

Universidade Federal do Rio de Janeiro

# **Análise e Previsão de Resultados de Partidas de Futebol**

Fábio Figueiredo Farias

Rio de Janeiro

2008

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Departamento de Métodos Estatísticos  
Instituto de Matemática  
Universidade Federal do Rio de Janeiro

Fábio Figueiredo Farias

## **Análise e Previsão de Resultados de Partidas de Futebol**

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Orientadora:

Flávia Maria Ferreira Pinto Landim

Novembro de 2008

À minha esposa, Priscila.

# Agradecimentos

- Agradeço à Deus, pois foi ele quem me capacitou e me guiou para que eu chegasse até essa etapa da minha vida;
- a todos da minha família, pois sempre estiveram do meu lado, dando apoio em todos os passos da minha vida. Especialmente minha esposa, minha mãe e meu padrasto, que acreditam em mim acima de tudo;
- à minha esposa Priscila, por abrir mão de momentos preciosos juntos para que pudesse finalizar o tão esperado mestrado;
- aos meus orientadores, pela paciência e pelo conhecimento passado durante todo o período em que estivemos envolvidos na dissertação;
- aos colegas de mestrado, pela convivência no primeiro ano de mestrado, que tornaram mais fácil e alegre aquele difícil ano de disciplinas. Em especial, à Flávia e ao Felipe “Fenômeno”, pelos bons momentos de estudo e descontração;
- ao Rodrigo Targino, aluno de iniciação científica, pelo conhecimento adquirido e o companherismo durante a dissertação;
- ao grande amigo Fellipe e a minha cunhada Simone, pela ajuda em tópicos importantes da dissertação;
- aos companheiros de trabalho, pelo incentivo, pela força e por me liberar em alguns dias para me dedicar à dissertação;
- à CAPES, pelo apoio financeiro;

- por fim, a todos aqueles que, de alguma forma, torçeram por mim e contribuíram para que eu concluísse esse curso de mestrado.

# Resumo

A classe de Modelos Lineares Dinâmicos Generalizados estende a modelagem dinâmica à família exponencial, e não ficamos mais restritos a assumir a normalidade dos dados. Esse tipo de modelagem é adequado para se trabalhar com dados provenientes de distribuições não-normais da família exponencial: Poisson para dados de contagem, Bernoulli para proporções, Gama para dados contínuos assimétricos, entre outras. Em particular, entre os dados provenientes de campeonatos de futebol, o número de gols é um candidato à distribuição de Poisson.

Essa dissertação se propõe a construir modelos para prever resultados de partidas de futebol, modelando o número de gols de cada equipe. Portanto, dois objetivos podem ser alcançados: fazer previsões e comparar modelos. Propomos o uso de dois modelos Bayesianos dinâmicos com coeficientes auto-regressivos de evolução para o número de gols feitos pelas equipes do campeonato brasileiro do ano de 2006. Além de ajustar esses dois modelos, também foram ajustados um modelo estático e um dinâmico propostos por RUE e SALVESEN (2000) e usados em SOUZA e GAMERMAN (2004). Para comparar os modelos, usamos a razão de verossimilhanças.

A inferência foi feita adotando-se a abordagem Bayesiana e, como a distribuição a posteriori não é analiticamente tratável, usamos métodos de Monte Carlo via Cadeias de Markov (MCMC).

# Abstract

The class of Dynamic Generalized Linear Models extends the dynamic modeling of exponential family, and we are no longer restricted to assuming normality of data. This type of modeling is suitable to work with data from non-normal sampling distributions of exponential family: Poisson for counting data, Bernoulli for proportions, Gamma for asymmetric continuous data, among others. In particular, among data from soccer championships, the number of goals is a candidate for a Poisson distribution.

The purpose of this research is to build models to predict results of soccer games, modeling the number of goals from each team. Therefore, two objectives can be achieved: making predictions and comparing models. The use of two dynamic bayesian models with autoregressive coefficients of evolution for the number of goals from different teams during Brazilian soccer championship in 2006 was proposed. In addition to adjusting these two first models, two other ones were also adjusted: a static model and a dynamic model proposed by RUE and SALVESEN (2000) and used in SOUZA and GAMERMAN (2004). In order to compare these models, the likelihood ratio was used.

The inference was made with the adoption of the bayesian approach and, as a posteriori distribution is not analytically tractable, methods of Markov chain Monte Carlo were used.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>5</b>
1.1	Considerações Gerais . . . . .	5
1.2	Objetivo do Estudo . . . . .	7
1.3	Dados Trabalhados . . . . .	7
1.4	Inferência Bayesiana . . . . .	10
1.5	Inferência por Simulação . . . . .	13
1.6	Organização do Trabalho . . . . .	15
<b>2</b>	<b>Modelos Lineares Dinâmicos</b>	<b>17</b>
2.1	Introdução . . . . .	17
2.2	Descrição do Modelo . . . . .	19
2.3	Inferência no MLD . . . . .	21
2.4	Previsão . . . . .	24
<b>3</b>	<b>Modelo Linear Dinâmico Generalizado</b>	<b>26</b>
3.1	Introdução . . . . .	26
3.2	Distribuições na Família Exponencial . . . . .	27
3.3	Descrição do Modelo . . . . .	28
3.4	Inferência no MLDG . . . . .	29
3.5	Previsão . . . . .	32
<b>4</b>	<b>Modelo para Dados Poisson</b>	<b>34</b>
4.1	Introdução . . . . .	34
4.2	Especificação dos Modelos Utilizados . . . . .	36
4.3	Processo de Inferência . . . . .	46
4.4	Previsão . . . . .	47
<b>5</b>	<b>Aplicação</b>	<b>49</b>
5.1	Introdução . . . . .	49

5.2 Resultados . . . . .	50
<b>6 Conclusões</b>	<b>65</b>
<b>Apêndice A: Gráficos diversos para o Modelo <math>D_3</math></b>	<b>69</b>
<b>Apêndice B: Algoritmo do WINBUGS para fazer inferência até a rodada <math>s</math> para o modelo <math>D_3</math></b>	<b>77</b>

# Capítulo 1

## Introdução

### 1.1 Considerações Gerais

O futebol é, sem dúvida, uma das maiores paixões dos brasileiros. Impressiona pela emoção, não só os que praticam, mas principalmente os seus inúmeros fãs. Assunto freqüente nos meios de comunicação, ele tanto encanta pela sua magia e dinamismo, que até aqueles que não são fãs declarados param para ver a seleção de seu país disputar uma Copa do mundo.

Quantas vezes clubes que fizeram grandes investimentos foram derrotados por outros que, “em teoria”, são mais fracos. Um elenco de ponta e um esquema tático infalível são primordiais nesse jogo, que às vezes parece um combate estratégico duelado pelos técnicos, mas nem sempre os melhores times vencem. É exatamente isso que o torna tão impressionante.

Esse esporte cada vez mais se torna alvo de especialistas que buscam na matemática e na estatística meios para explicar e prever os resultados. Tal fato é tão impressionante, que muitos especialistas da área esportiva dizem que o futebol é o esporte no qual a chance do mais fraco vencer o mais forte se maximiza em relação a qualquer outro esporte. A frase abaixo retrata bem essa propriedade:

Engendra imprecisão tática, exige uma grande qualidade técnica dos jogadores e faz com que o jogo decorra num ritmo de altas improbabilidades, mesmo quando um time muito superior joga com um time notavelmente inferior. Jogado com os pés, o futebol

fica menos previsível, o que faz com que nele se insinuem as idéias de sorte, destino, predestinação e vitória (DA MATTA, 1994: pag. 16).

Devido ao avanço que vem cada vez mais se intensificando em diversas áreas, a estatística hoje tem meios para tratar dados dessa natureza. Graças a ela, não só os resultados, mas também questões importantes podem ser estudadas, por exemplo, a quantidade de pontos que um time deve ter para ganhar o título com probabilidade razoavelmente alta e as chances de um time terminar na frente do outro.

Uma temporada de futebol dura aproximadamente 11 meses, independente do país. Ao longo desse período, as equipes disputam de 2 a 3 campeonatos. Algumas vezes chegam a jogar 3 vezes na semana em cidades diferentes, intensificando as chances de ocorrerem lesões e desgastes no elenco. Entretanto, este não é o único fator que influencia diretamente o desempenho da grande maioria das equipes durante uma temporada. Outros fatores, entre eles a troca de técnico e a presença dos torcedores, causam impactos significativos na campanha dos times no certame. Esses, ao contrário do primeiro, costumam ser cruciais no decorrer de uma temporada, quando o time ensaia uma recuperação no torneio.

A ação de fatores como esses é tão impressionante, que não se pode deixá-los de fora da análise, seja qual for o modelo escolhido. Eis que surge então a primeira razão para a escolha de uma modelagem baseada na classe dos Modelos Lineares Dinâmicos (MLD). A principal característica do uso dessa abordagem é capturar as alternâncias das equipes ao longo da disputa de um campeonato.

Por se tratar de um modelo dinâmico, os parâmetros são atualizados a cada instante de tempo, o que é razoável tendo em vista que o comportamento dos times varia bastante com o passar do tempo. Esse tipo de variação permite avaliar o desempenho das equipes ao longo das rodadas do campeonato. Os MLD serão descritos com mais detalhes no próximo capítulo.

A modelagem proposta aqui trabalha com os resultados de jogos anteriores dentro do mesmo campeonato para estudar os jogos seguintes, nos quais cada equipe se enfrenta duas vezes (1º turno e 2º turno).

Para simplificar as previsões, faz-se uso de resultados agregados: vitória, empate e

derrota. Entretanto, nada impede que seja feita uma análise preditiva mais detalhada com os placares dos jogos.

Vale ressaltar que essa aplicação vem se desenvolvendo e ganhando espaço no mundo acadêmico e no mercado. Trabalhos vêm sendo desenvolvidos, não só com o futebol, mas com outros esportes também, como por exemplo, KNORR-HELD (1999), RUE e SALVESEN (2000), ARRUDA (2000) e SOUZA e GAMERMAN (2004).

## 1.2 Objetivo do Estudo

Esta dissertação se propõe a construir modelos para prever resultados de partidas de futebol usando a modelagem dinâmica generalizada Bayesiana. A metodologia proposta em RUE e SALVESEN (2000) e mais tarde usada em SOUZA e GAMERMAN (2004) serviu de estímulo e será utilizada como pontapé inicial para os modelos construídos aqui.

Portanto, espera-se que as propriedades adicionais dos modelos aqui propostos, em relação aos já estudados, contribuam para esse tipo de estudo que visa à previsão de resultados em campeonatos de futebol.

Há também a intenção de que a modelagem proposta não seja restrita às condições trabalhadas nessa dissertação. Ou seja, que ela possa ser atribuída, efetuando os ajustes necessários, a qualquer campeonato do mundo, seja qual for o esporte, desde que haja confrontos pareados, como no futebol.

## 1.3 Dados Trabalhados

Esse estudo tem como base os dados referentes ao Campeonato Brasileiro de Futebol Profissional da Série A do ano de 2006. Essa base de dados foi coletada manualmente, rodada após rodada, diretamente do site da CBF (Confederação Brasileira de Futebol), disponível gratuitamente em <http://cbfnews.uol.com.br/seriea/>. Naquele ano, 20 equipes participaram do evento, totalizando 380 jogos.

Daqui por diante, para efeito de notação, as partidas serão denotadas da seguinte forma:

## Time A x Time B.

O Time A é classificado como mandante da partida e o Time B é classificado como visitante da partida. Todavia, existem alguns casos em que dois times enfrentam-se em um estádio neutro, ou seja, em um estádio que não é de domínio de nenhuma das duas equipes que estão se enfrentando ou é de domínio mútuo de tais equipes. Nesses casos, para efeito da modelagem, as duas equipes serão consideradas visitantes. Portanto, o banco de dados bruto é composto pelos resultados de todas as partidas realizadas nesse campeonato, ou seja, essa é toda a informação proveniente dos dados que será considerada para este estudo.

Os modelos propostos nesse estudo estão baseados na distribuição de Poisson, como será mostrado nos próximos capítulos. Essa distribuição, por todas as suas características, mostrou-se uma escolha razoável para tratar números de gols feitos por uma equipe de futebol. As características dessa distribuição, bem como os modelos citados, serão abordados com mais detalhes no capítulo 4.

A figura a seguir mostra essa adequação. Ela ilustra a comparação gráfica das freqüências esperadas (vermelho) e observadas (preto) para os números de gols feitos pelo mandante e visitante no campeonato de 2006.

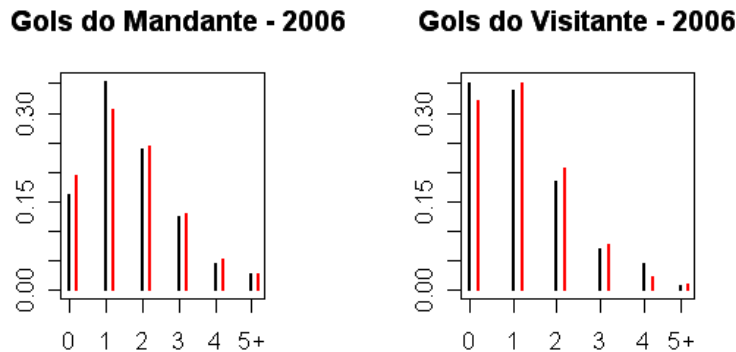


Figura 1.1: Comparação gráfica entre as distribuições dos números de gols de mandantes e visitantes do Campeonato de 2006 com amostras das distribuições teóricas de Poisson, cujos parâmetros ( $\lambda$ ) são as médias das distribuições dos números de gols de mandantes e visitantes.

Além de verificar graficamente a adequação à distribuição de Poisson, dois testes não-paramétricos foram realizados para testar essa hipótese. O primeiro, conhecido como

teste de Qui-quadrado, é aplicado quando estão em comparação duas ou mais categorias independentes, não necessariamente do mesmo tamanho. A hipótese nula ( $H_0$ ) estabelece que a distribuição do número de gols segue a distribuição teórica, nesse caso, a de Poisson.

Supondo  $H_0$  verdadeira, a estatística de teste é dada por:

$$Q = \sum_{i=1}^k \frac{(fo_i - fe_i)^2}{fe_i} \sim \chi_q^2, \text{ aproximadamente.}$$

Em que  $fo_i$  é a frequência observada para a categoria  $i$  (distribuição de gols feitos);  $fe_i$  é a frequência esperada para a categoria  $i$  (distribuição teórica);  $q = k - 1$  é o número de graus de liberdade da distribuição de Qui-quadrado.

As categorias ( $i$ ) escolhidas para os testes foram 0, 1, 2, 3, 4 e 5 ou mais gols, portanto, o valor de  $k$  é 6 e  $Q$  segue aproximadamente uma distribuição de Qui-quadrado com 5 graus de liberdade.

O segundo teste, conhecido como teste de Komolgorov-Smirnov, também se baseia na comparação das distribuições teórica e empírica. Ele compara, em relação ao maior desvio, a função de distribuição teórica da Poisson ( $F(x)$ ) e a função de distribuição empírica ( $F'(x)$ ). Os valores de  $x$  são os mesmos das categorias ( $i$ ) escolhidas para o teste anterior. Os desvios são os valores absolutos das diferenças entre as duas funções de distribuição e são calculados para cada valor de  $x$ . A estatística de teste é o maior entre esses desvios:

$$D = \max_x |F(x) - F'(x)|, x = 1, 2, \dots, k.$$

Sua distribuição, sob  $H_0$ , está bem difundida na literatura e pode ser encontrada em DE GROOT (2002). O valor crítico do teste para o nível de significância de 5% é dado por  $\frac{1,36}{\sqrt{N}}$ , em que  $N$  é número total de observações.

O campeonato de 2006 foi disputado por 20 equipes em dois turnos, resultando em 38 rodadas de 10 jogos, o que corresponde a um total de 760 observações para o número de gols. Portanto, o valor crítico para o teste de Komolgorov-Smirnov é igual a  $\frac{1,36}{\sqrt{380}} = 0,0698$ . Se o valor da estatística de teste for menor do que esse valor, não há razões para rejeitar a hipótese nula que, assim como no primeiro teste, estabelece que a distribuição

dos gols segue a distribuição de Poisson.

Os resultados dos testes para o campeonato brasileiro de 2006 estão dispostos na tabela 1.1.

Tabela 1.1: Testes não-paramétricos para a adequação do ajuste da distribuição do número de gols das equipes no campeonato de 2006 à distribuição de Poisson.

Teste de Qui-quadrado		
Distribuição	Estatística Teste ( $Q$ )	$P$ -valor
Gols Mandante - 2006	4,7780	<b>0,1888</b>
Gols Visitante - 2006	10,4306	0,0152
Teste de Komolgorov-Smirnov		
Distribuição	Estatística Teste ( $D$ )	Valor crítico
Gols Mandante - 2006	0,0325	<b>0,0698</b>
Gols Visitante - 2006	0,0323	<b>0,0698</b>

Tomando como base as evidências dos gráficos e dos testes, não há razões para rejeitar a hipótese de que o número de gols das equipes segue a distribuição de Poisson. Os valores em negrito na tabela 1.1 mostram os testes, sob os quais essa hipótese não foi rejeitada, considerando um nível de significância de 5%.

## 1.4 Inferência Bayesiana

Inferência estatística é o processo pelo qual tiramos conclusões acerca do estudo de interesse, usando informações de uma ou mais amostras e também informações subjetivas, no caso Bayesiano.

Esse tipo de informação pode ser carregada por um parâmetro  $\theta$ . Seu valor é desconhecido, mas o objetivo da inferência é diminuir esse grau de incerteza, seja na modelagem clássica ou Bayesiana.

Um problema importante na inferência estatística é a estimação dos parâmetros de um modelo, deduzidos de um conjunto de dados. Geralmente, inferências sobre os parâmetros que se baseiam em métodos estatísticos clássicos consistem na obtenção de estimadores



de mínimos quadrados ou estimadores de máxima verossimilhança e da suposição de propriedades assintóticas para esses estimadores.

Na abordagem clássica, os estimadores de mínimos quadrados dos parâmetros do modelo, apesar de não-tendenciosos, normalmente distribuídos e de variância mínima, encontram essas propriedades somente assintoticamente. Dessa forma, nem sempre é possível encontrar formas analíticas para os estimadores de mínimos quadrados; portanto, a solução das equações normais, usualmente, requer a utilização de métodos numéricos.

Na abordagem Bayesiana, elimina-se o uso de propriedades assintóticas. Não há distinção entre quantidades observáveis e os parâmetros de um modelo estatístico, todos são considerados quantidades aleatórias. A informação disponível sobre  $\theta$  é resumida probabilisticamente através de uma função de densidade de probabilidade, denotada distribuição a priori  $p(\theta)$ , a qual pode ser aumentada observando-se uma quantidade aleatória  $y$  relacionada com  $\theta$ . A distribuição amostral  $p(y|\theta)$  define essa relação. Para um valor fixo de  $y$ , a função  $L(\theta; y) = p(y|\theta)$  fornece a verossimilhança para cada valor possível de  $\theta$ .

A informação a priori e a verossimilhança são combinadas, levando à distribuição a posteriori de  $\theta$ ,  $p(\theta|y)$ . Esse processo é conhecido no universo estatístico como o teorema de Bayes, e pode ser formalizado por:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y, \theta)d\theta} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}.$$

O denominador dessa fórmula não depende de  $\theta$ , tornando-se assim uma constante normalizadora para a distribuição a posteriori. Sendo assim, esse teorema pode ser escrito em sua forma usual:

$$p(\theta|y) \propto L(\theta; y)p(\theta).$$

Essa forma deixa bem clara a idéia de que a distribuição de  $\theta$  a posteriori é fruto da combinação das informações provenientes dos dados e das informações subjetivas (posteriori  $\propto$  verossimilhança  $\times$  posteriori).

Após observar  $y$ , pode ser desejável prever o próximo instante dado por  $y'$ , também

relacionado com  $\theta$  através de  $p(y'|\theta)$ . Para isso, supõe-se que, condicionado a  $\theta$ ,  $y$  e  $y'$  são independentes, assim pode-se obter a distribuição  $p(y'|y)$ , dada por:

$$\begin{aligned}
 p(y'|y) &= \int p(y', \theta|y)d\theta; \\
 &= \int p(y'|\theta, y)p(\theta|y)d\theta; \\
 &= \int p(y'|\theta)p(\theta|y)d\theta.
 \end{aligned} \tag{1.1}$$

A partir daí, percebe-se a característica de atualização do modelo, tal que a distribuição a posteriori de  $\theta$ , após observar  $y$ , age como a priori no cálculo da posteriori no instante seguinte, em que  $y'$  ainda não foi observado. Após a observação de  $y'$ , uma nova posteriori é obtida reutilizando-se o teorema de Bayes.

Sendo assim, trazendo para uma forma mais geral, tem-se para as quantidades  $y_1, y_2, \dots, y_n$  independentes dado  $\theta$ :

$$\begin{aligned}
 p(\theta|y_1) &\propto l_1(\theta; y_1)p(\theta); \\
 p(\theta|y_2, y_1) &\propto l_2(\theta; y_2)p(\theta|y_1); \\
 &\propto l_2(\theta; y_2)l_1(\theta; y_1)p(\theta); \\
 &\vdots \qquad \qquad \qquad \vdots \quad \ddots \\
 p(\theta|y_n, y_{n-1}, \dots, y_1) &\propto l_n(\theta; y_n)p(\theta|y_{n-1}, \dots, y_1); \\
 &\propto \left[ \prod_{i=1}^n l_i(\theta; y_i) \right] p(\theta).
 \end{aligned}$$

O desafio da inferência Bayesiana é calcular os valores esperados de funções particulares do parâmetro  $\theta$  que resumem a densidade a posteriori. A informação, a priori a respeito da interpretação dos parâmetros, pode ser expressa utilizando-se prioris informativas ou prioris não-informativas, caso não haja opinião sólida sobre os parâmetros do modelo. No segundo caso, uma saída bastante usada consiste em considerar todos os

possíveis valores de  $\theta$  como igualmente prováveis.

