

Econometria e Séries Temporais - Aula 18/19 -

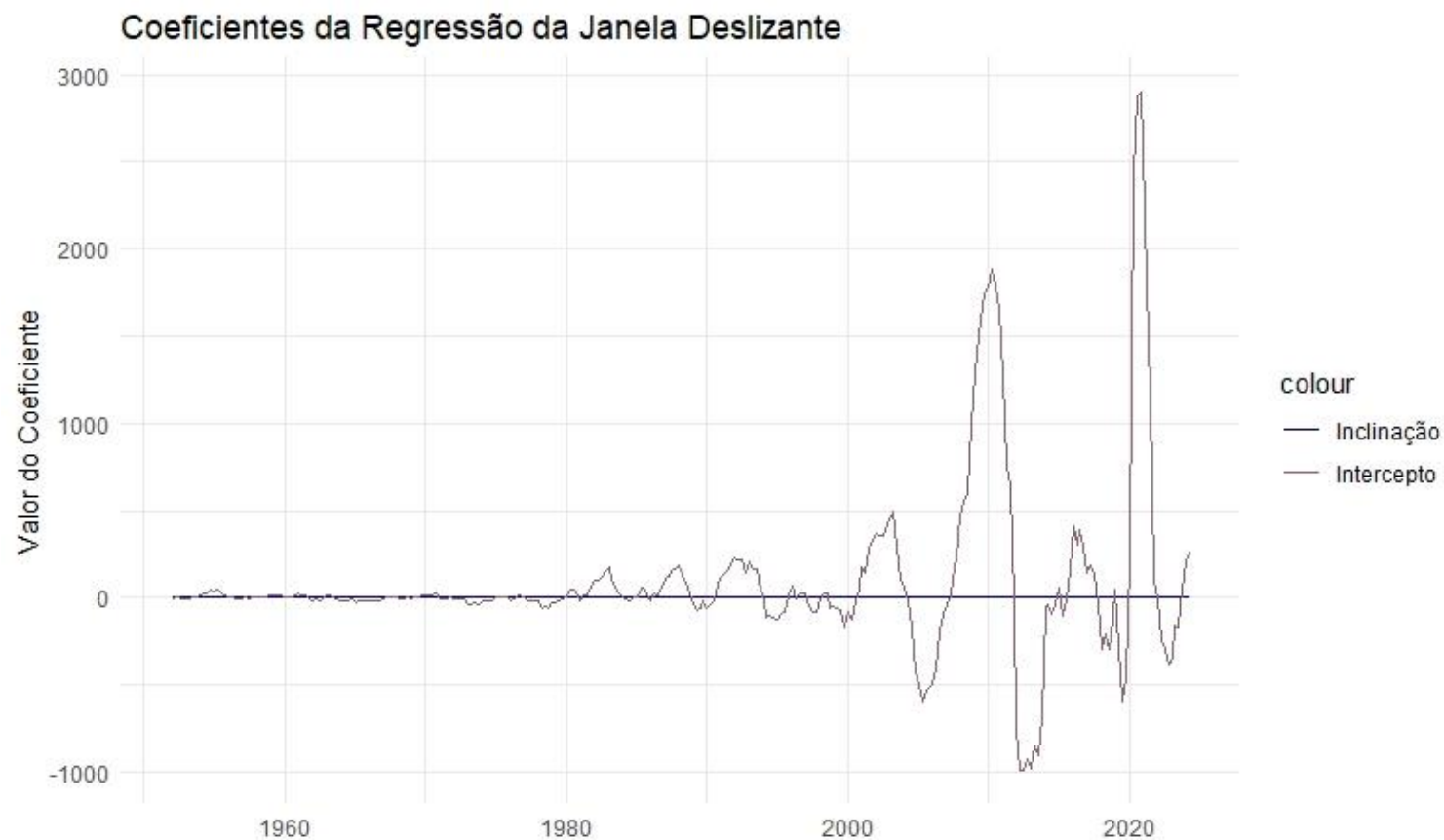
Prof. Mestre. Omar Barroso

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

Fazendo Testes de Projeções

- Vimos na aula passada que apesar de uma previsão econômica/financeira ser realizada com sucesso, necessitamos de testes para argumentar sobre a eficiência estatística do exercício.
- Em nosso caso, nossos testes partem do princípio que a nossa janela deslizando (rolagem) foi realizada com sucesso.
- Desta maneira, vamos realizar os testes discutidos na aula passada (DM, West e GW) para verificar a eficiência do nosso exercício.

