto aos resultados econométricos, os autores concordam que há espaço para aprofundar a sua análise, conforme sugerido pelo parecerista. Porém, preferiu-se não explorar mais a fundo estes resultados por dois motivos. Em primeiro lugar, o artigo já está bastante extenso, com mais de 30 páginas. Em segundo, dado o objetivos do artigo de conduzir uma primeira análise exploratória do tema a partir dos modelos de contagem, incluindo variáveis do ambiente escolar, acredita-se que a presente discussão já está satisfatória, especialmente pelo enfoque dado à análise da variável idade, que mostrou um comportamento interessante, como para as variáveis relativas a características da escola e como estas afetam os coeficientes estimados de outras variáveis dentro do modelo, em especial aquelas associadas ao background familiar.*

[P] Uma dúvida na questão econométrica é de que forma a seleção da sua amostra pode ter afetado os seus resultados. O fato da idade ter um comportamento diferenciado não pode estar sendo influenciado pela restrição a pessoas que estão frequentando a escola?

*[A] Outras regressões foram feitas utilizando os mesmo modelos mas diferentes amostras, incluindo uma com todas as crianças e jovens entre 8 a 18 anos. Nelas o comportamento da variável continuou o mesmo, o que indica a robustez do resultado obtido. Foi introduzida uma nota de rodapé na página 21 com este comentário.*