## 1.5 Inferência por Simulação

Na grande maioria dos estudos Bayesianos, a distribuição a posteriori é complexa e não pode ser obtida de forma fechada. Nesses casos, é comum optar por fazer inferência via simulações estocásticas. Um método bastante utilizado para realizar esse tipo de inferência é conhecido como método de simulação de Monte Carlo via cadeias de Markov (MCMC).

O fundamento desse método é estimar características probabilísticas de modelos. Eles consistem em técnicas de simulação iterativas para obter uma amostra da distribuição a posteriori e calcular estimativas amostrais de características dessa distribuição.

A idéia dos métodos de MCMC é construir uma cadeia de Markov, cuja distribuição de equilíbrio se aproxime o máximo possível da distribuição de interesse (posteriori), após um número finito de simulações dessa cadeia. Usualmente, opta-se por trabalhar com duas cadeias, com o objetivo de verificar a convergência delas. Atingida a convergência, dá-se origem a uma amostra da distribuição de interesse aproximada.

Considerando que a densidade conjunta de  $\theta_1, \dots, \theta_n$  é dada por  $p(\theta) = p(\theta_1, \dots, \theta_n)$  e que  $q(\theta'|\theta)$  define a distribuição condicional das transições do estado  $\theta$ , é possível construir uma cadeia com probabilidades de transição, em que cada estado pode ser obtido de outro com um número finito de iterações. Sendo assim, com uma quantidade razoável de iterações, dependendo de cada situação, chega-se à distribuição de equilíbrio.

Entre os métodos de construção da cadeia de Markov mais utilizados, estão o método de Metropolis-Hastings e o Amostrador de Gibbs. As descrições de ambos podem ser encontradas, mais detalhadamente, em GAMERMAN (1996) e MIGON e GAMERMAN (1999).

- Metropolis-Hastings

Nos algoritmos de Metropolis-Hastings, um valor é gerado a partir de uma distribuição auxiliar e aceito com uma dada probabilidade. Esse mecanismo de

correção garante a convergência da cadeia para a distribuição de equilíbrio que, neste caso, é a distribuição a posteriori (GAMERMAN, 1996). Suponha que no instante  $t$  a cadeia esteja no estado  $\theta_t$  e um valor  $\theta'$  é gerado de uma distribuição proposta: uma função de probabilidade de transição de estado apropriada:  $q(\cdot|\theta)$ . O novo valor é aceito com probabilidade:

$$\delta(\theta|\theta') = \min\left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)}\right),$$

como o novo estado da cadeia no instante  $t + 1$ , ou é rejeitado com probabilidade  $1 - \delta(\theta, \theta')$  e permanece no estado  $\theta$ . O algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

1. inicialize o contador de iterações e especifique um valor inicial  $t = 0$ ;
2. escolha o número de iterações para as cadeias a serem geradas;
3. gere um novo valor  $\theta'$  da distribuição  $q(\cdot|\theta)$  proposta;
4. calcule a probabilidade de aceitação  $\delta(\theta, \theta')$  e gere  $u \sim U[0, 1]$ ;
5. se  $u \leq \delta(\theta, \theta')$ , então aceite o novo valor e faça  $\theta^{t+1} = \theta'$ , caso contrário, rejeite e faça  $\theta^{t+1} = \theta$ ;
6. incremente o contador de  $t$  para  $t+1$  e volte ao passo 3 até completar o número de iterações;
7. analise se as cadeias convergiram. Caso contrário, retorne ao passo 2, aumentando o número de iterações. Deve-se ressaltar que é possível interromper o processo iterativo assim que a cadeia convirja para a distribuição de equilíbrio  $p$ .

- Amostrador de Gibbs

Trata-se de um esquema iterativo de amostragem de uma cadeia de Markov, cujas probabilidades de transição dos estados são realizadas a partir das distribuições condicionais completas. Supondo que  $p(\theta)$  é a distribuição de interesse, considera-se  $p_i(\theta_i) = p_i(\theta_i|\theta_{-i})$  como a distribuição condicional de  $\theta_i$ , dados os

valores de todos os outros  $\theta$ 's, e assume-se que é possível tirar amostras dessas distribuições para cada  $i = 1, \dots, n$ .

A partir daí, deve-se especificar os valores iniciais  $\theta^0 = (\theta_1^0, \dots, \theta_n^0)$  para todas as quantidades desconhecidas do modelo. Sendo assim, na  $j$ -ésima iteração, a cadeia se encontra no estado  $\theta^{(j)}$  e a posição da cadeia na iteração seguinte ( $j + 1$ ) é obtida da seguinte forma:

1. gere  $\theta_1^{(j+1)}$  de  $p_1(\theta_1 | \theta_2^{(j)}, \dots, \theta_n^{(j)})$ ;
2. gere  $\theta_2^{(j+1)}$  de  $p_2(\theta_2 | \theta_1^{(j+1)}, \theta_3^{(j)}, \dots, \theta_n^{(j)})$ ;
3. continue gerando para  $i = 3, \dots, n$  até o último passo  $\theta_n^{(j+1)}$ .

Por fim, chega-se ao vetor  $\theta^{(j+1)}$  e todo procedimento deve ser realizado para o próximo estado ( $j + 2$ ). Sob certas condições, a distribuição de  $\theta^{(j)}$  tenderá para a distribuição de interesse.

Nesse estudo, optou-se por uma análise bastante utilizada de avaliação da convergência: a inspeção gráfica. Nessa análise, duas cadeias são comparadas ao longo do tempo, esperando que ambas obtenham um comportamento convergente a medida que o número de iterações aumenta.

Todo o processo de inferência desse estudo foi feito com a ajuda do programa WINBUGS. Nele, é possível tratar modelos estatísticos complexos, além de estarem implementados os algoritmos de Metropolis-Hastings e o Amostrador de Gibbs.

## 1.6 Organização do Trabalho

No capítulo 2, inicia-se uma breve introdução aos MLD sob a hipótese de normalidade dos dados. Nas seções 2.2, 2.3 e 2.4 são apresentadas, respectivamente, a descrição, a inferência e a previsão para os MLD sob normalidade, que é o tipo de MLD mais simples.

No capítulo 3 são apresentados inicialmente os modelos de interesse do estudo: os Modelos Lineares Dinâmicos Generalizados (MLDG), nos quais estende-se a suposição de normalidade para dados provenientes da família exponencial. As seções 3.3, 3.4 e 3.5

apresentam respectivamente, a descrição, a inferência e a previsão para um caso geral dos MLDG.

O capítulo 4 descreve inicialmente algumas características da distribuição de Poisson. Depois, na seção 4.2, são apresentados os modelos propostos por RUE e SALVESEN (2000) e usados em SOUZA e GAMERMAN (2004) e, posteriormente, os modelos propostos por esse estudo.

O capítulo 5 apresenta os resultados obtidos, bem como a comparação entre modelos utilizados e de outros autores e a repercussão das alterações nos modelos.

Finalmente, as Conclusões e considerações finais do estudo são apresentadas no capítulo 6.

# Capítulo 2

## Modelos Lineares Dinâmicos

### 2.1 Introdução

Nas últimas décadas, a estatística vem se desenvolvendo de forma acelerada, principalmente quando se trata de modelagem estruturada em diversos níveis com influência do tempo. Os Modelos Lineares Dinâmicos (MLD), principais responsáveis por esse crescimento, permitem que os parâmetros evoluam com o passar do tempo e, com isso, propiciam um tratamento mais razoável ao estudo de séries temporais. Além disso, existe a possibilidade de combinar informações externas aos dados na análise e na previsão das variáveis de interesse, através da distribuição a priori.

Embora tenha sido introduzida em meados da década de 1970, o uso desta metodologia só passou a se difundir a partir da segunda metade da década de 1980, devido à inexistência, até então, de pacotes estatísticos que permitissem a sua utilização em larga escala. Em linhas gerais, essa metodologia permite estimar os componentes de tendência e sazonalidade das séries temporais de forma sequencial e dinâmica, através de um mecanismo que atualiza as estimativas a cada período de tempo. Isto confere à ela vantagens preditivas, além de permitir a análise do padrão dinâmico das séries através do estudo das variações dos parâmetros estimados.

Além das vantagens citadas acima, existem alguns diferenciais na modelagem dinâmica Bayesiana em relação à modelagem clássica que devem ser ressaltados:

- as séries são representadas através de componentes que são diretamente interpretáveis;
- explicita não apenas as estimativas dos componentes a cada instante de tempo, mas também indica uma medida de incerteza associada a tais estimativas;
- o modelo é capaz de se alterar ou se adaptar ao longo da amostra com base em um mecanismo de transição de um período para outro e gerando uma representação específica para cada instante de tempo.

A última característica é o principal destaque da modelagem Bayesiana. Ela é a responsável pela natureza dinâmica da metodologia e fornece a base para as intervenções subjetivas dos usuários do modelo.

Por exemplo, é comum os usuários terem conhecimento de eventos passados que explicam certas características apresentadas pelo histórico da série, ou de eventos futuros que irão alterar o comportamento da série. Tais informações devem ser incorporadas no sistema de análise e previsão, porém não estão presentes nos dados. A abordagem Bayesiana permite que estas informações sejam formalmente incorporadas no processo de estimação e previsão dos modelos.

Esses modelos têm sido amplamente aplicados em diversas áreas, como na economia, medicina, geologia e nos esportes. A última se destaca pelo crescimento do número de trabalhos científicos, visando principalmente à previsão de resultados.

Tal crescimento não se deu por acaso, pois trata-se de uma área ampla de informações e que há algumas décadas não era explorada cientificamente com tanto afinco. Haja vista a natureza inconstante da maioria dos esportes, a modelagem dinâmica se adequa sem maiores problemas, pois consegue explicar as variações temporais características dessa área.

Como em um bom modelo Bayesiano, no MLD, busca-se aliar todas as informações relevantes disponíveis nos dados (verossimilhança) com o conhecimento prévio acerca do estudo de interesse, seja ele teórico ou por experiência (priori).

A idéia básica aqui é que as observações flutuam em torno de uma média. Entretanto, diferentemente do que ocorre nos modelos estáticos, essa média está sujeita a pequenas



variações ao longo do tempo. Além disso, esse tipo de variação é essencialmente estocástica.

Nesse capítulo, trataremos três aspectos dos MLD: a descrição do modelo para dados normais, seu processo de inferência e a previsão. Vale ressaltar que toda a teoria exposta nesse capítulo é uma breve parte de todo o conhecimento sobre MLD explicado em WEST e HARRISON (1997).

## 2.2 Descrição do Modelo

Essa modelagem pressupõe que os dados sejam normalmente distribuídos. Como explicitado na seção 1.3, os modelos apresentados nesse estudo se baseiam na distribuição de Poisson. Portanto, começaremos descrevendo os MLD para dados normais no capítulo 2 e, mais adiante, nos capítulos 3 e 4, trataremos de casos mais gerais, nos quais não ficaremos mais restritos a normalidade dos dados.

Nos MLD restritos à normalidade, daqui por diante apenas MLD, a equação que descreve as observações é dada por:

$$(y_t|\theta_t) \sim N(F'_t\theta_t, V_t).$$

Para facilitar os cálculos, supõe-se que a matriz de covariâncias  $V_t$  é conhecida, assim como a matriz  $F_t$ , para  $t = 1, \dots, T$ . O vetor  $\theta_t$  é formado pelo conjunto de parâmetros do modelo no tempo  $t$  e  $F'_t\theta_t$  é o vetor das médias dos parâmetros do modelo e, como dito anteriormente, sofre variações ao longo do tempo. Esse processo é modelado por um passeio aleatório, conforme descrito abaixo:

$$\theta_t = G_t\theta_{t-1} + \omega_t, \text{ em que } \omega_t \sim N(0, W_t).$$

Portanto, a evolução nos parâmetros é controlada pelo termo aleatório  $\omega_t$ , através da sua variância  $W_t$ , também suposta conhecida para  $t = 1, \dots, T$ . Portanto, percebe-se que quanto maior o valor de  $W_t$ , maior será a variação entre os valores de um parâmetro em instantes consecutivos no tempo. Com um valor cada vez menor de  $W_t$ , o MLD descrito

acima se aproxima do modelo Bayesiano chamado estático, no qual os parâmetros não evoluem no tempo.

Assim como a matriz  $F_t$ , a matriz  $G_t$  também é conhecida para  $t = 1, \dots, T$ .

Como dito anteriormente, o MLD leva em conta não só informações provenientes dos dados, mas também considera o conhecimento do pesquisador ou de especialistas no assunto de interesse. Considere  $D_0$  toda essa informação inicial disponível. Logo, define-se  $D_1 = \{y_1, D_0\}$  como sendo toda a informação disponível sobre o processo em estudo, após observá-lo no seu primeiro instante. Esse modelo é conhecido como fechado. O mesmo raciocínio é aplicado para os demais instantes de tempo, no qual tem-se:

$$D_t = \{y_t, D_{t-1}\}, \text{ com } t = 1, 2, \dots, T.$$

Antes de observar o tempo  $t$ , toda a informação disponível é  $D_{t-1}$ . Essa informação é usada como priori para os parâmetros no próximo instante ( $\theta_t$ ). Logo, a distribuição a priori  $p(\theta_t|D_{t-1})$  e a distribuição a posteriori  $p(\theta_t|D_t)$  formam um consistente ciclo de passagem de informação ao longo do tempo.

É justamente nessa etapa que espera-se que passe de um instante para o outro todo o conhecimento adquirido pelo modelo sobre o processo. É também nessa etapa de passagem de conhecimento do modelo, que ele consegue indentificar comportamentos importantes nos dados, como tendência e sazonalidade.

A cada instante de tempo  $t$ , a estimação dos componentes não observáveis,  $\theta_t$ , tem que ser feita a partir do vetor de observações  $y = (y_1, y_2, \dots, y_t)$ . A proposta da metodologia dos MLD é tratar esta questão como um problema de estimação de modelos em espaço de estados, para o qual existe uma solução pronta, dada pela utilização de um algoritmo conhecido como Filtro de Kalman, que será apresentado na próxima seção.

Portanto, o MLD para cada instante  $t$  pode ser formalizado por:

$$\left\{ \begin{array}{l} \text{eq. de observação: } y_t = F_t' \theta_t + \nu_t, \quad \nu_t \sim N[0, V_t]; \\ \text{eq. do sistema: } \theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t]; \\ \text{informação inicial: } \theta_1 | D_0 \sim N[m_0, C_0]. \end{array} \right. \quad (2.1)$$

$y_t$  é o vetor de observações de dimensão  $r$ , para  $t = 1, \dots, T$ ,

$\theta_t$  é o vetor de parâmetros de dimensão  $n$ , para  $t = 1, \dots, T$ ,

$F_t$  é uma matriz conhecida de ordem  $n$  por  $r$ , para  $t = 1, \dots, T$ ,

$G_t$  é uma matriz conhecida de ordem  $n$  por  $n$ , para  $t = 1, \dots, T$ ,

$V_t$  é uma matriz de covariâncias conhecida de ordem  $r$  por  $r$ , para  $t = 1, \dots, T$ , e

$W_t$  é uma matriz de covariâncias conhecida de ordem  $n$  por  $n$ , para  $t = 1, \dots, T$ .

A primeira equação é chamada equação das observações, pois descreve a forma como cada observação é gerada para um dado estado do sistema no instante  $t$ . A segunda, por sua vez, é chamada equação do sistema ou de evolução, porque descreve a transição entre o instante  $t - 1$  e o instante  $t$ , do sistema de geração das observações, para  $t = 1, \dots, T$ .  $\theta_t$  é também conhecido como o vetor de estados do sistema.

O objetivo da estimação do MLD consiste em encontrar estimativas para o vetor de estados, a cada instante  $t$ , com base nas observações disponíveis, para todo  $t$ .

Assume-se que  $\nu_t$  e  $\omega_t$ , para  $t = 1, \dots, T$ , são independentes entre si e de  $\theta_1|D_0$ . Portanto, pode-se definir o modelo pela quádrupla  $\{F_t, G_t, V_t, W_t\}$ , pois esse conjunto contém toda a informação referente ao modelo.

Casos particulares do MLD geral apresentado são amplamente usados, inclusive para fins teóricos. O mais conhecido deles é o MLD de primeira ordem descrito por  $\{1, 1, V_t, W_t\}$ .

## 2.3 Inferência no MLD

O processo de inferência nos MLD, para  $t = 1, \dots, T$ , pode ser descrito por:

$$\theta_{t-1}|D_{t-1} \xRightarrow{\text{Evolução}} \theta_t|D_{t-1} \xRightarrow{\text{Atualização}} \theta_t|D_t.$$

A evolução e a atualização são feitas pela equação do sistema e o teorema de Bayes, respectivamente.

Assumimos verdadeiras as três propriedades a seguir para as perturbações definidas por  $\nu_t$  e  $\omega_t$ , para  $t = 1, \dots, T$ .

- Normalidade

$$\begin{aligned}\nu_t &\sim N[0, V_t]; \\ \omega_t &\sim N[0, W_t].\end{aligned}$$

- Ausência de auto-correlação

$$\begin{cases} E[\nu_t, \nu_{t-k}] = 0, & k = \pm 1, \pm 2, \dots; \\ E[\omega_t, \omega_{t-k}] = 0. \end{cases}$$

- Não auto-correlacionadas entre si

$$E[\nu_t, \omega_{t-k}] = 0, \quad k = 0, \pm 1, \pm 2, \dots$$

Portanto, o processo de estimação fica determinado pelo seguinte conjunto de equações, para  $t = 1, \dots, T$ :

$$a_t = E[\theta_t | D_{t-1}] = G_t m_{t-1};$$

$$R_t = V[\theta_t | D_{t-1}] = G_t C_{t-1} G' + W_t;$$

$$m_t = E[\theta_t | D_t] = a_t + \frac{R_t F_t (y_t - F_t' a_t)}{Q_t};$$

$$C_t = V[\theta_t | D_t] = R_t + \frac{R_t F_t F_t' R_t}{Q_t},$$

em que  $Q_t = F_t' R_t F_t + V_t$ .

Esse conjunto de equações forma o sistema chamado Filtro de Kalman. As duas primeiras equações são conhecidas como as equações de predição do Filtro. Com elas, obtém-se estimativas para o vetor de estados  $\theta_t$  e para sua matriz de variância  $W_t$ , usando-se apenas as informações disponíveis no instante  $t - 1$ .

As equações restantes são chamadas equações de atualização, pois atualizam as estimativas  $a_t$  e  $R_t$  usando a última observação disponível,  $y_t$ . A partir daí, estimativas de  $\theta_t$  podem ser geradas usando-se as informações disponíveis até o instante  $t$ .

Nota-se que  $m_t$  é a soma de  $a_t$  mais um termo de correção do erro de previsão do sistema em  $t$ , dado por  $y_t - F_t' a_t$ . Portanto, a estimativa  $a_t$  (que usa apenas a informação disponível no instante  $t - 1$ ) é atualizada após observar  $y_t$ . O fator de correção, dado por  $\frac{R_t F_t}{Q_t}$ , é chamado Ganho do Filtro.

O estimador gerado pelo Filtro de Kalman minimiza o erro quadrático médio de previsão dentre todos os estimadores lineares. Na hipótese de normalidade dos resíduos das observações ( $\nu_t$ ), ele estende esta propriedade para todos os estimadores, lineares ou não.

Na formulação Bayesiana, a aplicação do filtro de Kalman é representada em termos das distribuições a priori e a posteriori para o vetor de estados em  $t$ .

- Posteriori em  $t - 1$

$$\theta_{t-1} | D_{t-1} \sim N[m_{t-1}, C_{t-1}].$$

- Priori em  $t$

$$\theta_t | D_{t-1} \sim N[a_t, R_t], \text{ em que } a_t = G_t m_{t-1};$$

$$R_t = G_t C_{t-1} G_t' + W_t.$$

- Posteriori em  $t$

$$\theta_t | D_t \sim N[m_t, C_t], \text{ em que } m_t = a_t + \frac{R_t F_t (y_t - F_t' a_t)}{Q_t};$$

$$C_t = R_t + \frac{R_t F_t F_t' R_t}{Q_t}.$$

Até o momento não se havia falado muito sobre as especificações de  $V_t$  e  $W_t$ . A primeira, definida como variância das observações, geralmente é desconhecida, porém o pesquisador ou um especialista possui algum conhecimento a priori em torno desse parâmetro. Uma solução razoável para esse caso é adotar  $V_t = V$  e fazer  $W_t = V W_t^*$ , sendo  $W_t^*$  conhecido para  $t = 1, \dots, T$ .

No caso do termo  $W_t$ , conhecido como variância do sistema, não se aplica o mesmo tratamento analítico dado a  $V_t$ . Existe uma idéia bem definida, chamada fator de desconto, para processar a incerteza acerca de  $W_t$ . Essa prática concentra-se no objetivo

de descontar a informação contida nas observações mais antigas, ou seja, quanto maior for o valor do fator de desconto, maior será a quantidade de informação que passará do instante  $t - 1$  para  $t$ .

