

UMA ANÁLISE DOS DETERMINANTES DE DESEMPENHO EM CONCURSO PÚBLICO*

IVAN CASTELAR[†]
ALEXANDRE WEBER ARAGÃO VELOSO[‡]
ROBERTO TATIWA FERREIRA[§]
ILTON SOARES[¶]

Resumo

Este trabalho teve como objetivo examinar o papel dos fatores determinantes da aprovação em um concurso público. Tendo como amostra dados de um concurso realizado para o Banco Nordeste do Brasil, no qual concorreram 232.308 candidatos, e usando um modelo logit, as principais conclusões encontradas foram que renda familiar, tipo de educação, ser originário de região metropolitana, idade e ter cursado escola básica privada, são fundamentais para a aprovação do candidato. Enquanto que baixa renda pessoal e ter cursado apenas o ensino médio diminuem a chance do candidato ser aprovado. Os resultados reiteram, portanto, um ciclo vicioso de concentração de renda, o que pode ser extremamente danoso para uma região pobre como o Nordeste do Brasil.

Palavras-chave: Modelo Logit; Concurso Público; Mercado de Trabalho

Abstract

The main objective of this paper was to peruse the role of socio-economic characteristics on the chance of success in an aptitude test for admission to a public bank in the Northeast of Brazil. Using a logit model, and a sample of 232,308 candidates, the main results of the model point that family income, type of education, age and residential location of the candidate are the key factors influencing success in the test. The results, therefore, reiterate a vicious circle of income concentration which can be extremely harmful to a poor region as the Northeast of Brazil.

Keywords: Logit Model; Aptitude Test; Job Market

* Os autores agradecem os comentários de Ricardo Brito, Fabrício Linhares, do Editor desta Revista e de dois pareceristas anônimos. Todos os erros remanescentes, como de praxe, são de nossa exclusiva responsabilidade.

[†] Depto. de Economia Aplicada/UFC Adress:Av. da Universidade, 2700 – 2º andar, CEP: 60451-970 Fortaleza-CE – Fone: 288.77.51 Email: lume1250@yahoo.com.br

[‡] CEDEPLAR/UFGM, email:alexandre.weber@gmail.com

[§] Depto. de Economia Aplicada/UFC, email: rtf2@uol.com.br

[¶] University of Pennsylvania, email: iltonsoares@uol.com

1 Introdução

A instituição do concurso público para provimento de cargos no aparelho do Estado Brasileiro tem se constituído num forte atrator de recursos humanos desde meados do século passado. Assim, por este meio, uma elite egressa da classe média projetou-se na conformação da burocracia estatal brasileira.

Esta atração pode ser explicada por alguns fatores. Um deles é o fato de o setor público pagar melhores salários do que os praticados em níveis similares da iniciativa privada. Existem ainda, atrelados ao emprego público, benefícios indiretos tais como plano de saúde, duração das férias e previdência diferenciada. Outra explicação para esta atração é a de que a alocação no setor público proporciona um maior grau de segurança quanto à permanência no emprego.

O objetivo deste trabalho, por conseguinte, é o de investigar os fatores que levam determinadas pessoas a serem aprovadas em concursos públicos em detrimento de outras. Ou, dito de outra forma, indaga-se se existem fatores socioeconômicos que aumentam as chances de uma pessoa ter melhor desempenho no concurso. Mais especificamente, investiga-se se a renda, o nível de educação acima do exigido pelo concurso e outros fatores sócio-econômico implicam em maiores chances de aprovação. Para tanto, fez-se uso de um estudo de caso com base em dados de um concurso público para provimento de cargo administrativo de nível médio para o Banco do Nordeste do Brasil (BNB), realizado no ano de 2003.

Quanto ao caso estudado, tem-se a particularidade de ter sido um concurso para uma instituição localizada em uma região de baixa renda, quando comparada às regiões Sudeste e Sul do Brasil. Isto reforça a necessidade de entender o que explica a aprovação de um candidato, pois, de forma patente, a região Nordeste ressentia-se de oferta de vagas de trabalho atrativas. Assim, se os candidatos provenientes de famílias de melhor situação financeira tiverem maiores chances de serem aprovados, poderá existir um círculo vicioso de concentração de renda, o qual se torna mais grave em uma região pobre. Na sequência do trabalho essas questões serão devidamente desenvolvidas seguindo a estrutura descrita abaixo.

Além desta introdução, na seção 2, apresenta-se uma breve revisão da literatura sobre o problema. Na seção 3 discute-se a base de dados e na seção 4 expõe-se a metodologia adotada. Na seção 5 são apresentados os resultados empíricos obtidos; e, finalmente, na última seção são apresentadas as considerações finais.

2 Revisão da Literatura sobre Avaliação de Desempenho

O contexto teórico no qual pode ser enquadrado o desempenho de candidatos num concurso é relativamente amplo. As abordagens tradicionais descrevem este processo a partir de uma função de produção, em que o produto é o resultado do concurso, ou a nota do candidato, e os insumos são as variáveis que influenciam o desempenho (Hanushek 1986).

Outro vertente no qual o problema pode ser enquadrado é o da teoria microeconômica do consumidor, em que a equação que determina o resultado no concurso e a sua estática comparativa podem ser derivadas a partir de um problema de escolha entre consumo de bens, tempos de estudo, e trabalho (Fair 1978).

