

Pedro Maurício Ximenez da Silva

**Previsão do índice de preços ao consumidor:
uma abordagem via modelos dinâmicos**

Niterói - RJ, Brasil

11 de fevereiro de 2022

Pedro Maurício Ximenez da Silva

**Previsão do índice de preços ao
consumidor: uma abordagem via
modelos dinâmicos**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Jony Arrais Pinto Junior

Niterói - RJ, Brasil

11 de fevereiro de 2022

Pedro Maurício Ximenez da Silva

**Previsão do índice de preços ao consumidor:
uma abordagem via modelos dinâmicos**

Monografia de Projeto Final de Graduação sob o título “*Previsão do índice de preços ao consumidor: uma abordagem via modelos dinâmicos*”, defendida por Pedro Maurício Ximenez da Silva e aprovada em 11 de fevereiro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Jony Arrais Pinto Junior
Departamento de Estatística - UFF

Prof. Dr. Rafael Santos Erbisti
Departamento de Estatística - UFF

Profa. Dra. Patrícia Lusié Velozo da Costa
Departamento de Estatística - UFF

Niterói, 11 de fevereiro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

S586p Silva, Pedro Maurício Ximenez da
Previsão do índice de preços ao consumidor: uma abordagem
via modelos dinâmicos / Pedro Maurício Ximenez da Silva ;
Jony Arrais Pinto Junior, orientador. Niterói, 2022.
57 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística)-Universidade Federal Fluminense, Instituto de
Matemática e Estatística, Niterói, 2022.

1. Estatística. 2. Modelos lineares dinâmicos. 3.
Inferência Bayesiana. 4. IPCA. 5. Produção intelectual. I.
Junior, Jony Arrais Pinto, orientador. II. Universidade
Federal Fluminense. Instituto de Matemática e Estatística.
III. Título.

CDD -

Resumo

Com o surgimento da pandemia causada pelo vírus Sars-CoV-2, as economias mundiais foram afetadas com quedas nas suas produções, impactando preços de diversos produtos e serviços. Estes aumentos englobaram os preços dos alimentos, eletrodomésticos, serviços, energia elétrica, água e esgoto entre outros. Para avaliar esse aumento, é necessário entender a dinâmica temporal dos índices de preços. No Brasil, o índice oficial da inflação adotado pelo governo federal é o Índice Nacional de Preços ao Consumidor Amplo (IPCA), que por sua vez, aponta a variação do custo de vida médio de famílias com renda mensal de 1 e 40 salários mínimos. Além disso, serve de referência para as metas de inflação do país e para as alterações na taxa de juros. Com isso, a previsão desse índice pode antecipar os movimentos da economia nacional, podendo eliminar ou suavizar possíveis choques econômicos futuros. Nesse trabalho, serão utilizados os Modelos Lineares Dinâmicos (MLD) sob a perspectiva bayesiana, com o propósito de modelar e prever a série temporal do IPCA. Os modelos analisados neste trabalho são: o Modelo polinomial de 1° ordem (MP), caracterizado pela evolução temporal do nível ser modelada como um passeio aleatório simples, ou média localmente constante e o Modelo polinomial de 1° ordem com sazonalidade (MPS) que adiciona a equação de observação um termo de sazonalidade e a equação de evolução uma matriz que descreve a evolução dos parâmetros de estado no tempo. Os resultados dos ajustes destes modelos foram bem satisfatórios com desvio absoluto médio (MAD) e erro percentual absoluto médio (MAPE) baixos, principalmente os do Modelo Polinomial de 1° Ordem com Sazonalidade (MPS). A previsão obtida pelo MPS foi bem melhor que no modelo polinomial de 1° ordem, tendo em vista que, além de possuir o menor EQM entre os modelos, o intervalo de credibilidade (IC) da previsão cobria boa parte dos valores da série original e não ocorreu um aumento da incerteza, como no Modelo Polinomial de 1° Ordem (MP). Foi realizada uma previsão do IPCA para o ano de 2022 usando o modelo MPS.

Palavras-chave: Estatística. Modelos lineares dinâmicos. Inferência Bayesiana. IPCA.

Agradecimentos

Agradeço, primeiramente, a minha Mãe, Rosana V. Machado da Silva, minha Vó, Neuza Maria Machado da Silva, e a minha Tia/Avó(em memória), Maria Nilza da Cruz Bispo, por sempre me darem apoio e assim proporcionar a melhor educação possível e por me ensinar a ser sempre uma pessoa melhor.

Ao meu orientador o Prof. Dr. Jony Arrais Pinto Junior, agradeço por toda a paciência e dedicação para me auxiliar neste trabalho.

Agradeço também a todos os professores que me deram aula e que conheci durante a minha formação. Também agradeço aos meus amigos por estarem comigo ao longo desse jornada.

Por fim, agradeço a Deus por ter saúde no meio desse período pandêmico para concluir este trabalho.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviações	p. 12
1 Introdução	p. 13
2 Materiais e Métodos	p. 17
2.1 Material	p. 17
2.2 Modelos Lineares Dinâmicos	p. 17
2.2.1 Modelo polinomial de 1° ordem	p. 21
2.2.2 Modelo polinomial de 1° ordem com sazonalidade	p. 21
2.3 Inferência Bayesiana	p. 23
2.3.1 Filtro de Kalman	p. 24
2.3.2 Monte Carlo via Cadeias de Markov	p. 25
2.3.3 Previsão de distribuições <i>a posteriori</i>	p. 26
2.3.4 Previsão <i>k</i> -passos à frente	p. 27
2.4 Medidas de Comparação entre os Modelos	p. 27
2.4.1 Medidas de qualidade do ajuste	p. 28
2.4.2 Avaliação da previsão	p. 28
3 Análise dos Resultados	p. 30
3.1 Análise descritiva dos dados	p. 30

3.2	Resultados dos modelos lineares dinâmicos	p. 31
3.2.1	Ajustando o modelo polinomial de 1° ordem	p. 32
3.2.2	Ajustando o modelo polinomial de 1° ordem com sazonalidade	p. 34
3.2.3	Comparando os ajustes dos modelos	p. 36
3.3	Resultados das previsões dos modelos	p. 37
3.3.1	Previsão dos modelos 4 passos a frente para 2021	p. 37
3.3.2	Previsão dos modelos 12 passos a frente para 2021	p. 39
3.3.3	Comparando as previsões dos modelos	p. 41
3.3.4	Previsão do modelo polinomial de 1° ordem com sazonalidade para 2022	p. 42
4	Conclusão	p. 44
	Referências	p. 46
	Apêndice 1 – Tabela com a previsão de 2022 do modelo polinomial de 1° ordem com sazonalidade para o índice geral do IPCA	p. 49
	Apêndice 2 – Código em R do ajuste do modelo polinomial de 1° ordem	p. 50
	Apêndice 3 – Código em R do ajuste do modelo polinomial de 1° ordem com sazonalidade	p. 51
	Apêndice 4 – Código em R da previsão do modelo polinomial de 1° ordem	p. 52
	Apêndice 5 – Código em R da previsão do modelo polinomial de 1° ordem com sazonalidade	p. 53
	Apêndice 6 – Cadeias de Markov para μ_t do modelo polinomial de 1° ordem	p. 54
	Apêndice 7 – Cadeias de Markov para μ_t e para os γ_{12m} do modelo	

Lista de Figuras

1	Estrutura Índice de Preço ao Consumidor Amplo (IBGE) (IPCA)	p. 14
2	Série Temporal do IPCA (IBGE) de janeiro de 2012 à dezembro de 2021.	p. 31
3	Amostra <i>a posteriori</i> para W e V no modelo polinomial de 1° ordem. As linhas em vermelho são os Intervalo de credibilidade (IC) com nível de credibilidade de 95%, a esquerda o limite inferior e a direita o limite superior. A linha azul é a mediana.	p. 32
4	Série ajustada para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 33
5	Amostra <i>a posteriori</i> para $\tau_{\theta,1}^{-1}$ e V no modelo polinomial de 1° ordem com sazonalidade. As linhas em vermelho são os IC com nível de credibilidade de 95%, a esquerda o limite inferior e a direita o limite superior. A linha azul é a mediana.	p. 35
6	Série ajustada para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 36
7	Previsão 4 passos a frente para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 37
8	Previsão 4 passos a frente para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 38
9	Previsão 12 passos a frente para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho é intervalo de credibilidade de 95%.	p. 39

10	Previsão 12 passos a frente para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 40
11	Previsão de 2022 para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.	p. 42
12	Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para μ_t do modelo polinomial de 1° ordem.	p. 54
13	Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_2 do modelo polinomial de 1° ordem com sazonalidade.	p. 55
14	Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_5 do modelo polinomial de 1° ordem com sazonalidade.	p. 56
15	Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_9 do modelo polinomial de 1° ordem com sazonalidade.	p. 57

Lista de Tabelas

1	Estimativa pontual e intervalar para V e W	p. 33
2	Estimativa pontual e intervalar para V e $\tau_{\theta,1}^{-1}$	p. 35
3	Erro Percentual Absoluto Médio (MAPE) dos Modelos.	p. 36
4	Erro Quadrático Médio (EQM) das Previsões.	p. 41
5	Previsão de 2022 do modelo polinomial de 1° ordem com sazonalidade para o índice geral do IPCA.	p. 49

Lista de Abreviações

IPCA Índice de Preço ao Consumidor Amplo (IBGE)

POF Pesquisa de Orçamentos Familiares

EQM Erro Quadrático Médio

MAD Desvio Absoluto Médio

MAPE Erro Percentual Absoluto Médio

FFBS *Forward Filtering Backward Sampling*

MCMC Monte Carlo via Cadeias de Markov

MP Modelo Polinomial de 1° Ordem

MPS Modelo Polinomial de 1° Ordem com Sazonalidade

IC Intervalo de credibilidade

1 Introdução

A inflação é um termo que significa o aumento generalizado dos preços de bens e serviços em uma economia (BACEN, 2021). Com isso, para representá-la foram criados os índices de preços, que são calculados a partir de uma média ponderada de diversos preços resumidos em um número.

Os índices de preços podem ser construídos a partir de várias cestas de bens e serviços. Dependendo do objetivo, pode-se construir índices de preços ao consumidor, índices de preços ao produtor, índices de preços da construção civil, etc. Além disso, podem ser feitos recortes destes índices, visando uma melhor representação de um subgrupo específico. Por exemplo, no caso do Índice de Preço ao Consumidor (IPC - FGV IBRE) (IBRE, 2021a), há um recorte de famílias com renda mensal entre 1 e 2,5 salários mínimos, denominado, Índice de Preço ao Consumidor - C1 (IPC-C1 - FGV IBRE) (IBRE, 2021b). Desta forma, a percepção da população com relação aos resultados apurados nos índices é distinta.

No Brasil, o índice de preço referência do sistema de metas para a inflação no Brasil é o Índice Nacional de Preços ao Consumidor Amplo (IPCA - IBGE), que aponta a variação do custo de vida médio de famílias com renda mensal de 1 e 40 salários mínimos em 16 áreas geográficas: regiões metropolitanas de Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre, além do Distrito Federal e dos municípios de Goiânia, Campo Grande, Rio Branco, São Luís e Aracaju (IBGE, 2021c).

O IPCA é constituído por uma estrutura hierárquica, construída com base na Pesquisa de Orçamentos Familiares (POF - IBGE), esta, por sua vez, avalia as estruturas de gastos das famílias, assim, proporcionando a criação de um perfil do consumo familiar (IBGE, 2021a). Esse perfil é resumido em um conjunto de subitens consumidos pelas famílias, partindo de alimentos, eletrodomésticos, móveis residenciais, medicamentos, automóveis, combustíveis, passando por serviços como salão de beleza, reparos residenciais, consertos de itens de casa como eletrodomésticos e móveis, *streaming*, TV por assinatura, até tarifas

de eletricidade residencial, água e esgoto, aluguel residencial, entre outros.

A Pesquisa de Orçamentos Familiares (POF) é realizada por amostragem usando como unidade de investigação o domicílio, com isso os resultados dessa pesquisa permitem também estudar o consumo de famílias a partir de classes de rendimentos, as diferenças entre as classes de renda, as disparidades regionais, as situações rural e urbana e a dimensão do mercado consumidor para grupos de produtos e serviços (IBGE, 2021a).

Como o IPCA é um recorte das famílias com renda mensal de 1 a 40 salários mínimos, essa cesta é obtida a partir de um recorte da POF, e com isso, é construída uma estrutura para cada capital componente do índice, distribuindo um peso para cada subitem, por exemplo, gasolina, etanol, tarifas públicas. Desta forma, estes subitens são agregados por setor em itens, e após estes em subgrupos e em grupos, sendo realizado, por fim, a média ponderada de todos os grupos no índice geral. Para melhor entendimento, segue a Figura 1.

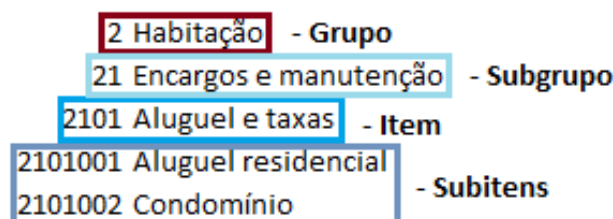


Figura 1: Estrutura IPCA

A variação destes subitens é obtida através da média da variação dos preços de cada subitem, por exemplo, em uma capital são coletados os preços de automóvel novo no período de coleta vigente do IPCA. Estes são comparados aos valores coletados no período anterior, assim obtém-se a variação desse subitem nessa capital.

Dado que o IPCA é constituído de subitens como tarifas, serviços, alimentos, produtos de linha branca e marrom, entre outros, a sua previsão se tornou muito relevante para economia, principalmente num momento de incertezas como o atual devido à pandemia causada pelo surgimento do vírus SARS-CoV-2. O Brasil sofreu bastante com a paralisação das atividades devido ao *lockdown* e, conseqüentemente, a pouca mobilidade da população trouxe uma queda na produção econômica do país (VALOR, 2021) e com a alta do dólar trouxe um aumento de preço em diversos produtos componentes do IPCA (CNN, 2021).

Alguns artigos abordam a previsão do IPCA por meio de vários métodos como a me-

todologia Box-Jenkins (JUNIOR, 2014), árvores de regressão (MATTOS; FIGUEIREDO, 2017) e métodos de fatores estimados combinados por meio da seleção de variáveis, de ponderação das previsões ou utilizando técnicas de encolhimento (BARBOSA; FERREIRA; SILVA, 2020).

A previsão do IPCA foi realizada por meio da metodologia de Box-Jenkins (ARIMA) conforme feito em Junior (2014), optando por não se utilizar covariáveis. Com isso, o autor testou se a série do IPCA possuía estacionariedade, obtendo um resultado negativo e, portanto, a partir das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) identificou o modelo ARIMA(14,0,4) como o melhor método segundo os critérios de Akaike, Schwarz, Hannan-Quinn e o R^2 . Na previsão do IPCA, este modelo obteve resultados moderados.

