Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Métodos alternativos de previsão de safras agrícolas

Daniel Lima Miquelluti

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Estatística
e Experimentação Agronômica

Piracicaba
2015

Daniel Lima Miquelluti
Engenheiro Agrônomo

Métodos alternativos de previsão de safras agrícolas

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **VITOR AUGUSTO OZAKI**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2015

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Miquelluti, Daniel Lima

Métodos alternativos de previsão de safras agrícolas / Daniel Lima Miquelluti. - - versão
revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.
86 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Modelos ARIMA 2. Modelos Lineares Dinâmicos 3. Inferência Bayesiana 4. Previsão
de safra agrícola I. Título

CDD 630.2195
M669m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

Aos meus pais, David e Jucléia, pelo apoio incondicional nessa jornada.

AGRADECIMENTOS

Ao Prof. Dr. Vitor Augusto Ozaki, professor e orientador, pelo seu apoio e atenção dispensada no decorrer deste trabalho. À sua forma amiga, exigente e crítica, fundamental contribuição no meu crescimento enquanto pesquisador.

Aos meus pais, David e Jucléia, pelo apoio, aconselhamento e compreensão ao longo desse caminho, bem como pela sólida formação que me foi dada.

Agradeço a todos professores do mestrado, amigos e funcionários do departamento de Ciências Exatas da ESALQ/USP, pelo auxílio e atenção ao longo deste período.

Obrigado aos colegas do GESER e de mestrado, em especial aos amigos: Bruna, Douglas (Maringá), Erasnilson (Mirtão), Gislaine, Otávio, Patrícia, Rick, Simone e Valiana.

À CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, que me concedeu uma bolsa durante a realização deste mestrado, auxílio financeiro que contribuiu para viabilização deste trabalho.

Por fim, deixo aqui minha sincera gratidão a todas as pessoas que, direta ou indiretamente, contribuíram para a concretização deste trabalho.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Modelos de Predição de Produção Agrícola	19
2.2 Modelos de Séries Temporais - ARIMA	20
2.3 Modelos Dinâmicos Bayesianos	21
2.4 Inferência Bayesiana	23
2.4.1 Distribuição a priori	24
2.4.2 Monte Carlo com Cadeias de Markov	27
3 METODOLOGIA	31
3.1 Descrição dos dados	31
3.2 Modelos ARIMA	31
3.2.1 Estacionariedade, Invertibilidade e Transformação da série	34
3.2.2 Identificação do modelo	37
3.2.3 Estimação de Parâmetros pelo método da Máxima Verossimilhança	38
3.2.4 Estimação de parâmetros utilizando inferência bayesiana	39
3.2.5 Diagnóstico	42
3.2.6 Previsão	43
3.2.7 Comparação de Modelos	45
3.3 Modelos Lineares Dinâmicos	45
3.3.1 Estimação recursiva e previsão dos estados	48
3.3.2 Estimação pelo método da Máxima Verossimilhança	53
3.3.3 Estimação por inferência Bayesiana	54
3.3.4 Diagnóstico	57
3.3.5 Regressão Dinâmica	57

3.3.6	Modelos Polinomiais de ordem n	58
3.3.7	Modelos de séries temporais	58
3.3.8	Combinação de Modelos	60
3.4	Comparação de Metodologias	61
4	RESULTADOS E DISCUSSÃO	63
4.1	Diagnóstico da série	63
4.2	Modelos ARIMA	63
4.3	Modelos Lineares Dinâmicos	65
4.4	Comparação dos Modelos	71
4.5	Utilização do Modelo Escolhido	73
5	CONCLUSÃO	75
	REFERÊNCIAS	77
	APÊNDICE	82

RESUMO

Métodos alternativos de previsão de safras agrícolas

O setor agrícola é, historicamente, um dos pilares da economia brasileira, e apesar de ter sua importância diminuída com o desenvolvimento do setor industrial e de serviços ainda é responsável por dar dinamismo econômico ao país, bem como garantir a segurança alimentar, auxiliar no controle da inflação e na formação de reservas monetárias. Neste contexto as safras agrícolas exercem grande influência no comportamento do setor e equilíbrio no mercado agrícola. Foram desenvolvidas diversas metodologias de previsão de safra, sendo em sua maioria modelos de simulação de crescimento. Entretanto, recentemente os modelos estatísticos vem sendo utilizados mais comumente devido às suas previsões mais rápidas em períodos anteriores à colheita. No presente trabalho foram avaliadas duas destas metodologias, os modelos ARIMA e os Modelos Lineares Dinâmicos (MLD), sendo utilizada tanto a inferência clássica quanto a bayesiana. A avaliação das metodologias deu-se por meio da análise das previsões dos modelos, bem como da facilidade de implementação e poder computacional necessário. As metodologias foram aplicadas a dados de produção de soja para o município de Mamborê-PR, no período de 1980 a 2013, sendo área plantada (ha) e precipitação acumulada (mm) variáveis auxiliares nos modelos de regressão dinâmica. Observou-se que o modelo ARIMA (2,1,0) reparametrizado na forma de um MLD e estimado por meio de máxima verossimilhança, gerou melhores previsões do que aquelas obtidas com o modelo ARIMA(2,1,0) não reparametrizado.

Palavras-chave: Modelos ARIMA; Modelos Lineares Dinâmicos; Inferência bayesiana;
Previsão de safra agrícola

ABSTRACT

Alternative Crop Prediction Methods

The agriculture is, historically, one of Brazil's economic pillars, and despite having its importance diminished with the development of the industry and services it still is responsible for giving dynamism to the country inland's economy, ensuring food security, controlling inflation and assisting in the formation of monetary reserves. In this context the agricultural crops exercise great influence in the behaviour of the sector and agricultural market balance. Diverse crop forecast methods were developed, most of them being growth simulation models, however, recently the statistical models are being used due to its capability of forecasting early when compared to the other models. In the present thesis two of these methodologies were evaluated, ARIMA and Dynamic Linear Models, utilizing both classical and bayesian inference. The forecast accuracy, difficulties in the implementation and computational power were some of the characteristics utilized to assess model efficiency. The methodologies were applied to Soy production data of Mamborê-PR, in the 1980-2013 period, also noting that planted area (ha) and cumulative precipitation (mm) were auxiliary variables in the dynamic regression. The ARIMA(2,1,0) reparametrized in the DLM form and adjusted through maximum likelihood generated the best forecasts, followed by the ARIMA(2,1,0) without reparametrization.

Keywords: ARIMA Models; Dynamic Linear Models; Bayesian Inference; Crop Forecast

LISTA DE FIGURAS

Figura 1 - Estrutura de independência condicional de um MLD	45
Figura 2 - Produção total de Soja (ton) 2a, Área plantada (ha) 2b e Precipitação acumulada (mm) 2c no município de Mamborê-PR	63
Figura 3 - Gráficos de Autocorrelação (a) e Autocorrelação Parcial (b) da série de produção total diferenciada	64
Figura 4 - QQ-plot dos resíduos do modelo (a) e Valores Preditos (b) e intervalos de confiança (5%) para o modelo ARIMA(2,1,0)	65
Figura 5 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado via inferência bayesiana	66
Figura 6 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado na forma de um MLD	67
Figura 7 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado na forma de um MLD bayesiano	67
Figura 8 - Valores do coeficiente β_1 (a) e intervalos de confiança (5%), QQ-plot dos resíduos (b) e Valores Preditos (c) e intervalos de confiança (5%) para o modelo de regressão linear dinâmica	68
Figura 9 - Valores do coeficiente β_1 (a) e intervalos de confiança (5%), QQ-plot dos resíduos (b) e Valores Preditos (c) e intervalos de confiança (5%) para o modelo de regressão linear dinâmica bayesiano	69
Figura 10 - Valores do coeficiente β_1 (a) e β_2 (b) com seus respectivos intervalos de confiança (5%) para o modelo de regressão múltipla dinâmica	70
Figura 11 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo de regressão múltipla dinâmica	70
Figura 12 - Valores do coeficiente β_1 (a) e β_2 (b) com seus respectivos intervalos de confiança (5%) para o modelo de regressão múltipla dinâmica bayesiano	71
Figura 13 - QQ-plot dos resíduos(a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo de regressão múltipla dinâmica bayesiano	72

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas da Produção total de Soja (ton), Área plantada (ha) e Precipitação acumulada (mm) no município de Mamborê-PR	31
Tabela 2 - Produção total de Soja (ton), Área plantada (ha) e Precipitação acumulada (mm) no município de Mamborê-PR	32
Tabela 3 - Coeficientes e respectivos intervalos de confiança (95%) do modelo ARIMA(2,1,0) ajustado	64
Tabela 4 - Estatísticas resumo e quantis para o modelo ARIMA(2,1,0) ajustado	65
Tabela 5 - Coeficientes e respectivos intervalos de confiança (95 %) do modelo ARIMA(2,1,0) ajustado na forma de um MLD	66
Tabela 6 - Estatísticas resumo e quantis para o modelo ARIMA(2,1,0) ajustado na forma de um MLD bayesiano	66
Tabela 7 - Erro Quadrático Médio para os modelos ajustados	72

1 INTRODUÇÃO

O setor agrícola é, historicamente, um dos pilares da economia brasileira, e apesar de ter sua importância diminuída com o desenvolvimento do setor industrial e de serviços é historicamente responsável por dar dinamismo econômico ao país, bem como garantir a segurança alimentar, auxiliar no controle da inflação e na formação de reservas monetárias (SCHNEIDER; FERREIRA; ALVES, 2013). Neste contexto as safras agrícolas exercem grande influência no comportamento do setor e equilíbrio no mercado agrícola.

Políticas públicas e decisões que irão afetar o setor como um todo são tomadas baseadas nas produções das principais culturas, soja, milho, arroz, trigo, café, laranja, para citar algumas. Numa tentativa de se antecipar o rendimento das culturas foram feitos estudos visando gerar projeções destas safras, para que tais decisões possam ser tomadas com a devida antecedência e da forma mais acertada possível (GAO *et al.*, 2010; CAPA-MOROCHO; RODRÍGUEZ-FONSECA; RUIZ-RAMOS, 2014). A tomada de decisões pressupõe que se conheça a área objeto, além da disponibilidade de métodos adequados que consigam prever, com margem de incerteza conhecida, os cenários futuros. Previsões de safra fornecem estimativas da produtividade de culturas agrícolas e auxiliam no processo de tomada de decisão quanto à definição de políticas de armazenagem e de comercialização, implicando em menores riscos de mercado e, em última instância, favorecendo à segurança alimentar.

Diversas metodologias de previsão de safra foram desenvolvidas, sendo em sua maioria modelos de simulação de crescimento. Geralmente, estes modelos são de natureza ecofisiológica e são baseados em processos que simulam os efeitos do clima, do solo e do manejo sobre a produtividade das culturas. Recentemente, tem sido utilizados modelos estatísticos, de maneira mais frequente, considerando-se que as suas predições são obtidas mais facilmente e de forma mais rápida, otimizando a tomada de decisão. Estes métodos estatísticos, no entanto, possuem características muito diferentes, gerando projeções, muitas vezes, não convergentes. Como exemplo, cita-se o estudo desenvolvido pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), (DOSSA, 2014). Nele são efetuadas projeções para diversas culturas num período de onze anos. Entretanto não se faz uma análise retrospectiva da acurácia das projeções, comparativamente a estudos anteriores conduzidos pelo mesmo órgão; bem como a me-

metodologia estatística é descrita de maneira sucinta. Evidencia-se, deste modo, incerteza sobre a qualidade e o desempenho destes modelos quanto à função que lhes é incumbida. Portanto o presente trabalho apresenta uma descrição detalhada das metodologias utilizadas, além de medidas de precisão e acurácia dos resultados obtidos. Este cenário possibilita a utilização cotidiana destes modelos, na prática, tomando como base o seu desempenho em situações reais.

Neste trabalho foi conduzida uma revisão dos métodos estatísticos de previsão de safras agrícolas, bem como o ajuste e discussão de dois deles, a saber: Modelos ARIMA e Modelos Lineares Dinâmicos. Foram discutidas as características gerais e peculiares de cada metodologia bem como se realizou um estudo comparativo do desempenho destes modelos quanto à predição de safra de soja no município de Mamborê-PR. Todos os procedimentos desenvolvidos foram implementados computacionalmente utilizando-se o pacote estatístico R e o software OpenBugs.

2 REVISÃO BIBLIOGRÁFICA

2.1 Modelos de Predição de Produção Agrícola

As primeiras tentativas de se modelar o crescimento e a produção agrícola remontam a década de 60 (WIT, 1966), época na qual se iniciaram os estudos em modelos de balanço hídrico e de taxa fotossintética. Embora o procedimento descrito pelo autor supracitado não seja um modelo de crescimento propriamente dito, este lançou a fundação para o desenvolvimento destes modelos. Posteriormente, na década de 70, foi desenvolvido o modelo conhecido como simulador de crescimento de safra elementar (ELCROS - Elementary Crop growth Simulator) por Wit *et al.* (1970), com complementações feitas por Vries, Brunsting e Laar (1974). Nos anos subsequentes foram feitas novas adições ao modelo, como a quantificação da energia necessária para o crescimento e manutenção da cultura e uma elaboração detalhada da micrometeorologia da lavoura. No modelo *MICROWEATHER* descrito em Goudriaan (1977), o microclima foi explicado como uma função de propriedades das plantas e do solo, assim como das condições climáticas prevalentes acima do dossel. Deste modo a transpiração e fotossíntese eram direcionadas pelas condições climáticas dentro e acima do dossel. Com o avanço dos estudos em micrometeorologia e a quantificação da resistência do dossel às trocas gasosas pôde-se incorporar variáveis climáticas e melhorar a simulação de transpiração, dando origem ao modelo BACROS (*BA*sic *CR*Op growth Simulator) no final da década de 70 (WIT, 1978). Este modelo simula o crescimento e a respiração de lavouras na fase vegetativa em condições de produção potencial; foi desenvolvido para gramíneas com parâmetros específicos para definir a espécie cujo crescimento se deseja modelar. Durante a década de 80 foi iniciado um projeto de pesquisa na área pela agência IBSNAT (*International Benchmark Sites Network for Agrotechnology Transfer*), sendo que este foi uma tentativa de demonstrar a efetividade de se estudar opções de plantio através de análises de sistemas e simulações. Os propósitos definidos neste projeto eram de entender os processos e mecanismos do ecossistema, e à partir destes ter a capacidade de prever os resultados para cada opção de plantio. Este projeto teve como fruto um sistema de suporte de decisão agrícola DSSAT (*Decision Support System for Agrotechnology Transfer*), sendo que dentro deste sistema estão incorporados os modelos de simulação de safra.

Desde então foram desenvolvidos diversos modelos que fazem interface com o DSSAT, dentre os quais modelos estocásticos, que necessitam de uma quantidade menor de informações, bem como possibilitam estimativas de produção antes do período de colheita. No Brasil o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) faz projeções de produção agrícola para um período de onze anos por meio de três destes modelos estocásticos, sejam eles modelos de suavização exponencial, modelos arima e modelos de espaço de estados, tomando como base um período de 37 anos. Neste trabalho, serão analisadas e comparadas duas destas metodologias, os modelos arima e os modelos dinâmicos.

2.2 Modelos de Séries Temporais - ARIMA

Os modelos ARIMA de séries temporais foram popularizados com o trabalho de Box & Jenkins em 1970 (BOX; JENKINS, 1970), aonde metodologias de análise de séries temporais foram organizadas, dando origem a estes modelos. A metodologia descrita e o processo iterativo de identificação, estimação e verificação de modelos foi amplamente aplicada como pode-se ver nos trabalhos de Caprio *et al.* (1983), Pflaumer (1992), Bianchi, Jarrett e Choudary (1998), Preez e Witt (2003), dentre outros. Entretanto embora amplamente utilizado faltava rigor matemático ao método de seleção de modelos, dado que esta seleção era feita de forma subjetiva, ficando a cargo do julgamento do pesquisador. A partir desta constatação, foram sugeridos diversos métodos para a seleção de modelos, dentre os quais o critério de Akaike (AIC), o critério de informação de Bayes (BIC), e um dos mais utilizados atualmente, o erro percentual absoluto médio (MAPE). Além disso a forma de identificação da estacionariedade da série, necessária para a utilização da metodologia Box-Jenkins, não possuía um critério com o rigor estatístico necessário. Este critério foi implementado por Dickey e Fuller (1979), sendo revisado e atualizado no trabalho de Xiao e Phillips (1998).

Outro ponto que pode levar a diferenças nas previsões são os métodos de estimação dos parâmetros do modelo (BOX; JENKINS; REINSEL, 2013). Apesar destes métodos serem equivalentes assintoticamente, suas propriedades em amostras finitas são diferentes, o que pode resultar em diferenças significativas nas estimativas. No trabalho de Newbold, Agiakloglou e Miller (1994) pode-se encontrar uma comparação dos métodos de estimação, sendo a máxima verossimilhança o método recomendado.

Mais recentemente Kim (2003) mostrou que o método bootstrap para correção de viés dos parâmetros leva a melhores previsões, no caso de pequenas amostras.

Além destes fatores, quando se trabalha com séries temporais multivariadas a identificação da ordem da função de transferência torna-se complexa e ineficiente quando se adota a metodologia tradicional. Na tentativa de solucionar este problema Edlund (1984) propôs a utilização de um estimador de regressão viesado para a estimação dos pesos das respostas aos impulsos.

Com a popularização da utilização de intervalos de confiança e densidades probabilísticas nas previsões dos modelos, gerou-se a necessidade de uma melhor avaliação dos mesmos, sendo o trabalho de Diebold, Gunther e Tay (1997) responsável pela introdução do método "probability integral transform", utilizado na avaliação de uma densidade univariada. Esforços na melhoria das previsões desta classe de modelos também podem ser encontrados na literatura, Pascual, Romo e Ruiz (2004) por exemplo, aplicam a metodologia bootstrap para a obtenção destes intervalos, indicando sua superioridade quando da incorporação da incerteza relativa à estimação e seleção do modelo.

