

Análise do artigo “Migração e inserção no mercado de trabalho: uma abordagem multinomial para a população economicamente ativa do Brasil”

AUTORES DO ARTIGO: ANA CAROLINA DA CRUZ LIMA, ANA MARIA HERMETO CAMILO DE OLIVEIRA, RODRIGO SIMÕES.

LOCAL DE PUBLICAÇÃO: ANAIS DO XXXIX CONGRESSO DA ANPEC EM FOZ DO IGUAÇU, 2011.

Objetivo

-
- Verificar a influência do processo migratório no Brasil sobre a condição de ocupação dos trabalhadores.

Diferenças na participação no mercado de trabalho

- Inter-relação entre as taxas de desemprego nacional e regionais.
 - Infraestrutura;
 - Mudanças econômicas nacionais (ex: Plano Real).
- Elevada heterogeneidade das taxas de desemprego e rendimento entre e dentro das regiões brasileiras.
 - Dentro das regiões devido a características não produtivas (sexo e cor de pele).
- Taxas regionais de desemprego tendem a convergir no longo prazo devido à migração.

Metodologia

- Processo migratório pode fazer com que os indivíduos mais qualificados se auto selecionem para migrar.
- A intuição é que esses indivíduos mais qualificados em regiões com poucas oportunidades econômicas tendem a migrar para buscar melhores oportunidades.

Metodologia

- Duas cidades: A e B.
 - Cidade A não tem empregos;
 - Cidade B tem empregos sobrando.

A

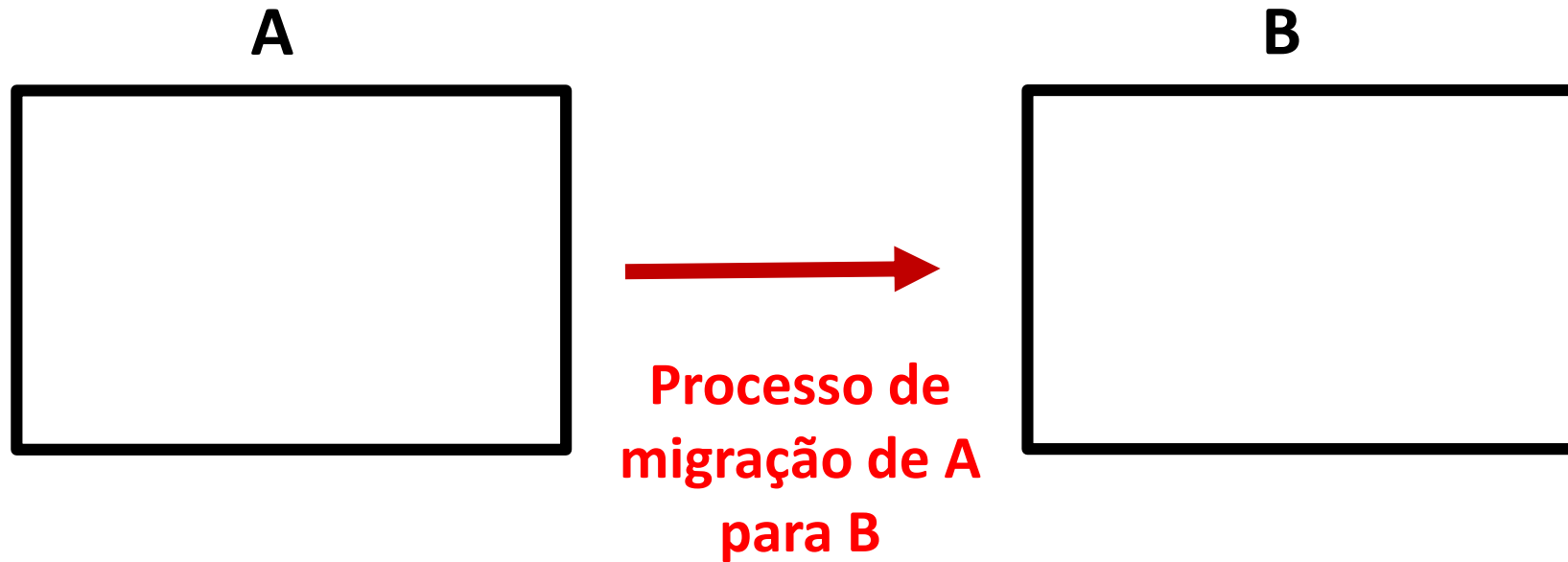


B



Metodologia

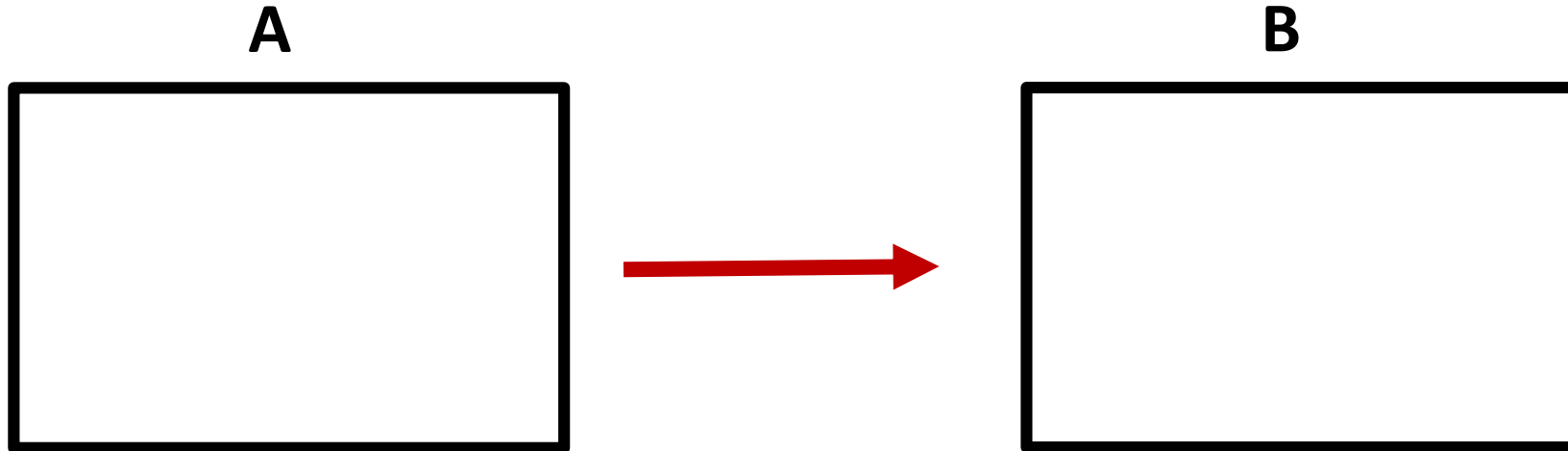
- Duas cidades: A e B.
 - Cidade A não tem empregos;
 - Cidade B tem empregos sobrando.



