Ensemble de Modelos de Fundação para Previsão da Inflação no Brasil

Previsão dos 377 subitens do IPCA, agregação com os pesos do IBGE e comparação com o Focus

Vítor Wilher

Análise Macro · 10 de novembro de 2026

Versão 2.0 — resultados exploratórios

Resumo. O IPCA é um índice agregado de 377 subitens. Este trabalho testa se nove modelos de fundação de séries temporais, usados sem treino, conseguem prever esses subitens bem o bastante para que a soma ponderada compita com o Focus, e conta o percurso da pesquisa decisão a decisão, à luz da literatura. Os modelos preveem os 377 subitens; as previsões são agregadas com os pesos do IBGE, ajustadas para combustíveis (ANP) e energia (bandeira e créditos de Itaipu), misturadas à média sazonal e combinadas num ensemble ponderado pelo inverso do RMSE passado. Num walk-forward com informação datada, o ensemble tem erro igual ao do Focus no mês seguinte (razão de RMSE 1,009) e de 0,911 a 0,998 do erro do Focus de 2 a 12 meses à frente, sem diferença significativa. Intervalos gaussianos construídos com os erros já conhecidos em cada data ficam perto da cobertura nominal, mas a correção de viés em tempo real piora a previsão. Com dados até set/2026, o ensemble projeta o IPCA em 12 meses em 5,14% em dez/2026 e 4,55% em set/2027, acima do teto da meta.

Palavras-chave: inflação; modelos de fundação; previsão desagregada; Focus; IPCA.

Índice

1 Introdução **3**

2 O IPCA: um índice agregado **3**
2.1 Breve história . . . 3
2.2 Como o índice é construído . . . 4
2.3 Por que prever pelos componentes . . . 5

3 Modelos de fundação para séries temporais **5**
3.1 Das redes recorrentes à atenção . . . 5
3.2 O que é um modelo de fundação . . . 6
3.3 A chegada às séries temporais . . . 6
3.4 Previsão probabilística e cuidados . . . 7

4 Os nove modelos usados **7**

5 Dados e desenho **8**

6 O percurso da pesquisa **9**
6.1 Primeiro resultado: os modelos puros . . . 9
6.2 Desagregar compensa . . . 9
6.3 Onde estava o erro . . . 9
6.4 Informação institucional: combustíveis e energia . . . 9
6.5 Combinar modelos . . . 10
6.6 Auditar as datas . . . 10
6.7 Medir a incerteza . . . 10
6.8 Da pesquisa à produção . . . 10

7 Resultados **11**
7.1 IPCA do mês . . . 11
7.2 Onde o erro sobra . . . 11
7.3 Incerteza . . . 13

8 Previsão atual **14**

9 Limitações **15**

10 Conclusão **15**

Referências **16**

1 Introdução

O IPCA é uma média ponderada de centenas de variações de preço. Prevê-lo pelos componentes é a ideia natural para quem conhece o índice por dentro, mas sempre esbarrou num custo prático: modelar centenas de séries curtas, heterogêneas e cheias de quebras. A literatura brasileira registra ganhos modestos da desagregação. No estudo do Banco Central com 50 métodos de aprendizado de máquina e combinações, o caminho por quatro componentes foi "um pouco decepcionante" e quase não bateu um ARMA (Araujo e Gaglianone 2022, 32). Os modelos de fundação de séries temporais — redes pré-treinadas em grandes coleções de séries e aplicadas sem estimação, *zero-shot* — mudam esse custo: prever os 377 subitens passa a ser uma única chamada por data (Ansari et al. 2024; Das et al. 2024).

A pergunta deste trabalho é simples: previsões *zero-shot* dos 377 subitens, agregadas com os pesos do IBGE, competem com a mediana do boletim Focus em horizontes de 1 a 12 meses, num desenho em que cada previsão usa apenas a informação disponível na data em que teria sido feita? A resposta curta é que sim no mês seguinte, com empate estatístico, e com erro até 9% menor de 2 a 12 meses, sem significância.

O texto tem uma particularidade: ele conta o percurso da pesquisa decisão a decisão, como ela de fato aconteceu. Cada passo — desagregar, olhar onde o erro estava, trazer informação institucional para combustíveis e energia, combinar modelos, auditar datas, medir a incerteza e, por fim, levar a previsão para produção — respondeu a um resultado do passo anterior e é comparado com o que a literatura recomendaria. Algumas decisões foram revertidas quando o teste honesto as contrariou; elas ficam registradas, porque mostram por que o protocolo importa mais que o modelo. A seção 2 conta a história e a construção do IPCA; a seção 3, a dos modelos de fundação; a seção 4 descreve os nove modelos usados; a seção 5, os dados e o desenho; a seção 6, o percurso; a seção 7 consolida os resultados; e a seção 8 apresenta a previsão atual.

2 O IPCA: um índice agregado

2.1 Breve história

Entre 1948 e 1978, o índice de preços oficial era calculado pelo Ministério do Trabalho, em 14 capitais. Em julho de 1978 o IBGE assumiu a tarefa, e a reformulação deu origem ao Sistema Nacional de Índices de Preços ao Consumidor (SNIPC), criado em março de 1979 (Instituto Brasileiro de Geografia e Estatística 2020, 17). No mesmo ano nasceu o INPC, voltado à correção de salários, e, em dezembro de 1979, o IPCA, "com a finalidade de prover uma medida da inflação brasileira [...] referentes ao consumo pessoal, independentemente da fonte do rendimento familiar" (Instituto Brasileiro de Geografia e Estatística 2020, 17; 2013). Entre novembro de 1985 e março de 1986 o IPCA chegou a ser o único indexador oficial do país (Instituto Brasileiro de Geografia e Estatística 2020, 19).

O índice mede a inflação das famílias residentes em áreas urbanas com rendimento de 1 a 40 salários mínimos, de qualquer fonte — um corte desenhado para cobrir cerca de 90% das famílias (Instituto Brasileiro de Geografia e Estatística 2020, pp. 25–37). A coleta cobre hoje 16 áreas: dez regiões metropolitanas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba, Vitória e Porto Alegre) e os municípios de Goiânia, Campo Grande, Rio Branco, São Luís, Aracaju e Brasília (Instituto Brasileiro de Geografia e Estatística 2020, 15). Os preços são

Tabela 1: Grupos do IPCA, pesos e número de subitens.

Grupo	Peso (%)	Subitens
Alimentação e bebidas	21,51	168
Transportes	20,16	28
Habitação	15,26	32
Saúde e cuidados pessoais	13,84	36
Despesas pessoais	10,41	24
Educação	6,21	19
Vestuário	4,62	31
Comunicação	4,51	8
Artigos de residência	3,48	31

Pesos do IPCA em set/2026 (SIDRA/IBGE, tabela 7060, variável 66). Estrutura da POF 2017-18, em vigor desde jan/2020. Contagem de subitens pelo código do grupo.

coletados ao longo do mês de referência e o resultado é divulgado nos primeiros dias do mês seguinte (Instituto Brasileiro de Geografia e Estatística 2020, pp. 25, 107).

Desde 1999 o IPCA é o índice do regime de metas para a inflação, instituído pelo Decreto 3.088 (Brasil 1999; Instituto Brasileiro de Geografia e Estatística 2020, 18). Em 2025 o regime passou a avaliar a meta de forma contínua: a inflação acumulada em 12 meses é apurada todo mês, e há descumprimento quando ela fica fora do intervalo por seis meses consecutivos (Brasil 2024). A meta é de 3,00%, com tolerância de 1,50 ponto percentual (Conselho Monetário Nacional 2024).

