

# Um estudo comparativo de técnicas econométricas e machine learning combinadas para estimar o PIB Brasileiro

Eduardo Palhares Júnior<sup>a,b,\*</sup>, Flavio A. M. Cipparrone<sup>b</sup>, Pedro Brito<sup>b</sup>,  
Rodrigo D. L. S. Bueno<sup>c</sup>

<sup>a</sup>*Federal Institute of Amazonas, Itacoatiara 69101-030, Brazil*

<sup>b</sup>*Escola Politécnica, University of São Paulo, São Paulo 05424-970, Brazil*

<sup>c</sup>*Faculdade de Economia Administração e Contabilidade, University of São Paulo  
05508-010, Brazil*

---

## Abstract

Este estudo tem como objetivo construir um modelo que utiliza indicadores macroeconômicos para estimar o comportamento futuro do PIB brasileiro. Dentre os modelos preditivos, são propostas técnicas econométricas como OLS e modelos VAR, além de técnicas de machine learn baseadas em árvores, redes neurais e vetores de suporte. Foi observado a relevância da decomposição sazonal bem como o efeito das variáveis exógenas.

*Keywords:* Machine Learning, Econometrics, Brazilian GDP, Forecast, Economic Cycle

---

## 1. Introdução

### 1.1. Contexto e Justificativa

A análise e previsão das fases do ciclo econômico são fundamentais para os formuladores de políticas, investidores e empresas, pois facilitam a tomada de  
5 decisões mais informadas e eficientes, permitindo a adaptação às mudanças nas condições econômicas e a minimização de riscos associados. Nesse contexto,

---

\*Corresponding author

*Email addresses:* [eduardo.palharesjr@ifam.edu.br](mailto:eduardo.palharesjr@ifam.edu.br) (Eduardo Palhares Júnior),  
[flavio@lps.usp.br](mailto:flavio@lps.usp.br) (Flavio A. M. Cipparrone), (Pedro Brito), [delosso@usp.br](mailto:delosso@usp.br) (Rodrigo D. L. S. Bueno)

diversos modelos e técnicas têm sido desenvolvidos e aprimorados para compreender e prever a amplitude e a duração das fases do ciclo econômico.

#### *1.1.1. Avanço das Tecnologias de Aprendizado de Máquina*

10 Com o progresso das tecnologias de aprendizado de máquina e o aumento da disponibilidade de dados, novos algoritmos e abordagens têm sido explorados na previsão de ciclos econômicos, proporcionando alternativas aos modelos econométricos tradicionais, como o modelo VAR (Vector Autoregression). Alguns estudos relevantes nesta área incluem os trabalhos de Gogas et al. (2015),  
15 que utilizaram algoritmos de florestas aleatórias e redes neurais artificiais para prever o ciclo econômico na Grécia; Chionis et al. (2010), que exploraram o uso de redes neurais artificiais e modelos VAR na previsão do ciclo econômico na Espanha; e Araújo et al. (2015), que propuseram um modelo híbrido de redes neurais artificiais e modelos VAR para prever o mercado de ações em alta  
20 frequência.

#### *1.1.2. Eficácia das Abordagens de Previsão*

A eficácia das abordagens baseadas em aprendizado de máquina em comparação aos modelos econométricos tradicionais ainda é objeto de debate na literatura acadêmica. Alguns estudos, como os de Pagan e Soybilgen (2019)  
25 Soybilgen (2019) e Chen et al. (2020), demonstram que as técnicas de aprendizado de máquina podem oferecer previsões mais precisas e robustas do que os modelos econométricos tradicionais, enquanto outros sugerem que os modelos econométricos, como o VAR, continuam a ser relevantes e eficientes na previsão de ciclos econômicos (Hansen et al., 2019; Banbura et al., 2010).

30 A diferença de resultados entre os estudos pode estar relacionada a fatores como a seleção de variáveis, a qualidade dos dados, a escolha dos hiperparâmetros dos modelos de machine learning e a metodologia de avaliação do desempenho dos modelos. Dessa forma, torna-se importante realizar uma análise comparativa abrangente que considere esses fatores e identifique a melhor abordagem de previsão para o ciclo econômico.  
35

### *1.1.3. Necessidade de Investigação Aprofundada*

Diante das divergências na literatura e do crescente interesse em utilizar técnicas de aprendizado de máquina para a previsão de ciclos econômicos, torna-se necessário realizar uma investigação mais aprofundada que compare e analise a eficácia dessas abordagens em relação aos modelos econométricos tradicionais. Por exemplo, estudos recentes como o de Hotta et al. (2021) demonstram que as técnicas de aprendizado de máquina podem oferecer previsões mais precisas e robustas em comparação com os modelos econométricos tradicionais na previsão de ciclos econômicos. Contudo, outros estudos como o de González et al. (2018) sugerem que os modelos econométricos tradicionais, como o VAR, continuam a ser relevantes e eficientes na previsão de ciclos econômicos. Essa investigação permitirá identificar as melhores técnicas de previsão, considerando as características específicas do ciclo econômico, e fornecerá informações valiosas para a tomada de decisões informadas e eficientes no âmbito das políticas econômicas, investimentos e gestão empresarial.

## *1.2. Objetivos*

O objetivo principal deste estudo é realizar uma comparação abrangente entre técnicas de regressão baseadas em machine learning e a abordagem econométrica (modelo VAR) para prever a amplitude das fases do ciclo econômico no Brasil. Para atingir esse objetivo, iremos:

### *1.2.1. Revisão da Literatura*

Revisar a literatura sobre ciclos econômicos, abordando suas características, fases e teorias explicativas, bem como os modelos de previsão desenvolvidos ao longo do tempo, incluindo técnicas de machine learning e modelos econométricos, como o VAR.

### *1.2.2. Dados e Preparação*

Selecionar e preparar os dados econômicos relacionados ao ciclo econômico brasileiro, incluindo variáveis macroeconômicas como inflação (IPCA), taxa básica de juros (SELIC) e índice de confiança do consumidor, entre outras.

65 *1.2.3. Implementação de Modelos*

Implementar e ajustar os algoritmos econométricos e de machine learning selecionados utilizando técnicas para seleção das variáveis e otimização de hiperparâmetros.

*1.2.4. Avaliação e Comparação*

70 Avaliar e comparar o desempenho dos algoritmos de machine learning e dos modelos VAR utilizando métricas apropriadas, como raiz do erro quadrático médio (RMSE), coeficiente de determinação ( $R^2$ ) e critérios de informação (AIC e BIC) para os modelos de regressão.

*1.2.5. Discussão e Implicações*

75 Discutir os resultados obtidos, identificando as implicações práticas e teóricas das descobertas, as limitações do estudo e possíveis direções para pesquisas futuras que aprofundem o entendimento das melhores técnicas de previsão para a amplitude das fases do ciclo econômico no Brasil.

*1.3. Estrutura do trabalho*

80 *1.3.1. Capítulo 2: Revisão da Literatura*

Este capítulo apresenta uma revisão da literatura sobre ciclos econômicos, modelos de previsão, técnicas de machine learning e modelos econométricos, como o VAR. A revisão aborda conceitos fundamentais, teorias explicativas e desenvolvimentos recentes no campo.

85 *1.3.2. Capítulo 3: Metodologia*

O capítulo 3 descreve a metodologia utilizada, incluindo a seleção e preparação dos dados, implementação dos algoritmos de machine learning e do modelo VAR, e a avaliação e comparação de desempenho entre os modelos. Também discute as técnicas de otimização de hiperparâmetros e as métricas de avaliação empre-  
90 gadas.

### 1.3.3. Capítulo 4: Resultados

Este capítulo apresenta os resultados dos algoritmos de machine learning e do modelo VAR, analisando o desempenho de cada modelo em termos das métricas de avaliação selecionadas. Além disso, realiza uma comparação de desempenho entre as técnicas, identificando possíveis vantagens e desvantagens de cada abordagem.

### 1.3.4. Capítulo 5: Discussão

O capítulo 5 discute os resultados obtidos, analisando as implicações teóricas e práticas dos achados, as limitações do estudo e sugestões para pesquisas futuras. Esta discussão permite identificar oportunidades de aprimoramento das técnicas de previsão e áreas em que novas abordagens podem ser exploradas.

### 1.3.5. Capítulo 6: Conclusão

O último capítulo conclui o trabalho, resumindo os principais achados e contribuições do estudo para o campo da previsão de ciclos econômicos. Além disso, destaca a relevância das descobertas para formuladores de políticas, investidores e empresas, enfatizando a importância de técnicas eficazes de previsão para a tomada de decisões informadas e eficientes.

## 2. Revisão da Literatura

### 2.1. Ciclos Econômicos

Ciclos econômicos são flutuações recorrentes e não regulares na atividade econômica ao longo do tempo. Eles são caracterizados por períodos alternados de expansão e contração, resultando em crescimento e declínio na produção econômica, emprego e outras variáveis macroeconômicas. A compreensão dos ciclos econômicos é essencial para a tomada de decisões econômicas, políticas e de investimento.

Existem diversas teorias que buscam explicar os ciclos econômicos. A hipótese dos choques externos, proposta por Kalecki (1935), sugere que flutuações na atividade econômica são resultado de mudanças no ambiente externo, como

variações no preço das commodities e fatores geopolíticos. A teoria das ondas  
120 longas, de Kondratiev (1925), sugere que os ciclos são causados por grandes mu-  
danças tecnológicas em períodos de 40 a 60 anos. Schumpeter (1911) Schum-  
peter (1934) argumenta que a inovação tecnológica impulsiona o crescimento  
econômico. Keynes (1936) propõe que os ciclos são causados por flutuações na  
demanda agregada e na oferta agregada, enquanto Hayek (1931) enfatiza a im-  
125 portância do crédito na geração dos ciclos econômicos. Por fim, a teoria dos  
ciclos políticos, de Nordhaus (1975) e Samuelson (1964), sugere que os ciclos  
estão ligados a períodos eleitorais e mudanças na política econômica.

### *2.1.1. Fases dos Ciclos Econômicos*

Os ciclos econômicos consistem em quatro fases distintas: expansão, pico,  
130 contração e vale. A expansão é um período de crescimento econômico, no qual  
a produção, o emprego e a renda aumentam. O pico é o ponto em que a  
atividade econômica atinge seu nível mais alto antes de começar a declinar. A  
contração ocorre quando a atividade econômica diminui, levando a uma queda  
na produção, no emprego e na renda. O vale representa o ponto mais baixo da  
135 atividade econômica antes de começar a se recuperar novamente.

As quatro fases dos ciclos econômicos têm sido objeto de estudo em diversas  
áreas da economia. Autores como Burns and Mitchell (1946) realizaram um es-  
tudo pioneiro sobre as fases do ciclo econômico, identificando a existência dessas  
quatro fases e estabelecendo o uso de indicadores econômicos para a medição  
140 das fases. Desde então, outros autores, como Fatas and Mihov (2003), têm  
se dedicado a aprimorar a análise das fases do ciclo econômico, desenvolvendo  
métodos mais precisos para a sua identificação e classificação.

A expansão é geralmente vista como uma fase positiva do ciclo econômico,  
pois a economia está crescendo e se expandindo, o que leva a um aumento na  
145 produção, no emprego e na renda. Por outro lado, a contração é vista como  
uma fase negativa, pois a economia está diminuindo e encolhendo, levando a  
uma queda na produção, no emprego e na renda. O pico e o vale são pontos  
de transição entre as fases, onde a atividade econômica atinge o seu máximo ou

mínimo e começa a se mover na direção oposta.

150 A compreensão das fases do ciclo econômico é fundamental para a análise e  
previsão das tendências econômicas e para a tomada de decisões informadas nos  
negócios e investimentos. A identificação das fases permite que sejam tomadas  
medidas proativas para se adaptar às mudanças nas condições econômicas e  
reduzir os riscos associados.

### 155 2.1.2. *Causas dos Ciclos Econômicos*

Os ciclos econômicos são um fenômeno complexo e multifacetado que tem  
sido objeto de estudo de muitos economistas ao longo dos anos. Uma das teorias  
que buscam explicar as causas dos ciclos econômicos é a hipótese dos choques  
externos, que sugere que as flutuações na atividade econômica são resultado  
160 de mudanças no ambiente externo, como variações no preço das commodities,  
política governamental e fatores geopolíticos (Kalecki, 1935). Outra teoria é a  
teoria das ondas longas, proposta por Nikolai Kondratiev, que sugere que os  
ciclos econômicos são causados por grandes mudanças tecnológicas que ocorrem  
em períodos de 40 a 60 anos. Segundo Kondratiev, as economias passam por  
165 fases de crescimento impulsionadas pela introdução de novas tecnologias, que  
aumentam a produtividade e criam novas oportunidades de negócios. No en-  
tanto, após um período de crescimento, a economia pode experimentar uma crise  
de superprodução, quando a oferta excede a demanda, levando a uma queda nos  
preços e uma desaceleração econômica (Kondratiev, 1925).