Metodologia

- Duas cidades: A e B.
 - Cidade A não tem empregos;
 - Cidade B tem empregos sobrando.

Se uma amostra representativa (em termos de características produtivas) da população de A migra para a cidade B, então não temos problemas de seleção!



Metodologia

- Endogeneidade no processo migratório:
 - Só se observa dados (condição de trabalho, salário, ...) apenas para aqueles que migraram.
 - Supõe-se que quem migrou apresenta as características produtivas “melhores”, em relação à população do seu município de origem.
 - Estimativas de MQO inconsistentes e viesadas.
- Essa análise lida com o problema de **censura** nos dados!

Metodologia

- Os autores utilizam uma variação do Modelo de auto seleção de Heckman (1976, 1979), compreendendo 2 etapas:
- I) Estimação da probabilidade de um indivíduo ser migrante;
- II) Modelo multinomial para verificar a influência dessa probabilidade sobre a condição do indivíduo no mercado de trabalho (ocupado no setor formal, ocupado no setor informal ou desocupado).

Metodologia

- Importante! Os autores não utilizam o Modelo de auto seleção de Heckman propriamente dito!
- Assim, vou apresentar os resultados deles e realizar uma análise própria utilizando o Modelo de auto seleção de Heckman.
- Primeiro, vou apresentar a metodologia por trás do:
 - Modelo de auto seleção de Heckman;
 - Modelo *logit* multinomial.

Modelo de auto seleção

Metodologia – Modelo de seleção

- Suponha a equação de salários (**EQUAÇÃO DE INTERESSE**):

$$W_i = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \quad , \quad \begin{array}{l} \varepsilon \sim N(0, \sigma_\varepsilon^2) \\ X \text{ independente de } \varepsilon \end{array}$$

- W_i só é observado para quem trabalha. Então, precisamos levar em consideração a decisão de entrar no mercado de trabalho dos indivíduos (auto seleção)!

- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad \begin{array}{l} u_i \sim N(0, \sigma_u^2) \\ Z \text{ independente de } u \end{array}$$

Metodologia – Modelo de seleção

-
- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad u_i \sim N(0, \sigma_u^2)$$

- Temos que \mathbf{Z} é um vetor de características individuais e produtivas.
- Assim, podemos representar essa variável latente pela seguinte função indicadora (**EQUAÇÃO DE SELEÇÃO**):

$$E_i = \mathbf{1}(\mathbf{Z}_i \boldsymbol{\gamma} + u_i > 0)$$

Metodologia – Modelo de seleção

- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad u_i \sim N(0, \sigma_u^2)$$

- Temos que \mathbf{Z} é um vetor de características individuais e produtivas.
- Assim, podemos representar essa variável latente pela seguinte função indicadora:

$$E_i = \mathbf{1}(\mathbf{Z}_i \boldsymbol{\gamma} + u_i > 0)$$

Função
indicadora

Metodologia – Modelo de seleção

$$W_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \quad , \quad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

- Tomando o valor esperado dessa equação, temos:

$$E(W_i | \mathbf{X}_i, E_i = 1) = E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | \mathbf{X}_i, \mathbf{Z}_i, u_i)$$

$$E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | \mathbf{Z}_i, u_i)$$

$$E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | u_i > \mathbf{Z}_i\boldsymbol{\gamma})$$

- A correção proposta por Heckman é a seguinte:

$$E(\varepsilon_i | u_i > \mathbf{Z}_i\boldsymbol{\gamma}) = \rho_{\varepsilon u} \sigma_\varepsilon \lambda_i(-\mathbf{Z}_i\boldsymbol{\gamma})$$

- $\rho_{\varepsilon u}$: correlação entre os termos de erro.
- $\lambda_i(-\mathbf{Z}_i\boldsymbol{\gamma})$: razão inversa de Mills.

Metodologia – Modelo de seleção

- A correção proposta por Heckman é a seguinte:

$$E(\varepsilon_i | u_i > \mathbf{Z}_i \boldsymbol{\gamma}) = \rho_{\varepsilon u} \sigma_{\varepsilon} \lambda_i(-\mathbf{Z}_i \boldsymbol{\gamma})$$

- $\rho_{\varepsilon u}$: correlação entre os termos de erro.
- $\lambda_i(-\mathbf{Z}_i \boldsymbol{\gamma})$: razão inversa de Mills.
- Como fazer essa correção na prática?

Metodologia – Modelo de seleção

- Como fazer essa correção na prática?
- **1º passo)** Estimar o seguinte modelo de regressão probit (**EQUAÇÃO DE SELEÇÃO**):

$$Prob(E = 1) = \Phi(\mathbf{Z}_i\boldsymbol{\gamma})$$

- E obter a estimativa da razão inversa de Mills, $\hat{\lambda}_i$.

Metodologia – Modelo de seleção

- Como fazer essa correção na prática?

- **2º passo)** Estimar por MQO a **EQUAÇÃO DE INTERESSE**, incluindo como variável independente a estimativa da razão inversa de Mills:

$$W_i = \mathbf{X}_i\boldsymbol{\beta} + \beta_\lambda \hat{\lambda}_i + \varepsilon_i$$

- Pronto! Agora você tem a equação de interesse, controlando pela possível auto seleção presente nos dados.
- β_λ mede a covariância entre os dois resíduos (ε_i e u_i) (Söderbom, 2011, pág. 12).
- Se $\hat{\beta}_\lambda$ for estatisticamente significativo podemos dizer que existe um processo de auto seleção.

Metodologia – Modelo de seleção

- A intuição do modelo de auto seleção de Heckman é que:
- No 1º estágio, estimamos a probabilidade de ser migrante dependendo de características individuais e produtivas e calculamos a razão inversa de Mills;
- No 2º estágio, estimamos a probabilidade de estar empregado dependendo de características individuais e produtivas, considerando a influência do processo de migração.

Metodologia – Modelo de seleção

- É importante não utilizar todas as variáveis independentes da equação de seleção na equação de interesse!
- Isso porque a sua inclusão leva a estimações imprecisas pelo modelo de Heckman (ver Wooldridge, 2010, pág 565).
- Essa imprecisão decorre da multicolinearidade gerada pela inclusão da razão inversa de Mills na equação.