As cestas e os pesos vêm das pesquisas de orçamentos familiares. Foram seis: ENDEF 1974-75 e POFs de 1987-88, 1995-96, 2002-03, 2008-09 e 2017-18, esta com cerca de 75 mil domicílios (Instituto Brasileiro de Geografia e Estatística 2020, 21). A estrutura da POF 2017-18 está em vigor desde janeiro de 2020 (Instituto Brasileiro de Geografia e Estatística 2020, 29), e é por isso que a série desagregada usada aqui começa nessa data.

2.2 Como o índice é construído

O IPCA tem quatro níveis abaixo do índice geral: grupo, subgrupo, item e subitem; o subitem é o nível mais desagregado com peso explícito vindo da POF (Instituto Brasileiro de Geografia e Estatística 2020, pp. 19–20). Nos dados usados (SIDRA, tabela 7060), a estrutura tem 9 grupos, 19 subgrupos, 51 itens e 377 subitens, os mesmos nos 81 meses de jan/2020 a set/2026 (Instituto Brasileiro de Geografia e Estatística 2026). A Tabela 1 mostra os pesos dos grupos.

O cálculo tem duas etapas. Na primeira, para cada subitem i e área, o índice elementar combina os preços dos produtos: a razão entre preços médios (Dutot) para cada produto e a média geométrica desses relativos (Jevons) para o subitem (Instituto Brasileiro de Geografia e Estatística 2020, pp. 68–69). Na segunda, os índices são agregados por média aritmética ponderada, numa fórmula "do tipo Lowe" na terminologia do IBGE, ou de Laspeyres na edição anterior (Instituto Brasileiro de Geografia e Estatística 2020, 70; 2013):

$$I_t^{\mathrm{ag}} = \frac{\sum_{\alpha} w_{\alpha}^{t-1} \, I_{\alpha,t}}{\sum_{\alpha} w_{\alpha}^{t-1}},$$

em que α percorre os componentes do agregado. Com os pesos em porcentagem, a variação mensal do índice geral é a soma das contribuições dos subitens:

$$\pi_t = \frac{1}{100} \sum_{i=1}^{377} w_{i,t}\, \pi_{i,t}. \tag{1}$$

Os pesos não são fixos: a cada mês são corrigidos pela variação do preço do subitem relativa à da sua área, de modo que o peso usado em t já incorpora os preços até $t-1$ (Instituto Brasileiro de Geografia e Estatística 2020, 70). No nível nacional, isso equivale aproximadamente a

$$w_{i,t+1} \;\propto\; w_{i,t} \left(1 + \frac{\pi_{i,t}}{100}\right). \tag{2}$$

A aproximação desconsidera que o IBGE normaliza dentro de cada área; a diferença é de segunda ordem. Nos dados, a Equação 1 com os pesos publicados reproduz o IPCA do mês com erro máximo de 0,005 ponto percentual, e a Equação 2 reproduz os pesos publicados dentro do arredondamento — usar o peso de $t-1$ no lugar do de t eleva o erro a 0,14 ponto. Duas consequências guiam o desenho: o peso de t é conhecido em $t-1$, e para horizontes maiores os pesos podem ser projetados com as próprias previsões, sem nunca usar peso futuro.

O índice é concentrado: os 10 maiores subitens somam 33,1% do peso, os 50 maiores, 67,6%, e 163 subitens têm peso abaixo de 0,05%. A gasolina pesa 5,13% e a energia elétrica residencial, 3,98%. Essa concentração é a razão por que errar poucos preços grandes custa mais que errar muitos pequenos — um ponto que voltará na seção 6.

2.3 Por que prever pelos componentes

Prever pelos componentes tem duas vantagens teóricas: cada série tem dinâmica própria (sazonalidade de alimentos, reajustes anuais de serviços, decisões administradas), e o agregado herda a soma ponderada dessas dinâmicas, que um modelo do índice cheio só vê misturadas. O custo é estatístico: são centenas de modelos, cada um com erro de estimação. A literatura de previsão em grupos formaliza esse dilema: um modelo *global*, que aprende um único conjunto de parâmetros para todas as séries, troca um pouco de viés por muita redução de variância, como um painel ou um BVAR (Montero-Manso e Hyndman 2021). O modelo de fundação é o caso extremo do modelo global — aprendido em milhões de séries de fora do problema — e é por isso que ele pode tornar a desagregação viável onde modelos estimados série a série falham.

3 Modelos de fundação para séries temporais

3.1 Das redes recorrentes à atenção

Uma rede recorrente recebe a própria saída anterior e usa os mesmos pesos em todos os passos; desdobrada no tempo, vira uma rede profunda com pesos compartilhados, treinada por retropropagação no tempo (Rumelhart et al. 1986; Werbos 1990). O gradiente, porém, é um produto de jacobianos que decai exponencialmente, o que impede aprender dependências longas (Bengio et al. 1994; Pascanu et al. 2013). A LSTM resolveu isso separando um estado de célula com atualização

aditiva, controlada por portas aprendidas (Hochreiter e Schmidhuber 1997), e a GRU fez o mesmo com menos parâmetros (Cho et al. 2014). Restou um limite: o processamento é sequencial.

O Transformer eliminou a recorrência e a substituiu pela atenção: cada passo consulta diretamente qualquer outro (Vaswani et al. 2017),

$$\text{Atenção}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}.$$

Três consequências explicam o salto: qualquer par de passos está a uma operação de distância, tudo vira produto de matrizes paralelizável, e o custo cresce com o quadrado do comprimento da sequência — o que motivou, em séries temporais, agrupar os pontos em blocos (*patches*) (Nie et al. 2023).

3.2 O que é um modelo de fundação

O termo foi cunhado para modelos treinados em dados amplos, em geral com autossupervisão em grande escala, e adaptáveis a muitas tarefas (Bommasani et al. 2021). Os autores alertaram para o outro lado da homogeneização: os defeitos do modelo base passam a todos os derivados. Em séries temporais, a autossupervisão é natural — o "rótulo" é o próprio futuro da série — e a mudança de paradigma é deixar de estimar um ARIMA ou uma LSTM para cada série e passar a consultar um modelo pré-treinado, que funciona porque tendências, sazonalidades e choques se repetem entre domínios.

3.3 A chegada às séries temporais

O caminho até os modelos de fundação de séries foi tudo menos linear. O N-BEATS mostrou que uma rede profunda só de camadas densas, sem componentes específicos de séries, superava o vencedor da competição M4 (Oreshkin et al. 2020). Em seguida, Zeng et al. (2023) mostraram que modelos lineares de uma camada batiam os Transformers de previsão de longo prazo em nove conjuntos de dados, abrindo o debate sobre se a atenção servia para séries. A resposta veio do PatchTST: cortar a série em blocos que viram *tokens* e tratar cada canal separadamente (Nie et al. 2023) — receita herdada por quase todos os modelos posteriores.