Barbosa, Ferreira e Silva (2020) utilizaram técnicas de aprendizado de máquinas para dividir a série temporal do IPCA em amostra treino e teste, respectivamente, para estimar os modelos e avaliar a previsão um tempo a frente. Além disso, a cada previsão realizada na série de teste, esta foi incorporada na série de treino e assim realizada uma nova previsão. Para estimar o modelo foram usadas 117 variáveis macroeconômicas como covariáveis visando explicar o comportamento do IPCA. A partir disso testaram a estacionariedade de todas e assim aplicaram técnicas de aprendizado de máquinas como: *benchmark*, *shrinkage*, supervisão de fatores, ponderação de previsões e seleção de previsores. Após a aplicação das técnicas de aprendizado de máquinas, foram testados algumas combinações destas técnicas com modelos como: modelo autoregressivo de ordem 4, de fator autoregressivo aumentado, fatorial, de ponderação *bagging*, *least angle regression*, entre outros. A partir dos resultados destas combinações de técnicas de aprendizado de máquinas com os modelos acima, concluiu-se que a combinação destas técnicas e a supervisão fatorial produzem melhores previsões que modelos que não utilizam fatores como: modelos fatoriais com ou sem supervisão e modelos que utilizam apenas aprendizado de máquinas sem supervisão dos fatores.

Outro exemplo de modelo para previsão do IPCA, foi apresentado em Mattos e Figueiredo (2017). Neste trabalho, os autores utilizaram modelos de classificação por duas técnicas: árvores de decisão e árvores de regressão. No primeiro momento, foi realizada a separação dos valores do IPCA em classes pelo método *K-Means*, sendo o número de grupos classificado por árvores de decisão. Com isso, definiu-se em qual grupo o valor do IPCA no próximo mês se enquadrava e foi realizada a previsão desse mês pelo modelo de Árvore de Regressão, considerando variáveis selecionadas a partir do método DTW.

Assim, concluiu-se que este modelo obteve bons resultados para a previsão do índice.

Portes (2017) aplicou a distribuição Generalizada de Valores Extremos (GEV) para modelar os níveis máximos de IPCA, escolhendo os quadrimestres, com o intuito de comparar a análise da previsão com os demonstrativos de cumprimento das metas fiscais. A partir dos resultados, chegou à conclusão de que o modelo que utilizou o dobro da matriz de covariâncias e variâncias e com a informação *a priori* sendo a média do IPC máximo quadrimestral da China, Rússia e Índia para construção da distribuição *a priori* Normal Trivariada, obteve melhor precisão e acurácia na predição do IPCA.

Diante da revisão de literatura feita, a grande maioria dos modelos fazem inferência sob a perspectiva frequentista. Com isso, a motivação deste trabalho está em utilizar o escopo de modelos lineares dinâmicos (WEST; HARRISON, 1997), apropriado para dados de séries temporais, sob a perspectiva bayesiana como mais uma ferramenta para modelagem e para fazer previsão do IPCA.

O trabalho está organizado da seguinte forma: no Capítulo 2 serão apresentados os conceitos de modelos lineares dinâmicos e os dois modelos que serão utilizados neste trabalho. Com isso, será apresentada uma revisão dos conceitos de inferência bayesiana e dos métodos de Monte Carlo via cadeias de Markov que serão aplicados na atualização e, conseqüentemente, na previsão desses modelos. No Capítulo 3, será realizada análise exploratória dos dados. Em seguida, serão apresentados os resultados dos modelos lineares dinâmicos apresentados no Capítulo 2. Por fim, no Capítulo 4 serão apresentadas as conclusões a partir dos resultados e possíveis trabalhos futuros.

2 Materiais e Métodos

Este trabalho tem por objetivo propor a previsão da série temporal do IPCA usando modelos lineares dinâmicos. Com isso, na Seção 2.1 será detalhada a série utilizada neste trabalho, a Seção 2.2 apresentará os modelos lineares dinâmicos que serão estudados, e, em seguida, na Seção 2.3 serão apresentados os conceitos de inferência bayesiana usados para a atualização da distribuição dos parâmetros com base na informação inicial com o intuito de prever o IPCA.

2.1 Material

Neste trabalho, será utilizada a série temporal do índice geral do IPCA, que pode ser obtida através do Sistema IBGE de Recuperação Automática (SIDRA) (IBGE, 2021b) (IBGE, 2019) (IBGE, 2011). Esses dados contém a variação mensal do IPCA de janeiro de 2012 à dezembro de 2021, a escolha deste período está ligada ao fato de serem as duas estruturas mais recentes do índice, vindas das últimas duas POF. Além disso, na aplicação de todos os modelos será utilizada a série temporal do índice geral do IPCA como a variável de interesse.

Na próxima seção será apresentada a classe de modelos lineares dinâmicos de forma geral para que sejam realizadas a modelagem e a previsão do IPCA.

2.2 Modelos Lineares Dinâmicos

Na Estatística há o interesse de se explicar uma variável de desfecho em relação a outras variáveis, como por exemplo, explicar o comportamento do tempo gasto para percorrer uma quilometragem e o consumo de oxigênio do corpo, assim a variável que se busca explicar é chamada de variável resposta e as demais de variáveis de explicativas ou covariáveis. Para isso, normalmente são utilizados modelos estatísticos para verificar essa relação, como, por exemplo, os modelos lineares. Além disso, estes modelos podem ser

utilizados para previsão do comportamento da variável resposta.

O IPCA, por ser uma série temporal, tem por característica a dependência do seu valor no mês atual de meses anteriores e, portanto, a utilização dos modelos lineares não é adequada, pois um dos seus pressupostos é a independência das observações. Por se tratar de dados temporais, serão utilizados os modelos lineares dinâmicos, que são adequados para este contexto.

Para a análise de dados de séries temporais, podem ser usados os modelos de espaço de estado que fornecem uma classe bem vasta de modelos. Estes são comumente aplicados em áreas como econometria, genética, processamento de sinais e dinâmica populacional. Os modelos lineares dinâmicos são uma classe particular dos modelos de espaço de estado, podendo ser definidos como clones de modelos que possuem parâmetros que variam no tempo, sendo muito úteis para lidar com séries temporais. Desta forma, esses modelos são flexíveis o suficiente para captar os principais recursos de uma ampla variedade de dados diferentes (PETRIS, 2010).

Considere uma sequência de variáveis aleatórias $\{Y_1, Y_2, \dots, Y_t\}$, em que Y_t denota o t -ésimo valor da série temporal do IPCA. Os **modelos lineares dinâmicos (MLD)**, para uma série temporal univariada, são caracterizados por um conjunto de quádruplos $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$, sendo definido da seguinte forma (WEST; HARRISON, 1997).

- **Equação de observação:** $Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N[0, V_t], \quad t = 1, \dots, T,$
- **Equação de evolução:** $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t], \quad t = 1, \dots, T,$
- **Informação inicial:** $(\boldsymbol{\theta}_0 | D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0],$

em que, $\boldsymbol{\theta}_t$ é um vetor p -dimensional dos parâmetros no tempo t , \mathbf{F}_t é um vetor $p \times 1$ de variáveis explicativas, cujos elementos são conhecidos, \mathbf{G}_t é uma matriz quadrada de ordem p que descreve a evolução dos parâmetros de estado no tempo, sendo conhecida como matriz de evolução, ϵ_t e $\boldsymbol{\omega}_t$, são respectivamente, o erro observado e o erro de evolução. Além disso, são independentes da informação inicial e para todo t e s com $t \neq s$, as combinações de ϵ_t e ϵ_s , $\boldsymbol{\omega}_t$ e $\boldsymbol{\omega}_s$ e ϵ_t e $\boldsymbol{\omega}_s$ são independentes.

Na equação de evolução, $\boldsymbol{\omega}_t$ segue uma distribuição normal multivariada e descreve as alterações nos elementos do vetor de parâmetros de estado no tempo $t = 2, \dots, T$. Além disso, o vetor de média zero representa a crença de que $\boldsymbol{\theta}_t$ seja constante ao longo do tempo t . Desta forma, a matriz de covariâncias \mathbf{W}_t reflete a extensão das mudanças em

$\boldsymbol{\theta}_t$, e conseqüentemente, o período de tempo em que a suposição de constância local de $\boldsymbol{\theta}_t$ é aceita.

As componentes \mathbf{F}_t e \mathbf{G}_t são escolhidas de acordo com o modelo e a natureza dos dados estudados. Em alguns casos, como, por exemplo, o modelo polinomial de 1° ordem, a matriz \mathbf{G}_t e o vetor \mathbf{F}_t são definidos como $\mathbf{G}_t = 1$ e $\mathbf{F}_t = 1$. Assim, o modelo é definido pela quádrupla $\{1, 1, V_t, \mathbf{W}_t\}$.

Na **equação de observação** é fornecida a relação entre o valor do IPCA no tempo t , Y_t e o vetor de parâmetros de estado p -dimensional no tempo t , $\boldsymbol{\theta}_t$. Além disso, os valores de Y_t são independentes dado $\boldsymbol{\theta}_t$ e Y_t depende apenas de $\boldsymbol{\theta}_t$, para $t = 1, \dots, T$. Na **equação de evolução**, os parâmetros de estado são atualizados por meio do tempo.

Assumindo que a variância observacional V_t é conhecida para cada tempo t , o precedente observacional e equação de evolução também podem ser expressas para cada $t = 1, 2, \dots, T$, como:

$$(Y_t | \boldsymbol{\theta}_t) \sim N[\mathbf{F}_t' \boldsymbol{\theta}_t, V_t]. \quad (2.1)$$

$$(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) \sim N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t]. \quad (2.2)$$

A **informação inicial** possui a distribuição *a priori* dos dados, sendo a representação probabilística da previsão inicial de $\boldsymbol{\theta}_t$ no tempo $t = 0$, com vetor de médias \mathbf{m}_0 conhecida, sendo uma crença inicial deste nível e a matriz de covariâncias \mathbf{C}_0 , conhecida, sendo a incerteza associada ao mesmo. O termo D_0 , representa toda a informação inicial relevante disponível sobre o IPCA, que será usada para formar visões iniciais sobre os dados futuros do índice.

O modelo definido acima considera tanto as variâncias observacionais, V_t , como a matriz \mathbf{W}_t conhecidas ao longo do tempo. No entanto, a variância do erro observacional é geralmente desconhecida e em relação a variância do erro de evolução é consideravelmente grande. Sendo assim, esta é considerada a principal fonte de incerteza da previsão, com isso serão apresentados procedimentos sob a perspectiva bayesiana apropriados para modelos com as variâncias dos erros das equações de observação e evolução consideradas desconhecidas.

Na prática não é tão razoável assumir a variância observacional conhecida, portanto há formas de se ajustar o modelo sem fazer esta suposição. Neste trabalho, o interesse reside em modelos lineares dinâmicos com variância observacional V_t e a matriz de covariâncias \mathbf{W}_t dos estados desconhecidas. Para tal, será feita duas simplificações no modelo. A

primeira será considerar que $V_t = V$ para todo tempo t , sendo $\tau = 1/V$ a precisão da observação. A segunda será considerar que a matriz de covariâncias do vetor $\boldsymbol{\theta}_t$ é constante e uma matriz diagonal.

A imposição da segunda simplificação, se justifica pelo pacote que será utilizado para modelagem e previsão neste trabalho. Deste modo, serão consideradas as seguintes propriedades dos modelos: (i) as variâncias observacional V e de evolução \mathbf{W} , serão desconhecidas e constantes; (ii) \mathbf{W} será considerada uma matriz diagonal e; (iii) as distribuições *à priori* das variâncias desconhecidas são gama inversas independentes (PETRIS, 2011).

Neste trabalho será assumido, que as observações são univariadas, dado que será usada apenas a série do índice geral do IPCA. Além disso, a matriz diagonal \mathbf{W} será definida como $\mathbf{W} = \text{diag}(\tau_{\theta,1}^{-1}, \tau_{\theta,2}^{-1}, \dots, \tau_{\theta,p}^{-1})$, assim cada parâmetro desconhecido tem a sua respectiva precisão $\tau_{\theta,i}$, onde $i = 1, \dots, p$. Além disso, as precisões são fixas ao longo do tempo, assim, \mathbf{W} também é fixa ao longo do tempo (PETRIS; PETRONE; CAMPAGNOLI, 2007).

Como as variâncias V e W serão assumidas como desconhecidas, estas serão estimadas pelo algoritmo de Gibbs presente na Seção 2.3.2, que por sua vez, extrai da distribuição condicional completa dos estados e das distribuições condicionais completas de $\{\tau, \tau_{\theta,1}, \dots, \tau_{\theta,p}\}$. Além disso, pelo fato de usarmos o pacote *dln* do R, a amostragem dos estados será feita usando o algoritmo *Forward Filtering Backward Sampling* (FFBS) criado por Carter e Kohn (1994), Frühwirth-Schnatter (1994) e Shephard (1994). Desta forma, será utilizado o modelo d-gama inversa *a priori* para $t \geq 1$ o modelo é definido como:

- **Equação de observação:** $Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N[0, V], \quad V = \tau^{-1}.$
- **Equação de evolução:** $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[0, \mathbf{W}], \quad \mathbf{W} = \text{diag}(\tau_{\theta,1}^{-1}, \tau_{\theta,2}^{-1}, \dots, \tau_{\theta,p}^{-1}).$
- **Informação inicial:**

$$(\boldsymbol{\theta}_0 | D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0],$$

$$(\tau | D_0) \sim G\left[\frac{n_0^2}{2}, \frac{n_0}{2}\right],$$

$$\tau_{\theta,i} \sim G(a_{\theta,i}, b_{\theta,i}), \quad i = 1, \dots, p.$$

As quantidades iniciais \mathbf{m}_0 , \mathbf{C}_0 , n_0 , $a_{\theta,0}$ e $b_{\theta,0}$ são especificadas, assim como a matriz \mathbf{G}_t e o vetor \mathbf{F}_t . Além disso, o vetor de parâmetros de estado $\boldsymbol{\theta}_0$ da **informação inicial** segue uma distribuição normal gama, e a precisão $\tau = 1/V$ segue uma distribuição gama e conseqüentemente a variância observacional possui uma distribuição gama inversa.

Neste trabalho, Y_t representa o valor do IPCA no tempo t que será utilizado em diferentes modelos lineares dinâmicos que serão apresentados nas próximas seções deste capítulo, eles são: (i) modelo polinomial de 1° ordem (MP) e (ii) modelo polinomial de 1° ordem com sazonalidade (MPS).

2.2.1 Modelo polinomial de 1° ordem

Inicialmente, será utilizado o **modelo polinomial de 1° ordem (MP)** que é caracterizado pela evolução temporal do nível ser modelada como um passeio aleatório simples, ou média localmente constante. Além disso, este modelo é um caso particular dos modelos lineares dinâmicos sendo caracterizado pela quádrupla $\{1, 1, V_t, W_t\}$ (WEST; HARRISON, 1997). Assim, neste trabalho serão assumidos $V_t = V$ e $W_t = W$. Além disso, é considerado que $V = \tau^{-1}$ e $W = \tau_{\theta,1}^{-1}$. Com isso, o modelo é caracterizado por $\{1, 1, V, W\}$ e definido por:

$$\begin{aligned}
 Y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N[0, V], & t &= 1, \dots, T, & V &= \tau^{-1} \\
 \mu_t &= \mu_{t-1} + \omega_t, & \omega_t &\sim N[0, W], & t &= 1, \dots, T, & V &= \tau_{\mu}^{-1} \\
 (\mu_0 | D_0) &\sim N[m_0, C_0], \\
 (\tau | D_0) &\sim G\left[\frac{n_0^2}{2}, \frac{n_0}{2}\right], \\
 \tau_{\mu} &\sim G(a_{\theta,i}, b_{\theta,i}),
 \end{aligned} \tag{2.3}$$

em que μ_t é o nível da série temporal do IPCA no tempo t , ϵ_t e ω_t , são respectivamente, o erro observado e o erro de evolução. Além disso, neste modelo $\theta_t = \mu_t$, $F_t = 1$ e $G_t = 1$.