Metodologia – Modelo de seleção

- Cuidado! O primeiro estágio de Heckman foi construído a partir de um modelo *probit*. Se você precisar utilizar outro tipo de modelo (ex: multinomial *probit/logit*), acesse: <https://core.ac.uk/download/pdf/43021578.pdf>
- Para dicas de como rodar modelos de seleção no R, acesse: http://rstudio-pubs-static.s3.amazonaws.com/108537_b7e700f5e367411a9f1431d6bb25a0fd.html

Modelo *logit* multinomial

Logit

-
- Assim como vimos na última monitoria, o logit também é uma função de transformação.

$$F(Y) = \log \left[\frac{Y}{(1 - Y)} \right]$$

- Aqui também temos que as estimativas obtidas por máxima verossimilhança não tem interpretação direta de probabilidade.
- A variação de 1 unidade de X (variável independente) influencia em β unidades o log das chances (odds) de acontecer Y (variável dependente).
- Para obter o impacto em termos de probabilidade, calculamos os efeitos marginais.

Logit

- Estimação de um modelo logit:

```
. logit pea homem
```

```
Iteration 0:  log likelihood = -223827.42
Iteration 1:  log likelihood = -217136.77
Iteration 2:  log likelihood = -217117.14
Iteration 3:  log likelihood = -217117.14
```

Logistic regression

```
Number of obs      =    337,510
LR chi2(1)         =   13420.56
Prob > chi2        =    0.0000
Pseudo R2         =    0.0300
```

Log likelihood = -217117.14

pea	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
homem	.8359028	.0073204	114.19	0.000	.821555	.8502506
_cons	.1171302	.004786	24.49	0.000	.1078099	.1265705

Esse resultado
está em termos
de log das
chances de
estar na PEA

Logit

- Existe a possibilidade de apresentar os resultados em termos de razão do log das chances (odds):

Esse resultado está em termos da RAZÃO dos logs das chances de estar na PEA. Ou seja, quais as chances dos homens de estar na PEA, em relação às chances das mulheres.

```
. logit pea homem, or
```

```
Iteration 0:    log likelihood = -223827.42
Iteration 1:    log likelihood = -217136.77
Iteration 2:    log likelihood = -217117.14
Iteration 3:    log likelihood = -217117.14
```

Logistic regression

```
Number of obs      =    337,510
LR chi2(1)         =   13420.56
Prob > chi2        =    0.0000
Pseudo R2         =    0.0300
```

Log likelihood = -217117.14

	pea	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
	homem	2.306896	.0168875	114.19	0.000	2.274033	2.340233
	_cons	1.124333	.005381	24.49	0.000	1.113836	1.134929

Logit

- Existe a possibilidade de apresentar os resultados em termos de razão do log das chances (odds):

Nesse caso, os homens têm 2,3 vezes mais chances de estarem na PEA, em relação às mulheres.

```
. logit pea homem, or
```

```
Iteration 0:  log likelihood = -223827.42
Iteration 1:  log likelihood = -217136.77
Iteration 2:  log likelihood = -217117.14
Iteration 3:  log likelihood = -217117.14
```

Logistic regression

```
Number of obs      =    337,510
LR chi2(1)         =   13420.56
Prob > chi2        =    0.0000
Pseudo R2         =    0.0300
```

Log likelihood = -217117.14

	pea	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
	homem	2.306896	.0168875	114.19	0.000	2.274033	2.340233
	_cons	1.124333	.005381	24.49	0.000	1.113836	1.134929

Logit

- Mas também temos os efeitos marginais (igual ao probit):

```
. margins, dydx(homem)
```

Average marginal effects

Number of obs = 337,510

Model VCE : OIM

Expression : Pr(pea), predict()

dy/dx w.r.t. : homem

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
homem	.1888432	.0015312	123.33	0.000	.1858422	.1918443

Os homens têm uma probabilidade de estar na PEA 18 pontos percentuais maior do que as mulheres.

Logit multinomial

- Qual a diferença das versões multinomiais (não ordenada e ordenada) do logit?
- Em termos de intuição, é só que os resultados são apresentados sempre em relação à categoria base.
- Por exemplo, suponha uma variável categórica (não ordenada) que indica:
 - Pessoa está ocupada no setor formal (valor = 1);
 - Pessoa está ocupada no setor informal (valor = 2);
 - Pessoa está descupada (valor = 3).

Logit multinomial

```
. mlogit c_ocup homem
```

Multinomial logistic regression

Number of obs = 209,866

LR chi2(2) = 1971.21

Prob > chi2 = 0.0000

Pseudo R2 = 0.0051

Log likelihood = -193475.94

c_ocup		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1	homem	.1620755	.0092406	17.54	0.000	.1439642	.1801868
	_cons	-.1326544	.0069932	-18.97	0.000	-.1463607	-.118948
2	(base outcome)						
3	homem	-.5596052	.0164232	-34.07	0.000	-.5917941	-.5274163
	_cons	-1.419561	.0108283	-131.10	0.000	-1.440784	-1.398337

Nesse caso, a categoria referência é a 2 (ocupado no setor informal)



Logit multinomial

```
. mlogit c_ocup homem
```

Multinomial logistic regression

Number of obs = 209,866

LR chi2(2) = 1971.21

Prob > chi2 = 0.0000

Pseudo R2 = 0.0051

Log likelihood = -193475.94

c_ocup	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1						
homem	.1620755	.0092406	17.54	0.000	.1439642	.1801868
_cons	-.1326544	.0069932	-18.97	0.000	-.1463607	-.118948
2	(base outcome)					
3						
homem	-.5596052	.0164232	-34.07	0.000	-.5917941	-.5274163
_cons	-1.419561	.0108283	-131.10	0.000	-1.440784	-1.398337

Os resultados
(em termos de
log das
chances) são
em relação à
categoria
referência

Logit multinomial

- Podemos apresentar em termos de probabilidades relativas:

Nesse caso, temos que a probabilidade de estar ocupado no setor formal (categoria 1) é 17,59% maior do que a probabilidade de estar ocupado no setor informal (categoria 2) para os homens, em relação às mulheres.

```
. mlogit c_ocup homem, rrr
Multinomial logistic regression

Log likelihood = -193475.94

Number of obs      =    209,866
LR chi2(2)         =    1971.21
Prob > chi2        =    0.0000
Pseudo R2         =    0.0051
```

c_ocup	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
1						
homem	1.175949	.0108665	17.54	0.000	1.154843	1.197441
_cons	.8757677	.0061244	-18.97	0.000	.8638461	.8878539
2	(base outcome)					
3						
homem	.5714346	.0093848	-34.07	0.000	.5533337	.5901277
_cons	.2418203	.0026185	-131.10	0.000	.2367422	.2470073