A partir de 2023 surgiram os modelos pré-treinados. O TimeGPT-1 foi apresentado como o primeiro modelo de fundação para séries, oferecido por API (Garza et al. 2023); o Lag-Llama, um decodificador aberto e probabilístico que usa defasagens como covariáveis (Rasul et al. 2023). Em 2024 vieram os três marcos que definiram a área. O Chronos tratou a série como linguagem: escala os valores pela média absoluta do contexto, quantiza-os em 4.094 faixas e treina um T5 com entropia cruzada, transformando a previsão em classificação (Ansari et al. 2024; Raffel et al. 2020). O TimesFM usou blocos de 32 pontos e um decodificador causal treinado em cerca de 100 bilhões de pontos de buscas, Wikipédia e dados sintéticos (Das et al. 2024). O Moirai introduziu blocos de tamanho dependente da frequência, atenção sobre qualquer número de variáveis e saída em mistura de distribuições, com o acervo LOTSA (Woo et al. 2024). Na mesma época, o MOMENT pré-treinou codificadores para várias tarefas (Goswami et al. 2024) e os Tiny Time Mixers mostraram que modelos minúsculos competem no *zero-shot* (Ekambaram et al. 2024). O benchmark GIFT-Eval, com um corpus de pré-treino separado do de teste, virou a régua comum (Aksu et al. 2024).

Tabela 2: Os modelos de fundação usados. Fontes: (Ansari et al. 2025; Google Research 2025, 2026; Auer et al. 2025; Liu et al. 2025; Cohen et al. 2025; IBM Research 2025; Garza et al. 2023).

Modelo	Organização	Arquitetura	Parâmetros	Saída
Chronos-2	Amazon, 2025	Codificador (T5), blocos de 16, atenção no tempo e entre séries do grupo	120 mi	21 quantis
TimesFM 2.5	Google, 2025	Decodificador causal, blocos de 32	200 mi	ponto e decis
TimesFM 3.0	Google, 2026	Decodificador com atenção entre variáveis	330 mi	ponto e 9 quantis
TiRex	NX-AI, 2025	xLSTM (sLSTM), blocos de 32	35 mi	decis e média
Moirai 2.0 (small)	Salesforce, 2025	Decodificador causal, blocos de 16	11 mi	9 quantis
Toto	Datadog, 2025	Decodificador, blocos de 64, atenção no tempo e entre variáveis	151 mi	mistura de t de Student
Granite TTM r2	IBM, 2025	Mixers (TSMixer), contexto de 52	0,07 mi	ponto
TimeGPT-1	Nixtla, 2023	Codificador-decodificador, por API	não informado	ponto e intervalos

Parâmetros dos checkpoints usados. Chronos-2 entra em duas versões: séries previstas isoladamente e com aprendizado entre séries.

3.4 Previsão probabilística e cuidados

Esses modelos produzem distribuições, não só pontos, o que exige métricas próprias. A perda *pinball* do quantil τ, $\rho_\tau(u) = u\,(\tau - \mathbb{1}\{u < 0\})$, é minimizada pelo quantil verdadeiro (Koenker e Bassett 1978), e o CRPS, regra estritamente própria, é o dobro da média da *pinball* sobre os quantis e se reduz ao erro absoluto quando a previsão é pontual (Gneiting e Raftery 2007). Dois cuidados acompanham o uso *zero-shot*. O primeiro é a contaminação: se as séries avaliadas estiverem no pré-treino, a "previsão" é uma lembrança; a recomendação é avaliar num período posterior ao corte do pré-treino. O segundo são as quebras: os modelos se adaptam rápido a um novo nível, mas não antecipam o choque. Um modelo de fundação é, portanto, mais um candidato no fluxo de previsão, sujeito à mesma disciplina: começar por referências simples, avaliar fora da amostra e no futuro, e só acreditar no ganho que sobrevive a essa disciplina (Hyndman e Athanasopoulos 2021; Wilher 2026).

4 Os nove modelos usados

A Tabela 2 resume os modelos; todos foram usados sem ajuste fino, com os pesos públicos nas versões indicadas. As descrições seguem os artigos e os cartões oficiais.

Chronos-2 (Ansari et al. 2025) abandonou a quantização do Chronos original: padroniza a série, corta-a em blocos de 16 pontos e usa um codificador em que cada camada combina atenção ao longo do tempo com uma atenção de grupo entre séries relacionadas. É isso que permite a versão "com aprendizado entre séries" usada aqui, em que os 377 subitens de uma mesma data compartilham informação. Foi pré-treinado em dados reais e sintéticos e devolve 21 quantis.

TimesFM 2.5 (Google Research 2025) mantém o desenho do TimesFM original (Das et al. 2024) — decodificador causal sobre blocos de 32 pontos — com 200 milhões de parâmetros, contexto de até 16 mil pontos e uma cabeça de quantis; o pré-treino inclui o corpus do GIFT-Eval, páginas da Wikipédia até 2023 e buscas até 2022. **TimesFM 3.0** (Google Research 2026) acrescenta atenção entre variáveis e covariáveis, com 330 milhões de parâmetros; tem licença não comercial.

TiRex (Auer et al. 2025) é a exceção arquitetural: em vez de um Transformer, usa células sLSTM do xLSTM (Beck et al. 2024), uma volta às redes recorrentes com memória que acompanha o estado da série. Com 35 milhões de parâmetros, prevê vários blocos à frente tratando o futuro como valor ausente, sem realimentar a própria previsão.

Moirai 2.0 (Liu et al. 2025) simplificou o Moirai original (Woo et al. 2024): um decodificador causal com tamanho de bloco único e saída em quantis. A versão pequena usada tem cerca de 11 milhões de parâmetros. **Toto** (Cohen et al. 2025), da Datadog, foi treinado majoritariamente em métricas de observabilidade de sistemas, com atenção alternada no tempo e entre variáveis e saída em mistura de distribuições t de Student. **Granite TTM** (Ekambaram et al. 2024; IBM Research 2025) é o menor: um *mixer* de cerca de 70 mil parâmetros na versão usada, sem dados macroeconômicos no pré-treino e, segundo o próprio cartão, sem suporte a frequência mensal — uma desvantagem que aparece nos resultados. **TimeGPT-1** (Garza et al. 2023) é acessado por API; seu corpus de pré-treino não é divulgado, o que o torna o único modelo com risco de contaminação não descartável.

Quanto à contaminação, nenhum corpus documentado contém o IPCA ou dados do IBGE; os dados macroeconômicos reais conhecidos terminam em 2018 ou antes, bem antes da janela de avaliação (2024–2026). A exposição é estrutural — séries macro mensais em geral — e não temporal.

5 Dados e desenho

Dados. Variação mensal e peso dos 377 subitens da tabela 7060 do SIDRA, de jan/2020 a set/2026. O IPCA em 12 meses é a série 13522 do SGS/BCB. As expectativas vêm da API Olinda do Focus (base de cálculo 0), sempre do primeiro relatório após a divulgação do IPCA da data de previsão.

Desenho. Avaliação por origem móvel com janela expansiva: em cada origem T, os modelos usam apenas os dados até T e preveem $T+1, \ldots, T+12$; as métricas são médias sobre as origens. As origens vão de dez/2022 a ago/2026; alvos de 2023 servem para escolher parâmetros e alvos de jan/2024 a set/2026 (33 por horizonte) para avaliar. Os subitens previstos são agregados pela Equação 1, com pesos projetados pela Equação 2. A comparação com o Focus usa o teste de Diebold-Mariano (Diebold e Mariano 1995) com a correção de amostra pequena de Harvey et al. (1997).