2.2.2 Modelo polinomial de 1° ordem com sazonalidade

Como o IPCA é constituído de uma cesta de produtos, em que alguns desses possuem sazonalidade, é interessante testar um modelo polinomial de 1° ordem com sazonalidade, com o objetivo de quantificar o impacto deste termo no ajuste e na previsão do IPCA. Com isso, será utilizado o **modelo polinomial de 1° ordem com sazonalidade (MPS)**

caracterizado pelo conjunto $\{\mathbf{F}_t, \mathbf{G}_t, V, \mathbf{W}\}$. Assim, o modelo é definido como:

$$\begin{aligned}
Y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N[0, V], \quad t = 1, \dots, T, \quad V = \tau^{-1} \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[0, \mathbf{W}], \quad t = 1, \dots, T, \quad \mathbf{W} = \text{diag}(\tau_{\theta,1}^{-1}, \tau_{\theta,2}^{-1}, \dots, \tau_{\theta,p}^{-1}) \\
(\boldsymbol{\theta}_0 | D_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0], \\
(\tau | D_0) &\sim G\left[\frac{n_0^2}{2}, \frac{n_0}{2}\right], \\
\tau_{\theta,i} &\sim G(a_{\theta,i}, b_{\theta,i}), \quad i = 1, \dots, p,
\end{aligned} \tag{2.4}$$

em que $\boldsymbol{\theta}_t$ é um vetor de parâmetros de estado p -dimensional no tempo t , \mathbf{F}_t é um vetor de dimensão $p \times 1$, \mathbf{G}_t é uma matriz quadrada de ordem p que descreve a evolução dos parâmetros de estado no tempo, ϵ_t e $\boldsymbol{\omega}_t$, são respectivamente, o erro observado e o erro de evolução. Além disso, são independentes da informação inicial e para todo t e s com $t \neq s$, as combinações de ϵ_t e ϵ_s , $\boldsymbol{\omega}_t$ e $\boldsymbol{\omega}_s$ e ϵ_t e $\boldsymbol{\omega}_s$ são independentes.

Será considerado uma sazonalidade anual do IPCA. Sendo assim, o vetor de parâmetros de estado $\boldsymbol{\theta}_t$ é composto de um componente de tendência e efeitos sazonais, $\boldsymbol{\gamma}_{12m}$.

Neste modelo, $\boldsymbol{\theta}_t = (\mu_t, \gamma_1, \dots, \gamma_{11})^T$ com dimensão de 12×1 , $\mathbf{F}_t = (1, 1, 0, \dots, 0)^T$ com dimensão 12×1 , $\mathbf{W} = \text{diag}(\tau_{\theta,1}^{-1}, \tau_{\theta,2}^{-1}, \dots, \tau_{\theta,p}^{-1})$ e;

$$\mathbf{G}_t = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}_{12 \times 12}, \tag{2.5}$$

de forma alternativa, o modelo pode ser escrito da seguinte forma:

$$\begin{cases} Y_t = \mu_t + \gamma_{12m} + \epsilon_t, & \epsilon_t \sim N(0, V), \quad t = 1, \dots, T \\ \mu_t = \mu_{t-1} + \omega_{1,t} \\ \gamma_{1,t} = -\gamma_{2,t-1} - \gamma_{3,t-1} - \dots - \gamma_{12,t-1} + \omega_{2,t} \\ \gamma_{2,t} = \gamma_{2,t-1} + \omega_{3,t} \\ \vdots \\ \gamma_{11,t} = \gamma_{11,t-1} + \omega_{12,t} \end{cases} \tag{2.6}$$

Para realizar a seleção correta dos efeitos sazonais na equação de observação será mantido o vetor \mathbf{F}_t e assim será reordenada o vetor de parâmetros de estado $\boldsymbol{\theta}_t$, em que o primeiro parâmetro μ_t será fixo e os parâmetros γ_i que serão realocados de posição. Para a equação de evolução será seguida a mesma ideia, onde a matriz \mathbf{G}_t será fixa e o vetor $\boldsymbol{\theta}_t$ será reordenado.

Com a definição dos modelos, o passo seguinte será realizar a inferência sobre os mesmos, para isso, será utilizada uma abordagem sob a ótica bayesiana. Assim, na próxima Seção serão apresentadas algumas definições essenciais e como será aplicada essa perspectiva nos modelos.

2.3 Inferência Bayesiana

Na inferência o objetivo está em fazer afirmações sobre quantidades de interesse $\boldsymbol{\phi} = (\phi_1, \dots, \phi_t)^T$, assim, pela perspectiva clássica, $\boldsymbol{\phi}$ é obtido a partir dos dados, sendo uma constante fixa. Há diversas formas de se fazer inferência estatística, as mais conhecidas são a frequentista (ou clássica) e a bayesiana. Pelo perspectiva bayesiana, combinam-se informações históricas sobre $\boldsymbol{\phi}$ e os dados. Com isso, nos modelos apresentados neste trabalho, $\boldsymbol{\phi}$ é tratado como um vetor aleatório.

Na inferência bayesiana a distribuição *a priori* contém toda a informação inicial sobre $\boldsymbol{\phi}$. Seja $\mathbf{y} = (y_1, \dots, y_t)^T$ a quantidade observável, toda a informação inicial sobre esta quantidade pode ser resumida, probabilisticamente por meio de $p(\boldsymbol{\phi})$. Com isso, a revisão da informação inicial é realizada utilizando o Teorema de Bayes.

$$p(\boldsymbol{\phi}|\mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\phi})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\phi})p(\boldsymbol{\phi})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\phi})p(\boldsymbol{\phi})}{\int p(\mathbf{y}|\boldsymbol{\phi})p(\boldsymbol{\phi})d\boldsymbol{\phi}}.$$

Ambas funções, $p(\boldsymbol{\phi})$ e $p(\mathbf{y}|\boldsymbol{\phi}) = L(\boldsymbol{\phi}; \mathbf{y})$ são combinadas pelo Teorema de Bayes resultando a distribuição *a posteriori* de $\boldsymbol{\phi}$, $p(\boldsymbol{\phi}|\mathbf{y})$. Desta forma, o Teorema de Bayes é apresentado como:

$$p(\boldsymbol{\phi}|\mathbf{y}) \propto L(\boldsymbol{\phi}; \mathbf{y})p(\boldsymbol{\phi}).$$

Portanto, o resultado desta combinação é a distribuição *a posteriori*, que por sua vez, é utilizada para toda a análise. Em muitos casos, estas distribuições são desconhecidas e difíceis de trabalhar. Uma alternativa bastante utilizada é recorrer aos métodos de Monte Carlo via cadeias de Markov e, mais especificamente, ao amostrador de Gibbs e realizar a amostragem através de distribuições conhecidas como condicionais completas *a posteriori*.

Estas distribuições são obtidas a partir da distribuição *a posteriori*. Assim, para cada parâmetro ϕ é calculada a condicional completa, considerando os demais parâmetros como constantes fixas.

2.3.1 Filtro de Kalman

Os modelos lineares de regressão dinâmica podem, de forma geral, ser representados em espaço de estados das observações e do estado do sistema. Desta forma, com os modelos definidos neste trabalho podem utilizar o filtro de Kalman para estimar os estados das observações.

A inferência bayesiana é uma aplicação estatística mais antiga que a usada atualmente, mas foi esquecida por possuir cálculos muito complexos de serem feitos algebricamente. Dado o surgimento dos computadores e a sua capacidade de realizar esses cálculos, foram criados métodos como o de filtro de Kalman, criado por Kalman (1960).

O filtro de Kalman refere-se a um conjunto de procedimentos recursivos para estimação em modelos dinâmicos. Considerando os modelos lineares dinâmicos definidos na Seção 2.2, a partir da informação inicial $(\theta_0|D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$ temos:

- (a) A distribuição *a priori*: $(\theta_t|D_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t)$, em que $\mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1}$ e $\mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$.
- (b) A previsão um passo a frente *a priori*: $(Y_t|D_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$, em que $\mathbf{f}_t = \mathbf{F}_t\mathbf{a}_t$ e $\mathbf{Q}_t = \mathbf{F}_t\mathbf{R}_t\mathbf{F}_t' + \mathbf{V}_t$
- (c) A distribuição *a posteriori*: $(\theta_t|D_t) \sim N(\mathbf{m}_t, \mathbf{C}_t)$, em que $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t'\mathbf{Q}_t^{-1}\mathbf{e}_t$ e $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t'\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{R}_t$, onde $\mathbf{e}_t = Y_t - \mathbf{f}_t$ é o erro de previsão.

Assim, o filtro de Kalman permite calcular as distribuições preditivas e filtradas de forma recursiva assim que novas observações se tornam disponíveis.

Para que sejam realizados os cálculos da atualização da informação inicial para a previsão do IPCA, será utilizado o algoritmo de *Forward Filtering Backward Sampling* (FFBS) e para a variância V e para os termos da diagonal da matriz \mathbf{W} será utilizado o algoritmo de Gibbs apresentadas na Seção 2.3.2, presentes na função `dlnGibbsDIG` do pacote `dln`.

2.3.2 Monte Carlo via Cadeias de Markov

A ideia do método Monte Carlo via Cadeias de Markov (MCMC) é obter uma amostra aleatória da distribuição e, com isso, simular um passeio aleatório no espaço paramétrico convergindo para uma distribuição de equilíbrio, podendo essa ser, a distribuição *a posteriori* (NOGUEIRA; FERREIRA; SAFADI, 2004). Desta forma, para realizar esse processo são utilizados algoritmos iterativos, que por sua vez, geram cadeias de Markov para cada parâmetro, utilizando as suas distribuições condicionais completas como núcleo de transição, e assim, convergindo para a distribuição *a posteriori*. Alguns métodos MCMC são o algoritmo de Metropolis-Hastings e amostrador de Gibbs.

Em Geman e Geman (1984), o amostrador de Gibbs é definido como um esquema amostral de cadeias de Markov. Considere $\boldsymbol{\tau} = (\tau_1, \dots, \tau_T)'$, assim são definidos valores para cada τ_t , $t = 1, \dots, T$ assim são definidos valores para cada componente de $\boldsymbol{\tau}$. Com isso, a partir das condicionais completas, são gerados novos valores considerando os valores definidos inicialmente. Este processo é repetido até que as cadeias convirjam para a distribuição *a posteriori*. Para ser utilizado este método é preciso que se seja capaz de gerar os valores da condicional completa, pois nem sempre se é capaz de gerar esses valores. Para simplificação, a seguir será apresentado o passo a passo do amostrador de Gibbs:

- 1) Selecionando inicialmente o valor de $i = 0$;
- 2) Determinado os valores iniciais de $\boldsymbol{\tau}^{(0)}$;
- 3) Com isso, será gerado um novo valor para $\boldsymbol{\tau}^{(i)}$ a partir do anterior $\boldsymbol{\tau}_{t-1}^{(i)}$ com a geração iterativa dos valores da seguinte forma:

$$\begin{aligned} \tau_1^{(i)} &\sim p\left(\tau_1 | \tau_2^{(i-1)}, \tau_3^{(i-1)}, \dots, \tau_T^{(i-1)}, \mathbf{y}\right) \\ \tau_2^{(i)} &\sim p\left(\tau_2 | \tau_1^{(i)}, \tau_3^{(i-1)}, \dots, \tau_T^{(i-1)}, \mathbf{y}\right) \\ &\dots \\ \tau_T^{(i)} &\sim p\left(\tau_T | \tau_1^{(i)}, \tau_2^{(i)}, \dots, \tau_{T-1}^{(i)}, \mathbf{y}\right) \end{aligned}$$

- 4) Por fim, i é incrementado para $i + 1$ e retorne ao passo 3.

Outro método de MCMC, é o algoritmo de Metropolis-Hastings proposto por Metropolis et al. (1953) e aprimorado por Hastings (1970). Neste método, a partir de uma distribuição proposta, são gerados valores e estes são aceitos ou não na cadeia por uma certa probabilidade. Para realizar esse processo, são definidos os valores iniciais de $\boldsymbol{\tau}$, e

com isso, são gerados os valores e comparados a probabilidade para serem aceitos ou não. A partir disso, se forem aceitos, o processo é o mesmo do amostrador de Gibbs, caso não, o valor anterior é repetido e continuam as iterações até que as cadeias convirjam para a distribuição *a posteriori*. Para simplificar, a seguir será apresentado o passo a passo do algoritmo de Metropolis-Hastings:

- 1) Selecionando inicialmente o valor de $i = 0$ e $j = 1$;
- 2) Determinando os valores iniciais de $\boldsymbol{\tau}^{(0)}$ e i é incrementado para $i + 1$;
- 3) São gerados valores a partir de uma distribuição proposta $q(\boldsymbol{\epsilon}^{(j)}|\boldsymbol{\tau}^{(i-1)})$, e com isso, calculada a probabilidade de aceitação:

$$\beta(\boldsymbol{\tau}^{(i-1)}|\boldsymbol{\epsilon}^{(j)}) = \min \left(1, \frac{p(\boldsymbol{\epsilon}^{(j)}|\mathbf{y})q(\boldsymbol{\tau}^{(i-1)}|\boldsymbol{\epsilon}^{(j)})}{p(\boldsymbol{\tau}^{(j-1)}|\mathbf{y})q(\boldsymbol{\epsilon}^{(j)}|\boldsymbol{\tau}^{(i-1)})} \right)$$

- 4) Após esse cálculo, será gerado $v \sim U(0, 1)$ para que seja aceito ou não. Caso $v \leq 1/\beta$ esse valor será aceito e $\boldsymbol{\tau}^{(i)} = \boldsymbol{\epsilon}^{(j)}$ e o contador será acrescido de 1, caso contrário, $\boldsymbol{\tau}^{(t)} = \boldsymbol{\tau}^{(t-1)}$ e o contador se mantém o mesmo. Retornando ao passo 2 até que convirja.

Para que o funcionamento do algoritmo de Metropolis-Hastings seja correto, deve-se propor uma distribuição adequada. No caso em que a distribuição proposta for inadequada, a probabilidade de aceitação, que considera a distribuição e o comportamento da condicional completa, pode ser consideravelmente pequena, o que, levaria a uma maior rejeição dos valores, e conseqüentemente, uma dificuldade na convergência das cadeias.

2.3.3 Previsão de distribuições *a posteriori*

A ideia da perspectiva bayesiana na previsão do IPCA é determinar uma distribuição *a posteriori* para o vetor de parâmetros de estado $\boldsymbol{\theta}_t$ com base em informações iniciais sobre os dados, sendo essas baseadas em resultados históricos ou em convicções. Com isso, para atualizar as informações iniciais utiliza-se o teorema de Bayes na forma de probabilidades *a posteriori*.