Alguns exemplos de aplicações na predição de produção podem ser vistas em Muhammad, Javed e Bashir (1992), Boken (2000), Awal e Siddique (2011), Padhan (2012) e outros.

2.3 Modelos Dinâmicos Bayesianos

A introdução dos modelos dinâmicos bayesianos deu-se em 1976 com o trabalho de Harrison e Stevens (1976), no qual foram definidas a estrutura e a aplicação desta classe de modelos, sendo a metodologia documentada em West e Harrison (1997). A ideia foi a de se utilizar a inferência bayesiana para a obtenção de previsões de séries temporais, adotando-se um modelo linear que se "atualizava" com o passar do tempo, por meio de um filtro de Kalman para as relações de recorrência. A matriz de variâncias e covariâncias da equação que relaciona os espaços de estado no tempo devia ser obtida de modo subjetivo pelo pesquisador. Neste aspecto, o trabalho de Ameen e Harrison (1985) é importante uma vez que introduziu o conceito de fatores de desconto, facilitando a obtenção dessas matrizes. Uma das grandes vantagens da aplicação desta metodologia no estudo de séries temporais é a possibilidade de se avaliar

constantemente o desempenho do modelo, por meio do método descrito em West e Harrison (1986), onde se utiliza o erro de previsão padronizado um passo a frente, que permite a identificação de mudanças de nível e de variância na série.

Outra extensão importante da metodologia de modelos dinâmicos ocorreu no trabalho de Migon e Harrison (1985) no qual foi desenvolvida a ideia de modelos dinâmicos não lineares aplicados ao estudo do impacto de propagandas televisivas. Uma aplicação diferente do método pode ser encontrada em Migon e Gamerman (1993) aonde se estudam modelos de crescimento. Utilizando-se a ideia de modelos lineares generalizados introduzida em Nelder e Wedderburn (1972), foi discutida a aplicação da metodologia para dados modelados por meio de distribuições pertencentes à família exponencial (WEST; HARRISON; MIGON, 1985).

O caso de várias séries evoluindo concomitantemente no tempo foi estudada em Gamerman e Migon (1993) gerando uma importante extensão da metodologia, os modelos dinâmicos hierárquicos. Uma equação estrutural é introduzida, além das já utilizadas equações de observação e evolução. Este arcabouço teórico permitiu o estudo de séries temporais multivariadas, como pode ser observado em Landim e Gamerman (2000).

Um problema inerente à metodologia é a inferência sobre os parâmetros, dado que na maioria dos casos a distribuição à posteriori não possui forma fechada. Para tanto pode-se utilizar o método MCMC conforme pode-se observar em Gamerman (1998), Carter e Kohn (1994), Shephard e Pitt (1997).

Entretanto, devido à necessidade de previsões em curtos prazos ou quase instantaneamente, gerou-se a necessidade da otimização da metodologia de inferência. Para tanto métodos sequenciais foram propostos, dentre eles podem-se citar aqueles descritos em Liu e West (2001), Gordon, Salmond e Smith (1993), Polson, Stroud e Müller (2002).

Extensões adicionais à metodologia podem ser encontradas em Ghil *et al.* (1981) no estudo de modelos espaço-temporais, Basseville *et al.* (1992), Chou, Willsky e Benveniste (1994) para modelos multi escala. Aplicações à estimação de produção agrícola podem ser encontradas em Kleshchenko, Goncharova e Naidina (2012), Dillon (2011), Sun (2000).

2.4 Inferência Bayesiana

A inferência bayesiana pode ser descrita como a “atualização” da incerteza a respeito de alguma quantidade de interesse quando da obtenção de novos dados ou resultados, sendo esta atualização feita pelo teorema de Bayes

$$p(\theta|Y) = \frac{f(Y|\theta)p(\theta)}{f(Y)}$$

Onde $f(Y|\theta)$ é a função de verossimilhança, $p(\theta)$ representa a informação a priori sobre a quantidade de interesse e $f(Y)$ é a densidade marginal de Y . A distribuição a posteriori $p(\theta|Y)$ representa o conhecimento adquirido a partir da obtenção dos dados, sendo o principal elemento da análise bayesiana, dado que à partir dela se faz a inferência sobre a quantidade de interesse. Esta seção é baseada nos trabalhos de Bernardo e Smith (2009) e Box e Tiao (2011).

Definição 1 *O valor que minimiza a perda esperada com relação à distribuição a posteriori é definido como estimador de Bayes, denotado por $\delta_B(Y)$*

$$\delta_B(Y) = \min_{\delta(Y)} \int L(\theta, \delta(\theta))p(\theta|Y)d\theta = E_{p(\theta|Y)} \{[L(\theta, \delta(Y))]\}$$

em que $\delta(Y)$ é um estimador de θ , $L(\theta, \delta(\theta))$ é a função que determina o cálculo da perda, $p(\theta|Y)$ é a distribuição a posteriori.

A definição 1 implica que o estimador de Bayes estará condicionado a posteriori e a função de perda utilizada. As duas funções de perda mais utilizadas são:

- i Perda quadrática: $L(\theta, \delta(Y)) = [\theta - \delta(Y)]^2$

Resolvendo-se a equação apresentada na definição com o uso da função perda quadrática chega-se a conclusão que o estimador de Bayes é $\delta_B(Y) = E_{p(\theta|Y)}[\theta]$, ou seja, a média da distribuição a posteriori.

- ii Perda absoluta: $L(\theta, \delta(Y)) = |\theta - \delta(Y)|$

A solução para o estimador neste caso é a mediana da distribuição a posteriori.

iii Perda 0 - 1:

$$L(\theta, \delta(\theta)) = \begin{cases} 0, & \text{se } \theta = \delta(Y) \\ 1, & \text{caso contrário} \end{cases}$$

O estimador de Bayes para este caso é a maior moda da distribuição a posteriori.

A variância a posteriori do estimador se dá por:

$$V_\delta(Y) = E_{p(\theta|Y)} \{[\theta - \delta(Y)]^2\} = V(\theta|Y) + [E_{p(\theta|Y)} - \delta(Y)]^2$$

O intervalo de credibilidade (ICr) é definido como

Definição 2 *O intervalo de credibilidade de 95% $ICr_{95\%}$ para θ é o intervalo delimitado pelos percentis 2,5% ($\theta_{[2,5\%]}$) e 97,5% ($\theta_{[97,5\%]}$) da distribuição a posteriori $p(\theta|Y)$ para θ .*

2.4.1 Distribuição a priori

A distribuição a priori reflete a incerteza sobre a quantidade de interesse antes da coleta dos dados. Serão apresentadas algumas formas de especificação desta distribuição.

Priori não informativa

Quando o conhecimento sobre o fenômeno em estudo é vago ou inexistente a distribuição a priori deve refletir esta ignorância. É caracterizada por total ou mínima quantidade de informação, não privilegiando qualquer valor do parâmetro.

- **Método de Bayes-Laplace**

O primeiro impulso na geração de distribuições não informativas foi o Princípio da Razão Insuficiente, elucidado por Bayes e Laplace. Este princípio traduz a ignorância a respeito do parâmetro por meio da equiprobabilidade. Seja $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$, finito, a distribuição não informativa gerada é a distribuição Uniforme Discreta

$$p(\theta) \propto c, \quad \theta \in \Theta$$

Onde c é uma constante.

No caso em que Θ é um conjunto enumerável, temos que a distribuição a priori é imprópria,

$$\int_{\Theta} p(\theta) d\theta = \infty$$

dado que não se garante a equiprobabilidade dos valores de θ . Sendo Θ infinito e não numerável, o Princípio da Razão Insuficiente conduz à distribuição Uniforme Contínua, a qual é imprópria se Θ não for limitado. Deste modo existe a necessidade de verificar se a distribuição a posteriori, obtida com o uso de priors impróprias, é própria, antes que se possa fazer qualquer inferência.

• Método de Jeffreys

Jeffreys utilizou-se da informação esperada de Fisher sobre $\theta \in \mathbb{R}$ para a geração de uma priori não informativa

$$I(\theta) = E \left[\left(\frac{d \log L(X|\theta)}{d\theta} \right)^2 \middle| \theta \right] = -E \left[\frac{d^2 \log L(X|\theta)}{d\theta^2} \middle| \theta \right]$$

Seja $\psi = g(\theta)$ e $\theta = g^{-1}(\psi)$, tem-se que

$$\begin{aligned} I(\psi) &= -E \left[\frac{d^2 \log L(X|\psi)}{d\psi^2} \right] \\ &= -E \left[\frac{d^2 \log L(X|\theta = g^{-1}(\psi))}{d\theta^2} \right] \left| \frac{dg^{-1}(\psi)}{d\psi} \right|^2 \\ &= I(\theta) \left| \frac{d\theta}{d\psi} \right|^2 \end{aligned} \tag{1}$$

Como $h(\psi) = h[\theta(\psi)] \left| \frac{d\theta}{d\psi} \right|$, tirando-se a raiz quadrada de 1 e comparando-se com esta última igualdade tem-se:

$$h(\theta) \propto |I(\theta)|^{1/2}$$

Definição 3 *Seja X uma observação com função densidade de probabilidade*

$h(X|\theta)$. A priori de Jeffreys tem função densidade de probabilidade dada por

$$h(\theta) \propto |I(\theta)|^{1/2}$$

Caso Θ seja um vetor multiparamétrico $\theta = \theta_1, \theta_2, \dots, \theta_k$, sua matriz de informação de Fisher é $I(\Theta) = \begin{pmatrix} I_{11}(\Theta) & I_{12}(\Theta) & \dots & I_{1k}(\Theta) \\ \vdots & \vdots & \vdots & \vdots \\ I_{k1}(\Theta) & I_{k2}(\Theta) & \dots & I_{kk}(\Theta) \end{pmatrix}$, sendo a priori não informativa de Jeffreys dada por $\pi(\Theta) \propto [|\det \mathbf{I}(\Theta)|]^{1/2}$. Supondo-se independência a priori entre os parâmetros $\theta_i, i = 1, 2, \dots, k$ tem-se

$$h(\theta) = \left[\prod_{i=1}^k |\text{diag}[I(\Theta)]| \right]^{1/2}$$

Priori Conjugada

Uma densidade a priori é dita conjugada com a função de verossimilhança quando a distribuição a posteriori resultante é da mesma família de distribuições da priori. Este fato reflete a situação em que há algum conhecimento sobre as quantidades de interesse do modelo em estudo.

Definição 4 *Seja $F = p(X|\theta), \theta \in \Theta$ uma classe de distribuições amostrais, então uma classe de distribuições P é conjugada a F se*

$$\nabla p(X|\theta) \in F \quad e \quad p(\theta) \in P \Rightarrow p(\theta|X) \in P$$

Uma família de distribuições conjugadas pode ser obtida considerando-se:

- i Identificar a classe P de distribuições para θ tal que $l(\theta; X)$ seja proporcional a um membro desta classe;*
- ii Verificar se P é fechada por amostragem, isto é, se:*

$$\nabla p_1, p_2 \in P \exists k \text{ tal que } kp_1p_2 \in P$$

Se além disso existir uma constante k tal que $k^{-1} = \int l(\theta; x)d\theta < \infty$ e todo $p \in P$ é

definido como $p(\theta) = kl(\theta; X)$ então P é a família conjugada natural ao modelo amostral gerador de $l(\theta; X)$. Uma forma alternativa de conjugação a priori são as priors conjugadas da família exponencial, descritas em Diaconis, Ylvisaker et al. (1979).

2.4.2 Monte Carlo com Cadeias de Markov

Metropolis -Hastings

O algoritmo de Metropolis-Hastings é um método algoritmo que permite a obtenção de amostras de uma densidade de forma indireta, quando a geração de amostras dessa densidade é impossibilitada devido ao conhecimento incompleto da distribuição a posteriori. O método consiste na escolha de um núcleo $q(\cdot|\theta)$, sendo gerados valores a partir deste núcleo e filtrados utilizando-se um critério de aceitação, assim garantindo a representatividade da amostra gerada. Seja $\pi(\theta)$ a densidade de probabilidade conjunta de interesse, em que $\theta = (\theta_1, \dots, \theta_k)$, com densidades condicionais a posteriori disponíveis e definidas por $\pi_i(\theta_i) = \pi_i(\theta_i|\theta_{-i}), i = 1, 2, \dots, k$.

Passo 1 Inicializar o contador de iterações da cadeia em $j = 1$ e escolher os valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)});$$

Passo 2 Gerar um novo valor θ^* da distribuição $q(\cdot|\theta)$;

Passo 3 Calcular a probabilidade de aceitação $\alpha(\theta, \theta^*)$ e gerar $u \sim U(0, 1)$;

Passo 4 Caso $u \leq \alpha$, aceitar o novo valor e definir $\theta^{j+1} = \theta^*$, caso contrário deve-se rejeitar o valor gerado e definir $\theta^{j+1} = \theta$;

Passo 5 Atualizar o contador de j para $j + 1$ e repetir o passo 2 até a convergência.

Onde o novo valor de θ^* é aceito com probabilidade dada por

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)} \right\}$$

A densidade de probabilidade $\pi(\theta)$ pode ser parcialmente conhecida, não sendo alterada a não ser por uma constante. É comum a permanência da cadeia no mesmo estado por várias iterações, sendo uma boa prática a monitoração do percentual de aceitação de novos valores. A verificação de convergência da cadeia pode ser feita por meio de

técnicas gráficas como: gráficos de séries temporais, onde a série deve se apresentar em uma faixa e sem a presença de tendências ou oscilações bruscas; gráficos de autocorrelações, onde a série não deve apresentar alta correlação, com exceção ao primeiro valor do gráfico que representa defasagem zero. Uma avaliação mais formal de convergência é feita pela utilização de técnicas numéricas, sendo os diagnósticos descritos por Gelman e Rubin (1992), Geweke (1991) os mais utilizados.

Amostrador de Gibbs

Em muitos casos a obtenção da densidade a posteriori analiticamente é impossível, e conseqüentemente a geração de amostras desta. No entanto, caso as densidades condicionais apresentem uma forma conhecida, o método de amostragem de Gibbs é uma alternativa. A metodologia consiste em um algoritmo de amostragem de uma cadeia de Markov cuja matriz de transição é formada pelas distribuições condicionais completas, podendo-se assim realizar a geração de distribuições que são aproximações das distribuições condicionais completas. Seja $\pi(\theta)$ a densidade de probabilidade conjunta de interesse com $\theta = (\theta_1, \dots, \theta_k)$ e supondo-se que as densidades condicionais completas a posteriori $\pi_i(\theta_i) = \pi_i(\theta_i|\theta_{-i})$, $i = 1, 2, \dots, k$ sejam conhecidas. O algoritmo é definido por:

Passo 1 Inicializar o contador de iterações da cadeia em $j = 1$ e escolher os valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)});$$

Passo 2 Obter um novo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$ à partir de $\theta^{(j-1)}$ por meio de gerações de valores sucessivas

$$\begin{aligned} \theta_1^j &\sim (\theta_1|\theta_2^{j-1}, \theta_3^{j-1}, \dots, \theta_k^{j-1}) \\ \theta_2^j &\sim (\theta_2|\theta_1^j, \theta_3^{j-1}, \dots, \theta_k^{j-1}) \\ &\vdots \\ \theta_k^j &\sim (\theta_k|\theta_1^j, \theta_2^j, \dots, \theta_{k-1}^j) \end{aligned}$$

Passo 3 Atualizar o contador de j para $j + 1$ e repetir o passo 2 até a convergência.

Pode-se avaliar a convergência da cadeia por meio dos mesmos métodos empregados

no caso do algoritmo Metropolis-Hastings.

3 METODOLOGIA

3.1 Descrição dos dados

Os dados utilizados neste trabalho consistiram em séries históricas de produção de soja (kg), área plantada de soja (ha) e precipitação acumulada nos meses de outubro a janeiro (mm) para o município de Mamborê, PR (tabelas 1 e 2). As séries compreendem o período entre os anos de 1980 e 2013 e foram obtidas junto ao Instituto Paranaense do Desenvolvimento Econômico e Social. Optou-se por adotar a série de produção municipal em detrimento da nacional devido a possibilidade da utilização de variáveis explicativas climáticas, que não estavam disponíveis no segundo caso, no nível de detalhamento que se necessitava. No presente trabalho, foram consideradas apenas as variáveis área e precipitação acumulada. A variável precipitação acumulada foi obtida somando-se as precipitações mensais dos meses de outubro a janeiro, sendo esta metodologia adotada de modo a simular um ciclo de cultivo de soja considerando-se a época de plantio recomendada para o município de Mamborê (AGRICULTURA, 2014).

Tabela 1 - Estatísticas descritivas da Produção total de Soja (ton), Área plantada (ha) e Precipitação acumulada (mm) no município de Mamborê-PR

Variável	Mínimo	Mediana	Média	Máximo	Desvio Padrão
Produção	111200,0	140400,0	133600,0	205600,0	33333,4
Área	49420,0	53000,0	51340,0	58000,0	5462,6
Precipitação	657,3	726,6	764,1	1082,0	169,2

Fonte: IPARDES/AGUASPARANA

3.2 Modelos ARIMA

De modo a se estudar os modelos de séries temporais deve-se primeiro definir uma série temporal. De acordo com Morettin e Toloi (2006) define-se.