Relembrando nossa projeção da aula passada



Teste DM

- Os resultados do teste Diebold-Mariano (DM) fornecem informações sobre a precisão de duas previsões concorrentes com base nos erros de previsão.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2  
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773  
alternative hypothesis: two.sided
```

Teste DM

- Estatística DM (DM = 0,55801):
Esta estatística é a estatística de teste para o teste Diebold-Mariano, que compara os erros de previsão de dois modelos. Um valor absoluto maior indica uma diferença mais forte entre as duas séries de erros de previsão. Aqui, a estatística DM é relativamente pequena, indicando uma diferença mínima entre os erros das duas previsões.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773
alternative hypothesis: two.sided
```

Teste DM

- Horizonte de previsão (Horizonte de previsão = 1): Isso especifica que os erros de previsão foram calculados para uma previsão de 1 passo (um período) à frente. Se sua previsão estivesse em um horizonte mais longo, isso também indicaria isso.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2  
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773  
alternative hypothesis: two.sided
```

Teste DM

- Poder da função de perda (poder da função de perda = 2): Este é o poder usado para a função de perda quadrada, que é típica para testes DM, pois enfatiza erros de previsão maiores.
- Valor p (valor p = 0,5773): O valor p aqui indica a probabilidade de observar uma estatística DM tão extrema quanto 0,55801 sob a hipótese nula de que ambos os modelos de previsão são igualmente precisos. Um valor p maior que 0,05 significa que não há evidências suficientes para rejeitar a hipótese nula.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2  
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773  
alternative hypothesis: two.sided
```

Teste DM

- Hipótese alternativa (hipótese alternativa: two.sided):
especifica que o teste verifica se um modelo é significativamente melhor ou pior que o outro em qualquer direção, tornando-o um teste bilateral.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2  
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773  
alternative hypothesis: two.sided
```


Teste DM

- **Conclusão**
- Como o valor p é 0,5773 (que está bem acima de um nível de significância padrão de 0,05), assim falhamos em rejeitar a hipótese nula.
- Isso significa que não há evidências suficientes para concluir que uma das previsões é significativamente mais precisa do que a outra - nesta comparação. Em termos práticos, os dois modelos estão tendo desempenho semelhante com base nos erros de previsão fornecidos.

Diebold-Mariano Test