Logit multinomial

- Ou podemos apresentar os efeitos marginais, que são em termos absolutos:

Os efeitos marginais representam qual a probabilidade dos homens estarem na categoria X (só em comparação com as mulheres). Aqui, os homens têm uma probabilidade 6 pontos percentuais maior de estarem ocupados no mercado formal, em relação às mulheres.

```
. margins, dydx(homem)
```

```
Average marginal effects  
Model VCE      : OIM
```

```
Number of obs   =    209,866
```

```
dy/dx w.r.t. : homem  
1._predict   : Pr(c_ocup==1), predict(pr outcome(1))  
2._predict   : Pr(c_ocup==2), predict(pr outcome(2))  
3._predict   : Pr(c_ocup==3), predict(pr outcome(3))
```

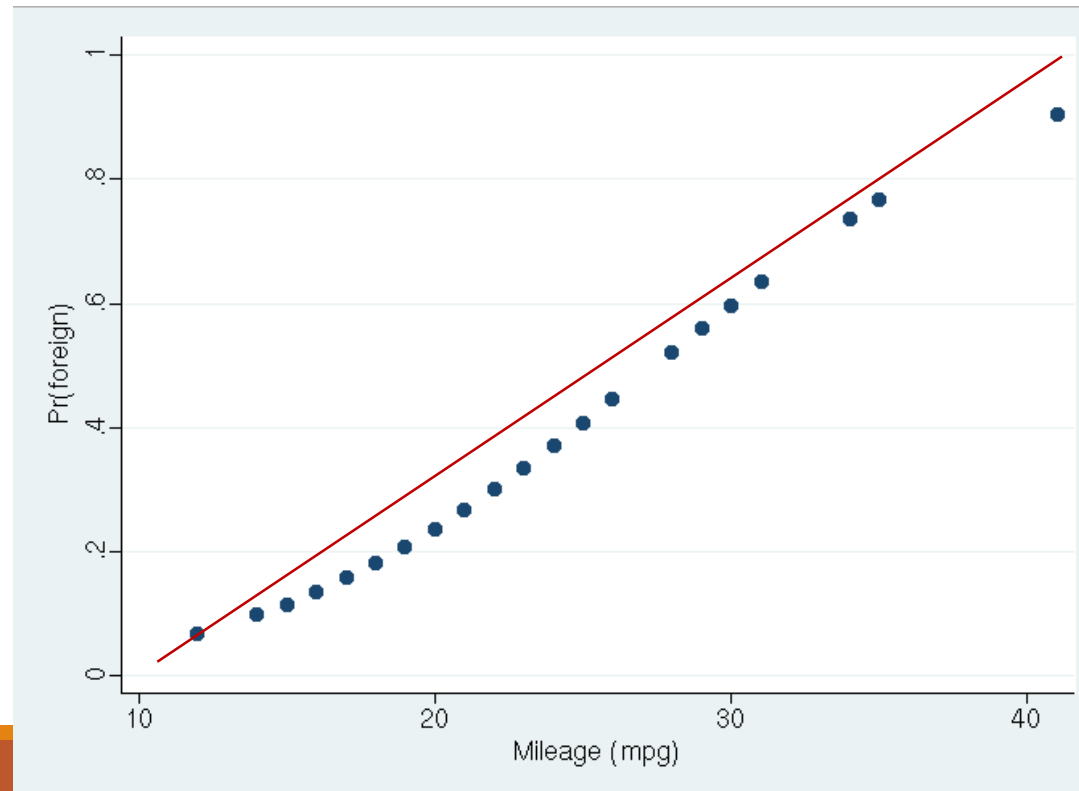
	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
homem						
predict						
1	.0610622	.0021666	28.18	0.000	.0568157	.0653087
2	-.0112886	.0021892	-5.16	0.000	-.0155793	-.0069979
3	-.0497736	.0012523	-39.75	0.000	-.0522281	-.0473191

Logit multinomial

-
- Importante! O logit e o probit são funções transformadoras limitadas entre 0 e 1. Entretanto, isso não significa que os seus efeitos marginais também estão limitados nesse intervalo!

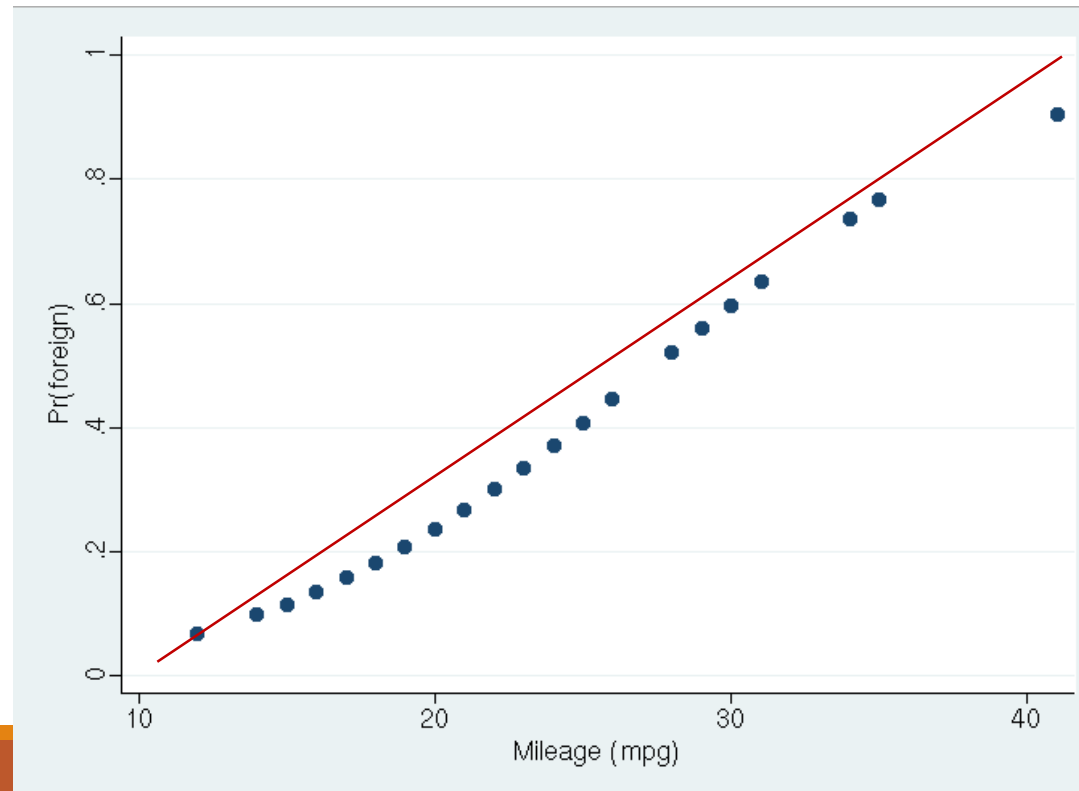
Logit multinomial

- No gráfico a seguir temos um modelo probit de regressão de “foreign” (Y) contra “mpg” (X). Os efeitos marginais (derivadas ao longo da curva de pontos) são semelhantes à inclinação da reta de 45 graus (linha vermelha).