170 O economista Joseph Schumpeter, em sua obra "Teoria do Desenvolvimento  
Econômico" (1911), sugeriu que a inovação tecnológica é um importante fator  
que impulsiona o crescimento econômico. Segundo ele, a introdução de no-  
vas tecnologias cria um aumento na produção e gera um ciclo de crescimento  
econômico. No entanto, esse crescimento não pode ser mantido por muito tempo  
175 e logo é seguido por uma fase de contração, uma vez que a economia se ajusta à  
nova tecnologia (Schumpeter, 1934). Alguns economistas, como John Maynard  
Keynes, argumentam que os ciclos econômicos são causados por fatores inter-  
nos, como flutuações na demanda agregada e na oferta agregada. De acordo

com a teoria keynesiana, durante uma recessão, a demanda agregada diminui,  
180 levando a uma queda na produção e no emprego. Para combater essa situação,  
o governo deve aumentar seus gastos, o que leva a um aumento na demanda  
agregada e, conseqüentemente, à recuperação econômica (Keynes, 1936).

Outros economistas, como Friedrich Hayek, enfatizam a importância do  
crédito na geração dos ciclos econômicos. Em sua obra "Preços e Produção"  
185 (1931), Hayek argumentou que a expansão do crédito leva a um aumento no  
investimento, que por sua vez estimula o crescimento econômico. No entanto,  
esse crescimento não pode ser sustentado por muito tempo e é seguido por uma  
fase de contração quando o crédito se torna mais difícil de ser obtido (Hayek,  
1931). Além disso, a teoria da escolha racional, desenvolvida por George Stigler  
190 e Gary Becker, sugere que as flutuações econômicas são resultado das decisões  
dos indivíduos, que tentam maximizar seu bem-estar econômico. Essa teoria  
sugere que as flutuações econômicas resultam de mudanças nas expectativas  
dos agentes econômicos, como consumidores e produtores, que tomam decisões  
com base em informações disponíveis e previsões sobre o futuro (Stigler, 1950;  
195 Becker, 1962).

A teoria dos ciclos políticos, proposta por William Nordhaus e Paul Samuelson,  
sugere que os ciclos econômicos estão ligados a períodos eleitorais e a mu-  
danças na política econômica. Segundo essa teoria, os governos tendem a adotar  
políticas econômicas expansionistas antes das eleições para aumentar o cresci-  
200 mento e reduzir o desemprego, o que pode levar a uma expansão insustentável  
e a uma eventual crise econômica (Nordhaus, 1975; Samuelson, 1964).

Hyman Minsky, por sua vez, argumenta que as crises financeiras são uma  
causa recorrente dos ciclos econômicos. Segundo a teoria de Minsky, períodos  
de estabilidade financeira levam a um aumento da tomada de riscos e da ala-  
205 vancagem, que eventualmente leva a uma crise financeira e uma desaceleração  
econômica (Minsky, 1975).

Compreender as múltiplas teorias e fatores que influenciam os ciclos econômicos  
é fundamental para a formulação de políticas eficazes e a tomada de decisões  
informadas no mundo dos negócios e investimentos. A abordagem holística



210 dessas teorias permite aos economistas ter uma visão mais ampla dos ciclos  
econômicos e de como diferentes fatores podem interagir para influenciar a ativi-  
dade econômica.

### *2.1.3. Indicadores dos Ciclos Econômicos*

Os economistas usam uma série de indicadores econômicos para analisar  
215 e prever ciclos econômicos. Esses indicadores podem ser classificados em três  
categorias: indicadores antecedentes, coincidentes e retardados. Indicadores an-  
tecedentes são aqueles que tendem a mudar antes das mudanças na atividade  
econômica, como a curva de rendimentos, índices de ações e índices de senti-  
mento dos consumidores. Indicadores coincidentes mudam ao mesmo tempo que  
220 a atividade econômica e incluem medidas como o Produto Interno Bruto (PIB)  
e a taxa de desemprego. Indicadores retardados mudam após as mudanças na  
atividade econômica e incluem estatísticas como a taxa de inflação e os salários  
médios.

Os indicadores dos ciclos econômicos são essenciais para a análise e previsão  
225 dos movimentos da economia. Dentre os indicadores antecedentes, podemos  
citar a curva de rendimentos, que é utilizada para prever mudanças na política  
monetária, assim como índices de ações, que refletem a confiança dos investi-  
dores em relação ao desempenho econômico futuro, e índices de sentimento dos  
consumidores, que podem ser um indicativo de futuras variações no consumo e  
230 investimento (Harvey, 1998).

Os indicadores coincidentes incluem medidas como o Produto Interno Bruto  
(PIB) e a taxa de desemprego. O PIB é um indicador importante que mede  
a produção econômica de um país em um determinado período de tempo e é  
utilizado para monitorar o desempenho da economia em relação ao seu potencial.  
235 Já a taxa de desemprego é um indicador do mercado de trabalho e pode ser  
utilizada para avaliar o grau de utilização dos recursos produtivos de um país  
(Okun, 1962).

Por fim, os indicadores retardados, que incluem a taxa de inflação e os  
salários médios, são utilizados para avaliar a dinâmica do mercado e como a

240 economia está se adaptando às mudanças. A taxa de inflação é um indicador importante que mede a variação dos preços dos bens e serviços e pode ser utilizada para avaliar o nível de estabilidade econômica. Os salários médios, por sua vez, são um indicador da dinâmica do mercado de trabalho e podem ser utilizados para avaliar a pressão inflacionária e a evolução dos custos das empresas  
245 (Phillips, 1958).

Em resumo, os indicadores dos ciclos econômicos são uma ferramenta importante para a análise e previsão do desempenho da economia. A escolha dos indicadores mais relevantes para essa tarefa dependerá do período de análise e das características específicas da economia em questão.

#### 250 2.1.4. *Teorias dos Ciclos Econômicos*

Diversas teorias foram propostas para explicar as flutuações na atividade econômica e suas causas. Entre as teorias mais conhecidas estão a teoria dos ciclos reais de negócios, a teoria do ciclo de crédito e a teoria das expectativas racionais.

255 A teoria dos ciclos reais de negócios, desenvolvida por Finn Kydland e Edward Prescott, enfatiza a importância dos choques de oferta na economia. Esses choques podem ser resultado de inovações tecnológicas, eventos naturais, mudanças nas políticas governamentais ou variações na produtividade. De acordo com essa teoria, os ciclos econômicos são gerados pelas respostas das empresas  
260 e dos trabalhadores a esses choques de oferta e pela interação dessas respostas no mercado.

A teoria do ciclo de crédito, proposta por economistas como Irving Fisher e Hyman Minsky, enfatiza o papel do crédito e dos mercados financeiros na geração de ciclos econômicos. Essa teoria sugere que os ciclos econômicos são  
265 causados pela expansão e contração do crédito na economia, que por sua vez afetam o consumo, o investimento e a produção. Durante períodos de expansão do crédito, o aumento da liquidez na economia estimula o consumo e o investimento, levando a um crescimento econômico. Por outro lado, a contração do crédito pode levar a uma diminuição no consumo e no investimento, resultando

270 em uma desaceleração econômica.

A teoria das expectativas racionais, proposta por Robert Lucas, sugere que as flutuações econômicas resultam das mudanças nas expectativas dos agentes econômicos, como consumidores e produtores, que tomam decisões com base em informações disponíveis e previsões sobre o futuro. Segundo essa teoria, as expectativas dos agentes econômicos afetam seu comportamento de consumo e investimento, o que influencia a atividade econômica.

Compreender essas teorias dos ciclos econômicos é essencial para a formulação de políticas econômicas eficazes e para a tomada de decisões informadas no mundo dos negócios e investimentos.

280 *Teoria dos Ciclos Reais de Negócios.* A teoria dos ciclos reais de negócios é uma teoria econômica que enfatiza a importância dos choques de oferta na economia e foi desenvolvida por Kydland and Prescott (1982). Os choques de oferta podem ser causados por inovações tecnológicas, eventos naturais, mudanças nas políticas governamentais ou variações na produtividade. Segundo essa teoria, os ciclos econômicos são gerados pelas respostas das empresas e dos trabalhadores a esses choques de oferta e pela interação dessas respostas no mercado.

A teoria dos ciclos reais de negócios destaca que os choques de oferta afetam a oferta de trabalho e de capital, bem como a produtividade e a produção, o que gera flutuações na economia que oscilam em torno de uma tendência de longo prazo. Quando ocorre um choque positivo, a economia experimenta um aumento na produtividade e no produto, o que gera um ciclo de expansão. Por outro lado, um choque negativo leva a uma diminuição na produtividade e no produto, o que leva a uma desaceleração econômica.

Os defensores da teoria dos ciclos reais de negócios argumentam que as políticas governamentais não são eficazes para estabilizar a economia em resposta às flutuações cíclicas. Em vez disso, as políticas governamentais devem se concentrar em promover a estabilidade monetária e fiscal de longo prazo, a fim de minimizar os efeitos desses choques na economia.

No entanto, a teoria dos ciclos reais de negócios tem sido criticada por alguns

300 economistas, que argumentam que ela não consegue explicar completamente as flutuações na atividade econômica e não leva em conta as políticas monetárias e fiscais implementadas pelos governos para estabilizar a economia.

Apesar disso, a teoria dos ciclos reais de negócios é uma importante contribuição para o entendimento dos ciclos econômicos, destacando a importância dos choques de oferta na economia e fornecendo uma base sólida para o desenvolvimento de políticas econômicas eficazes.

*Teoria do Ciclo de Crédito.* A teoria do ciclo de crédito é uma abordagem econômica que enfatiza a importância dos mercados financeiros na dinâmica econômica. Segundo Fisher (1933), a expansão do crédito é um dos principais 310 fatores que geram as flutuações econômicas. Ele destacou que a expansão do crédito leva a um aumento nos preços dos ativos, como ações e imóveis, que por sua vez leva a um aumento no consumo e no investimento. No entanto, quando a expansão do crédito se torna insustentável, ocorre um colapso dos preços dos ativos, levando a uma crise financeira e a uma desaceleração econômica.

315 Minsky (1992), outro economista que contribuiu para a teoria do ciclo de crédito, argumentou que a estabilidade financeira é um pré-requisito para a estabilidade econômica de longo prazo. Ele destacou a importância do endividamento excessivo e da especulação financeira na geração de crises financeiras e econômicas. Segundo ele, durante os períodos de prosperidade econômica, os investidores tendem a assumir riscos cada vez maiores, levando a um aumento 320 na especulação financeira e no endividamento excessivo. No entanto, quando ocorre uma crise financeira, ocorre um ajuste no mercado financeiro, levando a uma desaceleração econômica.

A teoria do ciclo de crédito tem sido usada como base para o desenvolvimento de políticas econômicas para evitar crises financeiras e estabilizar a economia. 325 As autoridades monetárias podem usar a política monetária para controlar a expansão do crédito e evitar a formação de bolhas financeiras. Além disso, as autoridades reguladoras podem implementar medidas para reduzir os riscos associados ao endividamento excessivo e à especulação financeira.

330 *Teoria das Expectativas Racionais.* A teoria das expectativas racionais, introduzida por Muth (1961) e desenvolvida posteriormente por Lucas e Sargent, é uma importante teoria para entender as flutuações econômicas. Essa teoria postula que os agentes econômicos formam expectativas sobre o futuro com base em informações disponíveis e previsões racionais. Por exemplo, as decisões de  
335 consumo de um indivíduo podem ser ajustadas com base em sua expectativa sobre o futuro da economia, como o aumento ou diminuição da renda ou a inflação. Da mesma forma, um produtor pode ajustar sua produção com base em suas expectativas sobre a demanda futura dos consumidores.

De acordo com a teoria das expectativas racionais, as flutuações econômicas  
340 resultam das mudanças nas expectativas dos agentes econômicos, que podem ser afetadas por diversos fatores, como políticas governamentais, choques externos e desenvolvimentos tecnológicos. Um estudo recente de Bom e Primiceri (2019), publicado na revista "Journal of Economic Dynamics and Control", utilizou um modelo dinâmico estocástico de equilíbrio geral (DSGE) para analisar  
345 como as expectativas dos consumidores sobre a inflação afetam a economia. Além disso, Carvalho et al. (2016), em um estudo publicado na revista "Econometrica", utilizaram a teoria das expectativas racionais para analisar o efeito das mudanças nas expectativas dos consumidores sobre a atividade econômica. O estudo mostrou que as expectativas dos consumidores são importantes para  
350 entender a dinâmica da economia, especialmente em períodos de incerteza.

Portanto, a teoria das expectativas racionais destaca a importância das expectativas dos agentes econômicos na dinâmica da economia e é fundamental para entender como as flutuações econômicas ocorrem. É importante ressaltar que políticas econômicas podem afetar as expectativas dos agentes e, consequentemente, a economia como um todo.  
355

*Teoria dos ciclos políticos.* A teoria dos ciclos políticos é uma abordagem que sugere que os ciclos econômicos estão relacionados a períodos eleitorais e a mudanças na política econômica. Segundo Nordhaus e Samuelson, os governos tendem a adotar políticas econômicas expansionistas antes das eleições para au-

360 mentar o crescimento e reduzir o desemprego, o que pode levar a uma expansão insustentável e eventual crise econômica (Nordhaus, 1975; Samuelson, 1964).