Para realizar a amostragem do θ_t será utilizado o algoritmo *Forward Filtering Backward Sampling* (FFBS) (CARTER; KOHN, 1994) (FRÜHWIRTH-SCHNATTER, 1994) (SHEPHARD, 1994). Este algoritmo pode ser considerado essencialmente uma versão de simulação das recursões de suavização. Desta forma, o algoritmo FFBS pode ser descrito como:

1. Executa o filtro de Kalman;
2. Desenha θ_t de uma distribuição $N(m_t, C_t)$;
3. Para $t = T - 1, T - 2, \dots, 0$, desenha θ_t de uma distribuição $N(h_t, H_t)$,

em que $h_t = m_t + C_t G_{t+1} R_{t+1}^{-1} (\theta_{t+1} - a_{t+1})$, $H_t = C_t - C_t G_{t+1} R_{t+1}^{-1} G_{t+1} C_t$ e $a_{t+1} = G_t m_t$.

Em Petris, Petrone e Campagnoli (2007) se encontram com mais detalhes os esquemas de atualização pelo algoritmo FFBS das Seções 2.2.1 e 2.2.2.

Com a inferência sobre os parâmetros dos modelos apresentada, serão abordados os procedimentos para a realização da previsão k -passos à frente de cada modelo apresentado na Seção 2.2.

2.3.4 Previsão k -passos à frente

O objetivo deste trabalho é a modelagem e a previsão do IPCA a partir dos modelos apresentados anteriormente. Desta forma, para a previsão de k -passos à frente nos modelos apresentados nas Seções 2.2.1 e 2.2.2, será utilizada a função *dlnForecast* presente no pacote *dln*, que utiliza uma distribuição de previsão no tempo t , dada por, $(Y_{t+k}|D_t)$, para $k \geq 1$. Com isso, a previsão de k -passos à frente segue a seguinte distribuição:

$$(Y_{t+k}|D_t) \sim N[f_t(k), Q_t(k)],$$

em que V_t é substituído pela estimativa S_t , $f_t(k) = \mathbf{F}'_{t+k} \mathbf{a}_t(k)$, $Q_t(k) = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + S_t$, $\mathbf{a}_t(k) = \mathbf{G}_{t+k} \mathbf{a}_t(k-1)$, $\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}'_{t+k} + \mathbf{W}_{t+k}$, onde $\mathbf{a}_t(0) = \mathbf{m}_t$, $\mathbf{R}_t(0) = \mathbf{C}_t$.

2.4 Medidas de Comparação entre os Modelos

Nesta monografia, serão utilizados modelos para séries temporais com o objetivo de prever o IPCA. Assim, após a modelagem, será verificado qual desses modelos obteve o melhor resultado, com isso, serão apresentados a seguir as medidas de qualidade que servirão de métricas nesta avaliação.

2.4.1 Medidas de qualidade do ajuste

Normalmente na avaliação dos modelos, são utilizadas técnicas científicas que permitem obter algumas medidas, estas usadas para a escolha do melhor modelo.

Para avaliar os modelos usados neste trabalho será utilizada como medida de qualidade do ajuste, o erro percentual absoluto médio (MAPE). Os modelos que possuem menores Erro Percentual Absoluto Médio (MAPE), serão considerados melhores ajustados.

Proposto em Lewis (1997), o erro percentual absoluto médio (MAPE) é uma das medidas mais utilizadas para verificar o ajuste e a previsão do modelo, por ser de fácil interpretação e compreensão. Esta medida é expressa em percentual do tamanho médio do erro. Desta forma, o MAPE pode ser calculado através da seguinte expressão:

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{Y}_t - Y_t|}{Y_t}$$

em que Y_t é o valor do IPCA no tempo t , \hat{Y}_t é o valor estimado pelo modelo para o IPCA e T é o tamanho da série temporal.

Outra medida de qualidade do ajuste que será utilizada para avaliar os modelos presentes neste trabalho será o desvio absoluto médio (MAD). Esta medida é o somatório da média absoluta entre o valor estimado pelo modelo e o valor real da série, e é dada pela seguinte equação:

$$MAD = \frac{1}{T} \sum_{t=1}^T |\hat{Y}_t - Y_t|$$

em que Y_t é o valor do IPCA no tempo t , \hat{Y}_t é o valor estimado pelo modelo para o IPCA e T é o tamanho da série temporal.

2.4.2 Avaliação da previsão

Para avaliar o melhor modelo em relação a previsão será utilizado o erro quadrático médio (EQM). Desta forma, o modelo que possuir o menor Erro Quadrático Médio (EQM) no período previsto será considerado o melhor.

Morettin e Bussab (2010) define o EQM como a média ao quadrado da diferença entre o valor estimado, \hat{Y}_t e o valor observado, Y_t . Assim, esta medida é dada por:

$$EQM = E[(\hat{Y}_t - Y_t)^2] = Var(\hat{Y}_t) + (E[\hat{Y}_t] - Y_t)^2$$

em que $E[\hat{Y}_t] - Y_t$ representa o viés do estimador. Com isso, quanto menor a variância, menor o viés e assim melhor será o modelo.

3 Análise dos Resultados

Neste Capítulo serão apresentados os resultados dos modelos lineares dinâmicos presentes na Seção 2.2. Para isso, foi utilizado o software R (TEAM, 2021). Para análise dos dados foram utilizados pacotes presentes no mesmo. Desta forma, para aplicar os ajustes dos modelos foi utilizado o pacote *dlm* (PETRIS; GILKS, 2018). Além disso, para a criação dos gráficos foi utilizado o pacote *ggplot2* (WICKHAM et al., 2021a) e para a realizar os tratamentos nos dados foram utilizados os pacotes *xlsx* (DRAGULESCU; ARENDT, 2020), *reshape2* (WICKHAM, 2020), *lubridate* (SPINU; GROLEMUND; WICKHAM, 2021) e *dplyr* (WICKHAM et al., 2021b). Para calcular o MAPE foi utilizado o pacote *MLmetrics* (YAN, 2016). Nos apêndices 2, 3, 4 e 5 estão os códigos que foram usados nessa análise.

3.1 Análise descritiva dos dados

Nesta seção serão apresentados alguns resultados referentes as análises das séries temporais no período de janeiro de 2012 à dezembro de 2021 para o índice geral do IPCA.

Na Figura 2 é apresentada a série do índice geral do IPCA. A partir de uma análise, pode verificar que a série parece ter sazonalidade anual, sendo que repete o comportamento ao longo do período, obtendo variações mais altas no início e no final do ano devido ao período de festas e de férias, onde alguns subitens como passagem aérea e produtos para presente sofrem uma alta nos preços.

Em 2018 pode-se observar um pico de variação, decorrente da crise dos caminhoneiros, que fez com o abastecimento de alimentos e de produtos que utilizavam prioritariamente o sistema rodoviário para ser distribuído pelo país sofresse um aumento considerável nos preços, desta forma, impactando o índice (FGV, 2018).

Outro impacto na série é observado a partir do início de 2020, em que começou a pandemia da Covid-19, elevando os preços de vários subitens componentes do índice, como

gasolina, alimentos em geral e principalmente os produtos que possuem algum componente que é exportado, devido a alta do dólar (CONSULTORIA, 2020).

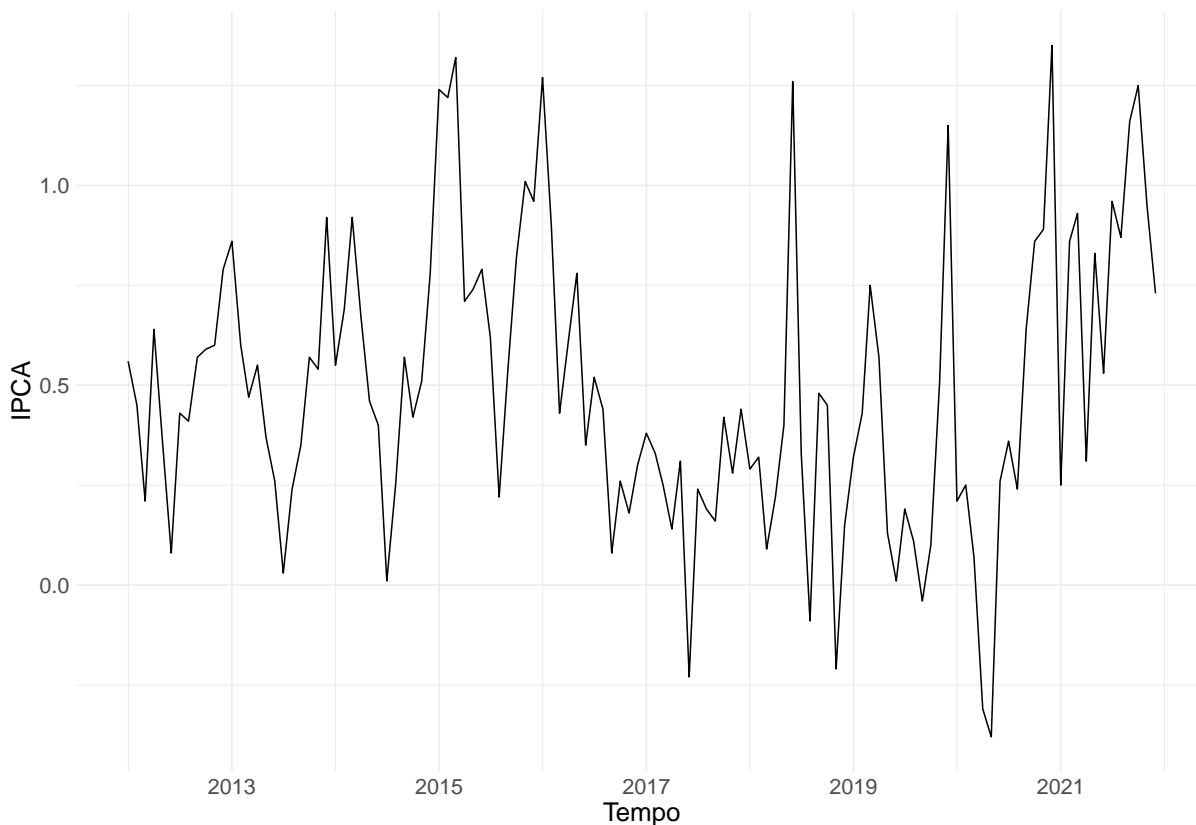


Figura 2: Série Temporal do IPCA (IBGE) de janeiro de 2012 à dezembro de 2021.

3.2 Resultados dos modelos lineares dinâmicos

Em todos os modelos apresentados na Seção 2.2 será utilizada a série do índice geral do IPCA como variável de interesse. Além disso, no ajuste do modelo polinomial de 1° ordem, foi gerada uma cadeia de tamanho 10.000, com um período de aquecimento de 2.000 e um espaçamento de 4. Esta configuração gerou uma amostra final de 2.000 valores para cada um dos parâmetros desconhecidos. Com isso, foi calculada uma estimativa pontual, sendo a mediana da amostra *a posteriori* para cada tempo t .

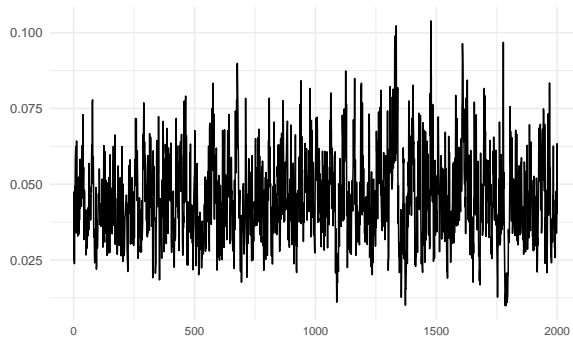
No modelos polinomial de 1° ordem com sazonalidade foi gerada uma cadeia de tamanho 100.000, com um período de aquecimento de 2.000 e um espaçamento de 40. Esta configuração gerou uma amostra final de 2.000 valores para cada um dos parâmetros desconhecidos. Assim, foi calculada uma estimativa pontual, sendo a mediana da amostra *a posteriori* para cada tempo t .

Para as variâncias desconhecidas V e W foram utilizados os mesmos métodos descritos acima para cada um dos modelos. Os valores utilizados para o tamanho da cadeia, o período de aquecimento e o espaçamento dos modelos, foi determinado a partir da visualização gráfico, para que os parâmetros desconhecidos convergissem para a distribuição *a posteriori*.

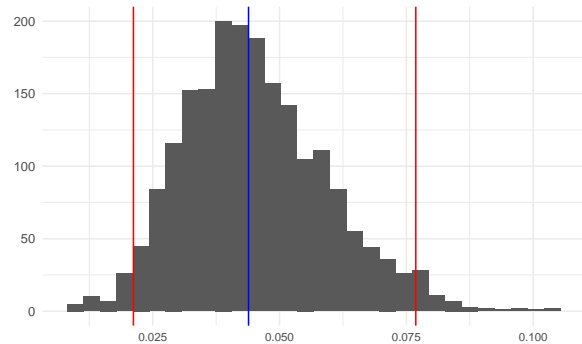
3.2.1 Ajustando o modelo polinomial de 1° ordem

Para o ajuste do IPCA, foi considerado o seguinte modelo polinomial de 1° ordem:

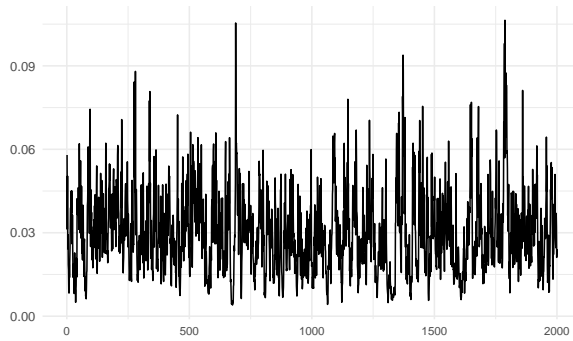
$$\begin{aligned}
 Y_t &= \mu_t + \epsilon_t, \quad \epsilon_t \sim N[0, V], \quad t = 1, \dots, 120, \\
 \mu_t &= \mu_{t-1} + \omega_t, \quad \omega_t \sim N[0, W], \quad t = 1, \dots, 120, \\
 (\mu_0 | D_0) &\sim N[0, 1e^{+7}], \\
 (\tau | D_0) &\sim G[1e^{-3}, 1e^{-3}], \quad \tau = 1/V \\
 \tau_\mu &\sim G(1e^{-3}, 1), \quad \tau_\mu = 1/W.
 \end{aligned} \tag{3.1}$$



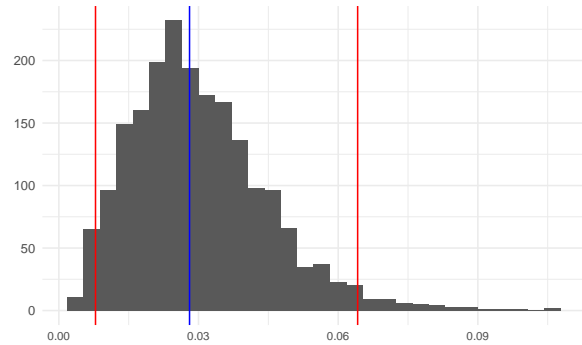
(a) Amostras *a posteriori* para V .



(b) Histograma para V .



(c) Amostras *a posteriori* para W .



(d) Histograma para W .

