

Correlação espúria em modelos de painel: uma análise do desmatamento na Amazônia*

CLIVE W. J. GRANGER**
NAMWON HYUNG**

O artigo analisa o efeito de fatores específicos individuais associados a tamanho em um modelo dinâmico de regressão em painel. A teoria e simulações estatísticas mostram que esses fatores, quando variantes no tempo ou dotados de distribuições com cauda longa, podem causar correlação espúria. Se um par de variáveis em painel depende, de algum modo, do tamanho e esta variável não for introduzida na regressão, então elas deverão apresentar uma forte correlação, mesmo se forem independentes. Além disso, previsões baseadas em modelos que omitem fatores associados a tamanho são seriamente afetadas por suas propriedades. Regressões a partir de dados agrupados com séries de tempo muito curtas apresentam um bom ajustamento dentro da amostra, mas as previsões fora da amostra serão insatisfatórias se os fatores associados ao tamanho dos indivíduos omitidos forem dotados de uma distribuição com cauda longa. Usando esses resultados, aplicamos um modelo dinâmico a um painel de dados da Amazônia. O desmatamento previsto é muito menor do que em estudos anteriores quando se controla para correlação espúria.

1 - Introdução

Inicialmente, considera-se um painel que tem uma grande dimensão longitudinal (N), mas uma pequena dimensão temporal (T), de apenas um dígito. Supõe-se que as variáveis de interesse dependam de uma variável associada ao tamanho dos indivíduos constante, como área, ou variável no tempo, como população ou renda total. É comum defrontar-se com dados em painel que apresentam variáveis associadas a tamanho cujas distribuições são dispersas ou dotadas de cauda longa, como, por exemplo, em um painel de países que incluisse Luxemburgo e Chipre ou Índia e China. Se $X_{j,t}$ e $Y_{j,t}$ formam um par de variáveis medidas na região j , no tempo t , e se cada uma depende de alguma forma do tamanho, então uma regressão de $X_{j,t}$ em $Y_{j,t}$ vai apresentar uma forte correlação, se a variável associada a tamanho não for incluída na regressão, mesmo se as variáveis forem independentes. Esse seria um exemplo de uma relação espúria decorrente de má especificação. Existem, é claro, várias maneiras bem conhecidas de lidar com esse problema,

* Os autores agradecem os comentários de um parecerista anônimo.

** Do Departamento de Economia da Universidade da Califórnia.

que são discutidas a seguir. Contudo, o problema é mais complicado se a variável associada a tamanho varia no tempo ou se o modelo possui termos dinâmicos. Uma regressão de $X_{j,t}$ em $X_{j,t-1}$, na qual a variável associada a tamanho é constante ou muda muito lentamente, apresenta uma alta correlação entre $X_{j,t}$ e sua defasagem, sugerindo, incorretamente, uma raiz unitária. Este artigo investiga a ocorrência do fenômeno de “correlação espúria”, principalmente através de simulações, mas também usando alguma teoria. Considera-se os efeitos de soluções clássicas alternativas para o problema, com atenção especial para suas implicações sobre a qualidade das previsões de longo prazo.

Assim sendo, a Seção 2 explica como resultados espúrios poderiam ocorrer em modelos de regressão mal-especificados e a Seção 3 aplica métodos similares em regressões dinâmicas em painel para ilustrar a correlação espúria. Nas Seções 4 e 5, adotam-se métodos de Monte Carlo para investigar esses problemas, usando-se os resultados das simulações para estimação, testes de raiz unitária e comparação de previsões. A Seção 6 examina as previsões de desmatamento utilizando uma base de dados da Amazônia e, por último, a Seção 7 apresenta as conclusões.

2 - Regressão de duas séries de painel

Suponha que S_j é uma variável associada ao tamanho dos indivíduos (tal como área, população ou renda total) e que $X_{j,t}$ e $Y_{j,t}$ formam um par de variáveis, para a região j e tempo t , dado por:

$$X_{j,t} = f_j(S_j) + \tilde{X}_{j,t} \quad (1)$$

$$Y_{j,t} = Af_j(S_j) + \tilde{Y}_{j,t}$$

onde \tilde{X} , \tilde{Y} não são associados a tamanho e $f_j(S_j)$ é uma função do tamanho.

Se rodarmos uma regressão em painel:

$$Y_{j,t} = \alpha + \beta X_{j,t} + e_{j,t} \quad (2)$$

então, se a variável associada ao tamanho tiver uma amplitude larga¹ (por exemplo, o painel inclui tanto países muito pequenos quanto países muito grandes), obteremos:

- a) $\hat{\beta} \cong A$;
- b) as estatísticas- t de $\hat{\beta}$ serão significativas; e
- c) R^2 grande.

¹ Essa condição é equivalente a $\text{corr}(X_{j,t}, f_j(S_j))$ próxima de um. Sob esta condição, $\hat{\beta} \cong A$, uma vez que $\hat{\beta} = A \text{Cov}(X_{j,t}, f_j(S_j)) / \text{Var}(X_{j,t}) = A [\text{corr}(X_{j,t}, f_j(S_j))]^2$.

Quer dizer, uma correlação aparente (ver Apêndice A para detalhes).

Poderíamos caracterizar esse resultado como uma relação espúria se a regressão a seguir tivesse $R^2 = 0$:

$$\tilde{Y}_{j,t} = \alpha + \beta \tilde{X}_{j,t} + \tilde{e}_{j,t} \quad (3)$$

A relação aparente (2) ocorre por causa da omissão da variável S_j na regressão. No entanto, termos como $f_j(S_j)$ são de fato efeitos fixos (porque eles ou não mudam com o tempo ou mudam muito lentamente). Poderíamos melhorar (2) incluindo esses efeitos e depois rodando a regressão:

$$Y_{j,t} = \alpha_j + \beta X_{j,t} + e_{j,t} \quad (4)$$

Se existem N regiões, temos agora de estimar $N+1$ parâmetros em (4) — fora as variâncias —, enquanto em (2) há apenas dois parâmetros. Se a base de dados for tal que $j = 1, \dots, N$ e $t = 1, \dots, T$, com N grande e T pequeno, então os α_j 's serão deficientemente estimados e a qualidade da previsão será ruim nesse caso.²

Suponha que pudéssemos “ajustar para o tamanho” cada variável:³

$$\begin{aligned} X_{j,t}^{SA} &= X_{j,t} - f(\hat{S}_j) \\ Y_{j,t}^{SA} &= Y_{j,t} - A f(\hat{S}_j) \end{aligned} \quad (5)$$

onde $\hat{\cdot}$ indica estimador. Agora poderíamos rodar uma regressão como (3):

$$Y_{j,t}^{SA} = \hat{\alpha} + \hat{\beta} X_{j,t}^{SA} + \hat{e}_{j,t} \quad (6)$$

na qual $\hat{\beta} \cong 0$ e $R^2 \cong 0$. Em certo sentido, a equação (2) seria uma regressão espúria, uma vez que a inferência poderia sugerir a existência de uma relação que, na verdade, não existe.

2 Na Tabela 5, o erro quadrático médio de previsão de (4) quando $h = 1$ e $T = 5$ é maior que em outros modelos.

3 Na teoria poderia considerar-se a decomposição em duas partes e estimar funções desconhecidas das variáveis de tamanho por algum método de suavização [ver Engle *et alii* (1986)] ou outro método não-paramétrico. Na prática, uma vez que não conhecemos a variável associada a tamanho nem sabemos como mensurá-la, não seria fácil fazer um ajustamento para o tamanho. Neste artigo não investigamos esse assunto mais a fundo. Nós o deixamos para pesquisas futuras.

Quando a variável associada ao tamanho não é uma constante, a situação torna-se mais complicada. Se termos como $f_j(S_{j,t})$ mudassem muito lentamente, ainda assim ocorreria uma aparente relação espúria em (4) devido à variável omitida nessa regressão, não incluída nos efeitos fixos.

Entorf (1996), investigando regressões espúrias em modelos de painel, mostrou que, ao se aplicar uma abordagem de Mínimos Quadrados Ordinários (MQO) usando variáveis *dummies* a um modelo de efeitos fixos erradamente conjecturados, quando, de fato, o Processo Gerador de Dados (PGD) para as duas séries são passeios aleatórios com tendência independentes, deve-se obter um modelo de efeitos fixos espúrio. Nós aplicamos um método similar, mas as hipóteses sobre o PGD para as duas séries são diferentes das de Entorf. Consideramos que as duas séries têm um fator comum associado a tamanho. Nesse caso, o modelo de efeitos fixos (4) também teria um problema de má especificação. Então esse modelo não seria capaz de captar corretamente o fator associado a tamanho variável no tempo e apresentaria uma relação espúria, $\beta \cong A$, entre duas séries por causa desse fator. Esse resultado é resumido a seguir.

Suponha que (4) é estimado por MQO. Se a variável associada a tamanho em (1) cresce com o tempo estocasticamente (passeio aleatório com tendência), sob as hipóteses (A.1) – (A.5), então a distribuição assintótica do estimador MQO de β pode ser escrita como:

$$T^{1/2} (\hat{\beta} - A) \Rightarrow N \left(0, \frac{12 \sum_{j=1}^N \delta_j^2 (\sigma_{y,j}^2 + A^2 \sigma_{x,j}^2)}{\left(\sum_{j=1}^N \delta_j^2 \right)} \right) \quad (7)$$

A comprovação desse resultado está no Apêndice B.