Becker (1974) provê outra abordagem baseada em incentivos e Bergstrom (1989) usa informação assimétrica. Abordagens mais modernas, como a da teoria dos leilões, podem também ser utilizadas. Um concurso pode ser visto como um *chance auction*¹ com preço de reserva, em que os candidatos oferecem lances, equivalentes ao esforço despendido em aprendizagem, e auferem em troca uma probabilidade de ser aprovado no concurso. Quanto maior o lance, maior será a probabilidade de ser aprovado, de forma que se estabelece um critério de dominância estocástica (Benegas 2009).

Os trabalhos empíricos sobre o tema, em geral, versam sobre aplicações de modelos de escolha discreta para determinar os fatores que influenciam o desempenho de alunos em cursos formais; ou numa situação de concursos, como o vestibular no Brasil, ou provas similares em outros países. A seguir faz-se uma breve resenha de trabalhos que se assemelham ao que se pretende desenvolver neste estudo.

Spector & Mazzeo (1980) avaliaram a efetividade de um novo método de ensino em um curso intermediário de macroeconomia. Para tanto, usaram um modelo de escolha binária probit, onde a variável dependente utilizada foi o conceito final no curso. Como regressores foram usados a nota média de admissão do aluno, a pontuação obtida em uma prova de conhecimentos prévios aplicada no início do curso e uma variável *dummy* para representar a diferença entre o novo método e o tradicional.

Brasfield et al. (1993), usam a estrutura educacional, adaptada numa função de produção, e um modelo probit para avaliar a importância do estudo de economia no curso secundário para o desempenho dos alunos de turmas introdutórias de economia na Murray State University. Nesse estudo os autores encontraram uma relação positiva entre os alunos que tiveram aulas de economia no secundário e suas notas nas classes introdutórias de macro e microeconomia.

Wood & Doyle (2002) usaram dados de um teste aplicado para empregados de grandes corporações norte-americanas. Os autores modelaram o número de respostas corretas como um processo de *Poisson*, com média condicional e variância determinadas por um vetor de variáveis explicativas. A variável dependente representava o número de questões corretas e como variáveis explicativas foram utilizadas *dummies* para gerência e outros cargos ocupados pelos empregados, níveis de educação, faixas de renda, origem (hispânica, afro-americana e asiática), faixas de idade, sexo, identidade política (liberal ou conservador), tipo de graduação e se o empregado em algum momento cursou uma disciplina de economia.

Os resultados obtidos revelam que um melhor desempenho nas respostas do questionário está associado a um maior nível de educação, ter feito disciplina de economia na faculdade, ter renda mais alta e ser do sexo masculino. As outras variáveis explicativas mostraram-se insignificantes.

Borg & Stranahan (2002) testaram se a personalidade é uma variável importante para explicar desempenho em cursos avançados de economia. O que estes autores queriam comprovar era se estudantes introvertidos se saem melhor do que estudantes extrovertidos. Para testar a hipótese aventada os autores usaram um modelo probit, onde os intervalos das notas nas provas foram utilizados como variável dependente. As variáveis explicativas foram *dummies* para sexo, raça, tipo de disciplina que cursava (macro, micro etc.),

¹Também conhecido como leilão chinês.

existência de monitor para a disciplina e os oito tipos de personalidades identificadas pelo Briggs Type Indicator (MTBI). Os autores concluíram que o tipo de personalidade é uma variável importante na explicação do desempenho de estudantes nos curso de economia; confirmando, portanto, a hipótese formulada.

Robb & Robb (1999) procuraram identificar o motivo do declínio do número de matrículas nos cursos de economia nos EUA e no Canadá. Para tanto, eles consideraram que a decisão do estudante em cursar economia envolve dois estágios. No primeiro estágio o estudante decide fazer economia no primeiro ano da universidade; e, no segundo estágio, o estudante decide se vai continuar a cursar economia depois dos estudos introdutórios.

A estrutura da estimação foi desenvolvida em duas etapas. Na primeira foi feita uma estimação de mínimos quadrados, em que a equação tinha como variável dependente a nota do aluno na disciplina de microeconomia e como variáveis explicativas um vetor de características dos estudantes, um vetor de características dos professores e *dummies* para alunos que estudaram cálculo, álgebra e outras disciplinas de economia.

Na segunda etapa da estimação os autores usaram um modelo probit, com a equação de regressão explicando se o estudante segue ou não no curso de economia. As variáveis explicativas utilizadas nesta segunda etapa foram as notas no curso introdutório, um vetor de características dos estudantes e um vetor de características dos professores.

Emilio et al. (2004) analisaram os determinantes do acesso à Universidade de São Paulo. O modelo utilizado foi um probit, corrigido pelo viés de seleção, com variável dependente definida como a probabilidade de sucesso no vestibular e um conjunto de características sócio-econômicas como variáveis explicativas. As principais conclusões do trabalho mostram que os fatores mais importantes para o sucesso no vestibular são a educação dos pais e o tipo de escola que os candidatos frequentaram. Os autores constataram ainda que o viés de seleção causa alterações nas magnitudes dos coeficientes estimados mas não muda seus sinais nem a significância estatística.

3 Base de Dados.

No concurso foram aplicadas provas objetivas de Língua Portuguesa, Matemática, Conhecimentos Gerais e Conhecimentos Bancários. Exigia-se ainda que o candidato fosse maior de 18 anos e houvesse completado o ensino médio.

Com base nas características do concurso e num questionário sócio-econômico, preenchido no ato da inscrição, foi definido um conjunto de variáveis que poderia explicar o desempenho dos candidatos. Descreve-se, na Tabela 1, este conjunto de variáveis binárias que comporá o vetor de variáveis independentes, com suas definições e o sinal esperado de cada um dos seus efeitos marginais.