Considerando a equação do sistema dada em 2.1, temos:

$$\begin{cases} V[\theta_{t-1}|D_{t-1}] = C_{t-1}; \\ V[\theta_t|D_{t-1}] = R_t = G_t C_{t-1} G_t' + W_t = P_t + W_t. \end{cases}$$

Logo, tem-se  $W_t = R_t - P_t$ , e definindo o fator de desconto ( $\delta$ ), tal que  $R_t = \frac{P_t}{\delta}$ , chega-se à forma final de  $W_t$ , dada por:

$$W_t = P_t(\delta^{-1} - 1).$$

Não existe um valor adequado para o fator de desconto, sua escolha vai depender do problema proposto, apenas é recomendado que alguns valores sejam testados. São considerados baixos os valores de  $\delta$  menores que 0,8, pois tendem a introduzir muita incerteza no sistema. Tipicamente em sistemas sem variações bruscas, são preferidos valores acima de 0,9.

Portanto, observa-se que no MLD, que supõe normalidade dos dados, todos os cálculos podem ser resolvidos de forma analítica, devido ao uso de prioris normais, já que o modelo normal é conjugado à distribuição normal. No próximo capítulo, uma classe mais geral de modelos dinâmicos é apresentada, na qual a hipótese de normalidade dos dados é relaxada, resultando em uma análise um pouco diferente da mostrada aqui.

## 2.4 Previsão

A previsão nos MLD segue o mesmo cálculo feito na equação 1.1. Passada a etapa de evolução na qual obtemos  $\theta_t|D_{t-1}$ , combinamos essa informação com a distribuição  $y_t|\theta_t$ , que é determinada pela equação de observação apresentada em 2.1.

Portanto, temos:

$$\begin{aligned}
p(y_t|D_{t-1}) &= \int p(y_t, \theta_t|D_{t-1})d\theta_t; \\
&= \int p(y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})d\theta_t; \\
&= \int p(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t.
\end{aligned}$$

Conforme dito anteriormente, a propriedade de normalidade de  $y_t|\theta_t$  e  $\theta_t|D_{t-1}$  permite que essa integral seja resolvida analiticamente. Essa solução é tal que:

$$y_t|D_{t-1} \sim N[f_t, Q_t], \text{ em que } f_t = F'_t a_t \text{ e } Q_t = F'_t R_t F_t + V_t.$$

Assim, supondo que deseja-se trabalhar com um horizonte de previsão de  $h$ -passos-a-frente, a partir de um instante  $t$  qualquer, temos:

$$\begin{aligned}
\theta_{t+h}|D_t &\sim N[a_t(h), R_t(h)], \text{ em que } a_t(h) = G_{t+h} a_t(h-1), \\
R_t(h) &= G_{t+h} R_t(h-1) G'_{t+h} + W_{t+h},
\end{aligned}$$

e

$$\begin{aligned}
y_{t+h}|D_t &\sim N[f_t(h), Q_t(h)], \text{ em que } f_t(h) = F'_{t+h} a_t(h), \\
Q_t(h) &= F'_{t+h} R_t(h) F_{t+h} + V_{t+h},
\end{aligned}$$

em que  $a_t(0) = m_t$ ,  $R_t(0) = C_t$  e  $h = 1, 2, \dots$

# Capítulo 3

## Modelo Linear Dinâmico Generalizado

### 3.1 Introdução

A teoria descrita no capítulo anterior está baseada na hipótese de normalidade dos dados com os quais se deseja trabalhar. Porém, como ocorre em diversas áreas de estudo, nem sempre os dados podem ser aproximados pela distribuição normal. Nesse caso, deve-se ascender a um nível mais geral, no qual se encontram os MLDG.

Assim como no capítulo 2, toda a teoria apresentada nesse capítulo está integralmente disponível em WEST e HARRISON (1997).

Esse tipo de modelagem preserva as mesmas características dos MLD. Entretanto, os dados são supostamente provenientes de uma distribuição pertencente a família exponencial. Portanto, pode-se afirmar que os MLD formam um caso particular dos MLDG, tendo em vista que a normal também é membro da família exponencial.

Ainda que a distribuição dos dados não siga o modelo normal, em alguns casos opta-se pela transformação dos dados para chegar a essa condição. Um exemplo para esse caso seria o modelo Gama, no qual a transformação para a normal pode ser bem útil, observando os ajustes necessários para a interpretação dos parâmetros fora de sua escala original.

Optando-se pela transformação dos dados para a distribuição normal, o modelo a ser



usado é o mesmo descrito no capítulo anterior. Deve-se ter o cuidado ao trabalhar com transformações, pois nem sempre encontramos valores consistentes quando os parâmetros retornam às suas escalas originais.

Quando dispõe-se de observações discretas, transformações geralmente não são apropriadas e também nesses casos a teoria dos MLDG é bem-vinda. Vale ressaltar que é sempre recomendável trabalhar com a escala original dos dados, pois torna a apresentação do modelo mais clara e a interpretação dos resultados fica facilitada.

A grande maioria das distribuições conhecidas, tanto discretas quanto contínuas, atendem à uma série de propriedades que as tornam membros de uma grande família de distribuições, conhecida como família exponencial. Essa característica faz com que o uso dessas distribuições seja permitido no conceito de Modelos Dinâmicos.

## 3.2 Distribuições na Família Exponencial

Seja  $\{y_1, y_2, \dots\}$  a série temporal das observações em estudo, a distribuição das observações é membro da família exponencial se a sua função de probabilidade (caso discreto) ou função densidade de probabilidade (caso contínuo) puder ser escrita da seguinte forma:

$$p(y_t|\eta_t, V_t) = \exp\{\phi_t[Y_t(y_t)\eta_t - a(\eta_t)]\}b(y_t, V_t). \quad (3.1)$$

$\eta_t$  é dito o parâmetro natural da distribuição e possui uma relação linear com o vetor de estados  $\theta_t$ .  $V_t > 0$ , supostamente conhecido para todo  $t$ , é o parâmetro de escala, para o qual  $\phi_t = V_t^{-1}$  define a precisão da distribuição.

Outras características das distribuições pertencentes à família exponencial são listadas a seguir:

- para valores fixos de  $y_t$ ,  $p(y_t|\eta_t, V_t)$  é a verossimilhança para  $\eta_t$ , dependendo de  $y_t$  apenas pelo valor  $Y_t(y_t)$ ;

- a função  $a(\eta_t)$  é assumida duas vezes diferenciável em  $\eta_t$ . Daí, tem-se a função da média:

$$\mu_t = E[Y_t(y_t)|\eta_t, V_t] = \frac{\partial a(\eta_t)}{\partial \eta_t} = a'(\eta_t);$$

- desse modo, tem-se também a função da variância:

$$V[Y_t(y_t)|\eta_t, V_t] = V_t a''(\eta_t).$$

Como dito anteriormente, muitas distribuições conhecidas fazem parte dessa extensa família, entre elas podemos citar as distribuições binomial, poisson, beta, gama e a própria normal.

### 3.3 Descrição do Modelo

Basicamente o raciocínio de construção do modelo é o mesmo apresentado no capítulo anterior para os MLD. A diferença aqui é que nem sempre chega-se à uma distribuição a posteriori analiticamente tratável. Quando isso ocorre, usamos métodos de simulação para amostrar da distribuição aproximada da posteriori, como foi mostrado na seção 1.5.

A equação de observação do MLDG é dada pela equação 3.1. Seja  $y_t$  o vetor de observações de dimensão  $r$ , para  $t = 1, \dots, T$ , o restante da descrição do modelo é dada por:

$$\left\{ \begin{array}{ll} \text{eq. de ligação:} & g(\eta_t) = \lambda_t = F_t' \theta_t; \\ \text{eq. do sistema:} & \theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t]; \\ \text{informação inicial:} & \theta_1 | D_0 \sim N[m_0, C_0]. \end{array} \right. \quad (3.2)$$

$\theta_t$  é o vetor de parâmetros de dimensão  $n$ , para  $t = 1, \dots, T$ ,

$F_t$  é uma matriz conhecida de ordem  $n$  por  $r$ , para  $t = 1, \dots, T$ ,

$G_t$  é uma matriz conhecida de ordem  $n$  por  $n$ , para  $t = 1, \dots, T$ ,

$W_t$  é uma matriz de covariâncias conhecida de ordem  $n$  por  $n$ , para  $t = 1, \dots, T$ ,  
 $\lambda_t = F_t' \theta_t$  é uma função linear do vetor de parâmetros, para  $t = 1, \dots, T$ , e  
 $g(\eta_t) = \lambda_t$  é uma função contínua, monótona e conhecida, chamada função de ligação,  
para  $t = 1, \dots, T$ .

Fica claro notar que trata-se de uma extensão do MLD, que supõe normalidade. Nota-se também que a única quantidade desconhecida em  $p(y_t | \eta_t, V_t)$  é o parâmetro natural  $\eta_t$ . Logo,  $\eta_t$  é a única fonte de incerteza sobre a distribuição  $p(y_t | D_{t-1})$ . A solução razoável adotada para esse caso é usar uma priori para  $\eta_t$ , dada por  $p(\eta_t | D_{t-1})$ . Além disso, se a priori escolhida para  $\eta_t$  é conjugada, então a posteriori  $p(\eta_t | D_t)$  pertence à mesma família que a priori.

Os erros  $w_t$  são supostos normais, não-correlacionados e independentes de  $y_t$ , condicional a  $\eta_t$ .

### 3.4 Inferência no MLDG

A inferência nos MLDG basicamente não se diferencia muito do que foi visto no capítulo anterior, no qual assumimos a hipótese de normalidade dos dados. Aqui, mesmo estando em termos mais gerais, os princípios utilizados na estimação dos parâmetros são os mesmos, salvo alguns ajustes que devem ser considerados quando se trabalha com a família exponencial.

No MLD, ficou claro que os cálculos necessários para definir todo o processo de inferência dos dados segue de maneira intuitiva e analítica. O mesmo nem sempre ocorre nessa classe de modelos.

Vale lembrar que, para os casos nos quais a distribuição a posteriori não pode ser obtida de forma analítica, será necessário recorrer a métodos numéricos para obter aproximações de tal distribuição. Métodos esses que foram descritos na seção 1.5.

Assim como no MLD, o processo de inferência do MLDG, para  $t = 1, \dots, T$ , é dado por:

$$\theta_{t-1} | D_{t-1} \xRightarrow{\text{Evolução}} \theta_t | D_{t-1} \xRightarrow{\text{Atualização}} \theta_t | D_t.$$

Como no capítulo 2 trabalhamos com a normalidade dos dados, foi possível especificar as distribuições dos vetores de parâmetros mostradas acima. No MLDG, especificaremos parcialmente essas distribuições, pois nem sempre trabalharemos com distribuições conjugadas, como foi o caso da normal.

Portanto, as distribuições dos vetores de parâmetros serão especificadas por suas médias e variâncias.

- Posteriori em  $t - 1$

$$\theta_{t-1} | D_{t-1} \sim [m_{t-1}, C_{t-1}].$$

- Priori em  $t$

$$\theta_t | D_{t-1} \sim [a_t, R_t], \text{ em que } a_t = G_t m_{t-1};$$

$$R_t = G_t C_{t-1} G_t' + W_t.$$

Além da priori para  $\theta_t$ , temos a distribuição a priori conjunta de  $\lambda_t$  e  $\theta_t$ , parcialmente especificada por seu vetor de médias e sua matriz de covariâncias:

$$\begin{pmatrix} \lambda_t \\ \theta_t \end{pmatrix} \Big| D_{t-1} \sim \left[ \begin{pmatrix} f_t \\ a_t \end{pmatrix}, \begin{pmatrix} q_t & F_t' R_t \\ R_t F_t & R_t \end{pmatrix} \right], \text{ em que } f_t = F_t' a_t \text{ e } q_t = F_t' R_t F_t.$$

Vale lembrar que,  $\lambda_t$  possui uma relação linear com o vetor de parâmetros, representada por  $\lambda_t = g(\eta_t) = F_t' \theta_t$ . Logo, fica claro que a média e a variância especificadas aqui coincidem com aquelas definidas para o MLD.

Após observado o valor de  $y_t$ , é possível obter, pela atualização do modelo, a distribuição a posteriori de  $\lambda_t$ , dada por:

$$\lambda_t | D_t \sim [f_t^*, q_t^*], \text{ em que } f_t^* = f_t + (y_t - f_t) \frac{q_t}{q_t + V_t} \text{ e } q_t^* = q_t - \frac{q_t^2}{q_t + V_t}. \quad (3.3)$$

Assim como ocorre no MLD, o objetivo do processo de inferência é, através das atualizações via teorema de Bayes, chegar à distribuição a posteriori de  $\theta_t$ . A distribuição a posteriori é obtida a partir da distribuição a posteriori conjunta dada por:

$$\begin{aligned}
p(\lambda_t, \theta_t | D_t) &\propto p(\lambda_t, \theta_t | D_{t-1})p(y_t | \lambda_t); \\
&\propto [p(\theta_t | \lambda_t, D_{t-1})p(\lambda_t | D_{t-1})]p(y_t | \lambda_t); \\
&\propto p(\theta_t | \lambda_t, D_{t-1})[p(\lambda_t | D_{t-1})p(y_t | \lambda_t)]; \\
&\propto p(\theta_t | \lambda_t, D_{t-1})p(\lambda_t | D_t).
\end{aligned}$$

Logo, dado  $\lambda_t$  e  $D_{t-1}$ ,  $\theta_t$  é independente de  $y_t$ , e portanto, podemos calcular a distribuição a posteriori para  $\theta_t$  da seguinte maneira:

$$\begin{aligned}
p(\theta_t | D_t) &= \int p(\lambda_t, \theta_t | D_t) d\lambda_t; \\
&= \int p(\theta_t | \lambda_t, D_{t-1})p(\lambda_t | D_t) d\lambda_t.
\end{aligned}$$

O segundo termo da integral,  $p(\lambda_t | D_t)$ , foi parcialmente especificado na equação 3.3. O primeiro termo fica parcialmente definido por  $E[\theta_t | \lambda_t, D_{t-1}]$  e  $V[\theta_t | \lambda_t, D_{t-1}]$ .

A esperança e a variância mencionadas são calculadas a partir da distribuição a priori conjunta de  $\lambda_t$  e  $\theta_t$ . Sob as hipóteses de generalidade que foram supostas até aqui, esses cálculos não saem de forma analítica e, por isso, deve-se estimá-los usando o estimador linear de Bayes. Os valores ótimos dos estimadores para a média e a variância são:

$$\hat{E}[\theta_t | \lambda_t, D_{t-1}] = a_t + \frac{R_t F_t (\lambda_t - f_t)}{q_t},$$

e

$$\hat{V}[\theta_t | \lambda_t, D_{t-1}] = R_t - \frac{R_t F_t F_t' R_t}{q_t}.$$

O cálculo dos estimadores acima é importante, pois como todo o raciocínio visto até aqui no caso generalizado para determinar a posteriori de  $\theta_t$ , precisa-se especificar sua média e variância. Esse cálculo é feito a partir de  $p(\theta_t | D_t)$  pelas seguintes equações:

$$E[\theta_t|D_t] = E[E(\theta_t|\lambda_t, D_{t-1})|D_t],$$

e

$$V[\theta_t|D_t] = V[E(\theta_t|\lambda_t, D_{t-1})|D_t] + E[V(\theta_t|\lambda_t, D_{t-1})|D_t].$$

Por fim, tem-se a posteriori para  $\theta_t$  parcialmente especificada.

- Posteriori em  $t$

$$\theta_t|D_t \sim [m_t, C_t], \text{ em que } m_t = a_t + \frac{R_t F_t (f_t^* - f_t)}{q_t},$$

$$C_t = R_t - \frac{R_t F_t F_t' R_t (1 - q_t^*/q_t)}{q_t}.$$

É com essa média e variância especificadas para a posteriori de  $\theta_t$  que chegamos, através de métodos de simulação, como aqueles mostrados na seção 1.5, a uma amostra da distribuição a posteriori ( $\theta_t|D_t$ ).

## 3.5 Previsão

Geralmente, não é trivial fazer a passagem da evolução para a previsão quando se trabalha com dados não-normais. Portanto, algumas medidas devem ser tomadas para contornar esse problema. Uma delas consiste em assumir que a distribuição  $\lambda_t|D_{t-1}$  é aproximadamente normal. Outra, bastante razoável, é trabalhar com prioris conjugadas adequadas aos valores especificados para a distribuição a priori de  $\lambda_t$ . Ou seja, assume-se que a priori abaixo seja conjugada:

$$p(\eta_t|D_{t-1}) = c(r_t, s_t) \exp\{r_t \eta_t - s_t a(\eta_t)\}.$$

Os parâmetros  $r_t$  e  $s_t$  são escolhidos de tal forma que atendam a média e a variância dadas, respectivamente, por  $E[g(\eta_t)|D_{t-1}] = f_t$  e  $V[g(\eta_t)|D_{t-1}] = q_t$ .

Sendo assim, a distribuição de previsão um-passo-a-frente possui densidade na forma:

$$p(y_t|D_{t-1}) = \frac{c(r_t, s_t)b(y_t, V_t)}{c(r_t + \phi_t y_t, s_t + \phi_t)}.$$

De modo análogo à distribuição de previsão para um-passo-a-frente, a distribuição de previsão para  $h$ -passos-a-frente é dada por:

$$p(y_{t+h}|D_t) = \frac{c(r_t(h), s_t(h))b(y_{t+h}, V_{t+h})}{c(r_t(h) + \phi_{t+h}y_{t+h}, s_t(h) + \phi_{t+h})}.$$

Portanto, chegamos as distribuições parcialmente especificadas do vetor de parâmetros e de observações, para o horizonte de previsão de  $h$ -passos-a-frente, para  $h = 1, 2, \dots$ , como segue abaixo:

$$\begin{aligned} \theta_{t+h}|D_t &\sim [a_t(h), R_t(h)], \text{ em que } a_t(h) = G_{t+h}a_t(h-1), \\ R_t(h) &= G_{t+h}R_t(h-1)G'_{t+h} + W_{t+h} \end{aligned}$$

e

$$\begin{aligned} \lambda_{t+h}|D_t &\sim [f_t(h), q_t(h)], \text{ em que } f_t(h) = F'_{t+h}a_t(h), \\ q_t(h) &= F'_{t+h}R_t(h)F_{t+h}, \end{aligned}$$

em que  $a_t(0) = m_t$  e  $R_t(0) = C_t$ .

No próximo capítulo, focaremos em um caso particular do conjunto de distribuições que pertencem à família exponencial. Falaremos com mais detalhes da distribuição que usaremos nesse estudo, a distribuição de Poisson.

# Capítulo 4

## Modelo para Dados Poisson

### 4.1 Introdução

Tendo em vista que a família exponencial abrange uma boa quantidade de distribuições, pode-se dizer que este foi um grande avanço na teoria de modelos dinâmicos. A distribuição de Poisson, uma distribuição bastante conhecida e também membro da família exponencial, será tratada nesse capítulo com mais detalhes.

Essa distribuição foi descoberta por Siméon-Denis (1781–1840) e publicada, conjuntamente com a sua teoria da probabilidade, em 1838. Ela expressa a probabilidade de um certo número de ocorrências num dado período tempo, caso estes ocorram com uma taxa média conhecida e caso cada evento seja independente do tempo decorrido desde o último evento. O modelo de Poisson segue uma distribuição de probabilidade discreta.

Seu espaço amostral é composto pelos números inteiros não negativos e sua função de probabilidade é dada por:

$$X \sim POI(\lambda) \Leftrightarrow P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ em que } x = 0, 1, 2, 3, \dots$$

Sua esperança e variância são dadas respectivamente por  $E[X] = \lambda$  e  $V[X] = \lambda$ .

Se fizermos  $X_t$  igual ao número de ocorrências durante o intervalo de tempo de 0 a  $t$ , então definimos  $p_n(t) = P(X_t = n), n = 0, 1, 2, 3, \dots$  como um processo de Poisson.



Assim, como descrito em MEYER (2000), esse processo possui 3 hipóteses:

- H1)** se  $\Delta_t$  for suficientemente pequeno, então  $p_1(\Delta_t) \sim \lambda\Delta_t$ . Ou seja, Se o intervalo for suficientemente pequeno, a probabilidade de obter exatamente uma emissão durante o intervalo é diretamente proporcional ao comprimento do intervalo;
- H2)**  $p_k(\Delta_t) \sim 0, k \geq 2$ . Ou seja, se o intervalo é suficientemente pequeno, a probabilidade de ocorrerem 2 ou mais eventos é desprezível;
- H3)** considerando os intervalos não-sobrepostos,  $(0, t_1)$  e  $(t_1, t_2)$ , para  $t_1 < t_2$ , o número de ocorrências nesses intervalos de tempo constituem variáveis aleatórias independentes.

Portanto, o fato desse modelo ser bastante usado para dados de contagem, o torna uma razoável opção na modelagem proposta aqui, tendo em vista que a variável resposta dos modelos trabalhados é o números de gols das equipes.

Na primeira hipótese é razoável imaginar que quanto menor o intervalo de tempo, menor é a probabilidade de um determinado time fazer um gol nesse intervalo. A segunda também faz sentido, pois em um intervalo suficientemente pequeno de tempo as chances de um time fazer 2 ou mais gols é realmente pequena. Já a terceira hipótese, se considerarmos um intervalo de tempo como sendo uma rodada de um campeonato de futebol, parece razoável pensar que o número de gols de uma equipe é independente do número de gols dessa mesma equipe em qualquer outra rodada do mesmo campeonato.

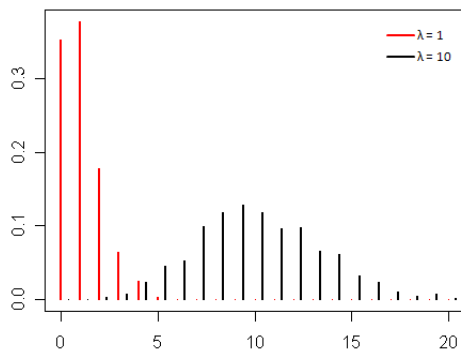


Figura 4.1: Função de probabilidade da distribuição de Poisson para 2 valores distintos de  $\lambda$ .