Definição 5 *Seja T um conjunto arbitrário. Um processo estocástico é uma família $Z = Z(t), t \in T$, de tal modo que para cada $t \in T, Z(t)$ seja uma variável aleatória.*

Supondo-se que estas variáveis estejam definidas num mesmo espaço de probabilidades $(\Omega, \mathcal{A}, \mathcal{P})$, $Z(t)$ é função de dois argumentos, $Z(t, \omega), t \in T, \omega \in \Omega$. Para cada $\omega \in \Omega$

Tabela 2 - Produção total de Soja (ton), Área plantada (ha) e Precipitação acumulada (mm) no município de Mamborê-PR

Ano	Produção Total (ton)	Area (ha)	Precipitação Acumulada (mm)
1980	141636	58000	680,7
1981	117992	56000	694,2
1982	102300	55000	779,9
1983	78520	39000	1077,9
1984	70122	43500	765,9
1985	92708	44000	682,3
1986	51552	32000	479,8
1987	120588	51300	647,5
1988	141360	57000	673
1989	101500	58000	718,7
1990	108300	57000	652,4
1991	105840	50400	530,3
1992	112925	47250	611,5
1993	110685	47000	734,5
1994	124800	48000	664,7
1995	131040	48000	777,5
1996	139830	49200	989,1
1997	135074	50100	1081,5
1998	148800	53400	862,2
1999	151262	53190	606,5
2000	148558	53150	466,7
2001	156370	52200	1038,3
2002	155250	54000	950,4
2003	167472	54060	988,5
2004	157080	54400	716,3
2005	143100	54000	867,3
2006	135572	52300	832,8
2007	167229	53150	948,2
2008	163862	52050	576,5
2009	140958	52500	675,4
2010	177200	52860	1046,4
2011	205600	57000	753,7
2012	145652	53300	754,8
2013	190921	53200	654,9

Fonte: IPARDES/AGUASPARANA

fixado, obtem-se uma função de t , definida como série temporal. Assim, pode-se passar a definição de modelos ARIMA(p,d,q).

Conforme encontrado em Box, Jenkins e Reinsel (2013) temos que um modelo ARIMA(p,d,q), constituído de um processo regressivo de ordem p , um processo de médias móveis de ordem q e um número de diferenciações d , denotados por AR(p)

e MA(q) respectivamente, pode ser escrito na forma:

$$\psi(B)e_t = \phi(B)\nabla^d Z_t + \theta(B)a_t \quad t = 1, 2, \dots, T \quad (2)$$

em que:

- $$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned}$$
1. B é o operador translação para o passado
 2. ψ são pesos da série
 3. ϕ é o vetor de parâmetros autorregressivos
 4. θ é o vetor de parâmetros de médias móveis
 5. e_t são choques aleatórios
 6. p representa o grau do processo autoregressivo
 7. q representa o grau do processo de médias móveis
 8. Z_t representa a série temporal no instante t
 9. $\phi(B)$ é um operador autorregressivo; assume-se que este seja estacionário, ou seja, as raízes de $\phi(B) = 0$ estão fora do círculo unitário
 10. $\psi(B) = \phi(B)\nabla^d$ é um operador autorregressivo generalizado; assume-se que este seja um operador não-estacionário com um número d de raízes de $\psi(B) = 0$ igual a unidade, ou seja, d raízes unitárias. Em que d é o número de defasagens da série
 11. $\theta(B)$ é um operador de médias móveis; assume-se que este seja invertível, ou seja, as raízes de $\theta(B) = 0$ estão fora do círculo unitário

3.2.1 Estacionariedade, Invertibilidade e Transformação da série

No campo de estudos de séries temporais, uma das suposições básicas trata da estacionariedade da série. Entretanto a maioria das séries encontradas na prática não possuem tal comportamento, apresentando não estacionariedade explosiva ou não estacionariedade não explosiva, sendo a última denominada homogênea. Este comportamento não estacionário se traduz na presença de tendências tanto crescentes quanto decrescentes e/ou variância não constante. Cabe dizer que a série pode ser estacionária em algum subperíodo específico, mesmo não o sendo no decorrer de todo o período em estudo. Formalizando-se a definição de estacionariedade temos a estacionariedade forte (estrita):

Definição 6 *Um processo estocástico $\{Z_t; t \in T\}$, é estacionário no sentido forte (ou estritamente estacionário) se $F_{Z_{t_1}, \dots, Z_{t_n}} = F_{Z_{t_1+\tau}, \dots, Z_{t_n}}(Z_1, \dots, Z_n)$ para todo $t_1 < t_2 < \dots \leq t_n \in T, n \geq 1$ e para todo $\tau \in T$, ou seja, se todas as distribuições finito-dimensionais $F_{Z_{t_1+\tau}, \dots, Z_{t_n}}(Z_1, \dots, Z_n) = P(Z_{t_1}(w) \leq Z_1, \dots, Z_{t_n}(w) \leq Z_n)$, permanecem as mesmas sob translações no tempo.*

Esta definição se mostra de difícil aplicação, portanto de maneira geral se define estacionariedade de modo menos restrito:

Definição 7 *Um processo estocástico $\{Z_t; t \in T\}$, é fracamente estacionário (ou estacionário de segunda ordem) se $E(Z_t) = \mu$, $Var(Z_t) = \sigma^2$ e $Cov(Z_t, Z_{t+k}) = \gamma(k)$, para todo $t \in T$ e $k \in \mathbb{N}$.*

Definição 8 *Um processo estocástico $\{Z_t; t \in T\}$, é Gaussiano se, para qualquer conjunto $t_1 < t_2 < \dots \leq t_n \in T, n \geq 1$ as variáveis aleatórias $Z_1, \dots, Z_n$ têm distribuição Normal n -variada.*

Utilizando-se o teorema de Wald (BRUSS; ROBERTSON, 1991) definimos estacionariedade para modelos lineares estacionários:

Teorema 1 *Se Z_t é um processo estocástico, então:*

$$Z_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j} = \mu + \psi(b)e_t$$

em que $\psi(b)$ é um polinômio definido por $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$, com $\psi_0 = 1$ constante de e_t .

Sendo e_t , com $t = 1, 2, \dots, n$, variáveis aleatórias independentes e identicamente distribuídas (iid), normalmente distribuídas com média zero e variância constante σ_a^2 , ou seja, $a_t \stackrel{iid}{\sim} N(0, \sigma_a^2)$, então neste caso, $\{e_t, t \geq 0\}$, é denominado ruído branco. Caso as raízes do polinômio $\psi(B) = 0$ estejam dentro ou sobre o círculo de raio unitário o processo $Z_t = \psi(B)e_t$ é considerado estacionário no sentido fraco. Se e_t é um processo i.i.d. $N(0, \sigma_a^2)$, a condição mencionada anteriormente garante a estacionariedade no sentido forte para o processo. Considere $\{Z_t, t \in T \subseteq Z\}$ uma série temporal, dado o Teorema de Wald, o modelo linear estacionário para Z_t é considerado estacionário se $|\psi(B)| < \infty$ para $|B| \leq 1$, com $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\psi_0 = 1$. O modelo ARIMA(p, d, q), dado em (2), é estacionário se o seu componente AR(p) é estacionário, ou seja, as raízes de $\phi(B) = 0$ estão fora do círculo de raio unitário. De modo a verificar a estacionariedade reescrevamos (2) na forma $Z_t = \psi(B)e_t$, em que $\psi(B) = \phi(B)^{-1}\theta(B)$. Entretanto $\psi(B) = \phi(B)^{-1}\theta(B) = \prod_{i=1}^p (1 - G_i B)^{-1}\theta(B)$, com $\phi(B) = 0$ tem raízes $G_i^{-1}, i = 1, 2, \dots, p$ e $\phi(B) = \prod_{i=1}^p (1 - G_i B)$, então fazendo-se a decomposição em frações parciais de $\psi(B)$:

$$\psi(B) = \left(\sum_{i=1}^p c_i (1 - G_i B)^{-1} \right) \theta(B) \implies |\psi(B)| \leq \left(\sum_{i=1}^p |c_i| \cdot |(1 - G_i B)^{-1}| \right) |\theta(B)|$$

para $|B| \leq 1$. Todavia,

$$|(1 - G_i B)^{-1}| \leq 1 + |G_i|B + |G_i|^2 B^2 + \dots \leq \sum_{j=0}^{\infty} |G_i|^j < \infty, \text{ se } |G_i| < 1$$

Deste modo, $|\psi(B)| < \infty$ para $|B| \leq 1$ se $|G_i| < 1$.

Definição 9 Um processo linear será invertível se $\pi(B)$ convergir para $|B| \leq 1$, em que $\pi(B) = \phi(B)\theta(B)^{-1}$.

O modelo ARIMA(p, q) dado em (2), é invertível se o seu componente MA(q) é invertível, ou seja, as raízes de $\theta(B) = 0$ estão fora do círculo de raio unitário. Reescrevendo (2) na forma $\pi(B)y_t = a_t$, em que $\pi(B) = \phi(B)\theta(B)^{-1} = (\prod_{i=1}^q (1 - H_i B)^{-1}\phi(B))$, em que H_i^{-1} são as raízes de $\phi(B) = 0, i = 1, 2, \dots, q$. Temos por decomposição em frações

parciais que

$$\pi(B) = \left(\sum_{i=1}^q c_i (1 - H_i B)^{-1} \right) \phi(B)$$

sendo $|\phi(B)| < \infty$ para $|B| \leq 1$. Tem-se $|\pi(B)| < \infty$ para $|B| \leq 1$ se $|H_i| < 1$.

No presente trabalho a estacionariedade dos processos em estudo é verificada por meio do teste de Dickey-Fuller Aumentado (DICKEY; FULLER, 1979):

$$\Delta Z_t = \alpha_0 + \delta Z_{t-1} + \gamma t + \beta_1 \Delta Z_{t-1} + \beta_2 \Delta Z_{t-2} + \cdots + \beta_p \Delta Z_{t-p} + \epsilon_t$$

Em que:

1. α é uma constante (*drift*)
2. γ é uma tendência temporal
3. δ é o coeficiente que representa a raiz do processo
4. β são parâmetros da regressão
5. ϵ_t são resíduos independentes e identicamente distribuídos
6. δZ_t é a série diferenciada

Sendo a hipótese nula:

$$H_0 : \delta = 0$$

A série precisa ser diferenciada para se tornar estacionária.

Contra:

$$H_1 : \delta < 0$$

A série possui tendência estacionária e precisa apenas de uma tendência temporal presente no modelo de regressão. No caso de não rejeição de H_0 aplicam-se diferenciações a série original de modo que esta se torne estacionária, podendo assim ser modelada conforme um modelo ARMA(p, q). Ainda há de se verificar a estacionariedade quanto

ao comportamento da variância da série, pois caso não haja homocedasticidade deve-se encontrar alguma forma de acomodar esta variância não constante no modelo ou transformar a série de maneira que ela passe a ter homogeneidade de variância. Uma transformação comumente utilizada é a introduzida por Box e Cox (1964), conforme definição de Wei (1994):

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda} \quad (3)$$

Para citar uma vantagem desta transformação, pode-se tratar λ como um parâmetro de transformação e realizar sua estimação a partir dos dados. A título de exemplo pode-se incluir λ como um parâmetro no modelo, $(1 - \phi_1 B - \dots - \phi_p B^p)(Z_T^{(\lambda)} - \mu) = (1 - \theta_1 B - \dots - \theta_q B^q)a_t$ sendo escolhido o valor de λ que resulta em menor erro quadrático médio residual.

$$Z_t^\lambda = \frac{Z_t^{\lambda-1}}{\lambda(\tilde{Z})^{\lambda-1}}$$

Em que:

$$\tilde{Z} = \left(\prod_{t=1}^n Z_t \right)^{1/n}$$

É a média geométrica dos dados, proveniente do jacobiano da transformação. Portanto, para $\lambda = 0$, o erro quadrático médio residual é computado do modelo ajustado:

$$Z_t^0 = \lim_{\lambda \rightarrow 0} Z_t^\lambda = (\ln Z_t) \tilde{Z}$$

3.2.2 Identificação do modelo

A identificação de modelos candidatos, deve ser feita através da análise das funções de autocorrelação e autocorrelação parcial por meio da plotagem dos valores de r_j obtidos nas equações:

$$r_j = \frac{\sum_{t=j+1}^T (Z_t - \bar{Z})(Z_{t-j} - \bar{Z})}{\sum_1^T (Z_t - \bar{Z})^2} \quad \forall j \quad j = 0, 1, \dots, t$$

A significância destas autocorrelações é testada individualmente por:

$$|r_j| > 1,96/\sqrt{T}$$

Em que 1,96 é o percentil da distribuição normal padrão correspondente a 95% de confiança e T é uma estimativa do desvio padrão de r_j . Para um conjunto M de correlações, pela utilização da estatística de Ljung-Box:

$$Q_{LB} = T(T+2) \sum_{j=1}^M \frac{r_j^2}{T-j}$$

Com $Q_{LB} \sim \chi_{(M)}^2$.

O coeficiente de autocorrelação parcial pode ser obtido utilizando-se:

$$Z_t = \alpha + \phi_{j1}Z_{t-1} + \phi_{j2}Z_{t-2} + \dots + \phi_{jj}Z_{t-j} + e_t$$

Em que ϕ_{jj} o coeficiente de autocorrelação parcial, possuindo as mesmas características do coeficiente de autocorrelação simples.

De posse dos gráficos de autocorrelação simples e parcial pode-se fazer a escolha de modelos candidatos a modelar a série em estudo. Sendo que um modelo candidato AR deverá apresentar FAC com decaimento exponencial e FACP diferente de zero nas primeiras p defasagens e um modelo MA uma FAC com valores diferentes de zero nas primeiras q defasagens. Um modelo ARMA irá apresentar FAC com decaimento exponencial ou em senoides amortecidas após a defasagem $q-p$ (MORETTIN; TOLOI, 2006).

3.2.3 Estimação de Parâmetros pelo método da Máxima Verossimilhança

Considerando-se a função de probabilidade conjunta de $a_t \stackrel{iid}{\sim} N(0, \sigma^2)$, ruído branco proveniente de um modelo ARMA(p,q), conforme demonstrado em Wei (1994) tem sua distribuição conjunta dada por:

$$P(\mathbf{a}|\phi, \mu, \boldsymbol{\theta}, \sigma_a^2) = \frac{1}{(2\pi\sigma_a^2)^{n/2}} \exp \left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^n a_t^2 \right]$$

Onde:

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + Z_t - \phi_1 Z_{t-1} + \cdots + \phi_q Z_{t-q}$$

Partindo-se destes resultados pode-se escrever a função de verossimilhança dos parâmetros $\{\phi, \mu, \theta, \sigma_a^2\}$. Se $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ e assumindo que $\mathbf{Z}_* = (Z_{1-p}, \dots, Z_{-1}, Z_0)$ e $\mathbf{a}_* = ((a_{1-q}, \dots, a_{-1}, a_0)')$ são conhecidos. O logaritmo da função de verossimilhança é

$$\ln L_*(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\phi, \mu, \theta)}{2\sigma_a^2} \quad (4)$$

Onde

$$\mathbf{S}_*(\phi, \mu, \theta) = \sum_{t=1}^n a_t^2(\phi, \mu, \theta | \mathbf{Z}_* \mathbf{a}_* \mathbf{Z})$$

É a função de soma de quadrados condicional. As quantidades de $\hat{\phi}$, $\hat{\mu}$ e $\hat{\theta}$ que maximizam a função 4 são os estimadores condicionais de máxima verossimilhança. Especificando-se $\mathbf{Z}_*$ e $\mathbf{a}_*$ como estacionário e uma série de variáveis aleatórias i.i.d $N(0, \sigma_a^2)$, respectivamente, pode-se substituir Z_t pela média amostral $\bar{Z}$ e a_t pelo seu valor esperado zero. Deste modo a soma de quadrados condicional torna-se:

$$\mathbf{S}_*(\phi, \mu, \theta) = \sum_{t=1}^n a_t^2(\phi, \mu, \theta | \mathbf{Z})$$

Após a obtenção de ϕ, μ e θ a estimativa $\hat{\sigma}_a^2$ de σ_a^2 é dada por:

$$\hat{\sigma}_a^2 = \frac{\mathbf{S}_*(\phi, \mu, \theta)}{g.l.}$$

Onde g.l. é o número de graus de liberdade utilizados na soma de $\mathbf{S}_*(\phi, \mu, \theta)$ menos o número de parâmetros estimados.

3.2.4 Estimação de parâmetros utilizando inferência bayesiana

Seja $\mathbf{S} = (Z_1, Z_2, \dots, Z_n)'$ o vetor de observações dado por um processo ARMA(p,q), pode-se reescrever 2 condicionando-se nas $r = \max(p, q)$ primeiras obser-

vações e obter os últimos $n - r$ erros

$$e_t = Z_t - \sum_{i=1}^p \phi_i Z_{t-i} + \sum_{j=1}^q \theta_j e_{t-j}$$

com $t = r + 1, r + 2, \dots, n$ e assumindo-se que $e_p = e_{p-1} = e_{p-2} = \dots = e_{i-q} = 0$. Partindo-se da pressuposição que os erros e_t são independentes e identicamente distribuídos com $N(0, \tau)$, em que $\tau = 1/\sigma^2$, tem-se uma amostra aleatória de tamanho $n - r$ de uma distribuição normal, $at \stackrel{iid}{\sim} N(0, \tau)$, com função de verossimilhança aproximada dada por

$$\begin{aligned} L(\phi, \theta, \tau | \mathbf{S}) &= \prod_{t=r+1}^n \sqrt{\frac{\tau}{2\pi}} \exp \left\{ -\frac{\tau}{2} a_t^2 \right\} \\ &\propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n a_t^2 \right\} \\ &\propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} \right]^2 \right\} \end{aligned}$$

Reescrevendo-se na forma vetorial com $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_p)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$, $\mathbf{W}_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$ e $\mathbf{b}_t = (a_{t-1}, a_{t-2}, \dots, a_{t-1})'$, para $t = r + 1, \dots, n$ temos

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau, | \mathbf{W}_t, \mathbf{b}_t) \propto \tau^{\frac{n-r}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \mathbf{W}_t' \boldsymbol{\phi} - \mathbf{b}_t' \boldsymbol{\theta} \right]^2 \right\} \quad (5)$$

em que

$$\begin{aligned} (y_t - \mathbf{W}_t' \boldsymbol{\phi} - \mathbf{b}_t' \boldsymbol{\theta})^2 &= (y_t - \mathbf{W}_t' \boldsymbol{\phi} - \mathbf{b}_t' \boldsymbol{\theta})' (y_t - \mathbf{W}_t' \boldsymbol{\phi} - \mathbf{b}_t' \boldsymbol{\theta}) \\ &= [(y_t - \mathbf{W}_t' \boldsymbol{\phi}) - (\mathbf{b}_t' \boldsymbol{\theta})'] (y_t - \mathbf{W}_t' \boldsymbol{\phi} - \mathbf{b}_t' \boldsymbol{\theta}) \\ &= y_t^2 - y_t \mathbf{W}_t' \boldsymbol{\phi} - y_t \mathbf{b}_t' \boldsymbol{\theta} - \boldsymbol{\phi}' \mathbf{W}_t y_t + \boldsymbol{\phi}' \mathbf{W}_t \mathbf{W}_t' \boldsymbol{\phi} + \boldsymbol{\phi}' \mathbf{W}_t \mathbf{b}_t' \boldsymbol{\theta} - \\ &\quad - \boldsymbol{\theta}' \mathbf{b}_t \mathbf{W}_t' \boldsymbol{\phi} + \boldsymbol{\theta}' \mathbf{b}_t \mathbf{b}_t' \boldsymbol{\theta} \end{aligned}$$

Utilizando-se densidades a priori não informativas assumindo que $\boldsymbol{\phi}, \boldsymbol{\theta}$ e τ são independentes tal que $\pi_0(\boldsymbol{\phi}) \propto \text{const.}$, $\pi_0(\boldsymbol{\theta}) \propto \text{const.}$ e $\pi_0(\tau) \propto \tau^{-1}$ (priori de Jeffreys).