```
data: forecast_errorsforecast_errors_2
DM = 0.55801, Forecast horizon = 1, Loss function power = 2, p-value = 0.5773
alternative hypothesis: two.sided
```

Teste de West

- Estatística de teste (`$statistic = -0,1323`): Este valor representa a estatística de teste de West (1996), que mede a diferença na precisão da previsão entre os dois modelos. Um valor próximo de zero, como $-0,1323$.
 - $-0,1323$, sugere que os erros de previsão dos dois modelos são muito semelhantes.
- `$statistic`
 - `[1] -0.1323426`
 - `$p_value`
 - `[1] 0.8947133`

Teste de West

- p-value (`$p_value = 0,8947`): Este alto p-value indica que não há evidências suficientes para rejeitar a hipótese nula de que ambos os modelos têm precisão de previsão igual.
 - Em outras palavras, em qualquer nível de significância comum (por exemplo, 0,05 ou 0,1), este resultado sugere que o desempenho dos dois modelos de previsão é estatisticamente indistinguível.
- `$statistic`
 - `[1] -0.1323426`
 - `$p_value`
 - `[1] 0.8947133`

Teste GW

- Os resultados do teste Giacomini-White (GW) fornecem insights sobre a previsibilidade condicional dos erros de previsão com base na variável GDP_Lag1.

Teste GW

- Intercepto (Estimativa = $5,9397e-05$): O intercepto representa a diferença média em erros de previsão quando o preditor GDP_Lag1 é zero. Este valor é muito próximo de zero, sugerindo que a diferença média em erros de previsão é mínima.
- Coeficiente de inclinação para GDP_Lag1 (Estimativa = $-1,0034e-07$): Este é o parâmetro-chave no teste GW, medindo o quanto a diferença em erros de previsão depende de GDP_Lag1. Uma estimativa próxima de zero indica que GDP_Lag1 tem pouco ou nenhum impacto nas diferenças de erro de previsão.

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	$5.9397e-05$	$6.5611e-03$	0.0091	0.9928
GDP_Lag1	$-1.0034e-07$	$4.7958e-07$	-0.2092	0.8344

Teste GW

- Valores de p ($\Pr(>|t|)$):
- Para o intercepto: $p = 0,9928$, e para GDP_Lag1: $p = 0,8344$.
- Esses altos valores de p significam que nem o intercepto nem o coeficiente GDP_Lag1 são estatisticamente significativos. Em outras palavras, não há evidências que sugiram que GDP_Lag1 afeta sistematicamente os erros de previsão ou que as diferenças na precisão da previsão sejam condicionalmente previsíveis.

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.9397e-05	6.5611e-03	0.0091	0.9928
GDP_Lag1	-1.0034e-07	4.7958e-07	-0.2092	0.8344

Afinal, os nossos testes falharam?

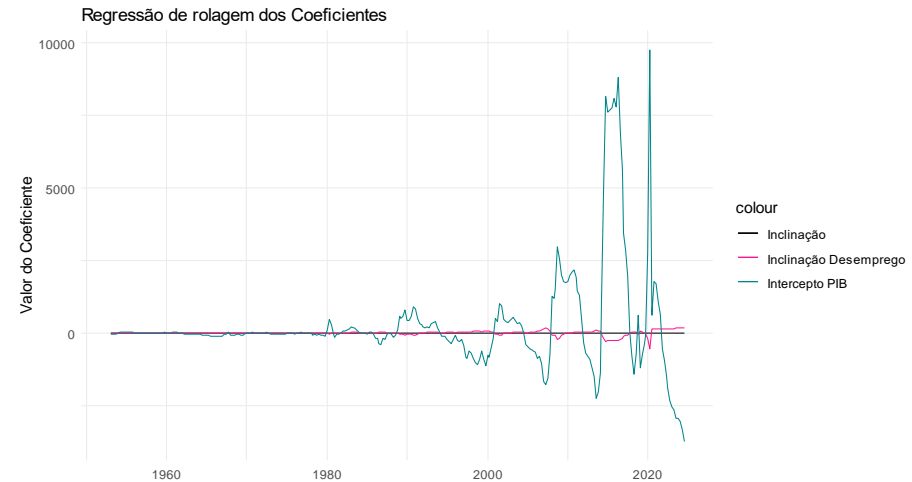
- A 'tríplice' dos testes de previsões nos dizem que nossas projeções não são muito diferentes após renderizamos diferentes variações da janela deslizando sobre o PIB.
- Podemos desta maneira sugerir que as projeções ficaram aceitáveis e podemos fechar o nosso argumento?

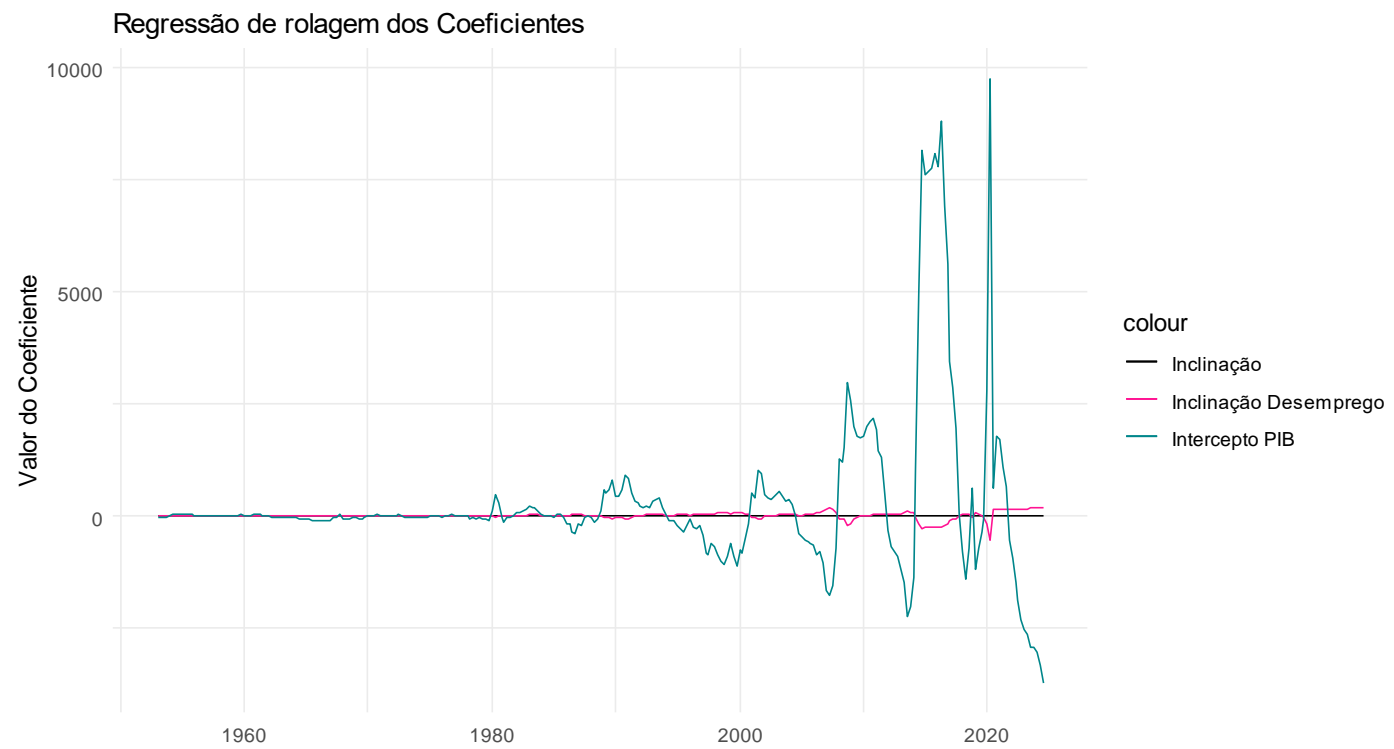
Afinal, os nossos testes falharam?

- Não!
- Esquecemos de comparar a nossa janela com um série temporal distinta...
- Desta maneira, mesmo após testar para diferentes versões, obtemos uma resposta parecida...
- Vamos comparar o PIB com um outra variável macroeconômica e ver os resultados...

Rolagem PIB + Desemprego

- Neste exercício, realizamos o mesmo método de rolagem para o PIB em conjunto com desemprego.
