

O impacto de indicadores macroeconômicos em ações no setor de concessão: uma análise de regressão múltipla.

JULIANA MIRANDA*

Resumo

Os indicadores macroeconômicos influenciam na evolução das ações de empresas de um país. Este estudo investiga a relação entre esses indicadores e as empresas do setor de concessões. O objetivo é identificar e quantificar as relações entre os conjuntos de dados utilizando regressões múltiplas. A validação dos modelos é realizada principalmente por meio da análise do coeficiente de determinação (R^2) e do Critério de Informação de Akaike (AIC).

1 Introdução

O setor de concessões no Brasil, que inclui empresas de infraestrutura como rodovias, saneamento e energia e desempenha um papel crucial na economia do país. O desempenho econômico dessas empresas está diretamente ligado às variáveis macroeconômicas que influenciam o ambiente de negócios, como inflação, taxa de juros, câmbio e PIB. Compreender essas relações é essencial para investidores, gestores e formuladores de políticas.

As variáveis macroeconômicas escolhidas para este estudo são a SELIC, o IPCA, o IBC-BR, o câmbio e o IGP-M, cada uma desempenhando um papel significativo na economia e, consequentemente, no desempenho das ações de empresas de concessão. A SELIC, taxa básica de juros, influencia diretamente o custo do crédito e o nível de endividamento das empresas, além de afetar a atividade econômica de forma indireta. O IPCA, Índice de Preços ao Consumidor Amplo, é o principal indicador da inflação no Brasil e impacta o poder de compra e a demanda agregada. O IBC-BR, Índice de Atividade Econômica do Banco Central, serve como uma proxy para o PIB e reflete a evolução da atividade econômica, sendo um indicador antecedente importante para as decisões de investimento. O câmbio, representando a taxa de conversão entre o real e outras moedas, afeta os custos de importação e exportação, influenciando a competitividade das empresas no mercado internacional. Por fim, o IGP-M, Índice Geral de Preços do Mercado, é amplamente utilizado como fator de correção nos contratos, impactando os custos operacionais e de financiamento das empresas. Cada uma dessas variáveis oferece uma perspectiva única sobre as condições econômicas que podem afetar o desempenho das ações das empresas de concessão.

Através da linguagem de programação python, serão realizados testes de estacionariedade (Dickey & Fuller, 1979), correlação e análises das premissas da regressão para entender a relação entre os ativos do setor de concessões e os indicadores macroeconômicos. Diferentes transformações serão aplicadas às séries temporais com base no comportamento específico de cada uma. Em seguida, o trabalho se concentra no objetivo específico de formular equações de regressão que expliquem os retornos acionários do setor de concessões em função das variações macroeconômicas. A validação dos modelos será feita por meio da análise do coeficiente de

determinação (R^2), do Critério de Informação de Akaike (AIC) e da visualização da distribuição dos resíduos.

Esse estudo tem como objetivo geral responder às perguntas: (i) O método de regressão múltipla é um bom método para analisar a relação entre indicadores macroeconômicos e empresas da área de concessão? (ii) Os indicadores são capazes de explicar a mudança no valor de fechamento das ações das empresas do setor de concessões?

2 Análise exploratória de dados

As séries escolhidas foram duas empresas de cada ramo do setor de concessão e podem ser visualizados na Tabela 1.

Área do setor	Empresa	Fonte
Ferrovias	CCR	Economática
	Ecorodovias (ECOR)	Economática
Energia	Taes (TAE)	Economática
	Eletrobras (ELET)	Economática
Saneamento	Sabesp (SBSP)	Economática
	Cosan S.A (CSAN)	Economática

Tabela 1. Empresas do setor de concessão e suas respectivas áreas e fonte dos dados.

Os indicadores escolhidos e suas fontes estão representados na Tabela 2.

Indicador	Sigla	Fonte
Inflação	IPCA	Economática
Juros	Selic (LFT)	Economática
Câmbio	CAMBIO	Banco Central do Brasil
Índice de atividade econômica	IBC-BR	Banco Central do Brasil
Índice Geral de Preços	IGP-M	FGV

Tabela 2. Empresas do setor de concessão e suas respectivas áreas e fonte dos dados.

Para todas as séries foram separados os dados nos mesmos períodos, isto é, desde janeiro de 2011 até março de 2024. De imediato já é possível se deparar com um problema que é: os dados possuem formatos diferentes, as empresas apresentam cotações diárias enquanto os indicadores macroeconômicos possuem distribuição mensal. A abordagem escolhida para lidar com esse problema foi transformar os dados diários em dados mensais escolhendo o último do mês. Dessa forma também foi possível lidar com o problema de feriados e finais de semana, quando o mercado acionário se encontra fechado. No geral, estamos lidando com cerca de 158 observações de cada série.

É importante ressaltar que os dados utilizados neste estudo são de natureza longitudinal, ou seja, são coletados ao longo do tempo,

*Professor: Luiz Max Carvalho

permitindo a análise das tendências e variações das variáveis macroeconômicas e do desempenho das ações das empresas de concessão. Isso significa que antes de checar qualquer relação entre os dados ou as premissas de uma relação linear, precisamos que as séries sejam estacionárias.