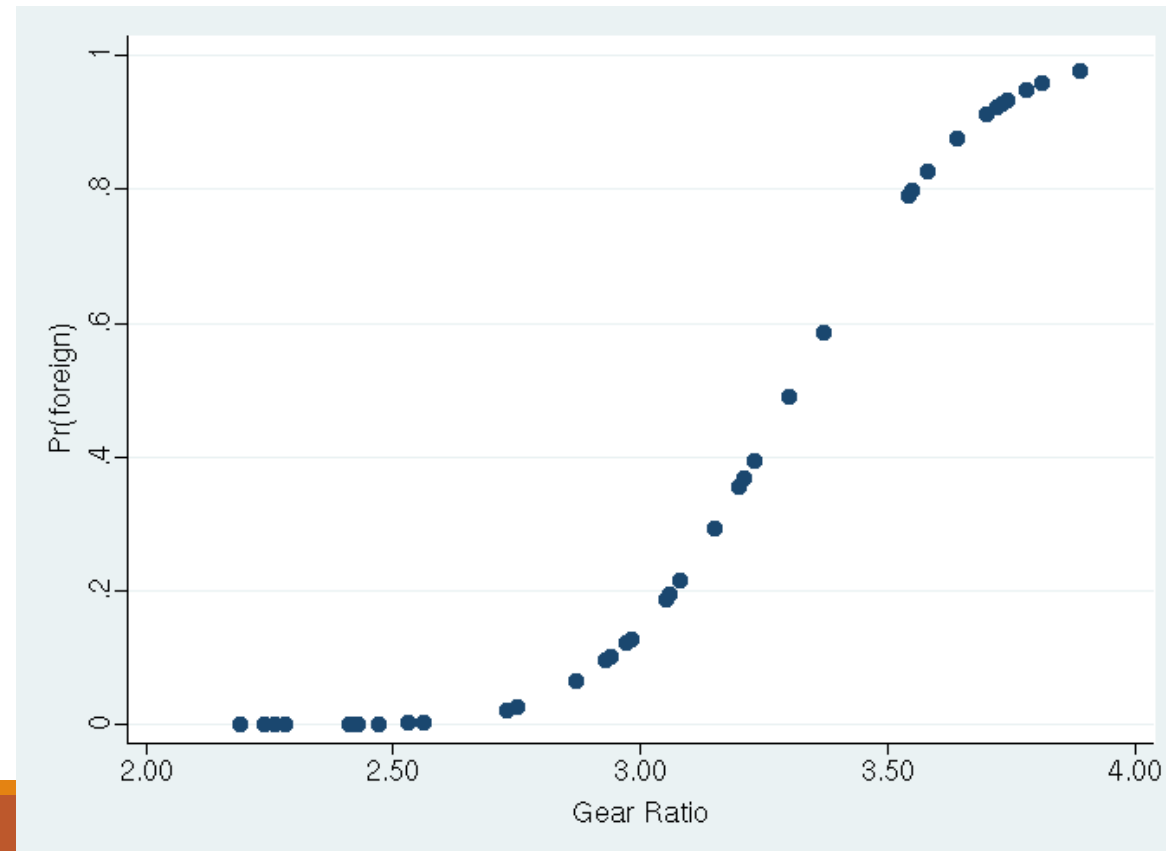
Logit multinomial

- Nesse caso, os efeitos marginais provavelmente estarão, aproximadamente, dentro do intervalo de 0 a 1.



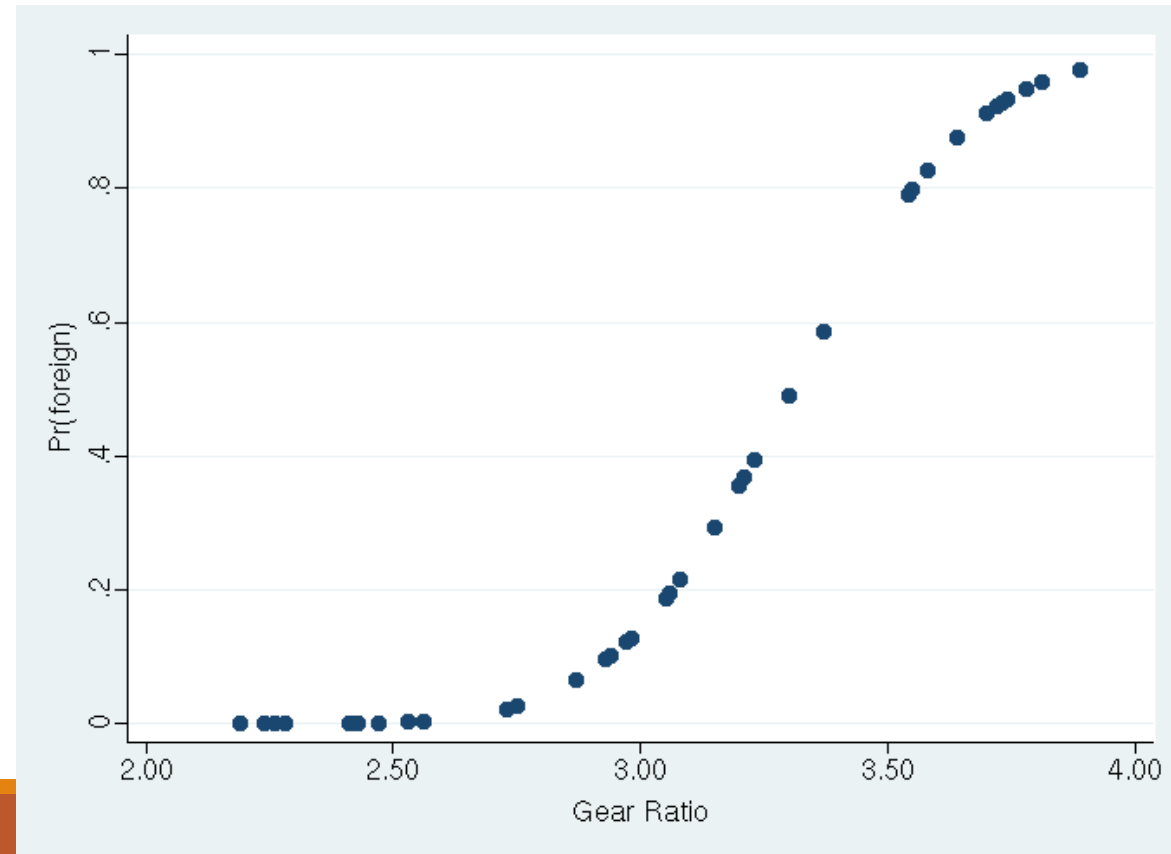
Logit multinomial

- E nesse exemplo, será que os efeitos marginais estão limitados entre 0 e 1?



