



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

Uma Aplicação do Teste Adaptativo  
Computadorizado via Filtro de Kalman  
Não-Linear

por

Júlio César Gomes Fonseca

Orientadora

Profa. Dra. Cibele Queiroz da Silva

Co-orientador

Prof. Dr. Raul Yukihiro Matsushita

Brasília, julho de 2019



# Agradecimentos

*O tempo nunca foi, e nunca será, o fim da linha, é apenas o começo ...*

... e cá estou, vencedor! Gostaria de aproveitar esta maravilhosa oportunidade para expressar meus mais sinceros agradecimentos a todos aqueles que fizeram parte desta jornada, cada um, a seu modo, contribui para que este momento fosse possível. Em especial, gostaria de agradecer:

A Deus, pois sem Ele e sua imensa bondade e glória, nossas vidas não fazem sentido algum. Graças a Ele conheci as pessoas certas, na hora certa, e isso fez e faz toda a diferença.

À minha família, sempre ao meu lado, apoio incondicional em simplesmente, TODOS os momentos, particularmente, nos mais difíceis. Não há palavra que eu conheça para traduzir minha gratidão a todos vocês, que sempre reduziram 750km de distância a palavras de fé, esperança e sustentação.

Aos meus pais, *José Geraldo Pereira da Fonseca* e *Maria Gomes da Fonseca*, minha eterna gratidão. Sem vocês meus queridos, a vida não teria e não tem o mesmo sabor. Minha determinação jamais seria a mesma sem nossas longas conversas. Prometo nunca parar de sonhar. Amo e respeito muito vocês!

À minha amada irmã, *Fernanda Gomes da Fonseca*, guerreira e parceira em tudo. Sempre o apoio incondicional, seja o que fosse. Eu contei e conto você. te amo!

Ao meu filho, *Michel Dassan de Jesus Fonseca*, a quem dedico este trabalho e todos os outros que virão. Se resisti até o fim, o fiz por você, e o faria novamente, quantas vezes fosse preciso, te amo!

À minha querida sobrinha, *Kauany Gomes Ribeiro*, que com seu jeito meigo e doce, sempre me fez pensar o quanto precisamos de gentileza nesta vida. Voe meu anjo, o futuro é seu, te amo!

À minha querida professora e orientadora, Profa. Dra. *Cibele Queiroz da Silva*, que iluminada por muita luz, sabedoria e dignidade, me conduziu ao sucesso. Sem a sua paciência e grande sabedoria, este trabalho jamais teria a mesma qualidade. Minha eterna gratidão pelo carinho e dedicação com que conduziu este trabalho, além da oportunidade de expandir meus conhecimentos sobre Teoria da Resposta ao Item, quando me convidou para participar do projeto de pesquisa realizado em conjunto com o CEBRASPE. Sem seus conselhos e orientação, nada seria o mesmo. Muito obrigado por acreditar em mim e proporcionar a oportunidade de chegar até onde cheguei.

Ao meu querido professor e coorientador, Prof. Dr. *Raul Yukihiro Matsushita*, por seu enorme conhecimento e dedicação a todos nós, alunos de pós-graduação. Sem sua inspiração, vasto conhecimento e paciência milenar, este trabalho não seria o mesmo. Muito obrigado por acreditar em mim e proporcionar a oportunidade de chegar até onde cheguei.

A todos os professores do Programa de Pós-graduação em Estatística (PPGEST-UnB), pela imensa contribuição com meus conhecimentos acadêmicos. Em especial aos professores, *André Cançado*, *Antônio Eduardo*, *Bernardo Borba*, *Cira Guevara*, *Eduardo Nakano* e *Gustavo Gilardoni*. Foram muitas horas de estudos e dedicação, valeu a pena cada segundo. Muito obrigado a todos!

A todos os funcionários do Departamento de Estatística, em especial, à *Tati*, *Ed*, *Lucas*, *André* e *Karen*. Sem a enorme paciência, prestatividade e a honrosa dedicação de vocês ao trabalho e a todos nós, alunos e incansáveis chatos, a pedir ajuda e conselhos, não teria participado de nenhum congresso e sequer conseguido resolver qualquer problema acadêmico. Vocês são demais!

Ao Programa de Pós-graduação em Estatística (PPGEST-UnB), pela honra em me acolher como parte deste time durante todo o tempo em que fui aluno e também representante dos discentes junto colegiado da pós-graduação.

Aos meus primeiros, queridos e eternos mestres em estatística, Profa. Dra. *Arminda Lucia Siqueira*, Prof. Dr. *Enrico Antônio Colosimo*, Prof. Dr. *Antônio Eduardo Gomes*, Prof. Dr. *Gregório Saravia Atuncar*, Prof. Dra. *Maria de Fátima Brant Drumond*, Prof. Dra. *Maria de Lourdes Moreira Fernandes*, Prof. Dra. *Rosângela Helena Loschi*, Prof. Dr. *Roberto da Costa Quinino*, Profa. Dra. *Sueli Aparecida Mingoti*, Prof. Dr. *Wanderley Ramalho* e em especial, ao Prof. Dr. *José Francisco Soares*, Prof. Dr. *Frederico Rodrigues Borges da Cruz*, e Prof. Dr. *Wilton de Oliveira Bussab*, com quem tive a honra e a oportunidade de trabalhar, respectivamente, em parceria SAGI-MDS/INEP, em projeto de pesquisa do Departamento de Estatística da UFMG e em projeto de extensão do Programa de Pós-graduação do Departamento de Sociologia da UFMG.

Aos meus queridos amigos, *Frederico Soares Ribeiro*, *William Lima*, *Vinicius Suzuki* e *Marcos Forti* pela presença e constante força nos meus passos nos últimos 10 anos, a quem revelei minhas frustrações, decepções, cansaços e alegrias, sempre com o incentivo incondicional de vocês.

Ao amigo e parceiro, *Suélio Alves de Moura*, pelas incansáveis horas de dedicação e atenção contínua, fosse nos estudos, nos congressos, nos cursos ou na convivência em Brasília.

À CAPES pelo apoio financeiro, que durante os primeiros três meses foi meu pão e vinho, e posteriormente, pela acolhida como parte do time de colaboradores, da qual tenho orgulho de pertencer. Minha gratidão eterna pela oportunidade de estudar.

Ao CEBRASPE pelo apoio financeiro e ampliação de meus conhecimentos, adquiridos em projeto de extensão executado com a orientação da Profa. Dra. *Cibele Queiroz da Silva*, intitulado “*Testes Adaptativos Computadorizados: calibração online de novos itens*”.

*“I have no special talent. I am only passionately curious.”*

(Albert Einstein)



*Ao meu filho,*

***Michel Dassin de Jesus Fonseca***





# Resumo

*Computerized Adaptive Test* (CAT) é um tipo de teste baseado em recursos computacionais que se adapta ao nível de habilidade dos examinandos. Essencialmente, uma implementação do CAT requer dois elementos fundamentais: um banco de itens calibrado e um algoritmo para seleção adaptativa dos itens. Este trabalho tem foco no segundo elemento, para o qual apresentamos uma adaptação à formulação não linear dos filtros de Kalman, prescrita por Harvey (1989), combinada com o conhecido princípio da máxima entropia (Kullback, 1997). Neste trabalho, estudos de simulação avaliam a acurácia da estimação de proficiências em aplicações CAT, desenvolvendo duas classes de estimadores utilizando a metodologia da Teoria da Resposta ao Item para construção dos modelos probabilísticos, e suavizações recursivas ancoradas nos modelos dinâmicos, para a formulação dos modelos de *Filtro de Kalman* (Harvey, 1989) e *Power Steady* (Smith, 1979). Como resultado são apresentados sete novos estimadores da proficiência, sendo o primeiro baseado no *Filtro de Kalman* e os demais no *Power Steady*, a saber: FK\_EAP, PS\_EAP, PS\_MAP, PS\_EAP\_AVG, PS\_EAP\_MEDIAN, PS\_MAP\_AVG, PS\_MAP\_MEDIAN. Os quatro últimos são suavizações baseadas na média e na mediana das estimativas correntes de  $\theta$ . Com as simulações, observou-se que os estimadores PS\_EAP e PS\_MAP geram estimativas cíclicas, sendo corrigidas com a suavização proporcionada pelo estimador PS\_EAP\_AVG. Portanto, com o estudo realizado nesta dissertação conclui-se que, dentre os sete estimadores avaliados, o PS\_EAP\_AVG e o FK\_EAP geram as melhores alternativas às estimativas de proficiência ao CAT tradicional.

**Palavras-chave:** Testes Adaptativos Computadorizados. Teoria Resposta ao Item. Filtro de Kalman. Modelos Dinâmicos. Estimação Bayesiana.



# Abstract

Computerized Adaptive Test (CAT) is a type of test based on computational resources that adapts to the level of ability of the examinees. Essentially, an implementation of CAT requires two fundamental elements: a calibrated item bank and a algorithm for adaptive selection of items. This work focuses on the second element, for which we present an adaptation to the nonlinear formulation of the Kalman filters, prescribed by [Harvey \(1989\)](#), combined with the known principle of maximum entropy ([Kullback, 1997](#)). In this work, simulation studies evaluate the accuracy of proficiency estimation in CAT applications, developing two classes of estimators using Item Response Theory methodology to construct the probabilistic models, and recursive smoothing anchored in the dynamic models, for the formulation of the models of Kalman Filter ([Harvey, 1989](#)) and Power Steady ([Smith, 1979](#)). As a result, seven new proficiency estimators are presented, the first one being based on the Kalman Filter and the others in Power Steady, namely: FK\_EAP, PS\_EAP, PS\_MAP, PS\_EAP\_AVG, PS\_EAP\_MEDIAN, PS\_MAP\_AVG, PS\_MAP\_MEDIAN. The last four are smoothing based on the mean and median. With the simulations, it was observed that the estimators PS\_EAP and PS\_MAP generate cyclic estimates, being corrected with the smoothing provided by the estimator PS\_EAP\_AVG. Therefore, the study concluded that among the seven estimators evaluated, PS\_EAP\_AVG and FK\_EAP generate the best alternatives to traditional CAT proficiency estimates.

**Keywords:** Computerized Adaptive Test. Item Response Theory. Kalman Filter. Dynamic Models. Bayesian Estimation.



# Lista de Figuras

2.1	Curva Característica do Item para respostas <i>corretas</i> ( $\beta_t = 0.5, \alpha_t = 1.0$ ) e <i>incorretas</i> ( $\beta_t = 0.5, \alpha_t = -1.0$ ) . . . . .	14
2.2	Curva Característica do Item - Ogiva Normal versus Ogiva Logística ( $\beta_t = 0.6, \alpha_t = 1.2$ ) . . . . .	16
2.3	Figura adaptada de (Andrade <i>et al.</i> , 2000, Pág. 11) . . . . .	19
5.1	<b>CAT EAP/3PL:</b> Trajetória dos $sd(\theta)$ . . . . .	87
5.2	<b>CAT EAP/3PL:</b> Estimativas de Proficiência . . . . .	88
5.3	<b>CAT EAP/3PL:</b> Trajetória das Proficiências Estimadas . . . . .	89
5.4	<i>Power Steady</i> - Comportamento dos parâmetros $\alpha$ e $\beta$ , quando $k = 0.85$ (Cenários 1 e 2). . . . .	92
5.5	<i>Power Steady</i> - Comportamento dos parâmetros $\alpha$ e $\beta$ , quando $k = 0.99$ (Cenários 1 e 2). . . . .	93
5.6	<b>Cenário 1:</b> <i>Power Steady</i> - Trajetória das estimativas de proficiência, quando $k = 0.81$ . . . . .	96
5.7	<b>Cenário 1:</b> <i>Power Steady</i> - Trajetória das estimativas de proficiência, quando $k = 0.90$ . . . . .	97
5.8	<b>Cenário 1:</b> <i>Power Steady</i> - Trajetória das estimativas de proficiência, quando $k = 0.99$ . . . . .	98
5.9	<b>Cenário 3:</b> <i>Power Steady</i> - Comportamento dos parâmetros $\alpha$ e $\beta$ , quando $\alpha_0 = \beta_0 = 7, k \in \{0.81, 0.85\}$ . . . . .	99
5.10	<b>Cenário 4:</b> <i>Power Steady</i> - Comportamento dos parâmetros $\alpha$ e $\beta$ , quando $\alpha_0 = \beta_0 = 50, k \in \{0.95, 0.99\}$ . . . . .	100
5.11	<b>Cenário 4:</b> <i>Power Steady</i> - Comportamento dos parâmetros $\alpha$ e $\beta$ , quando $\alpha_0 = \beta_0 = 50, k = 0.90$ . . . . .	101
5.12	<b>Cenário 3:</b> <i>Power Steady</i> - Estimativas de proficiência quando $\theta \in [-1, 1]$ , $k = 0.81$ . . . . .	102
5.13	<b>Cenário 3:</b> <i>Power Steady</i> - Estimativas de proficiência, quando $\theta \in [-1, 1]$ , $k = 0.85$ . . . . .	103
5.14	<b>Cenário 4:</b> <i>Power Steady</i> - Estimativas de proficiência, quando $\theta \in [-1, 1]$ , $k = 0.90$ . . . . .	104

5.15	<b>Cenário 4:</b> <i>Power Steady</i> - Estimativas de proficiência, quando $\theta \in [-1, 1]$ , $k = 0.95$ . . . . .	105
5.16	<b>Cenário 4:</b> <i>Power Steady</i> - Estimativas de proficiência quando $\theta \in [-1, 1]$ , $k = 0.99$ . . . . .	106
5.17	<b>Cenário 3:</b> <i>Power Steady</i> - Trajetória das estimativas de desvio-padrão de $\hat{\theta}$ , quando $k = 0.85$ . . . . .	107
5.18	<b>Cenário 4:</b> <i>Power Steady</i> - Trajetória das estimativas de desvio-padrão de $\hat{\theta}$ , quando $k = 0.99$ . . . . .	108
5.19	<b>Cenário 5:</b> <i>Power Steady</i> - Estimativas de proficiência quando $\theta \in [-3, 3]$ , $k = 0.85$ . . . . .	114
5.20	<b>Cenário 5:</b> <i>Power Steady</i> - Estimativas de proficiência quando $\theta \in [-3, 3]$ , $k = 0.90$ . . . . .	115
5.21	<b>Cenário 5:</b> <i>Power Steady</i> - Estimativas de proficiência quando $\theta \in [-3, 3]$ , $k = 0.99$ . . . . .	116
5.22	<b>Cenário 5b:</b> <i>Power Steady</i> - Trajetória dos estimadores $\hat{\theta}$ e $\tilde{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 0.5, \theta(\text{aleatório})$ . . . . .	119
5.23	<b>Cenário 5b:</b> <i>Power Steady</i> - Trajetória dos estimadores $\hat{\theta}$ e $\check{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 0.5, \theta(\text{aleatório})$ . . . . .	120
5.24	<b>Cenário 5:</b> <i>Power Steady</i> - Trajetória dos estimadores $\hat{\theta}$ e $\tilde{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 1, \theta(\text{fixo})$ . . . . .	123
5.25	<b>Cenário 5b:</b> <i>Power Steady</i> - Trajetória dos estimadores $\hat{\theta}$ e $\tilde{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 0.5, \theta(\text{aleatório})$ . . . . .	124
5.26	<b>Cenário 5:</b> <i>Power Steady</i> Trajetória dos estimadores $\hat{\theta}$ e $\tilde{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 1, \theta(\text{fixo})$ . . . . .	125
5.27	<b>Cenário 5:</b> <i>Power Steady</i> - Trajetória dos estimadores $\hat{\theta}$ e $\check{\theta}$ , quando $k = 0.99, \lambda = 1, \tau = 1, \theta(\text{fixo})$ . . . . .	126
5.28	<i>Power Steady</i> - Método EAP: Comparação entre $\theta$ vs $\hat{\theta}$ , quando $k = 0.99$ . . . . .	127
5.29	<i>Power Steady</i> - Método MAP: Comparação entre $\theta$ vs $\tilde{\theta}$ , quando $k = 0.99$ . . . . .	128
5.30	<i>Power Steady</i> - Método EAP: Comparação entre $\theta$ vs $\tilde{\theta}$ , quando $k = 0.99$ . . . . .	129
5.31	<i>Power Steady</i> - Método MAP: Comparação entre $\theta$ vs $\tilde{\theta}$ , quando $k = 0.99$ . . . . .	130
5.32	<i>Power Steady</i> - Método EAP: Comparação entre $\theta$ vs $\hat{\theta}$ , quando $k = 0.99$ . . . . .	131
5.33	<i>Power Steady</i> - Método MAP: Comparação entre $\theta$ vs $\check{\theta}$ , quando $k = 0.99$ . . . . .	132
5.34	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.99, \lambda = 1, \tau = 1, \theta > 0$ . . . . .	136
5.35	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.99, \lambda = 1, \tau = 1, \theta < 0$ . . . . .	137
5.36	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória dos $\alpha_t^*$ e $\beta_t^*$ , quando $\phi = 0.99, \lambda = 1, \tau = 1, \theta$ (fixo) . . . . .	138
5.37	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 1, \theta > 0$ . . . . .	140