Essa teoria tem sido objeto de intenso debate e pesquisa na área de economia política, com base na ideia de que os governos possuem incentivos políticos para influenciar a economia de modo a favorecer suas chances eleitorais (Drazen, 365 2000). No entanto, a validade empírica da teoria é controversa, com alguns estudos apoiando e outros refutando sua relevância (Alesina e Sachs, 1988; Hibbs, 1987)(Alesina and Sachs, 1988; Hibbs, 1977).

Compreender a influência da política sobre os ciclos econômicos é crucial para entender como as decisões políticas podem afetar a economia e, consequentemente, os negócios e investimentos. Portanto, é importante que os economistas 370 e os tomadores de decisão considerem as implicações políticas em suas análises e estratégias econômicas.

## *2.2. Modelos de Previsão de Ciclos Econômicos*

A previsão dos ciclos econômicos é essencial para a tomada de decisões em investimentos e políticas. Diversos modelos foram desenvolvidos ao longo dos anos 375 para prever esses ciclos, utilizando diferentes abordagens e variáveis. Abaixo, apresentamos alguns dos modelos mais relevantes e suas características:

### *2.2.1. Modelos Baseados na Curva de Rendimento*

Os modelos baseados na curva de rendimentos têm sido objeto de estudos empíricos realizados por diversos autores, como Estrella et al. (1991), Estrella 380 and Mishkin (1996), Estrella and Mishkin (1998), Estrella et al. (2003), Estrella and Trubin (2006), Chionis et al. (2010) e Gogas et al. (2015), que investigam a relação entre a curva de rendimentos e a atividade econômica. A curva de rendimentos, que é uma representação gráfica da relação entre os juros de diferentes prazos de títulos do governo, é uma variável importante nos modelos baseados 385 nessa abordagem, que buscam entender como essa relação entre os juros de curto e longo prazo pode ser usada para prever o comportamento futuro da economia.

Uma curva de rendimentos invertida, onde os juros de curto prazo são mais altos do que os de longo prazo, tem sido um indicador confiável de recessões futuras em muitos países. Segundo a teoria econômica, uma curva de rendimentos invertida reflete a expectativa de uma desaceleração econômica futura e, portanto, pode ser um sinal de alerta para os investidores e os formuladores de políticas econômicas.

Os modelos baseados na curva de rendimentos são úteis porque, diferentemente de outros modelos de previsão, não exigem a construção de modelos econômicos complexos ou a escolha de um grande número de variáveis econômicas. Em vez disso, esses modelos se baseiam em uma variável simples e facilmente mensurável - a curva de rendimentos - e têm mostrado boa precisão na previsão dos ciclos econômicos.

Em resumo, os modelos baseados na curva de rendimentos têm se mostrado eficazes na previsão dos ciclos econômicos e são amplamente utilizados por investidores e formuladores de políticas econômicas. É importante destacar que o entendimento desses modelos é fundamental para uma tomada de decisão informada e precisa no mundo dos negócios e investimentos.

### 2.2.2. Modelos de Indicadores Líderes

Os modelos de indicadores líderes são amplamente utilizados na previsão dos ciclos econômicos, uma vez que são capazes de antecipar mudanças na atividade econômica com base em variáveis que tendem a mudar antes do próprio ciclo. Tais variáveis podem incluir, por exemplo, dados sobre pedidos de bens duráveis, índices de confiança do consumidor e do empresário, e taxas de desemprego.

O Índice de Indicadores Líderes do Conference Board, criado em 1960, é um dos modelos mais conhecidos nessa categoria. Ele combina dez indicadores econômicos, como as taxas de juros, o número de novas construções de residências, as solicitações de seguro-desemprego e o índice de preços de ações, para prever a direção futura da economia dos Estados Unidos. Segundo um estudo de Filardo and Gordon (1998), o Índice de Indicadores Líderes tem apresentado um bom desempenho na previsão dos ciclos econômicos americanos.

Outro modelo conhecido na categoria de indicadores líderes é o Índice de Atividade Econômica do Banco Central do Brasil. Esse modelo é composto  
420 por uma combinação de indicadores econômicos, como a produção de bens de consumo duráveis e a utilização da capacidade instalada na indústria, que têm demonstrado uma forte relação com a atividade econômica no Brasil. De acordo com um estudo de Barreto et al. (2015), o Índice de Atividade Econômica tem sido um bom indicador para antecipar a direção do ciclo econômico brasileiro.

425 Em resumo, os modelos de indicadores líderes são importantes ferramentas para prever os ciclos econômicos, uma vez que são capazes de antecipar mudanças na atividade econômica com base em variáveis que tendem a mudar antes do próprio ciclo. O Índice de Indicadores Líderes do Conference Board e o Índice de Atividade Econômica do Banco Central do Brasil são exemplos de  
430 modelos bem-sucedidos nessa categoria.

### *2.2.3. Modelos Econométricos*

Os modelos econométricos utilizam técnicas estatísticas para estimar as relações entre variáveis econômicas e prever o comportamento futuro da economia. Um exemplo comum desses modelos é o modelo autoregressivo de médias  
435 móveis (ARMA), que combina processos autoregressivos e de médias móveis para modelar séries temporais econômicas.

### *2.2.4. Modelos Estruturais*

Os modelos estruturais são baseados em teorias econômicas e buscam capturar as relações fundamentais entre variáveis econômicas. Um exemplo conhecido é o modelo de equilíbrio geral dinâmico estocástico (DSGE), que utiliza equações baseadas na teoria econômica para modelar o comportamento de  
440 agentes econômicos e as interações entre eles.

### *2.2.5. Modelos de Aprendizado de Máquina*

O aprendizado de máquina tem se mostrado uma abordagem promissora  
445 na previsão de ciclos econômicos devido à sua capacidade de processar grandes volumes de dados e adaptar-se a padrões complexos e não lineares. Alguns dos



principais métodos de aprendizado de máquina aplicados à previsão de ciclos econômicos incluem:

*Florestas Aleatórias (RF)*. As Florestas Aleatórias são um método de aprendi-  
450 zado de conjunto que combina várias árvores de decisão para melhorar a pre-  
cisão e evitar o sobreajuste. Cada árvore é treinada em uma amostra aleatória  
dos dados, e a previsão final é obtida por meio de votação majoritária ou média  
das previsões individuais. As RF têm sido aplicadas na previsão de ciclos  
econômicos e na identificação de indicadores importantes para a detecção de  
455 recessões.

*Redes Neurais Artificiais*. As Redes Neurais Artificiais são modelos computa-  
cionais inspirados no funcionamento do cérebro humano. Elas consistem em  
camadas de neurônios artificiais interconectados que podem ser treinados para  
reconhecer padrões e fazer previsões com base em dados históricos. As RNA  
460 têm sido aplicadas com sucesso na previsão de várias variáveis econômicas, como  
crescimento do PIB, inflação e taxas de desemprego.

*Máquinas de Vetores de Suporte (SVM)*. As Máquinas de Vetores de Suporte  
são modelos de aprendizado supervisionado que podem ser utilizados para prob-  
lemas de classificação e regressão. Elas funcionam encontrando o hiperplano que  
465 melhor separa as classes ou minimiza o erro de previsão nos dados de treina-  
mento. As SVM têm sido aplicadas na previsão de ciclos econômicos, identifi-  
cando pontos de virada e classificando períodos de expansão e recessão.

Em conclusão, os modelos de aprendizado de máquina oferecem uma abor-  
dagem poderosa e flexível para a previsão de ciclos econômicos, complementando  
470 e, em alguns casos, superando as técnicas tradicionais. À medida que mais dados  
econômicos se tornam disponíveis e os algoritmos de aprendizado de máquina  
continuam a evoluir, é provável que essa área de pesquisa se expanda ainda mais  
no futuro.

### 2.3. Técnicas de Machine Learning

475 As técnicas de aprendizado de máquina têm sido aplicadas em diversos campos, como finanças, saúde, engenharia e previsão econômica. Alpaydin (2010) fornece uma introdução ao aprendizado de máquina, destacando suas principais técnicas e aplicações. Nesta seção, discutimos algumas aplicações específicas do aprendizado de máquina na previsão econômica e financeira.

#### 480 2.3.1. Previsão do Mercado de Ações

Várias técnicas de aprendizado de máquina têm sido empregadas na previsão de retornos e tendências no mercado de ações. Enke and Thawornwong (2005) aplicam redes neurais para prever retornos no mercado de ações e descobrem que esses modelos superam técnicas estatísticas tradicionais em termos de precisão. Güreşen et al. (2011) usam redes neurais artificiais para prever índices do mercado de ações e demonstram que essa abordagem melhora significativamente a precisão das previsões em comparação com outros métodos.

#### 2.3.2. Previsão de Alta Frequência no Mercado de Ações

Araújo et al. (2015) propõem um modelo híbrido para prever o mercado de ações em alta frequência, combinando técnicas de aprendizado de máquina, como redes neurais artificiais e algoritmos genéticos, com análise técnica. Esse modelo híbrido é capaz de capturar padrões complexos e não lineares nos dados de alta frequência, levando a melhores previsões do comportamento do mercado.

#### 2.3.3. Previsão de Falências

495 O aprendizado de máquina também tem sido utilizado para prever falências e insolvências de empresas. Sun et al. (2014) aplicam técnicas de aprendizado de máquina, como Support Vector Machines (SVM) e Random Forests (RF), para prever falências de empresas na China. Os resultados mostram que os modelos de aprendizado de máquina superam os métodos tradicionais de previsão de falências, como análise discriminante e regressão logística.

500

#### 2.3.4. *Previsão Macroeconômica*

O aprendizado de máquina também pode ser aplicado para prever variáveis macroeconômicas, como inflação, crescimento do PIB e taxa de desemprego. Choi and Varian (2012) utilizam métodos de aprendizado de máquina para prever a inflação nos Estados Unidos, utilizando dados de buscas do Google como uma fonte adicional de informação. Seus resultados mostram que a combinação de dados de buscas do Google com técnicas de aprendizado de máquina pode melhorar significativamente a precisão das previsões.

Em resumo, as técnicas de aprendizado de máquina têm mostrado um potencial considerável na previsão econômica e financeira, superando, em muitos casos, os métodos tradicionais. À medida que mais dados se tornam disponíveis e os algoritmos de aprendizado de máquina continuam a evoluir, é provável que o uso dessas técnicas na previsão econômica se torne cada vez mais comum.

#### 2.4. *Modelos Econométricos: VAR*

Os modelos econométricos desempenham um papel fundamental na análise e previsão de ciclos econômicos. Entre eles, os modelos VAR (Vector Autoregressive) são particularmente úteis para examinar as relações dinâmicas entre diferentes variáveis econômicas e financeiras.

##### 2.4.1. *Introdução aos Modelos VAR*

Os modelos VAR são uma classe de modelos econométricos multivariados que capturam a dinâmica entre várias séries temporais. Sims (1980) introduziu os modelos VAR como uma alternativa aos modelos estruturais tradicionais, argumentando que eles fornecem uma abordagem mais flexível e menos restritiva para modelar a interação entre variáveis econômicas.

##### 2.4.2. *Estimativa e Interpretação dos Modelos VAR*

A estimativa dos parâmetros em um modelo VAR é geralmente realizada através do método dos mínimos quadrados ordinários (OLS). Uma vez que os parâmetros são estimados, é possível analisar as relações entre as variáveis

através de funções de impulso-resposta e decomposições da variância do erro de  
530 previsão.

### 2.4.3. VARMAX

A metodologia VARMAX (Vector Autoregressive Moving Average with exogenous variables) é uma extensão do modelo VAR (Vector Autoregression), incorporando componentes de média móvel e variáveis exógenas (Petris et al.,  
535 2009). O método VARMAX tem sido amplamente utilizado em séries temporais multivariadas, permitindo a modelagem flexível de relações dinâmicas entre diversas variáveis. Ao incorporar a capacidade de lidar com efeitos sazonais, tendências e fatores externos, o VARMAX se destaca na captura da complexidade inerente a conjuntos de dados que exibem múltiplas interdependências  
540 temporais.

### 2.4.4. SARIMAX

O método SARIMAX (Seasonal Autoregressive Integrated Moving Average with exogenous variables) representa uma abordagem poderosa para a análise de séries temporais univariadas (Hyndman and Athanasopoulos, 2021). Integrando componentes sazonais ao tradicional modelo ARIMA (Autoregressive  
545 Integrated Moving Average), o SARIMAX torna-se particularmente eficaz na modelagem de dados com padrões cíclicos recorrentes. A inclusão de variáveis exógenas permite ao SARIMAX incorporar informações externas, fortalecendo sua capacidade de previsão em situações em que fatores adicionais influenciam  
550 a série temporal em estudo.

### 2.4.5. Testes de Causalidade de Granger

Granger (1969) introduziu o conceito de causalidade de Granger, que é uma abordagem estatística para testar se uma variável tem poder preditivo em relação a outra variável em um contexto de séries temporais. Os testes de  
555 causalidade de Granger são frequentemente aplicados em modelos VAR para identificar relações de causalidade entre variáveis econômicas e financeiras.