2.1 Estacionariedade: o teste de Dickey-Fuller aumentado

Uma série é chamada de estacionária se possui a mesma média e variância para todos os seus valores ao longo do tempo. A condição da estacionariedade é um requisito importante ao relacionar variáveis econômicas, pois quando uma série econômica é não-estacionária qualquer regressão realizada com outra variável é considerada espúria, mesmo que essa regressão mostre alta significância estatística. Para descobrir se uma série é estacionária, este trabalho utiliza o teste de raiz unitária de Dickey-Fuller Aumentado (ADF), que é derivado do teste de Dickey-Fuller (DF), apresentado em Dickey e Fuller (1979), e se baseia em testar os seguintes processos auto regressivos:

$$Y_t = \rho Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \epsilon_t$$

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \epsilon_t \quad (1)$$

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \epsilon_t$$

Com Y_t representando o valor da série no tempo t ; α representando o intercepto; βt sendo o termo que representa a tendência determinística; $\Delta Y_{t-i} = Y_{t-i} - Y_{t-i-1}$; ϵ_t é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid), sem autocorrelação e com distribuição $N(0, \sigma^2)$. Então, o objetivo do teste ADF é testar a hipótese nula de que a série é não estacionária, ou seja, $|\rho| = 1$, contra a hipótese alternativa de que a série é estacionária ($|\rho| < 1$).

Em termos práticos, utilizando a função ADF em Python, temos como resposta a estatística de teste, o p-valor e os valores críticos nos níveis de significância de 1%, 5% e 10%. A forma de interpretar os resultados é a seguinte: primeiro, olhamos para o p-valor. Se ele for menor do que 0,05, isso significa que a série é estacionária. Em seguida, comparamos o valor da estatística de teste com os valores críticos para definir o nível de significância. O valor da estatística de teste deve ser menor do que o valor do nível de significância.

A análise de estacionariedade foi realizada para todas as variáveis envolvidas no estudo, tanto independentes quanto dependentes. No caso das séries temporais das ações das empresas do setor, utilizamos o valor de fechamento das ações. Entre essas empresas, apenas a série temporal da empresa CCR3 apresentou resultados que atendem aos padrões de estacionariedade sem a necessidade de transformações adicionais. O resultado dos testes ADF nas séries temporais sem nenhuma transformação podem ser vistos na Tabela 3.

Empresa	p-valor	estatística de teste	nível crítico
Sabesp	0.9860	0.5411	-
Cosan S.A	0.9081	-0.4113	-
Taesa	0.9864	0.5567	-
Eletrobas	0.8751	-0.5813	-
Ecorodovias	0.1799	-2.2754	-
CCR	4.8228e-12	-7.8766e+00	1%

Tabela 3. Resultado do teste ADF para a série original das empresas

Após suas respectivas transformações por meio da diferenciação ou log-diferenciação temos na Tabela 4 os resultados finais com todas as séries estacionárias.

Empresa	p-valor	estatística de teste	nível crítico
Sabesp	8.3304e-14	-8.5682	1%
Cosan S.A	1.0702e-07	-6.0848	1%
Taesa	9.0904e-14	-8.5534	1%
Eletrobas	0.00007	-4.7192	1%
Ecorodovias	6.7058e-23	-1.2322e+01	1%
CCR	4.8228e-12	-7.8766e+00	1%

Tabela 4. Resultado do teste ADF após transformação na série temporal das empresas

Da mesma forma, essa análise foi realizada para os indicadores macroeconômicos. Deles, apenas no teste ADF do câmbio e do IBC-BR não foi possível rejeitar a hipótese nula de estacionariedade. Os resultados do teste ADF para as séries dos indicadores, sendo que nas séries do câmbio e do IBC-BR são diferenciadas pode ser observado na Tabela 5.

Indicador	p-valor	estatística de teste	nível crítico
IPCA	1.0931e-08	-6.5122	1%
Câmbio	2.2423e-17	-9.9708	1%
IBC-BR	0.0009	-4.1018	1%
CDI	0.0478	-2.8789	5%
SELIC	0.0468	-2.8869	5%
IGP-M	0.0000	-5.4807	1%

Tabela 5. Resultado do teste ADF após transformação nas séries dos indicadores

2.2 Gráficos

Foram plotadas todas as séries temporais das ações. Esses gráficos permitiram uma melhor visualização e identificação de fatores multiplicativos ou exponenciais presentes nas séries. Esta etapa não pode ser negligenciada na escolha do método mais adequado para transformar as séries em estacionárias, seja por meio de diferenciação ou log diferenciação. A identificação de padrões de crescimento exponencial, por exemplo, sugere a aplicação da log diferenciação, enquanto a presença de tendências lineares pode ser mais adequadamente tratada com a diferenciação simples.

Além disso, foram elaborados gráficos de dispersão para analisar a relação entre as ações e os indicadores econômicos selecionados.

Estes gráficos facilitaram a visualização das possíveis relações lineares ou não lineares entre as variáveis. Com base nessa análise, foi escolhida a série estacionária que apresentava a melhor relação visual linear com os indicadores.

Por último, foram utilizados histogramas e boxplots para melhor observação de outliers nas séries. Esses gráficos ajudaram a entender melhor a presença e impacto de outliers, ajudando na tomada de decisão sobre a necessidade de tratamento e qual seria mais preciso.