Observa-se que existem quatro categorias para classificar o grau de instrução de cada candidato; a saber, ensino médio, universitário incompleto, universitário completo e pós-graduação. Portanto, quando a variável EMED assume o valor um, ela exclui todas as outras três categorias. Já quando a variável ESUPO assume o valor um, ela exclui duas categorias – ensino médio e superior incompleto.

Tabela 1: Variáveis Explicativas e Sinais Esperados dos Efeitos Marginais

Variável	Notação	Especificação		Sinal esperado
		1	0	
Grau de escolaridade	EMED	Ensino Médio	Outros	-
	ESUPO	Superior completo ou com pós-graduação	Outros	+
Estado civil	SOLT	Solteiro	Outros	±
Renda mensal da família	YFALT	Maior que 10 SM. ^a	Menor ou igual a 10 SM. ^a	+
Renda pessoal mensal	YPBAI	Menor ou igual a 2 SM. ^a	Maior que 2 SM. ^a	-
Escola em que cursou o ensino médio	PRIV	Todo em escola privada	Outros	+
Situação no mercado de trabalho	DEST	Estagiário ou desempregado	Outros	+
Faixa etária	IDBAI	De 18 a 22 anos	Outros	+
Local de origem	RMETR	Oriundo de região metropolitana	Outros	+

^a Abreviamos salário mínimo por SM.

Fonte: Elaboração dos autores.

O efeito marginal do estado civil é, a princípio, ambíguo. Ao mesmo tempo em que um indivíduo casado tem o seu tempo de estudo limitado por obrigações domésticas, inexistentes para um indivíduo solteiro, ele deve encontrar maior motivação para ser aprovado no concurso e assim melhorar a sua situação familiar. O raciocínio inverso deve ser verdadeiro para um indivíduo sem vínculos conjugais.

A definição de desempregado usada aqui não deve ser confundida com a noção tradicional de desemprego formal ou friccional. Cabe observar que na pergunta do questionário sócio-econômico, que trata da condição no mercado de trabalho, não havia uma alternativa referente a estar desocupado; ou seja, não estar procurando emprego. Assim, dentre os que se denominaram desempregados é provável que haja uma parcela razoável que tenha optado por se dedicar aos estudos antes de entrar no mercado de trabalho, ou para retornar a esse mercado em melhores condições de colocação ou remuneração. Por conta disso, resolveu-se fundir as categorias de desempregados e estagiários numa só. A ideia é juntar categorias de pessoas que dispõem de tempo para o concurso maior do que a média dos demais candidatos. Verifica-se posteriormente se esta escolha foi acertada.

Vale ainda lembrar que a definição do corte para as variáveis renda familiar e pessoal foi baseada num índice de desempenho preliminar. Tal índice foi, a princípio de maneira *ad hoc*, definido como a razão entre percentual de candidatos aprovados num determinado intervalo de renda e o percentual do

total de candidatos no mesmo intervalo. Então, os cortes de 2 e 10 salários mínimos, foram estabelecidos de acordo com a condição do índice preliminar de desempenho ser menor ou maior do que a unidade. Por outro lado, os dados da PNAD de 2003, ano do concurso, mostram que 81% das pessoas ocupadas no Nordeste ganham até 2 salários mínimos mensais e 96,5% das famílias ganham menos de 10 salários mínimos. Portanto, o índice de desempenho reflete com fidelidade pessoas ou famílias mais ou menos abastadas para o padrão da região.

4 Metodologia

A explicação da probabilidade de sucesso num concurso é um caso típico de modelo de escolha binária. Assim, a equação que modela esta probabilidade é definida como,

$$P(\text{Sucesso}) = F(X\beta) \quad (1)$$

Onde X é o vetor de variáveis explicativas, definidas anteriormente, F é uma função cumulativa de probabilidade e β é um vetor de parâmetros que ultrapassa a dimensão de X por um termo que representa o intercepto.²

Há problemas na estimação desses modelos semelhantes aos encontrados em modelos de regressões em geral. No entanto, estes problemas podem, em alguns casos, ter consequências mais graves do que em modelos tradicionais.

Particularmente, este trabalho se preocupa com quatro tipos de erros comuns na especificação destes modelos. O primeiro é o de viés de seleção, ou erro de especificação na seleção da amostra. O segundo refere-se à omissão de variáveis explicativas, no qual se inclui o problema da heterogeneidade não observada. O terceiro tipo versa sobre a determinação simultânea das variáveis; ou seja, variáveis do lado direito da equação requerem equações adicionais para explicá-las. Finalmente, o quarto tipo de erro é sobre a especificação funcional; isto é, a função de probabilidade escolhida pode não ser a mais adequada. Comenta-se a seguir cada erro individualmente e as soluções propostas para contorná-los.

Especificamente, o viés de seleção emerge nesse caso porque anteriormente indivíduos tomaram a decisão de inscrever-se, ou não, no concurso. A questão crucial, portanto, é decidir até que ponto essa escolha prévia influencia a explicação do resultado do concurso.

O viés de seleção é prejudicial aos estimadores dos parâmetros de uma regressão quando as características não observáveis que afetam a seleção são correlacionadas com as características não observáveis que afetam a variável de interesse. Em outros termos, isto é equivalente à existência de correlação entre as características não observáveis do mecanismo de seleção e o erro da regressão (Vella 1998). Devido a essa correlação é necessário introduzir na regressão que explica o sucesso no concurso um termo de correção proveniente da modelagem do processo decisório de inscrição (Emilio et al. 2004). Nesse trabalho, preferiu-se não modelar a decisão prévia. Na apresentação dos resultados empíricos justifica-se a razão de não fazê-lo e qualificam-se os resultados obtidos à luz de possíveis vieses.