Portanto, a priori conjunta para ϕ, θ e τ é

$$\begin{aligned}\pi_0(\phi, \theta, \tau) &= \pi_0(\phi)\pi_0(\theta)\pi_0(\tau) \\ &\propto \tau^{-1}\end{aligned}$$

Combinando-se a função de verossimilhança, dada em 5, com a priori conjunta, expressa na última igualdade, temos a distribuição a posteriori

$$\pi(\phi, \theta, \tau) \propto \tau^{\frac{n-r}{2}-1} \exp \left\{ -\frac{\tau}{2} \sum_{t=r+1}^n \left[y_t - \mathbf{W}_t' \phi - \mathbf{b}_t' \theta \right]^2 \right\}$$

Sendo as distribuições condicionais

- Para τ

$$\tau | \phi, \theta, \mathbf{W}_t, \mathbf{b}_t \sim \text{Gamma} \left(\frac{n-r}{2}, \frac{1}{2} \sum_{t=r+1}^n \left[y_t - \mathbf{W}_t' \phi - \mathbf{b}_t' \theta \right]^2 \right)$$

- Para ϕ

Considerando-se

$$(\phi - \mu_\phi)' \sum (\phi - \mu_\phi) = \phi' \sum \phi - \phi' \sum \mu_\phi - \mu_\phi' \sum \phi + \mu_\phi' \mu_\phi \quad (6)$$

e seja

$$\begin{aligned}S^2(\phi | \theta) &\propto -y_t \mathbf{W}_t' \phi - \phi' \mathbf{W}_t y_t + \phi' \mathbf{W}_t \mathbf{W}_t' \phi + \phi' \mathbf{W}_t \mathbf{b}_t' \theta - \theta' \mathbf{b}_t \mathbf{W}_t' \phi \\ &\propto \phi' \mathbf{W}_t \mathbf{W}_t' \phi - \phi' (\mathbf{W}_t y_t - \mathbf{W}_t \mathbf{b}_t' \theta) - (y_t \mathbf{W}_t' - \theta' \mathbf{b}_t \mathbf{W}_t') \phi\end{aligned}$$

Comparando-se $S^2(\phi | \theta)$ com 6 tem-se

$$\begin{aligned}\Sigma_\phi &= \sum_{t=r+1}^n (\mathbf{W}_t \mathbf{W}_t'), \text{ de ordem } p \times p, \\ \mu_\phi &= \left[\sum_{t=r+1}^n (\mathbf{W}_t \mathbf{W}_t') \right]^{-1} \left[\sum_{t=r+1}^n (\mathbf{W}_t y_t - \mathbf{W}_t \mathbf{b}_t' \theta) \right], \text{ de ordem } p \times 1.\end{aligned}$$

Portanto,

$$\phi | \theta, \tau, \mathbf{W}_t, \mathbf{b}_t \propto N_p(\mu_\phi, \Sigma_\phi^{-1}) \quad (7)$$

- Para θ

Utilizando-se raciocínio análogo a densidade condicional para ϕ tem-se

$$\theta | \phi, \tau, \mathbf{W}_t, \mathbf{b}_t \propto N_q(\mu_\theta, \Sigma_\theta^{-1}) \quad (8)$$

com

$$\Sigma_\theta = \sum_{t=r+1}^n (\mathbf{b}_t \mathbf{b}_t'), \text{ de ordem } q \times q, \quad (9)$$

$$\mu_\theta = \left[\sum_{t=r+1}^n (\mathbf{b}_t \mathbf{b}_t') \right]^{-1} \left[\sum_{t=r+1}^n (\mathbf{b}_t y_t - \mathbf{b}_t \mathbf{W}_t' \phi) \right], \text{ de ordem } q \times 1. \quad (10)$$

3.2.5 Diagnóstico

Após a estimação dos parâmetros deve-se checar o ajuste do modelo. Isto é feito utilizando-se da premissa básica de que os a_t são ruído branco com média zero e variância constante. Isto pode ser feito através de testes de normalidade, como Shapiro e Wilk (1965), e através do gráfico de resíduos. Para se testar a pressuposição de que os erros são ruído branco, um teste muito utilizado é o teste de Ljung-Box. Este teste considera as FACs (Funções de Autocorrelação) residuais como uma unidade para verificar a hipótese nula:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$$

Com a estatística teste dada por

$$Q = n(n+2) \sum_{k=1}^K (n-k)^{-1} \hat{\rho}_k^2$$

Sendo que $Q \sim \chi^2(K-m)$, com $m = p + q$.

3.2.6 Previsão

De modo a obter as previsões de menor erro quadrático médio vamos reescrever o modelo 2 na forma de choques aleatórios:

$$Z_t = \psi(B)e_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots \quad (11)$$

Onde

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \frac{\theta(B)}{\phi(B)}$$

Com $\psi_0 = 1$. Para $t = n + l$, temos

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j a_{n+l-j} \quad (12)$$

Supondo que no tempo $t = n$ tenha-se as observações $Z_n, Z_{n-1}, Z_{n-2}, \dots$ e se deseja prever o valor futuro Z_{n+l} , l passos a frente, como uma combinação linear das observações $Z_n, Z_{n-1}, Z_{n-2}, \dots$. Utilizando-se de (11) pode-se determinar a previsão $\hat{Z}_n(l)$ de Z_{n+l} , de mínimo erro quadrático médio ser

$$\hat{Z}_n(l) = \psi_l^* a_n + \psi_{l+1}^* a_{n-1} + \psi_{l+2}^* a_{n-2} + \dots$$

Onde há de se determinar os ψ_j^* . O erro quadrático médio da previsão é:

$$E(Z_{n+l} - \hat{Z}_n(l))^2 = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} [\psi_{l+j} - \psi_{l+j}^*]^2$$

Em que este é minimizado quando $\psi_{l+j}^* = \psi_{l+j}$. Utilizando-se (12) e

$$E(a_{n+j} | Z_n, Z_{n-1}, \dots) = \begin{cases} 0, & j > 0, \\ a_{n+j}, & j \leq 0, \end{cases}$$

Temos que

$$E(Z_{n+l} | Z_n, Z_{n-1}, \dots) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \dots$$

Então, a previsão de mínimo erro quadrático médio de Z_{n+l} é dada pela sua esperança condicional:

$$\widehat{Z}_n(l) = E(Z_{n+l} | Z_n, Z_{n-1}, \dots)$$

Sendo o erro de previsão dado por:

$$e_n(l) = Z_{n+l} - \widehat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j} \quad (13)$$

Como $E(e_n(l) | Z_t, t \leq n) = 0$, a previsão é não viesada com a variância do erro dada por:

$$Var(e_n(l)) = \sigma_a^2 \sum_{j=1}^{l-1} \psi_j^2$$

Assumindo-se normalidade, os limites de previsão $(1 - \alpha)100\%$ são

$$\widehat{Z}_n(l) \pm N_{\alpha/2} \left[1 + \sum_{j=1}^{l-1} \psi_j^2 \right]^{1/2} \sigma_a$$

em que $N_{\alpha/2}$ é o desvio normal padrão tal que $P(N > N_{\alpha/2}) = \alpha/2$. O erro de previsão demonstrado em 13 é uma combinação linear de choques aleatórios futuros entrando no sistema após o tempo n . O erro de previsão um passo a frente é:

$$e_n(l) = Z_{n+1} - \widehat{Z}_n(l) = a_{n+1}$$

Assim, os erros de previsão um passo a frente são independentes, o que implica que $\widehat{Z}_n(1)$ é de fato a melhor previsão de Z_{n+1} . Entretanto, para tempos maiores os erros de previsão são correlacionados. Por exemplo:

$$e_n(l) = Z_{n+l} - \widehat{Z}_n(l) = a_{n+1} + \psi_1 a_{n+l-1} + \dots + \psi_{l-1} a_{n+1}$$

e

$$e_{n-j}(l) = Z_{n+l-j} - \widehat{Z}_{n-j}(l) = a_{n+l-j} + \psi_1 a_{n+l-j-1} + \dots + \psi_{l-1} a_{n-j+1}$$

que são para o mesmo horizonte l mas partindo de diferentes pontos n e $n - 1$ da série para $j < l$. Sendo que isto se verifica também para erros de diferentes horizontes para mesma origem:

$$Cov[e_n(2), e_n(1)] = E[(a_{n+2} + \psi_1 a_{n+1})(a_{n+1})] = \psi_1 \sigma_a^2$$

3.2.7 Comparação de Modelos

A comparação de modelos ARIMA candidatos foi feita por meio do erro percentual médio absoluto (MAPE):

$$MAPE = \frac{100}{T} \sum_1^T \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right|$$

Em que $\hat{Z}_t$ é o valor previsto da série.

3.3 Modelos Lineares Dinâmicos

Os modelos lineares dinâmicos (MLD) são um caso particular dos modelos de espaço de estado. Os modelos de espaço de estado baseiam-se na ideia de que séries temporais Y_t são funções incompletas e ruidosas de um processo latente não observável, chamado de processo de estado.

Figura 1 - Estrutura de independência condicional de um MLD

Pode-se pensar em (θ_t) como um processo aleatório auxiliar que facilita a tarefa de especificação da lei de probabilidade da série temporal; o processo de estados latente (θ_t) tem uma dinâmica Markoviana mais simples e pode-se assumir que a observação Y_t somente depende do estado do sistema no instante em que a medida é tomada, θ_t . De modo mais formal, as pressuposições de um modelo de espaço de

estados são:

A.1 $(\theta_t, t = 0, 1, \dots)$ é uma cadeia de Markov; isso é, θ_t depende dos valores passados $\theta_0, \theta_1, \dots, \theta_{t-1}$ somente através de θ_{t-1} . Assim, a lei de probabilidade do processo $(\theta_t, t = 0, 1, \dots)$ é especificada designando-se a densidade inicial $p_0(\theta_0)$ de θ_0 e as densidades de transição $p(\theta_t|\theta_{t-1})$ de θ_t condicionais a θ_{t-1} .

A.2 Condicionalmente a $(\theta_t, t = 0, 1, \dots)$, os Y_t 's dependem somente de θ_t . Segue que para qualquer $n \geq 1$, $(Y_1, \dots, Y_n|\theta_1, \dots, \theta_n)$ tem densidade condicional conjunta $\prod_{t=1}^n p(Y_t|\theta_t)$.

As pressuposições supracitadas e a especificação das densidades relevantes possibilita a escrita da lei de probabilidade do processo aleatório conjunto $((\theta_t, Y_t), t = 1, 2, \dots)$, da qual se pode deduzir as dependências entre as variáveis. Deve-se notar que a estrutura demonstrada na figura 1 tem de ser considerado como não definido, ou seja, as setas podem apontar para ambas direções. A partir de 1 pode-se verificar que para qualquer $n \geq 1$ que

$$\begin{aligned} (\theta_0, \theta_1, \dots, \theta_n, Y_1, \dots, Y_n) &\sim p_0(\theta_0) \prod_{t=1}^n p(\theta_t, Y_t|\theta_0, \theta_1, \dots, \theta_{t-1}, Y_1, \dots, Y_{t-1}) \\ &= p_0(\theta_0) \prod_{t=1}^n f(Y_t|\theta_0, \dots, \theta_t, Y_1, \dots, Y_{t-1}) p(\theta_t|\theta_0, \dots, \theta_{t-1}, Y_1, \dots, Y_{t-1}) \\ &= p_0(\theta_0) \prod_{t=1}^n f(Y_t|\theta_t) p(\theta_t|\theta_{t-1}) \end{aligned}$$

Percebe-se que Y_t é condicionalmente independente de observações passadas $(Y_1, \dots, Y_{t-1})$ dado o valor de θ_t .

A classe de modelos de espaço de estados lineares e gaussianos constituem os modelos lineares dinâmicos (WEST; HARRISON, 1997). Estes, são definidos por:

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

Em que:

1. $\mathbf{F}_t$ é uma matriz $(n \cdot r)$ conhecida;
2. $\mathbf{G}_t$ é uma matriz $(n \cdot n)$ conhecida;

3. $\mathbf{V}_t$ é uma matriz $(r \cdot r)$ conhecida;
4. $\mathbf{W}_t$ é uma matriz $(n \cdot n)$ conhecida;

Sendo as matrizes de observação e evolução do modelo definidas por:

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + \mathbf{e}_t \quad \mathbf{e}_t \sim N[\mathbf{0}, \mathbf{V}_t] \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]\end{aligned}\tag{14}$$

Em que:

1. $\mathbf{F}_t$ é a matriz de delineamento do modelo.
2. $\boldsymbol{\theta}_t$ é o vetor de parâmetros.
3. $\boldsymbol{\mu}_t = \mathbf{F}_t' \boldsymbol{\theta}_t$ define o nível da série.
4. $\mathbf{e}_t$ é o erro amostral com matriz de variâncias $\mathbf{V}_t$.
5. $\mathbf{G}_t$ é a matriz de evolução do sistema.
6. $\boldsymbol{\omega}_t$ é o erro de evolução com matriz de variância de evolução $\mathbf{W}_t$.

Sendo que a equação 14 é chamada de equação de observação e a subsequente de equação do sistema. Além disso, é assumido que $\boldsymbol{\theta}_0$ segue distribuição Normal

$$\boldsymbol{\theta}_0 \sim N_p(\mathbf{m}_0, \mathbf{C}_0)$$

para um vetor de médias não aleatório m_0 e matriz de variância de evolução C_0 , e é independente em $(\mathbf{e}_t)$ e $(\boldsymbol{\omega}_t)$.

Destaca-se que para modelos clássicos de séries temporais define-se $\mathbf{F}_t = \mathbf{F}$ e $\mathbf{G}_t = \mathbf{G}, \forall t$, sendo que para modelos clássicos de regressão linear temos $\mathbf{G}_t = \mathbf{I}_p$ e $\mathbf{V}_t = \mathbf{0}, \forall t$

3.3.1 Estimação recursiva e previsão dos estados

Uma das grandes dificuldades no processo de aplicação da metodologia aqui descrita é a especificação do modelo, pois nem sempre se pode obter uma interpretação dos estados e conhecimento de suas probabilidades de transição. De modo a determinar as equações básicas de recursão para estimação e previsão será considerado que as densidades $p(\mathbf{Y}_t|\boldsymbol{\theta}_t)$ e $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ foram especificadas. Para se estimar o vetor de estados são calculadas as densidades condicionais $p(\boldsymbol{\theta}_t|\mathbf{Y}_1, \dots, \mathbf{Y}_t)$, sendo que nos MLD o filtro de Kalman fornece as fórmulas de atualização de inferência sobre o estado atual do vetor, de modo que se pode passar de $p(\boldsymbol{\theta}_t|\mathbf{Y}_1, \dots, \mathbf{Y}_t)$ para $p(\boldsymbol{\theta}_{t+1}|\mathbf{Y}_1, \dots, \mathbf{Y}_{t+1})$. Entretanto, na análise de séries temporais a previsão é o maior objetivo, portanto a estimação dos estados é apenas um passo para que se possa prever o valor de futuras observações. Para previsões um passo a frente, isto é, prever $\mathbf{Y}_{t+1}$ baseado em $\mathbf{Y}_1, \dots, \mathbf{Y}_t$, deve-se primeiramente estimar o valor de $\boldsymbol{\theta}_{t+1}$ do vetor de estados, para que posteriormente se possa computar $\mathbf{Y}_{t+1}$. A densidade preditiva do estado um passo a frente é $p(\boldsymbol{\theta}_{t+1}|\mathbf{Y}_1, \dots, \mathbf{Y}_t)$, sendo baseada na densidade de filtragem de $\boldsymbol{\theta}_t$. A partir disto pode-se obter a densidade preditiva para um passo a frente $f(\mathbf{Y}_{t+1}|\mathbf{Y}_1, \dots, \mathbf{Y}_t)$. De modo a estimar as densidades de filtragem denotemos D_t a informação dada pelas primeiras t observações, $\mathbf{Y}_1, \dots, \mathbf{Y}_t$. Assumindo-se A.1 e A.2, a densidade filtrada e preditiva pode ser obtida por meio de um algoritmo recursivo. Iniciando-se por $\boldsymbol{\theta}_0 \sim p_0(\boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_0|D_0)$ pode-se obter, para $t = 1, 2, \dots$:

- a densidade preditiva para um passo a frente dado D_{t-1} , baseado na densidade de filtragem $p(\boldsymbol{\theta}_{t-1}|D_t)$ e no modelo de transição. Denotando-se $\boldsymbol{\theta}_{t+1} \Pi(\mathbf{Y}_1, \dots, \mathbf{Y}_t)|\boldsymbol{\theta}_t$;

$$\begin{aligned} p(\boldsymbol{\theta}_t|D_{t-1}) &= \int p(\boldsymbol{\theta}_t \boldsymbol{\theta}_{t-1}|D_{t-1})d(\boldsymbol{\theta}_{t-1}) \\ &= \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})p(\boldsymbol{\theta}_{t-1}|D_{t-1})d(\boldsymbol{\theta}_{t-1}) \\ &= \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1}|D_{t-1})d(\boldsymbol{\theta}_{t-1}) \end{aligned}$$

- a densidade preditiva para um passo a frente da próxima observação é:

$$\begin{aligned} p(\mathbf{Y}_t|D_{t-1}) &= \int p(\mathbf{Y}_t, \boldsymbol{\theta}_t|D_{t-1})d(\boldsymbol{\theta}_t) \\ &= \int p(\mathbf{Y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_{t-1})d(\boldsymbol{\theta}_t) \\ &= \int p(\mathbf{Y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_{t-1})d(\boldsymbol{\theta}_t) \end{aligned}$$

- a densidade de filtragem $p(\boldsymbol{\theta}_t|D_t)$, utilizando-se da regra de Bayes com $p(\boldsymbol{\theta}_t|D_{t-1})$ como a densidade a priori e a verossimilhança $L(\mathbf{Y}_t|\boldsymbol{\theta}_t)$ pode ser descrita por:

$$p(\boldsymbol{\theta}_t|D_t) = \frac{L(\mathbf{Y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_{t-1})}{L(\mathbf{Y}_t|\boldsymbol{\theta}_t, D_{t-1})} = \frac{L(\mathbf{Y}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t|D_{t-1})}{L(\mathbf{Y}_t|D_{t-1})}$$

Os resultados acima podem ser utilizados recursivamente para se computar as densidades preditivas k passos a frente, iniciando-se por $k = 1$:

$$p(\boldsymbol{\theta}_{t+k}|D_t) = \int p(\boldsymbol{\theta}_{t+k}|\boldsymbol{\theta}_{t+k-1})p(\boldsymbol{\theta}_{t+k-1}|D_t)d(\boldsymbol{\theta}_{t+k-1})$$

e

$$p(\mathbf{Y}_{t+k}|D_t) = \int p(\mathbf{Y}_{t+k}|\boldsymbol{\theta}_{t+k})p(\boldsymbol{\theta}_{t+k}|D_t)d(\boldsymbol{\theta}_{t+k})$$

Os resultados obtidos previamente solucionam os problemas de filtragem e previsão, entretanto deve-se notar que a obtenção das densidades condicionais relevantes não é uma tarefa simples. Para o caso em estudo entretanto, devido a utilização de alguns resultados comuns da distribuição Normal multivariada, pode-se provar que o vetor aleatório $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t, \mathbf{Y}_1, \dots, \mathbf{Y}_t)$ segue distribuição Normal para qualquer $t \geq 1$. Deste modo, as distribuições condicionais e marginais são Normais. A solução para o problema de filtragem se dá pelo filtro de Kalman.

