

# Dados em Painel e Causalidade Usando R - Aula 9-

Prof. Mestre. Omar Barroso Khodr

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

- GMM
- Método Arellano-Bond
- Painéis Dinâmicos

# Introdução

- O método de painéis dinâmicos de Arellano-Bond é uma técnica econométrica amplamente utilizada para estimar modelos de dados em painel quando há **dependência temporal (variáveis dependentes defasadas como regressores)** e **endogeneidade (correlação entre regressores e o erro)**.

# Objetivos

- **Em modelos de painel dinâmico como:**
- $y_{it} = \alpha y_{it,t-1} + \beta X_{it} + \aleph_{it} + \varepsilon_{it}$
- $y_{it}$ : Variável Dependente
- $y_{it,t-1}$ : Variável Dependente com uma defasagem (AR(1))
- $X_{it}$ : Variável(is) explicativa(s)
- $\aleph_{it}$ : Efeito individual (fixo)
- $\varepsilon_{it}$ : Termo de erro (idiossincrático)

# Objetivos

- **Em modelos de painel dinâmico como:**
- $y_{it} = \alpha y_{it,t-1} + \beta X_{it} + \mathfrak{N}_{it} + \varepsilon_{it}$
- Podem existir problemas como:
- Correlação entre defasagens (processo autorregressivo), efeitos fixos e o termo de erro.

# Solução Arellano-Bond (1991)

- O método propõe um estimador GMM (Método de Momentos Generalizado) em diferenças que usa instrumentos internos (valores defasados das variáveis) para lidar com a endogeneidade.
- Transformação em Diferenças:
- A equação é diferenciada para eliminar o efeito fixo  $\alpha_{it}$ :
- $\Delta y_{it} = \alpha \Delta y_{it,t-1} + \beta \Delta X_{it} + \Delta \varepsilon_{it}$

# Solução Arellano-Bond (1991)

- O método propõe um estimador GMM (Método de Momentos Generalizado) em diferenças que usa instrumentos internos (valores defasados das variáveis) para lidar com a endogeneidade.
- Transformação em Diferenças:
- A equação é diferenciada para eliminar o efeito fixo  $\alpha_{it}$  mas introduz correlação entre :  $\Delta y_{it,t-1}$  e  $\Delta \varepsilon_{it}$ .
- $\Delta y_{it} = \alpha \Delta y_{it,t-1} + \beta \Delta X_{it} + \Delta \varepsilon_{it}$

# Solução Arellano-Bond (1991)

- Defasagens como instrumentos...
- Arellano-Bond usam valores defasados em níveis (e.g.,  $y_{it,t-2}; \dots; y_{it,t-n}$ ) como instrumentos para  $\Delta y_{it,t-1}$ , pois:
- $y_{it,t-2}$  está correlacionado com  $\Delta y_{it,t-1} = y_{it,t-1} - y_{it,t-2}$ .
- Todavia, não está correlacionado com  $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ .
- Sob a hipótese de não-autocorrelação dos erros (ortogonalidade).



# Solução Arellano-Bond (1991)

- Defasagens como instrumentos...
- Arellano-Bond usam valores defasados em níveis (e.g.,  $y_{it,t-2}; \dots; y_{it,t-n}$ ) como instrumentos para  $\Delta y_{it,t-1}$ , pois:
- $y_{it,t-2}$  está correlacionado com  $\Delta y_{it,t-1} = y_{it,t-1} - y_{it,t-2}$ .
- Todavia, não está correlacionado com  $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ .
- Sob a hipótese de não-autocorrelação dos erros (ortogonalidade).

# Solução Arellano-Bond (1991)

- As condições de momento são:
- $E[y_{it-s} \cdot \Delta \varepsilon_{it-1}] = 0 \rightarrow s \geq 2 \text{ e } t \geq 3$ .
- O estimador GMM minimiza uma função de distância quadrática usando essas condições.
- Testes de Especificação:
- Teste de Sargan/Hansen: Verifica a validade dos instrumentos (sobre-identificação).
- Teste de Autocorrelação: Arellano-Bond testam se  $\Delta \varepsilon_{it}$  tem autocorrelação de ordem 1 (esperada) e de ordem 2 (não esperada).

# Solução Arellano-Bond (1991)

- Vantagens:
- Consistente para  $N \rightarrow \infty$  e  $T$  fixo.
- Eficiente ao explorar múltiplos instrumentos defasados.
- Limitações
- Fraqueza dos instrumentos: Se  $\alpha$  está próximo de 1, instrumentos defasados podem ser fracos (viés em amostras finitas).
- Variáveis não estritamente exógenas: Se  $X_{it}$  é predeterminando (correlacionado com erros passados), o método precisa ser estendido....

# Solução Arellano-Bond (1991)

- Resumo:
- O método de Arellano-Bond é essencial para estimar painéis dinâmicos com:
- Variável dependente defasada como regressor.
- Efeitos fixos individuais.
- Endogeneidade tratada via instrumentos defasados em GMM.