2.3 Correlação

Para entender melhor as relações entre os diferentes indicadores analisados, foi gerada uma matriz de correlação, conforme ilustrado na Figura 1. Este gráfico de correlação nos permite visualizar rapidamente a força e a direção das relações lineares entre as variáveis. Essa análise é fundamental para a construção de modelos preditivos, pois permite selecionar variáveis que têm relações significativas entre si, evitando problemas de multicolinearidade e potencializando a acurácia dos modelos.

Apesar da relação dos indicadores Selic e IGP-M apresentar uma correlação relativamente alta em comparação com as outras correlações apresentadas na matriz de correlação, os testes de multicolinearidade dos modelos implementados não apresentaram significância estatística.

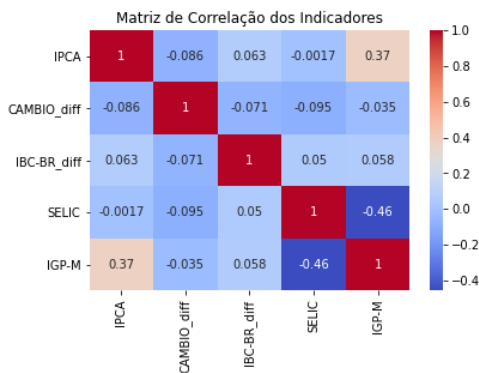


Fig. 1. Matriz de correlação dos indicadores macroeconômicos

Também foram calculadas correlações entre as séries de preço das ações e os indicadores econômicos e nenhuma das correlações foi significativa o suficiente para justificar o uso dos indicadores como preditores individuais no modelo.

3 Metodologia: Regressão Linear Múltipla

Será utilizada a análise de regressão linear múltipla para associar a série de fechamento das cotações das ações no setor de concessão com os indicadores macroeconômicos.

A regressão múltipla é um método usado para relacionar a variável dependente (ou resposta) com as variáveis independentes (ou preditoras). É dita linear justamente porque tem como objetivo montar uma equação com variáveis de primeiro grau que sirva para prever os valores que uma variável explicada terá, dado os valores das variáveis explicativas. Essa equação é dada por:

$$T_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon_i, \quad (2)$$

onde β_0 representa o intercepto da equação; β_i corresponde ao coeficiente linear da i -ésima variável; X_i representa a i -ésima variável explicativa e $\epsilon_i = Y_i - \hat{Y}_i$.

3.1 Premissas

Para que os resultados de uma regressão linear múltipla sejam válidos e interpretáveis, algumas premissas devem ser atendidas:

3.1.1 Linearidade. A relação entre a variável dependente e cada uma das variáveis preditoras precisa ser linear. Essa relação foi analisada por gráficos de dispersão durante a análise exploratória de dados. A série que foi escolhida foi a série estacionária que melhor representou uma relação linear.

3.1.2 Homocedasticidade. A variância dos resíduos deve ser constante para todos os valores das variáveis independentes. Em outras palavras, a dispersão dos resíduos deve ser aproximadamente a mesma em todos os níveis de predição.

3.1.3 Independência e Normalidade dos Erros. Os resíduos devem ser independentes uns dos outros. Isto significa que a ocorrência de um erro não deve influenciar a ocorrência de outro. Além disso, eles devem seguir uma distribuição normal.

3.1.4 Não multicolinearidade. As variáveis independentes não devem estar altamente correlacionadas entre si. Isso pois a multicolinearidade pode tornar difícil a determinação dos efeitos individuais das variáveis independentes na variável dependente podendo inflar a variância dos coeficientes estimados.

4 Implementação

Uma das primeiras decisões feitas na implementação do modelo foi a criação de uma variável dummy para capturar os efeitos da pandemia de COVID-19. Dado que os dados cobrem um período que inclui a pandemia, a inclusão desta dummy é necessária para isolar o impacto significativo deste evento nos dados. Esta variável dummy assume o valor de 1 durante o período da pandemia e 0 nos demais períodos, permitindo que o modelo de regressão capte melhor as variações atípicas. É importante ressaltar que os valores das estatísticas da dummy não devem ser analisados detalhadamente, pois o objetivo é que ela interfira na regressão para controlar os efeitos da pandemia, e não que seja considerada uma variável preditora principal.

No geral, foram realizados três modelos para cada série de ações. Todos os modelos começaram com a análise de estacionariedade e a visualização gráfica das séries. A inspeção gráfica é essencial, pois pode revelar tendências exponenciais, sugerindo que a melhor abordagem para tratar a série é a aplicação da transformação logarítmica seguida da diferenciação, um procedimento comum na modelagem de séries temporais. Em seguida, foram gerados gráficos de autocorrelação parcial para determinar a defasagem necessária para tornar a série estacionária.

Para todos os modelos foram testadas as séries diferenciadas e log diferenciadas e foi escolhido o modelo que possui melhor R^2 e menor AIC. O R^2 serve para medir o quanto da variabilidade total

das observações da série pode ser explicada pela reta de regressão traçada, o que significa que quanto maior o valor de R^2 , melhor será o ajuste do modelo. O coeficiente R^2 segue a seguinte fórmula:

$$R^2 = \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}, \quad (3)$$

ou seja, numerador representa a soma dos quadrados devido a regressão e o denominador é a soma dos quadrados totais.