Suponha que a variável associada a tamanho seja um processo estacionário em torno de uma tendência (*trend stationary*), $S_{j,t} = \delta_{j0} + \delta_{jt}t + \eta_{j,t}$, sujeito às hipóteses (A.1) – (A.5); então a distribuição assintótica do estimador MQO de β em (4) também será dada por (7). Essa proposição pode ser provada de forma análoga.

3 - Correlação espúria em um modelo dinâmico de painel

Para um modelo particular suponha que as variáveis explicativas de $Y_{j,t}$ sejam um vetor $X_{j,t}$, que pode incluir $Y_{j,t}$ defasado. Nesta seção admite-se que $X_{j,t}$ consiste somente em $Y_{j,t-1}$, pois o princípio da análise permanece o mesmo, mas as demonstrações são bastante simplificadas.

Seja $Y_{j,t}$ gerado como acima. Se regredirmos $Y_{j,t}$ em $Y_{j,t-1}$:

$$Y_{j,t} = m + \rho Y_{j,t-1} + e_{j,t} \quad (8)$$

então obteremos uma raiz unitária espúria (ver Apêndice A) e:

- a) $\hat{\rho} \cong 1$; ⁴
- b) as estatísticas- t de $\hat{\rho} = 1$ serão significativas;
- c) $\hat{m} \cong 0$; e
- d) $R^2 \cong 1$

Quando há uma raiz unitária espúria no sistema, o modelo resultante vai se ajustar bem aos dados, mas deverá gerar previsões ruins fora da amostra. Note-se que se $Y_{j,t-1}$ fosse substituído por $Y_{j,t-k}$, obteríamos resultados semelhantes, mas com uma dinâmica diferente. Devido ao fato de os dados não possuírem uma raiz unitária, esse fenômeno induzido pelo tamanho pode ser chamado de “correlação espúria”. Se as variáveis associadas ao tamanho $f_j(S_j)$ fossem efeitos fixos, poder-se-ia melhorar o modelo (8) através da inclusão de uma variável *dummy*, tal como em (4):

$$Y_{j,t} = m_j + \rho Y_{j,t-1} + e_{j,t} \quad (9)$$

Incorrer-se-ia, pois, nos mesmos problemas decorrentes de N grande e T pequeno. Haveria ainda um problema adicional causado por séries de tempo curtas em um modelo dinâmico em painel com efeitos específicos ao indivíduo, isto é, um viés para baixo⁵ de $\hat{\rho}$.

Quando o fator associado a tamanho for variável no tempo, ainda assim haverá uma aparente relação espúria em (9) por causa da variável omitida da regressão, não incluída nos efeitos fixos. Suponha que (9) seja estimado por MQO. Se a variável associada ao tamanho cresce com o tempo estocasticamente (passeio aleatório com tendência), então

4 Quando o PGD é (A.2'), se $\text{corr}(Y_{jt}, f_j(S_j)) \cong 1$, então $\rho \cong 1$, na medida em que $\rho = \rho_0 + (1 - \rho_0) \text{Cov}(Y_{jt}, f_j(S_j)) / \text{Var}(Y_{jt}) = \rho_0 + (1 - \rho_0) [\text{corr}(Y_{jt}, f_j(S_j))]^2$.

5 Ver Nickell (1981) para este detalhe. Ele calculou o viés para T pequeno. Seja o PGD:

$$Y_{j,t} = m_j + \rho Y_{j,t-1} + e_{j,t} = \frac{m_j}{1-\rho} + \sum_{i=0}^{\infty} \rho^i e_{j,t-i}$$

Então, o viés assintótico do estimador de mínimos quadrados de ρ é:

$$p \lim_{N \rightarrow \infty} (\hat{\rho} - \rho) = -\frac{1+\rho}{T-1} \left(1 - \frac{1-\rho^T}{T(1-\rho)} \right) \left\{ 1 - \frac{2\rho}{(1-\rho)(T-1)} \left[1 - \frac{1-\rho^T}{T(1-\rho)} \right] \right\}^{-1}$$

$\hat{\rho} \cong 1$. A prova disso decorre diretamente do Apêndice B, se admitirmos que a dinâmica de $Y_{j,t}$ é gerada por (A.2'), (A.4) e (A.5') e que $A = 1$, sem perda de generalidade.

4 - Estimação e testes de raiz unitária em modelos de painel: um estudo de simulação

Nas simulações apresentadas consideramos dois tipos de fatores específicos aos indivíduos, constantes e variáveis no tempo, como PGD. Uma variável associada a tamanho, como área, pode ser representada por um fator constante, enquanto população e renda total são classificadas como fatores variáveis no tempo. Nesta seção, apresentam-se resultados de estimação espúrios, derivados da má especificação causada pela omissão dos efeitos específicos aos indivíduos. Os processos geradores de dados $\{Y_{j,t}\}$ são definidos como:

- PGD-1 (dados em painel com efeitos específicos ao indivíduo constantes):

$Y_{j,t} = S_{j,t} + \tilde{Y}_{j,t}$, onde $\tilde{Y}_{j,t} = \rho \tilde{Y}_{j,t-1} + \varepsilon_{j,t}$,⁶ $\varepsilon_{j,t} \sim i.i.d.N(0, 1)$, e $S_j = S_{j,t} \sim \chi_{(6)}^2$, de modo que, entre regiões, a variável associada a tamanho constante é retirada de uma distribuição χ^2 ; e

- PGD-2 (dados em painel com efeitos específicos ao indivíduo variáveis no tempo):
PGD-1 com $S_{j,t} = \delta + S_{j,t-1} + \eta_{j,t}$, $S_{j0} \sim \chi_{(6)}^2$, $\delta = 0,05$ e $\eta_{j,t} \sim i.i.d.N(0; 0,1)$.

Três modelos de regressão são considerados: modelo 1 — $Y_{j,t} = \rho Y_{j,t-1} + e_{j,t}$; modelo 2 — $Y_{j,t} = m + \rho Y_{j,t-1} + e_{j,t}$; e modelo 3 — $Y_{j,t} = m_j + \rho Y_{j,t-1} + e_{j,t}$.

Os resultados das regressões em painel com fatores específicos ao indivíduo constantes são reportados na Tabela 1, que fornece valores médios da simulação baseados em amostras de tamanho $N = 100$ e $T = 5, 10, 25$ e 100 , em 100 replicações. As estimativas dos modelos 1 e 2 mostram uma clara tendência aos limites assintóticos calculados no Apêndice A. Quando $T = 5$, os modelos 1 e 2 ajustam-se muito bem em termos de R^2 , apesar de serem mal-especificados — como esperado a partir da Seção 3. Mas, fora da amostra, eles apresentam muitos erros de previsão (ver Tabela 5). Em modelos mal-especificados, que ignoram fatores específicos ao indivíduo constantes, $\hat{\rho}$ aproxima-se da unidade. Aqui, uma simples inspeção de $\hat{\rho}$ evidencia o problema de correlação espúria. Para examinar essa hipótese estatisticamente, uma análise do poder dos testes de raiz unitária é conduzida mais adiante nesta seção. A partir do modelo 3, vê-se que o problema da variável omitida pode ser resolvido usando-se efeitos fixos, apesar de ainda ocorrer um viés de pequenas amostras em $\hat{\rho}$. Para um T muito pequeno, ainda que o modelo seja correto, ele apresenta um forte viés para baixo, como apontado por Nickell (1981) e uma

6 Para evitar o problema de se fixar Y_{j0} , 20 pré-amostras são geradas em alguns casos. Exceto no caso dos testes de poder, os resultados não são muito diferentes do que quando se fixa $Y_{j0} = 0$.

TABELA 1

Estimação de dados em painel com efeitos específicos aos indivíduos fixos

Modelos	$\rho_0 = 0,7$	$T = 5$	$T = 10$	$T = 25$	$T = 100$	Asym ^a
1	ρ	0,9869	0,9877	0,9883	0,9883	0,9882
	s.e.	0,0076	0,0051	0,0031	0,0015	-
	R^2	0,9773	0,9763	0,9765	0,9769	-
2	m	0,2528	0,2544	0,2579	0,2586	0,2528
	s.e.	0,1015	0,0682	0,0419	0,0206	-
	ρ	0,9564	0,9570	0,9572	0,9571	0,9579
3	s.e.	0,0143	0,0096	0,0059	0,0029	-
	R^2	0,9176	0,9167	0,9161	0,9163	0,9176
	m_j	-	-	-	-	-
3	ρ	0,2177	0,4789	0,6254	0,6818	0,7000
	s.e.	0,0564	0,0310	0,0163	0,0074	-
	R^2	0,9568	0,9395	0,9307	0,9281	-
	ρ^b	0,3060	0,5034	0,6262	0,6826	0,7000

NOTAS: 1) Os limites assintóticos do modelo 1 podem ser derivados como:

$$\rho \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \hat{\rho} = \rho_0 + (1 - \rho_0) (\mu_s^2 + \sigma_s^2) / \left(\frac{\sigma_e^2}{1 - \rho_0^2} + \mu_s^2 + \sigma_s^2 \right)$$

2) Valores médios de 100 replicações, 100 indivíduos.

^a Os limites assintóticos ($T \rightarrow \infty$ e $N \rightarrow \infty$) são calculados pelas fórmulas (A.7)-(A.12).

^b Aproximação para pequenas amostras através da fórmula de Nickell da nota de rodapé 5.

capacidade preditiva muito ruim, como se pode ver na Tabela 5. O impacto no desempenho das previsões desses modelos é discutido na Seção 5.