# Exemplo

- Suponha que queremos estimar um modelo em que o investimento atual depende do investimento passado (dinâmica) e do valor de mercado:
- $invest_{it} = \alpha invest_{i,t-1} + \beta value_{it} + \aleph_{it} + \varepsilon_{it}$
- $\aleph_{it}$ : Efeito fixo não observado (e.g., características de uma empresa).
- $invest_{i,t-1}$ : Variável dependente defasada (causa endogeneidade).
- Nosso modelo contempla, dados anuais de investimento (invest), valor de mercado (value) e estoque de capital (capital) para 10 empresas (1935-1954).

# Exemplo

- Suponha que queremos estimar um modelo em que o investimento atual depende do investimento passado (dinâmica) e do valor de mercado:
- $invest_{it} = \alpha invest_{i,t-1} + \beta value_{it} + \aleph_{it} + \varepsilon_{it}$
- $\aleph_{it}$ : Efeito fixo não observado (e.g., características de uma empresa).
- $invest_{i,t-1}$ : Variável dependente defasada (causa endogeneidade).
- Nosso modelo contempla, dados anuais de investimento (invest), valor de mercado (value) e estoque de capital (capital) para 10 empresas (1935-1954).

# Exemplo

- **A saída incluirá:**
- Coeficientes estimados para  $\text{lag}(\text{invest}, 1)$  e  $\text{value}$ .
- **Teste de Sargan:** Para validade dos instrumentos ( $p\text{-valor} > 0.05$  indica instrumentos válidos).
- **Teste de Autocorrelação:**
- $\text{AR}(1)$  esperado ( $p\text{-valor} < 0.05$ ).
- $\text{AR}(2)$  não esperado ( $p\text{-valor} > 0.05$ ).

# Exemplo

- Tipo de modelo: GMM em dois passos com transformação em diferenças (para eliminar efeitos fixos individuais).
- Painei balanceado: 11 empresas (n), 20 anos (T), total de 220 observações (N).
- Observações usadas: 198 (devido à perda de dados ao usar defasagens e diferenças).

Balanced Panel: n = 11, T = 20, N = 220

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )	
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13	***
value	0.111276	0.017096	6.5090	7.565e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(170) = 9.421828 (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients: chisq(2) = 175.8223 (p-value = < 2.22e-16)



# Exemplo

- **lag(invest, 1) (0.920):**
- O investimento no ano anterior tem um efeito positivo e altamente significativo (p-valor  $\approx 0$ ) sobre o investimento atual.
- **Interpretação:** Um aumento de 1 unidade no investimento defasado está associado a um aumento de 0.92 unidades no investimento atual (persistência alta).
- **value (0.111):**
- O valor de mercado da empresa tem um efeito positivo e significativo (p-valor  $\approx 0$ ) sobre o investimento.
- **Interpretação:** Um aumento de 1 unidade no valor de mercado está associado a um aumento de 0.111 unidades no investimento atual.

Balanced Panel: n = 11, T = 20, N = 220

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13 ***
value	0.111276	0.017096	6.5090	7.565e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(170) = 9.421828 (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients: chisq(2) = 175.8223 (p-value =  $< 2.22e-16$ )

# Exemplo

- Teste de Sargan:
- Hipótese nula: Os instrumentos são válidos (não correlacionados com o erro).
- Resultado:  $p\text{-valor} = 1 \rightarrow$  Não rejeitamos  $H_0$ . Os instrumentos (valores defasados de invest) são válidos.
- Observação: Um  $p$ -valor alto (próximo de 1) é bom, mas valores exatamente 1 podem indicar excesso de instrumentos (possível sobreidentificação).

Balanced Panel:  $n = 11$ ,  $T = 20$ ,  $N = 220$

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )	
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13	***
value	0.111276	0.017096	6.5090	7.565e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test:  $\text{chisq}(170) = 9.421828$  (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients:  $\text{chisq}(2) = 175.8223$  (p-value =  $< 2.22\text{e-}16$ )

# Exemplo

- Teste AR(1):
- Esperamos autocorrelação de 1ª ordem (p-valor = 0.138) porque a transformação em diferenças introduz dependência entre os termos de erro. Assim, o teste sugere que o modelo esteja dentro do esperado.

Balanced Panel: n = 11, T = 20, N = 220

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )	
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13	***
value	0.111276	0.017096	6.5090	7.565e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test:  $\chi^2(170) = 9.421828$  (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients:  $\chi^2(2) = 175.8223$  (p-value =  $< 2.22e-16$ )

# Exemplo

- Teste AR(2):
- Não esperamos autocorrelação de 2ª ordem (p-valor = 0.289).
- O resultado não rejeita  $H_0$ , indicando que os erros não têm autocorrelação serial de ordem 2 (condição crítica para validade dos instrumentos).