O AIC é uma medida de ajuste do modelo que penaliza a complexidade, buscando encontrar o modelo que melhor se ajusta aos dados, proporcionando um equilíbrio entre robustez e parcimônia

$$AIC = 2k - 2\ln(L), \quad (4)$$

onde k representa o número de parâmetros do modelo e L a verossimilhança.

O primeiro modelo foi desenvolvido utilizando uma regressão múltipla stepwise para todas as variáveis, considerando as séries estacionarizadas e mantendo a variável dummy fixa. Isso significa que a dummy foi incluída no modelo de forma permanente e não pode ser eliminada, independentemente do ajuste realizado.

No segundo modelo de regressão múltipla, adotamos um procedimento específico para o tratamento de outliers. Utilizamos a eliminação de valores que se encontram fora dos limites estabelecidos pelos quantis inferiores (0.05) e superiores (0.95). Essa abordagem foi aplicada em um DataFrame contendo todas as variáveis relevantes para o modelo. Posteriormente, executamos o modelo de regressão múltipla utilizando as variáveis nesse novo formato, ou seja, após a remoção dos outliers de acordo com os critérios dos quantis.

No terceiro modelo, utilizamos os dados após a aplicação do procedimento para remoção de outliers, considerando o valor correto de defasagem para cada série temporal. Essa defasagem foi determinada através da análise da Função de Autocorrelação Parcial (PACF). Esta é uma função estatística que mede a relação entre duas observações de uma série temporal, levando em conta todas as observações intermediárias. Ao analisar a PACF, podemos identificar os lags ou defasagens que possuem uma correlação significativa com a variável de interesse, indicando o número correto de defasagens a serem consideradas na modelagem da série temporal.

É importante destacar que o uso de defasagem foi avaliado para as variáveis estacionárias sem a exclusão dos outliers, porém, em todas as séries, isso resultou em modelos com ajustes inferiores em comparação com os outros casos.

5 Análise de Resultados

Nesta seção, apresentamos e discutimos os resultados obtidos, explicando o significado das estatísticas de teste utilizadas para avaliar o modelo, bem como os procedimentos aplicados para cada variável. Os métodos do R^2 e do AIC foram utilizados para definir o melhor modelo e os métodos explicados nessa sessão são os que avaliam as premissas que os resultados devem seguir. É importante ressaltar que nem todos os indicadores econômicos foram incluídos em todas as regressões; os modelos foram ajustados iterativamente para eliminar indicadores que não contribuíam significativamente para a precisão do modelo.

5.1 Estatísticas de teste

5.1.1 Teste F : A estatística F é usada para testar a significância global do modelo de regressão. Ela compara a variância explicada pelo modelo com a variância residual. Um valor de F alto e um valor p associado baixo (< 0.05) indicam que o modelo, como um todo, é significativo para ao menos uma variável independente.

5.1.2 Durbin-Watson : O teste Durbin-Watson é usado para detectar a presença de autocorrelação dos resíduos em um modelo de regressão. Ele varia em um intervalo de $[0,4]$ e valores próximos de 2 indicam que os resíduos não estão correlacionados. Valores próximos de 0 indicam correlação positiva e valores próximos de 4 correlação negativa.

5.1.3 Skewness : Mede a simetria dos resíduos. Valores próximos de 0 indicam uma distribuição simétrica, valores negativos indicam assimetria a esquerda e valores positivos assimetria a direita.

5.1.4 Kurtosis : Um valor maior que 3 indica uma distribuição dos resíduos mais aguda do que o normal e um valor menor do que 3 indica uma distribuição mais achatada do que o normal.

5.1.5 Fator de inflação da variância (VIF) : É uma medida que quantifica o grau de multicolinearidade entre as variáveis independentes em um modelo de regressão. Se o VIF é igual a 1, então não existe correlação. Valores no intervalo (1,5) indicam que há uma correlação moderada mas que na maioria dos casos pode ser tolerada e valores maiores que 5 indicam uma multicolinearidade severa.

5.2 Resultados

5.2.1 Sabesp. A série da Sabesp que foi utilizada foi a série log-diferenciada. O modelo que obteve melhor ajuste foi o com o tratamento de outliers descrito na implementação e na Tabela 6 é possível ver os resultados.

Estatística	Resultado
R^2	0.283
R^2 ajustado	0.247
Log - Likelihood	102.74
AIC	-195.5
F-statistic	7.812
prob (F-statistic)	2.34e-5
Durbin-Watson	2.293
Kurtosis	3.613
Skew	-0.162

Tabela 6. Estatísticas da regressão múltipla da Sabesp

Na Tabela 7 podemos ver as variáveis que foram utilizadas com seus coeficientes, erro padrão, estatística t, p-valor e VIF.

O VIF de todos os indicadores, é muito próximo de 1 e isso ajuda a descartar a chance de multicolinearidade entre as variáveis independentes. A constante (intercepto) não é uma variável independente, então seu resultado é irrelevante. Essa conclusão também é indicada pelo teste de Durbin-Watson que apresenta um valor próximo de 2.