Regressões em painel com fatores específicos ao indivíduo variáveis no tempo são consideradas na Tabela 2. Uma vez que os fatores associados a tamanho são variáveis no tempo, o modelo 3 também tem um problema de variável omitida. Apesar de o fator específico ao indivíduo ter uma variância muito pequena, igual a 0,1, ele domina a estimação em grandes amostras, levando $\hat{\rho}$ à unidade. Sem se remover esse fator, não se pode obter a verdadeira estrutura dinâmica de \tilde{X} , que não é relacionada com a componente associada a tamanho. Em pequenas amostras, os valores estimados têm um viés ligeiramente maior em todos os modelos do que os valores estimados a partir de PGD-1, uma vez que PGD-2 contém uma componente de passeio aleatório. Contudo, tendo em vista que todos os modelos são mal-especificados nesse caso, eles têm uma capacidade

TABELA 2

*Estimação de dados em painel com efeitos específicos
aos indivíduos variáveis no tempo*

Modelos	$\rho_0 = 0,7$	$T = 5$	$T = 10$	$T = 25$	$T = 100$
1	ρ	0,9965	0,9958	0,9959	0,9983
	s.e.	0,0078	0,0052	0,0030	0,0012
	F^2	0,9094	0,9103	0,9160	0,9390
2	m	0,2689	0,2871	0,2991	0,2879
	s.e.	0,1076	0,0724	0,0451	0,0238
	ρ	0,9641	0,9619	0,9622	0,9720
3	s.e.	0,0151	0,0099	0,0059	0,0025
	F^2	0,9109	0,9120	0,9176	0,9399
	$\frac{1}{N} \sum m_j$	5,0672	3,1829	2,0872	1,0328
3	ρ	0,1827	0,5001	0,6935	0,8846
	s.e.	0,0584	0,0313	0,0152	0,0047
	F^2	0,9531	0,9346	0,9292	0,9427

NOTAS: 1) Valores médios de 100 replicações, 100 indivíduos.

2) $\delta = 0,05$ e $\text{Var}(\mu_{jt}) = 0,1$.

preditiva, vários passos à frente, muito diferente da de PGD-1, como será mostrado na Seção 5.

De acordo com as Tabelas 1 e 2, modelos em painel para fatores específicos ao indivíduo mal-especificados devem produzir “correlação espúria”. Para examinar essa hipótese estatisticamente, conduzimos uma análise do poder dos testes de raiz unitária de painel sugeridos por Levin e Lin (1992). Inicialmente, o poder dos testes é considerado na presença de erros *i.i.d.* e os valores críticos para os testes são dados por Levin e Lin, já mencionados.

Para fatores específicos ao indivíduo fixos, o teste de raiz unitária de painel para o modelo 2 tem baixo poder no caso de uma pequena amostra e um poder apropriado no caso de uma grande amostra. O baixo poder para o teste de raiz unitária de painel em uma pequena amostra sugere que séries em painel estacionárias poderiam ser incorretamente especificadas como processos de raiz unitária com probabilidade não-nula. O problema pode persistir em amostras maiores. Se a variância do fator associado a tamanho também for maior, digamos $\chi^2_{(6)} \rightarrow \chi^2_{(30)}$, então se encontrará correlação espúria com maior frequência, como indicado na Tabela 3, ou seja, menos poder para o caso de maior variância. O modelo 3, apesar de apresentar um severo viés para baixo, tem um

TABELA 3

Análise do poder do teste de raiz unitária de painel(Dados em painel com efeitos específicos aos indivíduos fixos, $\rho_0 = 0$)

Modelos	Tamanho (%)	$T = 5$	$T = 10$	$T = 25$
1	1	27,5 (0,0)	95,9 (0,0)	100,0 (0,0)
	5	74,5 (1,0)	100,0 (0,0)	100,0 (0,0)
	10	92,0 (5,0)	100,0 (2,8)	100,0 (15,8)
2	1	99,1 (13,2)	100,0 (75,2)	100,0 (100)
	5	100,0 (57,9)	100,0 (99,3)	100,0 (100)
	10	100,0 (80,5)	100,0 (100)	100,0 (100)
3	1	100,0 (100)	100,0 (100)	100,0 (100)
	5	100,0 (100)	100,0 (100)	100,0 (100)
	10	100,0 (100)	100,0 (100)	100,0 (100)

Notas: 1) Valores críticos do modelo 1: Tabela 1 de Levin e Lin (1992).

2) Valores críticos do modelo 2: Tabela 2 de Levin e Lin (1992).

3) Valores críticos do Modelo 3: Tabela 5 de Levin e Lin (1992).

4) 1.000 replicações geradas pelo PGD-1.

5) $S_j \sim \chi_{(6)}^2$; os valores entre parênteses são calculados quando $S_j \sim \chi_{(30)}^2$.

poder apropriado em pequenas amostras. Esse modelo não depende da variância do fator associado a tamanho, uma vez que os termos específicos ao indivíduo podem remover o efeito desse fator.

Considerando-se fatores específicos ao indivíduo variáveis no tempo, a partir da Tabela 4, vê-se que o poder do teste é menor do que no caso de fatores fixos (Tabela 3) porque o fator associado ao tamanho variável no tempo possui uma componente de raiz unitária fraca. Quanto maior a variância da componente associada a tamanho, menor o poder do teste, como era esperado. Quando T é grande, o termo constante na equação da componente associada a tamanho será uma componente dominante se $\text{Var}(\eta_{jt})$ for maior. Nesse caso, as séries individuais mostram uma clara tendência temporal, de modo que se prefere um modelo que contenha essa tendência.

TABELA 4

Análise do poder do teste de raiz unitária de painel(Dados em painel com efeitos específicos aos indivíduos variáveis no tempo, $\rho_0 = 0$)

Modelos	Tamanho (%)	$T = 5$	$T = 10$	$T = 25$
1	1	3,6 (4,2)	19,6 (13,9)	95,2 (46,9)
	5	29,0 (23,4)	75,1 (45,3)	99,7 (81,3)
	10	54,1 (41,8)	93,6 (67,3)	100,0 (92,7)
2	1	98,5 (80,7)	100,0 (98,9)	100,0 (100)
	5	99,9 (95,3)	100,0 (99,9)	100,0 (100)
	10	100,0 (98,0)	100,0 (100)	100,0 (100)
3	1	100,0 (94,4)	100,0 (100)	100,0 (100)
	5	100,0 (97,4)	100,0 (100)	100,0 (100)
	10	100,0 (98,2)	100,0 (100)	100,0 (100)

NOTAS: 1) Valores críticos do modelo 1: Tabela 1 de Levin e Lin (1992).

2) Valores críticos do modelo 2: Tabela 2 de Levin e Lin (1992).

3) Valores críticos do modelo 3: Tabela 5 de Levin e Lin (1992).

4) 1.000 replicações geradas pelo PGD-2.

5) Poder quando $\text{Var}(\eta_{jt}) = 0,1$; os valores entre parênteses são calculados quando $\text{Var}(\eta_{jt}) = 1$.**5 - Comparação de previsões do painel**

Uma das maneiras de se julgar a relevância de um modelo é avaliar suas previsões confrontando-as com as de outros modelos. Nesta seção compara-se o desempenho relativo de previsão dos três modelos apresentados. Explicam-se sucintamente as técnicas de avaliação de previsões de modelos em painel examinadas em Granger (1996) e depois se comparam os três modelos em painel usando simulações.

5.1 - Avaliação de previsões de modelos em painel

Considere um par de modelos que produzem previsões e, por conseguinte, erros de previsão. Sejam $e_{jt}^{(1)}, e_{jt}^{(2)}, j = 1, \dots, N$ e $t = 1, \dots, T$ os erros de previsão t períodos à frente dos modelos (1) e (2), respectivamente. Se as funções-perda são quadráticas e os erros

de previsão são não-viesados e normalmente distribuídos, então a hipótese nula de precisões de previsão iguais corresponde a variâncias do erro de previsão iguais:

$$H_0 : E [(e_{jt}^{(1)})^2] = E [(e_{jt}^{(2)})^2]$$

Defina $S_{jt} = e_{jt}^{(1)} + e_{jt}^{(2)}$ e $D_{jt} = e_{jt}^{(1)} - e_{jt}^{(2)}$. Um teste da hipótese nula antes mencionada é facilmente conduzido usando S_{jt} e D_{jt} , pois $E [(e_{jt}^{(1)})^2] = E [(e_{jt}^{(2)})^2]$ somente quando $\text{COV}(S_{jt}, D_{jt})$ é zero:

$$H_0 : \text{Cov}(S_{jt}, D_{jt}) = 0$$

Existe um teste bem conhecido (“teste do sinal”) que se baseia em p_1 e p_2 , onde p_1 é o número de vezes que $e_{jt}^{(1)}$ é maior que $e_{jt}^{(2)}$ e p_2 é o número de vezes que $e_{jt}^{(2)}$ é maior que $e_{jt}^{(1)}$. O número total de comparações é $Q = p_1 + p_2$. Sob a hipótese nula, a estatística do teste do sinal é:

$$\left(\frac{p_1}{Q} - \frac{1}{2} \right) / \sqrt{1/4Q} \sim N(0, 1) \quad (10)$$

Um teste alternativo da mesma hipótese nula é uma extensão do teste da *soma e diferença* [Granger e Newbold (1986, p. 279)]. Sob a hipótese de que a previsão individual é não-viesada e os erros de previsão são não-autocorrelacionados e independentes entre indivíduos, o teste da *soma e diferença* pode ser estendido a modelos de painel. A estatística do teste é:

$$\frac{\sqrt{NT-1}}{2} \log \left(\frac{1+r}{1-r} \right) \sim N(0, 1) \quad (11)$$

onde:

$$r = \frac{\sum_{j=1}^N \sum_{t=1}^T (e_{jt}^{(1)} + e_{jt}^{(2)}) (e_{jt}^{(1)} - e_{jt}^{(2)})}{\left[\sum_{j=1}^N \sum_{t=1}^T (e_{jt}^{(1)} + e_{jt}^{(2)})^2 \sum_{j=1}^N \sum_{t=1}^T (e_{jt}^{(1)} - e_{jt}^{(2)})^2 \right]^{1/2}}$$