#### 2.4.6. *Teste de co-integração de Johansen*

Desenvolvido por Søren Johansen, o teste é amplamente empregado para avaliar se existem vetores de co-integração entre diferentes séries temporais, indicando a presença de relações de equilíbrio de longo prazo (Johansen, 1991). Ao contrário de abordagens que se concentram apenas nas relações de curto prazo, o teste de co-integração de Johansen oferece uma visão mais abrangente, revelando a presença de padrões de equilíbrio a longo prazo entre variáveis economicamente relacionadas. Esse método é particularmente valioso em contextos nos quais as variáveis podem compartilhar uma tendência comum, permitindo uma compreensão mais profunda das dinâmicas subjacentes e facilitando análises mais precisas das relações de causalidade e previsões de longo prazo.

### 3. Metodologia

O objetivo geral desta metodologia é o estudo do comportamento dos ciclos econômicos, com base em técnicas econométricas e de aprendizado de máquina. Especificamente, o contexto escolhido para este estudo foi o cenário macroeconômico brasileiro, por meio de um modelo que descreve o comportamento do PIB em função de outras variáveis econômicas.

É importante ressaltar que a escolha do PIB como variável de interesse ocorreu devido ao contexto socioeconômico mundial decorrente da pandemia da Covid-19, dado o risco de recessão econômica generalizada. Contudo, o modelo é bastante flexível e permite análises mais amplas de outras variáveis de interesse, bastando alterar as hipóteses de pesquisa.

Para facilitar a análise e melhoria contínua do projeto, as etapas foram modularizadas para facilitar a liberação de estágios e compreensão dos fenômenos intermediários, conforme diagrama a seguir.

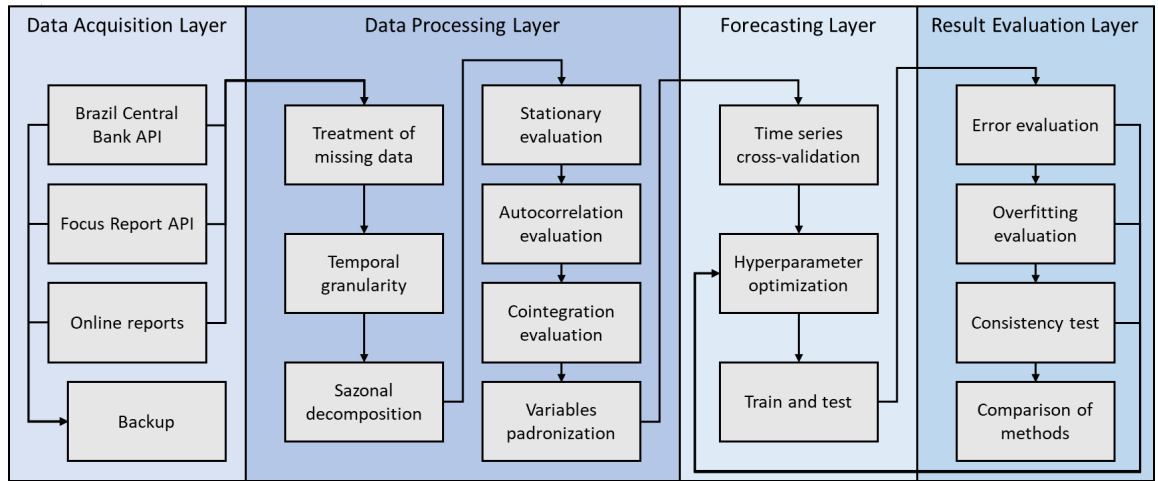


Figure 1: General architecture and methodology details

### 3.1. Dados e Variáveis

Para realizar uma análise abrangente e prever as fases do ciclo econômico no Brasil, este estudo utilizará dados macroeconômicos relevantes e confiáveis.

585 As variáveis selecionadas abrangem diferentes aspectos da economia e são amplamente utilizadas na literatura econômica.

Table 1: List and description of the economic variables

Economic Variable	Description
PIB	GDP monthly
IPCA	Broad National Consumer Price Index
IPA	Broad Producer Price Index
IPEM	Physical Production - Mineral extraction
IPIT	Physical Production - Capital goods
IPBC	Physical Production - Intermediate goods
IPBCD	Physical Production - Durable goods
IVVV	Sales volume index in the retail sector - Vehicles and motor-cycles, spare parts - Brazil

VVCCCL	Sales of factory authorized vehicle outlets - Light commercial cars sales
VVCC	Sales of factory authorized vehicle outlets - Trucks sales
IEF	Future expectations index
ICC	Consumer confidence index
SPUB	Credit operations outstanding of financial institutions under public control - Total
SPVT	Credit operations outstanding of financial institutions under private control - Total
M1	Money supply - M1 (working day balance average)
M2	Broad money supply - M2 (end-of-periodo balance)
PIB_0y	GDP expectations for the current year
PIB_1-0y	Short GDP expectation spread (1 year - current year)
PIB_4-1y	Long GDP expectation spread (4 years - 1 year)
Selic_0y	SELIC interest rate expectations for the current year
Selic_1-0y	Short SELIC expectation spread (1 year - current year)
Selic_4-1y	Long SELIC expectation spread (4 years - 1 year)
IPCA_0y	IPCA expectations for the current year
IPCA_1-0y	Short IPCA expectation spread (1 year - current year)
IPCA_4-1y	Long IPCA expectation spread (4 years - 1 year)
TC_0y	Exchange rate expectations for the current year
TC_1-0y	Short TC expectation spread (1 year - current year)
TC_4-1y	Long TC expectation spread (4 years - 1 year)
SBC_0y	Expectations of the trade balance balance for the current year
SBC_1-0y	Short SBC expectation spread (1 year - current year)
SBC_4-1y	Long SBC expectation spread (4 years - 1 year)
RP_0y	Primary outcome expectations for the current year
RP_1-0y	Short RP expectation spread (1 year - current year)
RP_4-1y	Long RP expectation spread (4 years - 1 year)
DLSP_0y	Public sector net debt expectations for the current year

DLSP_1-0y	Short DLSP expectation spread (1 year - current year)
DLSP_4-1y	Long DLSP expectation spread (4 years - 1 year)
ID_0y	Direct investment expectations for the current year
ID_1-0y	Short ID expectation spread (1 year - current year)
ID_4-1y	Long ID expectation spread (4 years - 1 year)
DL_6-1q	Interbank deposit rate spread (18 months – 3 months)
DL_5-1y	Interbank deposit rate spread (5 years – 1 year)
DL_10-2y	Interbank deposit rate spread (10 years – 2 years)
IBrX	IBrX100 - Brazilian Index
IBPO	Brazilian Corrugated Cardboard Index
Embi	Emerging Markets Bond Index Plus (Brazil Risk)
EEP	Electricity price
ITT	Terms of trade index
BOI	Live cattle futures
SOJA	Soybean futures
COBRE	Cooper futures
BRENT	Brent oil futures
BCOM	Bloomberg Commodity Index

---



Table 2: Summary statistics of the economic variables

Economic Variable	Min	Median	Max	Range	Mean	Std dev	Skewness	Kurtosis
PIB	159967,20	467253,00	927954,20	767987,00	457327,57	189127,43	0,31	-0,73
IPCA	-0,68	0,44	1,62	2,30	0,46	0,34	0,22	1,06
IPA	-1,77	0,50	5,92	7,69	0,62	1,16	1,14	3,04
IPEM	78,00	110,30	133,20	55,20	109,25	10,56	-0,27	-0,42
IPIT	67,30	107,20	132,50	65,20	107,40	11,78	-0,10	-0,27
IPBC	42,50	95,70	143,20	100,70	99,18	19,38	0,18	-0,60
IPBCD	20,00	130,00	182,70	162,70	129,83	26,49	-0,38	0,60
IVVV	42,00	95,46	149,08	107,08	93,63	20,35	-0,16	-0,32
VVCCL	11857,00	35059,00	79746,00	67889,00	38617,25	15933,56	0,88	-0,14
VVCC	2611,00	8740,00	17457,00	14846,00	8983,81	3208,87	0,21	-0,80
IEF	98,71	135,51	171,70	72,99	137,28	16,59	0,01	-0,60
ICC	84,55	126,49	170,18	85,63	126,66	21,76	0,16	-0,99
SPUB	196311,00	1413494,00	2306324,00	2110013,00	1184663,50	650419,43	-0,23	-1,45
SPVT	309382,00	1319846,00	3111224,00	2801842,00	1311691,32	678522,87	0,82	0,38
M1	117280591,00	316643361,00	644330793,00	527050202,00	322774325,12	143439889,95	0,80	-0,10
M2	488196033,00	1984883223,00	5259583670,00	4771387637,00	2098546881,79	1213447678,47	0,69	-0,26

PIB_0y	-6,55	2,39	7,55	14,1	1,9	2,76	-0,77	0,64
PIB_1-0y	-4,2	0,4	10,02	14,22	0,9	2,32	1,33	3,5
PIB_4-1y	-2,91	0,89	8,94	11,85	1,32	2,32	0,92	0,98
Selic_0y	1,95	10,99	18,14	16,19	10,42	3,68	-0,25	-0,32
Selic_1-0y	-3,84	-0,13	2,03	5,87	-0,37	1,35	-0,54	-0,46
Selic_4-1y	-6,33	-1,28	3,95	10,28	-1,37	2,45	0,02	-0,74
IPCA_0y	1,57	5,27	10,37	8,8	5,26	1,63	0,53	0,56
IPCA_1-0y	-5,1	-0,14	1,58	6,68	-0,56	1,2	-1,35	2,02
IPCA_4-1y	-6,96	-0,63	1,82	8,78	-1,02	1,51	-1,23	1,9
TC_0y	1,59	2,48	5,59	4	3,01	1,22	0,72	-0,75
TC_1-0y	-0,31	0,06	0,25	0,56	0,05	0,09	-0,96	1,89
TC_4-1y	-0,53	0,18	0,59	1,12	0,16	0,21	-0,84	1,36
SBC_0y	-0,3	37,34	70,91	71,21	34,24	19,99	-0,04	-1,27
SBC_1-0y	-14,05	-4,71	16,31	30,36	-3,7	5,34	0,95	1,01
SBC_4-1y	-31,32	-8,78	41,87	73,19	-6,68	11,51	1	1,18
RP_0y	-11,7	1,51	4,73	16,43	0,71	3,31	-1,47	3,19
RP_1-0y	-2,06	0,26	8,75	10,81	0,5	1,56	4,01	17,36
RP_4-1y	-1,36	0,48	10,97	12,33	1,15	2,16	2,71	9,26
DLSP_0y	34,01	43,58	68,12	34,11	46,92	9,66	0,39	-1,09

DLSP_1-0y	-2,04	0,18	5,01	7,05	0,56	2,08	0,38	-1,35
DLSP_4-1y	-9,51	-0,61	12,86	22,37	0,42	6,44	0,3	-1,45
ID_0y	13,2	56,48	86,14	72,94	50,28	20,98	-0,34	-1,05
ID_1-0y	-9,55	0,95	16,48	26,03	1,81	4,29	0,94	1,44
ID_4-1y	-9,26	3,28	55,48	64,74	6,03	8,93	1,52	4,28
DI_6-1q	-0,02	0	0,04	0,06	0	0,01	0,05	-0,37
DI_5-1y	-0,03	0,01	0,04	0,07	0,01	0,02	-0,04	-0,6
DI_10-2y	-0,02	0,01	0,04	0,06	0,01	0,01	0,57	-0,18
IBrX	7478,66	21562	54822,4	47343,74	26123,11	12007,67	0,74	-0,54
IBPO	157100	280560	371180	214080	274438,16	47365,2	-0,34	-0,53
Embi	136	252	532	396	271,32	79,98	0,88	0,57
EEP	12,2	109,79	804,54	792,34	176,16	178,02	1,62	2,28
ITT	85,77	101,28	122,43	36,66	102,19	9,17	0,25	-0,68
BOI	75,03	113,53	169,95	94,93	113,59	21,87	0,4	-0,55
SOJA	507	1017	1764,5	1257,5	1083,38	296,48	0,16	-0,79
COBRE	1,41	3,1	4,69	3,28	3,07	0,76	-0,11	-0,34
BRENT	24,74	71,59	140,67	115,93	76,09	24,69	0,36	-0,87
BCOM	190,55	350,25	664,37	473,82	368,81	89,28	0,67	0,46

---

### 3.1.1. *Indicadores de ciclo*

A análise dos indicadores de ciclo econômico brasileiro abrange quatro principais segmentos: preço, atividade econômica, confiança e monetários. No âmbito dos preços, busca-se compreender as tendências inflacionárias, enquanto a atividade econômica engloba indicadores que refletem o desempenho global da economia, como o Produto Interno Bruto (PIB). A confiança, medida por índices empresariais e do consumidor, oferece insights sobre as expectativas e a disposição para investir.