Já o teste de skew próxima de 0 indica pouca assimetria. O Kurtosis indica que a distribuição é um pouco pontiaguda em relação à uma normal. Isso pode ser visualizado na Figura 3.

Variável	coef	std err	t	p-valor	VIF
Constante	0.0907	0.031	2.943	0.004	14.804
Cambio	-0.5337	0.097	-5.521	0.000	1.121
Selic	-0.0787	0.037	-2.145	0.035	1.088
IGP-M	-0.0153	0.014	-1.129	0.262	1.020
Dummy	0.0117	0.028	0.420	0.674	1.037

Tabela 7. Coeficientes da regressão da Sabesp

Um Q-Q plot (Quantile-Quantile plot) é uma ferramenta gráfica utilizada para comparar a distribuição de um conjunto de dados com uma distribuição teórica, geralmente a distribuição normal. Se os dados seguem uma distribuição normal, eles devem acompanhar a linha diagonal representada por uma linha diagonal vermelha no gráfico na Figura 2. Nos

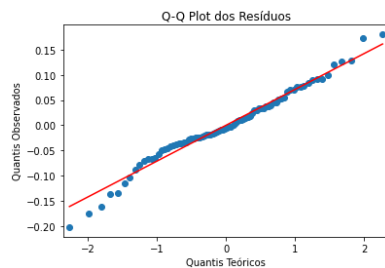


Fig. 2. Gráfico de QQ-plot do modelo de regressão múltipla para a SBSB

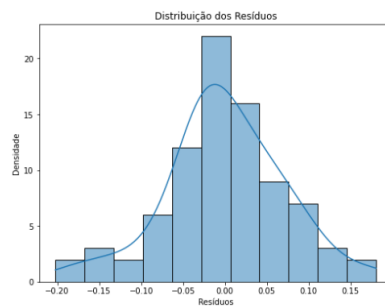


Fig. 3. Gráfico da distribuição do resíduo da Sabesp

Os resultados dos testes e do gráfico de distribuição do resíduo indicam a presença de outliers. O tratamento de outliers utilizando o método dos quantis foi realizado, mas mesmo assim, não foi suficiente para eliminar completamente a presença desses valores atípicos.

5.2.2 *Cosan S.A.* A série usada foi a log-diferenciada pelo comportamento exponencial da série crua e por ser a série que melhor estabeleceu uma relação com as variáveis durante a modelagem. Além disso, foi realizado o tratamento de outliers e o modelo que melhor se ajustou tem seu resumo nas Tabelas 8 e 9.

Estatística	Resultado
R^2	0.183
R^2 ajustado	0.143
Log - Likelihood	111.58
AIC	-213.2
F-statistic	4.596
prob (F-statistic)	0.00213
Durbin-Watson	2.605
Kurtosis	3.320
Skew	0.122

Tabela 8. Estatísticas da regressão múltipla da Cosan S.A

Na Tabela 9 podemos identificar que, entre os indicadores utilizados, o CÂMBIO foi o único que apresentou uma relação forte, enquanto o SELIC demonstrou uma relação moderada. Apesar disso, todos os indicadores apresentaram VIF próximos de um, o que nos permite acreditar na premissa de ausência de multicolinearidade.

Variável	coef	std err	t	p-valor	VIF
Constante	0.0769	0.035	2.227	0.029	21.7160
IPCA	-0.0329	0.033	-1.006	0.318	1.1844
CAMBIO	-0.4108	0.097	-4.230	0.000	1.2451
SELIC	-0.0636	0.038	-1.686	0.096	1.1546
Dummy	-0.035	0.053	-1.198	0.234	1.1481

Tabela 9. Coeficientes da regressão da Cosan S.A

Na Figura 4 Vemos o QQ-plot dos resíduos do modelo, no geral, apresentando poucos outliers.

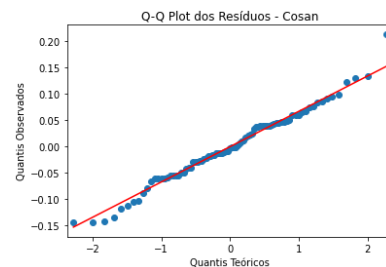


Fig. 4. QQ-plot dos resíduos da regressão da empresa Cosan S.A.

Na Figura 5 vemos o deslocamento da distribuição dos resíduos e que a curva possui uma cauda mais pesada no lado esquerdo,

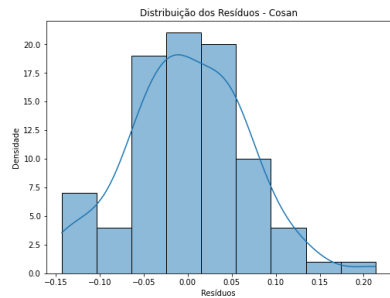


Fig. 5. Distribuição dos resíduos da empresa Cosan S.A.

5.2.3 *Taes*a. A série da empresa Taesa utilizada foi log-diferenciada com tratamento de outlier e apenas com as variáveis independentes CAMBIO e IBC-BR.