A figura 4.1 ilustra a função de probabilidade da distribuição de Poisson para dois valores diferentes do parâmetro  $\lambda$ . Portanto, é razoável esperar que se a variável tratada for o número de gols em partidas de futebol, a distribuição dos dados se aproxime da distribuição em vermelho na figura. Pois, sabe-se que valores acima de 5 são pouco prováveis para o número de gols em uma partida, se concentrando fortemente entre os valores 0, 1 e 2.

Conforme dito anteriormente, para uma distribuição pertencer a família exponencial, deve ser possível escrever sua função de probabilidade ou de densidade de probabilidade da seguinte forma:

$$p(y_t|\eta_t, V_t) = \exp\{\phi_t[Y_t(y_t)\eta_t - a(\eta_t)]\}b(y_t, V_t).$$

No caso da distribuição de Poisson isso é possível, e para isso, os valores dos termos são dados por:

$$\phi_t = 1,$$

$$\eta_t = \log\lambda_t,$$

$$a_t = e^{\eta_t},$$

$$Y_t(y_t) = y_t,$$

$$b(y_t, V_t) = \frac{1}{y_t!}.$$

Sua função de ligação é dada pela função logarítmica, por isso, o modelo de Poisson é conhecido na teoria dos Modelos Lineares generalizados como modelo log-linear.

Na seção 1.3, dois testes não-paramétricos realizados e comparações gráficas feitas foram apresentados para reforçar a idéia de que a Poisson é adequada para os objetivos desse estudo. Na próxima seção, descreveremos os modelos usados nesse estudo.

## 4.2 Especificação dos Modelos Utilizados

A motivação para o uso de modelos dinâmicos na modelagem de dados de partidas de futebol veio primeiramente do trabalho RUE e SALVESEN (2000), em que foi proposto

um MLDG para o número de gols das equipes da primeira divisão do campeonato inglês de futebol na temporada 1997-1998. Mais adiante, um modelo semelhante foi proposto em SOUZA e GAMERMAN (2004), desta vez visando os campeonatos brasileiros de 2002 e 2003.

Ao todo, esse estudo conta com 4 modelos: 2 modelos propostos, e outros 2, que foram propostos por RUE e SALVESEN (2000) e usados em SOUZA e GAMERMAN (2004).

Seja  $n$  o número de equipes de um campeonato de futebol,  $T$  o total de rodadas do campeonato e  $y_i^t$  o número de gols da equipe  $i$  na rodada  $t$ , considere a seguinte notação:

$$y^t = (y_1^t, \dots, y_n^t), \text{ para } t = 1, \dots, T.$$

O vetor  $y^t$  compreende toda informação disponível nos dados para a rodada  $t$ . Agora considere o vetor do número de gols, que consiste em toda a informação do campeonato até a rodada atual,  $t$ :

$$y = (y^1, \dots, y^t).$$

O vetor  $y$  corresponde ao vetor  $D_t$  definido nos capítulos 2 e 3, que determina toda informação proveniente dos dados até a rodada  $t$ .

A matriz conhecida  $F_t$  apresentada nos capítulos 2 e 3 para o MLD e o MLDG, respectivamente, é a mesma para os quatro modelos que serão descritos a seguir. A construção dessa matriz será detalhadas mais a frente nessa seção.

### 1. Estático ( $E$ )

O primeiro modelo recebe o nome de estático pelo fato dos parâmetros não variarem no tempo. Ele é semelhante ao MLDG proposto em SOUZA e GAMERMAN (2004), exceto claro pela ausência do índice de variação no tempo ( $t$ ).

O vetor de parâmetros desse modelo é dado por  $\theta = (\alpha, \beta, \gamma)$ , e seus elementos são, respectivamente,  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $\beta = (\beta_1, \dots, \beta_n)$  e  $\gamma = (\gamma_1, \dots, \gamma_n)$ .

Sua especificação é dada por:

$$\begin{cases} \text{eq. de observação: } y_i^t \sim Poi(\lambda_i^t), \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T, \\ \text{eq. de ligação: } \log \lambda_i^t = \alpha_i - \beta_j + \gamma_i. \end{cases} \quad (4.1)$$

$y_i^t$  é o número de gols da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$\alpha_i$  é o fator ataque da equipe  $i$ , para  $i = 1, \dots, n$ ,

$\beta_j$  é o fator defesa da equipe  $j$ , para  $j = 1, \dots, n$  e  $i \neq j$ , e

$\gamma_i$  é o fator campo da equipe  $i$ , para  $i = 1, \dots, n$ .

As priors do modelo  $E$  são dadas por:

$$\alpha_i \sim N(\mu_{\alpha_i}, \xi_{\alpha_i}), \text{ para } i = 1, \dots, n,$$

$$\beta_i \sim N(\mu_{\beta_i}, \xi_{\beta_i}), \text{ para } i = 1, \dots, n,$$

$$\gamma_i \sim N(\mu_{\gamma_i}, \xi_{\gamma_i}), \text{ para } i = 1, \dots, n,$$

em que  $\mu_{\alpha_i}$ ,  $\xi_{\alpha_i}$ ,  $\mu_{\beta_i}$ ,  $\xi_{\beta_i}$ ,  $\mu_{\gamma_i}$  e  $\xi_{\gamma_i}$  são constantes, para  $i = 1, \dots, n$ .

## 2. Dinâmico ( $D$ )

Esse é o mesmo modelo usado em SOUZA e GAMERMAN (2004). Ele é uma generalização do modelo estático, pois a única diferença é que os parâmetros passam a variar no tempo.

Considere o vetor de parâmetros desse modelo, agora variando no tempo:

$$\theta^t = (\alpha^t, \beta^t, \gamma^t), \text{ para } t = 1, \dots, T.$$

Analogamente ao modelo  $E$ , seus elementos são, respectivamente,  $\alpha^t = (\alpha_1^t, \dots, \alpha_n^t)$ ,  $\beta^t = (\beta_1^t, \dots, \beta_n^t)$  e  $\gamma^t = (\gamma_1^t, \dots, \gamma_n^t)$ , para  $t = 1, \dots, T$ .

Considere também o vetor de hiperparâmetros desse modelo, que é formado pelas variâncias da equação de evolução:

$$\psi^t = (\psi_\alpha^t, \psi_\beta^t, \psi_\gamma^t), \text{ para } t = 1, \dots, T,$$

em que  $\psi_\zeta^t = \sigma_\zeta^{2t}$ , para  $\zeta = \alpha, \beta$  e  $\gamma$ .

A especificação do modelo  $D$  é dada por:

$$\left\{ \begin{array}{l} \text{eq. de observação: } y_i^t \sim Poi(\lambda_i^t), \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T, \\ \text{eq. de ligação: } \log \lambda_i^t = \alpha_i^t - \beta_j^t + \gamma_i^t, \\ \text{eq. de evolução: } \theta_t = \theta_{t-1} + \omega_t, \text{ em que } \omega_t \sim [0, W_t]. \end{array} \right. \quad (4.2)$$

$y_i^t$  é o número de gols da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$\alpha_i^t$  é o fator ataque da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$\beta_j^t$  é o fator defesa da equipe  $j$  na rodada  $t$ , para  $j = 1, \dots, n, i \neq j$  e  $t = 1, \dots, T$ ,

$\gamma_i^t$  é o fator campo da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$W_t = \text{diag}(\sigma_\alpha^{2t} \mathbf{1}_n, \sigma_\beta^{2t} \mathbf{1}_n, \sigma_\gamma^{2t} \mathbf{1}_n)$ , para  $t = 1, \dots, T$ ,

em que  $\mathbf{1}_n$  é um vetor de ordem  $n$  com todos os elementos iguais a 1.

A equação do sistema, também conhecida como equação de evolução, pode ser escrita da seguinte maneira:

$$\left\{ \begin{array}{l} \alpha_i^{t+1} = \alpha_i^t + \omega_\alpha^t, \text{ com } \omega_\alpha^t \sim N[0, \sigma_\alpha^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \beta_i^{t+1} = \beta_i^t + \omega_\beta^t, \text{ com } \omega_\beta^t \sim N[0, \sigma_\beta^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \gamma_i^{t+1} = \gamma_i^t + \omega_\gamma^t, \text{ com } \omega_\gamma^t \sim N[0, \sigma_\gamma^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T. \end{array} \right.$$

As priors do modelo  $D$  são dadas por:

$$\alpha_i^1 \sim N(\mu_{\alpha_i}, \xi_{\alpha_i}), \text{ para } i = 1, \dots, n,$$

$$\beta_i^1 \sim N(\mu_{\beta_i}, \xi_{\beta_i}), \text{ para } i = 1, \dots, n,$$

$$\gamma_i^1 \sim N(\mu_{\gamma_i}, \xi_{\gamma_i}), \text{ para } i = 1, \dots, n,$$

em que  $\mu_{\alpha_i}, \xi_{\alpha_i}, \mu_{\beta_i}, \xi_{\beta_i}, \mu_{\gamma_i}$  e  $\xi_{\gamma_i}$  são constantes, para  $i = 1, \dots, n$ ,

$$\sigma_{\alpha}^{2t} \sim \text{Gama Inversa}(a_{\alpha}^t, b_{\alpha}^t), \text{ para } t = 1, \dots, T,$$

$$\sigma_{\beta}^{2t} \sim \text{Gama Inversa}(a_{\beta}^t, b_{\beta}^t), \text{ para } t = 1, \dots, T,$$

$$\sigma_{\gamma}^{2t} \sim \text{Gama Inversa}(a_{\gamma}^t, b_{\gamma}^t), \text{ para } t = 1, \dots, T,$$

em que  $a_{\alpha}^t, b_{\alpha}^t, a_{\beta}^t, b_{\beta}^t, a_{\gamma}^t$  e  $b_{\gamma}^t$  são constantes, para  $t = 1, \dots, T$ .

### 3. Dinâmico com 1 coeficiente auto-regressivo de evolução ( $D_1$ )

Esse é o primeiro dos 2 modelos propostos por esse estudo. O que diferencia esses modelos do modelo  $D$ , é o fato de trabalharem com coeficientes auto-regressivos na equação de evolução. O modelo  $D_1$  recebe esse nome, pois possui um coeficiente auto-regressivo.

Considere o vetor de parâmetros desse modelo, similar ao modelo  $D$ :

$$\theta^t = (\alpha^t, \beta^t, \gamma^t), \text{ para } t = 1, \dots, T.$$

Seus elementos são, respectivamente,  $\alpha^t = (\alpha_1^t, \dots, \alpha_n^t)$ ,  $\beta^t = (\beta_1^t, \dots, \beta_n^t)$  e  $\gamma^t = (\gamma_1^t, \dots, \gamma_n^t)$ , para  $t = 1, \dots, T$ .

O vetor de hiperparâmetros desse modelo é formado pelo coeficiente auto-regressivo de evolução e pelas variâncias da equação de evolução:

$$\psi^t = (\psi_{\alpha}^t, \psi_{\beta}^t, \psi_{\gamma}^t), \text{ para } t = 1, \dots, T,$$

em que  $\psi_{\zeta}^t = (\rho^t, \sigma_{\zeta}^{2t})$ , para  $\zeta = \alpha, \beta$  e  $\gamma$ .

A especificação do modelo  $D_1$  é dada por:

$$\left\{ \begin{array}{l} \text{eq. de observação: } y_i^t \sim \text{Poi}(\lambda_i^t), \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T, \\ \text{eq. de ligação: } \log \lambda_i^t = \alpha_i^t - \beta_j^t + \gamma_i^t, \\ \text{eq. de evolução: } \theta_t = \rho^t \theta_{t-1} + \omega_t, \text{ em que } \omega_t \sim [0, W_t]. \end{array} \right. \quad (4.3)$$

$y_i^t$  é o número de gols da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,  
 $\alpha_i^t$  é o fator ataque da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,  
 $\beta_j^t$  é o fator defesa da equipe  $j$  na rodada  $t$ , para  $j = 1, \dots, n$ ,  $i \neq j$  e  
 $t = 1, \dots, T$ ,

$\gamma_i^t$  é o fator campo da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$\varrho^t = \rho^t \mathbf{1}_{3n}$ , para  $t = 1, \dots, T$ ,

$W_t = \text{diag}(\sigma_\alpha^{2t} \mathbf{1}_n, \sigma_\beta^{2t} \mathbf{1}_n, \sigma_\gamma^{2t} \mathbf{1}_n)$ , para  $t = 1, \dots, T$ ,

em que  $\mathbf{1}_n$  é um vetor de ordem  $n$  com todos os elementos iguais a 1.

A equação do sistema, também conhecida como equação de evolução, pode ser escrita da seguinte maneira:

$$\begin{cases} \alpha_i^{t+1} = \rho^t \alpha_i^t + \omega_\alpha^t, & \text{com } \omega_\alpha^t \sim N[0, \sigma_\alpha^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \beta_i^{t+1} = \rho^t \beta_i^t + \omega_\beta^t, & \text{com } \omega_\beta^t \sim N[0, \sigma_\beta^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \gamma_i^{t+1} = \rho^t \gamma_i^t + \omega_\gamma^t, & \text{com } \omega_\gamma^t \sim N[0, \sigma_\gamma^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T. \end{cases}$$

As priors do modelo  $D_1$  são dadas por:

$\alpha_i^1 \sim N(\mu_{\alpha_i}, \xi_{\alpha_i})$ , para  $i = 1, \dots, n$ ,

$\beta_i^1 \sim N(\mu_{\beta_i}, \xi_{\beta_i})$ , para  $i = 1, \dots, n$ ,

$\gamma_i^1 \sim N(\mu_{\gamma_i}, \xi_{\gamma_i})$ , para  $i = 1, \dots, n$ ,

em que  $\mu_{\alpha_i}$ ,  $\xi_{\alpha_i}$ ,  $\mu_{\beta_i}$ ,  $\xi_{\beta_i}$ ,  $\mu_{\gamma_i}$  e  $\xi_{\gamma_i}$  são constantes, para  $i = 1, \dots, n$ ,

$\sigma_\alpha^{2t} \sim \text{Gama Inversa}(a_{\alpha^t}, b_{\alpha^t})$ , para  $t = 1, \dots, T$ ,

$\sigma_\beta^{2t} \sim \text{Gama Inversa}(a_{\beta^t}, b_{\beta^t})$ , para  $t = 1, \dots, T$ ,

$\sigma_\gamma^{2t} \sim \text{Gama Inversa}(a_{\gamma^t}, b_{\gamma^t})$ , para  $t = 1, \dots, T$ ,

em que  $a_\alpha^t$ ,  $b_\alpha^t$ ,  $a_\beta^t$ ,  $b_\beta^t$ ,  $a_\gamma^t$  e  $b_\gamma^t$  são constantes, para  $t = 1, \dots, T$  e,

$\rho^t \sim U[0, 1]$ , para  $t = 1, \dots, T$ .

#### 4. Dinâmico com 3 coeficientes auto-regressivos de evolução ( $D_3$ )

Esse é o modelo mais geral dentre os quatro. Ele não só varia no tempo como também trabalha com 3 coeficientes auto-regressivos de evolução, um para cada tipo de fator.

Considere o vetor de parâmetros desse modelo:

$$\theta^t = (\alpha^t, \beta^t, \gamma^t), \text{ para } t = 1, \dots, T.$$

Seus elementos são, respectivamente,  $\alpha^t = (\alpha_1^t, \dots, \alpha_n^t)$ ,  $\beta^t = (\beta_1^t, \dots, \beta_n^t)$  e  $\gamma^t = (\gamma_1^t, \dots, \gamma_n^t)$ , para  $t = 1, \dots, T$ .

O vetor de hiperparâmetros desse modelo é formado pelos coeficientes auto-regressivos de evolução e pelas variâncias da equação de evolução:

$$\psi^t = (\psi_\alpha^t, \psi_\beta^t, \psi_\gamma^t), \text{ para } t = 1, \dots, T,$$

em que  $\psi_\zeta^t = (\rho_\zeta^t, \sigma_\zeta^{2t})$ , para  $\zeta = \alpha, \beta$  e  $\gamma$ .

A especificação do modelo  $D_3$  é dada por:

$$\left\{ \begin{array}{l} \text{eq. de observação: } y_i^t \sim Poi(\lambda_i^t), \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T, \\ \text{eq. de ligação: } \log \lambda_i^t = \alpha_i^t - \beta_j^t + \gamma_i^t, \\ \text{eq. de evolução: } \theta_t = \varrho^t \theta_{t-1} + \omega_t, \text{ em que } \omega_t \sim [0, W_t]. \end{array} \right. \quad (4.4)$$

$y_i^t$  é o número de gols da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,  
 $\alpha_i^t$  é o fator ataque da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,  
 $\beta_j^t$  é o fator defesa da equipe  $j$  na rodada  $t$ , para  $j = 1, \dots, n$ ,  $i \neq j$  e  $t = 1, \dots, T$ ,

$\gamma_i^t$  é o fator campo da equipe  $i$  na rodada  $t$ , para  $i = 1, \dots, n$  e  $t = 1, \dots, T$ ,

$\varrho^t = (\rho_\alpha^t \mathbf{1}_n, \rho_\beta^t \mathbf{1}_n, \rho_\gamma^t \mathbf{1}_n)$ , para  $t = 1, \dots, T$ ,

$W_t = \text{diag}(\sigma_\alpha^{2t} \mathbf{1}_n, \sigma_\beta^{2t} \mathbf{1}_n, \sigma_\gamma^{2t} \mathbf{1}_n)$ , para  $t = 1, \dots, T$ ,

em que  $\mathbf{1}_n$  é um vetor de ordem  $n$  com todos os elementos iguais a 1.

A equação do sistema, também conhecida como equação de evolução, pode ser escrita da seguinte maneira:



$$\left\{ \begin{array}{l} \alpha_i^{t+1} = \rho_\alpha^t \alpha_i^t + \omega_\alpha^t, \text{ com } \omega_\alpha^t \sim N[0, \sigma_\alpha^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \beta_i^{t+1} = \rho_\beta^t \beta_i^t + \omega_\beta^t, \text{ com } \omega_\beta^t \sim N[0, \sigma_\beta^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T; \\ \gamma_i^{t+1} = \rho_\gamma^t \gamma_i^t + \omega_\gamma^t, \text{ com } \omega_\gamma^t \sim N[0, \sigma_\gamma^{2t}], \text{ para } i = 1, \dots, n \text{ e } t = 1, \dots, T. \end{array} \right.$$

As priors do modelo  $D_3$  são dadas por:

$$\begin{aligned} \alpha_i^1 &\sim N(\mu_{\alpha_i}, \xi_{\alpha_i}), \text{ para } i = 1, \dots, n, \\ \beta_i^1 &\sim N(\mu_{\beta_i}, \xi_{\beta_i}), \text{ para } i = 1, \dots, n, \\ \gamma_i^1 &\sim N(\mu_{\gamma_i}, \xi_{\gamma_i}), \text{ para } i = 1, \dots, n, \\ \text{em que } \mu_{\alpha_i}, \xi_{\alpha_i}, \mu_{\beta_i}, \xi_{\beta_i}, \mu_{\gamma_i} \text{ e } \xi_{\gamma_i} &\text{ são constantes, para } i = 1, \dots, n, \\ \sigma_\alpha^{2t} &\sim \text{Gama Inversa}(a_\alpha^t, b_\alpha^t), \text{ para } t = 1, \dots, T, \\ \sigma_\beta^{2t} &\sim \text{Gama Inversa}(a_\beta^t, b_\beta^t), \text{ para } t = 1, \dots, T, \\ \sigma_\gamma^{2t} &\sim \text{Gama Inversa}(a_\gamma^t, b_\gamma^t), \text{ para } t = 1, \dots, T, \\ \text{em que } a_\alpha^t, b_\alpha^t, a_\beta^t, b_\beta^t, a_\gamma^t \text{ e } b_\gamma^t &\text{ são constantes, para } t = 1, \dots, T \text{ e,} \\ \rho_\zeta^t &\sim U[0, 1], \text{ para } \zeta = \alpha, \beta \text{ e } \gamma \text{ e } t = 1, \dots, T. \end{aligned}$$

O modelo  $E$  é o mais simples, pois trabalha apenas com 3 parâmetros, que são os fatores ataque, defesa e campo. No modelo  $D$ , a descrição é semelhante ao modelo  $E$ , a única diferença é que os parâmetros passam a variar no tempo, e portanto, surge a equação de evolução. Os modelos  $D_1$  e  $D_3$  são idênticos ao modelo  $D$ , exceto pela presença de, respectivamente, 1 e 3 coeficientes auto-regressivos na equação de evolução.

Os três modelos dinâmicos, foram descritos de forma bem geral, com os hiperparâmetros variando no tempo. Nesse estudo optou-se por trabalhar com hiperparâmetros que não variam no tempo, para simplificar a modelagem.

As distribuições a priori dos parâmetros de um modelo descrevem nossa incerteza acerca desses parâmetros. Nesse estudo optou-se pelo uso de distribuições a priori não-informativas.