5.38	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 1, \theta < 0$ . . . . .	141
5.39	<b>Cenário 5:</b> <i>Filtro de Kalman</i> - Trajetória dos $\alpha_t^*$ e $\beta_t^*$ , quando $\phi = 0.65, \lambda = 1, \tau = 1, \theta$ (fixo) . . . . .	142
5.40	<b>Cenário 5a:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.99, \lambda = 1, \tau = 0.5, \theta < 0$ . . . . .	145
5.41	<b>Cenário 5a:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.99, \lambda = 1, \tau = 0.5, \theta > 0$ . . . . .	146
5.42	<b>Cenário 5a:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta < 0$ . . . . .	147
5.43	<b>Cenário 5a:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta > 0$ . . . . .	148
5.44	<b>Cenário 5b:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta$ (Aleatório) . . . . .	149
5.45	<b>Cenário 5b:</b> <i>Filtro de Kalman</i> - Trajetória do estimador $\hat{\theta}$ , quando $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta$ (Aleatório) . . . . .	150
5.46	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.65$ . . . . .	151
5.47	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.70$ . . . . .	152
5.48	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.75$ . . . . .	153
5.49	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.80$ . . . . .	154
5.50	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.85$ . . . . .	155
5.51	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.90$ . . . . .	156
5.52	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.95$ . . . . .	157
5.53	<i>Filtro de Kalman:</i> Comparação entre $\theta$ vs $\hat{\theta}$ , quando $\phi = 0.99$ . . . . .	158





# Lista de Tabelas

2.1	Sumarização das principais diferenças entre TCT e TRI . . . . .	9
5.1	<i>Power Steady</i> - Comparações por RMSE (Cenários 1 e 2) . . . . .	94
5.2	<i>Power Steady</i> - Comparações por RMSE (Cenários 3 e 4) . . . . .	101
5.3	Estatísticas de caracterização para $\alpha_t$ , quando $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP - Cenário 5) . .	110
5.4	Estatísticas de caracterização para $\beta_t$ , quando $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP - Cenário 5) . .	110
5.5	Estatísticas de caracterização para $\alpha_t$ , quando $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP - Cenário 5) . .	111
5.6	Estatísticas de caracterização para $\beta_t$ , quando $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP - Cenário 5) . .	111
5.7	Estatísticas de caracterização para $\alpha_t$ , quando $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP - Cenário 5) . .	112
5.8	Estatísticas de caracterização para $\beta_t$ , quando $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP - Cenário 5) . .	112
5.9	Estatísticas de caracterização para $\alpha_t$ , quando $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP - Cenário 5) . .	113
5.10	Estatísticas de caracterização para $\beta_t$ , quando $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP - Cenário 5) . .	113
5.11	<i>Power Steady</i> - Comparações por RMSE, quando $k = 0.99$ . . . . .	121
5.12	Estatísticas de caracterização para $\alpha_t^*$ , quando $\phi = 0.99, \lambda = 1, \tau = 1$ , com 5 estimativas provisórias de $\theta_{jt}$ ( <i>Filtro de Kalman</i> - Cenário 5) . . . . .	135
5.13	Estatísticas de caracterização para $\beta_t^*$ , quando $\phi = 0.99, \lambda = 1, \tau = 1$ , com 5 estimativas provisórias de $\theta_{jt}$ ( <i>Filtro de Kalman</i> - Cenário 5) . . . . .	139
5.14	Estatísticas de caracterização para $\alpha_t^*$ , quando $\phi = 0.65, \lambda = 1, \tau = 1$ , com 5 estimativas provisórias de $\theta_{jt}$ ( <i>Filtro de Kalman</i> - Cenário 5) . . . . .	139
5.15	Estatísticas de caracterização para $\beta_t^*$ , quando $\phi = 0.65, \lambda = 1, \tau = 1$ , com 5 estimativas provisórias de $\theta_{jt}$ ( <i>Filtro de Kalman</i> - Cenário 5) . . . . .	143
5.16	<i>Filtro de Kalman</i> - Comparações por RMSE . . . . .	144

A.1	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	185
A.2	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	186
A.3	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	186
A.4	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	186
A.5	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	187
A.6	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP) . . . . .	187
A.7	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ , com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método EAP) . . . . .	188
A.8	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 1$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	188
A.9	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	188
A.10	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	189
A.11	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	189
A.12	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 1$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	189
A.13	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 1$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	190
A.14	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 2$ , $\tau = 1$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	190
A.15	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	190
A.16	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	191
A.17	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 2$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	191
A.18	Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	191
A.19	Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	192
A.20	<i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	192

A.21 Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	192
A.22 Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 1$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	193
A.23 <i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 3$ , $\tau = 1$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	193
A.24 Estatísticas de caracterização para $\alpha_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	193
A.25 Estatísticas de caracterização para $\beta_t^*$ , quando $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ ( <i>Power Steady</i> - Método MAP) . . . . .	194
A.26 <i>Power Steady</i> - RMSE para $k = 0.99$ , $\lambda = 3$ , $\tau = 0.5$ com 3 estimativas provisórias de $\theta_{jt}$ . . . . .	194



# Lista de Abreviaturas

<b>3PL</b>	Modelos Logísticos de três parâmetros
<b>2PL</b>	Modelos Logísticos de dois parâmetros
<b>1PL</b>	Modelos Logísticos de um parâmetro
<b>CAT</b>	Testes Adaptativos Computadorizados
<b>CBT</b>	Testes Baseados em Computador
<b>TRI</b>	Teoria da Resposta ao Item
<b>TCT</b>	Teoria Clássica dos Testes
<b>CCI</b>	Curva Característica do Item
<b>FRI</b>	Função de Resposta ao Item
<b>FII</b>	Função de Informação do Item
<b>FIT</b>	Função de Informação do Teste
<b>CMI</b>	Critério da Máxima Informação
<b>CMIG</b>	Critério da Máxima Informação Global
<b>CMVP</b>	Critério da Máxima Verossimilhança Ponderada
<b>CIP</b>	Critério de Informação Pontual
<b>CII</b>	Critério de Informação Intervalar
<b>IKL</b>	Informação de Kullback-Leibler
<b>EAP</b>	Estimação da Esperança a Posteriori
<b>MAP</b>	Estimação da Moda a Posteriori
<b>MLE</b>	Estimação por Máxima Verossimilhança
<b>MMLE</b>	Estimação por Máxima Verossimilhança Marginal
<b>MMLE/EM</b>	Estimação por Máxima Verossimilhança Marginal utilizando algoritmo EM
<b>JMLE</b>	Estimação de Máxima Verossimilhança Conjunta
<b>EMV</b>	Estimador de Máxima Verossimilhança
<b>WLE</b>	Estimador de Máxima Verossimilhança Ponderada
<b>BME</b>	Estimador Bayesiano Modal
<b>EP</b>	Erro Padrão de Estimação
<b>PSD</b>	Desvio-padrão a Posteriori
<b>VAP</b>	Variância a Posteriori

<b>DLM</b>	Modelo Linear Dinâmico
<b>DLMG</b>	Modelos Lineares Generalizados Dinâmicos
<b>RMSE</b>	Raiz Quadrada do Erro Quadrático Médio
<b>FKNG</b>	Filtro de Kalman Não Gaussiano
<b>FK</b>	Filtro de Kalman
<b>FK_EAP</b>	Estimador do Filtro de Kalman pelo Método da Esperança a Posteriori
<b>PS</b>	Power Steady
<b>PS_EAP</b>	Estimador do Power Steady pelo Método da Esperança a Posteriori
<b>PS_MAP</b>	Estimador do Power Steady pelo Método da Moda a Posteriori
<b>PS_EAP_AVG</b>	Estimador do Power Steady pelo Método da Esperança a Posteriori suavizado pela Média
<b>PS_MAP_AVG</b>	Estimador do Power Steady pelo Método da Moda a Posteriori suavizado pela Média
<b>PS_EAP_MEDIAN</b>	Estimador do Power Steady pelo Método da Esperança a Posteriori suavizado pela Mediana
<b>PS_MAP_MEDIAN</b>	Estimador do Power Steady pelo Método da Moda a Posteriori suavizado pela Mediana

# Lista de Símbolos

$\theta_j$	Proficiência verdadeira do $i$ -ésimo indivíduo de um teste
$\boldsymbol{\theta}$	Vetor de proficiências de um teste
$P_g(\boldsymbol{\theta})$	Função Característica do Item
$P(\beta_i, \alpha_i, \theta_j)$	Curva Característica do Item $i$ , para o $j$ -ésimo indivíduo
$\beta_i$	Deslocamento na escala de proficiência, também conhecido por parâmetro de dificuldade
$\alpha_i$	Inclinação na escala de proficiência, também conhecido por parâmetro de discriminação
$c_i$	Parâmetro de acerto casual de um teste
$D$	Fator de escala, sendo 1 se considerada a função logística, ou 1.702 se considerada a ogiva normal
$P(U_{ij} = 1 \theta_j)$	Função de Resposta ao Item
$U_{ij}$	Variável aleatória dicotômica contendo a resposta do $j$ -ésimo indivíduo ao $i$ -ésimo item, sendo 1 resposta correta e 0 resposta incorreta
$I(\theta_j)$	Informação de Fisher do $j$ -ésimo indivíduo
$I_U(\boldsymbol{\theta})$	Informação de Fisher do teste
$\mathbf{U}_j$	Vetor aleatório de respostas do $j$ -ésimo indivíduo
$\mathbf{U}..$	Conjunto integral de respostas de um teste
$\boldsymbol{\zeta}_i$	Vetor dos parâmetros do $i$ -ésimo item
$\boldsymbol{\zeta}$	Vetor dos parâmetros de todos os itens
$L(\boldsymbol{\zeta})$	Função de verossimilhança dos itens de um teste
$l(\boldsymbol{\zeta})$	Log-verossimilhança dos itens de um teste
$P(\theta_j \mathbf{u}_j, \boldsymbol{\nu}, \boldsymbol{\zeta})$	Distribuição a posteriori da proficiência do $j$ -ésimo indivíduo
$\mathbf{u}_j$	Vetor de respostas dos itens respondidos pelo $j$ -ésimo indivíduo
$\boldsymbol{\nu}$	Vetor de parâmetros associados à distribuição da proficiência da população
$g(\theta_j \boldsymbol{\nu})$	Distribuição a priori da proficiência do $j$ -ésimo indivíduo

$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EMV}$	Estimador de Máxima Verossimilhança para o $j$ -ésimo indivíduo do CAT
$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{MAP}$	Estimador da Moda a Posteriori para o $j$ -ésimo indivíduo do CAT
$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EAP}$	Estimador da Esperança a Posteriori para o $j$ -ésimo indivíduo do CAT
$p_{jt}(\theta_{jt})$	Probabilidade do $j$ -ésimo indivíduo, com proficiência $\theta_{jt}$ , acertar o $t$ -ésimo item do teste,
$\eta_t$	Parâmetro de reparametrização do desvio normalizado para o $j$ -ésimo indivíduo, considerado o $t$ -ésimo item
$f(\eta_t)$	Função de densidade da distribuição a priori utilizada na construção dos modelos dinâmicos
$f(\eta_t X_t)$	Função de densidade da distribuição a posteriori utilizada na construção dos modelos dinâmicos
$f(\eta_{t+1} X_t)$	Função de densidade da distribuição a priori durante o estágio de transição dos modelos dinâmicos
$f(\eta_{t+1} X_{t+1})$	Função de densidade da distribuição a posteriori após a disponibilização do $(t + 1)$ -ésimo item
$l(\phi)$	Função de Máxima Entropia para o modelo Filtro de Kalman
$\phi$	Fator de desconto utilizado no modelo Filtro de Kalman
$k$	Fator de desconto utilizado no modelo Power Steady
$\dot{\alpha}_t$ e $\dot{\beta}_t$	Hiperparâmetros da distribuição a priori utilizada pelo modelo Filtro de Kalman
$\gamma_t$ e $\delta_t$	Hiperparâmetros da distribuição a priori utilizada pelo modelo Power Steady
$\alpha_{t+1}$ e $\beta_{t+1}$	Parâmetros utilizados pelas Equações de Atualização dos modelos dinâmicos
$\alpha_{t+1}^*$ e $\beta_{t+1}^*$	Parâmetros utilizados pelas Equações de Transição dos modelos dinâmicos após a inclusão das constantes $\lambda$ e $\tau$ nas equações
$\lambda$	Constante introduzida nas equações de transição dos modelos dinâmicos para tentar suavizar oscilações nos parâmetros $\alpha_t$ e $\beta_t$
$\tau$	Constante introduzida nas equações de transição dos modelos dinâmicos para tentar reduzir a variância dos estimadores
$\tilde{\theta}_{jt}$	Estimador da Esperança a Posteriori (EAP), condicional aos parâmetros $a_t$ e $b_t$ dos itens, para a proficiência do $j$ -ésimo indivíduo, no $t$ -ésimo item do teste
$\hat{\theta}_{jt}$	Estimador da Esperança a Posteriori (EAP) do modelo Filtro de Kalman, para a proficiência do $j$ -ésimo indivíduo, no $t$ -ésimo item do teste



$\hat{\eta}_{j,t+1}$	Estimador da Esperança a Posteriori (EAP) do desvio normalizado da proficiência do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$Var(\hat{\eta}_{j,t+1})$	Variância do estimador da Esperança a Posteriori (EAP) do desvio normalizado da proficiência do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$\hat{\theta}_{j,t+1}$	Estimador da Esperança a Posteriori (EAP) da proficiência do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$s.d.(\hat{\theta}_{j,t+1})$	Desvio-padrão do estimador da Esperança a Posteriori (EAP) da proficiência do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$\tilde{\eta}_{j,t}$	Estimador da Moda a Posteriori (MAP) do desvio normalizado da proficiência do $j$ -ésimo indivíduo considerado o $t$ -ésimo item
$Var(\tilde{\eta}_{j,t})$	Variância do estimador da Moda a Posteriori (MAP) do desvio normalizado da proficiência do $j$ -ésimo indivíduo considerado o $t$ -ésimo item
$\tilde{\theta}_{j,t+1}$	Estimador da Moda a Posteriori (MAP) da proficiência do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$s.d.(\tilde{\theta}_{j,t+1})$	Desvio-padrão do estimador da Moda a Posteriori da proficiência (MAP) do $j$ -ésimo indivíduo considerado o $(t + 1)$ -ésimo item
$\bar{\tilde{\theta}}_{jt}$	Suavização pela média, do estimador da Moda a Posteriori (MAP) para a proficiência do $j$ -ésimo indivíduo consideradas todas as proficiências avaliadas até o $t$ -ésimo item
$\dot{\tilde{\theta}}_{jt}$	Suavização pela mediana, do estimador da Moda a Posteriori (MAP) para a proficiência do $j$ -ésimo indivíduo consideradas todas as proficiências avaliadas até o $t$ -ésimo item
$\bar{\hat{\theta}}_{jt}$	Suavização pela média, do estimador da Esperança a Posteriori (EAP) para a proficiência do $j$ -ésimo indivíduo consideradas todas as proficiências avaliadas até o $t$ -ésimo item
$\dot{\hat{\theta}}_{jt}$	Suavização pela mediana, do estimador da Esperança a Posteriori (EAP) para a proficiência do $j$ -ésimo indivíduo consideradas todas as proficiências avaliadas até o $t$ -ésimo item



# Sumário

<b>Lista de Figuras</b>	<b>xiii</b>
<b>Lista de Tabelas</b>	<b>xvii</b>
<b>Lista de Abreviaturas</b>	<b>xxi</b>
<b>Lista de Símbolos</b>	<b>xxiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Avanços na Avaliação Educacional . . . . .	1
1.2 Tema de Estudo . . . . .	3
1.3 Objetivos . . . . .	4
1.4 Organização da Dissertação . . . . .	4
<b>2 Teoria da Resposta ao Item (TRI)</b>	<b>7</b>
2.1 Teoria Clássica dos Testes (TCT) <i>versus</i> Teoria da Resposta ao Item (TRI) .	7
2.2 Conceitos e Pressupostos . . . . .	9
2.2.1 Unidimensionalidade . . . . .	9
2.2.2 Independência Local . . . . .	11
2.2.3 Traço Latente e Curva Característica do Item . . . . .	12
2.2.3.1 O Traço Latente . . . . .	12
2.2.3.2 A Curva Característica do Item (CCI) . . . . .	13
2.2.3.3 O Modelo de Ogiva Normal . . . . .	15
2.2.3.4 O Modelo de Ogiva Logística . . . . .	16
2.2.4 Escala de Habilidade . . . . .	17
2.3 Estimação em TRI . . . . .	18
2.3.1 Função de Informação do Teste (FIT) . . . . .	20
2.3.2 Estimação dos Parâmetros dos Itens . . . . .	22
2.3.2.1 Estimação de Máxima Verossimilhança (MLE) . . . . .	23
2.3.2.2 Estimação de Máxima Verossimilhança Conjunta (JMLE) .	25
2.3.2.3 Estimação de Máxima Verossimilhança Marginal (MMLE) .	28
2.3.2.4 A Solução de Bock & Lieberman . . . . .	30
2.3.2.5 Método da Quadratura . . . . .	32