Por fim, os indicadores monetários, como taxas de juros e oferta de moeda, são cruciais para entender as condições financeiras e a eficácia das políticas monetárias. A análise conjunta desses segmentos proporciona uma visão abrangente do ciclo econômico, permitindo a identificação de fases de expansão, contração ou estabilização. (Mankiw; Samuelson, 2010; Giambiagi, 2005.; BAER, 2002)

### 3.1.2. *Expectativas futuras*

Os agentes econômicos coletam e publicam periodicamente dados sobre as expectativas do mercado para diversos indicadores econômicos e diferentes datas futuras. O spread de futuros reflete a diferença de preço entre contratos de futuros com diferentes datas de vencimento para um mesmo ativo subjacente. Essa diferença é influenciada pelas expectativas do mercado em relação aos preços futuros desse ativo.

Mediante a disponibilidade de dados para consulta, foram considerados diferentes spreads para cada tipo de dado. Para os dados fornecidos pelo boletim Focus foram considerados spreads de 1 ano e 4 anos em relação a expectativa do ano vigente. Para a taxa de juros foi proposto 1 spread curto entre 18m-3m e 2 spreads mais longos de 5a-1a e 10a-2a.

### 3.2. *Preparação dos Dados*

A preparação dos dados é uma etapa fundamental para garantir a qualidade e confiabilidade dos resultados obtidos nos modelos de previsão. O processo envolve várias etapas, que serão detalhadas a seguir.

### 3.2.1. Organização da Base de Dados

Inicialmente, os dados coletados são organizados em uma base de dados estruturada e consistente, facilitando a manipulação e a análise das informações. Nesta etapa, deve ser estabelecida uma nomenclatura adequada para as variáveis.

620 A análise exploratória dos dados envolve a verificação da distribuição das variáveis, a identificação de possíveis outliers e a análise de correlações entre as variáveis. Essa etapa é fundamental para a compreensão das características dos dados e a identificação de possíveis problemas que possam afetar os resultados dos modelos. O tratamento dos dados faltantes ou outliers associados a erros  
625 identificáveis, foram tratados através da interpolação exponencial.

$$r_i = \left( (1 + r_a)^{n_a} \times \left( \frac{(1 + r_p)^{n_p}}{(1 + r_a)^{n_a}} \right)^{\frac{n_i - n_a}{n_p - n_a}} \right)^{\frac{1}{n_i}} - 1$$

Durante a análise exploratória, foi necessário estabelecer o intervalo de dados e sua frequência temporal das séries. Percebeu-se que antes de 2005 os dados possuíam baixa disponibilidade e confiabilidade, portanto foi estabelecido uma frequência de dados mensal iniciando em janeiro de 2005.

### 630 3.2.2. Decomposição sazonal

Um dos objetivos deste trabalho é verificar o efeito da sazonalidade no problema estudado. Para isso, é proposto uma decomposição das componentes sazonais para as variáveis, de modo que cada componente será prevista separadamente e a previsão completa será reconstruída posteriormente, buscando uma melhor  
635 previsão do que as variáveis originais.

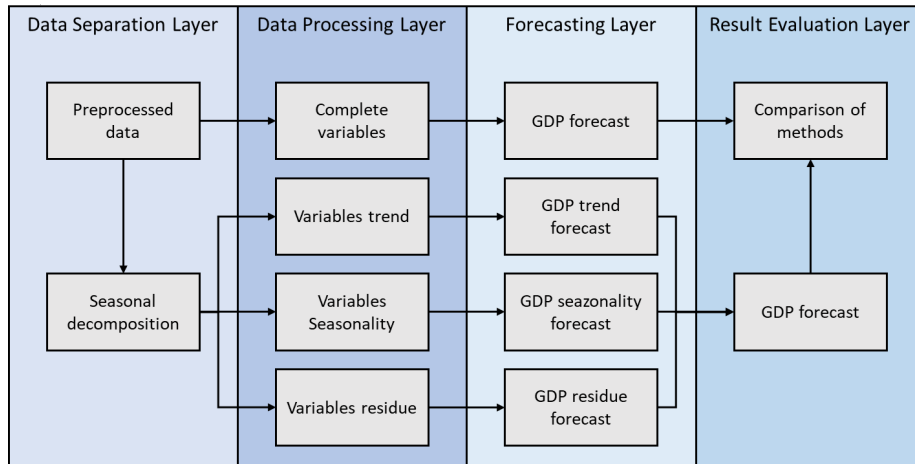


Figure 2: Seasonality decomposition forecast architecture

Neste trabalho foi utilizada a decomposição STL que é um método analítico para desagregar séries temporais em tendência, sazonalidade e resíduos.

$$y(t) = T_y(t) + S_y(t) + R_y(t)$$

O método utiliza a suavização local (Loess) em dois loops; o loop interno itera entre a suavização sazonal e de tendência e o loop externo minimiza o efeito de valores atípicos. Durante o loop interno, o componente sazonal é calculado primeiro e removido para calcular o componente de tendência. O restante é calculado subtraindo os componentes sazonais e de tendência da série de tempo. (Cleveland et al., 1990)

### 3.2.3. Estacionaridade e Autocorrelação

A estacionaridade em séries temporais refere-se à propriedade na qual as características estatísticas, como média e variância, permanecem constantes ao longo do tempo, facilitando a modelagem e análise devido a ausência de tendências sistemáticas ou padrões sazonais. (Hamilton, 1994)

Para avaliar a estacionaridade das séries foram utilizados os testes de raiz unitária como o KPSS (Kwiatkowski–Phillips–Schmidt–Shin) e o ADF (Augmented Dickey–Fuller) (Kwiatkowski et al., 1992), (Dickey and Fuller, 1979).

Além disso foram verificados os diagramas da função de autocorrelação simples (ACF) e parcial (PACF). Todas as séries avaliadas e suas componentes sazonais são estacionárias em até segunda diferença.

#### 655 3.2.4. Cointegração e Redução de Variáveis

Segundo Engle and Granger (1987) a análise de cointegração busca avaliar se há uma relação de longo prazo entre duas ou mais séries que têm tendências comuns, ou seja, mesmo que duas séries temporais se movam de maneira não estacionária, sua combinação linear pode ser estacionária. As primeiras tentativas de se identificar quais as variáveis estavam cointegradas baseou-se no teste de Engle-Granger, no entanto, foram verificados problema relacionados a componente de sazonalidade.

660 Devido a natureza multivariada do problema, foi necessário recorrer ao teste de Johansen, que é mais robusto em relação as hipóteses de homocedasticidade e normalidade dos erros, trazendo resultados significativamente melhores em todas as abordagens testadas.(Johansen, 1991)

#### 3.2.5. Padronização e Escalonamento dos Dados

Em alguns casos, pode ser necessário padronizar e escalonar os dados antes de aplicar os modelos de machine learning e VAR. Isso ocorre porque diferentes variáveis podem ter diferentes escalas e unidades de medida, o que pode afetar a precisão dos modelos. A padronização e o escalonamento consistem em ajustar a média e o desvio padrão das variáveis para que tenham valores comparáveis, acelerando a convergência dos métodos de previsão e minimizando o erro por overfitting.

$$X_i^* = \frac{X_i - \mu}{\sigma}$$

Após a conclusão dessas etapas, os dados estarão prontos para serem utilizados nos modelos de previsão de ciclos econômicos.

#### 670 3.3. Implementação dos Algoritmos de Machine Learning

Para a implementação dos algoritmos de machine learning selecionados, foi utilizada a linguagem de programação Python com as bibliotecas especializadas,

como scikit-learn e TensorFlow. O processo de implementação inclui várias etapas, descritas a seguir.

### 675 3.3.1. Implementação dos Algoritmos

Os algoritmos selecionados serão implementados utilizando as bibliotecas escolhidas (Pedregosa et al., 2011). O código será estruturado de forma clara e organizada, facilitando a compreensão e a manutenção futura.

680 *DT - Decision Tree.* é um método de aprendizado supervisionado não paramétrico usado para classificação e regressão, que utiliza diagramas para mapear as possíveis tomadas de decisão. O resultado de cada possível ação é determinado, ponderando a probabilidade associada e somando ao valor esperado de cada possível ação.

685 A partir de um vetor de rótulos, a árvore particiona de maneira recursiva o espaço de recursos de modo que as amostras com os mesmos rótulos ou valores alvo semelhantes sejam agrupadas.

$$Q_m^l(\theta) = \{(x, y) | x_j \leq t_m\}$$

$$Q_m^r(\theta) = Q_m \setminus Q_m^l(\theta)$$

A qualidade de uma divisão candidata de nó  $m$  é então calculado, buscando minimizar uma função de impureza ou função de perda  $H()$

$$G(Q_m, \theta) = \frac{n_m^l}{n_m} H(Q_m^l(\theta)) + \frac{n_m^r}{n_m} H(Q_m^r(\theta))$$

Dentre os critérios de regressão, destacam-se

- 690 • Erro quadrático médio

$$\bar{y}_m = \frac{1}{n_m} \sum_{y \in Q_m} y$$
$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} (y - \bar{y}_m)^2$$



- Desvio médio de Poisson

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} (y \log \frac{y}{\bar{y}_m} - y + \bar{y}_m)$$

*RF - Random Forest*:. trata-se de um método que se baseia em calcular a média de várias árvores de decisão profundas, treinadas em diferentes partes do mesmo conjunto de treinamento, com o objetivo de reduzir a variância. Isso ocorre às custas de um pequeno aumento no viés e alguma perda de interpretabilidade, mas geralmente aumenta muito o desempenho no modelo final.

- Critério de regressão

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2$$

- Critério de classificação

$$Gini = 1 - \sum_{i=1}^n (p_i)^2$$

*XGB - Extreme Gradient Boosting*:. O XGBoost, é um algoritmo baseado em árvores de decisão que utiliza a abordagem de impulsionamento de gradiente para construir um modelo robusto.

Durante o treinamento, o algoritmo otimiza uma função objetivo que incorpora termos de perda e termos de regularização.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$$

A construção do modelo envolve a adição sequencial de árvores, cada uma projetada para corrigir os erros do modelo anterior. Durante a construção de cada árvore, o algoritmo utiliza uma técnica de atualização gradiente para ajustar as previsões do modelo. Uma das grandes vantagens é que é possível media

705 a qualidade da árvore  $q$  através da função pontuação.

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

onde  $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$   
 $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$

*SVM - Support Vector Machine.* trata-se de um método que busca uma função, normalmente chamada de hiperplano, que consiga separar as classes de maneira a maximizar a distancia entre os pontos mais próximos. Pode ser utilizada para regressão, classificação e detecção de outliers.

$$\min_{w, b, \zeta, \zeta^*} \frac{1}{2} w^T w + C \sum_{i=1}^n (\zeta_i + \zeta_i^*)$$

$$\text{sujeito a } y_i - w^T \phi(x_i) - b \leq \varepsilon + \zeta_i,$$

$$w^T \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^*,$$

$$\zeta_i, \zeta_i^* \geq 0, i = 1, \dots, n$$

710 Aqui, estamos penalizando amostras cuja previsão é pelo menos  $\varepsilon$  longe de seu verdadeiro alvo. Essas amostras penalizam o objetivo por  $\zeta_i$  ou  $\zeta_i^*$ , dependendo se suas previsões estão acima ou abaixo do tubo.

Uma vez resolvido o problema de otimização, a saída de função-decisão para uma determinada amostra torna-se:

$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

*MLP - Multi-Layer Perceptron.* é um método que busca aproximar qualquer função contínua, bem como resolver problemas que não são linearmente separáveis. Se maneira suplementar às redes neurais feed forward, a MLP utiliza o algoritmo de backpropagation para minimizar o erro de treinamento de forma interativa.

$$Loss(\hat{y}, y, W) = \frac{1}{2n} \sum_{i=0}^n \|\hat{y}_i - y_i\|_2^2 + \frac{\alpha}{2n} \|W\|_2^2$$

A função de ativação bem como a quantidade de neuronios utilizada deve ser calibrada para as especificidades do problema, na etapa de ajuste dos hiper-  
 715 parâmetros. No entanto, é preciso tomar cuidado para não induzir o overfitting.

### 3.3.2. *Treinamento e Validação dos Modelos*

O treinamento dos métodos é realizado utilizando o subconjunto inicial do conjuntos de dados, através do conceito de validação cruzada. A validação cruzada consiste em dividir o conjunto de treinamento em várias partes, utilizando uma parte como conjunto de validação e o restante como conjunto de treinamento. O processo é repetido de forma iterativa, e os resultados são agregados para obter uma estimativa mais robusta do desempenho do modelo.

Como os dados utilizados são séries históricas, existe dependência temporal entre os eventos. Portanto, o processo de validação deve ocorrer necessariamente de forma progressiva. A figura ?? ilustra como o processo de validação cruzada é adaptado para garantir o encadeamento temporal dos dados. (Hyndman and Athanasopoulos, 2021)

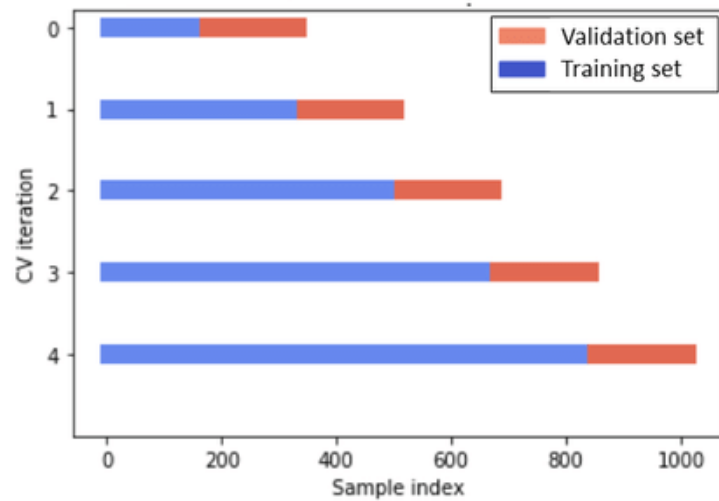


Figure 3: Cross-validation for time series

### 3.3.3. *Ajuste dos Hiperparâmetros*

O ajuste dos hiperparâmetros visa encontrar a combinação de parâmetros que otimize o desempenho do modelo, evitando o sobreajuste e melhorando a capacidade de generalização.