²A justificativa econométrica desta equação pode ser derivada de um random utility model (Train, 2003)

Na equação definida em (1) a probabilidade de sucesso em um concurso é modelada como função de um vetor de atributos, incluindo variáveis de grau de instrução. Portanto, não é possível observar a habilidade específica de cada candidato. Mesmo se fosse possível observá-la seria difícil mensurá-la, embora o quociente de inteligência (QI) seja uma boa *proxy* para habilidade.

No entanto, como a heterogeneidade não observada enquadra-se no caso mais geral de erro de omissão de variáveis, existe uma estratégia para testar a sua existência ou, mais importante, testar se a sua ocorrência é danosa para a estimação. O teste de White (1980) para heterocedasticidade é também um teste para erro de especificação. Portanto, um teste de White, caso se comprove a ausência de heterocedasticidade, poderá indicar também se a omissão é danosa para o processo de estimação.

Existe também a possibilidade de endogeneidade das variáveis do lado direito do modelo. Renda do candidato, renda da família, nível educacional e mesmo estado civil podem ser simultaneamente determinadas. Contudo, é extremamente improvável que os erros das equações que determinam estas variáveis sejam correlacionados com o erro da equação que explica o sucesso no concurso (Sargan 1975). A razão para isto é que cada uma dessas variáveis é determinada em diferentes mercados, como as rendas; ou por diferentes mecanismos de formação, como o grau de instrução e o estado civil. Portanto, o que se presume, é que exista correlação entre estas variáveis, mas não a existência de viés de simultaneidade.

O último tipo de erro de especificação a ser discutido é o do formato da escolha funcional. As formas funcionais mais populares para tratar variáveis dependentes binárias são os modelos probit e logit. Embora estes modelos forneçam resultados muito similares, nem por isso deve-se negligenciar um teste sobre a escolha funcional, principalmente pelo motivo da amostra nesse problema apresentar um número de valores nulos de Y substancialmente maior do que o número de 1's. Discriminar entre estes dois formatos requer um teste de modelos não entrelaçados. Silva (2001) mostra que, sob condições adequadas de regularidade, hipóteses alternativas sobre forma funcional em modelos binários podem ser entrelaçada em um modelo dado por

$$L_{\rho}(Y/X, \rho, \alpha, \beta, \gamma) = \frac{[(1 - \alpha)L_1(Y/X, \beta)^{\rho} + \alpha L_2(Y/X, \gamma)^{\rho}]^{1/\rho}}{\int [(1 - \alpha)L_1(Z/X, \beta)^{\rho} + \alpha L_2(Z/X, \gamma)^{\rho}]^{1/\rho} dZ} \quad (2)$$

onde L denota máxima verossimilhança. Para $\rho = 1$ a equação acima transforma-se em:

$$L(Y/X, \alpha, \beta, \gamma) = (1 - \alpha)L_1(Y/X, \beta) + \alpha L_2(Y/X, \gamma) \quad (3)$$

que é uma combinação linear convexa das formas funcionais alternativas e a decisão sobre a melhor hipótese recai sobre o valor de α .³ Este teste tanto é adequado para discriminar entre os modelos probit e logit, como para contrastar um desses modelos com um modelo de categoria assimétrica, como o gompit, por exemplo.

³Esta forma de entrelaçar é equivalente àquela proposta por Davidson e Mackinnon (1993).

5 Resultados Empíricos

Para explicar o resultado do concurso escolheu-se em (1) a função logit para representar F. O modelo foi estimado através de Máxima Verossimilhança e o algoritmo de maximização precisou de 10 interações para convergir. O valor da razão de verossimilhança para testar a restrição de que todos os coeficientes, exceto o intercepto, são iguais a zero foi de 453,3. O ponto crítico para, uma qui-quadrado com 9 graus de liberdade, dado um nível de significância de 5%, é de 16,9. Portanto, rejeita-se a hipótese de que conjuntamente as variáveis independentes não exercem influência sobre a probabilidade de sucesso no concurso.

Além do R^2 de McFadden, o qual é extremamente baixo nesse caso, outra medida de ajustamento é o percentual de valores corretamente preditos, o qual fornece o percentual corretamente predito de 1's, quando a probabilidade é maior do que um certo valor crítico; e de zeros, quando a probabilidade é menor do que este valor.

Usualmente este valor crítico é estabelecido como 0,5. No entanto, neste problema específico, como a proporção de zeros é substancialmente maior do que a de 1's, usou-se a probabilidade empírica de passar no concurso; ou seja, 0,00266. Assim a proporção de valores corretamente preditos para $Y=1$ é de 61,26% e a proporção de valores corretamente preditos para $Y=0$ é de 74,27%. Tais resultados levam a um valor ponderado de predições corretas de 61,29%. No entanto, é razoável ter em mente que medidas de ajustamento em modelos de escolha binária não são importantes nem do ponto de vista estatístico nem do ponto de vista econômico para avaliar a significância das variáveis explicativas (Wooldridge 2002). O teste da razão de máxima verossimilhança e os testes individuais de significância são tudo que se precisa para avaliar o modelo.

A razão entre o coeficiente estimado e o erro padrão representa uma estatística distribuída assintoticamente como um normal padrão. Esses valores são utilizados para a realização do teste de hipótese sobre a significância estatística de cada um dos parâmetros estimados. Como pode se observar na Tabela 2, somente a variável referente ao estado civil (SOLT) não se mostrou significativa, ao nível de 5%.

Quanto aos resultados para as outras variáveis estes são agrupadas em blocos, por um critério de similaridade, e seus efeitos marginais são comentados individualmente.