Para horizontes múltiplos de previsão (h passos à frente), seqüências de erros de previsão vão seguir um processo de médias móveis (MA) de ordem ($h-1$), violando a hipótese de que os erros de previsão são serialmente não-correlacionados, como no caso do teste da *soma e diferença*. Utilizando-se o teste de Meese-Rogoff, a hipótese de não-correlação serial pode ser relevada [ver Meese e Rogoff (1988)]. Esse teste pode ser estendido para modelos em painel com um N dado. Em se tratando de um painel, está fora da realidade se postular independência entre observações. Um método estendido pode ser aplicado a qualquer dependência arbitrária entre indivíduos (ver Apêndice C):

$$\frac{\hat{\gamma}_{SD}}{\sqrt{\sum / NT}} \sim N(0, 1) \quad (12)$$

onde:

$$\hat{\gamma}_{SD} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T S_{it} D_{it}$$

$$\hat{\Sigma} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{\tau=-S(T)}^{S(T)} \left(1 - \frac{|\tau|}{T} \right) \left\{ \hat{\gamma}_{S,S_j}(\tau) \hat{\gamma}_{D,D_j}(\tau) + \hat{\gamma}_{S,D_j}(\tau) \hat{\gamma}_{D,S_j}(\tau) \right\} \right]$$

$$\hat{\gamma}_{S,S_j}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} S_{it} S_{jt+\tau} \quad \hat{\gamma}_{S,D_j}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} S_{it} D_{jt+\tau}$$

e:

$$\hat{\gamma}_{D,S_j}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} D_{it} S_{jt+\tau} \quad \hat{\gamma}_{D,D_j}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} D_{it} D_{jt+\tau}$$

sendo que a defasagem truncada $S(T)$ cresce com o tamanho da amostra, mas a uma taxa menor.

Todos esses testes têm uma distribuição assintótica normal, mas em pequenas amostras eles terão propriedades distintas. No caso univariado, Diebold e Mariano (1995) mostram que o teste de Meese-Rogoff é robusto quanto à correlação serial em grandes amostras, mas superestima o tamanho em pequenas amostras na presença de correlação serial. O teste da *soma e diferença* será afetado pela autocorrelação serial, que induz a

um tamanho empírico acima do tamanho nominal; contudo, em pequenas amostras, este é melhor que o teste de Meese-Rogoff. Reportam-se apenas os resultados do teste de Meese-Rogoff Generalizado, mas os resultados dos testes do sinal e da *soma e diferença* são similares.⁷

5.2 - Previsões do modelo de painel: um estudo de simulação

Para investigar o efeito da variância e a constância da equação da componente associada a tamanho, simulam-se vários tipos de fatores específicos ao indivíduo variáveis no tempo. Os resultados, nas Tabelas 5, 6 e 7, mostram o erro quadrático médio dos três

TABELA 5

Comparação das previsões de modelos de painel com efeitos específicos aos indivíduos constantes

	EQM de Previsão			1 vs. 2 GMR	2 vs. 3 GMR	1 vs. 3 GMR
	Modelo 1	Modelo 2	Modelo 3			
<i>h</i> = 1						
<i>T</i> = 5	1,1777	1,1620	2,2860	0*(96**)	100(0)	94(0)
<i>T</i> = 10	1,1759	1,1578	1,3537	0(100)	90(0)	95(0)
<i>T</i> = 25	1,1755	1,1571	1,0784	0(100)	0(77)	0(99)
<i>h</i> = 5						
<i>T</i> = 5	3,2927	3,2384	3,3280	22(76)	57(39)	56(43)
<i>T</i> = 10	3,2493	3,1646	2,8357	9(91)	5(83)	2(98)
<i>T</i> = 25	3,2323	3,1477	2,2618	0(100)	0(92)	0(100)
<i>h</i> = 20						
<i>T</i> = 5	6,0579	6,7610	3,3688	65(35)	0(98)	0(100)
<i>T</i> = 10	5,5678	6,3906	2,9222	87(13)	0(97)	0(100)
<i>T</i> = 25	5,3186	6,2890	2,4048	99(1)	0(94)	0(100)

NOTAS: 1) Valor médio de 100 replicações, 100 cross-sections, tamanho fora da amostra igual a 50 e $\rho_0 = 0,7$.

2) *h* denota horizonte de previsão *h* passos à frente.

3) GMR significa teste de Meese-Rogoff Generalizado.

* Denota o número de vezes que o modelo 1 é superior ao modelo 2 a um nível de 5%, em 100 replicações.

** Denota o número de vezes que o modelo 2 é superior ao modelo 1 a um nível de 5%, em 100 replicações.

⁷ Os resultados das simulações dos testes da Subseção 5.2 estão disponíveis mediante solicitação.

modelos de regressão em painel, considerando-se efeitos específicos aos indivíduos constantes e variáveis no tempo.

Quanto aos efeitos constantes, os resultados podem ser resumidos como a seguir: a) o modelo 2 é o melhor num horizonte de previsão de um período à frente, mas será o pior modelo para previsões vários períodos à frente; e b) o modelo 3 é o melhor, exceto no caso de um horizonte curto de previsão em uma pequena amostra, mas será o pior modelo com $h = 1$ para um tamanho da amostra igual a 5 ou 10. Apesar do modelo 3 ter viés de pequenas amostras, ele parece ser útil para previsões vários períodos à frente.

TABELA 6

Comparação das previsões de modelos de painel com efeitos específicos aos indivíduos variáveis no tempo

	EQM de Previsão			1 vs. 2 GMR	2 vs. 3 GMR	1 vs. 3 GMR
	Modelo 1	Modelo 2	Modelo 3			
<i>h = 1</i>						
<i>T = 5</i>	1,2910	1,2765	5,3397	5*(92**)	100(0)	100(0)
<i>T = 10</i>	1,2884	1,2722	2,7411	0(99)	100(0)	100(0)
<i>T = 25</i>	1,2878	1,2737	1,8273	0(92)	100(0)	100(0)
<i>h = 5</i>						
<i>T = 5</i>	3,9860	4,0022	9,5982	42(55)	100(0)	100(0)
<i>T = 10</i>	3,9118	3,9140	8,9910	37(62)	100(0)	100(0)
<i>T = 25</i>	3,8831	3,9263	8,1500	58(29)	100(0)	100(0)
<i>h = 20</i>						
<i>T = 5</i>	10,2668	11,6686	14,2036	78(18)	76(24)	84(13)
<i>T = 10</i>	9,2413	11,2222	14,1362	98(1)	92(8)	94(2)
<i>T = 25</i>	8,6685	11,4939	15,5781	100(0)	100(0)	100(0)

NOTAS: 1) Valor médio de 100 replicações, 100 *cross-sections*, tamanho fora da amostra igual a 50 e $\rho_0 = 0,7$.

2) GMR significa teste de Meese-Rogoff Generalizado.

3) $\text{Var}(\eta_{it}) = 0,1$, isto é, os fatores específicos aos indivíduos têm uma variância pequena.

4) $\delta = 0,05$, isto é, os fatores específicos aos indivíduos têm um termo de tendência pequeno.

* Denota o número de vezes que o modelo 1 é superior ao modelo 2 a um nível de 5%, em 100 replicações.

** Denota o número de vezes que o modelo 2 é superior ao modelo 1 a um nível de 5%, em 100 replicações.

Com relação aos efeitos variáveis no tempo, nas Tabelas 6 e 7 eles são considerados para diferentes níveis de δ e $\text{Var}(\eta_{jt})$: variância pequena e tendência fraca ($\text{Var}(\eta_{jt}) = 0,1$ e $\delta = 0,05$) na Tabela 6; e variância pequena e tendência acentuada ($\text{Var}(\eta_{jt}) = 0,1$ e $\delta = 0,5$) na Tabela 7. Conclui-se que: a) nas Tabelas 6 e 7, tendo em vista que o PGD pode ser bem aproximado pelo modelo 1 em uma pequena amostra, esse modelo deve ter o melhor desempenho em previsões vários períodos à frente, independentemente da variância do fator associado a tamanho. No entanto, para horizontes de previsão mais curtos ($h = 1$), o modelo 2 tem um desempenho melhor que o do modelo 1; e b) quando o fator associado a tamanho possui uma tendência acentuada, como na

TABELA 7

Comparação das previsões de modelos de painel com efeitos específicos aos indivíduos variáveis no tempo

	EQM de Previsão			1 vs. 2 GMR	2 vs. 3 GMR	1 vs. 3 GMR
	Modelo 1	Modelo 2	Modelo 3			
<i>h = 1</i>						
<i>T = 5</i>	1,69	1,78	72,99	49*(50**)	100(0)	100(0)
<i>T = 10</i>	1,69	1,70	14,38	49(50)	100(0)	100(0)
<i>T = 25</i>	1,63	1,47	2,16	1(95)	100(0)	100(0)
<i>h = 5</i>						
<i>T = 5</i>	16,64	15,43	288,57	39(61)	100(0)	100(0)
<i>T = 10</i>	16,30	13,88	161,80	33(64)	100(0)	100(0)
<i>T = 25</i>	13,99	8,58	23,14	0(100)	100(0)	98(2)
<i>h = 20</i>						
<i>T = 5</i>	600,74	155,78	593,41	10(88)	100(0)	55(38)
<i>T = 10</i>	528,87	147,75	578,07	0(99)	100(0)	59(35)
<i>T = 25</i>	354,99	87,57	246,36	0(100)	100(0)	4(92)

NOTAS: 1) Valor médio de 100 replicações, 100 *cross-sections*, tamanho fora da amostra igual a 50 e $\rho_0 = 0,7$.