Teorema 2 (Filtro de Kalman) *Para o MLD definido em 14, se*

$$\boldsymbol{\theta}_{t-1}|D_{t-1} \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$$

em que $t \geq 1$, então

- A densidade preditiva para um passo a frente de $\boldsymbol{\theta}_t$ dado D_{t-1} é Normal, com parâmetros:

$$\begin{aligned}\mathbf{a}_t &= E(\boldsymbol{\theta}_t|D_{t-1}) = E(E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})|D_{t-1}) = E(\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}) = \mathbf{G}_t\mathbf{m}_{t-1} \\ \mathbf{R}_t &= \text{Var}(\boldsymbol{\theta}_t|D_{t-1}) = E(\text{Var}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})|D_{t-1}) + \text{Var}(E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})|D_{t-1}) \\ &= \mathbf{W}_t + \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t'\end{aligned}$$

- A densidade preditiva um passo a frente de $\mathbf{Y}_t$ dado D_{t-1} é Normal, com parâmetros:

$$\begin{aligned}\mathbf{f}_t &= E(\mathbf{Y}_t|D_{t-1}) = E(E(\mathbf{Y}_t|\boldsymbol{\theta}_t, D_{t-1})|D_{t-1}) = E(\mathbf{F}_t\boldsymbol{\theta}_t|D_{t-1}) = \mathbf{F}_t\mathbf{a}_t \\ \mathbf{Q}_t &= \text{Var}(\mathbf{Y}_t|D_{t-1}) = E(\text{Var}(\mathbf{Y}_t|\boldsymbol{\theta}_t, D_{t-1})|D_{t-1}) + \text{Var}(E(\mathbf{Y}_t|\boldsymbol{\theta}_t, D_{t-1})|D_{t-1}) \\ &= \mathbf{V}_t + \mathbf{F}_t\mathbf{R}_t\mathbf{F}_t'\end{aligned}$$

- A densidade de filtragem de $\boldsymbol{\theta}_t$ dado D_t é Normal com:

$$\begin{aligned}\mathbf{m}_t &= E(\boldsymbol{\theta}_t|D_t) = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t'\mathbf{Q}_t^{-1}\boldsymbol{\epsilon}_t \\ \mathbf{C}_t &= \text{Var}(\boldsymbol{\theta}_t|D_t) = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t'\mathbf{Q}_t^{-1}\mathbf{R}_t\end{aligned}$$

em que $\boldsymbol{\epsilon}_t = \mathbf{Y}_t - \mathbf{f}_t$ é o erro de previsão.

O filtro de Kalman possibilita a obtenção das densidades de filtragem e preditiva recursivamente, começando com $\boldsymbol{\theta}_0|D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ e obtendo-se $p(\boldsymbol{\theta}_1|D_1)$, procedendo de modo recursivo a medida que se obtém novas informações. A densidade condicional de $\boldsymbol{\theta}_t|D_t$ resolve o problema da filtragem, entretanto, é de interesse a obtenção de uma estimativa pontual. Considerando-se uma função de perda quadrática $L(\boldsymbol{\theta}_t, a) = (\boldsymbol{\theta}_t - a)'\mathbf{H}(\boldsymbol{\theta}_t - a)$, o estimador pontual Bayesiano de $\boldsymbol{\theta}_t$ dado D_t é a esperança condicional $\mathbf{m}_t = E(\boldsymbol{\theta}_t|D_t)$, sendo que a perda mínima esperada corresponde a matriz de covariância condicional $\text{Var}(\boldsymbol{\theta}_t|D_{t-1})$ para $\mathbf{H} = \mathbf{I}_p$. Pode-se notar pela expressão de $\mathbf{m}_t$ que esta apresenta uma forma de correção da estimativa, dado que é igual a média de previsão $\mathbf{a}_t$ mais uma correção que depende do quanto a nova observação difere da

sua previsão. O peso da correção é dado pela matriz de ganho:

$$K_t = \mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1}$$

Assim o peso da informação atual $\mathbf{Y}_t$ depende da matriz de covariância da observação $\mathbf{V}_t$ (por meio de $\mathbf{Q}_t$) e de $\mathbf{R}_t = \text{Var}(\theta_t | D_{t-1}) = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$. Deve-se notar que $\mathbf{m}_t = K_t \mathbf{Y}_t + (1 - K_t) \mathbf{m}_{t-1}$, é uma média ponderada de $\mathbf{Y}_t$ e $\mathbf{m}_{t-1}$, sendo que o peso $K_t = \mathbf{R}_t / \mathbf{Q}_t = (\mathbf{C}_{t-1} + \mathbf{W}) / (\mathbf{C}_{t-1} + \mathbf{W}_t + \mathbf{V}_t)$ da observação atual $\mathbf{Y}_t$ é também chamado de *coeficiente adaptativo*, e é $0 < K_t < 1$. Dado $\mathbf{C}_0$, se a relação $\mathbf{r}_t = \mathbf{W}_t / \mathbf{V}_t$, chamada de razão sinal-ruído, for pequena, K_t é pequeno e $\mathbf{Y}_t$ recebe pouco peso; se $\mathbf{V}_t = \mathbf{0}$, temos que $K_t = 1$ e $\mathbf{m}_t = \mathbf{Y}_t$, isto é, a previsão um passo a frente é dada pela observação mais recente. O cálculo das variâncias a posteriori $\mathbf{C}_t$ (e consequentemente de $\mathbf{R}_t$ e $\mathbf{Q}_t$) usando a fórmula de atualização iterativa contida no teorema 2, por mais simples que possa parecer, sofre de instabilidade numérica, podendo levar a matrizes de variância não simétricas e até mesmo negativas definidas. Um algoritmo mais robusto é o baseado na atualização sequencial da decomposição em valores singulares (DVS) de $\mathbf{C}_t$. De modo estrito, o filtro baseado em DVS pode ser visto como um filtro de raiz quadrada, de fato caso $\mathbf{A} = \mathbf{U} \mathbf{D}^2 \mathbf{U}'$ é a DVS de uma matriz de variâncias, então $\mathbf{U} \mathbf{D}$ é a raiz quadrada de $\mathbf{A}$.

As previsões um passo a frente, tanto de estados $\mathbf{f}_t = E(\boldsymbol{\theta}_t | D_t)$ como as de observações $\mathbf{f}_t = E(\mathbf{Y}_t | D_t)$, são obtidas como um subproduto do filtro de Kalman. Dado que para cada t , a previsão um passo a frente da observação f_t , é uma função linear da média de filtragem $\mathbf{m}_{t-1}$, a magnitude da matriz de ganho tem a mesma função de determinar o quão sensível $\mathbf{f}_t$ será a uma observação não esperada $\mathbf{Y}_{t-1}$, assim como foi para $\mathbf{m}_{t-1}$.

Como foi demonstrado anteriormente, para os MLD o filtro de Kalman fornece a estimativa de filtragem $\mathbf{m}_t$, dada a informação D_t , como a estimativa prévia $\mathbf{m}_{t-1}$ corrigida por um termo que depende dos erros de previsão:

$$\boldsymbol{\epsilon}_t = \mathbf{Y}_t - E(\mathbf{Y}_t | D_{t-1}) = \mathbf{Y}_t - \mathbf{f}_t$$

Estes erros podem ser escritos de modo alternativo, na forma de erros de estimação:

$$\begin{aligned}\epsilon_t &= \mathbf{Y}_t - \mathbf{F}_t \mathbf{a}_t = \mathbf{F}_t \boldsymbol{\theta}_t + \mathbf{e}_t - \mathbf{F}_t \mathbf{a}_t \\ &= \mathbf{F}_t (\boldsymbol{\theta}_t - \mathbf{a}_t) + \mathbf{e}_t = \mathbf{F}_t (\boldsymbol{\theta}_t - \mathbf{G}_t \mathbf{m}_{t-1}) + \mathbf{e}_t\end{aligned}$$

Para a sequência $(\epsilon_t, t \geq 1)$, algumas propriedades interessantes se mantêm.

1. O valor esperado de ϵ_t é zero, dado que $E(\epsilon_t) = E(E(\epsilon_t|D_{t-1})) = \mathbf{0}$.
2. O vetor aleatório ϵ_t é não correlacionado para qualquer função de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$. Em particular, caso $s < t$, então ϵ_t e $\mathbf{Y}_s$ não são correlacionados. Seja $\mathbf{Z} = g(\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1})$. Então

$$\begin{aligned}\text{Cov}(\epsilon_t, \mathbf{Z}) &= E(\epsilon_t \mathbf{Z}) = E(E(\epsilon_t \mathbf{Z} | D_{t-1})) \\ &= E(E(\mathbf{Y}_t - \mathbf{f}_t | D_{t-1}) \mathbf{Z}) = \mathbf{0}\end{aligned}$$

Isto é similar a se dizer que $E(\mathbf{Y}_t | D_{t-1})$ é a projeção ortogonal de $\mathbf{Y}_t$ no espaço vetorial linear de variáveis aleatórias que são funções de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$.

3. Para $s \neq t$, ϵ_s e ϵ_t não são correlacionados. Isto segue de 2 já que, se $s < t$ cada componente de ϵ_s é uma função de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$.
4. ϵ_t é uma função linear de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$. Visto que $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$ seguem uma distribuição conjunta Normal, $E(\mathbf{Y}_t | D_{t-1})$ é uma combinação linear de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$.
5. $(\epsilon_t, t \geq 1)$ é um processo Normal. A partir de 4 segue que, para todo t , $(\epsilon_1, \dots, \epsilon_t)$ é uma combinação linear de $(\mathbf{Y}_1, \dots, \mathbf{Y}_t)$ e portanto segue a distribuição Normal. Como consequência, dado que os ϵ_t 's não são correlacionados por 3, eles também são independentes. Ainda, como $\mathbf{Y}_t | D_{t-1} \sim N_m(\mathbf{f}_t, \mathbf{Q}_t)$, chega-se a $\epsilon_t | D_{t-1} \sim N_m(\mathbf{0}, \mathbf{Q}_t)$. Porém, $\mathbf{Q}_t$ não depende de $\mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}$, assim como a distribuição condicional $N_m(\mathbf{0}, \mathbf{Q}_t)$ também não depende, a qual deve ser também a distribuição não condicional de ϵ_t :

$$\epsilon_t \sim N_m(\mathbf{0}, \mathbf{Q}_t), \quad t = 1, 2, \dots$$

Os erros de previsão ϵ_t também são chamados de *inovações*. A representação $\mathbf{Y}_t = \mathbf{f}_t + \epsilon_t$ justifica essa terminologia, dado que pode-se pensar em $\mathbf{Y}_t$ como a soma de um componente que pode ser previsto por meio de observações passadas, $\mathbf{f}_t$, e outro componente, ϵ_t , que é independente do passado e portanto contém a verdadeira informação trazida pela observação $\mathbf{Y}_t$.

3.3.2 Estimação pelo método da Máxima Verossimilhança

Sejam n vetores aleatórios, $Y_1, \dots, Y_n$, cuja distribuição dependa de um parâmetro desconhecido ψ . A densidade conjunta das observações para um dado valor do parâmetro será denotado por $p(y_1, \dots, y_n; \psi)$. A função de verossimilhança é definida como, até um fator constante, a densidade de probabilidade dos dados observados como uma função de ψ , *i.e.*, denotando-se a verossimilhança por L , pode-se escrever $L(\psi) = \text{const.} \cdot p(Y_1, \dots, Y_n; \psi)$. Para um MLD é conveniente escrever a densidade conjunta das observações na forma

$$p(Y_1, \dots, Y_n; \psi) = \prod_{t=1}^n p(Y_t | D_{t-1}; \psi)$$

em que $p(Y_t | D_{t-1}; \psi)$ é a densidade condicional de Y_t dado a informação até o tempo $t-1$, assumindo-se que ψ é o valor do parâmetro desconhecido. Considerando-se o apresentado anteriormente, percebe-se que os termos do lado direito da equação configuram densidades Normais com média f_t e variância Q_t . Portanto pode-se escrever o logaritmo da função de verossimilhança como

$$\ell(\psi) = -\frac{1}{2} \sum_{t=1}^n \log |Q_t| - \frac{1}{2} \sum_{t=1}^n (y_t - f_t)' Q_t^{-1} (y_t - f_t)$$

em que f_t e Q_t dependem implicitamente de ψ . A expressão acima pode ser maximizada numericamente para se obter o estimador de máxima verossimilhança (EMV) de ψ

$$\hat{\psi} = \max_{\psi} \ell(\psi)$$

Denotando-se por $\mathbf{H}$ a matriz Hessiana de $-\ell(\psi)$, avaliada em $\psi = \hat{\psi}$. A matriz $\mathbf{H}^{-1}$ fornece uma estimativa da variância do EMV, $\text{Var}(\hat{\psi})$. Deve-se entretanto tomar cuidado com a otimização numérica, pois a função de verossimilhança de um MLD possui

vários máximos locais, implicando em diferentes máximos quando se parte de diferentes valores iniciais. Portanto deve-se utilizar diversos valores iniciais e fazer a comparação dos pontos máximos encontrados. Além disso pode-se encontrar o problema de uma verossimilhança plana, onde diferentes valores iniciais podem chegar a um mesmo valor de verossimilhança em diferentes pontos.

3.3.3 Estimação por inferência Bayesiana

A utilização dos EMV's $\hat{\psi}$ como se fossem os verdadeiros valores dos parâmetros na aplicação das filtragens recursivas sofre claramente da dificuldade de tratar apropriadamente da incerteza à respeito de ψ . A abordagem Bayesiana oferece uma formulação mais consistente ao problema. Os parâmetros ψ são considerados como um vetor aleatório. Assume-se que as hipóteses gerais de modelos de estado de espaço para os processos (Y_t) e (θ_t) , A.1 e A.2, mantém-se condicionalmente aos parâmetros ψ . O conhecimento a priori sobre ψ é expresso através de uma lei de probabilidade $p(\psi)$. Assim, para qualquer $n \geq 1$, assume-se que

$$(\theta_0, \theta_1, \dots, \theta_n, Y_1, \dots, Y_n, \psi) \sim p(\theta_0|\psi) \prod_{t=1}^n \ell(Y_t|\theta_t, \psi) p(\theta_t|\theta_{t-1}, \psi)$$

Considerando-se $D_t = (\mathbf{Y}_t, \dots, \mathbf{Y}_t)$, a inferência nos estados e parâmetros desconhecidos é feita pelo cálculo da distribuição a posteriori

$$p(\theta_s, \psi|D_t) = p(\theta|\psi, D_t)p(\psi|D_t)$$

(posteriori marginal, com $s = t$ para filtragem, $s > t$ para previsão de estados), ou a distribuição conjunta a posteriori do estado desconhecido até o tempo t e do parâmetro desconhecido ψ . Será utilizada a notação $\theta_{0:t}$ para denotar o vetor $(\theta_0, \theta_1, \dots, \theta_t)$, $t \geq 0$. Assim, dado a informação D_{t-1} , a distribuição conjunta de interesse a posteriori é

$$p(\theta_{0:t}, \psi|D_{t-1}) = p(\theta_{0:t}|\psi, D_{t-1})p(\psi|D_{t-1}) \quad (15)$$

Os resultados e fórmulas de recursão anteriormente apresentados para estimação e previsão continuam válidos condicionalmente a ψ , e podem ser utilizados no cálculo de

$p(\boldsymbol{\theta}_s|\psi, D_t)$; além disso, podem ser estendidos para a obtenção da densidade conjunta condicional $p(\boldsymbol{\theta}_{0:t}|\psi, D_t)$. Entretanto, eles são ponderados de acordo com a distribuição a posteriori de ψ dado a informação dos dados. Em princípio, a distribuição a posteriori é obtida utilizando-se o teorema de Bayes. Em alguns modelos simples e utilizando-se de prioris conjugadas, pode-se obtê-la na forma fechada. De maneira geral, os cálculos são intratáveis analiticamente, portanto deve-se utilizar metodologias como o amostrador de Gibbs.