A estatística F, por apresentar um p-valor de 0.059, indica que há uma probabilidade de 5.9% de observar um valor de F de 2.678 ou mais extremo, assumindo que a hipótese nula é verdadeira. Esse p-valor está ligeiramente acima do limiar comum de 0.05, sugerindo que a evidência contra a hipótese nula não é suficientemente forte para ser considerada estatisticamente significativa ao nível de 5%. No entanto, como o p-valor está muito próximo de 0.05, o resultado pode ser considerado marginalmente significativo em uma consideração com um nível de significância mais elevado, como 10%.

Pela estatística Kurtosis também observada na Tabela 10, podemos ver que ela apresenta um valor aproximado de 1.399 acima do valor de referência 3. Isso indica que há uma presença maior de valores extremos (outliers) do que o esperado em uma distribuição normal.

Estatística	Resultado
R^2	0.094
R^2 ajustado	0.058
Log - Likelihood	97.965
AIC	-187.9
F-statistic	2.613
prob (F-statistic)	0.0573
Durbin-Watson	2.036
Kurtosis	4.399
Skew	0.372

Tabela 10. Estatísticas da regressão múltipla da Taesa.

Conforme apresentado na Tabela 11, o indicador macroeconômico com maior impacto é o câmbio, possuindo o coeficiente mais elevado em termos absolutos, acompanhado de uma alta estatística de teste e um p-valor inferior a 0.05. Adicionalmente, os valores do VIF para todos os indicadores são próximos de 1, sugerindo que não há problemas de multicolinearidade entre as variáveis.

Variável	coef	std err	t	p-valor	VIF
Constante	0.0147	0.009	1.681	0.097	1.1551
CAMBIO	-0.2310	0.108	-2.135	0.036	1.0875
Dummy	-0.0022	0.030	-0.072	0.943	1.0938
IBC-BR	-0.0039	0.002	-1.684	0.096	1.0072

Tabela 11. Coeficientes da regressão da Taesa.

Agora, pelas Figuras 6 e 7 Podemos observar a presença de outliers, conforme sugerido pelo teste de Kurtosis. Além disso, notamos que a distribuição está deslocada o que também foi indicado pelo teste skew.

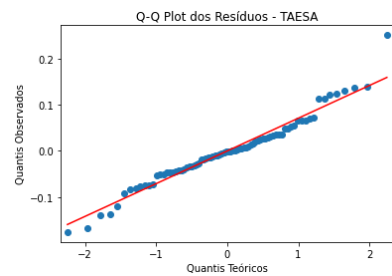


Fig. 6. QQ-plot dos resíduos da regressão da empresa Taesa

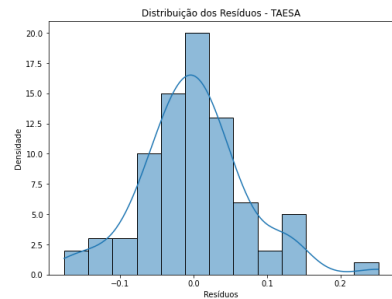


Fig. 7. Distribuição dos resíduos da empresa Taesa.

5.2.4 *Eletrobras*. Para a empresa Eletrobras, foi utilizada a série apenas diferenciada com tratamento de outlier e defasagem tanto dela quanto dos indicadores macroeconômicos.

Nas Tabelas 12 e 13 é possível perceber que, mesmo com o tratamento adequado de outliers, a utilização correta da defasagem e a transformação das séries em estacionárias, o método não foi capaz de capturar de maneira efetiva o efeito das variáveis independentes.

Isso é indicado pela presença significativa de outliers, conforme evidenciada pelos testes Durbin-Watson, Kurtosis e Skew, que estão fora dos valores padrão de referência.

Estatística	Resultado
R^2	0.092
R^2 ajustado	0.056
Log - Likelihood	-59.485
AIC	127.0
F-statistic	2.555
prob (F-statistic)	0.0616
Durbin-Watson	2.787
Kurtosis	10355
Skew	1.402

Tabela 12. Estatísticas da regressão múltipla da Eletrobras.

Variável	coef	std err	t	p-valor	VIF
Constante	1.6908	1.027	1.646	0.104	5.5575
CAMBIO	-14.5648	5.529	-2.634	0.010	1.0747
Dummy	0.4100	1.672	0.245	0.807	1.0218
IPCA	-2.7312	1.792	-1.524	0.132	1.0804

Tabela 13. Coeficientes da regressão da Eletrobras.

Essa situação pode ser mais claramente observada no gráfico QQ-plot, ilustrado na Figura 8, onde os valores cruzam a linha de referência, porém não seguem consistentemente a mesma direção que ela. Além disso, na Figura 9, que representa a distribuição dos resíduos, observamos valores consideravelmente acima na extremidade superior e uma cauda longa, indicando a presença de outliers e uma distribuição não normal dos resíduos.

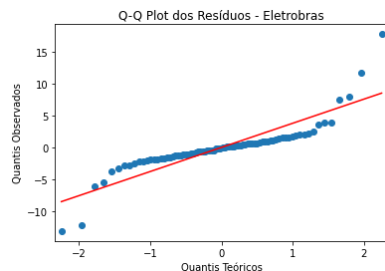


Fig. 8. QQ-plot dos resíduos da regressão da empresa Eletrobras.