2.3.3	Estimação Bayesiana . . . . .	33
2.3.3.1	Estimação Bayesiana de Parâmetros dos Itens . . . . .	34
2.3.3.2	Estimação Bayesiana Marginalizada . . . . .	35
2.3.3.3	O Componente da Verossimilhança . . . . .	37
2.3.3.4	A Componente da Distribuição à Priori . . . . .	38
2.3.4	Estimação Bayesiana via Algoritmo EM . . . . .	39
2.3.4.1	Algoritmo EM . . . . .	39
2.3.4.2	Estimação Máxima a Posteriori (MAP) . . . . .	41
2.3.4.3	Estimação da Esperança a Posteriori (EAP) . . . . .	42
<b>3</b>	<b>Teste Adaptativo Computadorizado (CAT)</b>	<b>45</b>
3.1	Componentes do CAT . . . . .	46
3.2	Métodos de Seleção de Itens . . . . .	49
3.2.1	Crítério da Máxima Informação (CMI) . . . . .	49
3.2.2	Aproximação de Owen . . . . .	50
3.2.3	Crítério da Máxima Informação Global (CMIG) . . . . .	51
3.2.4	Crítério da Máxima Verossimilhança Ponderada (CMVP) . . . . .	52
<b>4</b>	<b>CAT via Filtro de Kalman Não Linear</b>	<b>55</b>
4.1	O Modelo Dinâmico Gaussiano . . . . .	55
4.2	O Filtro de Kalman Não Linear (FKNG) . . . . .	57
4.3	Modelos Dinâmicos para o CAT . . . . .	58
4.4	Pressupostos diversos . . . . .	59
4.5	Estimação de proficiência via Filtro de Kalman com $c_t$ não nulo . . . . .	60
4.5.1	Distribuição a Priori . . . . .	64
4.5.2	Estimação de $\eta$ pelo Método EAP . . . . .	65
4.5.3	Elicitação da priori $BetaExp(\alpha_t, \beta_t)$ de $\eta_t$ . . . . .	66
4.5.4	Distribuição a Posteriori . . . . .	67
4.5.5	Distribuição Preditiva . . . . .	68
4.5.6	Um modelo dinâmico para o CAT via Filtro de Kalman . . . . .	69
4.6	Estimação de proficiência via Modelo Dinâmico Power Steady com $c_t$ nulo . . . . .	74
4.6.1	Distribuição a Priori de $\eta_{jt}$ . . . . .	75
4.6.2	Elicitação da priori $BetaExp(\gamma_t, \delta_t)$ de $\eta_{jt}$ . . . . .	76
4.6.3	Um Modelo <i>Power Steady</i> para o CAT . . . . .	79
4.6.4	Estimação de $\eta$ pelo Método EAP . . . . .	82
4.6.5	Estimação de $\eta$ pelo Método MAP . . . . .	82
<b>5</b>	<b>Simulações e Resultados</b>	<b>85</b>
5.1	Simulações com o Modelo <i>Power Steady</i> . . . . .	85
5.1.1	Avaliando o CAT tradicional . . . . .	87
5.1.2	Avaliando o comportamento de $\alpha_t$ e $\beta_t$ do modelo <i>Power Steady</i> . . . . .	90

5.1.3	Cenários 1 e 2 . . . . .	90
5.1.4	Cenários 3 e 4 . . . . .	95
5.1.5	Cenário 5 . . . . .	108
5.1.6	Obtenção de valores iniciais de $\alpha_{t t}$ e $\beta_{t t}$ . . . . .	108
5.1.7	Obtenção de valores iniciais de $\alpha_t^*$ e $\beta_t^*$ . . . . .	118
5.2	Simulações com o Modelo <i>Filtro de Kalman</i> . . . . .	134
5.2.1	Avaliando o comportamento de $\alpha_t$ $\beta_t$ . . . . .	134
<b>6</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>159</b>
6.1	Conclusões . . . . .	159
6.2	Trabalhos Futuros . . . . .	161
	<b>Referências Bibliográficas</b>	<b>163</b>
	<b>Apêndices</b>	<b>171</b>
	<b>Apêndice Teoria da Resposta ao Item</b>	<b>173</b>
A	Solução de Bock & Lieberman . . . . .	173
B	Solução de Bock & Aitkin . . . . .	175
	<b>Apêndice CAT via Filtro de Kalman Não-Linear</b>	<b>179</b>
C	Memória de Cálculo . . . . .	179
	<b>Apêndice Simulações e Resultados</b>	<b>185</b>
D	Cenário 5a: $\theta$ fixo, $\lambda = 1$ , $\tau = 0.5$ . . . . .	185
E	Cenário 5b: $\theta$ aleatório, $\lambda = 1$ , $\tau = 0.5$ . . . . .	186
F	Cenário 5c: $\theta$ fixo, $\lambda = 2$ , $\tau = 0.5$ . . . . .	188
G	Cenário 5d: $\theta$ fixo, $\lambda = 2$ , $\tau = 1$ . . . . .	189
H	Cenário 5e: $\theta$ aleatório, $\lambda = 2$ , $\tau = 0.5$ . . . . .	190
I	Cenário 5f: $\theta$ fixo, $\lambda = 3$ , $\tau = 0.5$ . . . . .	191
J	Cenário 5g: $\theta$ fixo, $\lambda = 3$ , $\tau = 1$ . . . . .	192
K	Cenário 5h: $\theta$ aleatório, $\lambda = 3$ , $\tau = 0.5$ . . . . .	193



# Capítulo 1

## Introdução

Nos dias atuais o uso de computadores e dispositivos eletrônicos tornou-se indispensável. No campo das avaliações educacionais, a tecnologia tem sido um instrumento de constante expansão e inovação. Em particular, o teste adaptativo computadorizado, *Computerized Adaptive Test* (CAT) em inglês, surgido nos anos 80 (Larson e Madsen, 1985), é um tipo de teste baseado em recursos computacionais, que se adapta ao nível de habilidade dos examinandos. Ancorado no uso das técnicas de teoria da resposta ao item, *Item Response Theory* (IRT) em inglês (Lord, 1980), sua principal característica é a inovação proporcionada pela rápida adaptação de itens extraídos de um banco de itens, aos examinandos de testes aplicados online.

Segundo Magis e Barrada (2017), no campo da psicometria, o teste adaptativo computadorizado (CAT) é uma importante área de pesquisas atuais, cuja implementação prática tem crescido expressivamente na última década. No campo da avaliação educacional, permite oferecer ao examinando, de maneira automatizada, itens que sejam compatíveis com sua proficiência, de modo que a massificação de processos avaliativos, juntamente com o aumento dos custos financeiros para a aplicação de testes tradicionais, corroboram para sua aplicação como alternativa promissora, tanto na redução de custos, quanto na rápida e adaptativa oferta de itens, àqueles que se submetem à avaliação.

Essencialmente, uma implementação do CAT requer dois elementos fundamentais: um banco de itens calibrado e um algoritmo para seleção adaptativa dos itens. Este trabalho tem foco no segundo elemento, para o qual é apresentada uma adaptação à formulação não linear dos filtros de Kalman, prescrita por Harvey (1989), combinada com o conhecido princípio da máxima entropia (Kullback, 1997).

### 1.1 Avanços na Avaliação Educacional

Com o rápido crescimento das tecnologias computacionais, assim como dos dispositivos móveis, torna-se cada vez mais desafiadora a tarefa de integrar testes baseados em computador - também conhecidos em inglês por *Computer Based Testing* (CBT) <sup>1</sup> - com o aprendizado via dispositivos móveis, especialmente para avaliações formativas e mesmo para auto-avaliações. Em termos das auto-avaliações, o CAT tem sido disponibilizado como meio ideal para tal realização. Sua natureza adaptativa, conjugado com os conceitos e aplicações, já bem conhecidos e bastante difundidos da metodologia da TRI, tem sido a principal razão para cada vez mais instituições de reconhecimento internacional adotarem seu uso.

Além do mais, um grande banco de itens é indispensável para a realização de um teste de

---

<sup>1</sup>Conceitos e aplicações podem ser encontrados em Mills *et al.* (2005)

excelência, mas quando um sistema de testes, tais como CAT, têm um grande banco de itens, a seleção de um item de teste pela TRI torna-se muito tediosa, se aplicado de forma não sistematizada e adaptativa. Além do grande banco de itens, o mecanismo de exposição de itens é essencial ao sistema de testes. Embora a metodologia empregada na TRI proponha formas de representar a relação entre a probabilidade de um indivíduo dar uma resposta certa a um item e seus traços latentes, isoladamente não possui um mecanismo de seleção de itens que sejam mais adequados às habilidades de examinandos, sendo portanto recurso meio para obtenção destes.

A relação entre CAT e TRI baseou-se na dependência direta de um pelo outro, pois o primeiro necessitava dos métodos empregados pelo segundo para sistematicamente estimar o melhor item baseado em habilidades latentes conhecidas. De maneira resumida, pode-se dizer que os papéis de ambos se complementam, de forma sequenciada. Enquanto a TRI nos últimos anos evoluiu quanto à especificação de modelos para a calibração dos parâmetros dos itens, abrangendo formas cada vez mais complexas de tratar os contextos de dependência destes, o CAT evoluiu no processamento da seleção destes itens, já calibrados, adotando algoritmos cada vez mais sofisticados para rapidamente associar um item dentre todos os disponíveis no banco de itens calibrados à habilidade do examinando em tempo real. Embora pareça um procedimento fácil para os avanços computacionais atuais, questões de segurança foram cuidadosamente observadas no processo de evolução da metodologia, desde os primeiros testes com o uso CAT.

Segundo [van der Linden e Glas \(2000\)](#), a capacidade dos examinandos para memorizar itens de teste, bem como a sua tendência de compartilhá-los com os futuros candidatos a estudos, pareceu ser muito maior do que o previsto inicialmente. Como consequência, surgiu a necessidade de métodos eficazes para controlar a exposição dos itens, bem como para detectar itens comprometidos. Além disso, a questão de como alinhar o conteúdo do teste com as especificações do teste e o conteúdo de equilíbrio entre as administrações de testes pareceu ser mais complicado do que o previsto. Esta questão levou a uma busca por novos algoritmos de teste, bem como a introdução de uma variedade de novas formas, como por exemplo, o teste adaptativo baseado em *testlets*.

Um *Testlet* ([Wainer e Kiely, 1987](#)) é definido como um agrupamento de itens que é baseado em um único estímulo, tal como em um teste de compreensão de leitura. Na prática, quando se utilizam modelos de resposta de item, é sempre assumido que as respostas a todos os itens são independentes umas das outras, sob o condicionamento da proficiência subjacente e latente do examinando. A experiência tem mostrado que, quando os testes são compostos por itens separados e não relacionados, essa suposição de independência condicional (CI) é suficiente o bastante para permitir que esses modelos sejam de grande utilidade prática. Há, no entanto, algumas circunstâncias razoavelmente comuns em que a hipótese de CI não é provável que seja verdade. A mais frequente de tais circunstâncias observadas é quando o teste é construído por *testlets*.

Um modelo proposto por [Wang et al. \(2002\)](#), forneceu um mecanismo para gerar escores de um teste que fosse composto por qualquer combinação ou agrupamento de itens de múltipla escolha, itens de preenchimento em branco e outros itens de resposta construídos que fossem julgados por avaliadores especialistas. Neste caso, um *testlet* poderia ser definido como a passagem e o conjunto de 4 para 12 itens que acompanham a passagem. Não é difícil imaginar, portanto, que questões como a interpretação errada da passagem, o conhecimento da matéria, a fadiga, e assim por diante, fariam com que as respostas a esses itens fossem mais altamente relacionadas do que o sugerido pela proficiência latente geral para o teste completo. Em algum sentido, essa falta de CI é uma forma de desajuste do modelo de proficiência unidimensional, o que deve ser explicado pela estrutura de teste (isto é, o planejamento



do *testlet*). Os autores concluíram que desenvolvedores de teste podiam se concentrar em qual estrutura de teste mede melhor a construção de interesse e confiar que em qualquer circunstância haverá um modelo de pontuação apropriado disponível.

## 1.2 Tema de Estudo

Segundo [Wainer \(2000\)](#), a noção básica de um teste adaptativo é imitar automaticamente o que um sábio examinador faria, de forma que o CAT tem por finalidade administrar itens, de um banco de itens previamente calibrados, que correspondam ao nível de proficiência do examinando. É assumido que tais itens descrevam o comportamento da resposta do indivíduo, sendo seus parâmetros estimados através de um dos modelos de Teoria da Resposta ao Item (TRI).

Considerando que a TRI é uma metodologia usada em avaliações educacionais desde os anos 50 ([Gulliksen, 1950](#)), o modelo padrão de medidas educacionais foi concebido unicamente sob a tendência dos examinandos em responder itens corretamente, isto é, em sua proficiência global.

De alguns anos para cá, outros modelos começaram a aparecer, com foco em outros padrões diferentemente da proficiência global. Como exemplo, [Mislevy e Verhelst \(1990\)](#) apresentam um modelo para respostas de itens quando indivíduos diferentes empregam diferentes estratégias, mas apenas as respostas, e não a escolha da estratégia, podem ser observadas.

Em argumento semelhante, no presente estudo os itens respondidos por dado examinando descrevem uma série temporal binária, em que  $X_t = 1$  para resposta correta e  $X_t = 0$  para resposta incorreta. Admite-se que a proficiência  $\theta$  do examinando, embora estática no intervalo de tempo em estudo do CAT, também descreve uma série temporal. Porém, devido ao aspecto iterativo do processo de estimação de  $\theta$ , é não observável.

Na estimação da proficiência  $\theta$  de dado examinando, a metodologia em estudo é baseada em um modelo dinâmico em que se explora a reparametrização da probabilidade de um item ser corretamente respondido, em particular para modelos unidimensionais de 3 parâmetros, definindo equações de transição de estados por meio do hiperparâmetro do Filtro de Kalman, criando assim, um sistema recursivo onde a distribuição à posteriori é avaliada no intuito de otimizar a convergência da estimação da proficiência de um dado examinando avaliado via CAT.

O objeto de estudo resume-se na proposta de uma metodologia alternativa na estimação da proficiência individual no processo de CAT, sendo esta alternativa baseada em princípios utilizados em filtros de Kalman. Assim, os principais focos de estudos concentram-se em:

- na aplicação dos métodos de TRI e CAT para modelos unidimensionais de 3 parâmetros com resposta dicotômica;
- na proposição de metodologia inédita, baseada em Filtro de Kalman, na estimação de proficiências via CAT;
- nas simulações de diferentes cenários para a busca da otimização do tempo de estimação das proficiências e correspondente avaliação da incerteza envolvida nesse processo;
- na comparação da metodologia proposta com metodologias já bastante disseminadas na área de CAT para a estimação de proficiências.

## 1.3 Objetivos

Neste trabalho, pretende-se apresentar um novo algoritmo para estimação de proficiências via CAT. A comparação da performance desta metodologia com as metodologias correntes possibilitará avaliar se, através desta nova proposta, é possível diminuir o número de questões a serem apresentadas ao examinando, de modo a tornar o processo de avaliação mais rápido e eficiente e, como consequência direta, a menor exposição dos itens contidos no banco de questões.

O principal objetivo concentra-se, portanto, na apresentação de uma nova forma de estimação que conduza a um ganho de performance na aplicação dos testes adaptativos, possibilitando novos resultados que abram novas frentes de pesquisas e melhorias no ramo da avaliação educacional.

## 1.4 Organização da Dissertação

Este trabalho pretende apresentar e discutir, de forma sistematizada, os principais aspectos técnicos utilizados ao longo da construção do mesmo. De forma que, ao longo do percurso da escrita são utilizados recursos computacionais para propiciar uma melhor organização e consequente otimização da construção do texto. Tais recursos, embora nem sempre totalmente gratuitos, estão disponíveis na internet e podem ser utilizados por quaisquer internautas para desenvolvimento ágil, a saber:

- **Evernote:** ferramenta utilizada para capturar, organizar e compartilhar anotações online (<https://evernote.com>).
- **Google Scholar:** ferramenta de busca utilizada para consulta de artigos e referências acadêmicas, sendo de valiosa importância na padronização destas no estilo L<sup>A</sup>T<sub>E</sub>X (<https://scholar.google.com.br/>).
- **GitHub:** repositório com integração ao RStudio, para armazenamento e versionamento de códigos (<https://github.com>).

Desta maneira, o texto está dividido em seis capítulos, sob a seguinte estrutura: o Capítulo 1 faz uma breve apresentação geral do assunto tratado e de alguns aspectos que norteiam o tema da dissertação. No Capítulo 2 a *Teoria da Resposta ao Item* é abordada de maneira que os principais tópicos pertinentes ao tema do trabalho são explicados em detalhes, a saber: métodos de estimação dos parâmetros, modelos de 1, 2 e 3 parâmetros, escalas de medida dos parâmetros, interpretação dos parâmetros, interpretação das curvas características do item e aspectos importantes sobre o tamanho de amostras na estimação de parâmetros para diferentes tipos de modelos.

O Capítulo 3 apresenta o *Teste Adaptativo Computadorizado* e seus aspectos relevantes, a saber: histórico evolutivo do CAT até os dias atuais, estimação dos parâmetros e seus trade-offs, vantagens e desvantagens do uso do CAT, principais avaliações conhecidas, implementações computacionais, pacotes conhecidos e implementações atualizadas.

O Capítulo 4 trata da *formulação matemática do problema* a ser tratado pela dissertação, apresentando as principais ideias e fundamentos necessários para concepção do *estado da arte* do método e seus principais impactos.

O Capítulo 5 apresenta os principais resultados obtidos com a estimação das proficiências utilizando os modelos *Filtro de Kalman* e *Power Steady*, e faz uma breve discussão sobre os aspectos positivos e dificuldades encontradas na utilização de cada um dos métodos.

O Capítulo 6 apresenta as considerações finais sobre o trabalho realizado, fazendo um balanço dos pontos mais relevantes encontrados com as simulações, fazendo um resumo dos questionamentos levantados durante toda a execução da dissertação e apresentando sugestões para trabalhos futuros.



## Capítulo 2

# Teoria da Resposta ao Item (TRI)

Em muitas situações de avaliação educacional, o objeto de estudo é uma variável subjacente, ou seja, uma medida que não necessariamente se tenha informação. Em geral, essa variável é algo que é intuitivamente entendido como “inteligência”, ou algum outro conceito não diretamente mensurável. Quando as pessoas são descritas como brilhantes ou medianas, não se tem uma idéia concreta deste conceito, uma vez que para cada um a “escala” de medida para definir tal conceito pode variar substancialmente. Da mesma forma, pode-se falar sobre a habilidade escolar e seus atributos, como obter boas notas, aprender um novo tema de forma autodidata, relacionar várias fontes de informação e usar o tempo de estudo de forma eficaz.