Vazamento. A literatura lista cinco caminhos pelos quais o futuro entra numa previsão: embaralhar a amostra, pré-processar com estatísticas da amostra inteira, usar variáveis que só existem depois, escolher o modelo no teste e contaminação do pré-treino (Kaufman et al. 2012). A pergunta que guiou cada escolha foi a das notas de aula (Wilher 2026): no instante em que a previsão seria usada para decidir, essa informação estaria disponível, com esse valor? Daí o Focus por safra de relatório, os pesos de t usados só a partir de $t-1$, a ANP restrita às semanas já publicadas e os créditos de energia só depois de anunciados.

6 O percurso da pesquisa

6.1 Primeiro resultado: os modelos puros

A primeira rodada aplicou os modelos de fundação aos subitens sem nenhum ajuste. Na janela de avaliação, o Chronos-2 com aprendizado entre séries teve, no mês seguinte, erro 1,53 vezes o do Focus quando desagregado, e 2,11 vezes quando aplicado ao índice cheio. Seis meses à frente, a razão caiu para 1,01. O padrão bate com o que o Banco Central documenta para seus próprios modelos: no horizonte muito curto, o mercado — que conhece anúncios de preços e coleta dados de alta frequência — é difícil de bater; o nicho dos modelos estatísticos está de 2–3 a 9–12 meses (Banco Central do Brasil 2025, 71).

6.2 Desagregar compensa

Comparando cada modelo desagregado com ele mesmo no índice cheio, a desagregação reduziu o erro do mês seguinte entre 17% e 23% nos sete modelos de uso geral, com significância a 5% em 5 deles (45 origens). As exceções foram o Granite TTM (razão 0,99), sem suporte à frequência mensal, e o TimeGPT (1,02). O resultado contrasta com o "decepcionante" da desagregação em quatro componentes do Banco Central (Araujo e Gaglianone 2022, 32) e é coerente com a leitura de Montero-Manso e Hyndman (2021): desagregar ao máximo só compensa quando o modelo das partes é global o bastante para não pagar o erro de estimação de cada série — exatamente o que um modelo de fundação oferece.

6.3 Onde estava o erro

Para decidir o passo seguinte, o erro do IPCA foi decomposto exatamente pelos subitens. Pela Equação 1, o erro de previsão é a soma de contribuições $c_i = (\hat{w}_i\hat{\pi}_i - w_i\pi_i)/100$, e a parcela do erro quadrático médio explicada por cada subitem é $\mathbb{E}[c_i e]/\mathbb{E}[e^2]$. No modelo puro, os preços administrados respondiam por 75% do erro do mês seguinte, e só a energia elétrica, por 58%. O diagnóstico apontou onde o Focus tinha vantagem: preços decididos por regra ou por autoridade, que o mercado conhece antes e que nenhum modelo puramente estatístico consegue antecipar.

6.4 Informação institucional: combustíveis e energia

O passo seguinte foi trazer essa informação para o modelo, como um analista faria. Para os combustíveis, o preço semanal da ANP já publicado na data de previsão substitui a previsão do modelo para gasolina, etanol, diesel e gás de botijão. O efeito no índice foi pequeno: a razão contra o Focus passou de 1,53 para 1,54, porque na data de divulgação do IPCA a ANP cobre só parte do mês seguinte. Para a energia, a bandeira tarifária do mês seguinte, já anunciada pela ANEEL, e os créditos do bônus de Itaipu, só depois de anunciados, entraram como ajustes. Aqui o ganho foi grande: a razão caiu para 1,05. É a mesma lição do WP 561: "a qualidade do conjunto de informação importa mais que o método de combinação" (Araujo e Gaglianone 2022, 32), e o box do RPM mostra a inflação implícita dos títulos públicos entre as variáveis mais relevantes no curto prazo (Banco Central do Brasil 2025).

6.5 Combinar modelos

Cada modelo híbrido foi então misturado à média sazonal (com o peso do mês seguinte escolhido só com 2023), e os nove foram combinados com pesos proporcionais ao inverso do erro passado em cada horizonte. A literatura de combinação de previsões diz que os pesos ótimos dependem das variâncias e covariâncias dos erros (Bates e Granger 1969) e que, na prática, pesos iguais costumam superar pesos estimados — o *forecast combination puzzle* (Timmermann 2006). Os resultados reproduzem o quebra-cabeça: a média simples e o ensemble ponderado ficam praticamente iguais (Tabela 4). A regra recursiva do Banco Central, média dos três melhores por horizonte (Banco Central do Brasil 2025), difere do ensemble em no máximo 0,025 na razão contra o Focus. A própria mediana do Focus é um *ensemble*, e a correlação entre os erros dos analistas, que leem as mesmas notícias, limita o ganho da agregação; o mesmo vale entre modelos de fundação treinados em corpora parecidos. Com tudo isso, o ensemble chegou a 1,009 do erro do Focus no mês seguinte.

6.6 Auditar as datas

Uma comparação justa exige que modelos e Focus usem a mesma informação. A auditoria mostrou que o campo de data da API do Focus é a data da pesquisa, não a da publicação, e que as atualizações da ANP precisam de data documentada. Os cortes foram refeitos com as datas efetivas, arquivadas com o seu código de verificação (*hash*). É o que a literatura chama de vazamento por variável que só existe depois (Kaufman et al. 2012), aplicado a dados de safra.

6.7 Medir a incerteza

Os intervalos de previsão seguiram o WP 561 do Banco Central: quantis gaussianos centrados na previsão, com o desvio-padrão dos erros passados por horizonte, suavizado para crescer com o horizonte (Araujo e Gaglianone 2022, seç. 2.3). Em seguida, testou-se corrigir a previsão pelo viés médio passado, já que os modelos superestimaram a inflação em 2024–2026. A correção chegou a ser adotada na previsão atual, mas o teste honesto — viés estimado em cada data só com os erros já conhecidos — mostrou que ela piorava a previsão e descalibrava os intervalos (seção 7). A decisão foi revertida: o viés passado não previu o viés seguinte, uma instância do dilema viés–variância (Geman et al. 1992). Os quantis nativos dos modelos também foram testados e se mostraram largos demais.

6.8 Da pesquisa à produção

Por fim, a previsão foi levada a produção: um único comando refaz a coleta, roda os nove modelos na data mais recente, aplica os ajustes, recalcula o ensemble e os intervalos, atualiza este texto e publica um painel. Cada rodada é gravada num registro que só recebe acréscimos, separado do *backtest*. A literatura de operação de modelos insiste que o código de aprendizado de máquina é uma fração pequena do sistema (Sculley et al. 2015) e que uma previsão só é reproduzível quando se sabe com que dados, código e artefato ela foi feita (Huyen 2022; Kreuzberger et al. 2023). O registro em tempo real é o teste definitivo: monitorar um modelo em produção é, no vocabulário econométrico, um teste sequencial de quebra estrutural, e é ele que dirá se o desempenho do *backtest* sobrevive (Lucas 1976).