5.1 Determinantes Gerais: Local de Origem, Faixa Etária e Situação no Mercado de trabalho

Os resultados apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos originários de regiões metropolitanas. Pode-se afirmar que o elemento diferencial desse determinante é a maior quantidade e melhor qualidade dos serviços ligados à educação e à cultura nos grandes centros urbanos, *vis-à-vis* os municípios do interior.

No quesito faixa de idade os resultados apontaram um efeito marginal positivo sobre a chance de aprovação no concurso para os candidatos na faixa de 18 a 22 anos. Este resultado pode ser justificado com base na argumentação de que estes candidatos possuem um melhor ritmo de estudo, por estarem ainda frequentando aulas ou por terem parado de estudar há pouco tempo.

Tabela 2: Coeficientes Estimados e Efeitos Marginais

Variável	Notação	Coeficiente estimado	P-valor	Efeito Marginal
Intercepto	C	-63,538	0,0000	-
Ensino Médio	EMED	-13,209	0,0000	-0,3301
Nível Superior Completo ou com Pós-graduação	ESUPO	0,2897	0,0039	0,0724
Estado Civil - Solteiro	SOLT	0,0042	0,9696	0,0010
Renda Familiar superior a 10 Salários Mínimos	YFALT	0,5459	0,0000	0,1364
Renda Individual abaixo de 2 Salários Mínimos	YPBAI	-0,1829	0,0266	-0,0457
Toda em escola Privada	PRIV	0,2503	0,0045	0,0626
Estagiário ou Desempregado	DEST	0,3199	0,0002	0,0800
Oriundos de Regiões Metropolitanas	RMETR	0,5034	0,0026	0,1258
Idade de 18 a 22	IDBAI	0,3259	0,0000	0,0815

Fonte: elaboração dos autores.

Um efeito marginal positivo sobre a chance de aprovação dos candidatos que estão desempregados ou são estagiários significa que essa condição no mercado de trabalho possibilita uma maior disponibilidade de tempo para dedicação aos estudos, o que vai se traduzir em uma maior chance de aprovação.

Com vistas a dirimir a dúvida sobre juntar estagiários e desempregados numa só variável, estimou-se a regressão separando as duas categorias. Os resultados mostram que os coeficientes para estagiários e desempregados têm o mesmo sinal positivo e não há diferença estatística sensível entre eles. Portanto, juntá-las acarreta o benefício adicional de redução na variância dos estimadores, pois isto equivale a estimar o modelo sob restrição.

5.2 Determinantes Relacionados à Renda: Renda Pessoal e Renda Familiar

Dispor de renda pessoal baixa produziu um efeito marginal negativo sobre a chance de aprovação. Este resultado está em conformidade com a argumentação de serem piores as condições para os que têm baixos rendimentos, visto que essa situação é limitante do acesso a materiais de estudo e a cursos preparatórios de qualidade. Fica clara, portanto, a relevância do acesso a produtos e serviços que reforcem, ou complementem o aprendizado já adquirido; ou que iniciem os candidatos em temas desconhecidos, como os exigidos na prova específica de conhecimentos bancários, por exemplo.

Possuir uma renda familiar alta redundou num efeito marginal positivo sobre a chance de aprovação. Todo o *background* dos candidatos oriundos de famílias de maior renda sugere o perfil de uma pessoa que cursou ensino formal de qualidade, com acesso a materiais didáticos e para-didáticos, ambiente propício ao desenvolvimento cognitivo, maior volume de informações, acesso a atividades extracurriculares, inclusão digital, e a outros elementos materiais e intangíveis. Todas estas características contribuem, naturalmente, para uma maior chance de aprovação no concurso.

5.3 Determinantes Relacionados à Educação Formal: Graduado ou Pós-graduado, Ensino Médio e Ensino Médio em Escola Privada.

Os candidatos com formação além do ensino médio, ou os “super-educados”, apresentam efeito marginal positivo sobre a chance de aprovação. Isso reflete, fundamentalmente, uma maior maturidade intelectual e conhecimento mais vasto sobre temas cobrados nas provas que iam além do que é ministrado no ensino médio.

Contudo, não se pode desprezar o fato de que num concurso que exige nível médio, no qual participaram praticamente 80.000 candidatos universitários, mais de 35.000 graduados e aproximadamente 5.600 pós-graduados, 28,5% das vagas ofertadas ficaram ociosas.

Ou seja, de um universo de cerca de 120.600 candidatos que são pelo menos universitários, somente 519 cumpriram os requisitos mínimos para serem aprovados no concurso, os demais perderam para si mesmos. Portanto, há uma forte evidência de que o baixo nível da qualidade do ensino tem transbordado para o nível superior.

O efeito marginal sobre a chance de aprovação de candidatos que cursaram apenas o nível médio foi negativo. Além de influenciar inversamente a probabilidade de passar no concurso, esta variável foi também a que produziu maior efeito marginal em valor absoluto. Este resultado se deve fundamentalmente à baixa qualidade do ensino médio público e à concorrência com candidatos universitários, ou até mesmo pós-graduados.

Ter cursado o ensino médio em escola privada apresentou efeito marginal positivo na chance de aprovação. Este resultado, como o anterior, também se deve às condições precárias do ensino público de nível médio e à ascensão da escola privada, onde são ofertadas metodologias de ensino, ambiente de estudo e acesso a materiais didáticos complementares incomparáveis com aqueles disponíveis na rede pública.