Balanced Panel: n = 11, T = 20, N = 220

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )	
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13	***
value	0.111276	0.017096	6.5090	7.565e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test:  $\text{chisq}(170) = 9.421828$  (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients:  $\text{chisq}(2) = 175.8223$  (p-value =  $< 2.22\text{e-}16$ )

# Exemplo

- **Persistência do investimento:** O coeficiente de lag(invest, 1) (0.92) indica que o investimento é altamente persistente (empresas tendem a manter seus níveis de investimento ao longo do tempo).
- **Efeito do valor de mercado:** O coeficiente de value (0.111) sugere que empresas com maior valor de mercado investem mais, tudo o mais constante.
- **Validade do modelo:** Os testes de Sargan e autocorrelação confirmam que o modelo está bem especificado e os instrumentos são válidos.

Balanced Panel: n = 11, T = 20, N = 220

Number of Observations Used: 198

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-215.7585	-14.3242	0.4702	-2.7724	12.6689	164.5442

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z )	
lag(invest, 1)	0.920231	0.128408	7.1665	7.695e-13	***
value	0.111276	0.017096	6.5090	7.565e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(170) = 9.421828 (p-value = 1)

Autocorrelation test (1): normal = -1.480322 (p-value = 0.13879)

Autocorrelation test (2): normal = -1.059406 (p-value = 0.28941)

Wald test for coefficients: chisq(2) = 175.8223 (p-value = < 2.22e-16)

# Exemplo

- **Pooled OLS (1):**
- **Viés esperado:** O Pooled OLS superestima o coeficiente da variável dependente defasada ( $\text{lag}(\text{invest}, 1)$ ), pois ignora a correlação entre  $\text{lag}(\text{invest}, 1)$  e o efeito fixo não observado.
- Aqui, o coeficiente é 0.512, menor que o do GMM (0.920), o que é incomum (geralmente o OLS superestima). Pode indicar problemas nos dados (ex.: autocorrelação negativa).
- **Viés em value:** O coeficiente (0.078) é subestimado comparado ao GMM (0.111), pois variáveis omitidas ou efeitos fixos distorcem a relação.
- **Problema:** O alto  $R^2$  (0.856) engana — o modelo não controla para heterogeneidade não observada.

## Comparação de Estimadores

Dependent variable:			
	invest		panel
	linear		GMM
	(1)	(2)	(3)
lag(invest, 1)	0.512*** (0.040)	0.413*** (0.042)	0.920*** (0.128)
value	0.078*** (0.007)	0.172*** (0.015)	0.111*** (0.017)
Constant	-12.300* (6.862)		
Observations	219	219	11
R2	0.856	0.555	
Adjusted R2	0.854	0.529	
F Statistic	639.418*** (df = 2; 216)	128.553*** (df = 2; 206)	
Note: *p<0.1; **p<0.05; ***p<0.01			

# Exemplo

- **Within (2):**
- **Viés esperado:** O estimador Within subestima o coeficiente da variável defasada devido ao viés de Nickell.
- Aqui, o coeficiente é 0.413 (vs. 0.920 no GMM), consistente com a teoria. A transformação Within gera correlação entre a primeira defasagem e o erro.
- **Viés em value:** O coeficiente (0.172) é superestimado comparado ao GMM (0.111), possivelmente porque o efeito fixo captura parte da relação entre value e invest.
- **Problema:** O  $R^2$  (0.555) é menor que no OLS, mas ainda não resolve a endogeneidade da variável defasada.

## Comparação de Estimadores

Dependent variable:			
	invest		panel
	linear		GMM
	(1)	(2)	(3)
lag(invest, 1)	0.512*** (0.040)	0.413*** (0.042)	0.920*** (0.128)
value	0.078*** (0.007)	0.172*** (0.015)	0.111*** (0.017)
Constant	-12.300* (6.862)		
Observations	219	219	11
R2	0.856	0.555	
Adjusted R2	0.854	0.529	
F Statistic	639.418*** (df = 2; 216)	128.553*** (df = 2; 206)	
Note: *p<0.1; **p<0.05; ***p<0.01			

# Exemplo

- **Arellano-Bond (3):**
- **Viés corrigido:** O GMM em diferenças usa instrumentos válidos (defasagens de invest) para lidar com a endogeneidade.
- O coeficiente de lag(invest, 1) (0.920) é maior que no Within e no OLS, refletindo a alta persistência real do investimento, livre de viés.
- O coeficiente de value (0.111) é intermediário entre OLS e Within, sugerindo que o GMM equilibra os vieses de ambos.
- **Força do modelo:** Os testes de Sargan e autocorrelação (vistos anteriormente) validam os instrumentos.

## Comparação de Estimadores

Dependent variable:			
	invest		panel
	linear		GMM
	(1)	(2)	(3)
lag(invest, 1)	0.512*** (0.040)	0.413*** (0.042)	0.920*** (0.128)
value	0.078*** (0.007)	0.172*** (0.015)	0.111*** (0.017)
Constant	-12.300* (6.862)		
Observations	219	219	11
R2	0.856	0.555	
Adjusted R2	0.854	0.529	
F Statistic	639.418*** (df = 2; 216)	128.553*** (df = 2; 206)	
Note: *p<0.1; **p<0.05; ***p<0.01			



# Bibliografia

- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*.
- Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata.
- Wooldridge, J.M. (2013) *Introductory econometrics: a modern approach*. 5th ed. Michigan State University.