Figura 3: Amostra *a posteriori* para W e V no modelo polinomial de 1° ordem. As linhas em vermelho são os IC com nível de credibilidade de 95%, a esquerda o limite inferior e a direita o limite superior. A linha azul é a mediana.

Variâncias	Estimativa pontual	Quantil: 2.5%	Quantil: 97.5%
V	0,0439	0,0212	0,0769
W	0,0281	0,0079	0,0642

Tabela 1: Estimativa pontual e intervalar para V e W.

Na Figura 3, pode-se verificar que as variâncias W e V convergem. Além disso, a variância observacional V aparenta ser maior que a variância de evolução W . É possível visualizar o ajuste do modelo polinomial de 1° ordem na Figura 4. Pela análise inicial do gráfico, é possível perceber que o modelo consegue captar bem o comportamento da série do IPCA, principalmente a tendência da mesma, porém não consegue acompanhar os picos de variações positivas e negativas da série ao longo do tempo. No Apêndice 6 encontram-se as cadeias em 10 tempos sorteados para μ , que mostram a convergência deste parâmetro.

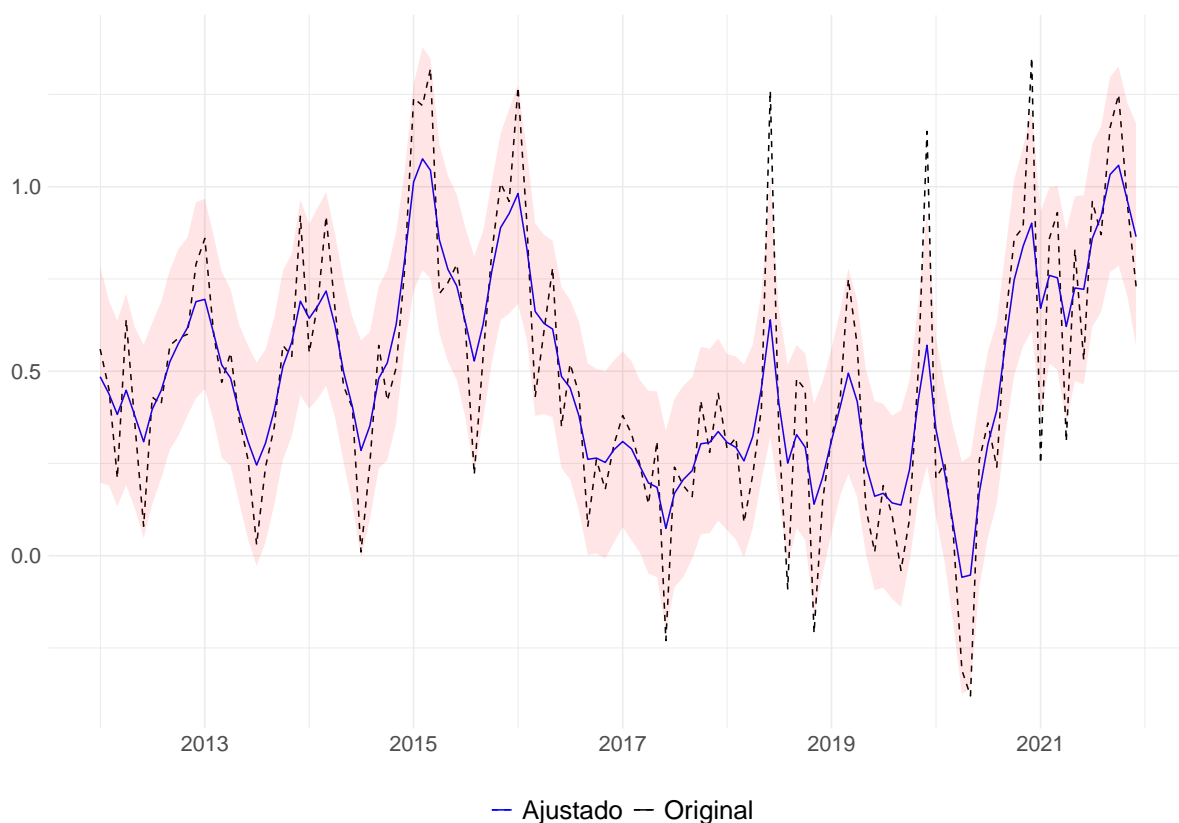


Figura 4: Série ajustada para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.

3.2.2 Ajustando o modelo polinomial de 1° ordem com sazonalidade

No modelo presente na Seção 2.2.2, foi considerada uma sazonalidade anual, pois, com base na Figura 2, o índice de preços ao consumidor amplo (IPCA) exibe um padrão sazonal, caracterizado por um padrão de comportamento do índice, com variações mais altas no início e no fim de cada ano, e mais moderadas ao longo do ano.

Desta forma, para o ajuste do IPCA, será considerado o seguinte modelo polinomial de 1° ordem com sazonalidade:

$$\begin{aligned}
Y_t &= \mu_t + \gamma_{12m} + \epsilon_t, \quad \epsilon_t \sim N(0, V), \quad t = 1, \dots, 120, \quad V = \tau^{-1} \\
\mu_t &= \mu_{t-1} + \omega_{1,t}, \quad \omega_{1,t} \sim N(0, \tau_{\theta,1}^{-1}), \quad t = 1, \dots, 120, \\
\gamma_{1,t} &= -\gamma_{2,t-1} - \gamma_{3,t-1} - \dots - \gamma_{12,t-1}, \quad t = 1, \dots, 120, \\
\gamma_{2,t} &= \gamma_{2,t-1}, \quad t = 1, \dots, 120, \\
&\vdots \\
\gamma_{11,t} &= \gamma_{11,t-1}, \quad t = 1, \dots, 120, \\
(\tau|D_0) &\sim G[1e^{-3}, 1e^{-3}], \\
\tau_{\theta,1} &\sim G(1e^{-3}, 1), \\
(\mu_0|D_0) &\sim N[0, 1e^{+7}], \\
(\gamma_{i,0}|D_0) &\sim N[0, 1e^{+7}], \quad i = 1, \dots, 11.
\end{aligned} \tag{3.2}$$

É possível perceber, pelo modelo definido acima, que será assumido que não existe covariância entre os componentes de $\theta_t = (\mu_t, \gamma_{1,t}, \dots, \gamma_{11,t})$, ou seja, $\omega_{2,t} = \omega_{3,t} = \dots = \omega_{11,t} = 0$. Com isso a matriz diagonal \mathbf{W} é dada por:

$$\mathbf{W} = \begin{pmatrix} \tau_{\theta,1}^{-1} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{12 \times 12} \tag{3.3}$$

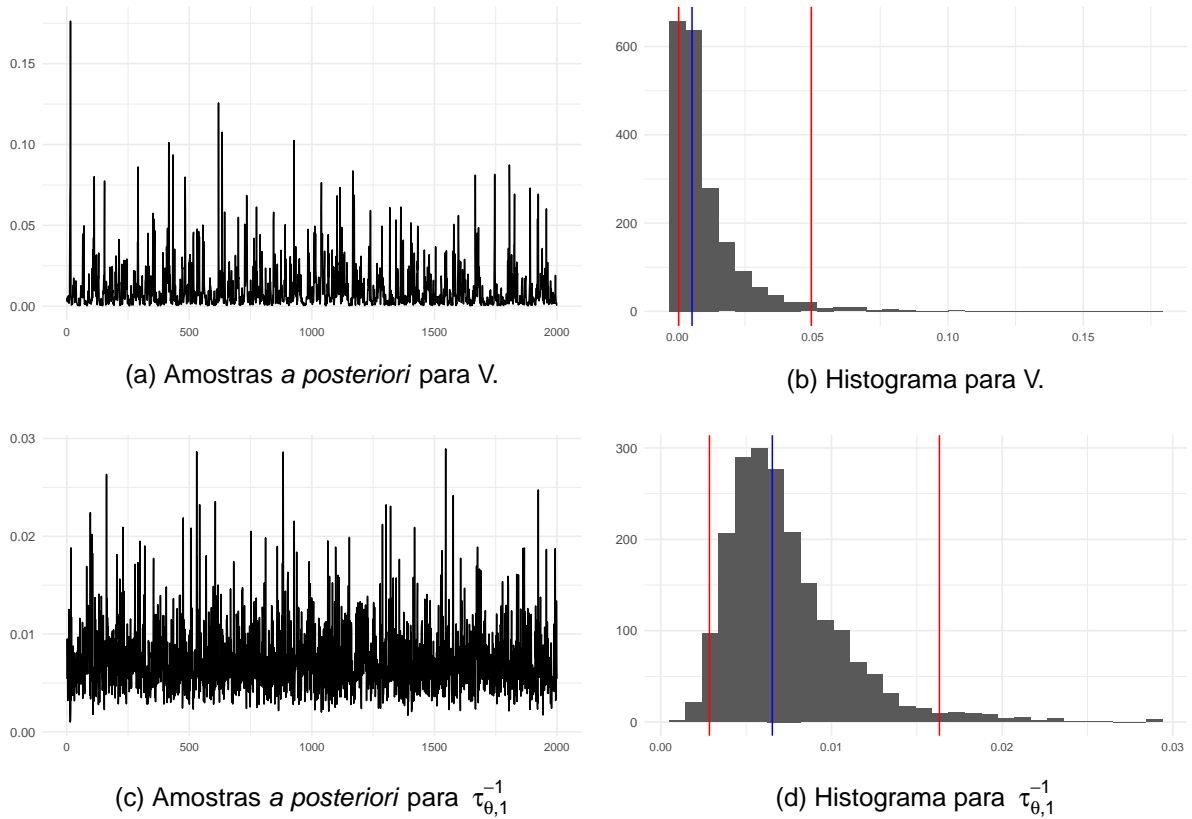


Figura 5: Amostra *a posteriori* para $\tau_{\theta,1}^{-1}$ e V no modelo polinomial de 1° ordem com sazonalidade. As linhas em vermelho são os IC com nível de credibilidade de 95%, a esquerda o limite inferior e a direita o limite superior. A linha azul é a mediana.

Variâncias	Estimativa pontual	Quantil: 2.5%	Quantil: 97.5%
V	0,0056	0,0006	0,0487
$\tau_{\theta,1}^{-1}$	0,0066	0,0030	0,0150

Tabela 2: Estimativa pontual e intervalar para V e $\tau_{\theta,1}^{-1}$.

Na Figura 5, pode-se verificar que as variâncias $\tau_{\theta,1}^{-1}$ e V convergem. Além disso, a variância observacional V aparenta ser menor que a variância $\tau_{\theta,1}^{-1}$. Na Figura 6 é possível visualizar o ajuste do modelo polinomial de 1° ordem com sazonalidade. Pela análise gráfica, é possível perceber que o modelo consegue captar de forma precisa o comportamento da série do IPCA, conseguindo representar os picos de variações positivas e negativas da série ao longo do tempo. Neste ajuste, pode-se considerar que haja *overfitting*. No Apêndice 7 encontram-se as cadeias em 10 tempos sorteados para μ e de algumas γ , que mostram a convergência destes parâmetros.

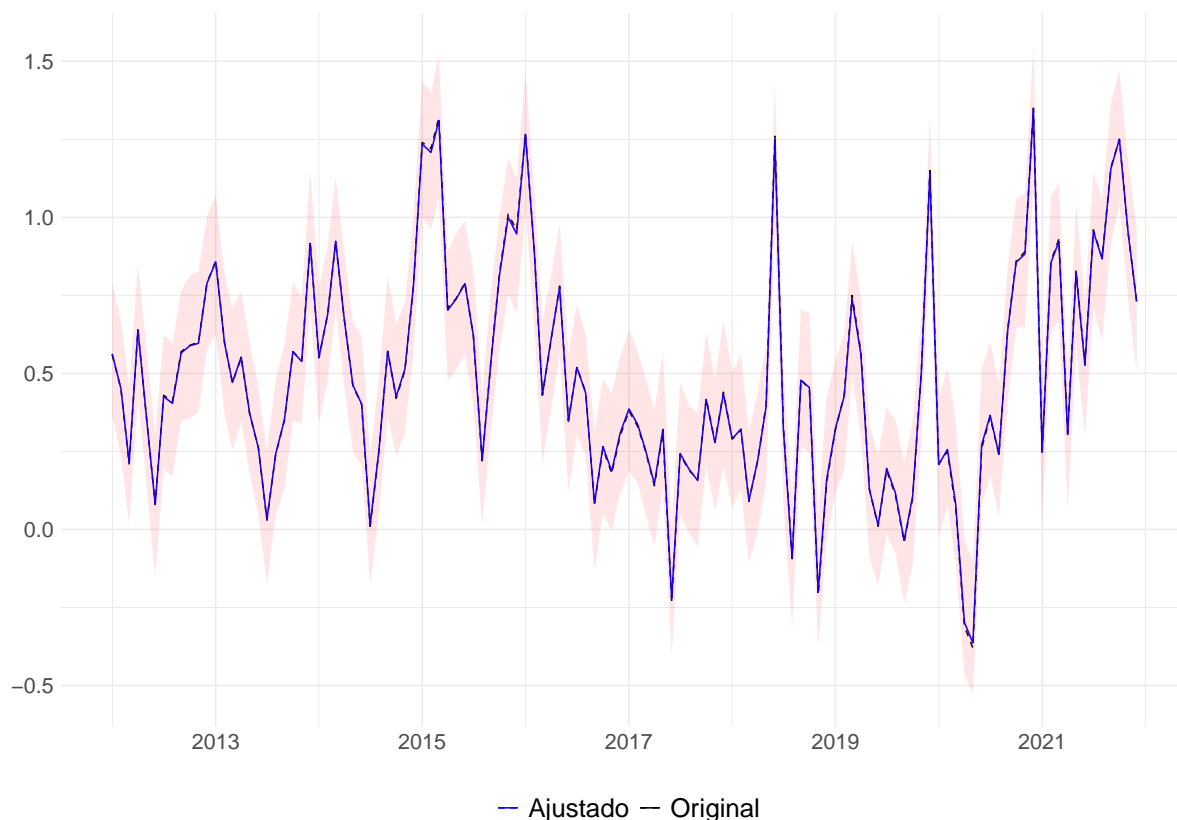


Figura 6: Série ajustada para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.

3.2.3 Comparando os ajustes dos modelos

Para analisar numericamente a assertividade dos modelos, foram utilizados o erro percentual absoluto médio (MAPE) e o desvio absoluto médio (MAD) apresentados, respectivamente, na Seção 2.4.1 e os resultados estão presentes na Tabela 3.

Modelos	MAPE	MAD
Modelo Polinomial de 1° Ordem	0,7948	0,1199
Modelo Polinomial de 1° Ordem com Sazonalidade	0,0222	0,0044

Tabela 3: Erro Percentual Absoluto Médio (MAPE) dos Modelos.

A partir dos resultados das tabelas acima, é possível verificar que o modelo polinomial de 1° ordem com sazonalidade representa melhor o comportamento da série do IPCA, pois possui os menores MAPE e Desvio Absoluto Médio (MAD) dentre os modelos testados.

3.3 Resultados das previsões dos modelos

Nesta seção, será assumido que a série de 2021 do IPCA seja desconhecida, pois desta forma, será possível verificar a capacidade preditiva desses modelos. Assim, será ajustado um novo modelo com a série do índice geral do IPCA de janeiro de 2012 até dezembro de 2020.