2) GMR significa teste de Meese-Rogoff Generalizado.

3) $\text{Var}(\eta_{jt}) = 0,1$, isto é, os fatores específicos aos indivíduos têm uma variância pequena.

4) $\delta = 0,5$, isto é, os fatores específicos aos indivíduos têm um termo de tendência grande.

* Denota o número de vezes que o modelo 1 é superior ao modelo 2 a um nível de 5%, em 100 replicações.

** Denota o número de vezes que o modelo 2 é superior ao modelo 1 a um nível de 5%, em 100 replicações.

Tabela 7, os modelos 2 e 3 produzem melhores previsões do que o modelo 1, exceto quando $h = 1$.

O modelo 3, com fatores específicos aos indivíduos fixos, ajusta-se muito bem fora da amostra se os fatores associados a tamanho não forem variáveis no tempo. Contudo, se esses fatores variarem, ainda que muito lentamente, então esse modelo gerará previsões insatisfatórias. Surpreendentemente, apesar dos modelos 1 e 2 serem mal-especificados, eles podem gerar previsões mais acuradas. Quando o fator associado a tamanho possui uma tendência acentuada comum a todas as séries individuais, o termo constante do modelo 2 pode captar esse fator muito bem. Por outro lado, se o fator associado a tamanho tiver uma tendência fraca mas for um passeio aleatório, então o modelo 1 deverá ter um desempenho melhor em termos de previsão.

6 - Aplicação aos dados do desmatamento da Amazônia⁸

Nesta seção utilizamos dados em painel do desmatamento da Amazônia para investigar os efeitos de fatores associados a tamanho nas estimativas dos coeficientes e previsões. Como visto em Anderson, *et alii*, a principal causa do desmatamento é a conversão de floresta em terra para agricultura. Uma estimativa da demanda por terras desmatadas na Amazônia nos fornece previsões grosseiras acerca do futuro da Floresta Amazônica. Para captar a dinâmica do desmatamento,⁹ as seguintes variáveis são importantes: tamanho do município, população, comprimento das estradas, nível de desmatamento defasado e assim por diante [ver Anderson e Reis (1996)]. Variáveis como tamanho do município e população são fatores associados a tamanho que diferem enormemente entre municípios. Esses fatores devem ter efeito crucial nas estimativas dos coeficientes e nas previsões que derivamos a partir dos dados do desmatamento da Amazônia.

Nosso modelo consiste em uma equação dinâmica que prevê novos desmatamentos, isto é, a primeira diferença das terras desmatadas, na região j e no tempo t , com base em informações passadas. Pela teoria e simulações das Seções 3 e 4, provavelmente iríamos concluir que, se há variáveis associadas a tamanho relevantes omitidas da regressão, então os coeficientes estimados seriam viesados, porque eles são correlacionados com novos desmatamentos em cada município. De fato, corroboramos essa hipótese a partir dos nossos modelos estimados.

8 Esta base de dados foi construída por Eustáquio J. Reis e colegas na Diretoria de Pesquisa do IPEA.

9 Refira-se a Anderson *et alii* (1996) para uma explicação detalhada da base de dados. Uma breve descrição é a seguinte: os dados foram coletados consistentemente para os anos de 1975, 1980 e 1985 para 316 municípios da Amazônia (506 municípios em 1991); as áreas desmatadas incluem terras de agricultura (culturas anuais, culturas permanentes e floresta plantada), pastagens (pastagens plantadas) e terras em pousio (pousio curto, pousio longo e terra não-utilizável); e as áreas virgens ou não-cultivadas (área total – área desmatada) cobrem áreas públicas, florestas naturais privadas e pastagens naturais privadas.

Diversos modelos são estimados na Tabela 8, onde se admite que as variáveis associadas a tamanho são constantes.¹⁰ O tamanho do município é não-significativo, mas o PIB, a população e o comprimento das estradas são altamente significativos. Note-se que o valor estimado da inclinação do modelo B cai de 0,58 para 0,10 quando se incluem três variáveis associadas a tamanho. Ao se adicionarem PIB, população ou comprimento das estradas individualmente, também se observa uma queda do coeficiente estimado para o termo dinâmico.

Como mencionado anteriormente, consideraremos as propriedades de fatores associados a tamanho variáveis no tempo. Para captar a dinâmica desses fatores, usamos uma estrutura auto-regressiva vetorial (VAR), apresentada na Tabela 9. População, comprimento das estradas e PIB são fatores variáveis no tempo importantes; tamanho do município e apoio financeiro do governo — Superintendência de Desenvolvimento da Amazônia (Sudam),¹¹ de 1985 — são tratados como constantes nos valores de 1985, para efeito de previsão. Inicialmente, PIB, comprimento de estradas e Sudam revelam-se não-significativos na equação de novos desmatamentos. Escolhemos um sistema bivariado com diferentes variáveis exógenas em que o coeficiente de novos desmatamentos defasados não difere significativamente de zero. Isso quer dizer que, se incluirmos a dinâmica dos fatores associados a tamanho no modelo, os novos desmatamentos serão determinados somente por esses fatores. A Tabela 10 apresenta as previsões vários passos

TABELA 8

Relação estimada entre desmatamento e variáveis associadas a tamanho

Modelos	Constante	Des _{i,t-1}	PIB _i	Rod _i	Pop _i	R ²
A	...	0,61 (0,034)	0,447
B	197,56 (47,02)	0,58 (0,034)	0,476
C	...	0,10 (0,040)	-4,85 (2,10)	2,54 (0,16)	1,87 (1,00)	0,706

NOTA: Os valores entre parênteses são os desvios padrão estimados.

Des = novos desmatamentos (km²).

PIB = PIB do município (US\$ 10 milhões), média de 1975, 1980 e 1985.

Rod = comprimento médio das rodovias estaduais e federais de 1975, 1980 e 1985 (km).

Pop = população do município (1.000 habitantes), média de 1975, 1980 e 1985.

¹⁰ PIB, população e comprimento de estradas são médias das observações de 1975, 1980 e 1985.

¹¹ Ver Anderson e Reis (1996). Créditos concedidos pela Sudam, área dos projetos aprovados por eta em 1985 em quilômetros quadrados.

TABELA 9

VAR estimado para desmatamento e variáveis associadas a tamanho

	Des_{jt-1}	Pop_{jt-1}	Área_j	Rod_j	PIB_j
Des_{jt}	0,0772 (0,039)	0,0202 (0,004)	-2,6117 (1,314)	2,5098 (0,207)	-1,1316 (0,203)
Pop_{jt}	-1,0609 (0,302)	0,7134 (1,208)	49,0253 (10,198)	5,3739 (1,607)	5,4048 (1,572)

NOTA: Os valores entre parênteses são os desvios padrão estimados.

Pop = população do município (habitantes).

Área = tamanho do município (1.000 km²).

Rod = comprimento médio das rodovias estaduais e federais de 1975, 1980 e 1985 (km).

PIB = PIB do município (US\$ milhões), média de 1975, 1980 e 1985.

TABELA 10

Previsões do desmatamento da Amazônia

(% de área desmatada sobre área total)

Ano	Valores observados			
1975	4,06			
1980	6,28			
1985	8,80			
Previsões	Modelo A	Modelo B	Modelo C	VAR
1990 (1 passo)	10,32 (9)	11,39 (33)	11,32 (13)	11,00 (13)
1995 (2 passos)	11,28 (10)	13,85 (73)	13,84 (29)	13,15 (24)
2000 (3 passos)	11,86 (13)	16,00 (107)	16,26 (52)	15,24 (37)
2005 (4 passos)	12,22 (17)	17,88 (128)	18,58 (76)	17,26 (54)
2010 (5 passos)	12,44 (18)	19,56 (147)	20,76 (97)	19,21 (71)

NOTA: Os valores entre parênteses correspondem ao número de municípios totalmente desmatados.

à frente dos modelos A, B e C e do VAR. Após considerar fatores associados a tamanho (constantes), o modelo C prevê que 20,76% da Floresta Amazônica estarão destruídos em 2010, nível muito próximo ao previsto pelo VAR (19,21%).¹²

Muitos estudos sobre o desmatamento da Amazônia têm ajustado tendências exponenciais ou lineares aos dados. As últimas estimativas do desmatamento mostram que essas simples projeções tendenciais eram muito pessimistas porque ignoravam mecanismos de realimentação que agem para diminuir o desmatamento.

Tendo em vista que a terra torna-se mais escassa, ocorre um processo natural de substituição em direção a métodos de produção mais intensivos em mão-de-obra e menos intensivos em terra. Uma curva com forma logística talvez fosse mais apropriada. Nesta seção, desenvolvemos modelos que usam variáveis *dummies* que nos permitem aproximar uma curva não-linear. Definimos as novas variáveis como se segue:

$$y_{1jt} = \begin{cases} y_{jt} & \text{se } \frac{x_{jt}}{S_j} \leq a \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad y_{2jt} = \begin{cases} y_{jt} & \text{se } a < \frac{x_{jt}}{S_j} \\ 0, & \text{caso contrário} \end{cases}$$

onde $0 \leq a \leq 1$. Na estimação apresentada, $a = 0,05$,¹³ S_j = tamanho do município e x_{jt} = área desmatada.