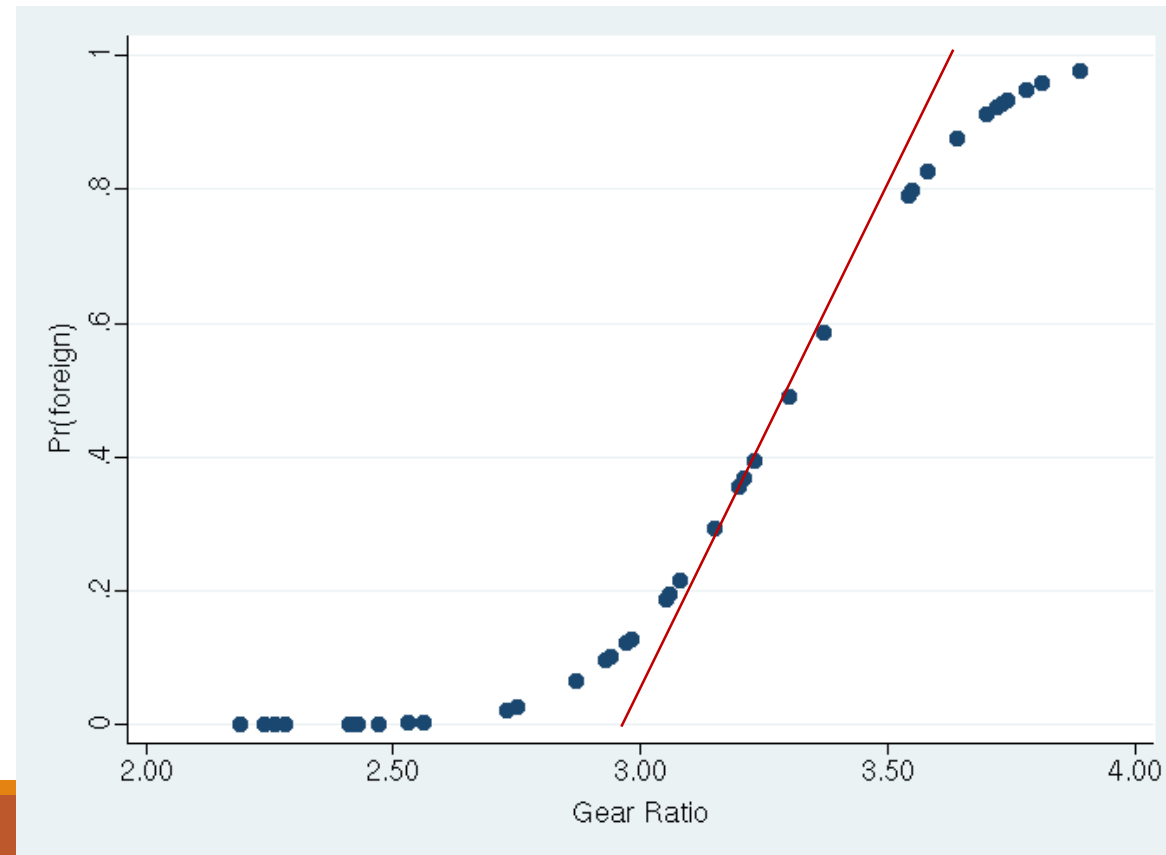
Logit multinomial

- A resposta é não!



Logit multinomial

- No ponto de “gear ratio” igual a 3,3, temos que o efeito marginal é igual a 1,38!



Resultados do artigo

Resultados do artigo

-
- Importante: os resultados a seguir são tentativas de replicação dos resultados apresentados no artigo. Existem diferenças devido a dificuldades em replicar a construção das variáveis utilizadas na análise.

Resultados do artigo

- Lembrando:
- Os autores utilizam uma variação do modelo de auto seleção de Heckman.
- 1º estágio) Estimação da probabilidade do indivíduo ser migrante, considerando características individuais e produtivas.
- 2º estágio) Estimação de um *logit* multinomial (não ordenado) da condição de ocupação do indivíduo, considerando características individuais, produtivas e a probabilidade de ser migrante (obtida no 1º estágio).

Resultados do artigo

-
- 1º estágio) Estimação da probabilidade do indivíduo ser migrante, considerando características individuais e produtivas.
 - O 1º estágio do artigo está na parte “Tabela 2” do do-file “Aula 05 – do file.do”.

Resultados – estimando a $\text{Prob}(\text{migrante} = 1)$

```
. probit migrante $x if pia == 1 & pea == 1
```

migrante	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.0600502	.1644939	0.37	0.715	-.2623519	.3824524
idade2	-.0004128	.0000253	-16.32	0.000	-.0004624	-.0003632
idltrab	-.00595	.1645562	-0.04	0.971	-.3284742	.3165741
anest	-.0192064	.1645464	-0.12	0.907	-.3417114	.3032985
sexo	-.076085	.0069374	-10.97	0.000	-.0896821	-.0624879
pref_fam	.0615444	.0072512	8.49	0.000	.0473322	.0757565
cor	.0567873	.0066636	8.52	0.000	.0437268	.0698478
trabalha	.0203683	.0119672	1.70	0.089	-.003087	.0438236
proc_trab	-.062629	.0098336	-6.37	0.000	-.0819025	-.0433556
rm	.057129	.0078848	7.25	0.000	.041675	.0725829
ar	.0703981	.0088877	7.92	0.000	.0529786	.0878176
urb	.2254975	.0102929	21.91	0.000	.2053237	.2456713
casalsf	.0236499	.009459	2.50	0.012	.0051105	.0421892
rendfampc	-.0000338	3.45e-06	-9.80	0.000	-.0000406	-.0000271
freq_esc	.0087713	.0114003	0.77	0.442	-.013573	.0311155
_cons	-2.185731	.9867122	-2.22	0.027	-4.119651	-.2518105