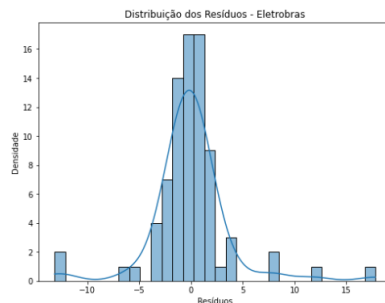


Fig. 9. Distribuição dos resíduos da empresa Eletrobras.

5.2.5 *Ecorodovias*. A série da empresa utilizada foi ela log-diferenciada sem tratamento de outlier.

Pela Tabela 14 podemos ver que ela apresentou um R^2 de 0.16 com uma Log-Likelihood de 139.68, ambos valores altos em comparação com os resultados das outras empresas. Além disso, apresenta uma teste de Durbin-Watson próximo de 2 indicando ausência de correlação entre os resíduos. Já o Kurtosis indica a presença de outliers.

Estatística	Resultado
R^2	0.184
R^2 ajustado	0.163
Log - Likelihood	139.68
AIC	-269.4
F-statistic	8.629
prob (F-statistic)	2.63e-06
Durbin-Watson	2.305
Kurtosis	3.979
Skew	0.021

Tabela 14. Estatísticas da regressão múltipla da Ecorodovias

Pela Tabela 15 podemos observar que as variáveis independentes utilizadas foram o IPCA, câmbio e SELIC, destacando o câmbio como aquela com a relação mais significativa. Além disso, ao considerarmos os valores do VIF, podemos concluir que não há preocupações significativas com multicolinearidade entre as variáveis.

Variável	coef	std err	t	p-valor	VIF
Constante	0.0563	0.030	1.895	0.060	13.5395
IPCA	-0.0284	0.027	-1.052	0.294	1.1545
CAMBIO	-0.3094	0.057	-5.396	0.000	1.0169
SELIC	-0.0347	0.034	-1.020	0.309	1.4562
Dummy	-0.0497	0.027	-1.860	0.065	1.595447

Tabela 15. Coeficientes da regressão da Ecorodovias

Pelo QQ-plot da Figura 10 percebemos que a maioria dos dados se ajusta bem à linha de referência, com exceção de três valores no início e no final que se destacam.

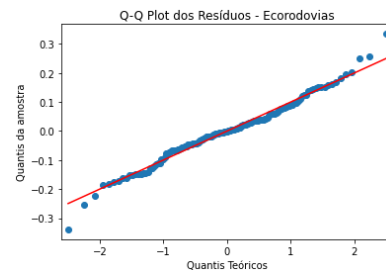


Fig. 10. QQ-plot dos resíduos da regressão da empresa Ecorodovias.

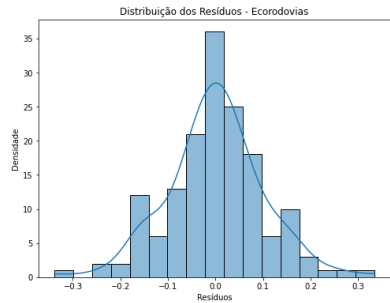


Fig. 11. Distribuição dos resíduos da empresa Ecorodovias.

Na Figura 11 conseguimos perceber principalmente um topo maior do que esperado.

5.2.6 CCR. Para a análise da empresa CCR, o valor de fechamento de suas ações foi utilizado sem qualquer transformação, pois essa série é naturalmente estacionária. No entanto, não foi observada uma relação linear evidente com as covariáveis. Embora a aplicação do log na variável independente pudesse potencialmente linearizar essa relação, tanto a série quanto as covariáveis não exibiam comportamento exponencial.

O intercepto foi retirado para essa modelagem pois resultou em uma melhora significativa em todas as estatísticas de análise e de resíduos. Isso implica que a linha de regressão passa pela origem, o que pode não ser realista em todos os contextos. Essa escolha pode resultar em uma modelagem menos flexível e a imposição de que a variável dependente seja zero quando todas as covariáveis forem zero, o que pode não refletir a realidade dos dados.

Apesar disso, o procedimento padrão utilizado para outras empresas também foi aplicado à CCR. O resultado mais eficaz foi obtido ao utilizar as séries com suas defasagens adequadas. Por meio de um método de regressão stepwise, além da variável dummy que foi forçada, apenas a variável SELIC foi considerada. Foram encontrados valores de coeficiente, p-valor e estatística t comparativamente satisfatórios, como mostrado na Tabela 17. É importante destacar que os intervalos de confiança (de 0.025 a 0.975) são amplos e não incluem o valor 0.

Na Tabela 16 podemos ver que o modelo possui um R^2 ajustado 0.817, ou seja, um valor muito superior ao valores encontrados para as outras empresas Além disso, a estatística F apresentou um valor significativamente alto de 339.9, com um Prob (F-statistic) muito baixo de 1.93e-56, confirmando a significância global do modelo.

A análise dos critérios de informação mostrou que o modelo proposto apresentou o menor AIC em comparação com outros modelos, evidenciando sua adequação e eficácia na explicação dos dados observados. Por outro lado, o teste de Durbin-Watson revelou uma autocorrelação positiva nos resíduos do modelo, sugerindo a necessidade de considerar essa questão ao interpretar os resultados.