Tabela 3: Erro da previsão mensal do ensemble contra o Focus.

h	n	RMSE Focus	RMSE ensemble	Razão	p DM-HLN	Razão média simples
1	33	0,163	0,164	1,009	0,899	1,009
2	33	0,199	0,197	0,994	0,951	0,995
3	33	0,226	0,225	0,998	0,972	1,000
4	33	0,259	0,250	0,964	0,602	0,966
5	33	0,274	0,251	0,915	0,333	0,916
6	33	0,281	0,256	0,911	0,396	0,913
7	33	0,278	0,257	0,924	0,397	0,926
8	33	0,274	0,255	0,932	0,504	0,934
9	33	0,273	0,251	0,921	0,414	0,921
10	33	0,272	0,254	0,933	0,498	0,933
11	33	0,274	0,259	0,945	0,622	0,946
12	33	0,271	0,252	0,930	0,476	0,930

IPCA mensal, alvos de jan/2024 a set/2026. Razão = RMSE da previsão / RMSE do Focus (mediana do primeiro relatório após a divulgação do IPCA da origem). Ajustes de ANP e Itaipu com corte na data da pesquisa Focus (script 17). DM-HLN: Diebold-Mariano com correção de Harvey, Leybourne e Newbold.

7 Resultados

7.1 IPCA do mês

A Tabela 3 resume o erro do ensemble. Em $h = 1$, a razão é 1,009: o ensemble empata com o Focus no mês seguinte. De 2 a 12 meses, a razão fica entre 0,911 e 0,998, sempre a favor do ensemble, mas o menor p-valor é 0,333: nenhuma diferença é significativa.

7.2 Onde o erro sobra

No ensemble final, o erro do mês seguinte se concentra em poucos preços (Figura 1). Passagem aérea lidera, com 22,8% do erro quadrático; alimentação no domicílio responde por 40%. Três meses à frente, Energia elétrica volta a liderar, com 34,6%, e seis meses à frente os preços administrados respondem por 63% do erro: os ajustes resolvem a bandeira do mês seguinte, mas as decisões regulatórias mais distantes continuam imprevisíveis para os modelos.

Tabela 4: Razão RMSE/Focus por modelo.

Modelo	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Chronos-2 CL	1,024	0,981	0,917	0,954
Chronos-2	1,023	0,980	0,913	0,950
TimesFM 3	1,001	0,985	0,925	0,929
TimesFM 2.5	1,063	1,008	0,939	0,945
TiRex	0,977	1,010	0,933	0,973
Moirai 2.0	1,042	0,979	0,885	0,907
Toto	1,043	0,972	0,906	0,930
Granite TTM	1,224	1,134	1,024	0,994
TimeGPT-1	1,155	1,075	0,912	0,935
Média simples	1,009	1,000	0,913	0,930
Ensemble 1/RMSE	1,009	0,998	0,911	0,930

Razão RMSE/Focus do IPCA mensal por modelo híbrido (ANP, bandeira, Itaipu e mistura sazonal), alvos desde jan/2024.

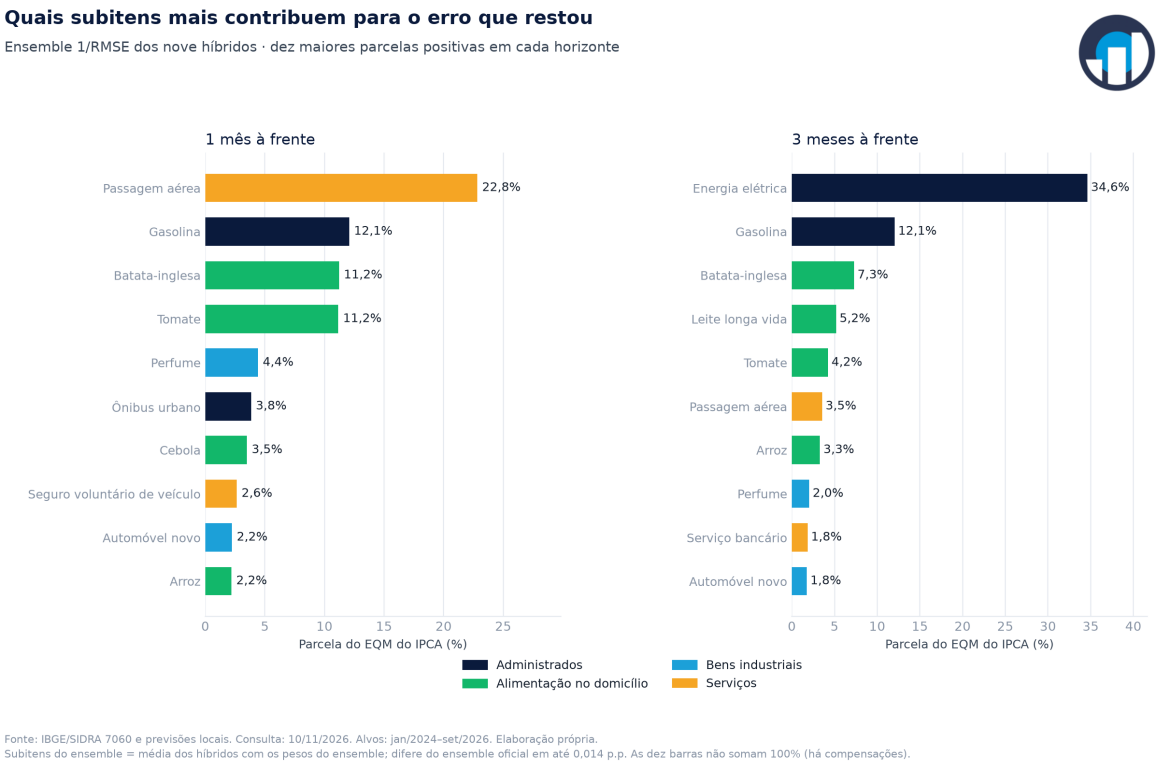


Figura 1: Subitens que mais contribuem para o erro do ensemble, 1 e 3 meses à frente (alvos desde jan/2024).

Tabela 5: Calibração dos intervalos no backtest.

Objeto	Centro	Cobertura 50%	Cobertura 80%	CRPS	Erro abs. médio
IPCA 12 meses	sem correção	52,4	78,4	0,345	0,484
IPCA 12 meses	corrigido	35,9	67,0	0,506	0,704
IPCA mensal	sem correção	51,1	74,6	0,140	0,194
IPCA mensal	corrigido	47,9	74,6	0,146	0,201

Intervalos gaussianos com viés e desvio-padrão por h estimados em cada origem só com erros já divulgados (mínimo 6), desvio-padrão monótono em h; alvos desde jan/2024, todos os horizontes empilhados.

7.3 Incerteza

Os intervalos seguem o WP 561: quantis gaussianos centrados na previsão, com desvio-padrão dos erros por horizonte, aqui tornado monótono em h. Para testá-los, cada origem usa apenas os erros já divulgados (Tabela 5). Sem correção de viés, o intervalo de 80% do IPCA em 12 meses cobriu 78% dos casos. Corrigindo pelo viés estimado em tempo real, a cobertura cai para 67% e o erro absoluto médio sobe de 0,48 para 0,70 p.p.: o viés passado não previu o viés seguinte.