Outra medida interessante que pode ser extraída da estimação do modelo é a do risco relativo, o qual mede a razão das probabilidades de sucesso, condicionadas a uma série específica de atributos (King & Zeng 2001). Assim, calculou-se a razão de probabilidades entre um indivíduo típico e um indivíduo atípico passarem no concurso. Isto corresponde, grosso modo, à razão entre probabilidades de um indivíduo rico e um indivíduo pobre obterem sucesso. A razão calculada foi de 41,14, o que significa que um indivíduo de boa situação financeira familiar, jovem, universitário ou pós-graduado, oriundo de região metropolitana e que cursou escola básica privada, tem cerca de 41 vezes mais chance de passar no concurso do que um indivíduo que não possui estes atributos.

5.4 Testes de Especificação

Para avaliar problemas de especificação no modelo dois tipos de teste foram realizados. O primeiro foi o teste de heterocedasticidade. Vale lembrar que modelo logit é essencialmente heterocedástico e sua estimação, através de mínimos quadrados não-lineares ponderados, leva em conta este fato. Portanto, o teste de White, neste caso, tem como objetivo central investigar uma possível omissão de variáveis. Como a suspeita sobre o erro de especificação recai sobre a ausência de uma medida de habilidade na equação, optou-se por definir

o vetor que modela a variância do erro como constituído apenas das variáveis de instrução

O resultado deste teste, baseado na regressão auxiliar apresentada no Apêndice [Apêndice A](#), mostra que o valor calculado para o teste de Lagrange é de 0,62. Já o valor crítico da distribuição qui-quadrado, a um nível de significância de 5% e com dois graus de liberdade, é de 5,99. Portanto, é aceita a hipótese nula de homocedasticidade; o que aqui significa, em última análise, uma indicação de que se houve omissão de variáveis esta não é danosa. Este resultado pode parecer surpreendente à primeira vista; no entanto, é em essência o mesmo encontrado por [Griliches & Mason \(1972\)](#), os quais não encontraram indícios de correlação entre habilidade e escolaridade.

O segundo teste, referente a erro de especificação, foi o da forma funcional. Apesar da argumentação de alguns autores ([Train 2003](#)) de que os modelos probit e logit são empiricamente indistinguíveis, existem situações em que eles podem divergir, requerendo nestes casos um teste formal de escolha.

Portanto, realizou-se um teste para discriminar entre os modelos logit e probit. Assim, o modelo probit foi estimado e seus resultados são apresentados no Apêndice [Apêndice B](#), já o resultado do teste é apresentado no Apêndice [Apêndice C](#). Como o valor do coeficiente da variável Z , correspondente a α na equação (3), é estatisticamente insignificante, não se rejeita a hipótese do modelo logit ser correto.

Escolher o modelo logit tendo como alternativa o modelo probit não leva à conclusão de que o primeiro é o mais apropriado. Realizou-se, então, o teste tendo como alternativa uma distribuição gompit, representativa da classe assimétrica.

O resultado do teste para discriminar entre o logit e o gompit, mostrado no Apêndice [Apêndice D](#), indica que o coeficiente estimado de Z é estatisticamente insignificante ao nível usual de 5%. Portanto, pode-se ter uma razoável confiança de que a forma funcional escolhida foi bem especificada.

Existe ainda a possibilidade de que as estimativas obtidas estejam sendo afetadas pelo viés de seleção, mencionado anteriormente. A alternativa para corrigi-las seria construir uma amostra censurada, com dados da PNAD, e modelar também o processo de inscrição ([Emilio et al. 2004](#)). No entanto, existem duas razões para não fazê-lo.

A primeira delas é a de que no questionário respondido pelos candidatos não há informação sobre o estado ou cidade de origem do entrevistado. Isto geraria certa imprecisão na construção de uma amostra que também incluísse pessoas que não se inscreveram no concurso.

A segunda é que a correção do viés de seleção usando dados de uma base diferente como a PNAD por si só não é garantia de bons resultados.⁴ Um exemplo de uso da PNAD para este fim, onde aparecem distorções inaceitáveis, pode ser constatado em [Duarte & Silveira \(2009\)](#).

Portanto, os resultados aqui apresentados devem ser vistos com a necessária prudência. Como o interesse principal deste artigo é distinguir diferenças de desempenho entre indivíduos com uma dotação de características superior a de indivíduos menos afortunados, um viés sistemático dos coeficientes não alteraria o ordenamento do desempenho e não comprometeria a substân-

⁴Embora que, ao preservar a significância e a ordem de grandeza dos coeficientes, a correção de Emilio, Belluzzo e Alves (2004) mostre-se robusta.

cia dos resultados. Já um viés errático poderia mudar fundamentalmente as conclusões obtidas.

6 Conclusões

Após a especificação de um modelo logit, onde a probabilidade de sucesso no concurso foi definida como função de um vetor de atributos sócio-econômicos, constatou-se que alta renda familiar, escolaridade acima do ensino médio, ser oriundo de região metropolitana, ter cursado o ensino médio em escola privada e ser jovem, são fatores que contribuem para aumentar a chance de passar no concurso. Por outro lado, ter renda pessoal abaixo de dois salários mínimos e ser detentor apenas de escolaridade de ensino médio contribui negativamente para aprovação no concurso. Constatou-se ainda que a variável ensino médio apresenta o maior efeito marginal em termos absolutos, o que dramatiza a situação dos candidatos de baixa escolaridade.

Delimita-se assim o perfil de um candidato jovem, pertencente a uma classe média de renda relativamente elevada, habitante de região metropolitana e que teve acesso à instrução e a meios culturais de qualidade como elemento típico de quem obteve sucesso no concurso. Este indivíduo, pela medida de risco relativo, tem 41 vezes mais chance de aprovação do que um indivíduo sem estes atributos. Existe, portanto, uma evidência razoável de persistência do status econômico de uma geração para outra.⁵ Aliás, é relevante frisar, que uma análise qualitativa dos resultados do ENEM para o ano de 2003 mostra conclusões inteiramente similares às obtidas neste trabalho (Veloso 2004).