Estatística	Resultado
R^2	0.819
R^2 ajustado	0.817
Log - Likelihood	-456.12
AIC	916.2
F-statistic	339.0
prob (F-statistic)	1.93e-56
Durbin-Watson	0.819
Kurtosis	3.283
Skew	-0.569

Tabela 16. Estatísticas da regressão múltipla da CCR

Variável	coef	std err	t	p-valor	VIF
SELIC	11.807	0.944	7.119	0.000	1.0411
Dummy	6.721	0.510	23.165	0.000	1.0411

Tabela 17. Coeficientes da regressão da CCR

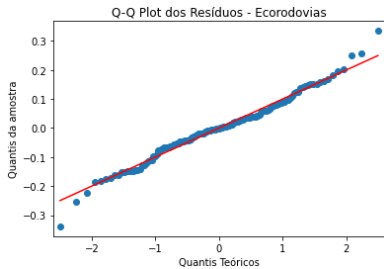


Fig. 12. QQ-plot dos resíduos da regressão da empresa CCR

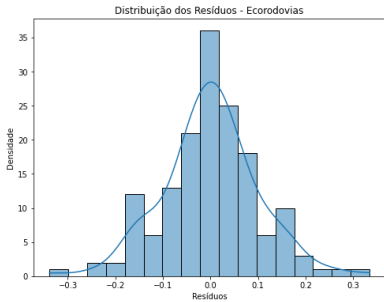


Fig. 13. Distribuição dos resíduos da empresa CCR.

6 Conclusão

Este estudo investigou a relação entre variáveis macroeconômicas e os retornos acionários das empresas do setor de concessões utilizando modelos de regressão múltipla. Os resultados obtidos foram mistos em relação ao que esperávamos responder.

No setor de ferrovias, a empresa CCR apresentou um resultado atípico, com uma série naturalmente estacionária e uma regressão

stepwise que eliminou todas as variáveis macroeconômicas exceto a SELIC, demonstrando uma relação de aproximadamente 81.7% (R^2 de 0.817). Por outro lado, para a Ecorodovias, o modelo com os indicadores Câmbio, IPCA e SELIC obteve um poder explicativo de 16.3% (R^2 de 0.163).

Analisando as duas empresas da área de saneamento, a Sabesp apresentou um melhor desempenho, alcançando um R^2 ajustado de 0.246 com os indicadores do Câmbio, SELIC e IGP-M, enquanto a Cosan obteve um R^2 ajustado de 0.143 para os indicadores do IPCA, Câmbio e SELIC.

Dos setores internos da área de concessão, o setor de energia foi o que teve mais dificuldade de adaptação ao modelo, com as variáveis que melhor se ajustaram à Taesa e Eletrobras resultando em um poder explicativo inferior a 10%.

Diante desses resultados, torna-se evidente que o modelo de regressão múltipla escolhido, embora tenha passado por diversas análises (Marco Bonobo, 2009), apresenta limitações significativas em sua capacidade de explicar as variações nos retornos acionários das empresas do setor de concessões. Essas descobertas destacam a complexidade e a natureza variável das relações entre variáveis macroeconômicas e o desempenho das empresas, sugerindo que uma abordagem exclusiva de regressão múltipla pode não ser adequada para capturar toda a diversidade e volatilidade dessas relações.

Note que, em alguns casos, o modelo demonstrou um desempenho relativamente satisfatório em comparação com estudos anteriores na área. Um poder explicativo próximo de 20% pode indicar que a empresa não está alheia às mudanças no cenário macroeconômico, fornecendo insights úteis sobre como essas variáveis influenciam seus retornos acionários. No entanto, devemos ressaltar que não há evidências suficientes para considerar este modelo como um bom modelo preditivo.

É importante destacar que uma abordagem alternativa, como o uso de modelos de Vetor Autoregressivo (VAR) ou de Volatilidade Condicionada Autorregressiva Generalizada (GARCH), poderia oferecer perspectivas adicionais sobre a relação dinâmica entre as variáveis macroeconômicas e os retornos das empresas de concessões. Esses modelos podem capturar melhor a dependência temporal e a volatilidade dos dados, permitindo uma análise mais completa das interações entre os diferentes fatores econômicos e o desempenho das empresas.

Assim, este estudo não apenas contribui para o entendimento da relação entre variáveis macroeconômicas e retornos acionários, mas também sugere direções para investigações futuras que utilizem abordagens mais avançadas e específicas para melhorar a precisão e a robustez das análises financeiras no setor de concessões.

7 Referências

- Economática
<https://economica.com/>.
- Banco Central do Brasil
<https://www.bcb.gov.br/>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427–431.
<https://doi.org/10.2307/2286348>
- Damodar N. Gujarati, Dawn C. Porter - *Econometria Básica*.
- Marcelo de Lima Balzana Filho, Ricardo Bordeaux (2013) - Uma análise da relação entre o retorno das ações do setor de construção civil brasileiro e indicadores macroeconômicos.
<https://doi.org/10.22409/engevista.v16i2.469>
- Marco Bonobo, Rebecca Barros, Silvia Matos, Carlos Carvalho (2009)- *Price Setting in a Variable Macroeconomic Environment: Evidence from Brazilian CPI*
- Repositório com as implementações do trabalho.
<https://github.com/jucarrica/Modelagem-Estatistica.git>