### *3.3.4. Análise dos Resultados e Conclusões*

Os resultados obtidos serão analisados e comparados, permitindo a identificação dos algoritmos mais eficazes para a previsão dos ciclos econômicos. A  
735 análise dos resultados também possibilitará a identificação de possíveis melhorias e ajustes nos modelos, visando aprimorar ainda mais o desempenho na previsão dos ciclos econômicos.

### *3.4. Implementação do Modelo VAR*

A implementação do modelo VAR envolve várias etapas, incluindo a seleção  
740 das variáveis, a determinação da ordem de integração e cointegração, o ajuste do modelo e a verificação de sua estabilidade e validade. Essas etapas são detalhadas a seguir:

#### *3.4.1. Seleção das Variáveis*

Inicialmente é necessário realizar a seleção das variáveis endógenas e exógenas,  
745 com base na teoria econômica e nas correlações observadas entre as variáveis durante a análise exploratória dos dados. A escolha das variáveis adequadas é crucial para garantir a qualidade das previsões do modelo VAR.

#### *3.4.2. Determinação da Ordem de Integração e Cointegração*

As séries temporais utilizadas no modelo VAR devem ser estacionárias. Para  
750 verificar a estacionariedade das séries, utilizou-se os testes de Dickey-Fuller Aumentado (ADF) e o teste KPSS. Em algumas situações, as séries devem ser diferenciadas até se tornarem estacionárias. A cointegração entre as variáveis será verificada por meio de testes como o teste de Johansen.

#### *3.4.3. Escolha da Ordem do Modelo VAR*

A ordem do modelo VAR, que indica o número de defasagens incluídas, será  
755 escolhida com base em critérios de informação, como o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). A ordem selecionada minimizará esses critérios e proporcionará o melhor equilíbrio entre a complexidade do modelo e o ajuste aos dados.

760 Nos modelos VARMAX e SARIMAX além das defasagens, existem outros  
parâmetros necessários para se definir como o número de diferenciações, a média  
móvel e periodicidade sazonal. Para a busca inicial dos parâmetros ótimos,  
foram utilizados os diagramas da função de autocorrelação ACF e PACF, mas  
o ajuste fino desses parâmetros se deu por tentativa e erro.

#### 765 3.4.4. Verificação da Estabilidade e Validade das Previsões

Após a estimação dos modelos VAR, foi verificada a estabilidade das pre-  
visões por meio da análise das raízes características do modelo. A validade  
das previsões será avaliada por meio da comparação entre as previsões geradas  
pelo modelo e os dados reais, utilizando a métrica Raiz Erro Quadrático Médio  
770 (RMSE).

### 3.5. Avaliação e Comparação de Desempenho

A avaliação e comparação do desempenho dos algoritmos de machine learning  
e do modelo VAR são etapas fundamentais para identificar qual abordagem é  
mais eficaz na previsão das fases do ciclo econômico no Brasil. Essa análise será  
775 realizada com base nas seguintes métricas e procedimentos:

#### 3.5.1. Métricas de Desempenho

Os modelos baseados em VAR podem utilizar métricas variadas para estimar  
de forma quantitativa a eficiência da estimativa em relação ao que foi observado.

780 Neste trabalho foram utilizados critérios de erro (MAE, MAPE, RMSE),  
critérios de correlação ( $R^2$ ) e critérios de informação (AIC e BIC). O critério  
adotado foi o RMSE, tendo em vista que nos melhores modelos ele concorda  
com os critérios de correlação e informação.

*Raiz do erro quadrático médio (RMSE)*. média dos quadrados dos erros entre  
as previsões e os valores reais;

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

*Coefficiente de Determinação  $R^2$ .* medida que indica a proporção da variância total explicada pelo modelo.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \Rightarrow R^2 = \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

*Critério de informação de Akaike (AIC).*

$$AIC = T \log \left( \frac{SSE}{T} \right) + 2(k + 2)$$

*Critério de informação de (BIC).*

$$BIC = T \log \left( \frac{SSE}{T} \right) + (k + 2) \log(T)$$

### 3.5.2. Comparação de Desempenho entre os Modelos

Com as métricas de desempenho calculadas para cada modelo, foi realizada  
785 uma análise comparativa entre os algoritmos de machine learning e os modelos econométricos. As análises comparativas foram realizadas tanto nas séries completas como nas suas componentes sazonais.

Essa comparação permite identificar qual abordagem é mais eficaz na previsão das fases do ciclo econômico no Brasil, levando em consideração aspectos  
790 como a capacidade de generalização, a precisão das previsões e a robustez dos modelos.

### 3.5.3. Análise dos Resultados e Conclusões

A análise dos resultados e das métricas de desempenho possibilitará a identificação de possíveis melhorias e ajustes nos modelos, visando aprimorar ainda  
795 mais o desempenho na previsão dos ciclos econômicos. Além disso, a comparação entre as diferentes abordagens permitirá extrair conclusões sobre os pontos fortes e fracos de cada método e orientar a escolha do melhor modelo para a previsão das fases do ciclo econômico no Brasil.

## 4. Resultados

800 Os resultados apresentados à seguir tem como principal objetivo demonstrar o potencial preditivo das técnicas escolhidas, bem como averiguar o efeito da decomposição sazonal na capacidade preditiva do modelo. Finalmente, buscase verificar se uma combinação de técnicas traz alguma vantagem, superando a melhor das técnicas apresentadas.

805 Dentre os métodos de machine learning, foram propostas duas redes neurais profundas implementadas de maneira bem diferente. O método chamado de "MLP" foi implementada utilizando como base as bibliotecas do scikit-learn, por ser uma ferramenta de simples implementação e otimização. O método chamado de "NN" foi implementado utilizando as bibliotecas do tensorflow, por  
810 ser uma ferramenta mais completa e robusta.

### 4.1. Combinação de variáveis explicativas

São propostos 3 cenários de análise, com o objetivo de avaliar o efeito das variáveis endógenas e exógenas. O método VARMAX já trabalha com a separação dos conjuntos endógenos e exógenos internamente, portanto só faz sentido compará-lo no cenário 3. O método SARIMAX já trata as componentes  
815 sazonais internamente, portanto não faz sentido aplicar a decomposição sazonal proposta em ???. O método dos mínimos quadrados (OLS) foi incluído como referência, dado sua simplicidade conceitual para interpretar os resultados.

#### 4.1.1. Variáveis endógenas

820 Analisando o cenário onde somente são consideradas as variáveis endógenas, verifica-se que o efeito da decomposição sazonal trouxe uma melhora significativa no desempenho das técnicas de machine learning e uma piora no modelo econométrico. É importante notar que mesmo com um desempenho melhor, somente o XGB conseguiu superar as técnicas econométricas originais, ou seja,  
825 a melhora de um desempenho muito ruim pode ser menos pior, mas não necessariamente é melhor.

Table 3: RMSE performance of methods using only endogenous variables

Method	Original	Decomposed	Diff	Best?	Trend	Seasonality	Residue
OLS	29.575,07	73.307,52	-147,87%	0	71.546,32	19.089,95	21.807,30
VAR	48.207,62	73.457,15	-52,38%	0	70.432,61	21.400,08	26.480,49
DT	76.429,96	59.114,59	22,66%	1	58.645,23	19.709,58	27.760,22
RF	79.290,91	64.367,25	18,82%	1	62.398,99	20.038,30	25.785,17
XGB	70.415,11	34.617,72	50,84%	1	28.650,65	19.764,91	25.690,05
MLP	131.465,75	49.491,59	62,35%	1	47.846,89	720.933,77	720.457,11
NN	173.154,65	42.569,40	75,42%	1	34.704,58	19.438,62	25.840,67
SVM	98.593,01	67.431,40	31,61%	1	53.097,06	24.745,70	35.551,33
Average	88.391,51	58.044,58	7,68%	0,75	53.415,29	108.140,11	113.671,54
Std	45.930,72	14.447,89	73,84%	0,46	15.679,23	247.612,74	245.209,12
Min	29.575,07	34.617,72	-147,87%	0,00	28.650,65	19.089,95	21.807,30
Max	173.154,65	73.457,15	75,42%	1,00	71.546,32	720.933,77	720.457,11

Olhando para as componentes sazonais, é possível perceber que os métodos que melhor performaram foram os que conseguiram capturar melhor as não linearidades da componente de tendência, uma vez que a sazonalidade e o resíduo são comparáveis entre quase todas as técnicas. Dessa forma, a abordagem se mostra válida quando se busca uma avaliação da direção do PIB em longo prazo.

#### 4.1.2. Variáveis exógenas

Considerando agora o cenário onde somente são consideradas as variáveis exógenas, verifica-se que a decomposição sazonal trouxe uma piora significativa no desempenho de todas as técnicas avaliadas. O SVM foi uma exceção mas que não deve ser considerada, já que a quebra estrutural causada pela pandemia de covid-19 acabou gerando overfitting que explica esses valores fora de escala.



Table 4: RMSE performance of methods using only exogenous variables

Method	Original	Decomposed	Diff	Best?	Trend	Seasonality	Residue
OLS	38.440,98	132.256,01	-244,05%	0	131.258,47	26.274,29	14.362,04
VAR	38.788,98	57.891,65	-49,25%	0	53.707,20	21.908,22	27.077,92
SARIMAX	38.438,02	-	-	-	-	-	-
DT	32.411,75	79.852,25	-146,37%	0	80.206,04	17.713,80	24.539,15
RF	57.377,21	70.458,83	-22,80%	0	69.359,37	17.609,88	24.539,82
XGB	37.759,84	55.269,06	-46,37%	0	52.392,69	17.818,30	24.611,16
MLP	47.651,21	72.169,42	-51,45%	0	76.938,51	716.371,10	722.240,54
NN	49.062,85	72.084,68	-46,92%	0	74.021,20	18.993,76	23.561,77
SVM	198.502,19	89.937,86	54,69%	1	87.792,51	17.458,07	28.895,09
Average	59.825,89	78.739,97	-69,06%	0,13	78.209,50	106.768,43	111.228,44
Std	52.555,84	24.296,31	89,25%	0,35	24.710,80	246.335,66	246.923,04
Min	32.411,75	55.269,06	-244,05%	0,00	52.392,69	17.458,07	14.362,04
Max	198.502,19	132.256,01	54,69%	1,00	131.258,47	716.371,10	722.240,54

Olhando para as componentes sazonais, é possível perceber que as técnicas de machine learning conseguiram capturar as não linearidades da componentes de sazonalidade melhor do que as técnicas econométricas. Como a amplitude 840 de sazonalidade é menor, essa diferença mostra-se significativa.

#### 4.1.3. Variáveis engógenas e endógenas

Considerando o cenário onde todas as variáveis explicativas são consideradas, podemos notar que a decomposição influenciou negativamente o desempenho dos métodos econométricos. Já em relação as técnicas de machine learning, não 845 foi possível chegar em um consenso, mas novamente o XGB foi o método que alcançou uma boa relação entre melhora e erro total. O método SARIMAX foi incluído nessa análise a título de curiosidade, mas não foi surpresa o desempenho ruim, já que foram incluídas variáveis endógenas em um método que só trabalha 850 com variáveis exógenas.

Table 5: RMSE performance of methods using all variables

Method	Original	Decomposed	Diff	Best?	Trend	Seasonality	Residue
OLS	21.023,35	51.939,36	-147,06%	0	45.791,37	27.260,42	15.000,20
VAR	35.517,25	55.019,62	-54,91%	0	50.593,57	15.557,31	26.980,56
VARMAX	22.107,80	124.073,85	-461,22%	0	127.429,71	23.874,37	18.170,79
SARIMAX	67.183,34	-	-	-	-	-	-
DT	45.901,91	58.268,85	-26,94%	0	56.516,56	18.465,30	24.733,51
RF	57.460,16	69.601,30	-21,13%	0	68.535,12	17.558,11	24.587,08
XGB	50.633,39	42.631,23	15,80%	1	38.234,64	20.503,85	24.998,07
MLP	97.455,64	75.885,58	22,13%	1	81.620,96	716.324,38	717.891,51
NN	34.661,39	102.143,22	-194,69%	0	100.615,60	23.860,45	21.315,33
SVM	111.122,56	84.033,25	24,38%	1	82.060,79	16.100,62	23.860,14
Average	54.306,68	73.732,92	-93,74%	0,33	72.377,59	97.722,76	99.726,35
Std	30.245,83	26.222,02	157,60%	0,50	28.767,98	232.009,25	231.842,64
Min	21.023,35	42.631,23	-461,22%	0,00	38.234,64	15.557,31	15.000,20
Max	111.122,56	124.073,85	24,38%	1,00	127.429,71	716.324,38	717.891,51

Olhando para as componentes sazonais, vemos que o XGB ao estimar a componente de tendência, consegue um bom desempenho mais próximo do cenário engógeno do que do exógeno, mostrando que o método é eficiente mas ainda possui limitações quanto ao peso das variáveis. Analogamente o SVM manteve o overfitting influenciado pelas variáveis exógenas na quebra estrutural referente a pandemia de covid-19. Finalmente, os métodos econométricos se mostraram eficientes para estimar o resíduo.