As taxas de desmatamento em 1980 e 1985 são altamente correlacionadas com variáveis associadas a tamanho. Espera-se que as variáveis associadas a tamanho negligenciadas venham a viesar o coeficiente do termo dinâmico em direção à unidade nos modelos D e E. A partir da Tabela 5, vê-se que as previsões dos modelos D e E não são tão boas porque eles não consideram variáveis associadas a tamanho. Depois de se adicionarem tais variáveis a esses modelos, constata-se que os coeficientes estimados são próximos de zero. O modelo F' é escolhido através de um procedimento de seleção de modelos e suas previsões são bastante similares às do modelo F.

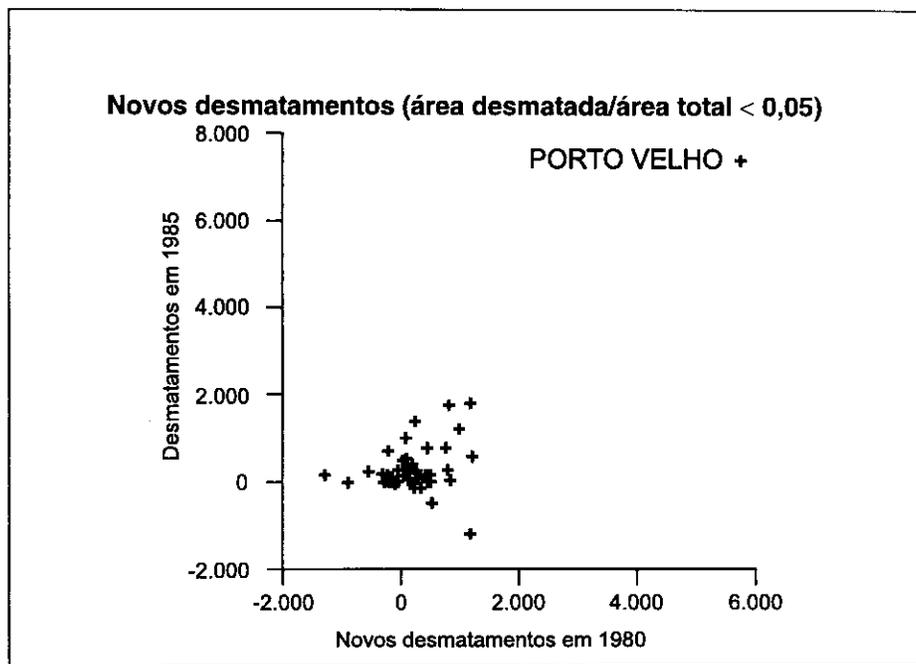
A estimativa do coeficiente do termo dinâmico é sensível à inclusão dos dados do Município de Porto Velho, em Rondônia (ver gráfico a seguir). Rondônia, localizada na parte sudoeste da Amazônia, experimentou um aumento dramático das atividades de agricultura de pequena escala¹⁴ e pecuária [ver Anderson *et alii* (1996, p. 45-47)]. Se excluirmos esse ponto da amostra, obteremos os resultados de D' e E' da Tabela 11, com valores do coeficiente de y_{1jt} menores do que os valores estimados na Tabela 8.

Após a inclusão dos fatores associados a tamanho, os resultados também sustentam uma curva do tipo logística. A inclusão desses fatores melhora bastante a capacidade

12. A título de comparação, as previsões de Anderson e Reis (1996) são 17,6% e 38,8%, respectivamente. Suas previsões são baseadas em dois cenários básicos: a) as políticas de desenvolvimento agressivas do período 1970/85 são continuadas nos próximos 25 anos (38,8%); e b) essas políticas são drasticamente reduzidas ou abandonadas (17,6%).

13. Os 25^o, 50^o e 75^o percentis da fração de área desmatada em 1980 são 0,05, 0,17, 0,33, respectivamente.

14. Culturas anuais ocupam até 15% da área desmatada e são primariamente cultivadas por pequenos fazendeiros que praticam agricultura de corte e queimada.



explicativa do modelo, ao mesmo tempo que a exclusão do *outlier* não afeta a estimativa do coeficiente do termo dinâmico. Assim, apesar de os valores estimados dos coeficientes dos termos defasados serem sensíveis a esses dados, as previsões do modelo F não são, conforme mostrado na Tabela 12.

Para testar a hipótese de taxa decrescente de desmatamento no que diz respeito à razão área desmatada/área total, inclui-se mais um ponto de fronteira. Definem-se as novas variáveis a seguir:

$$y_{1jt} = \begin{cases} y_{jt} & \text{se } \frac{x_{jt}}{S_j} \leq a \\ 0, & \text{caso contrário} \end{cases}$$

$$y_{2jt} = \begin{cases} y_{jt} & \text{se } a < \frac{x_{jt}}{S_j} \leq b \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad y_{3jt} = \begin{cases} y_{jt} & \text{se } b < \frac{x_{jt}}{S_j} \\ 0, & \text{caso contrário} \end{cases}$$

onde $a = 0,05$ e $b = 0,3$.

TABELA 11

Relação estimada entre desmatamento e variáveis associadas a tamanho

Modelos	Constante	Des _{1it-1}	Des _{2it-1}	PIB _i	Rod _i	Pop _i	R ²
D	-	1,053 (0,120)	0,585 (0,035)	-	-	-	0,472
D'	-	0,419 (0,223)	0,584 (0,032)	-	-	-	0,416
E	190,727 (46,083)	0,992 (0,118)	0,550 (0,035)	-	-	-	0,499
E'	206,451 (45,284)	0,278 (0,219)	0,548 (0,034)	-	-	-	0,453
F	-	0,088 (0,111)	0,104 (0,040)	-4,832 (2,118)	2,542 (0,173)	1,880 (1,009)	0,706
F'	-	-0,519 (0,175)	0,106 (0,039)	-5,752 (2,068)	2,543 (0,168)	2,007 (0,981)	0,683

NOTA: Os valores entre parênteses são os desvios padrão estimados; D', E' e F' excluem um outlier da regressão.

Des_{1it} = novos desmatamentos – área desmatada nos anos 80/área total < 0,05 (km²).

Des_{2it} = novos desmatamentos – área desmatada nos anos 80/área total ≥ 0,05 (km²).

PIB = PIB do município (US\$ 10 milhões), média de 1975, 1980 e 1985.

Rod = comprimento médio das rodovias estaduais e federais de 1975, 1980 e 1985 (km).

Pop = população do município (1.000 habitantes), média de 1975, 1980 e 1985.

TABELA 12

Previsões do desmatamento da Amazônia

(% de área desmatada sobre área total)

Previsões	Modelo D	Modelo E	Modelo E'	Modelo F	Modelo F'
1990 (1 passo)	10,35 (9)	39,55 (32)	40,39 (30)	11,32 (13)	11,25 (14)
1995 (2 passos)	11,29 (8)	48,06 (60)	44,22 (40)	13,85 (29)	13,75 (29)
2000 (3 passos)	11,84 (10)	53,05 (74)	43,94 (34)	16,28 (52)	16,16 (52)
2005 (4 passos)	12,16 (12)	57,30 (86)	44,12 (36)	18,60 (76)	18,46 (78)
2010 (5 passos)	12,34 (14)	60,70 (105)	44,17 (36)	20,79 (98)	20,61 (98)

NOTA: Os valores entre parênteses correspondem ao número de municípios que estão totalmente desmatados.

Os novos resultados são:

$$y_{jt} = 1,05 \cdot y_{1jt-1} + 0,62 \cdot y_{2jt-1} + 0,19 \cdot y_{3jt-1}$$

(0,118) (0,036) (0,123)

$$y_{jt} = 0,10 \cdot y_{1jt-1} + 0,14 \cdot y_{2jt-1} - 0,16 \cdot y_{3jt-1} - 5,58 \cdot GDP_j + 2,46 \cdot Road_j + 2,49 \cdot Pop_j$$

(0,111) (0,042) (0,096) (2,108) (0,174) (1,018)

Esses resultados também sustentam uma curva do tipo logística. Os coeficientes estimados das variáveis defasadas caem bastante após a inclusão das variáveis associadas a tamanho, o que sustenta que a omissão de variáveis associadas a tamanho causa um sério viés. Após a exclusão do *outlier*, os resultados da regressão são:

$$y_{jt} = 0,42 \cdot y_{1jt-1} + 0,62 \cdot y_{2jt-1} + 0,19 \cdot y_{3jt-1}$$

(0,220) (0,035) (0,121)

$$y_{jt} = -0,49 \cdot y_{1jt-1} + 0,13 \cdot y_{2jt-1} - 0,15 \cdot y_{3jt-1} - 6,17 \cdot GDP_j +$$

(0,173) (0,041) (0,093) (2,067)

$$+ 2,44 \cdot Road_j + 2,42 \cdot Pop_j + 1,87 \cdot Sudam_j$$

(0,169) (0,997) (1,399)

Em virtude de Rondônia ser o estado que recebe a menor quantidade de subsídios da Sudam, após a exclusão dos seus dados o coeficiente da Sudam torna-se significativo. O motivo do coeficiente negativo de y_{1it} da segunda regressão é o crescimento negativo de muitos municípios no Estado do Amazonas. Após sua exclusão da amostra¹⁵ o coeficiente de y_{1it} é não-significativamente diferente de zero.