No primeiro momento, para a previsão nos modelos apresentados neste trabalho serão considerados 4-passos a frente, desta forma, os dados previstos contemplam o período de janeiro de 2021 à abril de 2021. No segundo momento, será realizada previsão para 12 períodos a frente, com o intuito de verificar a capacidade preditiva destes modelos. Por fim, o modelo que obtiver a melhor previsão será utilizado para a previsão do ano de 2022 do índice geral do IPCA. Para a comparação com a previsão dos modelos serão consideradas as variações mensais do índice geral do IPCA no mesmo período de cada previsão.

3.3.1 Previsão dos modelos 4 passos a frente para 2021

Nas Figuras 7 e 8 estão os resultados das previsões 4 passos a frente para o modelo polinomial de 1° ordem e o modelo polinomial de 1° ordem com sazonalidade.

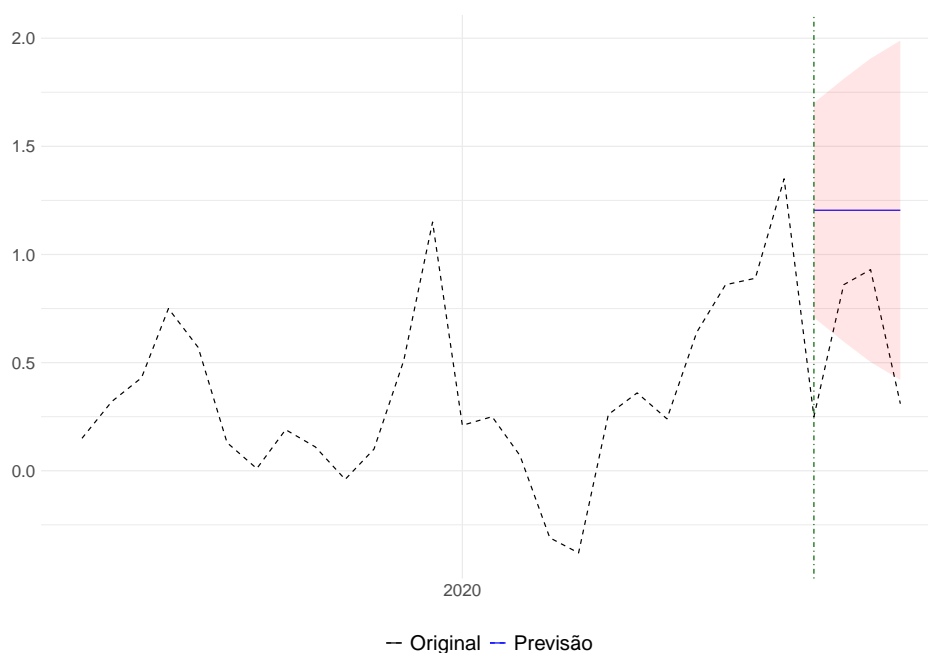


Figura 7: Previsão 4 passos a frente para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.



Figura 8: Previsão 4 passos a frente para o IPCA pelo modelo polinomial de 1º ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.

Pela análise da Figura 7, é possível perceber que a previsão do modelo é constante ao longo do período de janeiro de 2021 à abril de 2021. Este fato ocorre, pois este modelo, por ser definido pela quádrupla $\{1, 1, V, W\}$, possui o $F_t = 1$ e o $G_t = 1$, e assim faz com que o valor da previsão k -passos a frente seja um valor constante ao longo do período. Com isso, é possível verificar que a previsão do IPCA a partir deste modelo parece não ser a mais adequada. Além disso, é possível verificar pelo IC que este modelo tem por característica apresentar um aumento da incerteza com o passar do tempo.

Com base na Figura 8, é possível verificar que na previsão 4 passos a frente do modelo acompanha a tendência da série do IPCA, porém não consegue captar de forma adequada os picos de queda da série. Além disso, é possível verificar que na maior parte do tempo, a série original fica fora do IC.

3.3.2 Previsão dos modelos 12 passos a frente para 2021

Nas Figuras 10 e 9 é possível visualizar os resultados das previsões 12 passos a frente para o modelo polinomial de 1° ordem e o modelo polinomial de 1° ordem com sazonalidade.

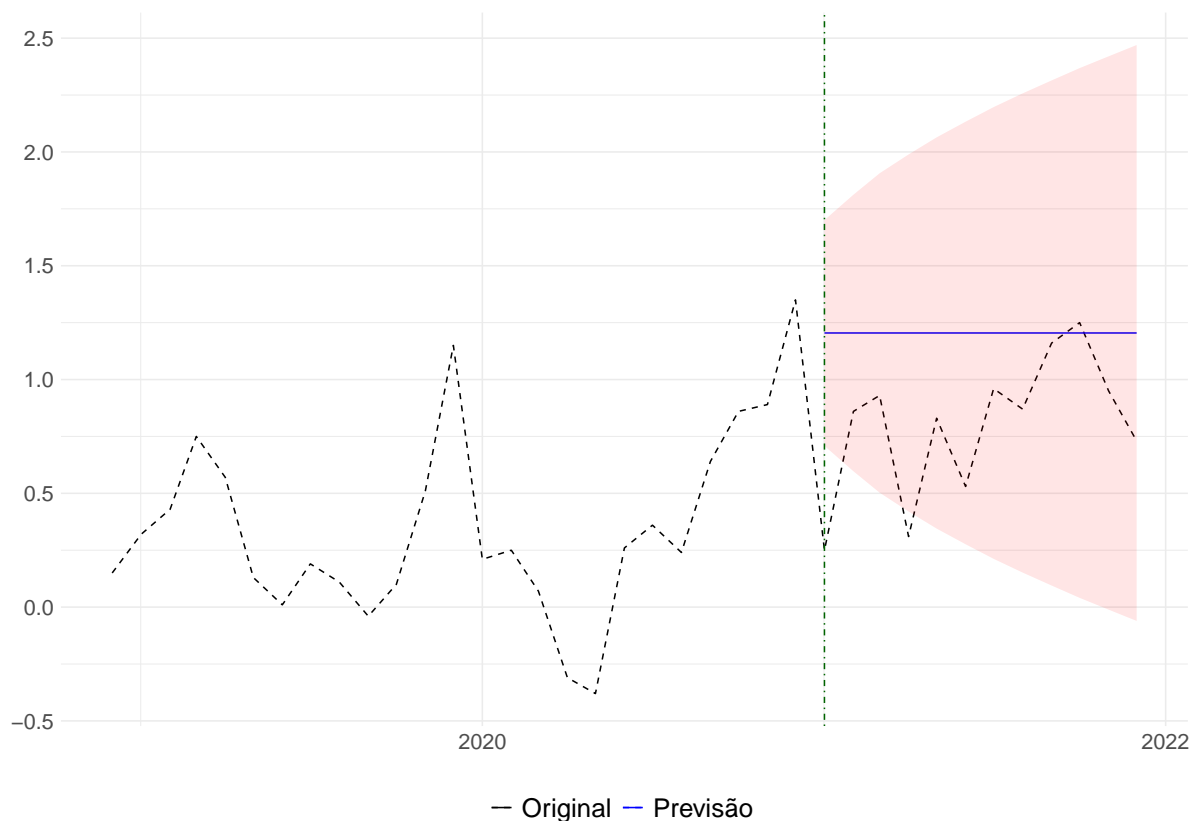


Figura 9: Previsão 12 passos a frente para o IPCA pelo modelo polinomial de 1° ordem, em que a parte sombreada em vermelho é intervalo de credibilidade de 95%.

Na Figura 9, é possível perceber que a previsão 12 passos à frente do modelo também é constante ao longo do período de janeiro de 2021 à dezembro de 2021. Com isso, tanto na previsão 4 passos à frente quanto na previsão 12 passos à frente, a previsão do IPCA a partir deste modelo parece não ser a mais adequada. Além disso, nesta previsão também é possível verificar o aumento da incerteza com o passar do tempo pelo IC.

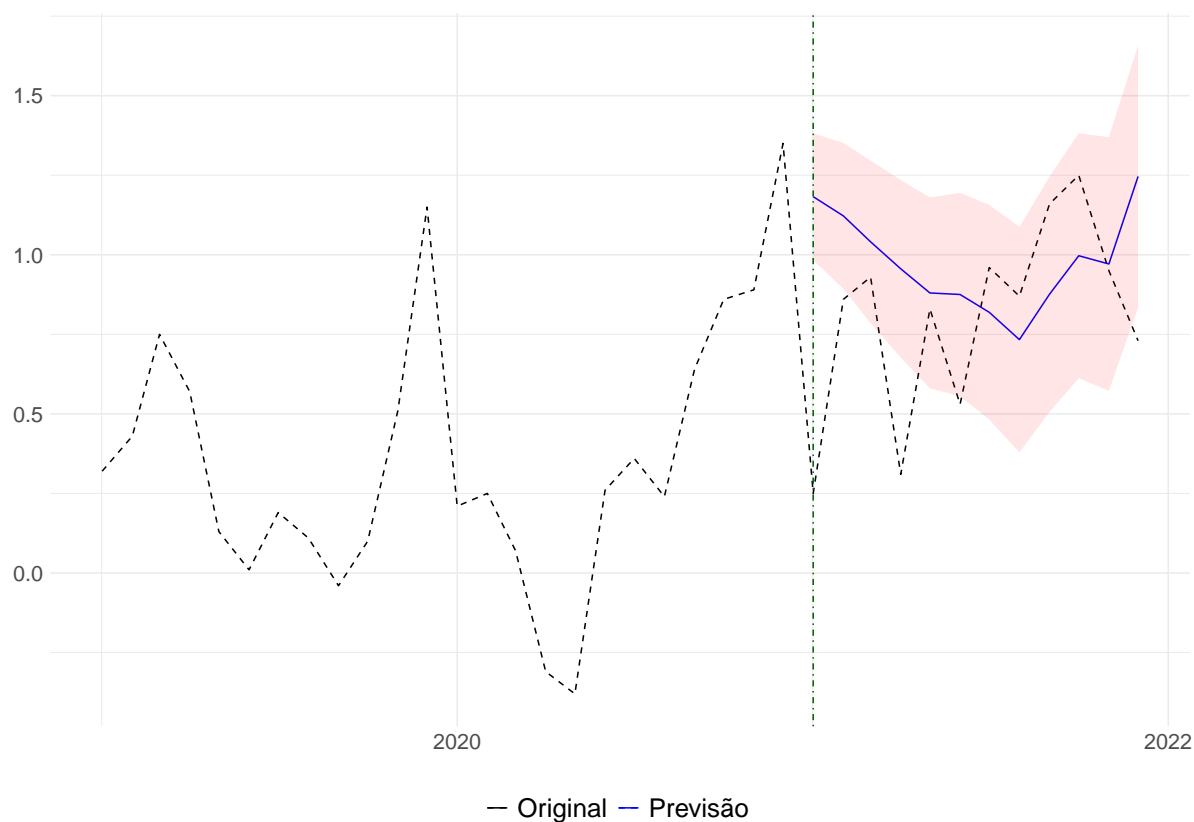


Figura 10: Previsão 12 passos a frente para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.

Com base na Figura 10, a previsão do modelo consegue acompanhar, inicialmente, a tendência da série porém a partir de novembro de 2021, a previsão superestima o valor da série original. Além disso, nesta previsão a maior parte da série se encontra dentro do intervalo de credibilidade.

3.3.3 Comparando as previsões dos modelos

Para analisar a assertividade da previsão dos modelos de forma numérica, foi utilizado o Erro Quadrático Médio (EQM) apresentado na Seção 2.4.2 e os resultados estão na Tabela 4.

Modelos	EQM 4-passos	EQM 12-passos
Modelo Polinomial de 1° Ordem	0,4991	0,2621
Modelo Polinomial de 1° Ordem com Sazonalidade	0,3293	0,1573

Tabela 4: Erro Quadrático Médio (EQM) das Previsões.

Com base nos resultados da tabela acima e da análise gráfica feita na Seção 3.3, é possível verificar que os MLD testados possuem algumas dificuldades na predição do IPCA, especialmente o modelo polinomial de 1ª ordem. Assim, dentre os modelos testados, o que obteve um menor EQM e conseqüentemente uma melhor previsão foi o modelo polinomial de 1° ordem com sazonalidade, tanto na previsão realizada para uma janela menor como para o cenário com uma janela maior.

Vale ressaltar que como os valores da previsão no modelo polinomial de 1° ordem são constantes ao longo do tempo, a incerteza associada é bem maior que a apresentada no modelo polinomial de 1° ordem com sazonalidade. Portanto, neste sentido, este modelo também é considerado melhor.

3.3.4 Previsão do modelo polinomial de 1° ordem com sazonalidade para 2022

Com base nos resultados obtidos na Seção 3.3.3, verificou-se que o modelo polinomial de 1° ordem com sazonalidade obteve os melhores ajustes e previsão, segundo, respectivamente, a Tabela 3 e 4. Desta forma, na Figura 11 é possível visualizar a previsão de 2022 do índice geral do IPCA.

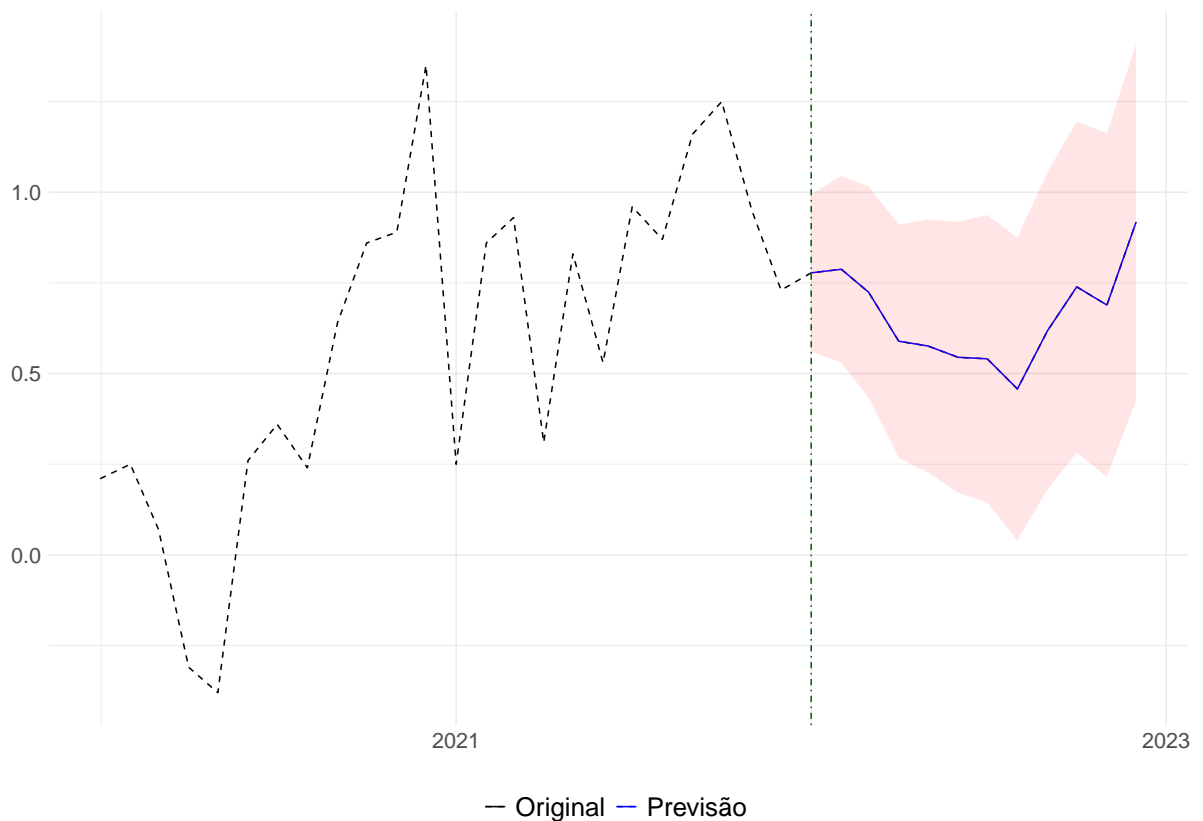


Figura 11: Previsão de 2022 para o IPCA pelo modelo polinomial de 1° ordem com sazonalidade, em que a parte sombreada em vermelho corresponde ao intervalo de credibilidade de 95%.