Obviamente, não é papel do sistema educacional preparar alunos para concursos, seus objetivos precípuos vão muito além disso. No entanto, deve haver certa conformidade entre a escola e o mercado de trabalho. Nesse sentido, investir na qualidade do ensino médio público pode ser uma política que ajude a contrabalançar os efeitos perversos das diferenças de dotações iniciais dos indivíduos. Ao contrário de políticas afirmativas, como a do sistema de quotas, com efeitos predominantemente de curto prazo, apostar numa política educacional de longo prazo pode ser uma maneira de amenizar o quadro de desigualdades que permeia o meio social brasileiro, e nordestino em particular.

Por outro lado, a presença de candidatos com qualificação além da requerida no concurso revela o fenômeno caracterizado como *overeducation* (Rumberger 1981). Preponderantemente esse fenômeno ocorre por desequilíbrios entre o mercado de trabalho e o sistema educacional, ou pela baixa qualidade deste último. As evidências apontam para a improdutividade, para os baixos salários e para a permanência dos indivíduos em situação de *overeducation* (Groot 1996). Portanto, preencher postos de trabalho, que requerem apenas educação de nível médio, com detentores de título universitário ou com pós-graduação, aponta para uma alocação ineficiente dos recursos públicos conjugada a uma falha estrutural do mercado de trabalho e do sistema de ensino regional.

⁵Este fenômeno de persistência é notado por Stokey (1998) para a realidade americana.

Referências Bibliográficas

- Becker, G. (1974), 'A theory of social interactions', *Journal of Political Economy* 82(6), 1063–1093.
- Benegas, M. (2009), Bidding behavior in chance auctions. Working Paper no. 18, CAEN/UC.
- Bergstrom, C. (1989), 'A fresh look at the rotten kid theorem – and other household mysteries', *Journal of Political Economy* 97(3).
- Borg, M. & Stranahan, H. (2002), 'Personality type and student performance in upper-level economics courses: the importance of race and gender', *Journal of Economic Education* pp. 3–14.
- Brasfield, D., Harrison, D. & McCoy, J. (1993), 'The impact of high school economics on the college principles of economics course', *Journal of Economic Education* 24(2), 99–111.
- Duarte, G. B. & Silveira, R. M. (2009), Avaliando o impacto do programa bolsa família sobre a frequência escolar: o caso da agricultura familiar no nordeste do brasil, in ANPEC, ed., 'Anais do XIII Encontro Regional de Economia', Fortaleza.
- Emilio, D. R., Belluzzo, W. & Alves, D. C. (2004), 'Uma análise econométrica dos determinantes do acesso à universidade de são paulo', *Economia Aplicada* 34(2), 275–305.
- Fair, R. C. (1978), 'A theory of extramarital affairs', *The Journal of Political Economy* 86(1), 45–61.
URL: <http://www.jstor.org/stable/1828758>
- Griliches, Z. & Mason, W. M. (1972), 'Education, income, and ability', *The Journal of Political Economy* 80(3), S74–S103.
URL: <http://www.jstor.org/stable/1831252>
- Groot, W. (1996), 'The incidence of, and returns to overeducation in the uk', *Applied Economics* 28(10), 1345–1350.
- Hanushek, E. (1986), 'The economics of schooling: production and efficiency in public schools', *Journal of Economic Literature* 24(3), 1141–1177.
- King, G. & Zeng, L. (2001), 'Logistic regression in rare events data', *Political Analysis* 9(2), 137.
- Robb, R. & Robb, A. (1999), 'Gender and the study of economics: The role of gender of the instructor', *The Journal of Economic Education* 30(1), 3–19.
- Rumberger, R. W. (1981), *Overeducation in the New York Labor Market*, Praeger, New York.
- Sargan, D. (1975), A comment, in G. A. Renton, ed., 'Modelling the economy', Heinemann Educational Books.
- Spector, L. & Mazzeo, M. (1980), 'Probit analysis and economic education', *Journal of Economic Education* 11(2), 37–44.

Stokey, L. C. (1998), Shirtsleeves to shirtleeves: The economics of social mobility, in N. Schwarz, D. Jacobs, E. Kalai & M. Kamien, eds, 'Frontiers of research in economic theory', Cambridge University Press.

Train, K. (2003), *Discrete choice methods with simulation*, Cambridge University Press.

Vella, F. (1998), 'Estimating models with sample selection bias: a survey', *The Journal of Human Resources* **33**(1), 127–169.

Veloso, A. W. A. (2004), Determinantes de desempenho em concursos públicos: um estudo de caso., Master's thesis, CAEN/UFC.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**(4), 817–838.