#### 4.1.4. Comparação das técnicas

Analisando inicialmente os modelos econométricos, percebemos que eles tem um desempenho similar entre si. Durante o ajuste fino dos métodos, os hiperparametros ótimos foram praticamente os mesmos em ambas as técnicas, de modo que as diferenças estruturais entre os métodos causaram pouca diferença

no desempenho. Quando analisamos a abordagem de decomposição sazonal, os resultados não são muito relevantes porque os métodos não foram desenhados para trabalhar dessa forma.

As técnicas de machine learning baseadas em árvore (DT, RF e XGB) performaram de maneira relativamente consistente em relação a outras técnicas. Embora o XGB tenha apresentado um desempenho superior as outras duas, sua complexidade operacional torna seus resultados pouco interpretáveis. Nesse sentido, estudar a solução das DT seria um primeiro passo para compreender o comportamento geral do fenômeno, e depois de compreendido o XGB seria capaz de extrair os melhores resultados.

Ambas as redes neurais propostas apresentaram resultados insatisfatórios, por diferentes motivos. O método MLP apresentou uma predição relativamente adequada mas com um delay, o que explica aqueles valores absurdos de erro nas componentes sazonais, mas que em certa medida se cancelavam na série reconstruída, mas mesmo assim o resultado geral é insatisfatório. O método NN por outro lado apresentava certa instabilidade nos resultados, tanto no que se refere a overfitting quanto a replicabilidade.

Os modelos baseados em SVM sofreram com overfitting, pois são muito sensíveis as quebras estruturais. Assim como as os modelos baseados em redes neurais, apesar de terem um bom potencial preditivo, em geral necessitam de projetos mais bem elaborados, específicos para uma certa demanda. A abordagem utilizada para tratamento das variáveis mostrou-se inadequada para essas classes de técnicas.

#### *4.2. Combinação de técnicas de predição*

A busca pelo modelo que melhor representa o fenômeno de interesse necessita de um critério objetivo, de modo que a medida RMSE foi adotada na busca de uma comparação quantitativa. Além disso, outras métricas como a correlação e os critérios de informação foram usando de forma complementar.

Os métodos econométricos quando subdimensionados, minimizavam o erro através de curvas simples que não representavam a complexidade do fenômeno,

induzindo a falsos positivos. Portanto, a análise gráfica foi fundamental para verificar a qualidade preditiva de cada técnica, bem como perceber os momentos em que as técnicas se perdiam, como a quebra estrutural em 2020.

#### 4.2.1. *Melhores modelos preditivos*

Considerando a abordagem tradicional onde as variáveis não são decompostas, os métodos econométricos apresentaram desempenho significativamente superior que os modelos de machine learning. Surpreendentemente o modelo que obteve melhor desempenho foi o OLS considerando apenas as variáveis exógenas.

Quando consideramos a abordagem utilizando decomposição sazonal, as técnicas de machine learning apresentaram vantagem, em especial para aquelas que são baseadas em árvores de decisão. Devido a sua boa qualidade preditiva da componente de tendência, o XGB considerando apenas as variáveis endógenas foi a técnica mais bem sucedida.

#### 4.2.2. *Modelo híbrido*

Olhando para as componentes sazonais separadamente, cada uma teve uma técnica que melhor se adequou na previsão. Portanto, buscou-se uma combinação das melhores soluções para tentar construir um modelo preditivo que superasse o melhor dos modelos alcançados até então.

A tendência foi melhor explicada pelo XGB endógeno, enquanto que a sazonalidade foi melhor explicada pelo VAR endógeno + exógeno, e finalmente os resíduos foram melhor explicados pelo OLS exógeno. Dessa forma, é proposto um modelo que combina essas 3 abordagens em um preditor híbrido.

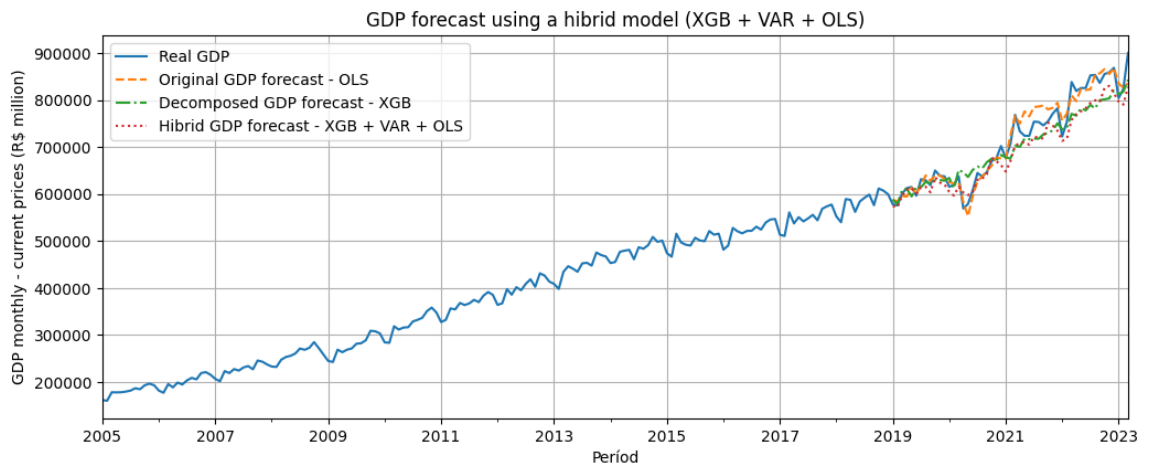


Figure 4: Comparison of GDP forecasting models – Original vs Decomposed vs Hybrid

915 Podemos observar no gráfico 4 que, de modo geral, a capacidade de predição do modelo foi adequada. O gráfico 5 restringe o período amostral para permitir uma observação maior dos detalhes.

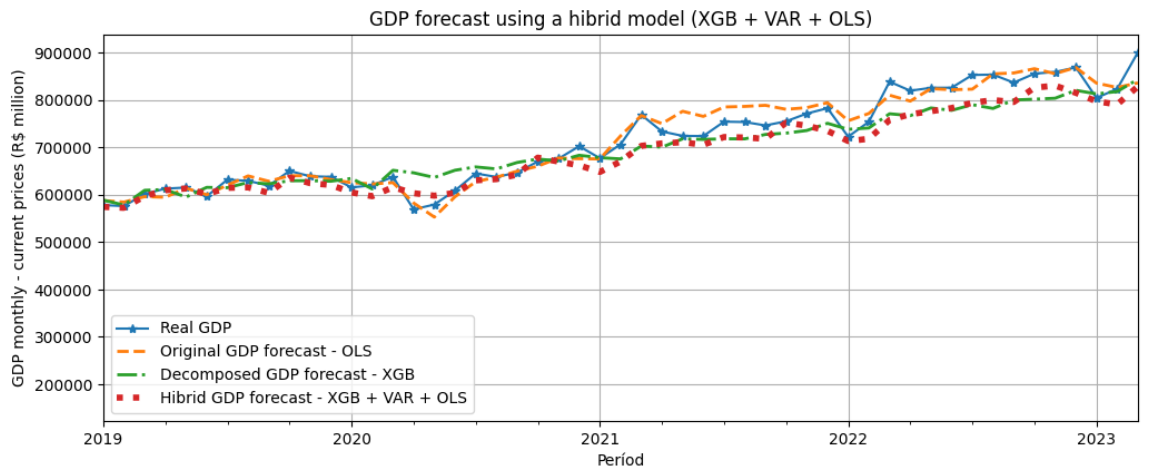


Figure 5: Comparison of models between 2019 and 2023

É possível observar qualitativamente que o modelo híbrido consegue melhorar diversas aspectos em relação ao modelo decomposto que utiliza somente o XGB. Podemos observar o primeiro trimestre de 2020 onde o modelo híbrido

920

consegue capturar parcialmente a quebra estrutural, ou os picos de transição de 2022 e 2023 onde o modelo híbrido se antecipa numa amplitude mais adequada. Apesar de não ter superado a qualidade do modelo OLS sem decomposição, considera-se que a abordagem foi bem sucedida.

925 Outro modelo que chamou a atenção foi um híbrido parecido, com XGB endógeno na tendência e OLS exógeno no resíduo, mas a sazonalidade foi estimada pelo SVM endógeno + exógeno. Se observarmos a tabela 5, notamos que ele representa o segundo melhor desempenho, com um erro muito parecido com o modelo VAR.

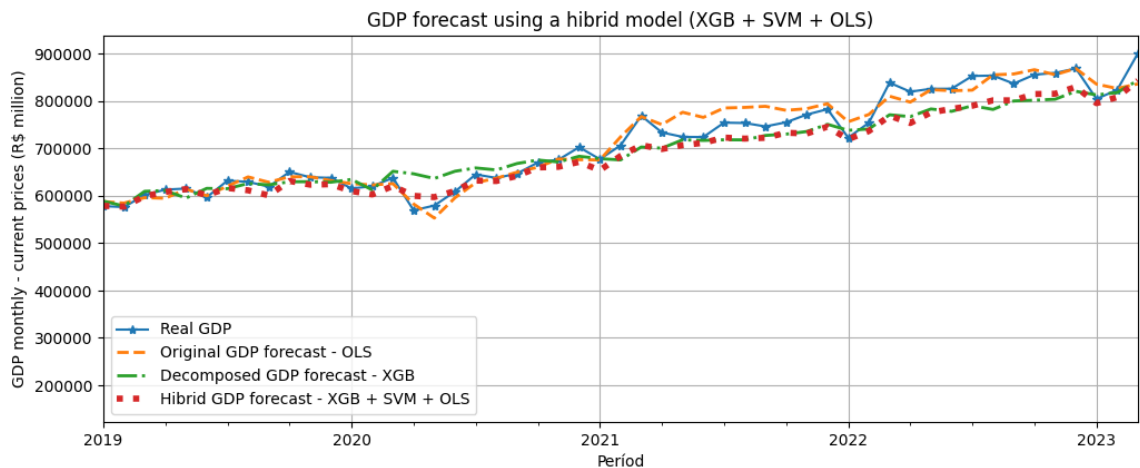


Figure 6: Híbrido model using SVC between 2019 and 2023

930 Observando o gráfico 6, é possível notar que as diferenças são bastante sutis em relação ao híbrido utilizando VAR. Em regiões críticas de picos ele desempenha ligeiramente melhor, alcançando a amplitude de forma mais adequada que o modelo anterior. Mas nas regiões de tendência ele apresenta um offset quase constante, mostrando ineficiência em capturar pequenas oscilações.

#### 935 4.2.3. Comparação de modelos

Quando comparamos os modelos que alcançaram melhor desempenho, tanto de forma qualitativa quanto quantitativa, é inegável que o OLS foi o grande

vencedor. Mas ainda assim, é relevante discutir o que foi observado nos modelos baseados na decomposição sazonal.

Table 6: RMSE performance of best models

Method	RMSE
OLS original: edg + exg	21.023,35
XGB decomposed: edg	34.617,72
Híbrid: XGB + VAR + OLS	32.319,26
Híbrid: XGB + SVM + OLS	30.814,75

940 Como esperado, os modelos híbridos superaram o melhor modelo decomposto. Apesar do modelo híbrido utilizando SVM ter superado quantitativamente o híbrido que utiliza VAR, acredito que uma análise qualitativa torna-se fundamental nesse caso.

945 O método VAR mostrou-se muito mais estável e robusto que o método SVM nos contextos em que se utilizavam somente variáveis endógenas e somente variáveis exógenas. Apesar do bom resultado no modelo completo, torna-se necessário investigar com mais cuidado o método SVM para confirmar se esse desempenho é replicável em outros cenários.

## 5. Conclusão

950 O presente trabalho propoe uma nova abordagem para estimar o PIB brasileiro, combinando técnicas de aprendizagem de máquina com técnicas econométricas. A preparação e tratamento dos dados é feito baseado nas condições exigidas pelo método VAR, trabalhando com as séries estacionárias. Verificou-se que os algoritmos de machine learning não lidam tão bem com essa abordagem, 955 provavelmente porque o processo de diferenciação remove informação sensível que dificulta a modelagem de não-linearidades.

A economia brasileira tem sido cada vez mais dependente do ciclo de commodities, portanto, buscou-se uma abordagem onde se pudesse tratar com mais

cuidado das componentes sazonais. Foi proposta uma metodologia que de-  
960 compõe e estima cada componente sazonal separadamente, buscando otimizar a  
modelagem. Verificou-se que as técnicas de machine learning tiveram uma mel-  
hora de performance utilizando essa abordagem, mas os métodos econométricos  
tiveram uma piora expressiva em todos os cenários.