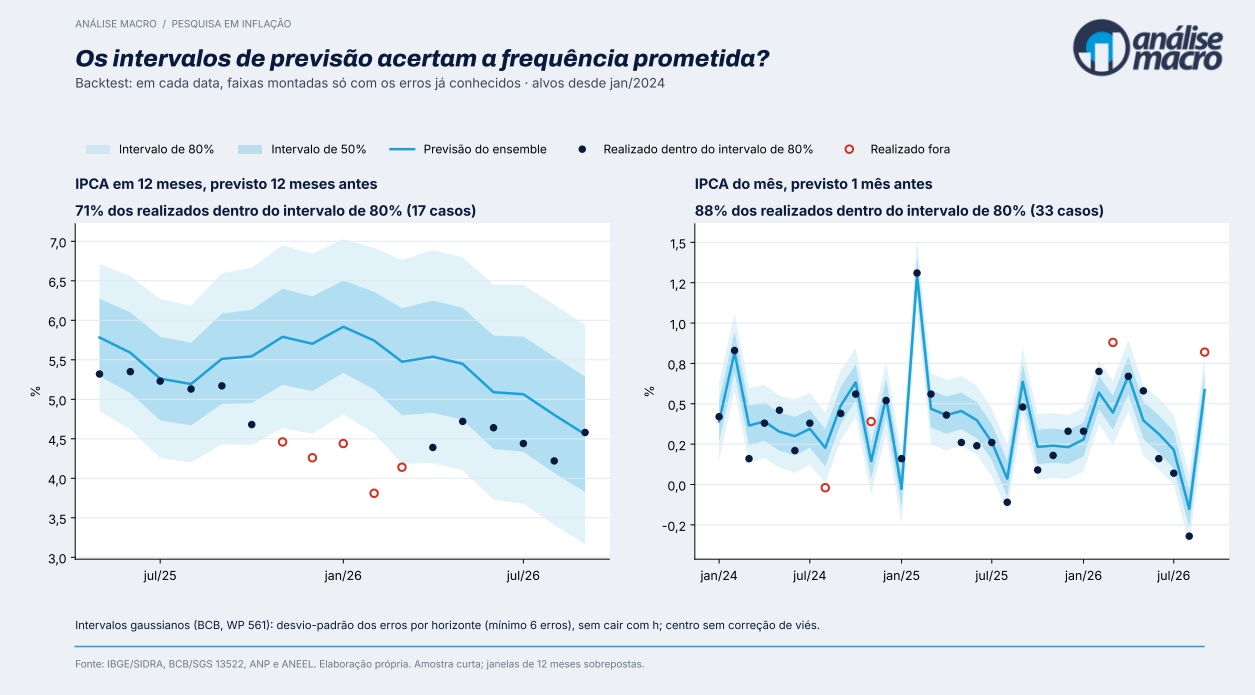


Figura 2: Previsões passadas do ensemble com intervalos de 50% e 80%, montados só com os erros conhecidos em cada data; pontos vazados: realizado fora do intervalo de 80%.

Os quantis nativos dos modelos de fundação, avaliados na previsão direta do índice cheio (Tabela 6), são largos demais: o intervalo de 80% cobre de 81% a 90% dos casos. Em 6 dos 6 modelos, o leque construído com os erros passados do próprio modelo tem CRPS igual ou menor, e o leque do ensemble híbrido (0,152) supera todos os nativos (0,196 a 0,217): os ajustes e a combinação valem mais que a distribuição de cada modelo.

Tabela 6: Quantis nativos contra o leque pelos erros passados.

Modelo	CRPS nativo	CRPS leque	Cob. 80% nativo	Cob. 80% leque	Larg. nativo	Larg. leque
Chronos-2	0,208	0,202	89,5	80,6	1,14	0,83
Moirai 2.0	0,211	0,206	90,4	80,2	1,15	0,85
TimeGPT-1	0,196	0,179	80,6	81,2	0,97	0,74
TimesFM 2.5	0,217	0,217	87,0	73,5	1,07	0,75
TiRex	0,216	0,208	88,9	78,1	1,21	0,82
Toto	0,214	0,206	81,2	82,1	0,97	0,90
Ensemble híbrido	—	0,152	—	75,0	—	0,57

IPCA mensal previsto diretamente no índice cheio, $h = 1$ a 12, alvos desde jan/2024. Nativo: decis do próprio modelo. Leque: gaussiano em torno da previsão pontual, com σ por h dos erros já divulgados (monótono em h). CRPS aproximado pela média de 2× perda pinball nos 9 decis.

Tabela 7: Trajetória prevista do IPCA em 12 meses.

Mês	Mês prev.	Focus mês	Mês que sai	IPCA 12m	Interv. 80%	P(>4,5%)	Ref. corrigida
out/2026	0,31	0,33	0,09	4,80	4,60–5,01	97	4,80
nov/2026	0,38	0,35	0,18	5,01	4,64–5,38	96	4,95
dez/2026	0,45	0,56	0,33	5,14	4,59–5,69	93	5,02
jan/2027	0,28	0,46	0,33	5,09	4,41–5,77	87	4,93
fev/2027	0,75	0,66	0,70	5,14	4,38–5,91	86	4,96
mar/2027	0,47	0,41	0,88	4,71	3,88–5,55	63	4,48
abr/2027	0,40	0,44	0,67	4,43	3,53–5,34	46	4,16
mai/2027	0,43	0,31	0,58	4,27	3,33–5,22	38	3,93
jun/2027	0,25	0,22	0,16	4,37	3,42–5,31	43	3,95
jul/2027	0,25	0,22	0,07	4,55	3,61–5,50	53	4,05
ago/2027	0,03	0,13	-0,32	4,92	3,98–5,87	72	4,31
set/2027	0,46	0,29	0,82	4,55	3,61–5,49	53	3,83

Previsão com dados até set/2026. IPCA 12m = meses realizados + previstos, compostos por produto; ponto = ensemble sem correção de viés; ref. corrigida = ponto menos o erro médio do backtest por horizonte; intervalo e probabilidade gaussianos com o desvio-padrão dos erros do backtest (monótono em h). Todas as colunas em %.

8 Previsão atual

Com dados até set/2026, o ensemble prevê IPCA de 0,31% em out/2026 (intervalo de 80% de 0,11 a 0,50%), contra 0,33% do Focus mensal. A trajetória do IPCA em 12 meses (Tabela 7, Figura 3) fica acima do teto de 4,5% em 9 dos 12 meses, 6 deles seguidos a partir de out/2026: 5,14% em dez/2026 (probabilidade de 93% de passar o teto) e 4,55% em set/2027 (intervalo de 80% de 3,61 a 5,49%). O Focus de 09/10/2026 indica 5,22% para 2026 e 4,48% nos 12 meses seguintes. A previsão pontual é o ensemble sem correção de viés. Descontando o erro médio do backtest em cada horizonte (0,72 p.p. a 12 meses), o fim da trajetória seria 3,83%; essa versão fica só como referência, porque, como mostra a seção anterior, a correção estimada em tempo real piorou a previsão. A primeira rodada registrada, de 10/11/2026, usou a versão corrigida antes dessa revisão; o registro identifica a versão de cada rodada.