A partir do gráfico acima, é possível verificar que a previsão do modelo polinomial de 1° ordem com sazonalidade, tem como expectativa um decréscimo na variação do índice geral do IPCA até o meio do ano, ficando em torno de 0,50, porém a partir deste período, o modelo prevê uma aceleração na inflação, chegando a quase 1,00% em dezembro de 2022. No Apêndice 1 está a tabela com as variações mensais, e os respectivos quantis 2,5% e 97,5% da previsão de 2022 do modelo polinomial de 1° ordem com sazonalidade para o índice geral do IPCA.

4 Conclusão

O objetivo principal deste trabalho foi o estudo da aplicação de modelos lineares dinâmicos pela perspectiva bayesiana para a modelagem e previsão do Índice de Preços ao Consumidor Amplo (IPCA - IBGE).

Neste trabalho foram aplicados dois modelos lineares dinâmicos: o modelo polinomial de 1° ordem e o modelo polinomial de 1° ordem com sazonalidade. No primeiro modelo, foi considerado que a variância observacional $V_t = V$ e de evolução $\mathbf{W}_t = \mathbf{W}$ desconhecidas e constantes. Já para o modelo polinomial de 1° ordem com sazonalidade, $V_t = V$ e $\mathbf{W} = \text{diag}(\tau_{\theta,1}^{-1}, 0, \dots, 0)$ uma matriz diagonal, com somente o primeiro elemento, associado a evolução do componente de tendência, diferente de 0. As distribuições a *priori* das variâncias desconhecidas foram consideradas gama inversas independentes, ou seja, foram consideradas distribuições gama para as precisões. Além disso, para a estimação destas variâncias e dos vetores de parâmetros de estado θ_t , foram aplicados os algoritmos de Monte Carlo via Cadeias de Markov.

No ajuste destes modelos, foi utilizado a série temporal no período de janeiro de 2012 à dezembro de 2021 do índice geral do IPCA como variável de interesse. Assim, para avaliar a qualidade do ajuste foram utilizadas as métricas MAD e MAPE. Desta forma, os resultados para MAD e MAPE foram, respectivamente, 0,0222 e 0,0044 no modelo polinomial de 1° ordem com sazonalidade, já no modelo polinomial de 1° ordem foram obtidos os valores de 0,7948 para o MAPE e 0,1199 para MAD. Com base nestes resultados, o modelo polinomial de 1° ordem não obteve um desempenho tão bom em relação ao modelo com sazonalidade, sendo um modelo que acompanha bem a tendência, porém não capta de forma adequada os picos da série.

No caso da previsão do IPCA, foi considerado desconhecido o período de 2021, pois desta forma, seria possível verificar a capacidade preditiva dos modelos. Os modelos foram avaliados em dois cenários, foram realizadas previsões com 4 e 12 passos à frente. Com isso, foi utilizado o EQM como métrica para a avaliação da previsão do IPCA. Desta forma,

os resultados para EQM para 4 passos e 12 passos à frente para o modelo polinomial de 1° ordem foram, respectivamente, 0,4991 e 0,2621, já no modelo polinomial de 1° ordem com sazonalidade foram obtidos, respectivamente, os valores de 0,3293 e 0,1573 do EQM para 4 e 12 passos à frente.

Com base nestes resultados, verifica-se que as previsões 4 e 12 passos à frente no modelo polinomial de 1° ordem com sazonalidade obteve melhores resultados. Além disso, o IC do modelo polinomial de 1° ordem possui um aumento da incerteza na previsão, o que não é observado para o modelo polinomial de 1° ordem com sazonalidade. Vale ressaltar ainda, que boa parte da série original é coberta por esse IC. Com base neste resultados, o modelo polinomial de 1° ordem com sazonalidade foi considerado o melhor. Desta forma, pode-se considerar que o objetivo principal deste trabalho foi cumprido, uma vez que o mesmo foi abordado.

Ao final do Capítulo 3, foi realizada uma previsão para 2022 do índice geral do IPCA utilizando o modelo polinomial de 1° ordem com sazonalidade. De acordo com o modelo, tem-se a expectativa um decréscimo na variação do índice geral do IPCA até o meio do ano, e a partir desse período um acréscimo na variação até dezembro de 2022.

Para trabalhos futuros, podem ser seguidas algumas linhas para o assunto abordado neste trabalho, como realizar um estudo mais aprofundado dos modelo lineares dinâmicos, ampliando as possibilidades de uso na série do IPCA, como, por exemplo, a utilização de variáveis explicativas. Seguindo esta linha, pode-se considerar a aplicação de modelos lineares dinâmicos generalizados, desta forma, podem ser mais apropriadas para aplicação dos modelos na série do IPCA. Pode-se também investigar outros cenários com outras componentes da diagonal principal de \mathbf{W} diferentes de 0, para ver o impacto que esta mudança teria no ajuste e na previsão do modelo com sazonalidade. Além disso, pode-se considerar a aplicação de outros modelos de espaço de estado, dado que esta classe de modelos é muito vasta.

Referências

BACEN. *O que é inflação?* 2021. <https://www.bcb.gov.br/controleinflacao/oqueinflacao>. Acessado em 13/08/2021.

BARBOSA, R. B.; FERREIRA, R. T.; SILVA, T. M. da. Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. 2020.

CARTER, C. K.; KOHN, R. On gibbs sampling for state space models. 1994.

CNN. *País vive choque inédito com dólar e commodities em alta - e isso afeta seu bolso.* 2021. <https://www.cnnbrasil.com.br/business/pais-vive-choque-inedito-com-dolar-e-commodities-em-alta-e-isso-afeta-seu-bolso/>. Acessado em 14/09/2021.

CONSULTORIA, N. *PANDEMIA: IMPACTO PARA A ECONOMIA BRASILEIRA 2020.* 2020. https://novoidealconsultoria.com.br/dicas/pandemia-impacto-para-a-economia-brasileira-2020/?pht=36801570733355363&utm_source=google&utm_medium=cpc&utm_campaign=google&gclid=Cj0KCQiAuP-OBhDqARIsAD4XHpcTaP2Vz1BelFOKAtxnVZCabz6MDjvdyE7nXCcbtphMPxwTQwcB. Acessado em 13/01/2022.

DRAGULESCU, A.; ARENDT, C. *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files.* 2020. <https://cran.r-project.org/web/packages/xlsx/index.html>. Acessado em 16/01/2022.

FGV. *Preços aceleram na semana da greve dos caminhoneiros.* 2018. <https://portal.fgv.br/noticias/precos-aceleram-semana-greve-caminhoneiros>. Acessado em 11/02/2022.

FRÜHWIRTH-SCHNATTER, S. Data augmentation and dynamic linear models. 1994.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, n. 6, p. 721–741, 1984.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Oxford University Press, 1970.

IBGE. *Série histórica do Índice Nacional de Preços ao Consumidor Amplo (IPCA) - julho de 2006 até dezembro de 2011.* 2011. <https://sidra.ibge.gov.br/tabela/2938>. Acessado em 19/01/2022.

IBGE. *Série histórica do Índice Nacional de Preços ao Consumidor Amplo (IPCA) - fevereiro de 2012 à dezembro de 2019.* 2019. <https://sidra.ibge.gov.br/tabela/1419>. Acessado em 21/08/2021.

- IBGE. *Pesquisa de Orçamentos Familiares (POF)*. 2021. <<https://www.ibge.gov.br/estatisticas/sociais/trabalho/9050-pesquisa-de-orcamentos-familiares.html?=&t=o-que-e>>. Acessado em 21/08/2021.
- IBGE. *Série histórica do Índice Nacional de Preços ao Consumidor Amplo (IPCA) - janeiro de 2020 à agosto de 2021*. 2021. <<https://sidra.ibge.gov.br/tabela/7060>>. Acessado em 21/08/2021.
- IBGE. *Índice Nacional de Preços ao Consumidor Amplo (IPCA)*. 2021. <<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=o-que-e>>. Acessado em 13/08/2021.
- IBRE, F. *Índice de Preços ao Consumidor (IPC)*. 2021. <<https://portalibre.fgv.br/ipc>>. Acessado em 21/08/2021.
- IBRE, F. *Índice de Preços ao Consumidor (IPC-C1)*. 2021. <<https://portalibre.fgv.br/estudos-e-pesquisas/indices-de-precos/ipc-c1>>. Acessado em 21/08/2021.
- JUNIOR, N. C. Utilização do método box-jenkins para previsão de indicadores econômicos (ipca, selic, câmbio e ibovespa). 2014.
- KALMAN, R. A new approach to linear filtering and prediction problems. *Trans. of the AMSE – Journal of Basic Engineering (Series D)*, 1960.
- LEWIS, C. D. Demand forecasting and inventory control. 1997.
- MATTOS, D. M. de; FIGUEIREDO, K. Previsão do ipca utilizando Árvores de regressão com variáveis selecionadas por dynamic time warping. 2017.
- METROPOLIS, N. et al. Equation of state calculations by fast computing machine. *American Institute of Physics and The journal of chemical physics*, v. 21, n. 6, p. 1087–1092, 1953.
- MORETTIN, P. A.; BUSSAB, W. O. *Estatística Básica*. 3. ed. [S.l.]: Editora Saraiva, 2010.
- NOGUEIRA, D. A.; FERREIRA, D. F.; SAFADI, T. Avaliação de critérios de convergencia univariados para o método de monte carlo via cadeias de markov. 2004.
- PETRIS, G. An r package for dynamic linear models. 2010.
- PETRIS, G. State space models in r. 2011.
- PETRIS, G.; GILKS, W. *dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models*. 2018. <<https://cran.r-project.org/web/packages/dlm/index.html>>. Acessado em 15/01/2022.
- PETRIS, G.; PETRONE, S.; CAMPAGNOLI, P. Dynamic linear models with r. 2007.
- PORTES, P. C. Modelagem bayesiana dos níveis máximos do Índice de preços ao consumidor. 2017.
- SHEPHARD, N. Partial non-gaussian state space models. 1994.

- SPINU, V.; GROLEMUND, G.; WICKHAM, H. *lubridate: Make Dealing with Dates a Little Easier*. 2021. <<https://cran.r-project.org/web/packages/lubridate/index.html>>. Acessado em 16/01/2022.
- TEAM, R. C. *R: A Language and Environment for Statistical Computing*. 2021. <<https://www.r-project.org/>>. Acessado em 15/01/2022.
- VALOR. *Marcas da pandemia*. 2021. <<https://valor.globo.com/coronavirus/a-economia-na-pandemia/>>. Acessado em 14/09/2021.
- WEST, M.; HARRISON, J. *Bayesian Forecasting and Dynamic Models*. 2. ed. [S.l.]: Springer, 1997.
- WICKHAM, H. *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. 2020. <<https://cran.r-project.org/web/packages/reshape2/index.html>>. Acessado em 16/01/2022.
- WICKHAM, H. et al. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. 2021. <<https://cran.r-project.org/web/packages/ggplot2/index.html>>. Acessado em 15/01/2022.
- WICKHAM, H. et al. *dplyr: A Grammar of Data Manipulation*. 2021. <<https://cran.r-project.org/web/packages/dplyr/index.html>>. Acessado em 16/01/2022.
- YAN, Y. *MLmetrics: Machine Learning Evaluation Metrics*. 2016. <<https://cran.r-project.org/web/packages/MLmetrics/index.html>>. Acessado em 27/01/2022.

APÊNDICE 1 – Tabela com a previsão de 2022 do modelo polinomial de 1° ordem com sazonalidade para o índice geral do IPCA

Mês/Ano	Variação Mensal	Quantil: 2,5%	Quantil: 97,5%
Janeiro/2022	0,7778	0,5608	0,9947
Fevereiro/2022	0,7876	0,5304	1,0448
Março/2022	0,7244	0,4329	1,0159
Abril/2022	0,5892	0,2675	0,9109
Maiio/2022	0,5760	0,2272	0,9249
Junho/2022	0,5448	0,1712	0,9185
Julho/2022	0,5407	0,1441	0,9372
Agosto/2022	0,4575	0,0396	0,8753
Setembro/2022	0,6183	0,1805	1,0560
Outubro/2022	0,7391	0,2826	1,1956
Novembro/2022	0,6889	0,2148	1,1631
Dezembro/2022	0,9177	0,4268	1,4086

Tabela 5: Previsão de 2022 do modelo polinomial de 1° ordem com sazonalidade para o índice geral do IPCA.