Os parâmetros definidos para os modelos podem ser trabalhados em sua escala original, porém para facilitar a etapa computacional do estudo, optou-se por fazer uma reparametrização nos modelos. Esse processo tem o objetivo de manter a primeira coluna

da matriz de planejamento  $(F_t, t = 1, \dots, T)$  preenchida com valores iguais a 1, e é dado por:

$$\log \lambda_i^t = \mu^t + \alpha_i^{*t} - \beta_j^{*t} + \gamma_i^t,$$

em que  $\mu^t = \alpha_1^t - \beta_1^t, \forall t = 1, \dots, T$ ,

$$\alpha_i^{*t} = \alpha_i^t - \alpha_1^t, \forall i = 1, \dots, n \text{ e } \forall t = 1, \dots, T, \text{ e}$$

$$\beta_i^{*t} = \beta_i^t - \beta_1^t, \forall i = 1, \dots, n \text{ e } \forall t = 1, \dots, T.$$

Logo, tem-se um total de  $(3n - 1)$  parâmetros, sendo: 1 nível,  $(n - 1)$  fatores ataque,  $(n - 1)$  fatores defesa e  $n$  fatores campo. Os valores ajustados não serão os valores reais dos parâmetros, e sim uma comparação entre esses e os fatores da equipe 1, equipe escolhida para ser o nível. Para recuperar a informação da equipe 1 e trazer todos os parâmetros para a mesma escala faremos como é mostrado abaixo para o fator ataque:

$$\begin{aligned} \alpha_i^{**} &= \alpha_i - \bar{\alpha}; \\ &= \alpha_i - \alpha_1 - \bar{\alpha} + \alpha_1; \\ &= \alpha_i^* - (\bar{\alpha} - \alpha_1); \\ &= \alpha_i^* - \left( \frac{\sum_{j=1}^n \alpha_j - n\alpha_1}{n} \right); \\ &= \alpha_i^* - \frac{\sum_{j=1}^n (\alpha_j - \alpha_1)}{n}; \\ &= \alpha_i^* - \frac{\sum_{j=1}^n \alpha_j^*}{n}. \end{aligned}$$

Para a equipe 1, temos:

$$\begin{aligned}
\alpha_1^{**} &= \alpha_1^* - \frac{\sum_{j=1}^n \alpha_j^*}{n}; \\
&= \alpha_1 - \alpha_1 - \frac{\sum_{j=1}^n \alpha_j^*}{n}; \\
&= -\frac{\sum_{j=1}^n \alpha_j^*}{n}.
\end{aligned}$$

O mesmo raciocínio se aplica ao fator defesa. Sendo assim, os fatores ataque e defesa passam a ser analisados em comparação a média global dos fatores, ou seja, se  $\alpha_i^{**}$  tem valor positivo, então a equipe  $i$ , obteve um fator ataque acima da média.

A seguir descrevemos a forma de construção da matriz  $F_t$ , para  $t = 1, \dots, T$ , considerando a reparametrização feita.

$$F'_t = \left[ \begin{array}{c|ccc|ccc|ccc} 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{array} \right]_{n \times (3n-1)},$$

em que  $f_{ij} = 1$ , se  $j = 1$ ,

$f_{ij} = 1$ ,  $\forall j = 2, \dots, n$ , se a equipe  $i$  for mandante,

$f_{ij} = -1$ ,  $\forall j = n + 1, \dots, 2n - 1$ , se a equipe  $i$  for visitante,

$f_{ij} = 1$ ,  $\forall j = 2n, \dots, 3n - 1$ , se a equipe  $i$  jogar em seu campo, e

$f_{ij} = 0$ ,  $\forall j = 2, \dots, 3n - 1$ , em todos os outros casos.

O número de linhas e colunas da matriz  $F'_t$  são dados respectivamente por  $n$  e  $3n - 1$ . O número de colunas na matriz é dividido em 4 partes, são elas: uma para o nível,  $n - 1$  para o fator ataque,  $n - 1$  para o fator defesa e  $n$  para o fator campo.

### 4.3 Processo de Inferência

Nas próximas seções, descreveremos brevemente a inferência e a previsão para o modelo  $D_3$ . Para os demais modelos, o raciocínio é análogo, pois tratam-se de casos particulares do modelo  $D_3$ .

Seja  $s$  a rodada atual, com portanto temos o vetor  $y = (y^1, \dots, y^s)$  contendo toda informação dos dados que dispomos e  $\psi$  o vetor de hiperparâmetros. Assim, podemos fazer inferência no modelo  $D_3$  usando a seguinte fórmula:

$$p(\theta^s, \psi | y) \propto L(\theta^s, \psi; y) p(\theta^s, \psi).$$

A verossimilhança, primeiro termo do lado direito da equação, pode ser obtida fazendo:

$$\begin{aligned} L(\theta^s, \psi; y) &= p(y | \theta^s, \psi); \\ &= p(y | \theta^s); \\ &= \prod_{t=1}^s p(y^t | \theta^t); \\ &= \prod_{t=1}^s \left[ \prod_{i=1}^n p(y_i^t | \theta^t) \right]. \end{aligned}$$

O segundo termo é dado por:

$$\begin{aligned} p(\theta^s, \psi) &= p(\theta^s | \psi) p(\psi); \\ &= \prod_{t=1}^s \left[ p(\theta^t | \theta^{t-1}, \psi) \right] p(\theta^1). \end{aligned}$$

Vale ressaltar que a distribuição  $y_i^t | \theta^t$  é dada pela equação de observações e a distribuição  $\theta^t | \theta^{t-1}, \psi$  pela equação de evolução. Chegamos a distribuição a posteriori de

$\theta^s$  fazendo:

$$p(\theta^s|y) = \int p(\theta^s, \psi|y) d\psi.$$

Conforme dito anteriormente, chegamos a uma amostra da distribuição aproximada da distribuição a posteriori acima, usando métodos numéricos.

## 4.4 Previsão

Considere  $h$  o horizonte de previsão, para  $h = 1, 2, 3, \dots$ . A distribuição preditiva para o modelo  $D_3$ , a partir da rodada  $s$ , segue o cálculo mostrado na equação 1.1, como mostrado abaixo:

$$\begin{aligned} p(y^{s+h}|y) &= \int \int \int p(y^{s+h}, \theta^{s+h}, \theta^s, \psi|y) d\theta^{s+h} d\theta^s d\psi; \\ &= \int \int \int p(y^{s+h}|\theta^{s+h}) p(\theta^{s+h}|\theta^s, \psi) p(\theta^s, \psi|y) d\theta^{s+h} d\theta^s d\psi; \\ &= \int \int \int p(y^{s+h}|\theta^{s+h}) p(\theta^{s+h}|\theta^s, \psi) p(\theta^s|y) p(\psi|y) d\theta^{s+h} d\theta^s d\psi. \end{aligned}$$

Como não é possível resolver a integral acima analiticamente, usa-se um dos métodos de MCMC citados no primeiro capítulo para amostrar da distribuição preditiva. Para obter uma amostra da preditiva, obtem-se uma amostra da distribuição conjunta dentro da integral. Essa amostra pode ser obtida através dos seguintes passos:

- i)  $\theta^s \sim p(\theta^s|y)$  e  $\psi \sim p(\psi|y)$ ;
- ii)  $\theta^{s+h} \sim p(\theta^{s+h}|\theta^s, \psi)$ ;
- iii)  $y^{s+h} \sim Poi(\lambda^{s+h})$ , em que  $\lambda^{s+h} = F'_{s+h} \theta^{s+h}$ .

Feita a previsão para uma determinada rodada, calculamos o valor da verossimilhança para que seja possível comparar as previsões dos modelos por rodada. Esse cálculo é dado por:

$$\begin{aligned} p(y^{s+h}|\theta^{s+h}) &= L(\theta^{s+h}; y^{s+h}); \\ &= \prod_{i=1}^n \frac{e^{-\lambda_i^{s+h}} (\lambda_i^{s+h})^{y_i^{s+h}}}{y_i^{s+h}!}, \end{aligned} \tag{4.5}$$

em que  $s$  representa a rodada atual,

$h = 1, 2, 3, \dots$  indica o horizonte de previsão,

$i = 1, \dots, n$  representa a equipe, e

$y_i^{s+h}$  é o número de gols da equipe  $i$  na rodada  $s + h$ .

Logo, quanto maior o valor da verossimilhança, melhor é o desempenho do modelo na previsão.

# Capítulo 5

## Aplicação

### 5.1 Introdução

Como foi comentado nos capítulos anteriores, a escolha dos parâmetros ( $\theta$ ) é de fundamental importância para o desenvolvimento do estudo. Ao trabalhar com dados de campeonatos de futebol, cada autor tem sua interpretação pessoal de quais variáveis podem ou não influenciar o resultado de uma partida. Porém, um deles é unanimidade e está sempre presente nas análises, trata-se do mando de campo. A tabela a seguir ilustra a importância de se jogar em casa, tomando por base 4 anos consecutivos do campeonato brasileiro.

Tabela 5.1: Resultados obtidos pelas equipes mandantes (%), por ano.

Resultado	Ano			
	2003	2004	2005	2006
Vitórias	54,0%	52,9%	51,0%	50,0%
Empates	25,6%	25,0%	21,6%	25,4%
Derrotas	20,4%	22,1%	27,4%	24,6%

Portanto, é bastante razoável incluir a variável mando de campo em qualquer que seja o modelo para dados provenientes de partidas de futebol.

Como exemplo de algumas outras variáveis que poderiam compor um modelo como esse, podemos citar a mudança de técnico, a distância que a equipe tem que viajar para

jogar e a quantidade de jogadores machucados. Nesse estudo, conforme visto no capítulo anterior, optou-se pelo uso de 3 fatores nos modelos, os fatores ataque, defesa e campo.

Na seção seguinte, os resultados dos modelos propostos são apresentados e analisados, principalmente para o modelo  $D_3$ , que é o mais geral. Entretanto, vale ressaltar que os modelos  $E$ ,  $D$  e  $D_1$  passaram pelo mesmo processo de ajuste e também foram analisados.

## 5.2 Resultados

Nesse estudo, trabalharemos com dados do campeonato brasileiro de 2006, o qual vem sendo disputado nos mesmos moldes dos torneios europeus desde 2003, usando o sistema de pontos corridos. Esse sistema contempla como campeão o clube que, ao final da última rodada do torneio, tiver o maior número de pontos.

No modelo  $D_3$ , assim como nos outros, utilizou-se o software WINBUGS (1996) na estimação e previsão. As demais operações realizadas nos dados, como gráficos, tabelas e análises foram feitas no software R (2008).

No WINBUGS, primeiro é preciso definir as prioris para os parâmetros e hiperparâmetros,  $\theta^1$  e  $\psi$ . Vale ressaltar que o WINBUGS trabalha com as precisões ( $\tau_\alpha$ ,  $\tau_\beta$  e  $\tau_\gamma$ ) e não com as variâncias ( $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  e  $\sigma_\gamma^2$ ). Feito isso, deve-se definir o modelo de acordo com as equações apresentadas na seção 4.2, para  $t = 1, \dots, s$ , em que  $s$  é a rodada atual.

$$y_i^t \sim Poi(\lambda_i^t);$$

$$\log \lambda_i^t = \alpha_i^t - \beta_j^t + \gamma_i^t.$$

$$\begin{cases} \alpha_i^{t+1} = \rho_\alpha \alpha_i^t + \omega_\alpha^t, \text{ com } \omega_\alpha^t \sim N[0, \tau_\alpha^{-1}]; \\ \beta_i^{t+1} = \rho_\beta \beta_i^t + \omega_\beta^t, \text{ com } \omega_\beta^t \sim N[0, \tau_\beta^{-1}]; \\ \gamma_i^{t+1} = \rho_\gamma \gamma_i^t + \omega_\gamma^t, \text{ com } \omega_\gamma^t \sim N[0, \tau_\gamma^{-1}]. \end{cases}$$

Terminada essa etapa, inicia-se o processo de simulação conforme descrito abaixo:

1. utiliza-se o chamado *Burn-in*, no qual um número inicial de iterações é desprezado



e diz-se que a cadeia é “aquecida” antes de serem feitas as iterações que serão analisadas;

2. faz-se um número razoável de iterações de modo que seja alcançada a convergência nas distribuições a posteriori amostradas. Essa verificação é feita por uma comparação visual das duas cadeias amostradas em um gráfico de linhas;
3. atingida a convergência, tem-se uma única amostra para cada posteriori formada pela união das duas cadeias.

Usando os dados do campeonato brasileiro de 2006, consideramos a rodada atual como sendo a de número 35, lembrando que ele teve um total de 38 rodadas. As amostras tiradas do WINBUGS já estão atualizadas até a rodada 35, ou seja, temos uma amostra da distribuição a posteriori dada por  $\theta^{35}|D_{35}$ .

Com essa informação passamos pela evolução para chegar a distribuição a priori para a rodada 36 amostrando de:

$$\theta^{36}|D_{35} \sim N(\theta^{35}, W).$$

Vale lembrar que a matriz de covariâncias da distribuição dos termos aleatórios da equação de evolução é dada por  $W = \text{diag}(\sigma_\alpha^2 1_n, \sigma_\beta^2 1_n, \sigma_\gamma^2 1_n)$ . Portanto, para amostrar da priori acima, será preciso obter amostras das distribuições dos hiperparâmetros  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  e  $\sigma_\gamma^2$ . Essas amostras são obtidas fazendo  $\sigma_\zeta^2 = \frac{1}{\tau_\zeta}$ , para  $\zeta = \alpha, \beta$  e  $\gamma$ .

Feita a amostragem dessa distribuição a priori, amostramos da distribuição preditiva:

$$y^{36}|D_{35} \sim \text{Poi}(\lambda^{36}),$$

em que  $\log \lambda^{36} = \alpha^{36} - \beta^{36} + \gamma^{36}$ .

O mesmo raciocínio é aplicado as rodadas seguintes, 37 e 38. O algoritmo utilizado no software WINBUGS está disponível no Apêndice B.

As priors escolhidas para os modelos foram:

$$\alpha_i^1 \sim N(0, (0,005)^{-1}), \text{ para } i = 1, \dots, n;$$

$$\beta_i^1 \sim N(0, (0,005)^{-1}), \text{ para } i = 1, \dots, n;$$

$$\gamma_i^1 \sim N(0, (0,005)^{-1}), \text{ para } i = 1, \dots, n;$$

$$\tau_\alpha \sim \text{Gama}(0,04, 0,15);$$

$$\tau_\beta \sim \text{Gama}(0,04, 0,15);$$

$$\tau_\gamma \sim \text{Gama}(0,04, 0,15);$$

$$\rho_\zeta \sim U[0,1], \text{ para } \zeta = \alpha, \beta \text{ e } \gamma.$$

Ao longo dessa seção alguns gráficos e tabelas serão apresentados, e para evitar a redundância nos gráficos comparativos mostrados por equipe, elegeu-se apenas três equipes. São elas, o São Paulo, que viria a ser o campeão daquele ano, o Flamengo, que terminou o campeonato no meio da tabela e o Santa Cruz que foi o último colocado na ocasião. Vale comentar que todos os gráficos mostrados aqui foram feitos de forma análoga para os times restantes do campeonato.

A figura abaixo mostra a tabela final do Campeonato Brasileiro de 2006.

Pos.	Clube	PG	J	V	GP	GC	SG
1º	SÃO PAULO/SP	78	38	22	66	32	34
2º	INTERNACIONAL/RS	69	38	20	52	36	16
3º	GRÊMIO/RS	67	38	20	64	45	19
4º	SANTOS/SP	64	38	18	58	36	22
5º	PARANÁ/PR	60	38	18	56	49	7
6º	VASCO/RJ	59	38	15	57	50	7
7º	FIGUEIRENSE/SC	57	38	15	52	44	8
8º	GOIÁS/GO	55	38	15	63	49	14
9º	CORINTHIANS/SP	53	38	15	41	46	-5
10º	CRUZEIRO/MG	53	38	14	52	45	7
11º	FLAMENGO/RJ	52	38	15	44	48	-4
12º	BOTAFOGO/RJ	51	38	13	52	50	2
13º	ATLÉTICO/PR	48	38	13	61	62	-1
14º	JUVENTUDE/RS	47	38	13	44	54	-10
15º	FLUMINENSE/RJ	45	38	11	48	58	-10
16º	PALMEIRAS/SP	44	38	12	58	70	-12
17º	PONTE PRETA/SP	39	38	10	45	65	-20
18º	FORTALEZA/CE	38	38	8	39	62	-23
19º	SÃO CAETANO/SP	36	38	9	37	53	-16
20º	SANTA CRUZ/PE	28	38	7	41	76	-35

Figura 5.1: Classificação final do Campeonato Brasileiro de 2006. Legenda: PG - Pontos Ganhos; J - Jogos; V - Vitórias; GP - Gols pró; GC - Gols Contra; SG - Saldo de Gols. Fonte: <http://www.cbf.com.br/>

Nos modelos, as 20 equipes participantes do torneio foram ordenadas em uma seqüência numérica de 1 a 20 para facilitar a nomenclatura dos fatores no momento

de escrever as linhas de programação nos programas de ajuste dos modelos. A tabela a seguir descreve essa ordenação:

Tabela 5.1: Ordenação das equipes em uma seqüência numérica.

Equipe	Sigla	Índice	Equipe	Sigla	Índice
Botafogo	BOT	1	Internacional	INT	11
Atlético-PR	CAP	2	Juventude	JUV	12
Corinthians	COR	3	Palmeiras	PAL	13
Cruzeiro	CRU	4	Paraná	PAR	14
Figueirense	FIG	5	Ponte Preta	PON	15
Flamengo	FLA	6	Santos	SAN	16
Fluminense	FLU	7	São Caetano	SCA	17
Fortaleza	FOR	8	São Paulo	SPA	18
Goiás	GOI	9	Santa Cruz	STA	19
Grêmio	GRE	10	Vasco	VAS	20

Visando padronizar a análise dos modelos, após verificada a convergência dos parâmetros, analisou-se o conjunto das 1000 últimas iterações do modelo. Isso foi feito para todos os parâmetros dos 4 modelos, e ao final foi tomada a união das duas cadeias. Portanto, as amostras de todas as distribuições trabalhadas têm tamanho 2000.

Os gráficos de linhas que ilustram o comportamento dos 3 fatores (ataque, defesa e campo) para as equipes na rodada de número 35 do modelo  $D_3$ , bem como dos parâmetros  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$ ,  $\sigma_\gamma^2$ ,  $\rho_\alpha$ ,  $\rho_\beta$  e  $\rho_\gamma$  estão no Apêndice A.

A função desse tipo de gráfico é verificar a convergência dos parâmetros ao longo das 1000 iterações pelo comportamento das duas cadeias. Observando-se os gráficos, comprova-se a convergência dos parâmetros.

Tabela 5.2: Resumo do ajuste das variâncias da equação de evolução do modelo  $D_3$

Parâmetro	Média	Erro Padrão	$Q_{2,5\%}$	Mediana	$Q_{97,5\%}$
$\sigma_\alpha^2$	0,128	0,028	0,101	0,120	0,204
$\sigma_\beta^2$	0,128	0,028	0,101	0,119	0,203
$\sigma_\gamma^2$	0,124	0,026	0,101	0,116	0,195

Observando a tabela 5.2, notamos que as variâncias dos termos aleatórios da equação de evolução assumiram valores próximos a 0,12. O que é bom, pois tratam-se de valores que não são muito baixos e nem muito altos em relação a média, que é 0. Valores muito baixos anulariam a parte aleatória da equação de evolução, e por outro lado, valores altos fariam com que essa equação se torna-se totalmente aleatória.

A figura abaixo também apresenta as distribuições das variâncias, porém em outra visão, utilizando histogramas.

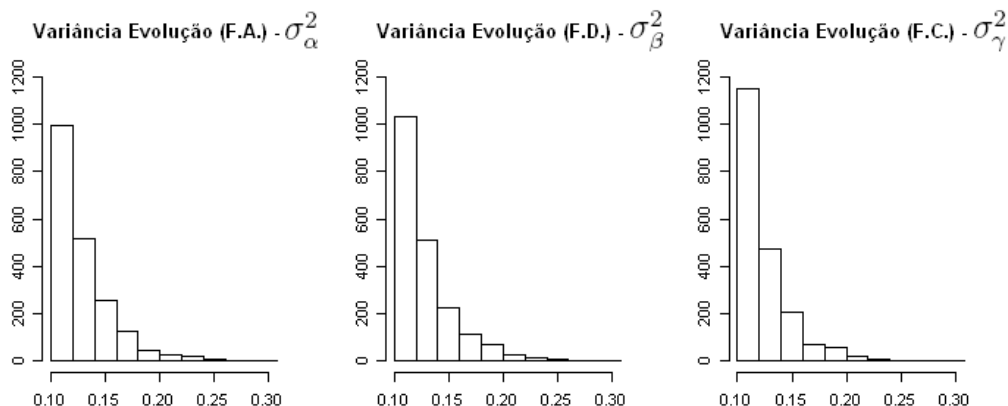


Figura 5.2: Histogramas das variâncias da equação de evolução do modelo  $D_3$  ( $\rho_\alpha$ ,  $\rho_\beta$  e  $\rho_\gamma$ ).

Os histogramas mostram uma visão mais clara de que ocorreu realmente o esperado no comportamento das variâncias.

Tabela 5.3: Resumo do ajuste dos coeficientes auto-regressivos de evolução do modelo  $D_3$

Coeficiente	Média	Erro Padrão	$Q_{2,5\%}$	Mediana	$Q_{97,5\%}$
$\rho_\alpha$	0,762	0,146	0,445	0,781	0,979
$\rho_\beta$	0,733	0,157	0,396	0,747	0,979
$\rho_\gamma$	0,697	0,155	0,361	0,705	0,967

Os valores relativamente altos dos 3 coeficientes, mostrados na tabela 5.3, em torno de 0,8, indicam um ajuste em relação ao modelo dinâmico (D), no qual, esses valores são iguais a 1. A magnitude desses valores, de certa forma, era esperada, pois valores muito baixos resultariam na completa aleatoriedade da equação de evolução e valores muito próximos a 1 ratificariam o uso do modelo dinâmico (D) como o mais adequado.