Nas áreas acadêmicas, é comum se usar termos descritivos, tais como habilidade de leitura, capacidade aritmética, domínio de um certo tema, entre outros, para exemplificar níveis de conhecimento. Cada um destes níveis é o que os psicometristas chamam de traço não observável ou latente. Embora tal variável seja facilmente descrita, e pessoas experientes possam listar seus atributos ou características, não há uma maneira única de ser mensurada, uma vez que, representa um conceito e não uma dimensão física ou matemática. Como o objetivo básico na avaliação educacional tem foco na determinação de quanta característica latente uma pessoa possui, as pesquisas científicas que tratam deste assunto se basearam nos métodos de estimação de variáveis, tais como, habilidades escolares, de leitura, matemáticas e aritméticas, etc. Neste sentido, a *Teoria Clássica dos Testes* (TCT) e a *Teoria da Resposta ao Item* (TRI) compreendem os dois métodos estatísticos tradicionais de planejamento e análise de tais habilidades. O termo genérico “proficiência” é usado na Teoria da Resposta do Item para se referir a essa habilidade como *traço latente*.

### 2.1 Teoria Clássica dos Testes (TCT) versus Teoria da Resposta ao Item (TRI)

Em comparação à teoria clássica dos testes (TCT), a teoria da resposta ao item (TRI) é considerada como o método padrão, senão o preferido para realizar avaliações psicométricas de novas e medidas já estabelecidas (Embretson e Reise, 2000). Entretanto, embora a TRI tenha evoluído nos últimos 50 anos, a TCT ainda tem sido pesquisada e aplicada continuamente. Muitos programas de teses ainda implementam TCT em seus planejamentos e avaliação de resultados. Isto se deve a algumas vantagens da TCT em relação à TRI. Por exemplo, a TCT descreve a relação entre o escore verdadeiro e o escore observado de maneira linear, fazendo com que os modelos de TCT sejam de fácil entendimento e aplicação para muitos pesquisadores. Isto se baseia inteiramente no total de escores ou nos escores do

número de respostas corretas. Um escore observado de um examinando é definido como o escore total obtido por cada examinando e é diferente do escore verdadeiro por um escore de erro comum. Esta metodologia de pontuação gerou um número de vantagens assim como um número de limitações. Além dos resultados sintetizados na Tabela 2.1, abaixo são listadas algumas das vantagens do uso da TCT em relação à TRI:

- A análise requer tamanhos amostrais menores que os requeridos pela TRI.
- O procedimento matemático é muito mais simples se comparado ao da TRI, uma vez que os modelos da TCT são lineares enquanto da TRI não-lineares.
- A estimação de parâmetros é conceitualmente direta e requer um mínimo de suposições, fazendo os modelos úteis e largamente aplicáveis.
- As análises não requerem rigorosos estudos de bons ajustes como na TRI.

Entretanto, segundo Le (2013), a TCT apresenta um número maior de desvantagens em relação à TRI. A pedra angular de muitas análises da TCT são as características dos itens de teste, dificuldade e confiabilidade. Esses índices são medidos pela proporção do item,  $p$ , dos examinandos que respondem corretamente às questões, e da correlação total do item,  $r$ . No entanto, os índices não são constantes, pois são totalmente dependentes da amostra de examinandos de onde são obtidos. Eles não podem ser usados para indicar a característica ou a qualidade de um teste. Outra desvantagem é que as pontuações dos examinandos são testáveis. Ou seja, os examinandos podem obter pontuações mais altas em um teste mais fácil e pontuações menores em um teste mais difícil e, portanto, nenhum resultado verdadeiro pode ser extraído. Isso não permite uma base sólida capaz de combinar itens de teste e níveis de habilidade.

Nesse sentido, Le (2013) argumenta que a TRI agrega grandes benefícios em relação à TCT. Na estrutura da TRI, as características do item são independentes da amostra e as pontuações latentes de uma pessoa são independentes do teste, desde que os modelos selecionados se encaixem bem aos dados. Assim, as pontuações que descrevem a proficiência do examinando não dependem da dificuldade do teste. Suas pontuações podem ser menores em testes mais difíceis e maiores em testes mais fáceis, mas suas pontuações de habilidade permanecem constantes em relação a qualquer teste no momento do teste ou no levantamento do mesmo.

A TRI também permite o cálculo da probabilidade de um respondente específico selecionar uma categoria em um item de teste. Os itens de teste independentes da amostra facilitam o desenho do CAT, o que permite uma comparação ou identificação mais precisa dos examinandos. Além disso, a TRI pode ser usada para refinamento ou desenvolvimento de escala, pois é capaz de calcular erros padrão e, portanto, fornece informações sobre a qualidade de cada item. Isso ajuda a tomar decisões na seleção de itens para excluir ou incluir em um instrumento de teste ou pesquisa. Adicionalmente, os itens também são selecionados com base em seus índices de dificuldade e discriminação, ou seja, sua capacidade de discriminar grupos de traços latentes altos e baixos.

Apesar dessas vantagens, os modelos de TRI também têm suas deficiências. Do ponto de vista técnico, os modelos são mais complexos e os métodos de estimação de parâmetros geralmente envolvem métodos numéricos complicados. Os traços latentes, bem como os parâmetros do item também podem ser difíceis de interpretar de forma gráfica e numérica. Os modelos Rasch (modelos de um parâmetro) são mais diretos para se aplicar do que outros modelos. No entanto, devido às restrições impostas pelos pressupostos do modelo, podem surgir problemas com o ajuste dos modelos Rasch ao testar os dados. Consequentemente,

apesar de algumas vantagens da TRI sobre a TCT, o aumento da popularidade da TRI e o crescente desenvolvimento de pacotes computacionais, faz com que a metodologia apresente uma série de desafios únicos. Se por um lado, a especificação de modelos em TRI é mais complicada em relação à TCT, por outro, a interpretação adequada dos resultados exige atenção cuidadosa a várias características únicas das saídas gráficas e numéricas da TRI, especialmente para modelos mais desafiadores. Por fim, o processamento de dados e a preparação para uso de um software de TRI podem ser mais complexos e demorados.

**Tabela 2.1:** *Sumarização das principais diferenças entre TCT e TRI*

	TCT	TRI
Dependência Populacional	As estatísticas dos itens são dependentes da população	Os parâmetros do item são dependentes da população
Dependência de itens	Escore do examinando é dependente do teste	Estimativas das habilidades do examinando são dependentes dos itens
Confiabilidade (erro-padrão de medição)	Todos os escores tem a mesma confiabilidade	Cada valor de habilidade tem sua própria confiabilidade
Suposições	Unidimensionalidade	Independência Local
Dificuldade do item	Todos os itens tem a mesma contribuição para o escore total, independente da dificuldade, se não for ponderada	A dificuldade dos itens contribuem para a estimação da habilidade do examinando
Discriminação do item	A discriminação do item não é considerada no escore do examinando	A discriminação do item contribui para estimação da habilidade do examinando, a menos que o modelo se mantenha constante
Escala de comparabilidade	A dificuldade do item e o escore do examinando estão em escalas diferentes escalas	A dificuldade do item e o escore do examinando estão na mesma escala

Fonte: Zheng (2014).

## 2.2 Conceitos e Pressupostos

### 2.2.1 Unidimensionalidade

A unidimensionalidade é um dos pressupostos da maioria dos modelos da *Teoria de Resposta ao Item* (Fernández, 1990, as cited in Vitória *et al.*, 2006), o que confere, inicialmente, atualidade ao seu estudo e aprofundamento. Contudo, foi desde sempre, considerada desejável num teste, mesmo na *Teoria Clássica dos Testes* (TCT), pois os conceitos de homogeneidade e dificuldade só fazem sentido quando se assume avaliar um único atributo. Por um lado, tem-se a identidade do próprio teste (*face validity*) e, por outro, quando se assume obter pontuações globais num teste, o que significa medir a mesma coisa ao longo dos vários itens, de forma que, o teste em questão deverá ser unidimensional (Cuesta, 1996, as cited in Vitória *et al.*, 2006).

Segundo Hambleton e Swaminathan (1985), a unidimensionalidade não pode ser completamente conseguida e apontam alguns fatores que podem atuar concomitantemente com o traço em avaliação. Na mesma linha, Pasquali (2009) defende que, para satisfazer o postulado de unidimensionalidade é suficiente admitir que haja um fator dominante. Assim, o desempenho num teste pode ser influenciado por outras variáveis cognitivas, pela personalidade do indivíduo ou, mesmo, por fatores inerentes à própria aplicação do teste, nomeadamente a motivação, a ansiedade, a capacidade para trabalhar em velocidade ou o uso correcto das folhas de respostas, entre outros. Portanto, a questão principal não é avaliar se um teste é estritamente unidimensional mas sim em que medida as dimensões adicionais geram distorções na medida principal do fator dominante que os itens estão mais ou menos relacionados (Wright e Stone, 2004).

Não há uma definição unânime de unidimensionalidade, dado que a dificuldade de se obter uma definição universal prende-se ao fato de que a maioria dos autores a define em função da forma como será avaliada. O conceito de unidimensionalidade tem surgido, ainda, muito associado ao de independência local e em algumas ocasiões tomados como sinônimos (Cuesta, 1996, as cited in Vitória *et al.*, 2006). Na opinião de Hattie (1984), apesar do conceito de independência local ser essencial para a definição de traço latente e de unidimensionalidade, não podem ser tomados como sinônimos. A independência local pode ser conseguida com 1, 2, . . . ,  $k$  traços latentes ou dimensões.

Uma das abordagens conceituais de unidimensionalidade está baseada no *padrão de respostas*. Dentro desta abordagem, os itens colocados por ordem crescente de dificuldade dariam lugar a padrões de resposta tais que, o indivíduo que acertasse o item 3, acertaria também o item 1 e 2, e assim sucessivamente. Ou seja, se um indivíduo possuísse uma quantidade  $X$  do traço medido, então, acertaria todos os itens cuja dificuldade fosse menor que  $X$ . Uma prova que se afastasse deste padrão não seria unidimensional (Cuesta, 1996, as cited in Vitória *et al.*, 2006).

Na abordagem conceitual de dimensionalidade baseada nas *teorias do traço latente*, assume-se que  $k$  traços latentes explicam a realização dos indivíduos, de tal forma que, para um nível fixo de capacidade, as respostas dos indivíduos aos itens seriam estatisticamente independentes (quer dizer, cumprem a condição de independência local). Quando um único traço explica a realização dos indivíduos, então esse conjunto de itens considera-se unidimensional (Hambleton e Swaminathan, 1985, as cited in Vitória *et al.*, 2006).

Segundo Vitória *et al.* (2006), se num espaço de  $k$  dimensões um indivíduo for representado por um ponto no espaço e se cada uma das dimensões interfere na realização dos indivíduos em, pelo menos, 2 itens do teste, então  $k$  é a dimensão desse espaço latente (Cuesta, 1996, as cited in Vitória *et al.*, 2006). Esta definição de dimensionalidade liga-se com a independência local através da característica de independência estatística entre as respostas dos indivíduos, que se dá quando se especificam todas as dimensões do espaço latente que explicam as respostas dos indivíduos (Cuesta, 1996, as cited in Vitória *et al.*, 2006).

Formalmente,

$$P(X = x|\theta) = \prod_{t=1}^n P(X_t = x_t|\theta) \quad (2.1)$$

Isto significa que a distribuição conjunta das respostas a um conjunto de itens  $X = x_1, x_2, \dots, x_n$ , num dado vetor de traços latentes,  $\theta = \theta_1, \theta_2, \dots, \theta_k$ , é igual ao produto das distribuições marginais dos itens dado  $\theta$ , ou seja, são estatisticamente independentes. Quando  $k = 1$  trata-se de um espaço unidimensional (Cuesta, 1996, as cited in Vitória *et al.*, 2006). Apesar de baseada na independência local, não são aqui tomadas como sinônimos, uma vez que, como afirmado antes, a independência local pode ser conseguida com  $k$  dimensões.

Para Cuesta (1996), a exigência de unidimensionalidade feita pela TRI não coincide com a forma como esta é, usualmente, entendida. De fato, “quando na TRI se diz que um teste é unidimensional o que se quer dizer é que todos os indivíduos que possuem a mesma capacidade estimada têm a mesma probabilidade de dar uma resposta correta a cada item” (Cuesta, 1996, p. 247). Esta definição não pressupõe que a capacidade seja função de um único traço, basta que a combinação de traços seja idêntica para todos os itens.



### 2.2.2 Independência Local

A função característica do item é a regressão do escore do item nas variáveis latentes  $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \dots, \theta_k)$ . O problema prático de especificar  $\boldsymbol{\theta}$  é parte do problema de escolher um modelo apropriado (Lord e Novick, 1968).

Sempre quando se trata de diversos itens simultaneamente, a suposição de *independência local* é usualmente considerada necessária para a efetividade teórica do trabalho com as curvas características do item. De modo que, independência local significa que dentro de qualquer grupo de examinandos, todos caracterizados pelos mesmos valores  $\theta_1, \theta_2, \dots, \theta_k$ , as distribuições condicionais dos escores dos itens são todas independentes umas das outras. O que isto significa é que as pontuações dos itens estão relacionadas entre si somente através das variáveis latentes  $\theta_1, \theta_2, \dots, \theta_k$ .

A definição formal de independência local para itens  $g = 1, 2, \dots, n$  é que para quaisquer valores fixados de  $\theta_1, \theta_2, \dots, \theta_k$ , a distribuição conjunta  $f$  dos escores dos itens  $y_{g*}$  é igual ao produto das distribuições marginais  $f_g$ :

$$f(y_{1*}, y_{2*}, \dots, y_{n*} | \boldsymbol{\theta}) = \prod_{g=1}^n f_g(y_{g*} | \boldsymbol{\theta}). \quad (2.2)$$

O que se observa é que o argumento utilizado em (2.2) é o mesmo apresentado por (2.1) para se definir dimensionalidade (Vitória *et al.*, 2006).

Quando  $g = 2, 3, \dots, n$ , (2.2) torna-se:

$$f(y_{2*}, y_{2=3*}, \dots, y_{n*} | \boldsymbol{\theta}) = \prod_{g=2}^n f_g(y_{g*} | \boldsymbol{\theta}). \quad (2.3)$$

Dividindo (2.3) por (2.2) obtêm-se:

$$h_1(y_{1*} | \boldsymbol{\theta}; y_{2*}, y_{3*}, \dots, y_{n*}) = f_1(y_{1*} | \boldsymbol{\theta}). \quad (2.4)$$

Isto mostra que, sobre a independência local, a distribuição condicional de  $h_1$  de  $y_{1*}$  para valores fixos de  $\boldsymbol{\theta}, y_{2*}, y_{3*}, \dots, y_{n*}$  não dependem de  $y_{2*}, y_{3*}, \dots, y_{n*}$ , e este resultado se mantém não importando qual item seja escolhido primeiro.

Por outro lado, pode ser mostrado que caso (2.4) se mantenha para cada item, então (2.4) implica em (2.2). Portanto, ambas as equações constituem uma definição de independência local.

Supondo que (2.4) não se mantém, então para um grupo de examinandos todos com o mesmo  $\boldsymbol{\theta}$  e com o mesmo  $y_{g*}$  para  $g = 2, 3, \dots, n$ , a distribuição condicional de  $y_{1*}$  seria diferente do que para outros grupos de examinandos com o mesmo  $\boldsymbol{\theta}$  mas diferente  $y_{g*}$ . Porém, isto não pode acontecer no espaço latente completo, desde que o espaço seja definido de modo que a distribuição de  $y_{1*}$ , seja a mesma para todos os grupos pertinentes de examinandos. A suposição de independência local é então equivalente a suposição de que os  $\theta_1, \theta_2, \dots, \theta_k$  sob consideração abrangem o espaço latente completo.

Para itens binários, a suposição de independência local pode ser escrita como:

$$P(U_{1*} = u_1, U_{2*} = u_2, \dots, U_{n*} = u_n | \boldsymbol{\theta}) = \prod_{g=1}^n P(U_{g*} = u_g | \boldsymbol{\theta}) \quad (2.5)$$

onde  $u_g$  é igual a zero ou um. Denotando a probabilidade condicional por um dado  $\boldsymbol{\theta}$  de

uma resposta correta para um item binário  $g$  por  $P_g(\boldsymbol{\theta})$ , ou simplesmente  $P_g$ , dado por:

$$P_g \equiv P_g(\boldsymbol{\theta}) \equiv P(U_{g*} = 1|\boldsymbol{\theta}). \quad (2.6)$$

Esta função  $P_g(\boldsymbol{\theta})$  é a função característica do item para um item binário, isto é, é a regressão do escore do item sobre  $\boldsymbol{\theta}$ . A distribuição de frequência de escore de item binário para um  $\boldsymbol{\theta}$  fixado pode ser escrita como:

$$f_g(u_g|\boldsymbol{\theta}) \equiv P_g^{u_g} Q_g^{1-u_g}, \quad (2.7)$$

onde  $Q_g \equiv 1 - P_g$ . Escrevendo (2.7) de maneira simplificada, tem-se:

$$f_g(u_g|\boldsymbol{\theta}) \equiv \begin{cases} P_g & \text{se } u_g = 1, \\ Q_g & \text{se } u_g = 0. \end{cases} \quad (2.8)$$

O padrão de resposta do item, a ser denotado pelo vetor

$$\mathbf{V}_* \equiv (U_{1*}, U_{2*}, \dots, U_{n*})', \quad (2.9)$$

é um vetor de variáveis aleatórias. de (2.7) e (2.8) conclui-se que a distribuição condicional de  $\mathbf{V}_*$  para um dado  $\boldsymbol{\theta}$  é:

$$P(\mathbf{v}|\boldsymbol{\theta}) \equiv \prod_{n=1}^n P_g^{u_g} Q_g^{1-u_g}. \quad (2.10)$$

Se para alguma população de examinandos,  $\boldsymbol{\theta}$  tiver a distribuição de frequência denotada por  $g(\boldsymbol{\theta})$ , então  $P(\mathbf{v})$ , a distribuição não condicional de  $\mathbf{V}_*$  para o grupo total de examinandos, é dada pela equação:

$$P(\mathbf{v}) \equiv \int g(\boldsymbol{\theta}) \prod_{n=1}^n P_g^{u_g} Q_g^{1-u_g} d\boldsymbol{\theta}, \quad (2.11)$$

onde a integração é sobre o conjunto  $k$ -dimensional correspondendo aos elementos de  $\boldsymbol{\theta}$ . Observada uma amostra retirada de  $P(\mathbf{v})$ , pode-se usar (2.11) para fazer inferências sobre a distribuição desconhecida  $g$  das variáveis não observáveis  $\boldsymbol{\theta}$ , desde que a função  $P_g(\boldsymbol{\theta})$  seja conhecida.