Resultados – estimando a $\text{Prob}(\text{migrante} = 1)$

```
. probit migrante $x if pia == 1 & pea == 1
```

migrante	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.060050				.9	.3824524
idade2	-.000412				.4	-.0003632
idltrab	-.0059				.2	.3165741
anest	-.019200				.4	.3032985
sexo	-.07608				.1	-.0624879
pref_fam	.061544				.2	.0757565
cor	.056787				.8	.0698478
trabalha	.020368				.7	.0438236
proc_trab	-.06262				.5	-.0433556
rm	.05712				.5	.0725829
ar	.070398				.6	.0878176
urb	.2254975	.0102929	21.91	0.000	.2053237	.2456713
casalsf	.0236499	.009459	2.50	0.012	.0051105	.0421892
rendfampc	-.0000338	3.45e-06	-9.80	0.000	-.0000406	-.0000271
freq_esc	.0087713	.0114003	0.77	0.442	-.013573	.0311155
_cons	-2.185731	.9867122	-2.22	0.027	-4.119651	-.2518105

Variável binária que assume valor 1 se o indivíduo é migrante (não nasceu no município atual de residência, mas mora nesse local pelo menos a 5 anos) e valor 0 caso contrário

Resultados – estimando a Prob(migrante = 1)

```
. probit migrante $x if pia == 1 & pea == 1
```

migrante	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.0600502				.023519	.3824524
idade2	-.0004128				.004624	-.0003632
idltrab	-.00595				.084742	.3165741
anest	-.0192064				.17114	.3032985
sexo	-.076085				.096821	-.0624879
pref_fam	.0615444				.173322	.0757565
cor	.0567873				.137268	.0698478
trabalha	.0203683	.0119672	1.70	0.089	-.003087	.0438236
proc_trab	-.062629	.0098336	-6.37	0.000	-.0819025	-.0433556
rm	.057129	.0078848	7.25	0.000	.041675	.0725829
ar	.0703981	.0088877	7.92	0.000	.0529786	.0878176
urb	.2254975	.0102929	21.91	0.000	.2053237	.2456713
casalsf	.0236499	.009459	2.50	0.012	.0051105	.0421892
rendfampc	-.0000338	3.45e-06	-9.80	0.000	-.0000406	-.0000271
freq_esc	.0087713	.0114003	0.77	0.442	-.013573	.0311155
_cons	-2.185731	.9867122	-2.22	0.027	-4.119651	-.2518105

Vetor de características individuais e produtivas que podem influenciar a decisão de migrar

Resultados – estimando a $\text{Prob}(\text{migrante} = 1)$

```
. probit migrante $x if pia == 1 & pea == 1
```

migrante	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.0600502	.1644939				.24524
idade2	-.0004128	.0000253				.03632
idltrab	-.00595	.1645562				.55741
anest	-.0192064	.1645464				.32985
sexo	-.076085	.0069374				.24879
pref_fam	.0615444	.0072512				.57565
cor	.0567873	.0066636	8.52	0.000	.0437268	.0698478
trabalha	.0203683	.0119672	1.70	0.089	-.003087	.0438236
proc_trab	-.062629	.0098336	-6.37	0.000	-.0819025	-.0433556
rn	.057129	.0078848	7.25	0.000	.041675	.0725829
ar	.0703981	.0088877	7.92	0.000	.0529786	.0878176
urb	.2254975	.0102929	21.91	0.000	.2053237	.2456713
casalsf	.0236499	.009459	2.50	0.012	.0051105	.0421892
rendfampt	-.0000338	3.45e-06	-9.80	0.000	-.0000406	-.0000271
frequ_esc	.0087713	.0114003	0.77	0.442	-.013573	.0311155
_cons	-2.185731	.9867122	-2.22	0.027	-4.119651	-.2518105

- Lembre que essas
estimativas representam
variações no z-score!

Resultados – estimando a Prob(migrante = 1)

. margins, dydx(\$x) atmeans vsquish

	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
idade	Com o comando margins conseguimos as estimativas em termos de probabilidade (efeitos marginais)				.0819553	.1194731
idade2					.0001444	-.0001135
idltrab					.1026112	.0988938
anest					.1067463	.0947466
sexo					.0280148	-.0195212
pref_fam					.0147866	.0236648
cor					.01366	.0218192
trabalha	.0063628	.0037384	1.70	0.089	-.0009643	.0136899
proc_trab	-.0195645	.0030717	-6.37	0.000	-.0255849	-.0135441
rm	.0178464	.0024631	7.25	0.000	.0130188	.0226739
ar	.0219915	.0027763	7.92	0.000	.01655	.0274329
urb	.0704426	.0032135	21.92	0.000	.0641442	.0767409
casalsf	.0073879	.0029548	2.50	0.012	.0015966	.0131792
rendfampc	-.0000106	1.08e-06	-9.80	0.000	-.0000127	-8.45e-06
freq_esc	.00274	.0035614	0.77	0.442	-.0042401	.0097202

Resultados – estimando a Prob(migrante = 1)

`. margins, dydx($x) atmeans vsquish`

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.0187589	.0513857	0.37	0.715	-.0819553	.1194731
idade2	-.000129	7.89e-06	-16.35	0.000	-.0001444	-.0001135
idltrab	-.0018587	.0514053	-0.04	0.971	-.1026112	.0988938
anest	-.0059998	.0514022	-0.12	0.907	-.1067463	.0947466
sexo	-.023768	.0021668	-10.97	0.000	-.0280148	-.0195212
pref_fam	.0192257	.0022649	8.49	0.000	.0147866	.0236648
cor	.0177396	.0020815	8.52	0.000	.01366	.0218192
trabalha	.0063628	.0037384	1.70	0.089	-.0009643	.0136899
proc_trab	-.0195645	.0030717	-6.37	0.000	-.0255849	-.0135441
rm	.0178464	.0024631	7.25	0.000	.0130188	.0226739
ar	.0219915	.0027763	7.92	0.000	.01655	.0274329
urb	.0704426	.0032135	21.92	0.000	.0641442	.0767409
casalsf	.0073879	.0029548	2.50	0.012	.0015966	.0131792
rendfampc	-.0000106	1.08e-06	-9.80	0.000	-.0000127	-8.45e-06
freq_esc	.00274	.0035614	0.77	0.442	-.0042401	.0097202

Resultados do artigo

- Com o comando “predict”, obtemos a estimativa da probabilidade de ser migrante.

predict p_migrante, pr

Resultados do artigo

- 2º estágio) Estimação de um *logit* multinomial (não ordenado) da condição de ocupação do indivíduo, considerando características individuais, produtivas e a probabilidade de ser migrante (obtida no 1º estágio).
- Com a probabilidade obtida no 1º estágio, os autores estimam o *logit* multinomial incluindo essa probabilidade como variável independente (parte “Tabela 04” do do-file “Aula 05 – do file”).

```
mlogit c_ocup $x2 p_migrante if pia == 1 & pea == 1, base(2)
```