7 - Resumo e conclusões

O artigo analisa os efeitos de fatores específicos individuais associados a tamanho em um modelo dinâmico de regressão em painel. Resultados teóricos mostram que esses fatores, quando variáveis no tempo ou dotados de distribuições com cauda longa, tendem a causar correlação espúria em uma regressão. Além disso, previsões baseadas em modelos que omitem os fatores associados a tamanho são seriamente afetadas por suas propriedades. Um modelo estimado a partir de dados agrupados (*pooling regression*) apresenta um bom ajustamento dentro da amostra, mas as previsões fora da amostra são insatisfatórias se os fatores associados a tamanho têm uma distribuição com cauda longa.

15 Suspeitamos da possibilidade de erros de medida nos dados da pesquisa sobre terras nesta área.

Um modelo de painel com efeitos associados ao tamanho dos indivíduos pode apresentar muitos problemas se o painel de dados tiver séries de tempo muito curtas. Uma vez que os termos individuais constantes são estimados deficientemente, as previsões também serão deficientes. Esses problemas podem se agravar se os fatores associados ao tamanho dos indivíduos não forem constantes, mas variáveis no tempo. Usando esses resultados, aplicamos um modelo dinâmico de painel a dados da Amazônia. Sem a inclusão de fatores associados a tamanho na regressão, os novos desmatamentos apresentam uma correlação muito alta com os seus valores defasados. Após a inclusão desses fatores, a correlação é próxima de zero, o que sugere que esses dados contêm correlação espúria no fase inicial do desmatamento. O modelo, após controlar-se para correlação espúria incluindo variáveis associadas a tamanho, prevê que aproximadamente 20% da Amazônia Legal estarão desmatados no ano 2010, o que constitui uma previsão otimista se comparada às de estudos anteriores.

Apêndice A

Suponha que $\tilde{X}_{j,t}$ e $\tilde{Y}_{j,t}$ sejam gerados por processos estacionários independentes:

$$\phi_x(L)\tilde{X}_{j,t} = \theta_x(L)\varepsilon_{j,t}^x \quad (\text{A.1})$$

$$\phi_y(L)\tilde{Y}_{j,t} = \theta_y(L)\varepsilon_{j,t}^y \quad (\text{A.2})$$

onde $\varepsilon_{j,t}^z$ é *i.i.d.*(0, σ_z^2), $\phi_z(L)$ e $\theta_z(L)$ são polinômios no operador de defasagens L , estáveis e inversíveis para $z = x$ ou y . Admita que $\phi_z(L)$ e $\theta_z(L) = I$ por simplicidade algébrica. Se consideramos uma função do fator associado a tamanho linear, então (1) pode ser reescrito como:

$$X_{j,t} = S_{j,t} + \tilde{X}_{j,t} \quad (\text{A.3})$$

$$Y_{j,t} = AS_{j,t} + \tilde{Y}_{j,t} \quad (\text{A.4})$$

$S_{j,t}$ é retirado (independentemente) de uma distribuição com cauda longa, onde:

$$E[S_{j,t}] = \mu_s \text{ e } \text{Var}[S_{j,t}] = \sigma_s^2 \text{ para todo } j \text{ e } t \quad (\text{A.5})$$

Admita que os fatores associados ao tamanho dos indivíduos são constantes:

a) O estimador de MQO de $\hat{\beta}$ em (2) pode ser reescrito como:

$$\hat{\beta} = \frac{\sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})(Y_{j,t} - \bar{Y})}{\sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})^2} = \frac{\sum_{j=1}^N \sum_{t=1}^T [S_j - \bar{S} + \varepsilon_{j,t}^x - \bar{\varepsilon}^x] [A(S_j - \bar{S}) + \varepsilon_{j,t}^y - \bar{\varepsilon}^y]}{\sum_{j=1}^N \sum_{t=1}^T [S_j - \bar{S} + \varepsilon_{j,t}^x - \bar{\varepsilon}^x]^2}$$

onde:

$$\bar{X} = \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T X_{j,t}$$

$$\bar{\varepsilon}^x = \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T \varepsilon_{j,t}^x \quad \text{e} \quad \bar{S} = \frac{1}{N} \sum_{j=1}^N S_j$$

Agora calcule o viés assintótico de $\hat{\beta}$ tomando os limites de probabilidade¹⁶ quando $T \rightarrow \infty$ e $N \rightarrow \infty$, de modo que $p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \hat{\beta} \rightarrow A$ quando a variância do tamanho é maior

($\sigma_x^2 \rightarrow \infty$), onde:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \hat{\beta} = \frac{A \sigma_s^2}{\sigma_s^2 + \sigma_x^2} \quad (\text{A.6})$$

b) Primeiro, defina:

$$s^2 = \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (Y_{j,t} - \hat{\alpha} - \hat{\beta} X_{j,t})^2 = \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (Y_{j,t} - \bar{Y})^2 - \hat{\beta}^2 \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})^2$$

¹⁶ Os limites de probabilidade são obtidos primeiro deixando $T \rightarrow \infty$ e depois $N \rightarrow \infty$.

Tome os limites de probabilidade para s^2 quando $T \rightarrow \infty$ e $N \rightarrow \infty$:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} s^2 = (A^2 \sigma_x^2 + \sigma_y^2) - \left[\frac{A \sigma_x^2}{\sigma_x^2 + \sigma_x^2} \right]^2 (\sigma_x^2 + \sigma_x^2) = \frac{A \sigma_x^2 \sigma_x^2}{\sigma_x^2 + \sigma_x^2} + \sigma_y^2$$

onde:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})^2 = \sigma_x^2 + \sigma_x^2$$

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{NT} \sum_{j=1}^N \sum_{t=1}^T (Y_{j,t} - \bar{Y})^2 = A^2 \sigma_x^2 + \sigma_y^2$$

A estatística- t é:

$$t_{\beta} = \frac{\hat{\beta}}{s_{\hat{\beta}}} = \hat{\beta} / s \left[\sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})^2 \right]^{-1/2}$$

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{\sqrt{NT}} t_{\beta} = \frac{a \sigma_x^2}{\left[A^2 \sigma_x^2 \sigma_x^2 + \sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_y^2 \right]^{1/2}} \quad (\text{A.7})$$

as estatísticas- t vão divergir, isto é, $t_{\beta} = O_p((NT)^{1/2})$, de modo que as estatísticas- t de β serão significativas.

c) Em seguida, considere o coeficiente de determinação:

$$R^2 = \frac{\sum_{j=1}^N \sum_{t=1}^T (\hat{Y}_{j,t} - \bar{Y})^2}{\sum_{j=1}^N \sum_{t=1}^T (Y_{j,t} - \bar{Y})^2} = \frac{\hat{\beta}^2 \sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X})^2}{\sum_{j=1}^N \sum_{t=1}^T (Y_{j,t} - \bar{Y})^2}$$

Inserindo $\hat{\beta}$ de a) e usando os limites de probabilidade anteriormente derivados, obtemos:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} R^2 = \frac{a^2 \sigma_s^4}{(A^2 \sigma_s^2 + \sigma_y^2) (\sigma_s^2 + \sigma_x^2)} \quad (\text{A.8})$$

de modo que R^2 se aproxima de 1 quando $\sigma_s^2 \rightarrow \infty$.

Agora consideremos o caso do painel dinâmico. Seja $Y_{j,t}$ gerado por (A.4) e (A.5) e seja $A = 1$, sem perda de generalidade:

$$\tilde{Y}_{j,t} = \rho_0 \tilde{Y}_{j,t-1} + \varepsilon_{j,t} \quad (\text{A.2'})$$

onde $\varepsilon_{j,t}$ é *i.i.d.*(0, σ_ε^2) e $|\rho_0| < 1$. Então, seu PGD pode ser reescrito como:

$$Y_{j,t} = (1 - \rho_0) S_j + \rho_0 Y_{j,t-1} + \varepsilon_{j,t}$$

a) O estimador de MQO de $\hat{\rho}$ na equação (8) se aproxima de um, quando $\sigma_s^2 \rightarrow \infty$, dado que:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \hat{\rho} = \rho_0 + (1 - \rho_0) \frac{\sigma_s^2}{\frac{\sigma_\varepsilon^2}{1 - \rho_0^2} + \sigma_s^2} \quad (\text{A.9})$$

b) A estatística-t é:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{\sqrt{NT}} t_\rho = \frac{p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} (\hat{\rho} - 1)}{p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} s} \left[\frac{\sigma_\varepsilon^2}{1 - \rho_0^2} + \sigma_s^2 \right]^{1/2} \quad (\text{A.10})$$

c) \hat{m} se aproxima de zero quando a variância do tamanho é maior, dado que:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \hat{m} = \frac{(1 - \rho_0) \mu_x}{1 + (1 - \rho_0^2) \sigma_s^2 / \sigma_\varepsilon^2} \quad (\text{A.11})$$

d) O coeficiente de determinação, R^2 , tende à unidade quando $\sigma_y^2 \rightarrow \infty$:

$$p \lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} R^2 = \left[\rho_0 + (1 - \rho_0) \sigma_y^2 / \left(\frac{\sigma_\varepsilon^2}{1 - \rho_0^2} + \sigma_y^2 \right) \right]^2 \quad (\text{A.12})$$

Apêndice B

Admita que (A.1) ~ (A.4) e que o PGD de fatores associados ao tamanho dos indivíduos, S_{jt} , segue um passeio aleatório com tendência:

$$S_{j,t} = \delta_j + S_{j,t-1} + \eta_{j,t} = \delta_j t + S_{j,0} + \sum_{s=1}^t \eta_{j,s} \quad (\text{A.5'})$$

S_{j0} é retirado (independentemente) de uma distribuição com cauda longa (por exemplo, uma qui-quadrada com seis graus de liberdade), para todo $j = 1, N$.