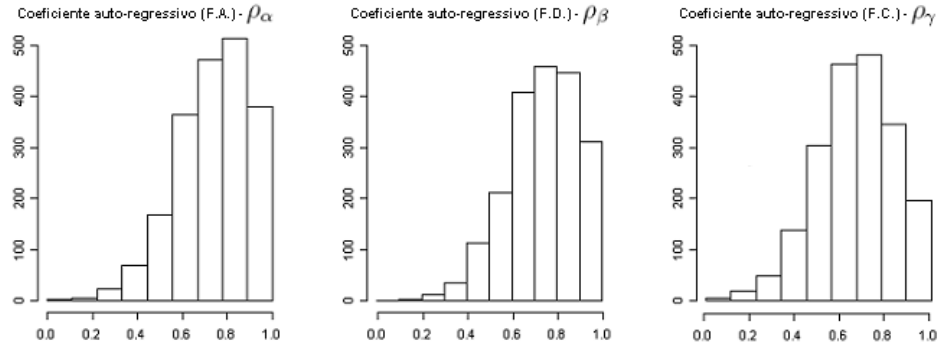


Figura 5.3: Histogramas dos coeficientes auto-regressivos de evolução do modelo  $D_3$  ( $\rho_\alpha$ ,  $\rho_\beta$  e  $\rho_\gamma$ ).

A figura 5.3 evidencia bem o comportamento dos valores ajustados para os 3 coeficientes auto-regressivos de evolução do modelo  $D_3$ , valores em torno de 0,8.

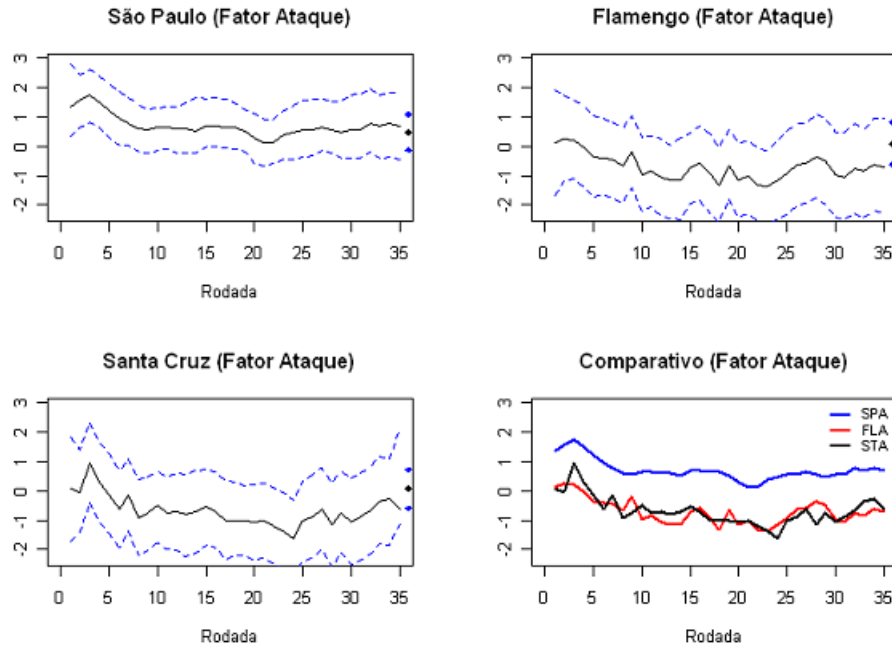


Figura 5.4: Gráficos de linhas dos fatores ataque de 3 equipes selecionadas. As linhas cheias representam a mediana e as tracejadas os quantis 97,5% e 2,5%. Os pontos azuis e pretos cheios, para efeito de comparação, indicam a mesma informação para o modelo  $E$ .

O uso de um modelo dinâmico quando se trabalha com esse tipo de dado é bem recomendável em comparação ao modelo estático. Uma característica do modelo dinâmico que evidencia bem esse ponto de vista, é o fato de podermos analisar o desempenho dos fatores das equipes ao longo das rodadas. As figuras 5.4, 5.5 e 5.6

apresentam o comportamento dos fatores ataque, defesa e campo, respectivamente, para as equipes do São Paulo, Flamengo e Santa Cruz.

No fator ataque fica claro ver que a equipe do São Paulo se manteve relativamente sempre acima da média de todas as equipes, ou seja, valores acima de zero. Esse comportamento era o esperado para a equipe campeã do torneio, e além disso, vale ressaltar que o São Paulo terminou o campeonato com o melhor ataque (66 gols feitos). Santa Cruz e Flamengo alternaram bastante ao longo das 35 rodadas, mas se mantiveram sempre em torno do mesmo nível. Ao final do torneio as duas equipes totalizaram quase o mesmo número de gols feitos, 41 para o Santa Cruz e 44 para o Flamengo. Valores esses abaixo da média que foi de 52 gols feitos por equipe.

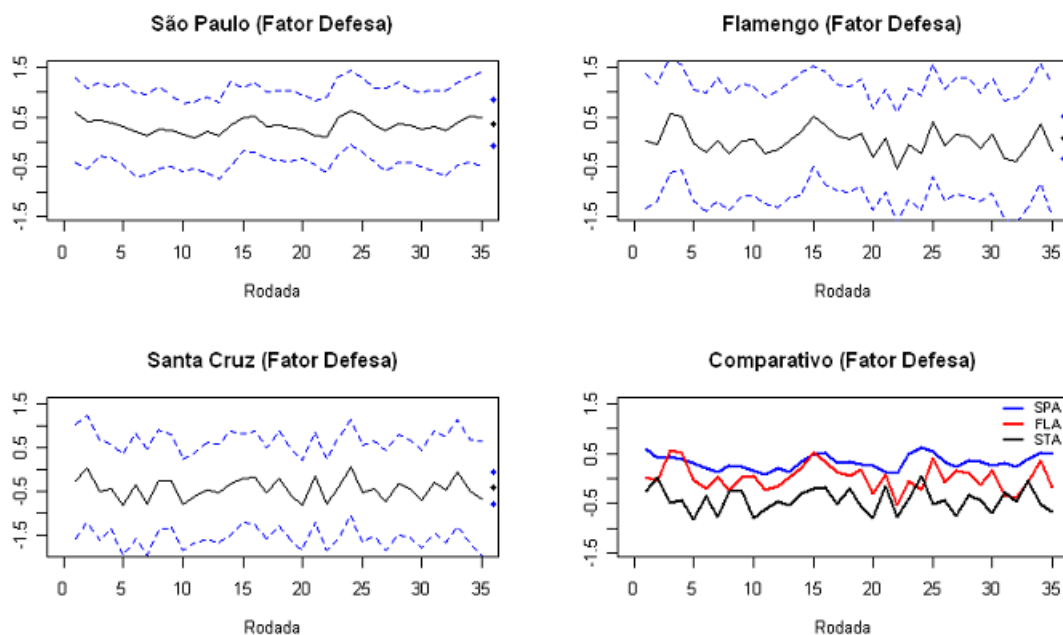


Figura 5.5: Gráfico de linhas dos fatores defesa de 3 equipes selecionadas. As linhas cheias representam a mediana e as tracejadas os quantis 97,5% e 2,5%. Os pontos azuis e pretos cheios, para efeito de comparação, indicam a mesma informação para o modelo  $E$ .

A equipe do São Paulo também terminou o campeonato com a melhor defesa (32 gols sofridos). O fator defesa, representado na figura 5.5, assim com o fator ataque ilustra bem a superioridade do São Paulo em relação aos demais clubes.

Já que no fator ataque, Flamengo e Santa Cruz tiveram comportamento semelhante, o que determinou a classificação final das duas equipes no campeonato foi o fator defesa,

no qual, o time carioca se mostrou mais eficiente que a equipe pernambucana.

O excelente aproveitamento do São Paulo nos dois fatores justifica, não só a sua colocação final na tabela, como o fato de ter garantido o título com duas rodadas de antecedência. Isso reforça a idéia de que um bom time deve ser forte tanto no ataque, quanto na defesa.

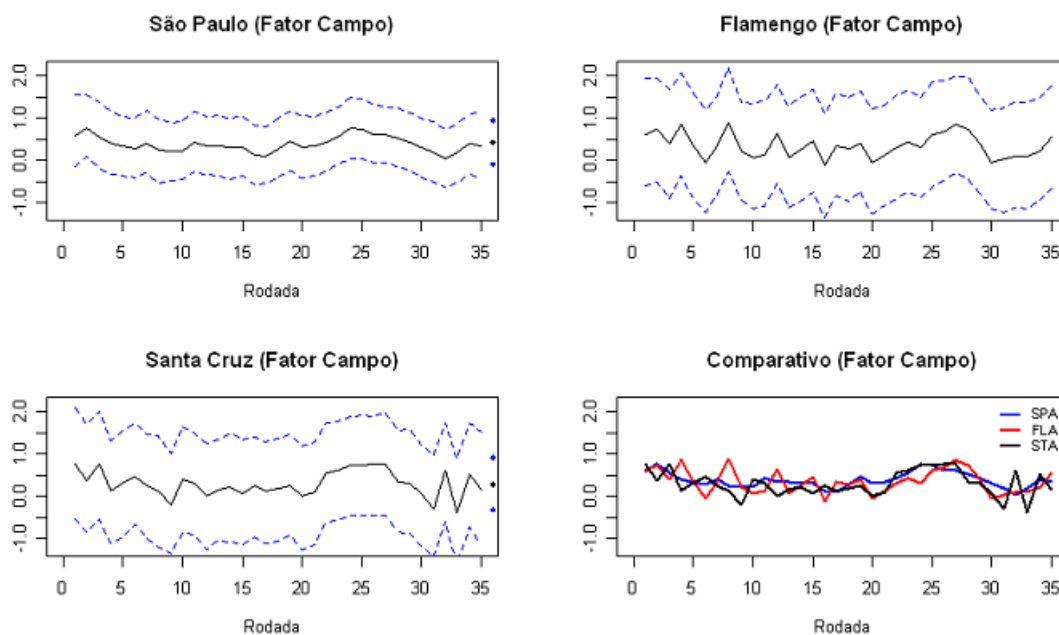


Figura 5.6: Gráfico de linhas dos fatores campo de 3 equipes selecionadas. As linhas cheias representam a mediana e as tracejadas os quantis 97,5% e 2,5%. Os pontos azuis e pretos cheios, para efeito de comparação, indicam a mesma informação para o modelo  $E$ .

No fator campo também ocorreu o esperado, já que os valores estiveram próximos de zero. Isso indica que o São Paulo não dependeu tanto de seu campo para fazer mais gols, e conseqüentemente vencer as partidas, reforçando também a idéia de que time que pretende ser campeão joga bem em qualquer lugar. Flamengo e Santa Cruz também se mostraram não ser tão dependentes de seus campos, a equipe carioca um pouco mais do que as duas outras equipes, porém essa diferença não foi tão grande.

Vale ressaltar que nos três fatores, o comportamento apresentado pelo modelo  $D_3$  oscilou em torno daquele apresentado no modelo estático.

O comportamento das demais equipes nos fatores ataque, defesa e campo também seguiu um padrão coerente com a classificação e o desempenho delas no torneio.

Os gráficos das figuras 5.7, 5.8 e 5.9 apresentam respectivamente os fatores ataque, defesa e campo das 20 equipes na rodada de número 35, para o modelo  $D_3$ .

Nesses gráficos, as equipes (eixo das abcissas) foram dispostas propositalmente de acordo com a classificação final do campeonato.

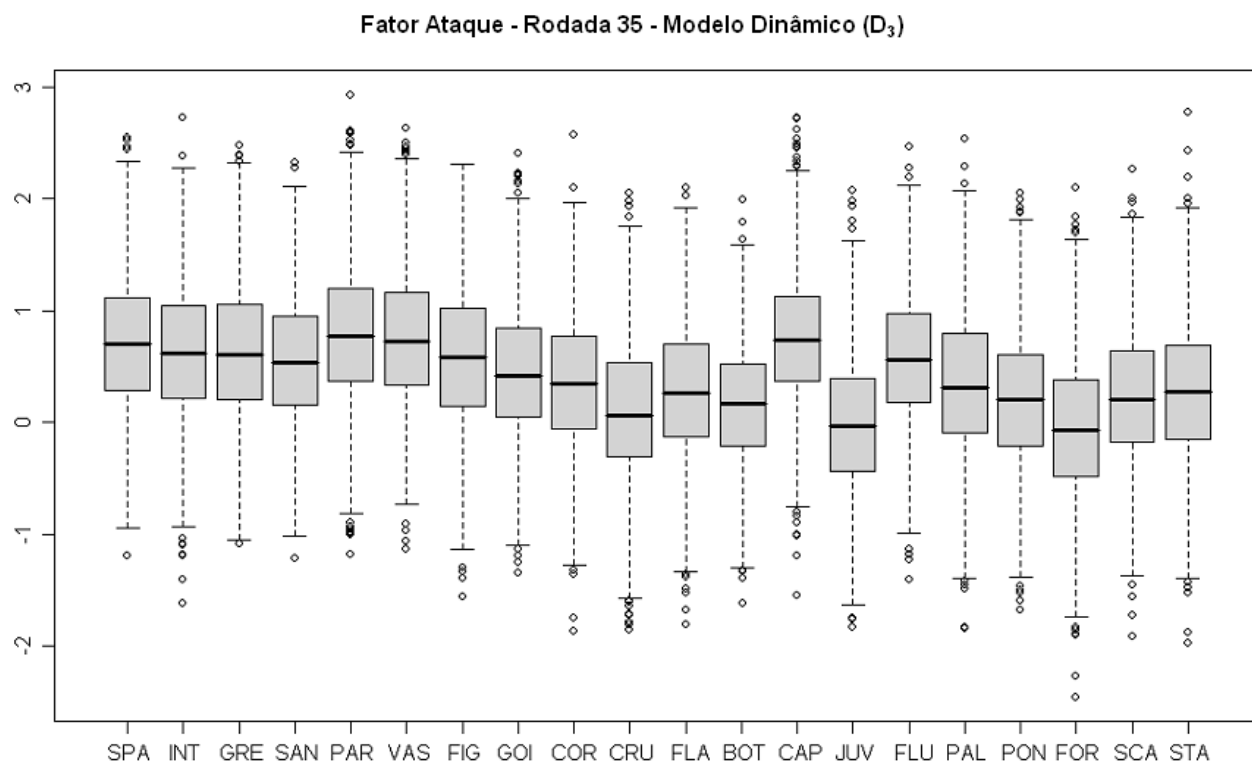


Figura 5.7: Gráficos do tipo *box-plot* dos fatores ataque ( $\alpha$ ) para as 20 equipes do campeonato referente a rodada de número 35.

Tanto na figura 5.7 quanto na 5.8, fica clara a idéia de que os times com melhores índices de ataque e defesa estão mais bem posicionados na tabela. Apesar de ainda faltarem 3 rodadas para acabar o campeonato, a classificação dos fatores ataque e defesa já possui uma correlação positiva com a classificação final do torneio.

Isso acontece porque o campeonato brasileiro é muito competitivo, e qualquer equipe que conseguir manter um bom nível, tanto no ataque quanto na defesa, até um certo estágio do campeonato, esse nível fará com que essa equipe se mantenha bem colocada por um certo número de rodadas.

Vale destacar nesses gráficos o comportamento da equipe do Atlético-PR, que contrastou um excelente fator ataque com um péssimo fator defesa, resultando em uma



posição mediana na tabela de classificação.

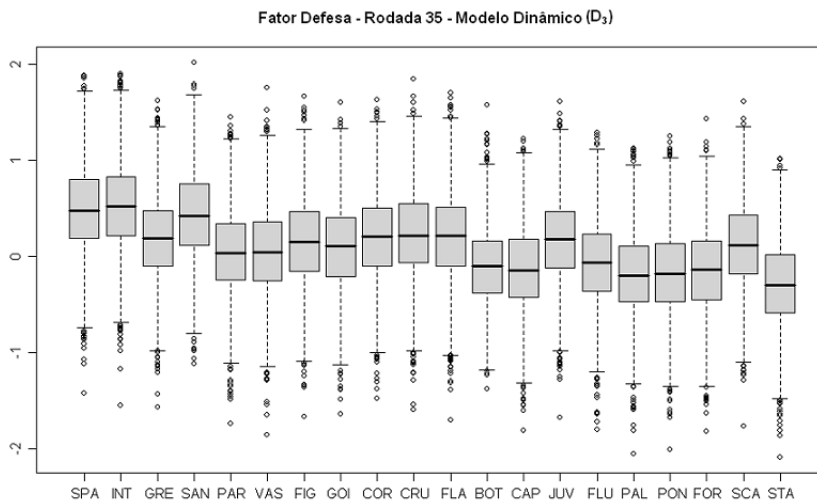


Figura 5.8: Gráficos do tipo *box-plot* dos fatores defesa ( $\beta$ ) para as 20 equipes do campeonato referente a rodada de número 35.

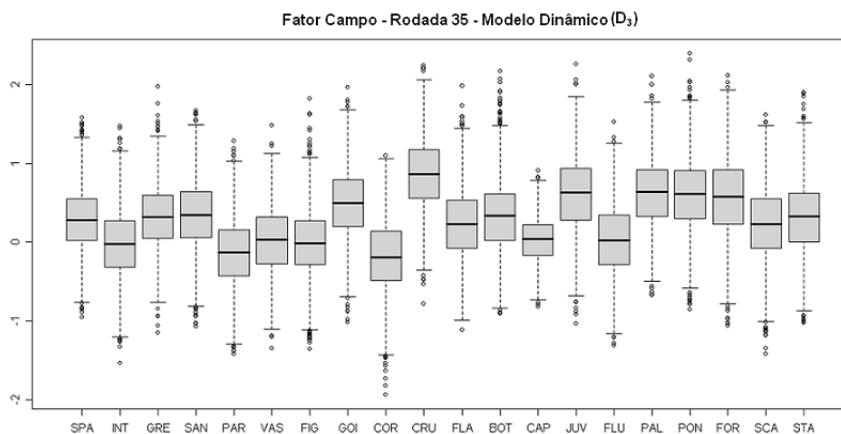


Figura 5.9: Gráficos do tipo *box-plot* dos fatores campo ( $\gamma$ ) para as 20 equipes do campeonato referente a rodada de número 35.

Na figura 5.9 ganham destaque os fatores campo, do Cruzeiro que demonstrou ter uma grande dependência de jogar em seu estádio, onde fez 67,3% (35 gols) do seu total de gols no campeonato, e do Corinthians. O último, em 15 dos 19 jogos que disputou em casa marcou no máximo 1 gol, o que é muito pouco para os padrões brasileiros de um time considerado de grande porte que joga em casa com o apoio de sua torcida. Os demais demonstraram comportamentos semelhantes.

Conforme visto anteriormente, a forma escolhida de comparação dos modelos na

previsão foi o cálculo da verossimilhança. Quanto maior esse valor, melhor é o poder de previsão do modelo.

A tabela 5.4 apresenta as verossimilhanças dos quatro modelos para cada uma das 3 rodadas finais, bem como a porcentagem de acerto na previsão.

Tabela 5.4: Valores das verossimilhanças para os modelos, por rodada.

Modelo	Rodada 36		Rodada 37		Rodada 38	
	Verossim.	Acerto	Verossim.	Acerto	Verossim.	Acerto
$E$	$7,061e^{-13}$	40%	$2,234e^{-14}$	60%	$3,259e^{-17}$	40%
$D$	$2,235e^{-12}$	40%	$9,612e^{-14}$	60%	$8,046e^{-17}$	40%
$D_1$	$6,296e^{-13}$	30%	$1,907e^{-14}$	60%	$3,162e^{-17}$	30%
$D_3$	$6,332e^{-12}$	40%	$5,039e^{-13}$	70%	$1,469e^{-16}$	40%

Observa-se que os valores das verossimilhanças são menores a medida que se passam as rodadas, independente do modelo. Fica clara, mais uma vez, a superioridade dos modelos dinâmicos em relação ao estático. Já o modelo dinâmico  $D_1$ , que possui um único coeficiente para os três fatores, obteve um rendimento ligeiramente inferior ao estático. A figura abaixo mostra o comportamento quase aleatório desse coeficiente.

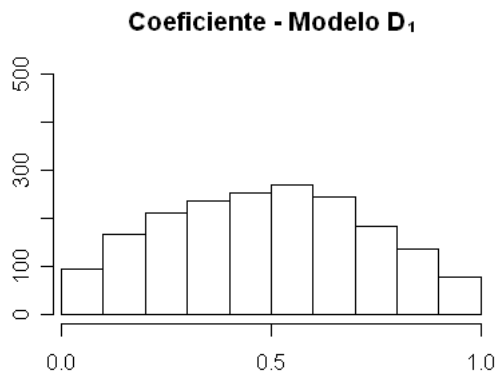


Figura 5.10: Histograma do coeficiente do modelo  $D_1$  ( $\rho$ ).

Como era esperado, o modelo dinâmico  $D_3$ , que é o mais geral, se mostrou melhor em comparação aos outros. A figura a seguir ilustra bem essa característica, pois ela mostra um gráfico comparativo do logaritmo das razões das verossimilhanças em relação ao modelo estático ( $E$ ).