Estimação utilizando-se prioris conjugadas

Seja $(Y_t, \boldsymbol{\theta}_t)$, $t \geq 1$ descrito por um MLD com

$$\mathbf{V}_t = \sigma^2 \tilde{\mathbf{V}}_t \quad , \quad \boldsymbol{\omega}_t = \sigma^2 \tilde{\mathbf{W}}_t \quad , \quad \mathbf{C}_0 = \sigma^2 \tilde{\mathbf{C}}_0$$

Assume-se neste caso que todas as matrizes sejam conhecidas, entretanto, o parâmetro de escala σ^2 , é desconhecido. Define-se também $\phi = \sigma^{-2}$ como a precisão. Portanto, a incerteza está nos vetores de estados e no parâmetro ϕ . O MLD fornece a lei de probabilidade condicional de $(Y_t, \boldsymbol{\theta}_t)$ dado ϕ e em particular o modelo assume, para qualquer $t \geq 1$

$$\begin{aligned} Y_t | \boldsymbol{\theta}_t, \phi &\sim N_m(\mathbf{F}_t \boldsymbol{\theta}_t, \phi^{-1} \tilde{\mathbf{V}}_t) \\ \boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \phi &\sim N_p(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \phi^{-1} \tilde{\mathbf{W}}_t) \end{aligned}$$

A escolha natural de distribuições a priori para $(\phi, \boldsymbol{\theta}_0)$ são as prioris hierárquicas

$$\begin{aligned} f(x, \phi | m_0, C_0, \alpha, \beta) &= \frac{\beta^\alpha}{\gamma(\alpha) \sqrt{2\pi C_0}} \phi^{\alpha-1/2} e^{-\beta\phi} e^{-\frac{\phi(x-m_0)^2}{2C_0}} \\ \phi &\sim \Gamma(\alpha_0, \beta_0) \\ \boldsymbol{\theta}_0 | \phi &\sim N(\mathbf{m}_0, \phi^{-1} \tilde{\mathbf{C}}_0) \end{aligned}$$

Tem-se portanto as seguintes fórmulas recursivas para filtragem

- A densidade de previsão um passo a frente de $(\boldsymbol{\theta}_t, \phi) | D_{t-1}$ é uma Gamma-Normal

com parâmetros $(\mathbf{a}_t, \tilde{\mathbf{R}}_t, \alpha_{t-1}, \beta_{t-1})$, em que

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_t \quad , \quad \tilde{\mathbf{R}}_t = \mathbf{G}_t \tilde{\mathbf{C}}_{t-1} \mathbf{G}_t' + \tilde{\mathbf{W}}_t$$

sendo que a densidade condicional preditiva um passo a frente de $Y_t | \phi, D_{t-1}$ é Normal, com parâmetros

$$f_t = \mathbf{F}_t \mathbf{a}_t \quad , \quad Q_t = \phi^{-1} \tilde{Q}_t = \phi^{-1} \mathbf{F}_t \tilde{\mathbf{R}}_t \mathbf{F}_t' + \tilde{V}_t$$

- A distribuição de filtragem de $(\boldsymbol{\theta}_t, \phi | D_t)$ é Gamma-Normal, com parâmetros

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \tilde{\mathbf{R}}_t \mathbf{F}_t \tilde{\mathbf{Q}}^{-1} (Y_t - f_t) & \tilde{\mathbf{C}}_t &= \tilde{\mathbf{R}}_t - \tilde{\mathbf{R}}_t \mathbf{F}_t \tilde{\mathbf{Q}}^{-1} \tilde{\mathbf{R}}_t' \\ \alpha_t &= \alpha_{t-1} + \frac{m}{2} & \beta_t &= \beta_{t-1} + \frac{1}{2} (Y_t - f_t)' \tilde{\mathbf{Q}}^{-1} (Y_t - f_t) \end{aligned}$$

De acordo com o segundo item segue que a densidade marginal de filtragem de $\boldsymbol{\theta}_t | D_t$ é uma Student-t multivariada, e a densidade de σ^2 dado D_t é uma Gamma-Inversa.

Especificação de $\mathbf{W}_t$ utilizando-se fatores de desconto

A estrutura e magnitude da matriz de covariância de estados $\mathbf{W}_t$ tem um papel fundamental na determinação do impacto das informações passadas na estimação de estados e na previsão. Se $\mathbf{W}_t$ é grande, existe grande incerteza na evolução de estados, e muita informação é perdida na passagem de $\boldsymbol{\theta}_{t-1}$ para $\boldsymbol{\theta}_t$. Nas recursões do filtro de Kalman, a incerteza sobre $\boldsymbol{\theta}_{t-1}$ dado D_t é resumida na matriz de covariância condicional $\mathbf{C}_{t-1} = V(\boldsymbol{\theta}_{t-1} | D_{t-1})$, sendo que indo de $\boldsymbol{\theta}_{t-1}$ para $\boldsymbol{\theta}_t$ por meio da equação de estado $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, a incerteza aumenta e tem-se $V(\boldsymbol{\theta}_t | D_{t-1}) = \mathbf{R}_t = \mathbf{G}_t' \mathbf{C}_{t-1} \mathbf{G}_t + \mathbf{W}_t$. Assim, se $\mathbf{W}_t = \mathbf{0}$, i.e., tem-se $\mathbf{R}_t = V(\mathbf{G}_t \boldsymbol{\theta}_{t-1} | D_{t-1}) = \mathbf{P}_t$. Caso $\mathbf{W}_t > 0$, $\mathbf{P}_t$ é aumentada em $\mathbf{R}_t = \mathbf{P}_t + \mathbf{W}_t$. Nesse caso $\mathbf{W}_t$ expressa a perda de informação na passagem de $\boldsymbol{\theta}_{t-1}$ para $\boldsymbol{\theta}_t$ devido ao componente de erro estocástico na evolução de estado, sendo que a perda depende da magnitude de $\mathbf{W}_t$ em relação a $\mathbf{P}_t$. Pode-se portanto expressar $\mathbf{W}_t$ como uma proporção de $\mathbf{P}_t$

$$\mathbf{W}_t = \frac{1 - \delta}{\delta} \mathbf{P}_t$$

Onde $\delta \in (0, 1]$. Segue que $\mathbf{R}_t = 1/\delta \mathbf{P}_t$, com $1/\delta > 1$. O parâmetro δ é chamado de *fator de desconto* dado que ele “desconta” a matriz $\mathbf{P}_t$, com uma evolução de estado determinística, da matriz $\mathbf{R}_t$.

3.3.4 Diagnóstico

No caso univariado, os resíduos, definidos por $\tilde{e}_t = e_t/\sqrt{Q_t}$, são ruído branco Normal, *i.e.* uma sequência de variáveis Normais independentes e identicamente distribuídas com média zero. Esta propriedade pode ser explorada para se chegar as pressuposições do modelo: se modelo é correto, a sequência $\tilde{e}_1, \dots, \tilde{e}_t$ calculada à partir dos dados deve-se parecer com uma amostra de tamanho t de uma distribuição normal padrão. Pode-se utilizar diversos testes para a realização destas verificações, entretanto, pode-se também utilizar formas mais informais, como o *QQ-plot* e o gráfico das FACs dos resíduos. O primeiro é utilizado para se verificar a normalidade, sendo o segundo utilizado na verificação de correlação entre os erros. Pode-se também utilizar o gráfico da série temporal dos erros $\tilde{e}_t$ para determinar a presença de *outliers*, pontos de mudança de comportamento da série e outros padrões não esperados.

3.3.5 Regressão Dinâmica

Uma forma de se tratar os MLDs seria como uma generalização do modelo de regressão linear, permitindo que os coeficientes variem no tempo. Exemplificando com o modelo de regressão linear estático:

$$\begin{aligned} Y_t &= \theta_{1,t} + \theta_{2,t}x_t + e_t, & e_t &\sim N(0, \sigma_t^2) \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t & \boldsymbol{\omega}_t &\sim N_2(\mathbf{0}, \mathbf{W}_t) \end{aligned}$$

Sendo $(Y_t, X_t), t = 1, 2, \dots$ observado ao longo do tempo. Permitindo-se que os parâmetros variem ao longo do tempo pode-se modelar não linearidade na relação entre X e Y , bem como mudanças estruturais no processo. O processo descrito é um MLD com $\mathbf{F}_t = (1, X_t)$ e estados $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{1,t}, \boldsymbol{\theta}_{2,t})'$. No caso de $\mathbf{G}_t = \mathbf{I}$, $\mathbf{I}$ uma matriz identidade, temos $\sigma_t^2 = \sigma^2$ e $\boldsymbol{\omega}_t = \mathbf{0}$ para qualquer tempo t , voltando ao modelo de regressão linear estático.

3.3.6 Modelos Polinomiais de ordem n

O modelo polinomial geral de ordem n tem um espaço de estados n -dimensional e é descrito pelas matrizes:

$$\begin{aligned} \mathbf{F} &= (1, 0, \dots, 0) \\ \mathbf{G} &= \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 1 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix} \\ \mathbf{W} &= \text{diag}(W_1, \dots, W_n) \end{aligned}$$

Em termos de seus componentes, o modelo pode ser escrito na forma

$$\begin{cases} Y_t = \theta_{t,1} + e_t \\ \theta_{t,j} = \theta_{t-1,j} + \theta_{t-1,j+1} + \omega_{t,j} & j = 1, \dots, n-1 \\ \theta_{t,n} = \theta_{t-1,n} + \omega_{t,n} \end{cases}$$

Então, para $j = 2, \dots, n$, o j -ésimo componente do vetor de estados para qualquer tempo t representa, até um erro aleatório, o incremento do componente $(j-1)$ durante o próximo intervalo de tempo, enquanto o primeiro componente representa a resposta média, ou o nível da série. A função de previsão, $f_t(k)$, é um polinômio de grau $n-1$ em k .

3.3.7 Modelos de séries temporais

Existem diversas representações dos modelos ARIMA na forma de MLD, entretanto a aqui apresentada é provavelmente a mais utilizada. Considerando o processo ARMA(p,q) definido em 2, a relação de definição pode ser escrita como

$$Y_t = \sum_{j=1}^r \phi_j Y_{t-j} + \sum_{j=1}^{r-1} \psi_j e_{t-j} + e_t \quad (16)$$

com $r = \max(p, q + 1)$, $\phi_j = 0$ para $j > p$ e $\psi_j = 0$ para $j > q$. Definindo as matrizes:

$$\mathbf{F} = (1, 0, \dots, 0) \quad (17)$$

$$\mathbf{G} = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \dots & 0 & 1 \\ \phi_r & 0 & \dots & 0 & 0 \end{bmatrix} \quad (18)$$

$$\mathbf{R} = [1, \psi_1, \dots, \psi_{r-2}, \psi_{r-1}]' \quad (19)$$

Introduzindo-se um vetor de estado r -dimensional $\theta_t = (\theta_{1,t}, \dots, \theta_{r,t})'$, então o modelo ARMA apresentado possui a seguinte representação como MLD:

$$\begin{cases} Y_t = \mathbf{F}\theta_t \\ \theta_{t+1} = \mathbf{G}\theta_t + \mathbf{R}e_t \end{cases} \quad (20)$$

Este é um MLD com $V = 0$ e $W = \mathbf{R}\mathbf{R}'\sigma^2$, em que σ^2 é a variância da sequência de erros (e_t). De modo a verificar esta igualdade, deve-se notar que a equação observacional fornece $Y_t = \theta_{1,t}$ e a equação de estado é

$$\begin{aligned} \theta_{1,t} &= \phi_1\theta_{1,t-1} + \theta_{2,t-1} + e_t \\ \theta_{2,t} &= \phi_2\theta_{1,t-1} + \theta_{3,t-1} + \psi_1e_t \\ &\vdots \\ \theta_{r-1,t} &= \phi_{r-1}\theta_{1,t-1} + \theta_{r,t-1} + \psi_{r-2}e_t \\ \theta_{r,t} &= \phi_r\theta_{1,t-1} + \psi_{r-1}e_t \end{aligned}$$

Substituindo-se a expressão de $\theta_{2,t-1}$ na primeira equação, temos

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \phi_2\theta_{1,t-2} + \theta_{3,t-2} + \psi_1e_{t-1} + e_t$$

e procedendo por substituições sucessivas chega-se a

$$\theta_{1,t} = \phi_1 \theta_{1,t-1} + \cdots + \phi_r \theta_{1,t-r} + \psi_1 e_{t-1} + \cdots + \psi_{r-1} e_{t-r} + e_t \quad (21)$$

Lembrando que $r = \max(p, q + 1)$ e $y_t = \theta_{1,t}$ pode-se ver que este é o modelo ARMA.

Por exemplo, o modelo ARMA(1,1) pode ser escrito na forma de um MLD como:

$$Y_t = \phi_1 Y_{t-1} + e_t \psi_1 e_{t-1} \quad e_t \sim N(0, \sigma^2) \quad (22)$$

sendo $r = q + 1 = 2$ e as matrizes do MLD:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{V} = 0 \quad (23)$$

$$\mathbf{G} = \begin{bmatrix} \phi_1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 1 & \psi_1 \\ \psi_1 & \psi_1^2 \end{bmatrix} \sigma^2 \quad (24)$$

3.3.8 Combinação de Modelos

Os modelos apresentados anteriormente captam apenas alguns componentes de uma série temporal (tendência, regressão), mas no campo dos MLD estes podem ser combinados para se modelar séries com um comportamento mais complexo. Supondo, por exemplo, que uma série Y_t é a soma de um componente de tendência polinomial $Y_{P,t}$ e um componente AR $Y_{AR,t}$:

$$Y_t = Y_{P,t} + Y_{AR,t} + e_t \quad (25)$$

Pode-se construir um MLD para cada componente tal que

$$Y_{P,t} = \mathbf{F}_{P,t} \boldsymbol{\theta}_{P,t} \quad (26)$$

$$\boldsymbol{\theta}_{P,t} = \mathbf{G}_{P,t} \boldsymbol{\theta}_{P,t-1} + \boldsymbol{\omega}_{L,t}, \quad \boldsymbol{\omega}_{P,t} \sim N(0, \mathbf{W}_{P,t}) \quad (27)$$

e

$$Y_{AR,t} = \mathbf{F}_{AR,t} \boldsymbol{\theta}_{AR,t} \quad (28)$$

$$\boldsymbol{\theta}_{AR,t} = \mathbf{G}_{AR,t} \boldsymbol{\theta}_{AR,t-1} + \boldsymbol{\omega}_{AR,t}, \quad \boldsymbol{\omega}_{AR,t} \sim N(0, \mathbf{W}_{AR,t}) \quad (29)$$

Definindo-se $\mathbf{F}_t$ e $\boldsymbol{\theta}_t$ como matrizes particionadas

$$\mathbf{F}_t = (\mathbf{F}_{P,t} \quad \mathbf{F}_{AR,t}), \quad \boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{P,t} \\ \boldsymbol{\theta}_{AR,t} \end{pmatrix} \quad (30)$$

com $\mathbf{G}$ e $\mathbf{W}_t$ como matrizes bloco-diagonais

$$\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{P,t} & \\ & \mathbf{G}_{AR,t} \end{pmatrix}, \quad \mathbf{W}_t = \begin{pmatrix} \mathbf{W}_{P,t} & \\ & \mathbf{W}_{AR,t} \end{pmatrix} \quad (31)$$

Então Y_t é escrito como um MLD com equação observacional

$$Y_t = \mathbf{F}_t \boldsymbol{\theta}_t + e_t = \mathbf{F}_{P,t} \boldsymbol{\theta}_{P,t} + \mathbf{F}_{AR,t} \boldsymbol{\theta}_{AR,t} + e_t, \quad e_t \sim N(0, V_t) \quad (32)$$

e equação de estados

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t) \quad (33)$$

3.4 Comparação de Metodologias

A análise de qualidade de predição das três metodologias será efetuada utilizando-se o erro quadrático médio (EQM) das cinco últimas previsões, definido por:

$$EQM = \frac{1}{5} \sum_{i=n-4}^n (\hat{Y}_i - Y_i)^2$$

Onde n é o número de observações, destacando-se que as previsões são obtidas para um passo a frente apenas. É indicada a metodologia que apresenta o menor valor de EQM. Também serão levadas em consideração a facilidade de aplicação e a exigência em termos de poder computacional necessário à implementação da rotina. Todas as rotinas desenvolvidas foram implementadas no software R, com auxílio do pacote *dln* (PETRIS *et al.*, 2010) e do software OpenBugs no caso da inferência bayesiana.

Considerações Gerais

Tanto os modelos ARIMA quanto os MLD partem do pressuposto que

se possua uma longa série de observações para o processo em estudo, sendo que a não observância deste pressuposto pode levar a estimativas não acuradas e grandes erros de previsão. Além disto considera-se que as observações sejam obtidas de modo adequado em todos os casos, sendo este um grande problema no Brasil, especialmente para dados de produção agrícola em geral, onde as séries são curtas e obtidas de modo indireto na maioria dos casos, o que leva a grandes erros na estimação dos valores verdadeiros da série.

4 RESULTADOS E DISCUSSÃO

4.1 Diagnóstico da série

O comportamento da série de produção total de soja do município de Mamborê-PR pode ser visualizado na figura 2a. Observa-se a presença de tendência, bem como a existência de dois períodos turbulentos, no início e no final da série. A figura 2b também evidencia a presença de dois períodos turbulentos na área plantada de soja para o mesmo município, o que não acontece quando se analisa a figura 2c, que apresenta a precipitação acumulada. Aplicando-se o teste de Dickey-Fuller Aumentado, obteve-se o valor da estatística ADF igual a 0,2316 ($p = 0,6919$). Este resultado indica a ausência de estacionariedade e, portanto, há a necessidade de se diferenciar a série. A partir da série de produção total diferenciada foram construídos os gráficos de autocorrelação e autocorrelação parcial. Verifica-se a presença de um processo autoregressivo (AR) de segunda ordem, pois nota-se o decaimento em senóides amortecidas no gráfico de autocorrelações (Figura 3a) e duas defasagens significativas no gráfico de autocorrelações parciais (Figura 3b).