Outro aspecto a se analisar foi o efeito das variáveis endógenas e exógenas nas  
965 diferentes componentes sazonais. Podemos verificar que as variáveis endógenas  
influenciam significativamente nas componentes de tendência, enquanto que as  
variáveis exógenas possuem mais influência na componentes residual. Por fim, a  
sazonalidade possui influência significativa tanto das variáveis endógenas quanto  
exógenas.

970 Por fim, foi proposto um modelo híbrido que combina diferentes técnicas  
para cada componente sazonal. Verificou-se que essa abordagem trouxe mel-  
horia de performance em relação as decomposições baseadas em uma única tec-  
nica, tanto do ponto de vista quantitativo como qualitativo. Apesar das técnicas  
econométricas clássicas sem decomposição terem demonstrado desempenho sig-  
975 nificativamente superior, a abordagem se mostra válida no sentido de investigar  
aspectos intrínsecos de cada componente sazonal, que pode ser de valia para  
compreender o complexo fenômeno como um todo.

A integração das diferentes técnicas de previsão também pode oferecer uma  
visão mais abrangente das interações entre as variáveis econômicas, permitindo  
980 aos tomadores de decisão compreender melhor os mecanismos subjacentes aos  
ciclos econômicos e, assim, implementar políticas e estratégias mais efetivas.  
Além disso, vale ressaltar que a seleção de variáveis, a qualidade dos dados e a  
escolha dos hiperparâmetros podem ter um impacto significativo no desempenho  
dos modelos. Portanto, a realização de uma análise de sensibilidade e a busca  
985 contínua pela otimização dos modelos são cruciais para garantir a robustez das  
previsões e melhorar a eficácia das técnicas de previsão.

Em suma, os resultados desta pesquisa mostram que a combinação de técnicas  
de aprendizado de máquina e modelos econométricos, podem proporcionar pre-  
visões mais precisas e insights mais profundos sobre os ciclos econômicos, auxili-



990 iando na tomada de decisões informadas e eficientes por formuladores de políticas,  
investidores e empresas.

### *Trabalhos futuros*

Os resultados obtidos com os métodos econométricos foram superiores em  
relação as técnicas de machine learning, provavelmente pelo fato do tratamento  
995 feito nas variáveis. Repetir a análise de machine learning utilizando as séries  
temporais originais (em nível) pode melhorar significativamente o desempenho  
desses métodos. Além disso, cada técnica possui uma metodologia própria para  
a seleção de variáveis, o que deve reduzir o overfitting em técnicas como a SVM  
e as redes neurais. Essa abordagem mais simplificada pode trazer vantagens  
1000 sensíveis à abordagem de decomposição sazonal.

Com relação a otimização dos hiperparâmetros, cada método foi avaliado  
em relação aos seus parâmetros principais. A grande quantidade de cenários  
avaliados, o tempo de simulação e a diversidade de parâmetros para algumas  
técnicas tornou inviável uma análise exploratória profunda. Para trabalhos  
1005 futuros, sugere-se um projeto mais detalhado para cada técnica, em especial as  
redes neurais e a SVM.

A respeito dos modelos híbridos, outras combinações podem ser testadas.  
Muitas vezes a combinação linear de soluções individualmente piores pode levar  
a uma solução melhor, como discutido no caso do SVM. Desenvolver uma técnica  
1010 que consiga combinar de forma automática todas as inúmeras combinações pos-  
sibilita um olhar diferente para o problema.

Por fim, a inclusão de outras variáveis e outras técnicas pode ser utilizadas  
tanto na busca por melhorar a estimativa do PIB brasileiro, como avaliar a  
estimativa de outras variáveis econômicas, bem como, outras economias.

### 1015 **References**

Alesina, A., Sachs, J., 1988. Political parties and the business cycle in  
the united states, 1948-1984. *Journal of Money, Credit and Banking* 20,

63–82. URL: <https://EconPapers.repec.org/RePEc:mcb:jmoncb:v:20:y:1988:i:1:p:63-82>.

1020 Alpaydin, E., 2010. Introduction to Machine Learning. Adaptive computation and machine learning, MIT Press. URL: <https://books.google.com.br/books?id=4j9GAQAIAAJ>.

Araújo, R.A., Oliveira, A.L., Meira, S., 2015. A hybrid model for high-frequency stock market forecasting. *Expert Systems with Applications* 42, 4081–  
1025 4096. URL: <https://www.sciencedirect.com/science/article/pii/S0957417415000093>, doi:<https://doi.org/10.1016/j.eswa.2015.01.004>.

BAER, W., 2002. A economia brasileira. 2. ed. rev. atual. e ampl. ed., Nobel, São Paulo :.

Banbura, M., Giannone, D., Reichlin, L., 2010. Large bayesian vector auto  
1030 regressions. *Journal of Applied Econometrics* 25, 71–92. URL: <https://EconPapers.repec.org/RePEc:jae:japmet:v:25:y:2010:i:1:p:71-92>.

Becker, G.S., 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70, 9–49. URL: <http://www.jstor.org/stable/1829103>.

1035 Burns, A.F., Mitchell, W.C., 1946. Measuring Business Cycles. Number burn46-1 in NBER Books, National Bureau of Economic Research, Inc. URL: <https://ideas.repec.org/b/nbr/nberbk/burn46-1.html>.

Choi, H., Varian, H., 2012. Predicting the present with  
google trends. *Economic Record* 88, 2–9. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4932.2012.00809.x>,  
1040 doi:<https://doi.org/10.1111/j.1475-4932.2012.00809.x>,  
arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-4932.2012.00809.x>.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. Stl:  
A seasonal-trend decomposition procedure based on loess (with discussion).  
1045 *Journal of Official Statistics* 6, 3–73.

- 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, 427–431. URL: <http://www.jstor.org/stable/2286348>.
- Drazen, A., 2000. *Political Economy in Macroeconomics*. Princeton University Press. URL: <http://www.jstor.org/stable/j.ctv346rwp>.  
1050
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55, 251–276. URL: <https://ideas.repec.org/a/ecm/emetrp/v55y1987i2p251-76.html>.
- Enke, D., Thawornwong, S., 2005. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications* 29, 927–940. doi:10.1016/j.eswa.2005.06.024.  
1055
- Estrella, A., Mishkin, F., 1996. The yield curve as a predictor of u.s. recessions. *Current Issues in Economics and Finance* 2, 7. URL: <https://EconPapers.repec.org/RePEc:fip:fednci:y:1996:i:jun:n:v.2no.7>.
- Estrella, A., Mishkin, F.S., 1998. Predicting u.s. recessions: Financial variables as leading indicators. *The Review of Economics and Statistics* 80, 45–61. URL: <http://www.jstor.org/stable/2646728>.  
1060
- Estrella, A., Trubin, M.R., 2006. The yield curve as a leading indicator: some practical issues. *Current Issues in Economics and Finance* 12. URL: <https://ideas.repec.org/a/fip/fednci/y2006ijulnv.12no.5.html>.  
1065
- Fatas, A., Mihov, I., 2003. The case for restricting fiscal policy discretion. *The Quarterly Journal of Economics* 118, 1419–1447. URL: <https://EconPapers.repec.org/RePEc:oup:qjecon:v:118:y:2003:i:4:p:1419-1447>.
- Filardo, A.J., Gordon, S.F., 1998. Business cycle durations. *Journal of Econometrics* 85, 99–123. URL: <https://www.sciencedirect.com/science/article/pii/S0304407697000961>, doi:[https://doi.org/10.1016/S0304-4076\(97\)00096-1](https://doi.org/10.1016/S0304-4076(97)00096-1).  
1070

- Fisher, I., 1933. The debt-deflation theory of great depressions. *Econometrica* 1, 337–357. URL: <http://www.jstor.org/stable/1907327>.  
1075
- Giambiagi, F., 2005. *Economia brasileira contemporânea* : Elsevier,, Rio de Janeiro :.
- Gogas, P., Papadimitriou, T., Chrysanthidou, E., 2015. Yield curve point triplets in recession forecasting. *International Finance* 18. doi:10.1111/infi.12067.  
1080
- Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. URL: <http://www.jstor.org/stable/1912791>.
- Güreşen, E., Kayakutlu, G., Daim, T., 2011. Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.* 38, 10389–10397. doi:10.1016/j.eswa.2011.02.068.  
1085
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press. URL: [https://www.worldcat.org/title/time-series-analysis/oclc/1194970663&referer=brief\\_results](https://www.worldcat.org/title/time-series-analysis/oclc/1194970663&referer=brief_results).
- Hansen, S., McMahon, M., Tong, M., 2019. The long-run information effect of central bank communication. *Journal of Monetary Economics* 108, 185–202. URL: <https://www.sciencedirect.com/science/article/pii/S0304393219301606>, doi:<https://doi.org/10.1016/j.jmoneco.2019.09.002>. “Central Bank Communications:From Mystery to Transparency”May  
1090 23-24, 2019Annual Research Conference ofthe National Bank of UkraineOrganized in cooperation withNarodowy Bank Polski.
- Hayek, F., 1931. *Prices and Production*. Studies in economics and political science ... no. 107 in the series of monographs by writers connected with the London School of Economics and Political Science, G. Routledge. URL:  
1100 <https://books.google.com.br/books?id=K5PUAAAAMAAJ>.

- Hibbs, D.A., 1977. Political parties and macroeconomic policy. *The American Political Science Review* 71, 1467–1487. URL: <http://www.jstor.org/stable/1961490>.
- Hyndman, R., Athanasopoulos, G., 2021. *Forecasting: Principles and Practice*. 3rd ed., OTexts, Australia. 1105
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59, 1551–80. URL: <https://EconPapers.repec.org/RePEc:ecm:emetrp:v:59:y:1991:i:6:p:1551-80>.
- 1110 Kalecki, M., 1935. A macrodynamic theory of business cycles. *Econometrica* 3, 327–344. URL: <http://www.jstor.org/stable/1905325>.
- Keynes, J.M., 1936. *The General Theory of Employment, Interest and Money*. Macmillan. 14th edition, 1973.
- Kondratiev, N.D., 1925. The major economic cycles. *Voprosy Konjunktury* 1, 1115 28–79.
- Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178. URL: <https://www.sciencedirect.com/science/article/pii/030440769290104Y>, doi:[https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y). 1120
- Kydland, F.E., Prescott, E.C., 1982. Time to build and aggregate fluctuations. *Econometrica* 50, 1345–1370. URL: <http://www.jstor.org/stable/1913386>.
- 1125 Mankiw, N.G., . *Macroeconomics*. 8th ed ed., Worth.
- Minsky, H.P., 1975. *John Maynard Keynes / by Hyman P. Minsky*. Columbia University Press New York.

- Minsky, H.P., 1992. The Financial Instability Hypothesis. Economics Working Paper Archive wp 74. Levy Economics Institute. URL: [https://ideas.repec.org/p/lev/wrkpap/wp\\_74.html](https://ideas.repec.org/p/lev/wrkpap/wp_74.html).  
1130
- Muth, J.F., 1961. Rational expectations and the theory of price movements. *Econometrica* 29, 315–335. URL: <http://www.jstor.org/stable/1909635>.
- Nordhaus, W.D., 1975. The political business cycle. *The Review of Economic Studies* 42, 169–190. URL: <http://www.jstor.org/stable/2296528>.
- 1135 Okun, A., 1962. M. 1962, “potential gnp: Its measurement and significance”, in: *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 89–104.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.  
1140
- Petris, G., Petrone, S., Campagnoli, P., 2009. *Dynamic Linear Models with R. useR!*, Springer-Verlag, New York.
- 1145 Phillips, A.W., 1958. The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957. *Economica* 25, 283–299. URL: <https://ideas.repec.org/a/bla/econom/v25y1958i100p283-299.html>, doi:10.1111/j.1468-0335.1958.
- Samuelson, P., 2010. *Economics*. Tata McGraw Hill. URL: <https://books.google.com.br/books?id=gzqXdHXxeAC>.  
1150
- Samuelson, P.A., 1964. Theoretical notes on trade problems. *The Review of Economics and Statistics* 46, 145–154. URL: <http://www.jstor.org/stable/1928178>.

- Schumpeter, J., 1934. The Theory of Economic Development: An Inquiry Into  
1155 Profits, Capital, Credit, Interest, and the Business Cycle. Harvard University. Dept. of Economics. Economic studies, Harvard University Press. URL:  
<https://books.google.com.br/books?id=rf2ZAAAAIAAJ>.
- Sims, C.A., 1980. Macroeconomics and reality. *Econometrica* 48, 1–48. URL:  
<http://www.jstor.org/stable/1912017>.
- 1160 Soybilgen, B., 2019. Identifying turkish business cycle regimes in real time.  
*Applied Economics Letters* 27, 1–5. doi:10.1080/13504851.2019.1607243.
- Stigler, G.J., 1950. The development of utility theory. i. *Journal of Political Economy* 58, 307–327. URL: <http://www.jstor.org/stable/1828885>.