## 2.2.3 Traço Latente e Curva Característica do Item

### 2.2.3.1 O Traço Latente

Quando é necessário mensurar a quantidade de um traço latente que uma pessoa possui, faz-se necessário ter uma escala de medida correspondente, isto é, uma régua com uma métrica dada. A tarefa de definir a escala de medida, os números na escala e a quantidade de característica que cada um dos números representam é, na *Teoria da Resposta ao Item*, uma tarefa que requer detalhes técnicos sofisticados.

Segundo [Andrade et al. \(2000\)](#), a escala da habilidade é uma escala arbitrária onde o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. Uma vez que existe uma unidade de medida e um ponto zero arbitrário, tal escala é referida como existente em um nível de medição de intervalo. A idéia básica é que se alguém pudesse verificar (medir) fisicamente ou matematicamente a “capacidade” de uma

pessoa, essa régua seria usada para dizer a quantidade de habilidade que a pessoa tem, de modo que a “capacidade” de várias pessoas possam ser comparadas.

Embora em termos teóricos a habilidade deva ser vista como pertencente ao conjunto dos números reais ( $\mathbb{R}$ ), considerações técnicas geralmente tratam de valores dentro do range  $-3$  a  $3$ . Cuidados devem ser tomados quanto às interpretações, haja vista que valores além desta faixa podem ser encontrados e não estarão incorretos. Sua representação mais geral encontrada na literatura é dada pela letra grega theta ( $\theta$ )

Uma maneira formal de definir o traço latente, segundo Lord e Novick (1968), se dá quando se considera um conjunto de  $n$  itens e um conjunto de  $k'$  características ou habilidades individuais, e supondo que cada uma das habilidades afeta o desempenho dos examinandos em pelo menos um item do conjunto, de forma que se possa denotar  $\theta$  por um vetor  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{k'})$ . Assim, cada examinando pode ser representado por um ponto no espaço  $k'$ -dimensional, chamado de *espaço latente*.

Assumindo que cada item é administrado uma única vez para cada examinando, e considerando a distribuição de frequência condicional (sobre os indivíduos) do escore do item  $Y_{g*}$  para qualquer valor fixo de  $\theta$ , se esta distribuição não observável não for a mesma para todas as populações de examinandos, então há uma ou mais dimensões adicionais para  $\theta_1, \theta_2, \dots, \theta_{k'}$  que discriminam entre as populações de interesse. Assim, por definição, no espaço latente completo, a distribuição condicional do escore do item para um  $\theta$  fixado é a mesma para todas as populações de interesse. De maneira que  $k$  denota a dimensionalidade do espaço latente completo.

Portanto, a natureza e a dimensionalidade do espaço latente completo são definidas de tal forma que dependem tanto dos traços particulares, como das populações particulares de interesse. Entretanto, apesar da impossibilidade de se distinguir objetivamente quais, dentre as variáveis que compõe o espaço latente completo, são aquelas associadas aos erros de medida, é usual para a compreensão do traço latente, incluir todas as dimensões importantes que afetam a performance sob um dado conjunto de itens e desconsiderando as definições subjetivas de erros de medida.

### 2.2.3.2 A Curva Característica do Item (CCI)

Segundo Lord e Novick (1968), a regressão do escore do item sobre  $\theta$  é conhecida por *Função Característica do Item*, ou *Curva Característica do Item*, de forma que por conta de sua definição, permaneça invariante entre os grupos de examinandos considerados ao definir o espaço latente completo. Isso significa que qualquer parâmetro que descreva a função característica do item é um parâmetro de item invariante.

Para itens binários, a função característica do item especifica exatamente como as respostas observadas de uma população de examinandos dependem dos traços latentes  $\theta_1, \theta_2, \dots, \theta_k$ . Genericamente, estes traços latentes são tratados como “habilidades” embora seja necessário manter em mente que uma grande variedade de capacidades cognitivas ou habilidades físicas possam ser rotuladas como habilidade (Baker e Kim, 2004).

Para o presente propósito, embora seja possível considerar diversos traços latentes de um mesmo indivíduo, assume-se que  $\theta_j$  seja o traço latente (*habilidade*) conhecido para cada indivíduo na população,  $j = 1, 2, \dots, N$ . De forma que a métrica da distribuição de frequência dos escores dos examinandos sobre a escala de habilidades é assumida ter média  $\mu_\theta$  e variância  $\sigma_\theta^2$ .

Em termos populacionais,  $P_t(\theta_j)$  denota a probabilidade de uma resposta correta em qualquer ponto da escala de habilidade e  $t$  denota um item,  $t = 1, 2, \dots, n$ . Dessa maneira, a curva característica do item (CCI) pode ser definida como um membro da família de funções

monótonas de dois parâmetros da variável habilidade, sendo os dois parâmetros definidos por:

$\beta_t$ : é o parâmetro de locação, também conhecido por parâmetro de *dificuldade* do item  $t$ , expresso em unidades de  $\theta$ , indicando o ponto na escala de habilidade onde  $P_t(\theta_j) = 0.5$   $-\infty \leq \beta_t \leq \infty$ .

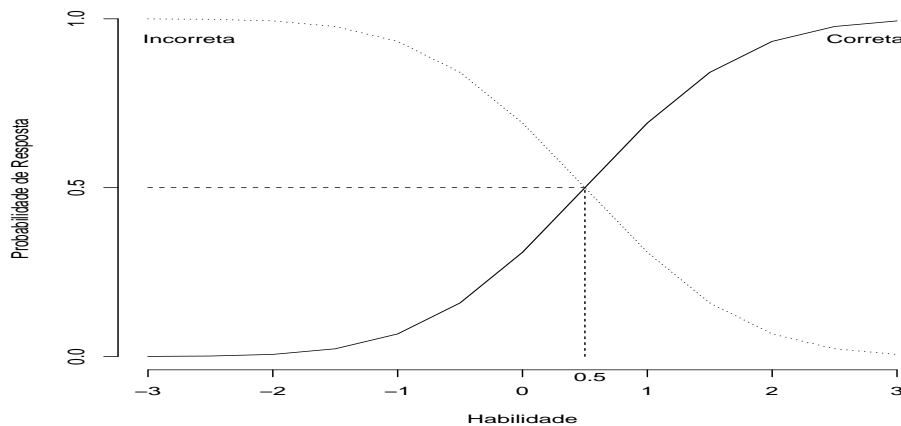
$\alpha_t$ : é o parâmetro de escala que indexa o poder discriminante (*discriminação*) de um item,  $-\infty \leq \alpha_t \leq \infty$

Formalmente, a curva característica do item pode ser definida pela função matemática

$$P_t(\theta_j) = P(\beta_t, \alpha_t, \theta_{jt}). \quad (2.12)$$

Diferentes famílias de funções de dois parâmetros, com aparência similar a funções de distribuição acumulada podem ser utilizadas para modelar as curvas características do item, embora as mais comuns encontradas na literatura sejam o *modelo de ogiva normal* (Lord e Novick, 1968) e o *modelo de ogiva logística* (Birnbaum, 1957).

Segundo Baker e Kim (2004), embora a prática comum seja discutir somente os resultados da CCI das respostas corretas para um item, as curvas são avaliadas para cada categoria dos itens, de forma que itens dicotômicos resultam em duas curvas características do item, uma para respostas certas e outra para respostas erradas. Assim, a probabilidade de uma resposta incorreta  $Q_t(\theta_j)$  também pode ser plotada como uma função dos escores da habilidade. A Figura 2.1 apresenta duas curvas para itens dicotômicos de modo que  $P_t(\theta_j) + Q_t(\theta_j) = 1$  é o resultado da CCI para cada nível da habilidade do indivíduo  $j$ . De forma semelhante, Samejima (1969) usou a terminologia “curvas características das categorias de resposta ao item” para identificar que havia uma curva característica do item associada para cada categoria de resposta. O modelo proposto baseia-se no fato de que indivíduos com maior habilidade possuem maior probabilidade de acertar o item e que esta relação não é linear (Andrade *et al.*, 2000). Fato este que pode ser constatado pela Figura 2.1, onde observa-se que a CCI tem um formato de “S” com inclinação ( $\alpha_t$ ) e deslocamento ( $\beta_t$ ) na escala de habilidade definida pelos parâmetros do item.



**Figura 2.1:** Curva Característica do Item para respostas corretas ( $\beta_t = 0.5, \alpha_t = 1.0$ ) e incorretas ( $\beta_t = 0.5, \alpha_t = -1.0$ )

### 2.2.3.3 O Modelo de Ogiva Normal

Na literatura encontram-se diversas justificativas para o uso da ogiva normal como modelo para curva característica do item. As explicações mais pragmáticas datam de período anterior aos anos 50, quando basicamente assumia-se que a CCI pudesse ser modelada por uma ogiva normal, de modo que esta era ajustada para a proporção de respostas corretas ao longo da escala de habilidade e a avaliação da qualidade de ajuste feita por testes chi-quadrado.

Segundo Baker e Kim (2004), a formulação para representar matematicamente a expressão (2.12) é proveniente de uma função de distribuição normal acumulada, dada por:

$$P_t(\theta_j) \equiv P(\mu_t, \sigma_t, \theta_{jt}) = \Phi(Z_t) = \int_{-Z_t = -(\theta_{jt} - \mu_t)/\sigma_t}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \quad (2.13)$$

onde  $\mu_t$  é a média,  $\sigma_t$  é o desvio padrão,  $\theta_{jt}$  é a habilidade do  $j$ -ésimo indivíduo, e  $Z_t$  é o desvio normal, ambos para o  $t$ -ésimo item.

Devido a natureza simétrica da distribuição normal, tem-se que  $\Phi(Z_t) = 1 - \Phi(-Z_t)$ . Na seção 2.2.3.2 apresentou-se  $\beta_t$  como parâmetro de locação e  $\alpha_t$  como parâmetro de discriminação. Em (2.13) observa-se que os parâmetros de locação e escala são  $\mu_t$  e  $\alpha_t$ , respectivamente. No caso da ogiva normal, a média e a mediana são idênticas, e assim, que uma vez  $\beta_t$  é definido como o ponto na escala da habilidade cuja probabilidade de resposta correta é 0.5, o que corresponde à média da ogiva normal. Portanto,  $\beta_t = \mu_t$ .

Em uma distribuição normal,  $\sigma_t$  representa o desvio padrão da distribuição, sendo consequentemente a medida de dispersão da distribuição. Assim, quando  $\sigma_t$  é grande, a ogiva normal tende a ser mais achatada e não muito inclinada próximo a  $\beta_t$ . Quando  $\sigma_t$  é pequeno, a ogiva normal tende a ser mais inclinada na região de  $\beta_t$ . Quanto mais inclinado o meio da ogiva normal, menor o valor de  $\sigma_t$ . Esta relação inversa explica o fato de  $\alpha_t$  ser uma função de  $\sigma_t$ , e como consequência, os parâmetros  $\beta_t$  e  $\alpha_t$  da curva característica do item são funções diretas dos parâmetros de *locação* e *escala* da função de distribuição normal acumulada. Além disso, de (2.13), tem-se:

$$Z_t = \frac{\theta_{jt} - \mu_t}{\sigma_t}, \quad (2.14)$$

mas  $\beta_t = \mu_t$  e  $\alpha_t = 1/\sigma_t$ . Então o desvio normal pode ser escrito como:

$$Z_t = \frac{\theta_{jt} - \mu_t}{\sigma_t} = \alpha_t(\theta_{jt} - \beta_t). \quad (2.15)$$

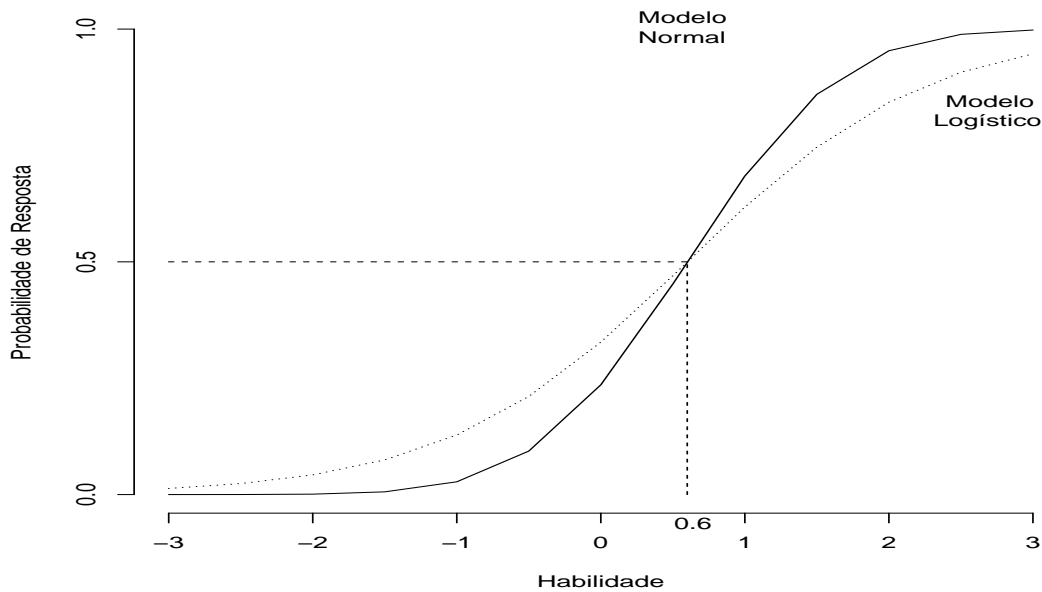
Uma apresentação apropriadamente matemática da discussão sobre os parâmetros da ogiva normal, podem ser vistos em Lord e Novick (1968, p. 376). Um dos resultados encontrados é que em TRI o conceito de locação da curva característica do item não é o mesmo de dificuldade do item no contexto da TCT, entretanto, por motivos de compatibilidade, o termo “dificuldade” é utilizado em conjunto com  $\beta_t$ . Vale ressaltar que na região de  $\beta_t$  a CCI é aproximadamente linear com uma inclinação de  $\alpha_t/\sqrt{2\pi}$  (Lord e Novick, 1968, p. 368). Uma interpretação rudimentar do parâmetro de discriminação é que ele é proporcional à inclinação da CCI no ponto sob a escala de habilidade que corresponde à  $\beta_t$ .

### 2.2.3.4 O Modelo de Ogiva Logística

O uso de uma ogiva logística como um modelo para a CCI é atribuído a [Birnbbaum \(1957\)](#). A forma acumulada da função logística, frequentemente chamada de lei logística é dada por:

$$P_t(\theta_j) = P(\alpha_t^*, \beta_t, \theta_{jt}) = \Psi(Z_t) = \frac{e^{Z_t}}{1 + e^{Z_t}} = \frac{1}{1 + e^{-Z_t}}, \quad (2.16)$$

onde  $Z_t = \alpha_t^*(\theta_{jt} - \beta_t)$  é chamado de *logito*;  $\beta_t$  é o parâmetro de locação, e representa o ponto na escala da habilidade no qual  $P_t(\theta_j) = 0.5$ ;  $\alpha_t$  é o parâmetro de discriminação e é o recíproco do desvio padrão da função logística.



**Figura 2.2:** Curva Característica do Item - Ogiva Normal versus Ogiva Logística ( $\beta_t = 0.6, \alpha_t = 1.2$ )

Observando as duas curvas apresentadas na Figura 2.2 nota-se que as curvas são similares, compartilhando o mesmo parâmetro de escala, embora tenha-se considerado o parâmetro de discriminação o mesmo ( $\alpha_t = \alpha_t^*$ ), as inclinações das curvas são diferentes. Porém, manteve-se o mesmo ponto de inflexão, de forma que o valor numérico de  $\beta_t = 0.6$  tem o mesmo significado para ambos os modelos. Estes resultados devem-se à diferença na variância das duas funções, de modo que, enquanto uma função de densidade  $N(0, 1)$  tem desvio unitário, a função correspondente logística terá a mesma média, porém, variância na ordem de  $\pi^2/3$ , evidenciando a maior dispersão desta. [Haley \(1952\)](#) mostrou que quando o desvio logístico  $Z_t = \alpha_t^*(\theta_{jt} - \beta_t)$  mantém o parâmetro de escala constante em 1.702, a diferença absoluta entre a  $P_t(\theta_j)$  da ogiva normal e da ogiva logística é menor do que 0.01 para todos os valores de  $\theta_{jt}$ , de modo que:

$$|\Phi(Z_t) - \Psi(1.702Z_t)| < 0.01 \quad \forall -\infty < \theta_{jt} < \infty. \quad (2.17)$$

A característica mais importante da função logística é que sua distribuição acumulada tem forma fechada, de forma que comparada à (2.13), a expressão de  $P_t(\theta_j)$  em (2.16) é prontamente avaliada, sem a necessidade de qualquer integração numérica, dado que se conheça o valor do desvio  $Z_t$ . A função logística está também relacionada à chance de

respostas corretas. Em qualquer ponto da escala de habilidade  $\theta_{jt}$ , a probabilidade de uma resposta correta é dada por  $P_t(\theta_j)$ , de modo que a probabilidade de uma resposta incorreta é dada por  $Q_t(\theta_j) = 1 - P_t(\theta_j)$ . Assim, as chances de se obter uma resposta correta são  $P_t(\theta_j)/Q_t(\theta_j)$ . Se o logaritmo natural desta razão é tomado, tem-se:

$$\ln \left[ \frac{P_t(\theta_j)}{Q_t(\theta_j)} \right] = \ln(P_t(\theta_j)) - \ln(Q_t(\theta_j)) = Z_t. \quad (2.18)$$

Para mostrar que (2.18) se mantém sob o modelo logístico, considere a função acumulada dada em (2.16), de forma que:

$$\begin{aligned} e^{-Z_t} &= \frac{1 - P_t(\theta_j)}{P_t(\theta_j)} = \frac{Q_t(\theta_j)}{P_t(\theta_j)} \\ Z_t &= -\ln \left[ \frac{Q_t(\theta_j)}{P_t(\theta_j)} \right] = \alpha_t(\theta_{jt} - \beta_t) \end{aligned} \quad (2.19)$$

A relação apresentada em (2.19), entre o desvio logístico e a proporção de respostas corretas, é conhecida como transformação logística. Para o propósito de estimar os parâmetros dos itens, é conveniente usar a transformação para obter uma representação linear da CCI. De modo que com (2.19) a ogiva logística relativa à  $P_t(\theta_j)$  e  $\theta_{jt}$  torna-se uma linha de regressão linear que relaciona  $Z_t$  e  $\theta_{jt}$ .