Comparação do logaritmo das razões das verossimilhanças com o modelo Estático (E)

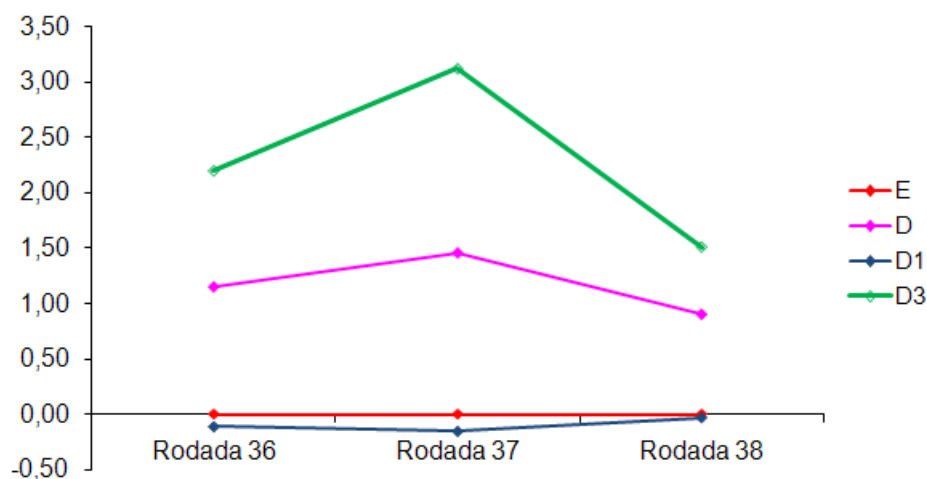


Figura 5.11: Gráfico de linhas do logaritmo das razões das verossimilhanças calculadas para os modelos  $E$ ,  $D$ ,  $D_1$  e  $D_3$  em relação ao modelo  $E$ .

Nessa figura, vale destacar que na rodada em que todos os modelos apresentaram os melhores percentuais de acerto, a rodada 37, o modelo proposto  $D_3$  foi aproximadamente 3 vezes melhor do que o modelo  $E$  na escala logrítmica. Isso corresponde a 22 vezes na escala decimal.

Tabela 5.5: Previsões do modelo  $D_3$  para a rodada de número 36.

Rodada 36	Vitória	Empate	Derrota
GRE 3 x 1 STA	<b><u>78,8%</u></b>	12,9%	8,3%
FLA 0 x 2 FIG	40,3%	27,4%	32,4%
COR 1 x 1 FLU	32,4%	31,5%	36,2%
JUV 3 x 2 PAL	<b><u>57,7%</u></b>	23,0%	19,4%
SPA 1 x 1 CAP	71,4%	16,0%	12,7%
SCA 0 x 1 VAS	33,6%	26,2%	<b><u>40,3%</u></b>
PAR 1 x 0 INT	27,4%	28,3%	44,4%
BOT 2 x 2 GOI	56,8%	20,8%	22,5%
CRU 1 x 1 SAN	44,9%	27,4%	27,7%
FOR 1 x 0 PON	<b><u>54,3%</u></b>	23,3%	22,5%

As tabelas 5.5, 5.6 e 5.7 apresentam as previsões detalhadas por rodada e por jogo

para o modelo  $D_3$ .

Tabela 5.6: Previsões do modelo  $D_3$  para a rodada de número 37.

Rodada 37	Vitória	Empate	Derrota
BOT 0 x 0 COR	35,0%	24,6%	40,5%
GRE 3 x 0 FLA	<u>44,7%</u>	25,7%	29,6%
JUV 2 x 2 FOR	36,4%	33,7%	30,0%
SPA 2 x 0 CRU	<u>50,8%</u>	25,6%	23,6%
SCA 0 x 2 PAR	25,8%	23,0%	<u>51,2%</u>
VAS 1 x 1 SAN	38,9%	24,3%	36,9%
GOI 3 x 0 PON	<u>45,8%</u>	24,1%	30,1%
STA 1 x 2 FLU	32,3%	22,4%	<u>45,4%</u>
PAL 1 x 4 INT	31,1%	22,5%	<u>46,5%</u>
CAP 1 x 4 FIG	37,8%	20,7%	<u>41,6%</u>

Tabela 5.7: Previsões do modelo  $D_3$  para a rodada de número 38.

Rodada 38	Vitória	Empate	Derrota
INT 1 x 4 GOI	39,9%	28,1%	32,0%
FLA 4 x 1 SCA	<u>41,5%</u>	26,6%	32,0%
FLU 1 x 1 PAL	43,6%	21,2%	35,3%
FIG 0 x 0 VAS	39,3%	23,9%	36,8%
PON 1 x 1 CAP	41,2%	21,2%	37,7%
PAR 0 x 0 SPA	34,7%	23,8%	41,6%
CRU 3 x 1 BOT	<u>48,1%</u>	24,3%	27,6%
COR 5 x 3 JUV	<u>39,8%</u>	27,7%	32,5%
SAN 3 x 1 STA	<u>51,1%</u>	22,5%	26,5%
FOR 1 x 0 GRE	33,4%	24,6%	42,1%

Dos últimos 30 jogos o modelo  $D_3$  acertou 15, o que corresponde a 50% de acerto no geral. Vale destacar o ótimo aproveitamento do modelo na rodada de número 37, na qual ele obteve um percentual de acerto equivalente a 70%.

A seguir, a tabela 5.8 apresenta uma comparação entre a tabela de classificação final do campeonato e uma tabela criada a partir da rodada de número 35 e completada pelos resultados previstos pelo modelo  $D_3$ .

Tabela 5.8: Previsão do modelo  $D_3$  para a classificação final das equipes

Classificação Real				Classificação Prevista			
Pos.	Clube	PG	V	Pos.	Clube	PG	V
1°	São Paulo	78	22	1°	São Paulo	82	24
2°	Internacional	69	20	2°	Internacional	75	22
3°	Grêmio	67	20	3°	Grêmio	70	21
4°	Santos	64	18	4°	Santos	62	18
5°	Paraná	60	18	5°	Vasco	60	16
6°	Vasco	59	15	6°	Paraná	56	17
7°	Figueirense	57	15	7°	Figueirense	56	15
8°	Goiás	55	15	8°	Flamengo	55	16
9°	Corinthians	53	15	9°	Cruzeiro	55	15
10°	Cruzeiro	53	14	10°	Corinthians	54	16
11°	Flamengo	52	15	11°	Botafogo	52	14
12°	Botafogo	51	13	12°	Goiás	51	14
13°	Atlético-PR	48	13	13°	Juventude	49	14
14°	Juventude	47	13	14°	Fluminense	49	13
15°	Fluminense	45	11	15°	Atlético-PR	46	13
16°	Palmeiras	44	12	16°	Palmeiras	43	12
17°	Ponte Preta	39	10	17°	Ponte Preta	41	11
18°	Fortaleza	38	8	18°	São Caetano	36	9
19°	São Caetano	36	9	19°	Fortaleza	34	7
20°	Santa Cruz	28	7	20°	Santa Cruz	28	7

Esse tipo de visualização das previsões permite que um outro tipo de análise seja feita, o que não dava para ser feito olhando somente para aos resultados dos jogos previstos por rodada.

Observando a tabela podemos ver que o modelo  $D_3$  acertou:

- 8 posições na tabela;
- o campeão;
- 3 dos 4 clubes que se classificaram para disputar a Taça Libertadores do ano

seguinte;

- os 4 clubes que foram rebaixados para a série B do ano seguinte;
- 6 dos 8 clubes que se classificaram para disputar a Copa Sul-Americana do ano seguinte.

# Capítulo 6

## Conclusões

O modelo proposto por esse trabalho obteve resultados satisfatórios. Para as três últimas rodadas do campeonato brasileiro de 2006 ele foi, respectivamente, 2,83, 5,24 e 1,83 vezes melhor que o modelo  $D$ , comparando suas verossimilhanças. O último, foi proposto inicialmente por RUE e SALVESEN (2000), e reescrito por SOUZA e GAMERMAN (2004) para dados do campeonato brasileiro de 2002 e 2003.

A grande diferença em favor do modelo  $D_3$  foi o uso de 3 coeficientes auto-regressivos ( $\rho_\alpha$ ,  $\rho_\beta$  e  $\rho_\gamma$ ) na equação de evolução do modelo. Esses coeficientes apresentaram valores entre 0,35 e 0,98, o que confronta diretamente com os respectivos valores do modelo  $D$ , que são todos iguais a 1.

O modelo  $D_3$  obteve performance razoável em suas previsões, tanto nos resultados das partidas quanto nas informações finais, como o campeão e os rebaixados. Além disso, ele se mostrou bem coerente ao descrever o comportamento das equipes em relação as suas classificações.

Vale lembrar que o modelo proposto possui a importante característica de poder ser usado para analisar qualquer temporada do campeonato brasileiro de futebol, bem como, campeonatos de outros países e até outros esportes, considerando confrontos pareados e sistemas de rodadas, como no futebol.

É possível adicionar informações de campeonatos anteriores como informação inicial para os parâmetros das equipes que permaneceram no torneio, já para as equipes novatas, um procedimento razoável seria adotar o uso de prioris não-informativas. Fica como

sugestão para trabalhos futuros usar as informações de campeonatos anteriores como prioris para as equipes que se mantiveram no campeonato.

Portanto, ficaria também como sugestão estender o estudo para o campeonato de 2008, e fazer previsões para as rodadas finais.

Nesse trabalho optou-se pelo uso dos métodos MCMC disponíveis para estimar as distribuições a posteriori. Em RUE, MARTINO e CHOPIN (2007) sugere-se que o uso de aproximações integrais encaixadas de Laplace (*Integrated Nested Laplace Approximations*) e suas versões simplificadas pode, na maioria dos casos, produzir diretamente aproximações bastante acuradas das distribuições a posteriori. O principal benefício desse método estaria no corte dramático dos custos computacionais. Uma sugestão para trabalhos futuros é usar esse método para obter amostras da distribuição a posteriori.



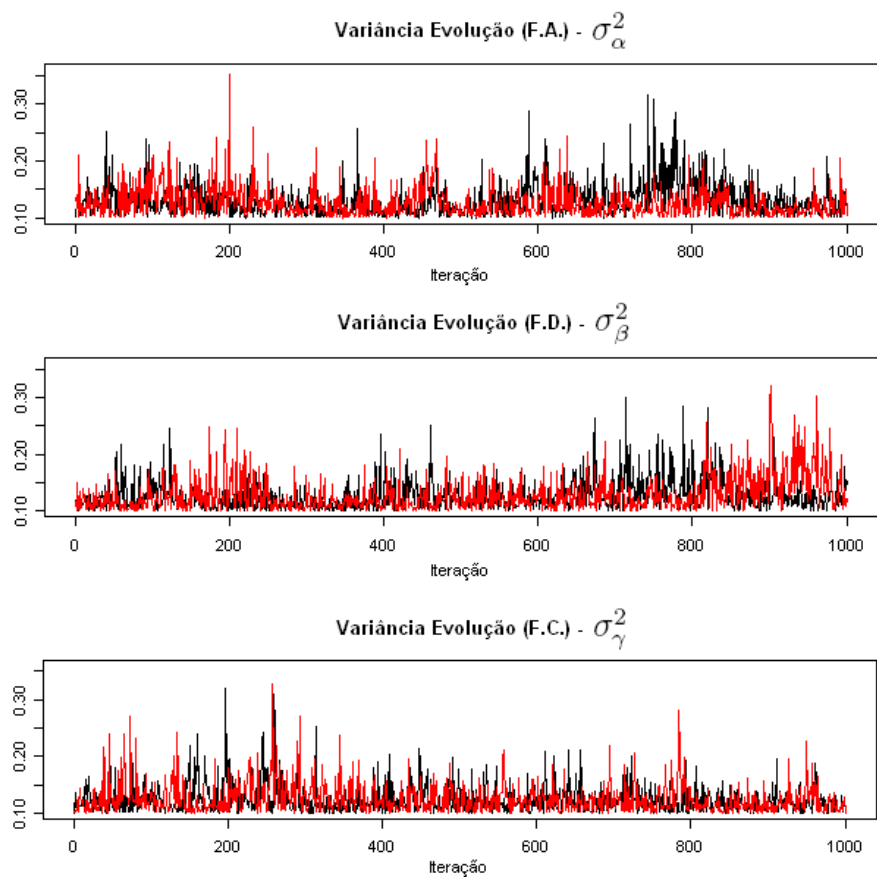
# Referências Bibliográficas

- [1] ARRUDA, M. L. (2000). Poisson, Bayes, Futebol e DeFinetti. Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para a obtenção do grau de Mestre em Estatística.
- [2] DA MATTA, R. (1994). Antropologia do óbvio: notas em torno do significado social do futebol brasileiro. Dossiê Futebol, Revista USP, p. 16.
- [3] DE GROOT, M. H. e SCHERVISH, M. J. (2002). Probability and Statistics 3rd ed. Addison-Wesley.
- [4] GAMERMAN, D. (1996). Simulação Estocástica Via Cadeias de Markov. ABE - Associação Brasileira de Estatística.
- [5] KNORR-HELD, L. (1999). Dynamic rating of sports teams. Ludwig-Maximilians University. Munique, Alemanha.
- [6] MEYER, P. L. (2000). Probabilidade Aplicações À Estatística. Editora LTC. Segunda Edição.
- [7] MIGON, H. S. e GAMERMAN, D. (1999). Statistical Inference - an integrated approach. Arnold.
- [8] RUE, H e SALVESEN, O. (2000). Prediction and retrospective analysis of soccer matches in a league. Norwegian University of Science and Technology. Trondheim, Noruega.

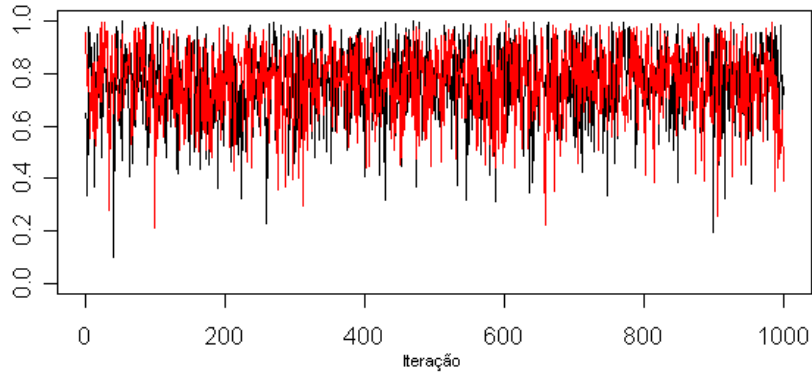
- [9] RUE, H. , MARTINO, S e CHOPIN, N. (2007) Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. Norwegian University of Science and Technology. Trondheim, Noruega.
- [10] SOUZA JR, O. G. e GAMERMAN, D. (2004). Previsão de partidas de futebol usando modelos dinâmicos. Anais do XXXVI SBPO. São João del Rey - MG.
- [11] THOMAS, A., SPIEGELHALTER, D. J. e GILKS, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs Sampling. . Em Bayesian Statistics 4, 837-842 (Editors Bernardo, J.M., Berger, J.O., Dawid, A.P. e Smith, A.F.M.).
- [12] WEST, M. e HARRISON, J. (1997). Bayesian Forecasting and Dynamic Models. *Springer-Verlag, Springer Series in Statistics - New York*. Second Edition.
- [13] R version 2.7.1 (2008): The R Foundation for Statistical Computing.

# Apêndice A:

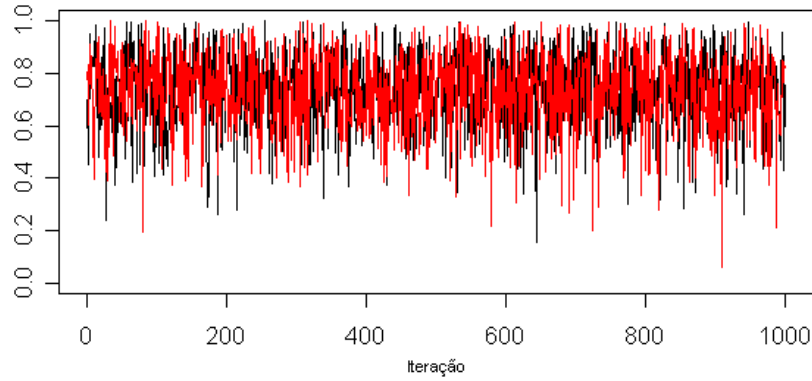
## Gráficos diversos para o Modelo $D_3$



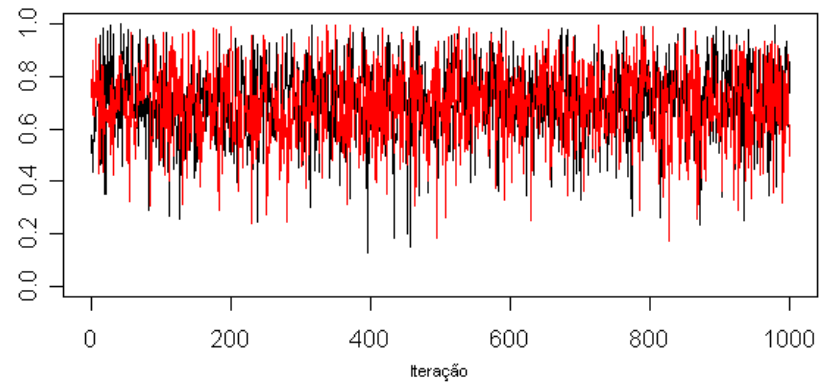
**Coeficiente Autoregressivo (F.A.) -  $\rho_\alpha$**

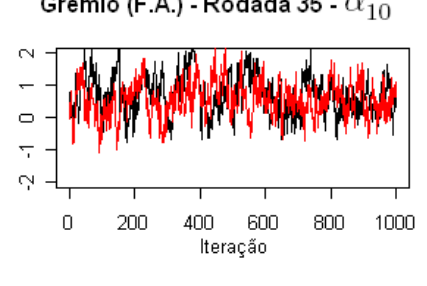
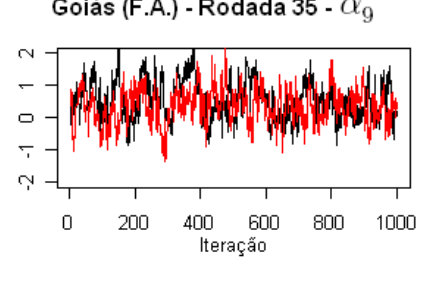
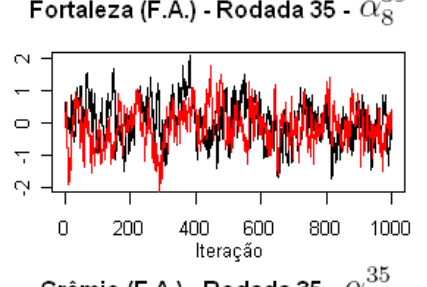
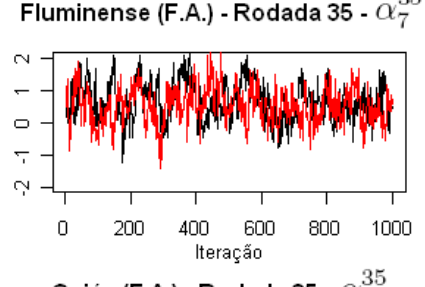
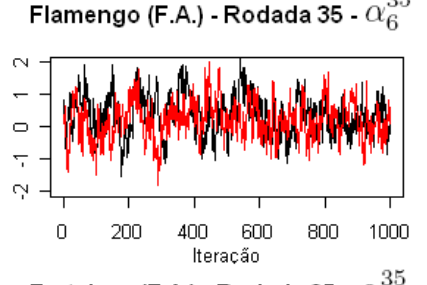
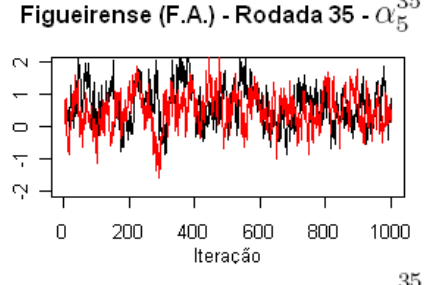
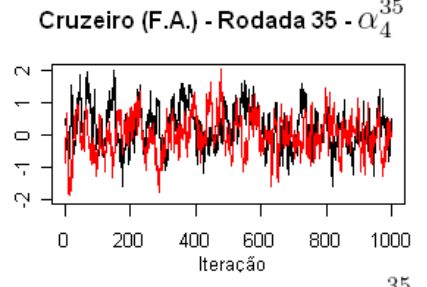
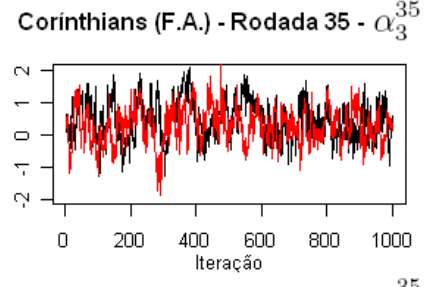
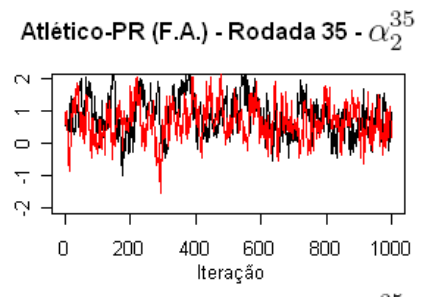
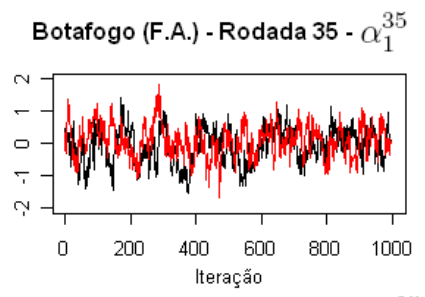