Figura 2 - Produção total de Soja (ton) 2a, Área plantada (ha) 2b e Precipitação acumulada (mm) 2c no município de Mamborê-PR

4.2 Modelos ARIMA

As estimativas dos parâmetros do modelo ARIMA(2,1,0) são apresentadas na tabela 3. Pode-se verificar que ambos os parâmetros foram significativos com um baixo erro padrão. A análise dos resíduos do modelo ajustado, apresentados na

Figura 3 - Gráficos de Autocorrelação (a) e Autocorrelação Parcial (b) da série de produção total diferenciada

Tabela 3 - Coeficientes e respectivos intervalos de confiança (95%) do modelo ARIMA(2,1,0) ajustado

Parâmetro	Estimativa	Intervalo de Confiança (95%)
ϕ_1	-0,4813	$[-0,8093 ; -0,1533]$
ϕ_2	-0,3904	$[-0,7435 ; -0,0373]$

figura 4a, indica que o ajuste foi adequado, não permanecendo correlação ou tendência. As previsões um passo à frente para a série podem ser observadas na figura 4b. Percebe-se que somente uma observação não está no intervalo de predição, entretanto em apenas dois anos as estimativas estavam próximas dos valores observados. Também obtiveram-se estimativas dos parâmetros do modelo ARIMA (2,1,0) utilizando-se a metodologia de inferência bayesiana, sendo que as distribuições a priori utilizadas foram:

$$\phi_1 \sim N(0, 10^{-6}) \quad \phi_2 \sim N(0, 10^{-6}) \quad \tau \sim G(0.01, 0.01)$$

Os resultados são apresentados na tabela 4. Nota-se que eles possuem pouca diferença daqueles obtidos pelo método clássico. No entanto há maior dispersão das estimativas obtidas através do método bayesiano. De maneira similar ao modelo ajustado via inferência clássica os resíduos do modelo bayesiano (Figura 5a) não apresentam qualquer

Figura 4 - QQ-plot dos resíduos do modelo (a) e Valores Preditos (b) e intervalos de confiança (5%) para o modelo ARIMA(2,1,0)

Tabela 4 - Estatísticas resumo e quantis para o modelo ARIMA(2,1,0) ajustado

Parâmetro	Média	Desvio Padrão	Quantis		
			2,5%	50%	97,5%
ϕ_1	-0,4909	0,1752	-0,8303	-0,4891	-0,1401
ϕ_2	-0,3922	0,2017	-0,7806	-0,3972	-0,0020

comportamento que indique autocorrelação ou fuga de normalidade. A maior diferença entre os modelos deu-se nas previsões um passo a frente; neste caso o modelo ajustado via inferência bayesiana apresentou estimativas mais distantes dos valores observados (Figura 5b).

4.3 Modelos Lineares Dinâmicos

O modelo ARIMA(2,1,0) escrito na forma de um MLD é representado utilizando-se as seguintes matrizes de delineamento e evolução:

$$\mathbf{F} = (1, 0, \dots, 0) \quad \mathbf{G} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix}$$

As estimativas foram bastante próximas das anteriores, tendo um maior grau de precisão, conforme constata-se pelo menor intervalo de confiança 5. Também

Figura 5 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado via inferência bayesiana

Tabela 5 - Coeficientes e respectivos intervalos de confiança (95 %) do modelo ARIMA(2,1,0) ajustado na forma de um MLD

Parâmetro	Estimativa	Intervalo de Confiança (95 %)
ϕ_1	-0,5240	[-0,5316 ; -0,5163]
ϕ_2	-0,4458	[-0,4536 ; -0,4379]

não há indicativo de não normalidade dos resíduos (Figura 6a). A análise do gráfico de previsões (Figura 6b) indica um comportamento mais suave das previsões, não ocorrendo valores extremos comparativamente àqueles das metodologias anteriores e ainda a maioria dos valores observados estava contido nos intervalos de previsão. Esse comportamento também ocorreu quando se adotou o método bayesiano tabela 6 e figura 7a utilizando-se as seguintes distribuições a priori:

$$\phi_1 \sim N(0, 10^{-3}) \quad \phi_2 \sim N(0, 10^{-3}) \quad \tau \sim G(0.3, 0.3)$$

Tabela 6 - Estatísticas resumo e quantis para o modelo ARIMA(2,1,0) ajustado na forma de um MLD bayesiano

Parâmetro	Média	Desvio Padrão	Quantis		
			2,5%	50%	97,5%
ϕ_1	-0,5016	0,1695	-0,8340	-0,5018	-0,1757
ϕ_2	-0,3883	0,1875	-0,7534	-0,3903	-0,0158

Figura 6 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado na forma de um MLD

Figura 7 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo ARIMA(2,1,0) ajustado na forma de um MLD bayesiano

As previsões um passo a frente foram semelhantes às obtidas por meio de inferência clássica, à parte do intervalo de previsão, que se mostrou mais amplo na inferência bayesiana, provavelmente devido ao uso de distribuições a priori não informativas (Figura 7b).

Modelos de Regressão Dinâmica

Inicialmente ajustou-se um modelo de regressão linear dinâmica, considerou-se $\alpha = 0$, pois o coeficiente β representa o impacto da área plantada e conforme pode-se esperar, com área igual zero não há produção. O modelo de regressão linear dinâmica ajustado tem a seguinte representação:

$$\mathbf{F} = \mathbf{X}_t \quad \mathbf{G} = \mathbf{I}_1 \quad W = 10$$

O ajuste da regressão linear dinâmica gerou as estimativas de β observadas na figura 8a, nas quais se verifica que o coeficiente da variável área apresentou inicialmente uma leve queda seguida por uma fase de crescimento, resultando embora num comportamento diferente àquele da produtividade (total produzido/área plantada total). Verifica-se também a significância do coeficiente em todo o período de estudo. Na figura 8b percebe-se que os resíduos do modelo não apresentam tendência e/ou com-

Figura 8 - Valores do coeficiente β_1 (a) e intervalos de confiança (5%), QQ-plot dos resíduos (b) e Valores Preditos (c) e intervalos de confiança (5%) para o modelo de regressão linear dinâmica

portamento que indique não normalidade. Entretanto, quando se analisam as previsões obtidas com uso do modelo de regressão linear dinâmica constata-se grandes erros de previsão para o início do período, aumentando a qualidade das previsões à medida que o processo segue no tempo. Estimando-se os parâmetros do modelo por meio de inferência bayesiana chegou-se às estimativas de β_1 apresentados na figura 9a, na qual se verifica o mesmo comportamento encontrado por meio da inferência clássica, entretanto captando-se melhor as pequenas variações temporais na evolução do parâmetro,

bem como um menor intervalo de credibilidade para as estimativas deste. Os resíduos não apresentam qualquer indício de fuga de normalidade como se observa na figura 9b.

Figura 9 - Valores do coeficiente β_1 (a) e intervalos de confiança (5%), QQ-plot dos resíduos (b) e Valores Preditos (c) e intervalos de confiança (5%) para o modelo de regressão linear dinâmica bayesiano

O modelo de regressão linear dinâmica ajustado via inferência bayesiana resultou em menores intervalos de previsão para as produções estudadas, entretanto isto acarretou num menor número de observações contidas neste intervalo (Figura 9c) utilizando-se as seguintes distribuições a priori:

$$\beta_1 \sim N(0, 10^{-3}) \quad \tau \sim G(0.01, 0.01)$$

De modo similar ao procedido para a análise de um modelo de regressão linear simples procedeu-se a análise do modelo de regressão múltipla, no qual foi adicionada uma variável preditora, no caso, o impacto da precipitação acumulada no período de cultivo da soja. Também se considerou $\alpha = 0$, dado que espera-se que não haja produção no caso de ausência de chuva e área plantada. Este modelo possui a seguinte representação:

$$\mathbf{F} = \mathbf{X}_{1t}, \mathbf{X}_{2t} \quad \mathbf{G} = \mathbf{I}_2 \quad \mathbf{W} = \text{diag}\{3.334, 100\}$$

Os valores estimados e seus respectivos intervalos de confiança para a estimativa de β_1 podem ser observados na figura 10a. O comportamento foi similar àquele do modelo de regressão linear, o que deve-se ao fato do coeficiente β_2 (Figura 10b) não ter sido significativo no período estudado.

Figura 10 - Valores do coeficiente β_1 (a) e β_2 (b) com seus respectivos intervalos de confiança (5%) para o modelo de regressão múltipla dinâmica

Conforme a análise dos resíduos (Figura 11a) há problemas de ajuste no modelo de regressão linear múltipla. O aumento do número de parâmetros resultou em boas previsões pontuais, mas com um intervalo de previsão extenso, o que torna inviável o seu uso para fins práticos (Figura 11b). As estimativas pontuais dos parâ-

Figura 11 - QQ-plot dos resíduos (a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo de regressão múltipla dinâmica

metros obtidas por meio de inferência bayesiana foram similares àquelas obtidas pela

máxima verossimilhança. Nota-se diferença na estabilidade do parâmetro β_2 (Figura 12). Utilizaram-se as seguintes distribuições a priori:

$$\beta_{11} \sim N(0, 10^{-3}) \quad \beta_{21} \sim N(0, 10^{-3}) \quad \tau \sim G(0.01, 0.01)$$

Observando-se o comportamento dos resíduos (figura 13a), percebe-se que estes apre-

Figura 12 - Valores do coeficiente β_1 (a) e β_2 (b) com seus respectivos intervalos de confiança (5%) para o modelo de regressão múltipla dinâmica bayesiano

sentam comportamento similar aos obtidos por meio da inferência clássica. Entretanto, analisando-se a figura 13b verifica-se que as previsões possuem intervalos de previsão mais restritos, assim como no caso da regressão linear dinâmica ajustada por inferência bayesiana.

4.4 Comparação dos Modelos

Dificuldade de implementação e poder computacional

Os modelos ARIMA possuem maiores facilidades computacionais, uma vez que há diversas rotinas já implementadas para ajustá-los. Além de resolverem grande parte dos problemas de previsão, por serem, relativamente, mais simples necessitam de um processador muito menos potente do que aquele necessário para o ajuste

Figura 13 - QQ-plot dos resíduos(a) e Valores Preditos (b) e intervalos de confiança (5%) do modelo de regressão múltipla dinâmica bayesiano

de modelos lineares dinâmicos. Outro ponto importante a ressaltar, é a dificuldade de estimação dos parâmetros dos modelos dinâmicos por meio da inferência clássica, pois em muitos casos, observados no presente trabalho, as funções logaritmo de máxima verossimilhança convergiram para máximos locais e geraram estimativas absurdas.

Qualidade das previsões

Na tabela 7 verifica-se que o modelo com menor EQM e portanto melhores previsões (cinco últimas) foi o modelo ARIMA(2,1,0) ajustado na forma de um MLD, sendo seguido pelo mesmo modelo ajustado via inferência bayesiana.

Tabela 7 - Erro Quadrático Médio para os modelos ajustados

Modelo	EQM
ARIMA(2,1,0)	911049873
ARIMA(2,1,0) Bayesiano	2977163664
ARIMA (2,1,0) na forma de um MLD	709864781
ARIMA (2,1,0) na forma de um MLD bayesiano	768924355
Regressão Linear Dinâmica	1736932804
Regressão Linear Dinâmica bayesiana	1259423166
Regressão Múltipla Dinâmica	1213681063
Regressão Múltipla Dinâmica bayesiana	1270675646

4.5 Utilização do Modelo Escolhido

O modelo ARIMA(2,1,0) escrito na forma de um MLD é o indicado para a previsão de safra no município de Mamborê-PR, pois devido ao seu menor EQM este levará a previsões mais acuradas, assim sendo as decisões tanto econômicas quanto logísticas tomadas com base neste modelo serão, em teoria, as melhores possíveis na gama dos modelos estudados neste trabalho. Um exemplo de setor a se beneficiar de melhores previsões de safra é o setor de seguros, podendo derivar a partir das projeções deste modelo a produtividade esperada para o município e assim calcular um prêmio de seguro mais adequado ao município. Também ressalta-se a sua utilização para o provimento de condições necessárias ao escoamento da futura produção.

5 CONCLUSÃO

Foram abordadas diferentes metodologias para a obtenção de previsões para séries de produção agrícola. Com base nos resultados encontrados pode-se depreender que:

- i. O modelo ARIMA (2,1,0) tanto na sua forma tradicional quanto na forma de um MLD forneceu as melhores estimativas, indicando que a autocorrelação entre os valores da própria série de produção geram melhores previsões que o uso de variáveis explicativas.
- ii. De maneira geral as estimativas obtidas por meio de inferência bayesiana tiveram amplos intervalos de credibilidade, o que em parte é explicado pela utilização de prioris não informativas. Entretanto isto pode ser alterado com o conhecimento adquirido no presente trabalho, bem como do engajamento de especialistas do setor agrícola, gerando prioris informativas que representem com maior precisão e acurácia o cenário agrícola.
- iii. As dificuldades na implementação dos MLDs por meio de inferência clássica geraram sérios problemas na definição dos valores iniciais a serem utilizados nos algoritmos de otimização. Portanto, recomenda-se a utilização de inferência bayesiana para modelar a incerteza a respeito dos parâmetros de maneira adequada.
- iv. Os MLDs, embora não tenham apresentado um desempenho satisfatório, apresentam vantagens quanto ao seu uso na modelagem de séries de produção, principalmente pela possibilidade de se intervir em qualquer instante no tempo e se ajustar o modelo para acomodar novos parâmetros, bem como alterações nos existentes. Este fato é de grande valia na previsão de safras, pois a produção agrícola sofre constantes alterações devido ao surgimento de novas tecnologias, incentivos governamentais, variações de preço, clima em geral e variáveis de cunho comportamental do agricultor.
- v. Todos os modelos foram afetados pela baixa quantidade e qualidade de observações, sendo importante o desenvolvimento de um sistema confiável de amostragem para as diversas variáveis que impactam o setor agrícola brasileiro, de modo que futuros estudos levem à inferências mais precisas e acuradas.

- vi. Pôde-se perceber o comportamento divergente dos modelos, portanto reafirma-se mais uma vez a necessidade da indicação do modelo utilizado bem como da acurácia de suas previsões para que se possam fazer projeções de qualidade e que tenham a precisão necessária para a tomada de decisões do setor agrícola como um todo.

O estudo desenvolvido se limitou a estudar duas metodologias de previsão e apenas um município, entretanto podem ser considerados diversos modelos e mais de um município, ou um estado, até mesmo o país inteiro. Entretanto este serve como estímulo e base para futuros trabalhos no campo da previsão de safras agrícolas, de modo a fornecer procedimentos minimamente necessários para a obtenção de estimativas confiáveis de produção agrícola quando da utilização de modelos estocásticos de previsão.

REFERÊNCIAS

AGRICULTURA, PECUÁRIA E ABASTECIMENTO MINISTÉRIO DA. **Zoneamento Agrícola de Risco Climático para a cultura de soja no Estado do Paraná**. [S.l.], 2014.

AMEEN, J.R.M.; HARRISON, P.J. Normal discount bayesian models. **Bayesian statistics**, Valência, v. 2, n. 2, p. 271–198, 1985.

AWAL, M.A.; SIDDIQUE, M.A.B. Rice production in bangladesh employing by arima model. **Bangladesh Journal of Agricultural Research**, Bangladesh, v. 36, n. 1, p. 51–62, 2011.

BASSEVILLE, M.; BENVENISTE, A.; CHOU, K. C; GOLDEN, S. A.; NIKOUKHAH, R.; WILLSKY, A. S. Modeling and estimation of multiresolution stochastic processes. **Information Theory, IEEE Transactions on**, New Jersey, v. 38, n. 2, p. 766–784, 1992.

BERNARDO, J. M.; SMITH, A. F. M. **Bayesian theory**. [S.l.]: John Wiley & Sons, 2009.

BIANCHI, L.; JARRETT, J.; CHOUDARY, R. Improving forecasting for telemarketing centers by arima modeling with intervention. **International Journal of Forecasting**, Washington, v. 14, n. 4, p. 497–504, 1998.

BOKEN, V. Forecasting spring wheat yield using time series analysis. **Agronomy Journal**, Washington, v. 92, n. 6, p. 1047–1053, 2000.

BOX, G.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, Londres, v. 26, n. 2, p. 211–252, 1964.

BOX, G; JENKINS, G. Time series analysis: Forecasting and control. **Holden-D. iv, San Francisco**, 1970.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis**. Londres: John Wiley & Sons, 2013. 784 p.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. [S.l.]: John Wiley & Sons, 2011.

BRUSS, F. T.; ROBERTSON, J. B. Wald's lemma'for sums of order statistics of iid random variables. **Advances in applied probability**, JSTOR, p. 612–623, 1991.

CAPA-MOROCHO, M.; RODRÍGUEZ-FONSECA, B.; RUIZ-RAMOS, M. Crop yield as a bioclimatic index of el niño impact in europe: Crop forecast implications. **Agricultural and Forest Meteorology**, Washington, v. 198, n. 1, p. 42–52, 2014.

CAPRIO, U.; GENESIO, R.; POZZI, S.; VICINO, A. Short term load forecasting in electric power systems: A comparison of arma models and extended wiener filtering. **Journal of Forecasting**, Washington, v. 2, n. 1, p. 59–76, 1983.