## 2.2.4 Escala de Habilidade

A escala de habilidade na TRI é um dos conceitos mais importantes quando se trata da interpretação dos resultados do método. Ao contrário do score de um teste com número finito de questões dicotômicas, que assume valores entre 0 e 1, a habilidade pode, teoricamente, assumir quaisquer valores reais entre  $-\infty$  e  $+\infty$ .

É necessário que se defina uma origem e uma unidade de medida, para a definição da escala. Em [Andrade et al. \(2000\)](#), utilizou-se a escala com média igual a 0 e desvio-padrão igual a 1, sendo representada por escala (0,1). Essa escala é bastante utilizada pela TRI, e neste caso, os valores do parâmetro  $\beta_t$  variam aproximadamente entre -2 e +2, e o parâmetro  $\alpha_t$  entre 0 e +2, sendo que os valores mais apropriados de  $\alpha_t$  seriam aqueles maiores do que 1.

Embora em termos práticos, não haja diferença entre estabelecer estes valores ou outros quaisquer, o importante são as relações de ordem existentes entre seus pontos. Por exemplo, na escala (0,1) um indivíduo com habilidade 1,20 está 1,20 desvios-padrão acima da habilidade média. Este mesmo indivíduo teria a habilidade 248, e conseqüentemente estaria também 1,20 desvios-padrão acima da habilidade média, se a escala utilizada para esta população fosse a escala (200;40). Isto pode ser visto, considerando os desvios apresentados em (2.15) e (2.19), a partir da transformação de escala<sup>1</sup>:

$$\alpha_t(\theta_{jt} - \beta_t) = \left( \frac{\alpha_t}{40} \right) [(40 \times \theta_{jt} + 200) - (40 \times \beta_t + 200)] = \alpha_t^*(\theta_{jt}^* - \beta_t^*), \quad (2.20)$$

onde  $\alpha_i(\theta_j - \beta_i)$  é a parte do modelo probabilístico proposto envolvida na transformação. Com a reparametrização tem-se:

1.  $\theta_{jt}^* = 40 \times \theta_{jt} + 200,$

---

<sup>1</sup>Para maiores detalhes sobre transformações em distribuições da família locação-escala, ver [\(Lehmann e Casella, 1998, Pág. 16-23\)](#)



2.  $\beta_t^* = 40x + 200$ ,
3.  $\alpha_t^* = \frac{\alpha_t}{40}$ ,
4.  $P_t(\theta_j) = P_t(\theta_j^*)$ .

De forma que a probabilidade de um indivíduo responder corretamente a um certo item é sempre a mesma, independente da escala utilizada para medir sua habilidade, o que equivale a dizer que a habilidade de um indivíduo é *invariante* à escala de medida. A conclusão é que não faz sentido analisar itens a partir dos valores de seus parâmetros, sem conhecer a escala a qual pertencem.

## 2.3 Estimação em TRI

Segundo Zheng (2014), a *Teoria da Resposta ao Item* possui diversas formas de modelar a probabilidade de respostas corretas, entre elas o caso em que o item é pontuado dicotomicamente, ou de diferentes níveis de resposta, caso em que o item tem mais de duas categorias de resposta. Essas probabilidades dependem dos parâmetros do item específico e do examinando. Quando os itens são dicotômicos, assim como de múltipla escolha, os modelos de TRI mais comumente utilizados são os modelos *logísticos de 1 parâmetro* (1PL), modelos *logísticos de 2 parâmetros* (2PL), e os modelos *logísticos de 3 parâmetros* (3PL).

Considerando  $U_{jt}$  ser uma variável dicotômica que assume o valor 1, quando o indivíduo  $j$  responde corretamente o item  $t$ , ou 0 quando o indivíduo  $j$  não responde corretamente ao item  $t$ , a correspondente probabilidade  $P(U_{jt} = 1|\theta_{jt})$ , avaliada sobre o nível de habilidade ( $\theta_{jt}$ ) do  $j$ -ésimo examinando, quando responde ao  $t$ -ésimo item, é modelada pelas seguintes *Funções de Resposta do Item* (FRI):

1PL

$$P(U_{jt} = 1|\theta_{jt}) = \frac{1}{1 + e^{[-(\theta_{jt} - \beta_t)]}}; \quad (2.21)$$

2PL

$$P(U_{jt} = 1|\theta_{jt}) = \frac{1}{1 + e^{[-\alpha_t(\theta_{jt} - \beta_t)]}}; \quad (2.22)$$

3PL

$$P(U_{jt} = 1|\theta_{jt}) = c_t + (1 - c_t) \frac{1}{1 + e^{[-\alpha_t(\theta_{jt} - \beta_t)]}}. \quad (2.23)$$

Nas equações acima, todos os parâmetros do item podem variar de acordo com os itens individuais, que descrevem as características de cada item. Além dos parâmetros especificados em (2.12), quando definindo a CCI, as equações acima introduzem o parâmetro  $c_t$  como sendo o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade de responderem corretamente o item  $t$ . Muitas vezes é referido como a probabilidade de acerto casual (Andrade *et al.*, 2000).

É importante notar que  $P(U_{jt} = 1|\theta_{jt})$  pode ser vista como a proporção de respostas corretas ao item  $t$  dentre todos os indivíduos da população com habilidade  $\theta_{jt}$ . A relação entre  $P(U_{jt} = 1|\theta_{jt})$  e os parâmetros do modelo é apresentada na Figura 2.3, de onde observa-se que o parâmetro  $\beta_j$  é medido na mesma unidade de habilidade e que o parâmetro



$c_t$  não depende da escala, tratando-se de uma probabilidade. Na realidade, o parâmetro  $\beta_t$  representa a habilidade necessária para uma probabilidade de acerto da ordem de  $(1 + c_t)/2$ . De modo que, quanto maior o valor  $\beta_t$ , mais difícil é o item, e vice-versa. O parâmetro  $\alpha_t$ , como mostra a Figura 2.3, é proporcional à derivada da tangente da curva no ponto de inflexão. Assim, itens com  $\alpha$  negativo, não são esperados sob esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade. De modo que valores baixos de  $\alpha$ , indicam que o item tem pouco poder de discriminação, enquanto valores muito altos, indicam itens com curvas características muito íngremes.

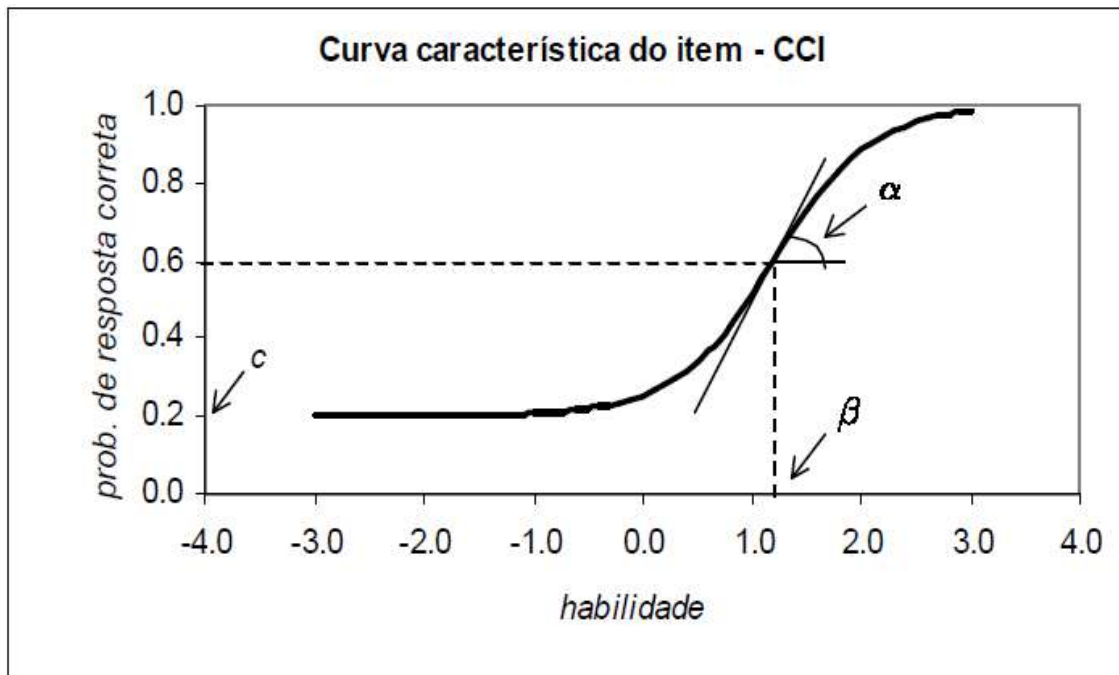


Figura 2.3: Figura adaptada de (Andrade et al., 2000, Pág. 11)

O modelo 1PL (2.21) é o mais simples entre os três apresentados, mas tem a suposição mais forte: todos os itens são assumidos ter igual poder de discriminação e nenhuma chance de resposta casual. O modelo 2PL (2.22) assume nenhuma chance para resposta casual, mas permite um poder de discriminação variado, quando modelado pelos parâmetros  $\alpha_t$ . O modelo 3PL (2.23) inclui todos os três parâmetros, de modo que podem delinear um perfil mais robusto de um item.

Quando as respostas para um item podem ser pontuadas com mais de dois níveis, assim como pequenas respostas com créditos parciais, há uma variedade de modelos politômicos para modelar as respostas dos dados. Embora este trabalho não tenha foco em tais modelos, concentrando-se apenas nos modelos dicotômicos de três parâmetros (3PL), citam-se, a seguir, os mais conhecidos e, para os leitores interessados em mais detalhes, recomenda-se a leitura de Baker e Kim (2004), Embretson e Reise (2000), Samejima (1997), Hambleton et al. (1991), Lord (1980).

Há muitos trabalhos na literatura que abordam não apenas a classe de modelos politômicos, assim como, todo o processo de estimação dos mesmos, a saber:

- Modelo de Resposta Gradual, (Samejima, 1969);
- Modelo de Crédito Parcial, (Masters, 1982);
- Modelo de Crédito Parcial Generalizado, (Muraki, 1992);

- *Modelo de Taxa de Escala*, (Andrich, 1978);
- *Modelo de Resposta Nominal*, (Bock, 1972).

### 2.3.1 Função de Informação do Teste (FIT)

Na TCT, o índice de dificuldade de um item é definido como a proporção de examinandos que respondem ao item corretamente. Como resultado, nem o escore do examinando, nem o índice de dificuldade dos itens, são invariantes na amostra. Em outras palavras, o escore do examinando depende da dificuldade dos itens no teste, e o índice de dificuldade do item depende do nível da habilidade dos examinandos que respondem ao item. De forma que o erro padrão de medida é usualmente calculado através do índice de confiança do teste completo, conduzindo assim ao fato contraintuitivo de que o erro padrão de medida não varia para diferentes examinandos.

Na TRI, o erro padrão de medida não é mais mantido constante através dos diferentes níveis de habilidade dos examinandos. Ao invés disso, o clássico índice estatístico, *Informação de Fisher*, foi trazido para dentro da TRI para fornecer os limites inferiores do quadrado do erro padrão de medida em cada nível da habilidade ( $\theta_j$ ) do  $j$ -ésimo examinando. Definida por:

$$I(\theta_j) = -E \left[ \frac{\partial^2}{\partial \theta_j^2} \log f(X; \theta_j) \right] \quad (2.24)$$

onde  $\theta_j$  denota o parâmetro desconhecido,  $X$  o conjunto de dados da amostra e  $\log f(X; \theta_j)$  a função de log-verossimilhança, a Informação de Fisher, no contexto da TRI, também é conhecida por *Função de Informação do Teste* (FIT) de examinandos com habilidade  $\theta_j$ , e é dada por :

$$I(\theta_j) = \sum_{t=1}^n \frac{[P'_t(\theta_j)]^2}{P_t(\theta_j)[1 - P_t(\theta_j)]}, \quad (2.25)$$

onde  $t = 1, 2, \dots, n$ , denotando  $n$  itens respondidos pelos examinandos e  $P_t(\theta_j)$  a *Função de Resposta ao Item* (FRI) de alguma das formulações das equações 2.21-2.23. Considerando-se apenas um dos itens da equação 2.25, tem-se a *Função de Informação do Item* (FII), medida bastante utilizada em conjunto com a CCI, que permite analisar quanto um item contém de informação para a medida de habilidade. No caso do modelo logístico 3PL, pode ser escrita como:

$$I_t(\theta_j) = D^2 \alpha_t^2 \frac{Q_t(\theta_j)}{P_t(\theta_j)} \left[ \frac{(P_t(\theta_j) - c_t)}{1 - c_t} \right]^2. \quad (2.26)$$

Segundo Andrade *et al.* (2000), esta equação mostra a importância que têm os três parâmetros sobre o montante de informação do item. Isto é, a informação é maior:

- (i) quando  $\beta_t$  se aproxima de  $\theta_j$ ;
- (ii) quanto maior for o  $\alpha_t$ ;
- (iii) e quanto mais  $c_t$  se aproximar de 0.

Outra maneira de representar a FIT é através do erro padrão de medida, conhecido na TRI por erro padrão de estimação, dado por:

$$EP(\theta_j) = \frac{1}{\sqrt{I(\theta_j)}}. \quad (2.27)$$

Como essas medidas dependem do valor de  $\theta_j$ , a amplitude do intervalo de confiança para  $\theta_j$  dependerá também do seu valor. À medida em que o número de itens aumenta, a variância da estimativa de  $\theta_j$  aproxima-se de  $1/I(\theta_j)$ . Portanto,  $1/\sqrt{I(\theta_j)}$  fornece o limite inferior do erro padrão de medida sob  $\theta_j$  (Zheng, 2014).

Como a TRI coloca o parâmetro de dificuldade do item na mesma escala do parâmetro de habilidade do examinando, a idéia de combinar a dificuldade do item com a habilidade do examinando, torna-se simplesmente combinar aos valores do parâmetro de dificuldade do item ao parâmetro de habilidade do examinando. Considerando o conhecido limite inferior do erro padrão de medida citado anteriormente, itens com maiores valores de informação, dados pelo  $I_t(\theta_j)$ , produzirão menores erros padrão de medida sob  $\theta_j$ , ou seja, melhor precisão de medição para o nível de proficiência do examinando. Na verdade, itens muito fáceis ou muito difíceis fornecerão pouca informação para estimar o nível de proficiência do examinando, enquanto os itens com dificuldades correspondentes geralmente são mais informativos. Isto pode ser matematicamente mostrado para os modelos 1PL, 2PL e 3PL através de suas funções de informação (Chang e Ying, 2009).

Para o modelo 1PL,

$$I(\theta_j|\beta_t) = \frac{e^{(\theta_j-\beta_t)}}{[1 - e^{(\theta_j-\beta_t)}]^2}. \quad (2.28)$$

Para um dado examinando,  $I(\theta_j)$  atinge seu valor máximo em  $1/4$  quando  $\beta_t = \theta_j$ .

Para o modelo 2PL,

$$I(\theta_j|\alpha_t, \beta_t) = \frac{\alpha_t^2 e^{\alpha_t(\theta_j-\beta_t)}}{[1 - e^{\alpha_t(\theta_j-\beta_t)}]^2}. \quad (2.29)$$

Para um dado examinando e um fixo valor do parâmetro  $\alpha_t$ ,  $I(\theta_j)$  atinge seu valor máximo em  $\alpha_t^2/4$  quando  $\beta_t = \theta_j$ .

Para o modelo 3PL,

$$I(\theta_j|\alpha_t, \beta_t, c_t) = \frac{(1 - c_t)\alpha_t^2 e^{2\alpha_t(\theta_j-\beta_t)}}{\{[c_t + e^{\alpha_t(\theta_j-\beta_t)}]\}[1 - e^{\alpha_t(\theta_j-\beta_t)}]^2}. \quad (2.30)$$

Para um dado examinando e fixos valores do parâmetro  $\alpha_t$  e parâmetro  $c_t$ ,  $I(\theta_j)$  atinge seu valor máximo quando

$$\beta_t = \theta_j - \frac{1}{\alpha_t} \log \left( \frac{1 + \sqrt{1 + 8c_t}}{2} \right). \quad (2.31)$$

Isto está tipicamente na vizinhança do valor de  $\theta_j$ . Por exemplo, quando  $\alpha_t = 2$  e  $c_t = 0.25$ , a dificuldade ótima do item selecionado,  $\beta_t \approx \theta_j - 0.15$ . Uma discussão mais detalhada das expressões 2.28-2.30, bem como a justificativa matemática para 2.31 podem ser obtidos em

Chang e Ying (2009).