Sob essa hipótese o estimador de MQO de β do modelo (4) pode ser escrito como:

$$\hat{\beta} = \frac{\sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X}_j)(Y_{j,t} - \bar{Y}_j)}{\sum_{j=1}^N \sum_{t=1}^T (X_{j,t} - \bar{X}_j)^2} = \frac{\sum_{j=1}^N \left(\frac{1}{T^3} \sum_{t=1}^T X_{j,t} Y_{j,t} - \frac{1}{T^2} \bar{X}_j \bar{Y}_j \right)}{\sum_{j=1}^N \left(\frac{1}{T^3} \sum_{t=1}^T X_{j,t}^2 - \frac{1}{T^2} \bar{X}_j^2 \right)}$$

Considere as propriedades de convergência das somas individuais tal como em Entorf (1997):

$$\begin{aligned} \sum_{t=1}^T X_{j,t} Y_{j,t} &= \sum_{t=1}^T A X_{j,t}^2 + \sum_{t=1}^T (\varepsilon_{j,t}^y - a \varepsilon_{j,t}^x) \delta_j t + O_p(T^2) \\ T^2 \bar{X}_j \bar{Y}_j &= T^2 A \bar{X}_j^2 + T^2 \frac{\delta_j}{2} \sum_{t=1}^T (\varepsilon_{j,t}^y - a \varepsilon_{j,t}^x) + O_p(T^2) \\ \sum_{t=1}^T X_{j,t}^2 &= T^3 \frac{\delta_j^2}{3} + O_p(T^{5/2}), \quad \bar{X}_j^2 = T^2 \frac{\delta_j^2}{4} + O_p(T^{3/2}) \end{aligned}$$

Inserindo em $\hat{\beta}$ como definido anteriormente, obtemos:

$$\hat{\beta} = \frac{\sum_{j=1}^N \left\{ \frac{1}{T^3} A \sum_{j=1}^N X_{j,t}^2 - \frac{1}{T^2} A \bar{X}_j^2 + \frac{1}{T^3} \delta_j \sum_{t=1}^T t(\epsilon_{j,t}^y - A\epsilon_{j,t}^x) - \frac{1}{T^2} \frac{1}{2} \delta_j \sum_{t=1}^T (\epsilon_{j,t}^y - A\epsilon_{j,t}^x) + O_p(T) \right\}}{\sum_{j=1}^N \left\{ \frac{1}{T^3} \sum_{j=1}^N X_{j,t}^2 - \frac{1}{T^2} \bar{X}_j^2 \right\}}$$

Segue-se que:

$$T^{3/2}(\hat{\beta} - A) \approx \frac{\sum_{j=1}^N \left\{ \frac{1}{T^{3/2}} \delta_j \sum_{t=1}^T t(\epsilon_{j,t}^y - A\epsilon_{j,t}^x) - \frac{1}{T^{1/2}} \frac{1}{2} \delta_j \sum_{t=1}^T (\epsilon_{j,t}^y - A\epsilon_{j,t}^x) \right\}}{\sum_{j=1}^N \frac{1}{12} \delta_j^2}$$

Podemos usar resultados conhecidos de convergência para momentos amostrais em funcionais de processos de Wiener:

$$\frac{1}{T^{3/2}} \sum_{t=1}^T t \epsilon_{j,t}^z \Rightarrow \sigma_{zj} W_{zj}(1) - \sigma_{zj} \int_0^1 r \sigma_{zj} w_{zj}(r) dr$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^T \epsilon_{j,t}^z \Rightarrow \sigma_{zj} W_{zj}(1)$$

onde $z = x$ ou y , \Rightarrow implica convergência fraca e $W(\bullet)$ é um processo de Wiener padrão. Então, sua distribuição assintótica é:

$$T^{3/2}(\hat{\beta} - A) \Rightarrow \frac{12 \sum_{j=1}^N \delta_j \left\{ \frac{1}{2} \sigma_{yj} W_{yj}(1) - \int_0^1 \sigma_{yj} W_{yj}(r) dr \right\} - A \left\{ \frac{1}{2} \sigma_{xj} W_{xj}(1) - \int_0^1 \sigma_{xj} W_{xj}(r) dr \right\}}{\sum_{j=1}^N \delta_j^2}$$

onde $\left(\frac{1}{2} \sigma_j W_j(1) - \int_0^1 \sigma_j W_j(r) dr\right)$ tem distribuição normal com $(0, \frac{\sigma_j^2}{12})$.

Apêndice C

$$\begin{aligned} \text{Var}(\hat{\gamma}_{SD}) &= E[\hat{\gamma}_{SD}^2] - (E[\hat{\gamma}_{SD}])^2 \\ &= \frac{1}{(NT)^2} \sum_{t=1}^T \sum_{\tau=1}^T \sum_{i=1}^N \sum_{j=1}^N E[S_{it} D_{it} S_{j\tau} D_{j\tau}] - \left(\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N E[S_{it} D_{it}]\right)^2 \\ &= \frac{1}{(NT)^2} \sum_{t=1}^T \sum_{\tau=1}^T \sum_{i=1}^N \sum_{j=1}^N \{E[S_{it} D_{it} S_{j\tau} D_{j\tau}] - E[S_{it} D_{it}] E[S_{j\tau} D_{j\tau}]\} \\ &= \frac{1}{(NT)^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{t=1}^T \sum_{\tau=1}^T \{E[S_{it} D_{it} S_{j\tau} D_{j\tau}] - E[S_{it} D_{it}] E[S_{j\tau} D_{j\tau}]\} \right] \\ &= \frac{1}{(NT)^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{t=1}^T \sum_{\tau=1}^T \{E[S_{it} S_{j\tau}] E[D_{it} D_{j\tau}] + E[S_{it} D_{j\tau}] E[D_{it} S_{j\tau}]\} \right] \end{aligned}$$

onde consideramos normalidade multivariada, de modo que o quarto cumulativo conjunto da distribuição de $[S_{it}, S_{j\tau}, D_{it}, D_{j\tau}]$ é zero. Para T grande, obtemos:

$$\text{Var}(\hat{\gamma}_{SD}) \approx \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[\frac{1}{T} \sum_{\tau=-\infty}^{\infty} \{ \gamma_{S_i S_j}(\tau) \gamma_{D_i D_j}(\tau) + \gamma_{S_i D_j}(\tau) \gamma_{D_i S_j}(\tau) \} \right]$$

onde:

$$\gamma_{S_i S_j}(\tau) = \text{Cov}(S_{it}, S_{jt+\tau}), \gamma_{S_i D_j}(\tau) = \text{Cov}(S_{it}, D_{jt+\tau})$$

$$\gamma_{D_i S_j}(\tau) = \text{Cov}(D_{it}, S_{jt+\tau}), \gamma_{D_i D_j}(\tau) = \text{Cov}(D_{it}, D_{jt+\tau})$$

Um estimador consistente de Σ é:

$$\hat{\Sigma} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{\tau=-S(T)}^{S(T)} \left(1 - \frac{|\tau|}{T} \right) \left\{ \hat{\gamma}_{S_i S_j}(\tau) \hat{\gamma}_{D_i D_j}(\tau) + \hat{\gamma}_{S_i D_i}(\tau) \hat{\gamma}_{D_i S_j}(\tau) \right\} \right]$$

Abstract

This paper analyzes the effects of individual-specific size factors in a dynamic panel regression model. Theory and simulation show that a size-factor, with a long-tailed distribution or a time-varying property, may cause spurious stochastics. If a pair of panel variables depends on size in some way, then they appear to find a strong relationship, if the size variable is not used in the regression, even if the variables are otherwise independent. Moreover, forecasts based on models that have omitted size-factors are affected seriously by the property of the size-factors. A pooling regression with very short time-series appears to fit well in sample, but forecasts poorly out-of-sample if the neglected individual-specific size-factor has a long-tailed distribution. Using these results, we apply a dynamic panel model to Amazon data. Predicted deforestation is much lesser than previous studies after accounting for spurious stochastics.

Bibliografia

- ANDERSON, L. E. *et alii*. *Report on Amazon deforestation*. San Diego: University of California/Department of Economics, 1996 (Discussion Paper).
- ANDERSON, L. E., REIS, E. J. *Predicting Amazon deforestation through the determinants of demand for agricultural land*. Denmark: University of Aarhus/Department of Economics, 1996, mimeo.
- DIEBOLD, F. X., MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, v. 113, n. 3, p. 253-263, 1995.
- ENGLE, R. F. *et alii*. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, v. 81, p. 310-320, 1986.
- ENTORF, H. Random walks with drifts: nonsense regressions and spurious fixed-effect estimation, forthcoming. *Journal of Econometrics*, 1996.
- GRANGER, C. W. J. *Evaluation of panel data models: some suggestions from time series*. UCSD, 1996, mimeo.

GRANGER, C. W. J., NEWBOLD, P. *Forecasting Economic Time Series*. Orlando: Academic Press, 1986.

LEVIN, A., LIN, C. F. *Unit root test in panel data: asymptotic and finite-sample properties*. 1992 (UCSD, Working Paper).

MEESE, R., ROGOFF, K. Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *Journal of Finance*, v. 43, p. 933-948, 1988.

NICKELL, S. Biases in dynamic models with fixed effects. *Econometrica*, v. 49, n. 6, p. 1.417-1.426, 1981.

PRIESTLEY, M. B. *Spectral analysis and time series*. New York: Academic Press, 1981.

(Originais recebidos em setembro de 1997. Revistos em novembro de 1997.)