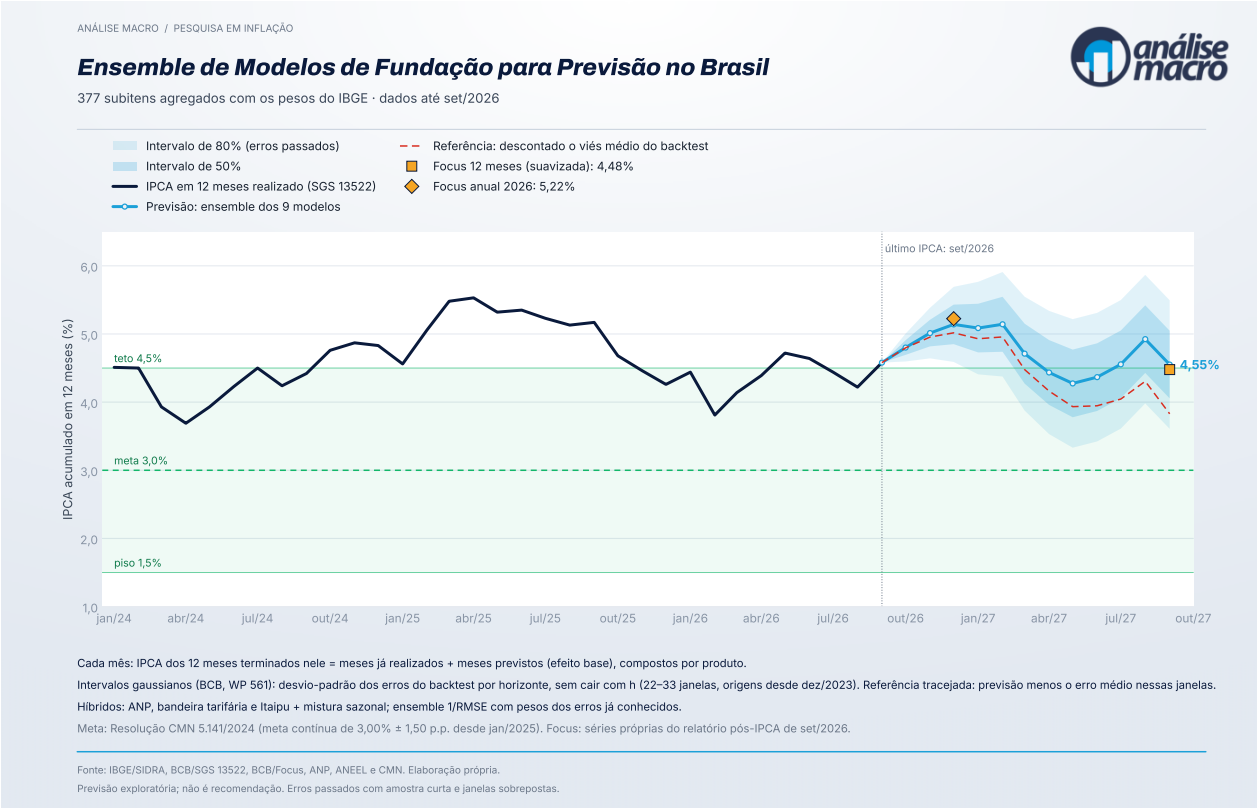


Figura 3: Trajetória prevista do IPCA em 12 meses, intervalos de 50% e 80%, meta e Focus.

9 Limitações

A amostra de avaliação tem 33 meses e as janelas de 12 meses se sobrepõem. Parte do desenho (ajustes de energia, mistura sazonal, correção de viés) foi formulada depois de observar 2024–2026, embora os parâmetros numéricos venham de 2023. Os modelos podem ter visto o IPCA no pré-treino. As datas de atualização da ANP não garantem os valores de cada vintage, e a autenticação documental dos créditos de Itaipu está incompleta. TimesFM 3 (na versão instalada) e Granite TTM não têm distribuição nativa, e a distribuição bottom-up conjunta dos 377 subitens não foi construída.

10 Conclusão

A ideia de partida — prever o IPCA pelos seus 377 componentes — só se tornou viável com modelos globais o bastante para não pagar o erro de estimação de cada série, e os modelos de fundação cumprem esse papel: desagregar reduziu o erro em todos os modelos de uso geral. Sozinhos, porém, eles não alcançam o mercado no curto prazo; a diferença estava nos preços administrados, e foi a informação institucional sobre combustíveis e, sobretudo, energia que fechou a distância. Combinada e ajustada por essas regras simples, a previsão desagregada alcança o Focus no IPCA mensal e fica à frente dele, sem significância, de 2 a 12 meses. Os intervalos baseados em erros passados são bem calibrados e superam as distribuições nativas dos modelos; a correção de viés não ajuda. O

próximo passo é acompanhar o registro de previsões ao vivo, que separa o desempenho genuíno do backtest.

Referências

Aksu, Taha, Gerald Woo, Juncheng Liu, et al. 2024. *GIFT-Eval: A Benchmark For General Time Series Forecasting Model Evaluation*. arXiv preprint arXiv:2410.10393. https://arxiv.org/abs/2410.10393.

Ansari, Abdul Fatir, Oleksandr Shchur, Jaris Küken, et al. 2025. *Chronos-2: From Univariate to Universal Forecasting*. arXiv preprint arXiv:2510.15821. https://arxiv.org/abs/2510.15821.

Ansari, Abdul Fatir, Lorenzo Stella, Caner Turkmen, et al. 2024. "Chronos: Learning the Language of Time Series". *Transactions on Machine Learning Research*. https://openreview.net/forum?id=gerNCVqqtR.

Araujo, Gustavo Silva, e Wagner Piazza Gaglianone. 2022. *Machine Learning Methods for Inflation Forecasting in Brazil: New Contenders versus Classical Models*. Working Paper Series N. 561. Banco Central do Brasil. https://www.bcb.gov.br/pec/wps/ingl/wps561.pdf.

Auer, Andreas, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, e Sepp Hochreiter. 2025. "TiRex: Zero-Shot Forecasting Across Long and Short Horizons with Enhanced In-Context Learning". *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*. https://openreview.net/forum?id=v7UqniC9pF.

Banco Central do Brasil. 2025. "Tópicos especiais de previsão de inflação com aprendizado de máquina". Em *Relatório de Política Monetária*, v. 1. Banco Central do Brasil. https://www.bcb.gov.br/content/ri/relatorioinflacao/202512/rpm202512p.pdf.

Bates, J. M., e C. W. J. Granger. 1969. "The Combination of Forecasts". *Operational Research Quarterly* 20 (4): 451–68.

Beck, Maximilian, Korbinian Pöppel, Markus Spanring, et al. 2024. "xLSTM: Extended Long Short-Term Memory". *Advances in Neural Information Processing Systems* 37. https://proceedings.neurips.cc/paper_files/paper/2024/hash/c2ce2f2701c10a2b2f2ea0bfa43cfaa3-Abstract-Conference.html.

Bengio, Yoshua, Patrice Simard, e Paolo Frasconi. 1994. "Learning Long-Term Dependencies with Gradient Descent is Difficult". *IEEE Transactions on Neural Networks* 5 (2): 157–66.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, et al. 2021. *On the Opportunities and Risks of Foundation Models*. arXiv preprint arXiv:2108.07258. https://arxiv.org/abs/2108.07258.

Brasil. 1999. *Decreto nº 3.088, de 21 de junho de 1999: estabelece a sistemática de "metas para a inflação" como diretriz para fixação do regime de política monetária*. Https://www.planalto.gov.br/ccivil_03/decreto/d3088.htm.

Brasil. 2024. *Decreto nº 12.079, de 26 de junho de 2024: estabelece nova sistemática de meta para a inflação como diretriz para fixação do regime de política monetária*. Https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/decreto/D12079.htm.

Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, et al. 2014. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation". *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 1724–34.

Cohen, Ben, Emaad Khwaja, Youssef Doubli, et al. 2025. "This Time is Different: An Observability Perspective on Time Series Foundation Models". *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*. https://openreview.net/forum?id=1jDAYXfcS2.