### 2.3.2 Estimação dos Parâmetros dos Itens

Uma das etapas mais importantes da TRI é a estimação dos parâmetros dos itens e das habilidades dos respondentes. Segundo Andrade *et al.* (2000), é possível dividir o problema em três situações:

- Quando já se conhecem os parâmetros dos itens, é necessário estimar apenas as habilidades.
- Se já são conhecidas as habilidades, é necessário estimar apenas os parâmetros dos itens.
- E no caso mais comum, é necessário estimar os parâmetros dos itens e as habilidades dos indivíduos simultaneamente.

A estimação dos parâmetros dos itens ou dos examinandos dependem de algoritmos estatísticos. Quando os parâmetros dos itens são conhecidos, ou seja, calibrados, algoritmos comuns para estimação da habilidade dos examinandos  $\theta_j$  incluem a *estimação por máxima verossimilhança* (MLE), *estimação da esperança a posteriori* (EAP) e *estimação da moda a posteriori* (MAP). Quando ambos os parâmetros dos itens e parâmetros dos examinandos estão para ser estimados, o algoritmo de *estimação de máxima verossimilhança conjunta* (JMLE) pode ser empregado. Enquanto JMLE foi provado ser bem sucedido para o modelo 1PL, o mesmo encontrou dificuldades com modelos mais complexos e números maiores de examinandos. Nestes casos, a rotina mais comumente utilizada para a estimação é a *estimação por máxima verossimilhança marginal* (MMLE), empregada conjuntamente com o algoritmo EM, introduzido por Dempster *et al.* (1977). O método MMLE-EM integra o parâmetro  $\theta_j$  dos examinandos, usando sua distribuição à posteriori obtida a partir dos dados e, em seguida, encontra os valores dos parâmetros do item que maximizam essa verossimilhança da esperança a posteriori.

Embora existam outras propostas de estimação para os parâmetros dos itens e habilidades, as citadas acima podem ser consideradas as mais importantes e, portanto, este estudo irá explorá-los. Para melhor compreensão dos métodos a serem apresentados, adotam-se as convenções aplicadas por Azevedo (2003) e as notações e desenvolvimentos apresentados em Andrade *et al.* (2000), para descrever o cenário onde serão avaliados os métodos de estimação.

Suponha que um conjunto de  $N$  indivíduos foi submetido a um teste (prova) de  $n$  itens (questões) dicotômicos ou dicotomizados (itens de múltipla escolha corrigidos como certo/errado), onde  $\theta_{jt}$  é a habilidade e  $U_{jt}$  a variável aleatória que representa a resposta (binária) do indivíduo  $j$  ao item  $t$ , com

$$U_{jt} = \begin{cases} 1, & \text{resposta correta,} \\ 0, & \text{resposta incorreta.} \end{cases}$$

Os processos de estimação apresentados a seguir levam em consideram o fato de que o foco do estudo são os modelos dicotômicos de 3PL, de forma que a FRI considerada para os desenvolvimentos apresentados referem-se ao modelo (2.23), conjuntamente com as seguintes notações:

- $\mathbf{U}_j = (U_{j1}, U_{j2}, \dots, U_{jn})'$ , o vetor aleatório de respostas do  $j$ -ésimo indivíduo.

- $\mathbf{U}_{..} = (\mathbf{U}'_1, \dots, \mathbf{U}'_N)'$ , o conjunto integral de respostas.
- $\zeta_t = (\alpha_t, \beta_t, c_t)'$  vetor dos parâmetros do  $t$ -ésimo item.
- $\zeta = (\zeta'_1, \dots, \zeta'_n)'$  vetor dos parâmetros de todos os itens.
- $\theta = (\theta_1, \dots, \theta_N)'$ , vetor de habilidades dos  $N$  indivíduos.

No intuito de ser mais específico com respeito à notação, a qual evidencia a relação entre o modelo e o vetor de parâmetros dos itens, denota-se, a partir deste ponto, a Função de Resposta (FRI) por:

$$P(U_{jt} = 1 | \theta_{jt}, \zeta_t) = c_t + (1 - c_t) \frac{1}{1 + e^{[-D\alpha_t(\theta_{jt} - \beta_t)]}}, \quad (2.32)$$

onde  $D$  é um fator de escala igual a 1, quando se deseja obter o resultado da função logística, ou 1,7 caso o objetivo for obter uma aproximação para a distribuição normal. Vale ressaltar que o modelo (2.32) possui duas principais suposições que são imprescindíveis para o desenvolvimento dos processos de estimação:

- (S1) as respostas oriundas de indivíduos diferentes são independentes.
- (S2) os itens são respondidos de forma independente por cada indivíduo (independência local) da a sua habilidade.

### 2.3.2.1 Estimação de Máxima Verossimilhança (MLE)

Considerando as habilidades  $\theta_j$  conhecidas, e os pressupostos (S1) e (S2), é possível escrever a verossimilhança  $L(\zeta) = P(\mathbf{U}_{..} = \mathbf{u}_{..} | \theta, \zeta)$ , como

$$\begin{aligned} L(\zeta) &= \prod_{j=1}^N P(\mathbf{U}_j = \mathbf{u}_j | \theta_{jt}, \zeta) \\ &= \prod_{j=1}^N \prod_{t=1}^n P(\mathbf{U}_{jt} = \mathbf{u}_{jt} | \theta_{jt}, \zeta_t), \end{aligned} \quad (2.33)$$

onde na última igualdade, tem-se a distribuição de  $U_{jt}$  dependendo unicamente de  $\zeta$  através de  $\zeta_t$  (pelo modelo). Usando a notação  $P_{jt} = P(U_{jt} = 1 | \theta_{jt}, \zeta_t)$  e  $Q_{jt} = 1 - P_{jt}$ , tem-se

$$\begin{aligned} P(\mathbf{U}_{jt} = \mathbf{u}_{jt} | \theta_{jt}, \zeta_t) &= P(\mathbf{U}_{jt} = 1 | \theta_{jt}, \zeta_t)^{\mathbf{u}_{jt}} P(\mathbf{U}_{jt} = 0 | \theta_{jt}, \zeta_t)^{1 - \mathbf{u}_{jt}} \\ &= P_{jt}^{\mathbf{u}_{jt}} Q_{jt}^{1 - \mathbf{u}_{jt}}. \end{aligned} \quad (2.34)$$

Portanto, a verossimilhança pode ser escrita como

$$L(\zeta) = \prod_{j=1}^N \prod_{t=1}^n P_{jt}^{\mathbf{u}_{jt}} Q_{jt}^{1 - \mathbf{u}_{jt}}. \quad (2.35)$$

De forma que a log-verossimilhança  $[\ln(L(\zeta)) = l(\zeta)]$  pode ser escrita como

$$l(\zeta) = \sum_{j=1}^N \sum_{t=1}^n \{ \mathbf{u}_{jt} \ln P_{jt} + (1 - \mathbf{u}_{jt}) \ln Q_{jt} \}. \quad (2.36)$$

Como os Estimadores de Máxima Verossimilhança (EMV) de  $\zeta_t$ ,  $t = 1, \dots, n$ , são os valores que maximizam a verossimilhança, são por consequência, as soluções da equação

$$\frac{\partial l(\zeta)}{\partial \zeta_t} = \mathbf{0}, \quad t = 1, \dots, n. \quad (2.37)$$

Resolvendo a derivada em (2.37), tem-se

$$\begin{aligned} \frac{\partial l(\zeta)}{\partial \zeta_t} &= \sum_{j=1}^N \left\{ \mathbf{u}_{jt} \frac{\partial(\ln P_{jt})}{\partial \zeta_t} - (1 - \mathbf{u}_{jt}) \frac{\partial(\ln Q_{jt})}{\partial \zeta_t} \right\} \\ &= \sum_{j=1}^N \left\{ \mathbf{u}_{jt} \frac{1}{P_{jt}} \left( \frac{\partial P_{jt}}{\partial \zeta_t} \right) - (1 - \mathbf{u}_{jt}) \frac{1}{Q_{jt}} \left( \frac{\partial P_{jt}}{\partial \zeta_t} \right) \right\} \\ &= \sum_{j=1}^N \left\{ \mathbf{u}_{jt} \frac{1}{P_{jt}} - (1 - \mathbf{u}_{jt}) \frac{1}{Q_{jt}} \right\} \left( \frac{\partial P_{jt}}{\partial \zeta_t} \right) \\ &= \sum_{j=1}^N \left\{ \frac{\mathbf{u}_{jt} - P_{jt}}{P_{jt} Q_{jt}} \right\} \left( \frac{\partial P_{jt}}{\partial \zeta_t} \right). \end{aligned} \quad (2.38)$$

Por conveniência, considera-se a ponderação:

$$W_{jt} = \frac{P_{jt}^* Q_{jt}^*}{P_{jt} Q_{jt}}, \quad (2.39)$$

onde

$$P_{jt}^* = \left\{ 1 + e^{-D\alpha_t(\theta_{jt} - \beta_t)} \right\}^{-1} \quad \text{e} \quad Q_{jt}^* = 1 - P_{jt}^*. \quad (2.40)$$

Possibilitando assim, reescrever a equação (2.38)

$$\frac{\partial l(\zeta)}{\partial \zeta_t} = \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \left( \frac{\partial P_{jt}}{\partial \zeta_t} \right). \quad (2.41)$$

De forma que para obter as equações de estimação, serão necessárias as seguintes expressões:

$$\frac{\partial P_{jt}}{\partial \alpha_t} = D(1 - c_t)(\theta_{jt} - \beta_t) P_{jt}^* Q_{jt}^*, \quad (2.42)$$

$$\frac{\partial P_{jt}}{\partial \beta_t} = -D\alpha_t(1 - c_t) P_{jt}^* Q_{jt}^*, \quad (2.43)$$

$$\frac{\partial P_{jt}}{\partial c_t} = Q_{jt}^*. \quad (2.44)$$

Para o parâmetro de discriminação, de (2.41) e (2.42) têm-se que

$$\begin{aligned}
\frac{\partial l(\zeta)}{\partial \alpha_t} &= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \left( \frac{\partial P_{jt}}{\partial \alpha_t} \right) \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) D(1 - c_t) (\theta_{jt} - \beta_t) P_{jt}^* Q_{jt}^* \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= D(1 - c_t) \sum_{j=1}^N \{ (\mathbf{u}_{jt} - P_{jt}) (\theta_{jt} - \beta_t) W_{jt} \}.
\end{aligned} \tag{2.45}$$

Para o parâmetro de dificuldade, de (2.41) e (2.43) têm-se que

$$\begin{aligned}
\frac{\partial l(\zeta)}{\partial \beta_t} &= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \left( \frac{\partial P_{jt}}{\partial \beta_t} \right) \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) (-1) D \alpha_t (1 - c_t) P_{jt}^* Q_{jt}^* \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= -D \alpha_t (1 - c_t) \sum_{j=1}^N \{ (\mathbf{u}_{jt} - P_{jt}) W_{jt} \}.
\end{aligned} \tag{2.46}$$

Para o parâmetro de acerto ao acaso, de (2.41) e (2.44) têm-se que

$$\begin{aligned}
\frac{\partial l(\zeta)}{\partial c_t} &= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \left( \frac{\partial P_{jt}}{\partial c_t} \right) \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) Q_{jt}^* \frac{W_{jt}}{P_{jt}^* Q_{jt}^*} \right\} \\
&= \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \frac{W_{jt}}{P_{jt}^*} \right\}.
\end{aligned} \tag{2.47}$$

Por fim, as equações de estimação para os parâmetros  $\alpha_t$ ,  $\beta_t$  e  $c_t$  são, respectivamente,

$$\alpha_t : D(1 - c_t) \sum_{j=1}^N \{ (\mathbf{u}_{jt} - P_{jt}) (\theta_{jt} - \beta_t) W_{jt} \} = 0, \tag{2.48}$$

$$\beta_t : -D \alpha_t (1 - c_t) \sum_{j=1}^N \{ (\mathbf{u}_{jt} - P_{jt}) W_{jt} \} = 0, \tag{2.49}$$

$$c_t : \sum_{j=1}^N \left\{ (\mathbf{u}_{jt} - P_{jt}) \frac{W_{jt}}{P_{jt}^*} \right\} = 0. \tag{2.50}$$

### 2.3.2.2 Estimação de Máxima Verossimilhança Conjunta (JMLE)

Considerando que os procedimentos apresentados nesta seção são representações de dois blocos - um para a estimação dos parâmetros dos itens e outro para a estimação dos parâmetros das habilidades dos examinandos - os detalhes algébricos são apenas abstrações dos resultados obtidos na seção (2.3.2.1).

Segundo Baker e Kim (2004), sob condições típicas de teste, uma amostra aleatória de  $N$  examinandos é retirada de uma população de examinandos onde o traço latente (habilidade) é desconhecido. Estes examinandos respondem aos  $n$  itens do teste, e as respostas, representadas pela variável aleatória binária  $U_{jt}$ , onde  $t$  denota o item,  $t = 1, \dots, n$ , e  $j$  denota o examinando,  $j = 1, \dots, N$ . Para cada examinando, considera-se um vetor de respostas aos itens de tamanho  $n$ , denotado por  $(u_{1j}, \dots, u_{nj} | \theta_j)$ . Sob a suposição **(S2)**, apresentada na seção (2.3.2), as  $U_{jt}$  são estatisticamente independentes para todos os examinandos que possuem a mesma habilidade.

Pela construção formulada, há  $N$  vetores de respostas para cada examinando, resultando em uma matriz  $n$  por  $N$  denotada por  $\mathbf{U} = [u_{jt}]$ , sendo  $\theta$ , o vetor de  $N$  escores de habilidades dos examinandos  $(\theta_1, \dots, \theta_N)$ . Desta forma, a probabilidade da matriz  $n \times N$  de resposta dos itens é dada pela função de verossimilhança:

$$P(\mathbf{U} = u_{jt} | \theta) = \prod_{j=1}^N \prod_{t=1}^n P_t^{u_{jt}}(\theta_j) Q_t^{1-u_{jt}}(\theta_j). \quad (2.51)$$

Simplificando a notação, onde  $P_t(\theta_j) = P_{jt}$  e  $Q_t(\theta_j) = Q_{jt}$ , tem-se

$$P(\mathbf{U} = u_{jt} | \theta) = \prod_{j=1}^N \prod_{t=1}^n P_{jt}^{u_{jt}} Q_{jt}^{1-u_{jt}}. \quad (2.52)$$

Neste ponto, observa-se que a expressão (2.52), é a mesma apresentada em (2.35), de modo que o procedimento algébrico de estimação é o mesmo, assim, tomando o logaritmo da função de verossimilhança, tem-se

$$\mathbf{L} = \ln P(\mathbf{U} = u_{jt} | \theta) = \sum_{j=1}^N \sum_{t=1}^n [u_{jt} \ln P_{jt} + (1 - u_{jt}) \ln Q_{jt}]. \quad (2.53)$$

Tal como postulado na seção (2.2.3.2), as  $P_{jt}$  são funções da curva característica do item (CCI), e os desvios  $Z_{jt}$  (normal ou logístico), são expressos em termos dos parâmetros lineares dos itens  $\alpha_t$  e  $\beta_t$ ,  $Z_{jt} = \alpha_t + \beta_t \theta_{jt}$ . Como os parâmetros para todos os  $n$  itens e as  $N$  habilidades dos examinandos, são desconhecidos, ao tomar a derivada de  $\mathbf{L}$  com respeito a todos os parâmetros (habilidades e parâmetros dos itens), igualando-os a zero, e resolvendo as  $2n + N$  equações simultâneas, obtêm-se as estimativas de máxima verossimilhança dos parâmetros desconhecidos. Tomando as derivadas com respeito aos parâmetros dos itens, têm-se,

$$\frac{\partial \mathbf{L}}{\partial \alpha_t} = \sum_{j=1}^N u_{jt} \frac{1}{P_{jt}} \left( \frac{\partial P_{jt}}{\partial \alpha_t} \right) + \sum_{j=1}^N (1 - u_{jt}) \frac{1}{Q_{jt}} \left( \frac{\partial Q_{jt}}{\partial \alpha_t} \right), \quad (2.54)$$

$$\frac{\partial \mathbf{L}}{\partial \beta_t} = \sum_{j=1}^N u_{jt} \frac{1}{P_{jt}} \left( \frac{\partial P_{jt}}{\partial \beta_t} \right) + \sum_{j=1}^N (1 - u_{jt}) \frac{1}{Q_{jt}} \left( \frac{\partial Q_{jt}}{\partial \beta_t} \right), \quad (2.55)$$

e estas serão os  $n$  pares de cada derivada. A sumarização das equações (2.54) e (2.55) são feitas sob os  $N$  examinandos.

Tomar as derivadas da log-verossimilhança com respeito às habilidades dos examinandos, resulta em

$$\frac{\partial \mathbf{L}}{\partial \theta_{jt}} = \sum_{j=1}^N u_{jt} \frac{1}{P_{jt}} \left( \frac{\partial P_{jt}}{\partial \theta_{jt}} \right) + \sum_{j=1}^N (1 - u_{jt}) \frac{1}{Q_{jt}} \left( \frac{\partial Q_{jt}}{\partial \theta_{jt}} \right), \quad (2.56)$$



de forma que haverá uma derivada para cada um dos  $N$  parâmetros de habilidade. Nesta formulação, as derivadas de  $P_{jt}$  e  $Q_{jt}$  com respeito aos parâmetros serão dependentes da particular curva característica do item (CCI) empregada.