URL: <http://www.jstor.org/stable/1912934>

Wood, W. & Doyle, J. (2002), 'Economic literacy among corporate employees', *Journal of Economic Education* pp. 195–205.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Apêndice A Regressão Auxiliar para Teste de Heteroscedasticidade

Variável	Coefficiente Estimado	Erro Padrão	Estatística t	p-valor
FP	0,0140	0,150	0,093	0,9255
FP*EMED	-0,7609	1,677	-0,453	0,6500
FP*ESUPO	0,4716	0,919	0,512	0,6082
FP*SOLT	0,0005	0,109	0,004	0,9961
FP*YFALT	-0,0127	0,104	-0,122	0,9029
FP*YPBAI	-0,0009	0,084	-0,011	0,9907
FP*PRIV	-0,0028	0,089	-0,032	0,9744
FP*DEST	-0,0008	0,090	-0,009	0,9923
FP*IDBAI	-0,0071	0,112	-0,063	0,9495
FP*RMETR	-0,0036	0,095	-0,038	0,9697
FP*(-X β)*EMED	0,1090	0,242	0,449	0,6531
FP*(-X β)*ESUPO	-0,0909	0,177	-0,512	0,6084
R^2				0,0000003
Soma do quadrado dos resíduos				234307,7
R^2 Ajustado				-0,0000045
Critério de informação de Akaike				2,846552
Média da Variável Dependente				0,000056
Critério de Schwarz				2,847087
S.D. da Variável Dependente				1,004298
Estatística Durbin-Watson				0,292656
Erro Padrão da Regressão				1,004321
Log-verossimilhança				-330626,4
Estatística de Razão de verossimilhança ($\ln \lambda$) - 2 graus de liberdade				0,62341
p-valor de λ				0,73219

Fonte: Elaboração dos autores

Apêndice B Resultados do Modelo *Probit*

Variável	Coefficiente Estimado	Erro Padrão	Estatística <i>t</i>	p-valor
C	- 2,919	0,043	-67,75	0,0000
EMED	- 0,412	0,036	-11,41	0,0000
ESUPO	0,095	0,034	2,75	0,0058
SOLT	-0,004	0,036	-0,11	0,9093
YFALT	0,190	0,034	5,51	0,0000
YPBAI	-0,06	0,027	-2,30	0,0210
PRIV	0,088	0,029	3,03	0,0024
DEST	0,102	0,028	3,54	0,0004
IDBAI	0,108	0,036	3,00	0,0027
RMETR	0,165	0,028	5,75	0,0000
Média da Variável Dependente				0,002660
Critério de informação de Akaike				0,034998
S.D. da Variável Dependente				0,051509
Critério de Schwarz				0,035444
Erro Padrão da Regressão				0,051452
Critério de Hannan-Quinn				0,035128
Soma do quadrado dos resíduos				614,9534
R^2 McFadden				0,052872
Log-verossimilhança				-4.054,7414
Log-verossimilhança restrita				-4.281,5040
Estatística de Razão de verossimilhança ($\ln \lambda$) – 9 graus de liberdade				452,7414
p-valor de λ				0,0000
Obs. com $y=0$				231.690
Total de observações				232.308
Obs. com $y=1$				618

Fonte: Elaboração dos autores.

Apêndice C Resultados de Teste de Especificação para Forma Funcional – Logit x Probit

Variável	Coefficiente Estimado	Erro Padrão	Estatística <i>t</i>	p-valor
p-valor				
C	-6,3547	0,136	-46,69	0,0000
EMED	-1,3210	0,118	-11,15	0,0000
ESUPO	0,2901	0,101	2,86	0,0042
SOLT	0,0048	0,111	0,04	0,9654
YFALT	0,5460	0,098	5,54	0,0000
YPBAI	-0,1824	0,08	-2,21	0,0266
PRIV	0,2498	0,089	2,77	0,0055
DEST	0,3202	0,088	3,62	0,0003
IDBAI	0,3259	0,108	3,00	0,0027
RMETR	0,5034	0,088	5,69	0,0000
Z	-0,0403	1,402	-0,03	0,9770
Média da Variável Dependente				0,002660
Critério de informação de Akaike				0,035004
S.D. da Variável Dependente				0,051509
Critério de Schwarz				0,035494
Erro Padrão da Regressão				0,051451
Critério de Hannan-Quinn				0,035147
Soma do quadrado dos resíduos				614,9495
R ² McFadden				0,052937
Log-verossimilhança				4.054,8530
Log-verossimilhança restrita				4.281,5040
Estatística de Razão de verossimilhança (ln λ) – 10 graus de liberdade				453,4006
p-valor de λ				0,0000
Obs. com $y = 0$	231.690			
Total de observações				232.308
Obs. com $y = 1$				618

Fonte: Elaboração dos autores

Apêndice D Resultados de Teste de Especificação para Forma Funcional – *Logit x Gompit*

Variável	Coefficiente Estimado	Erro Padrão	Estatística t	p-valor
C	-6,3564	0,136	-46,44	0,0000
EMED	-1,3215	0,118	-11,15	0,0000
ESUPO	0,2908	0,101	2,86	0,0041
SOLT	0,0059	0,111	0,05	0,9575
YFALT	0,5463	0,098	5,54	0,0000
YPBAI	-0,1816	0,083	-2,16	0,0303
PRIV	0,2490	0,089	2,77	0,0055
DEST	0,3211	0,088	3,62	0,0003
IDBAI	0,3261	0,108	3,00	0,0026
RMETR	0,5036	0,088	5,69	0,0000
Z	-0,0614	0,787	-0,07	0,9379
Média da Variável Dependente				0,002660
Critério de informação de Akaike				0,035004
S.D. da Variável Dependente				0,051509
Critério de Schwarz				0,035494
Erro Padrão da Regressão				0,051451
Critério de Hannan-Quinn				0,035147
Soma do quadrado dos resíduos				614,9495
R^2 McFadden				0,052938
Log-verossimilhança				4.054,8530
Log-verossimilhança restrita				4.281,5040
Estatística de Razão de verossimilhança ($\ln \lambda$) – 10 graus de liberdade				453,3077
p-valor de λ				0,0000
Obs. com $y=0$				231.690
Total de observações				232.308
Obs. com $y = 1$				618

Fonte: Elaboração dos autores