Conselho Monetário Nacional. 2024. *Resolução CMN nº 5.141, de 26 de junho de 2024: fixa o índice de preços, a meta para a inflação e seu intervalo de tolerância aplicáveis a partir de 1º de janeiro de 2025*. Diário Oficial da União, 27 jun. 2024, Seção 1, p. 71.

Das, Abhimanyu, Weihao Kong, Rajat Sen, e Yichen Zhou. 2024. "A decoder-only foundation model for time-series forecasting". *Proceedings of the 41st International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, v. 235: 10148–67. https://arxiv.org/abs/2310.10688.

Diebold, Francis X., e Roberto S. Mariano. 1995. "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics* 13 (3): 253–63.

Ekambaram, Vijay, Arindam Jati, Pankaj Dayama, et al. 2024. "Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series". *Advances in Neural Information Processing Systems* 37. https://proceedings.neurips.cc/paper_files/paper/2024/hash/874a4d89f2d04b4bcf9a2c19545cf040-Abstract-Conference.html.

Garza, Azul, Cristian Challu, e Max Mergenthaler-Canseco. 2023. *TimeGPT-1*. arXiv preprint arXiv:2310.03589. https://arxiv.org/abs/2310.03589.

Geman, Stuart, Elie Bienenstock, e René Doursat. 1992. "Neural Networks and the Bias/Variance Dilemma". *Neural Computation* 4 (1): 1–58.

Gneiting, Tilmann, e Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation". *Journal of the American Statistical Association* 102 (477): 359–78.

Google Research. 2025. *TimesFM 2.5 (google/timesfm-2.5-200m-pytorch)*. Hugging Face model card, https://huggingface.co/google/timesfm-2.5-200m-pytorch.

Google Research. 2026. *TimesFM 3.0 (google/timesfm-3.0-pytorch)*. Hugging Face model card, https://huggingface.co/google/timesfm-3.0-pytorch.

Goswami, Mononito, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, e Artur Dubrawski. 2024. "MOMENT: A Family of Open Time-series Foundation Models". *Proceedings of the 41st International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, v. 235: 16115–52. https://arxiv.org/abs/2402.03885.

Harvey, David, Stephen Leybourne, e Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors". *International Journal of Forecasting* 13 (2): 281–91.

Hochreiter, Sepp, e Jürgen Schmidhuber. 1997. "Long Short-Term Memory". *Neural Computation* 9 (8): 1735–80.

Huyen, Chip. 2022. *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*. O'Reilly Media.

Hyndman, Rob J., e George Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3º ed. OTexts.

IBM Research. 2025. *Granite TimeSeries TinyTimeMixer R2 (ibm-granite/granite-timeseries-ttm-r2, branch 52-16-ft-r2.1)*. Hugging Face model card, https://huggingface.co/ibm-granite/granite-timeseries-ttm-r2.

Instituto Brasileiro de Geografia e Estatística. 2013. *Sistema Nacional de Índices de Preços ao Consumidor: métodos de cálculo*. 7º ed. V. 14. Relatórios Metodológicos. IBGE, Coordenação de Índices de Preços. https://biblioteca.ibge.gov.br/visualizacao/livros/liv65477.pdf.

Instituto Brasileiro de Geografia e Estatística. 2020. *Sistema Nacional de Índices de Preços ao Consumidor: métodos de cálculo*. 8º ed. V. 14. Relatórios Metodológicos. IBGE, Coordenação de Índices de Preços. https://biblioteca.ibge.gov.br/visualizacao/livros/liv101767.pdf.

Instituto Brasileiro de Geografia e Estatística. 2026. *Tabela 7060: IPCA, variação mensal e peso mensal por subitem (a partir de janeiro/2020)*. SIDRA/IBGE, API https://servicodados.ibge.gov.br/api/v3/agregados/7060, variáveis 63 e 66, classificação 315.

Kaufman, Shachar, Saharon Rosset, Claudia Perlich, e Ori Stitelman. 2012. "Leakage in Data Mining: Formulation, Detection, and Avoidance". *ACM Transactions on Knowledge Discovery from Data* 6 (4): 15:1–21.

Koenker, Roger, e Gilbert Bassett. 1978. "Regression Quantiles". *Econometrica* 46 (1): 33–50.

Kreuzberger, Dominik, Niklas Kühl, e Sebastian Hirschl. 2023. "Machine Learning Operations (MLOps): Overview, Definition, and Architecture". *IEEE Access* 11: 31866–79. https://doi.org/10.1109/ACCESS.2023.3262138.

Liu, Chenghao, Taha Aksu, Juncheng Liu, et al. 2025. *Moirai 2.0: When Less Is More for Time Series Forecasting*. arXiv preprint arXiv:2511.11698. https://arxiv.org/abs/2511.11698.

Lucas, Robert E. 1976. "Econometric Policy Evaluation: A Critique". *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.

Montero-Manso, Pablo, e Rob J. Hyndman. 2021. "Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality". *International Journal of Forecasting* 37 (4): 1632–53.

Nie, Yuqi, Nam H. Nguyen, Phanwadee Sinthong, e Jayant Kalagnanam. 2023. "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers". *International Conference on Learning Representations (ICLR)*. https://iclr.cc/virtual/2023/poster/10876.

Oreshkin, Boris N., Dmitri Carpov, Nicolas Chapados, e Yoshua Bengio. 2020. "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting". *International Conference on Learning Representations (ICLR)*. https://iclr.cc/virtual/2020/poster/1743.

Pascanu, Razvan, Tomas Mikolov, e Yoshua Bengio. 2013. "On the Difficulty of Training Recurrent Neural Networks". *Proceedings of the 30th International Conference on Machine Learning*, 1310–18.

Raffel, Colin, Noam Shazeer, Adam Roberts, et al. 2020. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *Journal of Machine Learning Research* 21 (140): 1–67.

Rasul, Kashif, Arjun Ashok, Andrew Robert Williams, et al. 2023. *Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting*. arXiv preprint arXiv:2310.08278. https://arxiv.org/abs/2310.08278.

Rumelhart, David E., Geoffrey E. Hinton, e Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors". *Nature* 323: 533–36.

Sculley, D., Gary Holt, Daniel Golovin, et al. 2015. "Hidden Technical Debt in Machine Learning Systems". *Advances in Neural Information Processing Systems* 28: 2503–11.

Timmermann, Allan. 2006. "Forecast Combinations". Em *Handbook of Economic Forecasting*, organizado por Graham Elliott, Clive W. J. Granger, e Allan Timmermann, v. 1. Elsevier.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. "Attention is All you Need". *Advances in Neural Information Processing Systems* 30. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Werbos, Paul J. 1990. "Backpropagation Through Time: What It Does and How to Do It". *Proceedings of the IEEE* 78 (10): 1550–60.

Wilher, Vítor. 2026. *Notas de aula de Decisões Baseadas em Dados I*. Doutorado em Economia, FGV EPGE.

Woo, Gerald, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, e Doyen Sahoo. 2024. "Unified Training of Universal Time Series Forecasting Transformers". *Proceedings of the 41st International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, v. 235: 53140–64. https://arxiv.org/abs/2402.02592.

Zeng, Ailing, Muxi Chen, Lei Zhang, e Qiang Xu. 2023. "Are Transformers Effective for Time Series Forecasting?" *Proceedings of the AAAI Conference on Artificial Intelligence* 37: 11121–28. https://doi.org/10.1609/aaai.v37i9.26317.