- CARTER, C. K; KOHN, R. On gibbs sampling for state space models. **Biometrika**, Londres, v. 81, n. 3, p. 541–553, 1994.
- CHOU, K. C.; WILLSKY, A. S.; BENVENISTE, A. Multiscale recursive estimation, data fusion, and regularization. **Automatic Control, IEEE Transactions on**, Washington, v. 39, n. 3, p. 464–478, 1994.
- DIACONIS, P.; YLVISAKER, D. *et al.* Conjugate priors for exponential families. **The Annals of statistics**, Institute of Mathematical Statistics, v. 7, n. 2, p. 269–281, 1979.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, Washington, v. 74, n. 366a, p. 427–431, 1979.
- DIEBOLD, F. X; GUNTHER, T. A.; TAY, A. S. Evaluating density forecasts. **International Economic Review**, Washington, v. 39, n. 1, p. 863–883, 1997.
- DILLON, B. Estimation of a dynamic agricultural production model with observed, subjective distributions. In: **Proceedings of Northeast Universities Development Consortium Conference**. New Haven: Yale University, Economic Growth Center, 2011. v. 1, n. 1. Disponível em: 2014-08-11. Disponível em: <www.econ.yale.edu/conference/neudc11/schedule.htm>.
- DOSSA, D. **Projeções do Agronegócio Brasil 2013/14 a 2023/24. 2014, Ministério da Agricultura, Pecuária e Abastecimento (MAPA) and Assessoria de Gestão Estratégica (AGE)**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2014. 200 p.
- EDLUND, P. Identification of the multi-input box-jenkins transfer function model. **Journal of Forecasting**, Washington, v. 3, n. 3, p. 297–308, 1984.
- GAMERMAN, D. Markov chain monte carlo for dynamic generalised linear models. **Biometrika**, Washington, v. 85, n. 1, p. 215–227, 1998.
- GAMERMAN, D.; MIGON, H. S. Dynamic hierarchical models. **Journal of the Royal Statistical Society. Series B (Methodological)**, Londres, v. 55, n. 3, p. 629–642, 1993.
- GAO, L.; WANG, C.; SI, Z.; CHEN, J.; LIU, M. Development of crop pest forecast and information management system in heilongjiang province. **China Plant Protection**, Pequim, v. 2, n. 1, p. 005, 2010.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, Washington, v. 7, n. 4, p. 457–472, 1992.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. **Federal Reserve Bank of Minneapolis Research Department Staff Report**, Minneapolis, v. 196, n. 1, p. 1–30, 1991.
- GHIL, M.; COHN, S.; TAVANTZIS, J.; BUBE, K.; ISAACSON, E. Applications of estimation theory to numerical weather prediction. In: **Dynamic meteorology: Data assimilation methods**. Londres: Springer, 1981. p. 139–224.

GORDON, N. J.; SALMOND, D. J.; SMITH, A. F. M. Novel approach to nonlinear/non-gaussian bayesian state estimation. In: **IEE Proceedings F (Radar and Signal Processing)**. Washington: [s.n.], 1993. v. 140, n. 2, p. 107–113, 1993.

GOUDRIAAN, J. **Crop micrometeorology**. Pudoc: Center for Agricultural Publishing and Documentation, 1977. 249 p.

HARRISON, P. J.; STEVENS, C. F. Bayesian forecasting. **Journal of the Royal Statistical Society. Series B (Methodological)**, Londres, v. 38, n. 3, p. pp. 205–247, 1976.

KIM, J. H. Forecasting autoregressive time series with bias-corrected parameter estimators. **International Journal of Forecasting**, Washington, v. 19, n. 3, p. 493–502, 2003.

KLESHCHENKO, A.D.; GONCHAROVA, T.A.; NAIDINA, T.A. Using the satellite data in dynamic models of crop yield forecasting. **Russian Meteorology and Hydrology**, Moscou, v. 37, n. 4, p. 279–285, 2012.

LANDIM, F.; GAMERMAN, D. Dynamic hierarchical models: an extension to matrix-variate observations. **Computational statistics & data analysis**, Washington, v. 35, n. 1, p. 11–42, 2000.

LIU, J.; WEST, M. Combined parameter and state estimation in simulation-based filtering. In: **Sequential Monte Carlo methods in practice**. Londres: Springer, 2001. p. 197–223.

MIGON, H.S.; HARRISON, P.J. An application of non-linear bayesian forecasting to television advertising. **Bayesian Statistics**, Washington, v. 2, n. 1, p. 681–696, 1985.

MIGON, H. S.; GAMERMAN, D. Generalized exponential growth models a bayesian approach. **Journal of Forecasting**, Washington, v. 12, n. 7, p. 573–584, 1993.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. [S.l.]: Blucher, 2006.

MUHAMMAD, F.; JAVED, M.S.; BASHLR, M. Forecasting sugarcane production in pakistan using arima models. **Pak Journal of Agricultural Sciences**, Bangladesh, v. 9, n. 1, 1992.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society. Series A (General)**, Londres, v. 135, n. 3, p. pp. 370–384, 1972.

NEWBOLD, P.; AGIAKLOGLOU, C.; MILLER, J. Adventures with arima software. **International Journal of Forecasting**, Washington, v. 10, n. 4, p. 573–581, 1994.

PADHAN, P.C. Application of arima model for forecasting agricultural productivity in india. **Journal of Agricultural Social Sciences**, Bangladesh, v. 2, n. 1, p. 50–56, 2012.

PASCUAL, L.; ROMO, J.; RUIZ, E. Bootstrap predictive inference for arima processes. **Journal of Time Series Analysis**, Washington, v. 25, n. 4, p. 449–465, 2004.

- PETRIS, G. *et al.* An r package for dynamic linear models. **Journal of Statistical Software**, American Statistical Association, v. 36, n. 12, p. 1–16, 2010.
- PFLAUMER, P. Forecasting us population totals with the box-jenkins approach. **International Journal of Forecasting**, Washington, v. 8, n. 3, p. 329–338, 1992.
- POLSON, N. G.; STROUD, J. R.; MÜLLER, P. Practical filtering with sequential parameter learning. **University of Chicago Technological Report**, Chicago, v. 70, n. 2, p. 1–18, 2002.
- PREEZ, J. Du; WITT, S. F. Univariate versus multivariate time series forecasting: an application to international tourism demand. **International Journal of Forecasting**, Elsevier, Washington, v. 19, n. 3, p. 435–451, 2003.
- SCHNEIDER, S.; FERREIRA, B.; ALVES, F. **Aspectos multidimensionais da agricultura brasileira: diferentes visões do Censo Agropecuário 2006**. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea), 2013. 402 p.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Washington, v. 52, n. 3, p. 591–611, 1965.
- SHEPHARD, N.; PITT, M. K. Likelihood analysis of non-gaussian measurement time series. **Biometrika**, Washington, v. 84, n. 3, p. 653–667, 1997.
- SUN, J. Dynamic monitoring and yield estimation of crops by mainly using the remote sensing technique in china. **Photogrammetric engineering and remote sensing**, Washington, v. 66, n. 5, p. 645–650, 2000.
- VRIES, F. W. T. De; BRUNSTING, A. H. M.; LAAR, H. H. Van. Products, requirements and efficiency of biosynthesis a quantitative approach. **Journal of theoretical Biology**, Elsevier, New York, v. 45, n. 2, p. 339–377, 1974.
- WEI, W. W. **Time series analysis**. Londres: Addison-Wesley publ, 1994. 624 p.
- WEST, M.; HARRISON, P. J. Monitoring and adaptation in bayesian forecasting models. **Journal of the American Statistical Association**, Washington, v. 81, n. 395, p. 741–750, 1986.
- WEST, M.; HARRISON, P. J. **Bayesian Forecasting & Dynamic Models**. 2nd. ed. Londres: Springer Verlag, 1997. 682 p.
- WEST, M.; HARRISON, P. J.; MIGON, H. S. Dynamic generalised linear models and bayesian forecasting (with discussion). **Journal of the American Statistical Association**, Washington, v. 80, n. 389, p. 73–97, 1985.
- WIT, C. T. **Photosynthesis of leaf canopies**. Washington: Centre for Agricultural Publications and Documentation, 1966. 64 p.
- WIT, C. T.; BROUWER, R.; VRIES, F. W. T. P.; SETLÍK, I *et al.* The simulation of photosynthetic systems. In: WAGENINGEN, NETHERLANDS, PUDO. **Prediction and measurement of photosynthetic productivity. Proceedings of the IBP/PP Technical Meeting, Tfebon, [Czechoslovakia], 14-21 September, 1969**. [S.l.], 1970. p. 47–70.

WIT, C. T. De. Simulation of assimilation, respiration and transpiration of crops. **Quarterly Journal of the Royal Meteorological Society**, Londres, v. 105, n. 445, p. 728–729, 1978.

XIAO, Z.; PHILLIPS, P. C. B. An adf coefficient test for a unit root in arma models of unknown order with empirical applications to the us economy. **The Econometrics Journal**, Washington, v. 1, n. 2, p. 27–43, 1998.

APÊNDICE

Estimação de um modelo ARIMA (2,1,0) por meio de inferencia bayesiana - WinBUGS

```
model{
mu[1] <- phi1*y1+phi2*y0
dy[1] ~ dnorm(mu[1],tau2)
mu[2] <- phi1*y1+phi2*y0
dy[2] ~ dnorm(mu[2],tau2)
for(t in 3:T1) {
mu[t] <- phi1*dy[t-1] + phi2*dy[t-2]
dy[t] ~ dnorm(mu[t],tau2)
}
# Predictions
for (t in 3:T1) {
yp[1+t] ~ dnorm(mup[1+t],tau2)
mup[1+t] <- phi1*dy[t]+phi2*dy[t-1]
}
# Prior distribution
phi1 ~ dnorm(0.0, 1.0E-6)
phi2 ~ dnorm(0.0, 1.0E-6)
tau2 ~ dgamma(0.01, 0.01)
y0 ~ dnorm(0.0,1.0E-6)
y1 ~ dnorm(0.0,1.0E-6)
}
```

Ajuste de um modelo ARIMA (2,1,0) reescrito na forma de um MLD

```
buildfun = function(theta){
theta <- parm_rest(theta)
dlm <- dlmModARMA(ar=c(theta[1],theta[2]),dV=1e-7, ma=NULL, sigma2=theta[3])

tmp0 <- matrix(c(theta[1],theta[2],1,0),nr=2)
tmp1 <- matrix(c(theta[3],0,0,0),nc=1)
tmp <- solve(diag(4)-tmp0%x%tmp0)%*%tmp1
dlm$C0[1:2,1:2] <- matrix(tmp,nr=2)
return( dlm )
}

estMLE = dlmMLE (diff(y), parm = c(0,0,1), build=buildfun,hessian=T)
hs=hessian(function(theta) dlmLL(y, buildfun(theta)), estMLE$par)
all(eigen(hs, only.values = TRUE)$values > 0)
aVar=solve(hs)
```

```

aVar
sqrt(diag(aVar))
modpar=buildfun(estMLE$par)
drop(V(modpar))
drop(W(modpar))

LSARMod2 <- buildfun(estMLE$par)
LSARf2 <- dlmFilterDF(y,LSARMod2,DF=.9)
nobs=length(y)

coef <- LSARf2$m[nobs+1,1:2]
coef[1] <- coef[1]-nobs*coef[2]
covar <- dlmSvd2var(LSARf2$U.C[[nobs+1]],LSARf2$D.C[nobs+1,])
coef.se <- c(sqrt(covar[1,1]),sqrt(covar[2,2]))
cbind(coef, coef.se)

```

Ajuste de um modelo ARIMA (2,1,0) reescrito na forma de um MLD por meio de inferência bayesiana - WinBUGS

```

model {

  for (t in 3:T){
    dy[t]~dnorm(mu[t],tau.obs)
    mu[t] <- phi1*(dy[t-1])+phi2*(dy[t-2])
  }

  for (t in 3:T){
    dyp[t]~dnorm(mup[t],tau.obs)
    mup[t] <-phi1*(dy[t-1])+phi2*(dy[t-2])
  }

  phi1 ~ dnorm(0, 0.001)I(-1,1)
  phi2 ~ dnorm(0, 0.001)I(-1,1)
  tau.obs ~dgamma(.3,.3)
}

```

Ajuste de um modelo de regressão linear dinâmica

```

buildfun2 = function(theta){
  dlm <- dlmModReg(area,addInt=FALSE,dV=theta[1]^2)
}

estMLE2 = dlmMLE (y, parm = 1, build=buildfun2)
exp(estMLE2$par)
(estMLE2$par)^2

modpar2=buildfun2(estMLE2$par)
hs=hessian(function(theta) dlmLL(y, buildfun2(theta)), estMLE2$par)
all(eigen(hs, only.values = TRUE)$values > 0)

```

```

aVar=solve(hs)
aVar
sqrt(diag(aVar))
V(modpar2)

```

Ajuste de um modelo de regressão linear dinâmico por meio de inferência bayesiana - WinBUGS

```

model {

for (t in 1:T){
y[t]~dnorm(mu[t],tau.obs[t])
mu[t] <- alpha+beta1[t]*area[t]
}
for (t in 2:T){
tau.obs[t] <- tau * pow(0.95,t-1)
}

for (t in 2:T){
beta1[t] ~ dnorm(beta1[t-1],tau.state)
}
for (t in 2:T){
yp[t]~dnorm(mup[t],tau.p[t])
mup[t] <-alpha+ beta1[t-1]*area[t]
tau.p[t] <- 1/((1/tau.state+1/tau.state)*(area[t]*area[t]))+tau.obs[t]
}
alpha~dnorm(0,0.001)
beta1[1] ~dnorm(0,.0001)
tau ~dgamma(.01,.01)
tau.obs[1]<-tau
tau.state<-0.3
W <-1/tau.state
}

```

Ajuste de um modelo de regressão múltipla dinâmico

```

x=matrix(c(as.numeric(area),pp[,1]), nc = 2)
buildfun3 = function(theta) {
dlmModReg(x,addInt=FALSE,dV=theta[1]^2)
}
exp(estMLE3$par)
estMLE3 = dlmMLE (y, parm = 1, build=buildfun3)
modpar3=buildfun3(estMLE3$par)
hs3=hessian(function(theta) dlmLL(y, buildfun3(theta)), estMLE3$par)
all(eigen(hs3, only.values = TRUE)$values > 0)
aVar3=solve(hs3)
aVar3
sqrt(diag(aVar3))
V(modpar3)
coef=estMLE3$par

```

Ajuste de um modelo de regressão múltipla dinâmico por meio de inferência bayesiana - WinBUGS

```

model {
  for (t in 1:T){
    y[t]~dnorm(mu[t],tau.obs[t])
    mu[t] <-beta0+ beta1[t]*x[t,1]+beta2[t]*x[t,2]
  }
  for (t in 2:T){
    tau.obs[t] <- tau * pow(0.95,t-1)
  }

  for (t in 2:T){
    beta1[t] ~ dnorm(beta1[t-1],tau.state)
    beta2[t] ~ dnorm(beta2[t-1],tau.state2)
  }
  for (t in 2:T){
    yp[t]~dnorm(mup[t],tau.p[t])
    mup[t] <-beta0+ beta1[t-1]*x[t,1]+beta2[t-1]*x[t,2]
    tau.p[t] <- 1/((1/tau.state+1/tau.state)*(x[t,1]*x[t,1]))
    +1/((1/tau.state2+1/tau.state2)*(x[t,2]*x[t,2]))+tau.obs[t]
  }

  beta0~dnorm(0,0.001)
  beta1[1] ~dnorm(0,0.001)
  beta2[1] ~dnorm(0,0.001)
  tau ~dgamma(.01,.01)
  tau.obs[1]<-tau
  tau.state <- 0.3
  tau.state2<- 0.01
  W <-1/tau.state
}

```

Procedimentos gerais para ajuste e verificação dos modelos por meio de inferência bayesiana - R script (Modelo de regressão múltipla)

```

parameters2=c("mu","beta0","beta1","beta2","yp")
T=length(y)
pp1=pp[,1]
x=cbind(as.numeric(area),pp1)
y=dados[which(dados$Municipio==levels(dados$Municipio)[16]),4]
data2 = list ("y", "x", "T")
prod.sim3 = bugs(data2,init=NULL, parameters2,
model.file="D:\\Dropbox\\Dropbox\\ESALQ_13_14\\Dissertação\\Programas\\R\\regressaomult3.txt",
n.chains=2, n.iter=500000,codaPkg=TRUE)
out.coda3 = read.bugs(prod.sim3,thin=500)
out.coda3
xyplot(out.coda3)
plot(out.coda3)
densityplot(out.coda3)
acfplot(out.coda3)
autocorr.diag(out.coda3)
gelman.diag(out.coda3)
gelman.plot(out.coda3)

```

```

out.summary3 = summary(out.coda3, q=c(0.025,0.5, 0.975))
out.summary3$quantiles[2,]
names(out.summary3)
plot(y,type="o",col=2,ylim=c(40000,220000))
points(out.summary3$quantiles[-1,2])
lines(out.summary3$quantiles[-1,1])
lines(out.summary3$quantiles[-1,3])

dim(out.summary3$quantiles)
plot(y1,ylim=c(40000,250000),ylab="Produção de Soja em ton", xlab="Ano", main="",type = 'o',
     col = "seagreen")
lines(ts(out.summary3$quantiles[105:137,2],start=1981), type = 'o',
     pch = 20, col = "brown")
lines(ts(out.summary3$quantiles[105:137,1],start=1981), lty = 2, col = "brown")
lines(ts(out.summary3$quantiles[105:137,3],start=1981), lty = 2, col = "brown")

ts.plot(ts(out.summary3$quantiles[2:35,2],start=1980),ylim=c(0,5),main="",xlab="Ano",ylab="")
lines(ts(out.summary3$quantiles[2:35,1],start=1980),lty=2,col="black")
lines(ts(out.summary3$quantiles[2:35,3],start=1980),lty=2,col="black")

ts.plot(ts(out.summary3$quantiles[36:69,2],start=1980),ylim=c(-100,100),main="",xlab="Ano",ylab="")
lines(ts(out.summary3$quantiles[36:69,1],start=1980),lty=2,col="black")
lines(ts(out.summary3$quantiles[36:69,3],start=1980),lty=2,col="black")

erro=y[2:34]-out.summary3$quantiles[105:137,2]
qqnorm(erro,xlab="Quantis Teóricos",main="",ylab="Quantis Amostrais")
qqline(erro,distribution=qnorm)

mse_reg_mult=sum((y[30:34]-out.summary3$quantiles[133:137,2])^2)/5

```