- Em comparação o PIB tende a ser mais volátil do que o desemprego ao longo prazo.





Intuição

- Podemos supor diversos fatores por trás das distorções do PIB.
- Sejam elas por mudanças estruturais significativas ao longo das décadas.
- Assim, como choques externos por conta de disputas geo-políticas e crises.
- Todavia, devemos testar se nossa previsão do PIB é mais eficaz do que a do desemprego.
- Talvez as previsões se sobreponham ou não tenham nenhuma relação.

Resultados Testes...

Teste de DM

- Estatística: $DM=0,96166$
- p-valor: 0,337
- Interpretação: O p-valor está acima de 0,05, indicando que não há diferença estatisticamente significativa entre a precisão da previsão dos dois modelos (PIB vs. Desemprego).
- Em outras palavras, a previsão atual e o modelo de referência (previsão somente do PIB) parecem ter desempenho semelhante, pelo menos em termos de erro médio quadrático de previsão.

Teste de West

- Estatística: 3,88
- p-valor: 0,0001
- Interpretação: Com um p-valor bem abaixo de 0,05, o teste West sugere uma diferença estatisticamente significativa nos erros de previsão, indicando que pode haver autocorrelação nos erros de previsão no qual o modelo não capturou, o que pode justificar ajustes de modelo ou variáveis adicionais.

Teste de GW

- Coeficiente de atraso do PIB: 0,00046 (valor de p: 0,60)
- Coeficiente de atraso do UNRATE: -2,77 (valor de p: 0,22)
- Interpretação: Os valores de p para os atrasos do PIB e do Desemprego são maiores que 0,05, o que implica que esses atrasos não adicionam valor preditivo significativo para explicar as diferenças nos erros de previsão. Portanto, incluir a taxa de desemprego ou o atraso do PIB como variáveis condicionantes não melhora o poder preditivo condicional.

Resumo dos Resultados

- O teste DM indica desempenho de previsão comparável **sem melhora** significativa em relação ao benchmark somente do PIB.
- O teste West destaca que **pode haver autocorrelação nos erros**, sugerindo possível melhoria do modelo com uma estrutura mais avançada.
- Já o teste GW indica que adicionar as variáveis defasadas não afeta significativamente a precisão da previsão. No entanto, a condicionante pode trazer resultados diferentes.

Discussão dos Resultados

- Podemos explorar mais ajustes de modelo, como tentar estruturas de defasagem alternativas ou adicionar novas variáveis explicativas.
- As variáveis explicativas podem ser de acordo com uma gama de diferentes dados macroeconômicos e macrofinanceiros (câmbio, juros, SWAPs e CDFs).
- No contexto, do teste de GW uma variável condicionada (instrumentalizada) em referência a ciclos de negócios e/ou bolhas econômicas pode trazer resultados mais robustos!

Referências

Diebold, F. X., and Mariano, R. S. (1995), Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, 130-141

Diebold, F. X., and Mariano, R. S. (2015), Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests, Journal of Business and Economic Statistics, 33, 1-9

Elliott, G., and Timmermann, A. (2016), Economic Forecasting, Princeton University Press. (Ch 2, Ch16, Ch17)

Giacomini, F., and White, H. (2006), Tests of Conditional Predictive Ability, Econometrica 74, 1545–78

West, K. D. (1996), Asymptotic Inference About Predictive Ability, Econometrica, 64, 1067–1084.