Resultados do artigo

```
. mlogit c_ocup $x2 p_migrante if pia == 1 & pea == 1, base(2)
```

c_ocup	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1						
idade	.8037638	.2308201	3.48	0.000	.3513647	1.256163
idltrab	-.9214242	.2307988	-3.99	0.000	-1.373782	-.4690669
anest	-.6983944	.2308007	-3.03	0.002	-1.150756	-.2460332
sexo	.7517478	.0126374	59.49	0.000	.7269791	.7765166
cor	-.010841	.0124132	-0.87	0.382	-.0351705	.0134885
rm	.0234386	.0140304	1.67	0.095	-.0040606	.0509378
ar	-.0674255	.016207	-4.16	0.000	-.0991905	-.0356604
urb	-.2624506	.0276434	-9.49	0.000	-.3166307	-.2082705
casalsf	-.0229226	.0159733	-1.44	0.151	-.0542296	.0083845
rendfampc	.0003194	8.55e-06	37.33	0.000	.0003026	.0003361
freq_esc	-.5158789	.0176365	-29.25	0.000	-.5504458	-.4813119
p_migrante	16.74131	.336038	49.82	0.000	16.08269	17.39993
_cons	-7.838281	1.384887	-5.66	0.000	-10.55261	-5.123953
2	(base outcome)					

Resultados do artigo

3

idade	1.444102	.3025303	4.77	0.000	.8511529	2.03705
idltrab	-.7519221	.3024439	-2.49	0.013	-1.344701	-.159143
anest	-1.000605	.3024161	-3.31	0.001	-1.59333	-.4078805
sexo	-2.19636	.0257146	-85.41	0.000	-2.24676	-2.14596
cor	1.558536	.0243571	63.99	0.000	1.510797	1.606275
rm	2.055301	.0269348	76.31	0.000	2.00251	2.108092
ar	2.220085	.0314989	70.48	0.000	2.158348	2.281822
urb	7.360399	.0676676	108.77	0.000	7.227773	7.493025
casalsf	.41741	.0339178	12.31	0.000	.3509324	.4838875
rendfampc	-.002451	.0000345	-71.04	0.000	-.0025186	-.0023833
freq_esc	-.1639226	.0270444	-6.06	0.000	-.2169286	-.1109166
p_migrante	-96.05195	.7710155	-124.58	0.000	-97.56312	-94.54079
_cons	-10.94028	1.816045	-6.02	0.000	-14.49966	-7.380896

Resultados do artigo

- Esses resultados são interpretados como os resultados de um probit, com a diferença de que eles são sempre em relação à categoria base.
- O cálculo dos efeitos marginais permite calcular as probabilidades para todas as categorias separadamente!

Resultados do artigo

- Aqui temos os efeitos marginais da categoria 1 (ocupado no setor formal). Todos os resultados tem interpretação em termos de probabilidade.

```
. margins, dydx($x2 p_migrante) atmeans predict(outcome(1))
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.1872456	.0571156	3.28	0.001	.0753011	.2991901
idltrab	-.2231955	.0571104	-3.91	0.000	-.3351298	-.1112612
anest	-.1651009	.0571109	-2.89	0.004	-.2770361	-.0531656
sexo	.2086504	.0031315	66.63	0.000	.2025128	.214788
cor	-.0174392	.0030617	-5.70	0.000	-.02344	-.0114384
rm	-.0135662	.0034645	-3.92	0.000	-.0203566	-.0067759
ar	-.0378343	.0040024	-9.45	0.000	-.0456787	-.0299898
urb	-.1351598	.0068806	-19.64	0.000	-.1486455	-.1216742
casalsf	-.0096742	.0039422	-2.45	0.014	-.0174008	-.0019476
rendfampc	.000103	2.13e-06	48.32	0.000	.0000988	.0001072
freq_esc	-.1273902	.004352	-29.27	0.000	-.13592	-.1188603
p_migrante	5.092123	.0836575	60.87	0.000	4.928157	5.256089

Resultados do artigo

- Note que o efeito marginal da “probabilidade de ser migrante” é maior que 1. Conforme discutimos anteriormente, isso se deve à essa própria variável.

```
. margins, dydx($x2 p_migrante) atmeans predict(outcome(1))
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
idade	.1872456	.0571156	3.28	0.001	.0753011	.2991901
idltrab	-.2231955	.0571104	-3.91	0.000	-.3351298	-.1112612
anest	-.1651009	.0571109	-2.89	0.004	-.2770361	-.0531656
sexo	.2086504	.0031315	66.63	0.000	.2025128	.214788
cor	-.0174392	.0030617	-5.70	0.000	-.02344	-.0114384
rm	-.0135662	.0034645	-3.92	0.000	-.0203566	-.0067759
ar	-.0378343	.0040024	-9.45	0.000	-.0456787	-.0299898
urb	-.1351598	.0068806	-19.64	0.000	-.1486455	-.1216742
casalsf	-.0096742	.0039422	-2.45	0.014	-.0174008	-.0019476
rendfampc	.000103	2.13e-06	48.32	0.000	.0000988	.0001072
freq_esc	-.1273902	.004352	-29.27	0.000	-.13592	-.1188603
p_migrante	5.092123	.0836575	60.87	0.000	4.928157	5.256089

Resultados do artigo

- Para obter os efeitos marginais para as demais categorias da variável dependente, mude a opção “predict(outcome(N))”. N indica o valor da categoria.

```
. margins, dydx($x2 p_migrante) atmeans predict(outcome(1))
```

	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
idade	.1872456	.0571156	3.28	0.001	.0753011	.2991901
idltrab	-.2231955	.0571104	-3.91	0.000	-.3351298	-.1112612
anest	-.1651009	.0571109	-2.89	0.004	-.2770361	-.0531656
sexo	.2086504	.0031315	66.63	0.000	.2025128	.214788
cor	-.0174392	.0030617	-5.70	0.000	-.02344	-.0114384
rm	-.0135662	.0034645	-3.92	0.000	-.0203566	-.0067759
ar	-.0378343	.0040024	-9.45	0.000	-.0456787	-.0299898
urb	-.1351598	.0068806	-19.64	0.000	-.1486455	-.1216742
casalsf	-.0096742	.0039422	-2.45	0.014	-.0174008	-.0019476
rendfampc	.000103	2.13e-06	48.32	0.000	.0000988	.0001072
freq_esc	-.1273902	.004352	-29.27	0.000	-.13592	-.1188603
p_migrante	5.092123	.0836575	60.87	0.000	4.928157	5.256089