**Coeficiente Autoregressivo (F.D.) -  $\rho_\beta$**

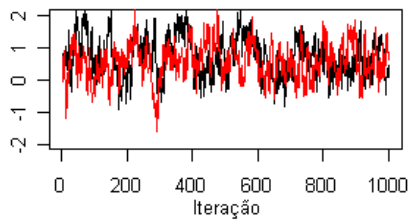


**Coeficiente Autoregressivo (F.C.) -  $\rho_\gamma$**

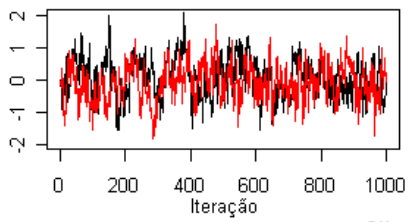




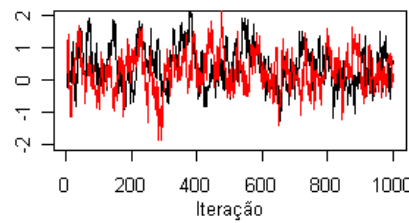
Internacional (F.A.) - Rodada 35 -  $\alpha_{11}^{35}$



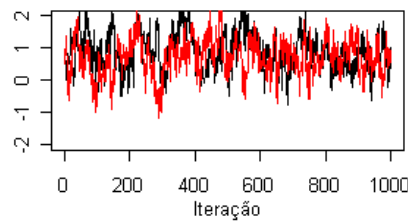
Juventude (F.A.) - Rodada 35 -  $\alpha_{12}^{35}$



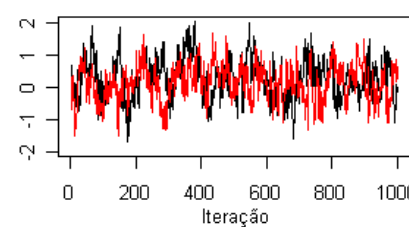
Palmeiras (F.A.) - Rodada 35 -  $\alpha_{13}^{35}$



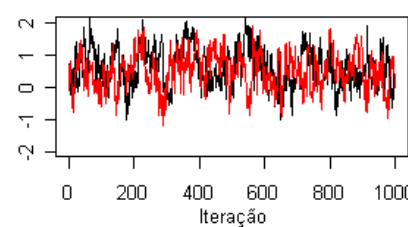
Paraná (F.A.) - Rodada 35 -  $\alpha_{14}^{35}$



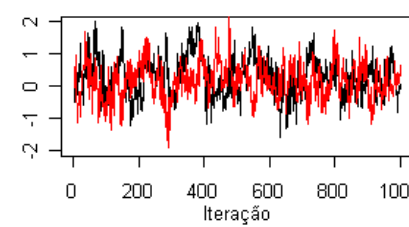
Ponte Preta (F.A.) - Rodada 35 -  $\alpha_{15}^{35}$



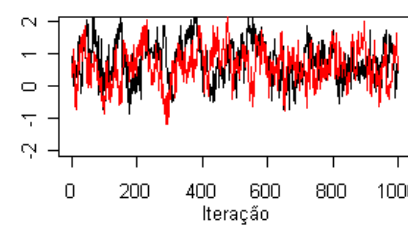
Santos (F.A.) - Rodada 35 -  $\alpha_{16}^{35}$



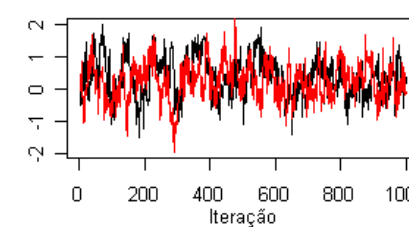
São Caetano (F.A.) - Rodada 35 -  $\alpha_{17}^{35}$



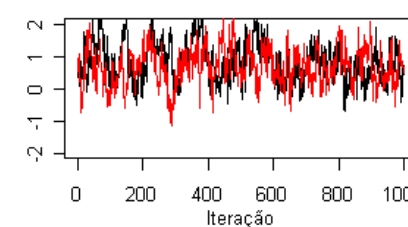
São Paulo (F.A.) - Rodada 35 -  $\alpha_{18}^{35}$

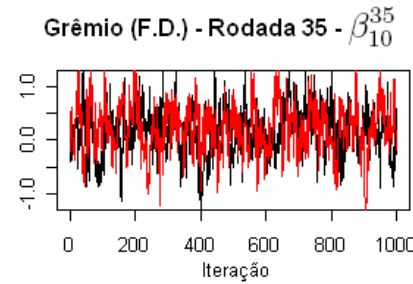
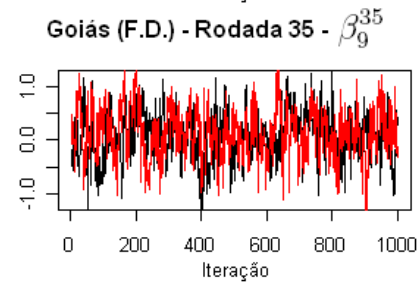
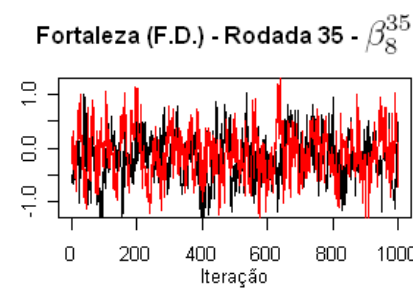
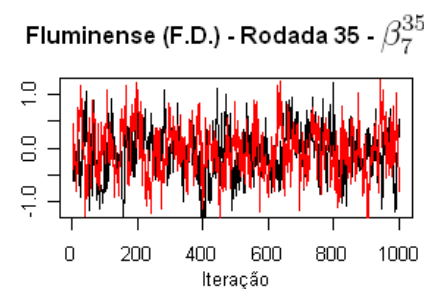
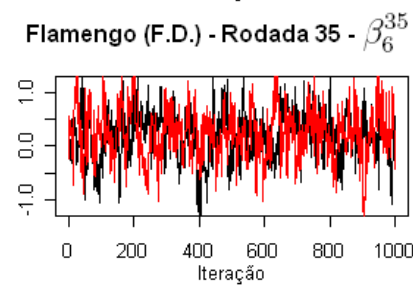
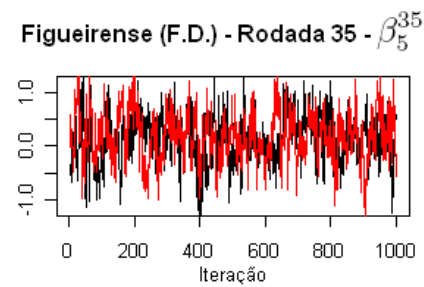
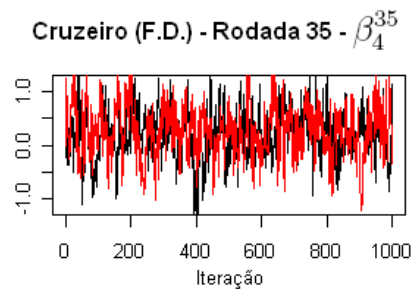
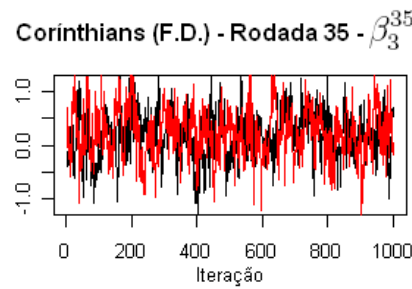
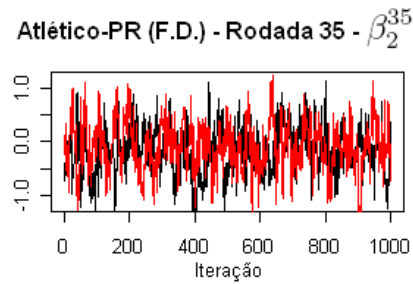
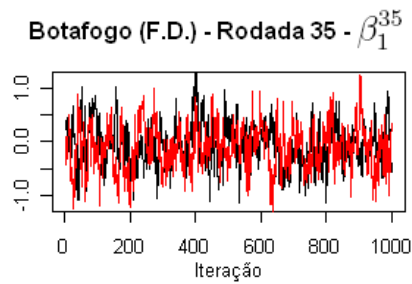


Santa Cruz (F.A.) - Rodada 35 -  $\alpha_{19}^{35}$

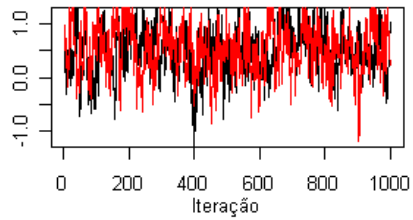


Vasco (F.A.) - Rodada 35 -  $\alpha_{20}^{35}$

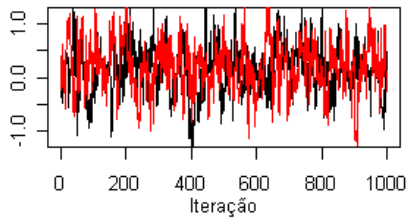




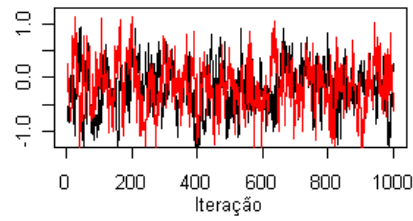
Internacional (F.D.) - Rodada 35 -  $\beta_{11}^{35}$



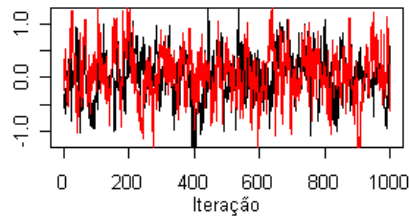
Juventude (F.D.) - Rodada 35 -  $\beta_{12}^{35}$



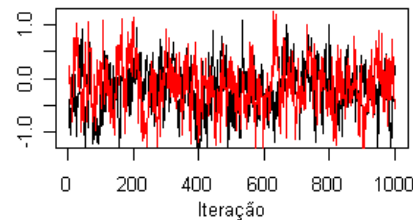
Palmeiras (F.D.) - Rodada 35 -  $\beta_{13}^{35}$



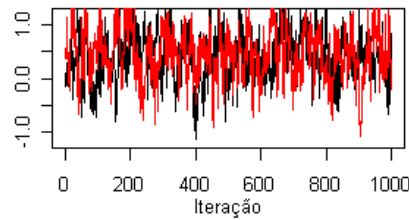
Paraná (F.D.) - Rodada 35 -  $\beta_{14}^{35}$



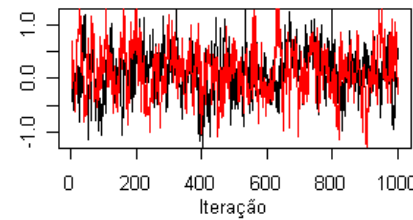
Ponte Preta (F.D.) - Rodada 35 -  $\beta_{15}^{35}$



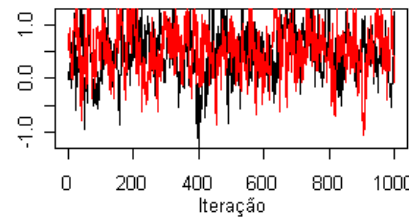
Santos (F.D.) - Rodada 35 -  $\beta_{16}^{35}$



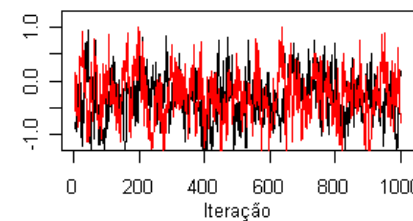
São Caetano (F.D.) - Rodada 35 -  $\beta_{17}^{35}$



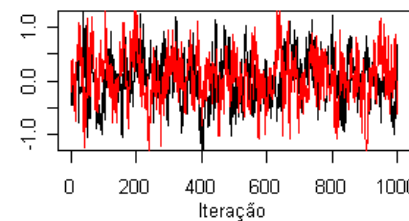
São Paulo (F.D.) - Rodada 35 -  $\beta_{18}^{35}$



Santa Cruz (F.D.) - Rodada 35 -  $\beta_{19}^{35}$

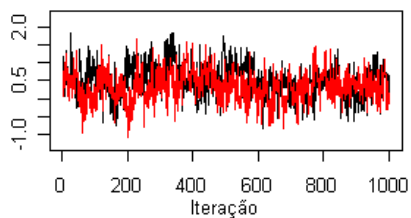


Vasco (F.D.) - Rodada 35 -  $\beta_{20}^{35}$

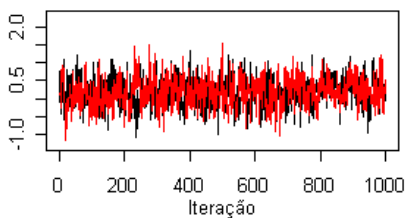




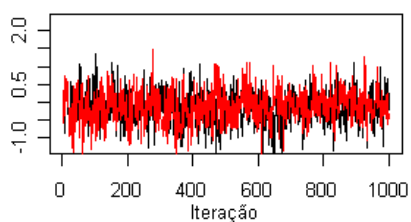
**Botafogo (F.C.) - Rodada 35 -  $\gamma_1^{35}$**



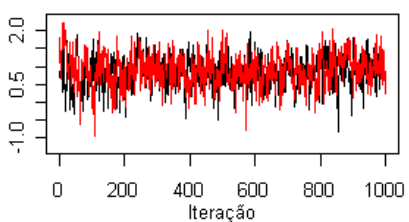
**Atlético-PR (F.C.) - Rodada 35 -  $\gamma_2^{35}$**



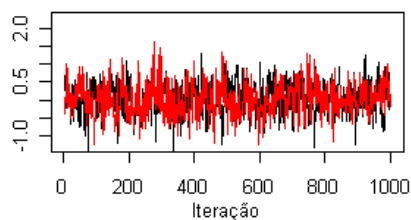
**Corinthians (F.C.) - Rodada 35 -  $\gamma_3^{35}$**



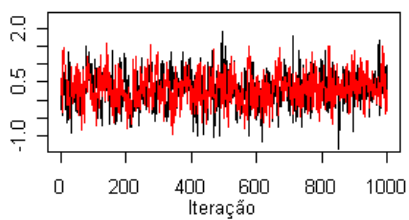
**Cruzeiro (F.C.) - Rodada 35 -  $\gamma_4^{35}$**



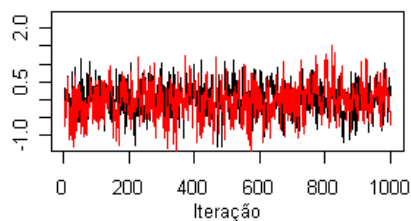
**Figueirense (F.C.) - Rodada 35 -  $\gamma_5^{35}$**



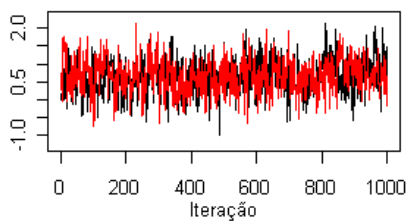
**Flamengo (F.C.) - Rodada 35 -  $\gamma_6^{35}$**



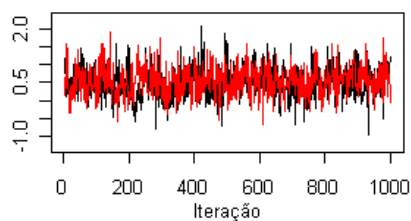
**Fluminense (F.C.) - Rodada 35 -  $\gamma_7^{35}$**



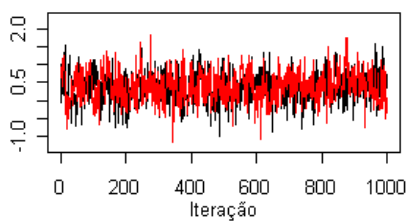
**Fortaleza (F.C.) - Rodada 35 -  $\gamma_8^{35}$**



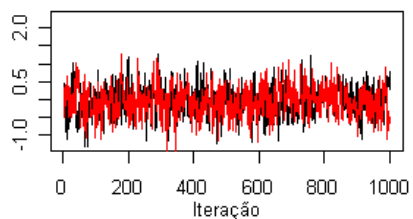
**Goiás (F.C.) - Rodada 35 -  $\gamma_9^{35}$**



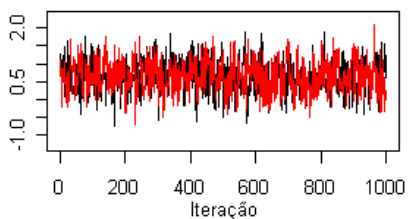
**Grêmio (F.C.) - Rodada 35 -  $\gamma_{10}^{35}$**



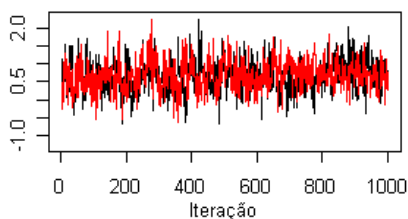
**Internacional (F.C.) - Rodada 35 -  $\gamma_{11}^{35}$**



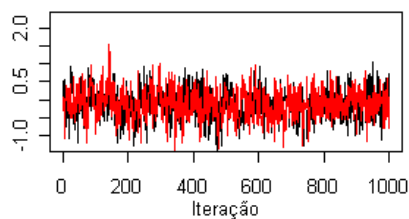
**Juventude (F.C.) - Rodada 35 -  $\gamma_{12}^{35}$**



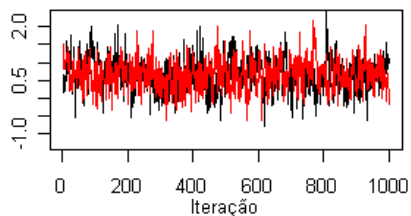
**Palmeiras (F.C.) - Rodada 35 -  $\gamma_{13}^{35}$**



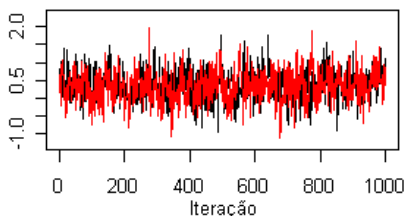
**Paraná (F.C.) - Rodada 35 -  $\gamma_{14}^{35}$**



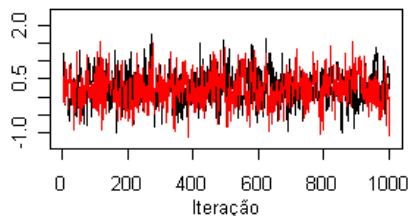
**Ponte Preta (F.C.) - Rodada 35 -  $\gamma_{15}^{35}$**



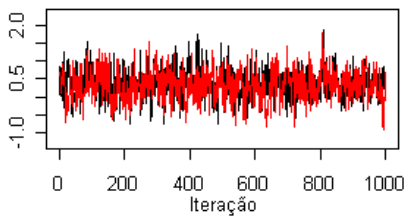
**Santos (F.C.) - Rodada 35 -  $\gamma_{16}^{35}$**



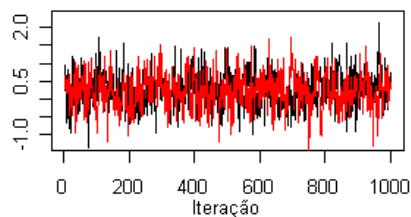
**São Caetano (F.C.) - Rodada 35 -  $\gamma_{17}^{35}$**



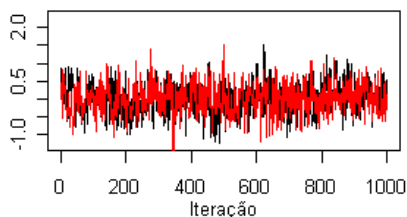
**São Paulo (F.C.) - Rodada 35 -  $\gamma_{18}^{35}$**



**Santa Cruz (F.C.) - Rodada 35 -  $\gamma_{19}^{35}$**



**Vasco (F.C.) - Rodada 35 -  $\gamma_{20}^{35}$**



## Apêndice B:

# Algoritmo do WINBUGS para fazer inferência até a rodada $s$ para o modelo $D_3$

```
model;
{
tau1 ~ dgamma(a_tau1, b_tau1)
tau2 ~ dgamma(a_tau2, b_tau2)
tau3 ~ dgamma(a_tau3, b_tau3)
tau4 ~ dgamma(a_tau4, b_tau4)
for(t in 1 : s) {
w1[t] ~ dnorm(0, tau1)
w2[t] ~ dnorm(0, tau2)
w3[t] ~ dnorm(0, tau3)
w4[t] ~ dnorm(0, tau4)
}
theta[1, 1] ~ dnorm(mu_theta_1, xi_theta_1)
for(t in 2 : s) {
theta[1, t] < -theta[1, t - 1] + w1[t]
}
for(j in 2 : n) {
theta[j, 1] ~ dnorm(mu_theta_2, xi_theta_2)
}
for(t in 2 : s) {
for(j in 2 : n) {
theta[j, t] < -rho1 * theta[j, t - 1] + w2[t]
}
}
for(j in (n + 1) : (2 * n - 1)) {
```

```

theta[j, 1] ~ dnorm(mu_theta_3, xi_theta_3)
}
for(t in 2 : s) {
  for(j in (n + 1) : (2 * n - 1)) {
    theta[j, t] < -rho2 * theta[j, t - 1] + w3[t]
  }
}
for(j in (2 * n) : (3 * n - 1)) {
  theta[j, 1] ~ dnorm(mu_theta_4, xi_theta_4)
}
for(t in 2 : s) {
  for(j in (2 * n) : (3 * n - 1)) {
    theta[j, t] < -rho3 * theta[j, t - 1] + w4[t]
  }
}
rho1 ~ dunif(0, 1)
rho2 ~ dunif(0, 1)
rho3 ~ dunif(0, 1)
for(j in 1 : (n * s)) {
  y[j] ~ dpois(lambda[j])
}
for(t in 1 : s) {
  for(j in (n * t - (n - 1)) : (n * t)) {
    log(lambda[j]) < -inprod(X[j, ], theta[, t])
  }
}
}
}

```

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)