APÊNDICE 2 – Código em R do ajuste do modelo polinomial de 1° ordem

```

set.seed(999)
mcmc <- 2000
burn <- 500

### Ajustando o modelo polinomial de 1 ordem
outGibbsIRW <- dlmGibbsDIG(y = ipcatotal,
                          mod = dlmModPoly(1),
                          shape.y = 1e-3, rate.y = 1e-3,
                          shape.theta = 1e-3, rate.theta = 1e-3,
                          n.sample = mcmc + burn, thin = 4)

## Estimativa pontual dos thetas (mediana) (y-estimado)
thetaMedian <- ts(apply(outGibbsIRW$theta[, 1, -burn], 1, median),
                 end = end(ipcatotal),
                 freq = frequency(ipcatotal))
ThetaMedian2 = as.data.frame(thetaMedian)
ThetaMedian2.1 = ThetaMedian2[2:121,]

## MAPE
ipcaMAPEMAD = ipcatotal
MAPEPoli = round(MAPE(ThetaMedian2.1, ipcaMAPEMAD), 4)
MAPEPoli

## MAD
MADPoli = round(mean(abs(ThetaMedian2.1 - ipcaMAPEMAD)), 4)
MADPoli

```

APÊNDICE 3 – Código em R do ajuste do modelo polinomial de 1° ordem com sazonalidade

```

set.seed(999)
mcmc <- 2000
burn <- 500

### Ajustando o modelo polinomial de 1 ordem com sazonalidade
outGibbsIRW2 <- dlmGibbsDIG(y = ipcatotal,
                           mod = dlmModPoly(1) + dlmModSeas(12),
                           shape.y = 1e-3, rate.y = 1e-3,
                           shape.theta = 1e-3, rate.theta = 1e-3,
                           n.sample = mcmc + burn, thin = 40, ind = c(1))

## Calculando o y_estimado
mat_y = matrix(NA, ncol = 120, nrow = 2500)
for(j in 1:2500){
  theta10 = outGibbsIRW2$theta[,j]
  aux = (dlmModPoly(1) + dlmModSeas(12))$FF%*%t(theta10[2:121,])
  mat_y[j,] = c(aux)
}

y_med = apply(mat_y, 2, median)
y_medts = ts(y_med, start = c(2012, 1), frequency = 12)

## MAPE
ipcaMAPEMAD = ipcatotal
MAPESazo = round(MAPE(y_medts, ipcaMAPEMAD), 4)
MAPESazo

## MAD
MADSazo = round(mean(abs(y_medts - ipcaMAPEMAD)), 4)
MADSazo

```

APÊNDICE 4 – Código em R da previsão do modelo polinomial de 1º ordem

```

### Previsao MP
PoliV = median(outGibbsIRW$dV[-(1 : burn)])
PoliW = median(outGibbsIRW$dW[-(1 : burn)])

## Forecast 4 passos
PoliC = dlmFilter(ipca4.1, dlmModPoly(1, dV = PoliV, dW = PoliW))
ForePoli = dlmForecast(PoliC, nAhead = 4)

## Forecast 12 passos
PoliC2 = dlmFilter(ipca4.1, dlmModPoly(1, dV = PoliV, dW = PoliW))
ForePoli2 = dlmForecast(PoliC2, nAhead = 12)

## EQM Previsao Poli 4 passos
ThetaPrevPoli = ts(ForePoli$f, start = c(2021, 1), frequency = 12)
EQMPrevPoly = round(mean((ThetaPrevPoli - ipcaprev4.0)^2, na.rm = TRUE), 4)
EQMPrevPoly

## EQM Previsao Poli 12 passos
ThetaPrevPoli2 = ts(ForePoli2$f, start = c(2021, 1), frequency = 12)
EQMPrevPoly2 = round(mean((ThetaPrevPoli2 - ipcaprev4.1)^2, na.rm = TRUE), 4)
EQMPrevPoly2

```

APÊNDICE 5 – Código em R da previsão do modelo polinomial de 1° ordem com sazonalidade

```
### Previsao MPS
```

```
SazoV = median(outGibbsIRW2$dV[-(1 : burn)])
```

```
SazoW = median(outGibbsIRW2$dW[-(1 : burn)])
```

```
## Forecast 4 passos
```

```
SazoC = dlmFilter(ipca4.1, dlmModPoly(1, dV = SazoV, dW = SazoW) +  
                 dlmModSeas(12, dV = 0, dW = c(0, rep(0, 12 - 2))))
```

```
ForeSazo = dlmForecast(SazoC, nAhead = 4)
```

```
## Forecast 12 passos
```

```
SazoC12 = dlmFilter(ipca4.1, dlmModPoly(1, dV = SazoV, dW = SazoW) +  
                  dlmModSeas(12, dV = 0, dW = c(0, rep(0, 12 - 2))))
```

```
ForeSazo12 = dlmForecast(SazoC12, nAhead = 12)
```

```
## EQM Previsao Sazo 4 passos
```

```
ThetaPrevSazo = ts(ForeSazo$f, start = c(2021, 1), frequency = 12)
```

```
EQMPrevSazo = round(mean((ThetaPrevSazo - ipcaprev4.0)^2, na.rm = TRUE), 4)
```

```
EQMPrevSazo
```

```
## EQM Previsao Sazo 12 passos
```

```
ThetaPrevSazo2 = ts(ForeSazo12$f, start = c(2021, 1), frequency = 12)
```

```
EQMPrevSazo12 = round(mean((ThetaPrevSazo2 - ipcaprev4.1)^2, na.rm = TRUE), 4)
```

```
EQMPrevSazo12
```

APÊNDICE 6 – Cadeias de Markov para μ_t do modelo polinomial de 1° ordem

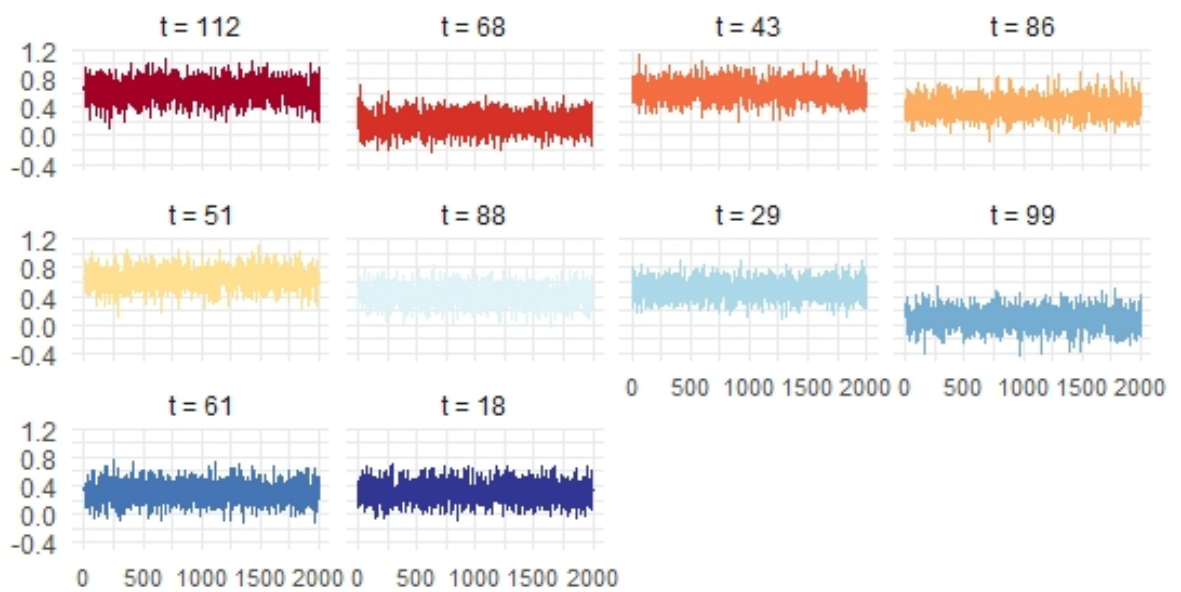


Figura 12: Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para μ_t do modelo polinomial de 1° ordem.

APÊNDICE 7 – Cadeias de Markov para μ_t e para os γ_{12m} do modelo polinomial de 1° ordem com sazonalidade

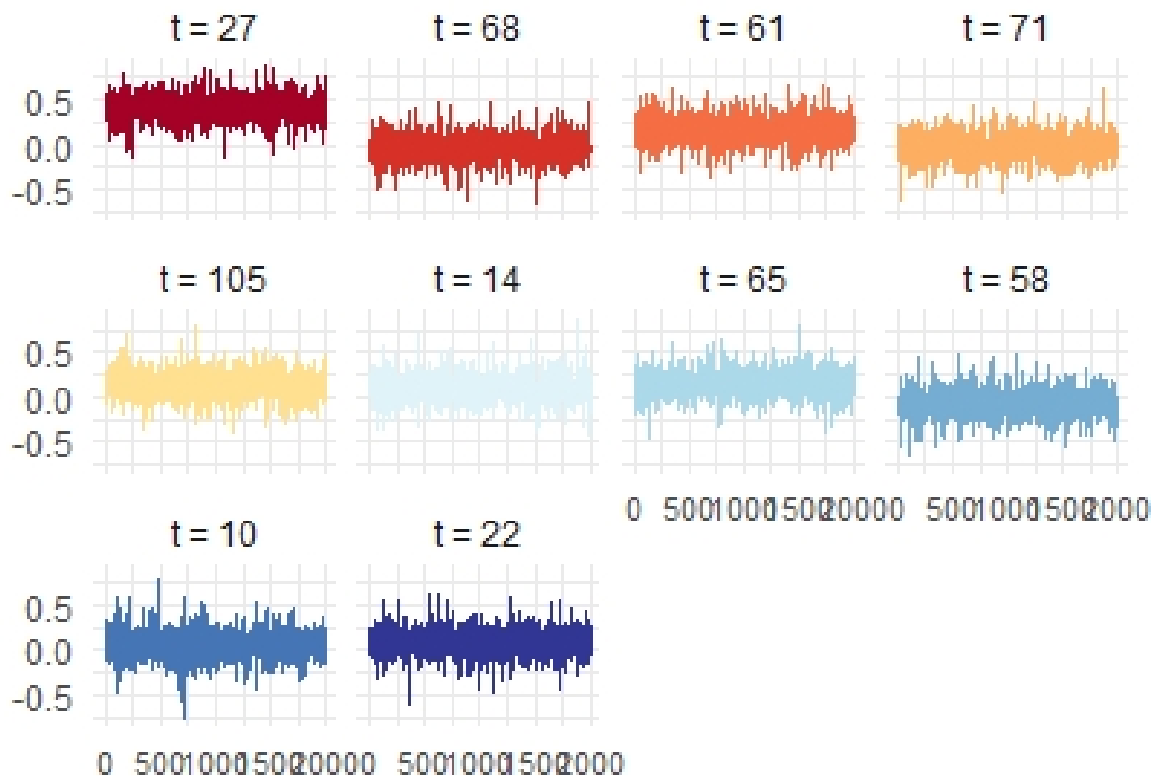


Figura 13: Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_2 do modelo polinomial de 1° ordem com sazonalidade.

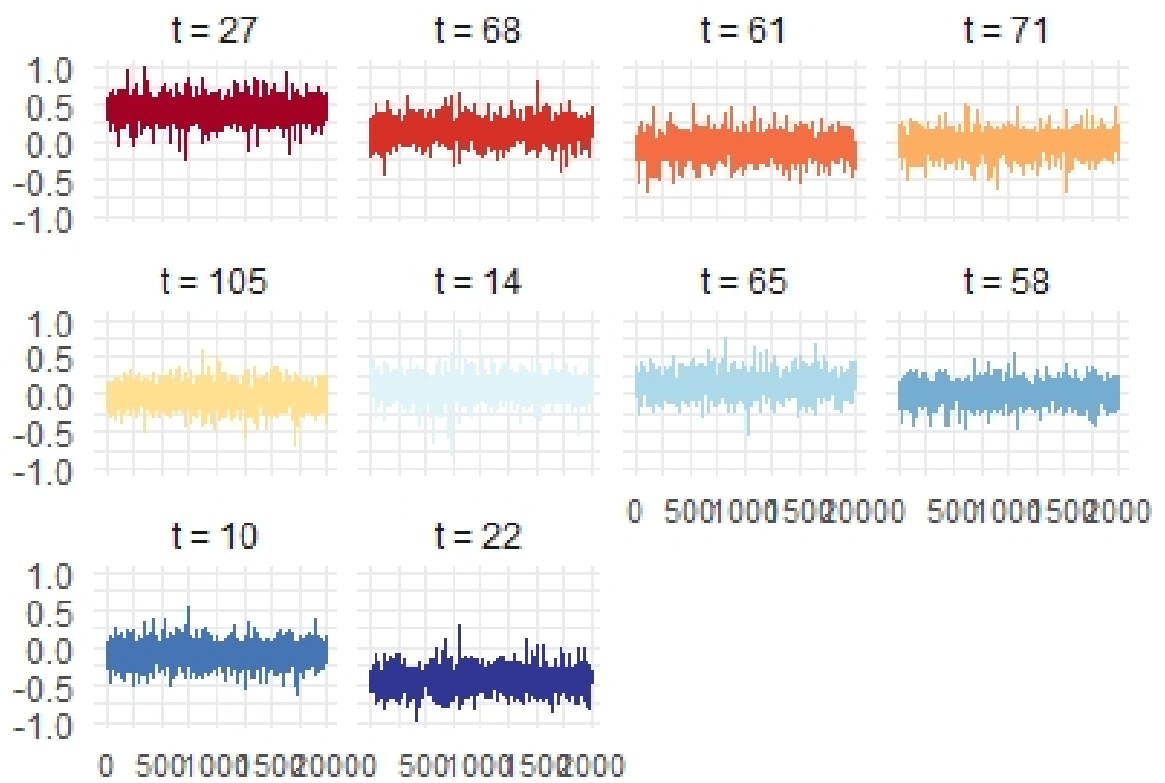


Figura 14: Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_5 do modelo polinomial de 1° ordem com sazonalidade.

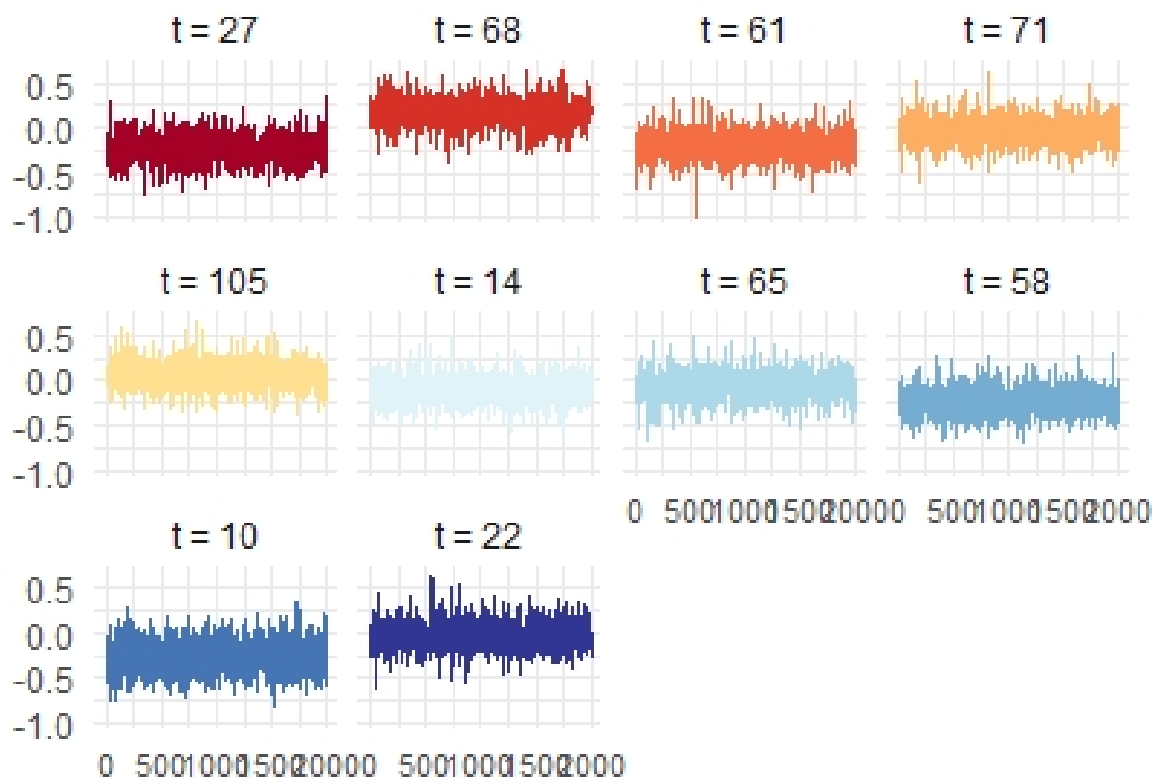


Figura 15: Sorteio de 10 tempos dentre $t = 1, \dots, 120$ para o γ_9 do modelo polinomial de 1° ordem com sazonalidade.