O método de Newton-Raphson iterativo é utilizado para obter as estimativas dos itens e das habilidades, de modo que, as segundas derivadas de ordem parcial da função de logverossimilhança com respeito aos parâmetros são necessárias. A equação de Newton-Raphson será da forma:

$$\mathbf{A}_{t+1} = \mathbf{A}_t + \mathbf{B}_t^{-1} \mathbf{F}_t, \quad (2.57)$$

onde  $\mathbf{A}$  é o vetor coluna da estimativa dos parâmetros dos itens e das habilidades, de tamanho  $2n + N$ ,  $\mathbf{B}$  é a matriz das derivadas de segunda ordem, de tamanho  $(2n + N) \times (2n + N)$ , dada pelos termos

$$\frac{\partial^2 \mathbf{L}}{\partial \alpha_t^2}, \quad \frac{\partial^2 \mathbf{L}}{\partial \beta_t^2}, \quad \frac{\partial^2 \mathbf{L}}{\partial \theta_{jt}^2}, \quad \frac{\partial^2 \mathbf{L}}{\partial \theta_{jt} \alpha_t} \quad \text{e} \quad \frac{\partial^2 \mathbf{L}}{\partial \theta_{jt} \beta_t}, \quad (2.58)$$

$\mathbf{F}$  é o vetor coluna das derivadas de primeira ordem da função de verossimilhança, de tamanho  $2n + N$  e  $t$  os índices de interação.

Devido à dimensionalidade de todos os termos da equação (2.57) ser de ordem  $2n + N$ , quando o número de itens ou o número de examinandos é muito grande, uma redução de dimensionalidade deve ser encontrada para tornar plausível a estimação. Em função da estrutura da matriz e de sua alta dimensionalidade quando a ordem dos seus termos a serem estimados atingiam limites além das capacidades computacionais normais da época em que os estudos foram realizados, três suposições foram feitas para obter as equações que forneciam soluções plausíveis:

1. *Desde que cada examinando é retirado de uma amostra aleatória da população, eles são independentes.* Esta suposição implicou na eliminação dos termos das derivadas cruzadas entre os pares de examinandos.
2. *Não havia razão para acreditar que deveria haver qualquer covariação entre os examinandos individuais e os parâmetros de um dado item.* Desta maneira, examinandos e itens foram tratados como independentes, eliminando  $(2n \times N)$  derivadas cruzadas correspondentes.
3. *Itens são independentes.* Elimina as correspondentes derivadas de segunda ordem, deixando somente submatrizes  $2 \times 2$  para cada item.

Mesmo após aplicar as três suposições citadas acima, a estrutura da matriz  $\mathbf{B}$  torna-se esparsa, contendo matrizes  $2 \times 2$  ao longo da diagonal principal. Neste ponto, foi proposto um procedimento em dois estágios (Birnbaum, 1968, pág. 420), conhecido por "Vai e Volta", no qual no primeiro estágio os parâmetros dos itens são estimados assumindo que as habilidades dos examinandos são conhecidas; e no segundo estágio as habilidades dos examinandos são estimadas supondo que os parâmetros dos itens são conhecidos. Para iniciar o processo, uma estimativa rude de cada habilidade dos examinandos é obtida. Em geral os escores padronizados do teste são utilizados como estimativa do valor "conhecido" inicial para os valores das habilidades (Wood *et al.*, 1976). De modo que para cada item haverá uma equação separada de escore de Fisher, dada por

$$\begin{bmatrix} \hat{\alpha}_t \\ \hat{\beta}_t \end{bmatrix}_{t+1} = \begin{bmatrix} \hat{\alpha}_t \\ \hat{\beta}_t \end{bmatrix}_t - \begin{bmatrix} E \left( \frac{\partial^2 \mathbf{L}}{\partial \alpha_t^2} \right) & E \left( \frac{\partial^2 \mathbf{L}}{\partial \alpha_t \beta_t} \right) \\ E \left( \frac{\partial^2 \mathbf{L}}{\partial \alpha_t \beta_t} \right) & E \left( \frac{\partial^2 \mathbf{L}}{\partial \beta_t^2} \right) \end{bmatrix}_t^{-1} \begin{bmatrix} \frac{\partial \mathbf{L}}{\partial \alpha_t} \\ \frac{\partial \mathbf{L}}{\partial \beta_t} \end{bmatrix}_t. \quad (2.59)$$

Tais equações são resolvidas individualmente para cada um dos  $n$  itens. No segundo estágio estes parâmetros estimados dos itens são considerados “verdadeiros” e o processo prossegue com a estimação, para cada um dos  $N$  examinandos, associando uma nova equação de Newton-Raphson para resolver as estimativas de máxima verossimilhança das habilidades, onde os dados observados das derivadas de segundo ordem são trocadas por suas respectivas esperanças sob o método de escore de Fisher, gerando as seguintes equações:

$$[\hat{\theta}_j]_{t+1} = [\hat{\theta}_j]_t - \left[ E \left( \frac{\partial^2 L}{\partial \theta_j^2} \right) \right]_t^{-1} \left[ \frac{\partial L}{\partial \theta_j} \right]_t. \quad (2.60)$$

Após este processo de estimação ser executado para cada examinando, têm-se um vetor de estimativas  $\hat{\theta}_j$  de máxima verossimilhança de tamanho  $N$ . Neste ponto, um ciclo é completado, de forma que as estimativas iniciais das habilidade dos examinandos, inicializadas de modo grosseiro, foram trocadas pelo conjunto de estimativas de máxima verossimilhança das habilidades  $\hat{\theta}_j$ . Adicionalmente, o conjunto de estimativas dos parâmetros dos itens estão disponíveis para cada um dos  $n$  itens.

Em função do processo iterativo de estimação conjunta dos parâmetros, o procedimento executado até este ponto é conhecido por *paradigma de Birnbaum*. A métrica das habilidades estimadas  $\hat{\theta}_j$  é única somente até a transformação linear, que pode ser através da ogiva normal ou logística. Em TRI isto é conhecido por “*problema de identificação*”, desde que a locação e a escala desta métrica é desconhecida, havendo somente  $N - 2$  estimativas de  $\hat{\theta}_j$  independentes, mesmo considerando os  $N$  valores obtidos. Dessa forma, é necessário criar valores para a locação e escala das métricas de habilidades obtidas, isto é, “âncoras” das habilidades estimadas  $\hat{\theta}_j$ . O processo de ancoragem de habilidades no JMLE é apresentado por Baker e Kim (2004, págs. 96-99). Outros pontos adicionais sobre o *paradigma de Birnbaum* não são abordados neste texto, tais como: o uso de agrupamento no paradigma de estimação conjunta (Bock, 1972; Kolakowski e Bock, 1973; Wood *et al.*, 1976); limitando as estimativas (Wright e Panchapakesan, 1969; Wright, 1977) e dados faltosos (Lord, 1974).

### 2.3.2.3 Estimação de Máxima Verossimilhança Marginal (MMLE)

Sobre a estimação conjunta dos parâmetros, apresentada na seção 2.3.2.2, na literatura os parâmetros dos itens são também conhecidos por “*parâmetros estruturais*”, que são fixos em número pelo tamanho do teste, e os parâmetros dos examinandos por “*parâmetros incidentais*”, uma vez que dependem do tamanho da amostra. Segundo Andrade *et al.* (2000), as habilidades são denominadas de parâmetros incidentais, pois o número destes parâmetros ( $\theta_j$ ), cresce com o número de indivíduos. Já os parâmetros dos itens são denominados de parâmetros estruturais pois o número desses parâmetros não se altera quando a amostra cresce.

Em outro contexto, anterior à publicação do *paradigma de Birnbaum*, Neyman e Scott (1948) mostraram que quando parâmetros estruturais são estimados simultaneamente com parâmetros incidentais, as estimativas de máxima verossimilhança dos parâmetros estruturais não necessitam ser consistentes quando o tamanho de amostra aumenta. Se há estatísticas suficientes disponíveis para os parâmetros incidentais, o procedimento de máxima verossimilhança condicional (Andersen, 1972) pode ser utilizado para a estimação consistente dos parâmetros estruturais. Como resultado, um procedimento de estimação para modelos de TRI de 2 e 3 parâmetros que evitasse o problema de estimativas inconsistentes dos parâmetros estruturais, tornou-se de grande valia. Apesar de problemas na implementação, para a qual verificou-se funcionar para pequenos testes, o procedimento do MMLE foi inicialmente proposto a partir do artigo de Bock e Lieberman (1970), sendo posteriormente

reformulado em uma nova proposta teórica e computacionalmente mais acessível, a qual sob certas condições, é uma instância do algoritmo EM, de forma que o método passa a ser conhecido na literatura como método de estimação MMLE/EM (Bock e Aitkin, 1981).

Embora não seja considerada uma técnica de estimação bayesiana, o processo de marginalização incorpora o teorema de Bayes em seus cálculos. No MMLE, o estimador pontual da habilidade de um indivíduo ( $\hat{\theta}_j$ ), é trocado por uma distribuição que aloca os dados do  $j$ -ésimo indivíduo em toda a escala da habilidade na proporção da probabilidade de estarem em um determinado ponto ao longo da escala da habilidade. Assim, em vez de um simples valor de  $\hat{\theta}_j$  para um examinando, a probabilidade é descoberta a partir de um examinando que tem cada possível valor de  $\theta_j$  ao longo da escala de habilidade condicional ao vetor  $\mathbf{u}_j$  de resposta do item, ao vetor  $\boldsymbol{\zeta}$  de parâmetros dos itens e ao vetor de parâmetros  $\boldsymbol{\nu}$  da distribuição da habilidade da população. Em estatística bayesiana, esta probabilidade é conhecida como probabilidade à posteriori. Como a habilidade é assumida ser uma contínua, um gráfico desta probabilidade como função da habilidade, resulta em uma curva suave. Esta curva representa a sumarização da informação sobre  $\theta$  contida nos dados de resposta de itens observados para o examinando  $j$ . A probabilidade à posterior empregada neste caso é o veículo para libertar o processo de estimação do item de sua dependência sobre a estimação de cada habilidade dos examinandos, como é o caso no JMLE. Isto é visto ao empregar o teorema de Bayes para calcular a distribuição da habilidade à posteriori completa, dada por:

$$P(\theta_j | \mathbf{u}_j, \boldsymbol{\nu}, \boldsymbol{\zeta}) = \frac{P(\mathbf{u}_j | \theta_j, \boldsymbol{\zeta}) g(\theta_j | \boldsymbol{\nu})}{\int_{\mathbb{R}} P(\mathbf{u}_j | \theta_j, \boldsymbol{\zeta}) g(\theta_j | \boldsymbol{\nu}) d\theta_j}. \quad (2.61)$$

Onde pela suposição (S2), apresentada na seção (2.3.2), o primeiro termo do numerador é dado por

$$P(\mathbf{u}_j | \theta_j, \boldsymbol{\zeta}) = \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}} \quad (2.62)$$

que é a função de verossimilhança, sendo esta a probabilidade do vetor  $\mathbf{u}_j$  de resposta dos itens dos examinandos, condicional à habilidade  $\theta_j$  dos examinandos e aos parâmetros dos itens em  $\boldsymbol{\zeta}$ .

O segundo termo no numerador,  $g(\theta_j | \boldsymbol{\nu})$ , é a função de densidade de probabilidade da habilidade na população de examinandos. Uma vez que a forma desta distribuição é desconhecida antes dos dados serem coletados, o julgamento sobre qual distribuição utilizar é livre. Embora não seja uma regra, frequentemente utiliza-se a distribuição normal padrão. Esta distribuição, com vetor de parâmetros  $\boldsymbol{\nu}$ , define a probabilidade relativa dos valores de  $\theta_j$ . No contexto bayesiano isto é conhecido por *distribuição à priori*, de forma que para a formulação dada esta é a *distribuição a priori da habilidade*.

Segundo Baker e Kim (2004), uma suposição necessária é a de que a distribuição à priori da habilidade seja a mesma para todos os examinandos. Assim, o produto no numerador é a distribuição conjunta de  $\mathbf{u}_j$  e  $\theta_j$ . A expressão no denominador de (2.61) é a probabilidade marginal (não condicional) do vetor  $\mathbf{u}_j$  de resposta dos itens com respeito aos parâmetros dos itens e à densidade populacional das habilidades.

A distribuição da probabilidade à posteriori em (2.61) combina a informação da distribuição à priori da habilidade com a informação da função de verossimilhança. O que a distribuição à priori essencialmente faz é distribuir os dados, nos vetores de resposta dos itens, através da escala de habilidade proporcionalmente à probabilidade posteriori dos examinandos, em cada ponto da escala de habilidade.

### 2.3.2.4 A Solução de Bock & Lieberman

Para estimar os parâmetros dos itens, a solução inicial foi proposta por **Bock e Lieberman (1970)**, onde apesar do artigo original considerar uma CCI de uma ogiva normal, são apresentados a seguir os cálculos para o modelo logístico de 3 parâmetros  $(\alpha_t, \beta_t, c_t)$ .

Para fins de simplificação da notação, considera-se

$$P(\mathbf{u}_j) = \int P(\mathbf{u}_j|\theta_j, \boldsymbol{\zeta})g(\theta_j|\boldsymbol{\nu})d\theta_j. \quad (2.63)$$

De forma que a função de verossimilhança marginal é dada por

$$\mathbf{L} = \prod_{j=1}^N P(\mathbf{u}_j). \quad (2.64)$$

E para encontrar as equações de verossimilhança marginal para  $\alpha_t$ , toma-se

$$\frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) = 0.$$

Onde

$$\begin{aligned} \frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) &= \sum_{j=1}^N \frac{\partial}{\partial \alpha_t}[\ln P(\mathbf{u}_j)] \\ &= \sum_{j=1}^N [P(\mathbf{u}_j)]^{-1} \frac{\partial}{\partial \alpha_t} \left[ \int_{\mathbb{R}} P(\mathbf{u}_j|\theta_j, \boldsymbol{\zeta})g(\theta_j|\boldsymbol{\nu})d\theta_j \right]. \end{aligned} \quad (2.65)$$

Convenientemente, em razão de  $\theta_j$  poder ser visto como uma amostra aleatória de indivíduos de uma população, o índice  $j$  é retirado de  $\theta$  em (2.65), resultando na expressão

$$\frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) = \sum_{j=1}^N [P(\mathbf{u}_j)]^{-1} \int_{\mathbb{R}} \frac{\partial}{\partial \alpha_t} [P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})g(\theta|\boldsymbol{\nu})d\theta]. \quad (2.66)$$

Usando a relação

$$\frac{\partial}{\partial \alpha_t}[P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})] = \frac{\partial}{\partial \alpha_t}[\ln P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})]P(\mathbf{u}_j|\theta, \boldsymbol{\zeta}),$$

resultando em

$$\begin{aligned} \frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) &= \sum_{j=1}^N [P(\mathbf{u}_j)]^{-1} \int_{\mathbb{R}} \frac{\partial}{\partial \alpha_t}[\ln P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})]P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})g(\theta|\boldsymbol{\nu})d\theta \\ &= \sum_{j=1}^N \int_{\mathbb{R}} \frac{\partial}{\partial \alpha_t}[\ln P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})] \left[ \frac{P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})g(\theta|\boldsymbol{\nu})}{P(\mathbf{u}_j)} \right] d\theta. \end{aligned} \quad (2.67)$$

E substituindo (2.61) em (2.67), têm-se

$$\frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) = \sum_{j=1}^N \int_{\mathbb{R}} \frac{\partial}{\partial \alpha_t}[\ln P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})][P(\theta_j|\mathbf{u}_j, \boldsymbol{\nu}, \boldsymbol{\zeta})]d\theta. \quad (2.68)$$

Trocando  $P(\mathbf{u}_j|\theta, \boldsymbol{\zeta})$  na equação (2.68), pelo resultado obtido após a suposição de indepen-

dência local, obtida em (2.62), tem-se

$$\begin{aligned}
\frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) &= \sum_{j=1}^N \int_{\mathfrak{X}} \frac{\partial}{\partial \alpha_t} \left[ \ln \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] \times \\
&\quad [P(\theta | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})] d\theta \\
&= \sum_{j=1}^N \int_{\mathfrak{X}} \left[ \ln \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right]^{-1} \times \\
&\quad \frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] \times [P(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})] d\theta. \tag{2.69}
\end{aligned}$$

É necessário calcular a derivada do primeiro termo à direita em (2.69), de maneira que se obtém

$$\frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] = \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \frac{\partial}{\partial \alpha_t} [P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}}]. \tag{2.70}$$

Onde

$$\begin{aligned}
\frac{\partial}{\partial \alpha_t} [P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}}] &= \frac{\partial}{\partial \alpha_t} [P_t(\theta)^{u_{jt}}] Q_t(\theta)^{1-u_{jt}} + P_t(\theta)^{u_{jt}} \frac{\partial}{\partial \alpha_t} [Q_t(\theta)^{1-u_{jt}}] \\
&= u_{jt} P_t(\theta)^{u_{jt}-1} \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] Q_t(\theta)^{1-u_{jt}} + \\
&\quad P_t(\theta)^{u_{jt}} (1-u_{jt}) Q_t(\theta)^{1-u_{jt}-1} \left[ \frac{\partial Q_t(\theta)}{\partial \alpha_t} \right] \\
&= u_{jt} P_t(\theta)^{u_{jt}-1} \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] Q_t(\theta)^{1-u_{jt}} - \\
&\quad P_t(\theta)^{u_{jt}} (1-u_{jt}) Q_t(\theta)^{1-u_{jt}-1} \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] \\
&= \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] [P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}}] \left[ \frac{u_{jt}}{P_t(\theta)} - \frac{1-u_{jt}}{Q_t(\theta)} \right] \\
&= \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] [P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}}] \left[ \frac{u_{jt} - P_t(\theta)}{P_t(\theta) Q_t(\theta)} \right] \\
&= \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] \left[ \frac{u_{jt} - P_t(\theta)}{P_t(\theta) Q_t(\theta)} \right]. \tag{2.71}
\end{aligned}$$

Substituindo (2.71) em (2.70) têm-se

$$\frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] = \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] \left[ \frac{u_{jt} - P_t(\theta)}{P_t(\theta) Q_t(\theta)} \right]. \tag{2.72}$$

Por fim, com mais algumas manipulações algébricas, apresentadas no Apêndice (A), são

obtidas as equações de verossimilhança para  $(\alpha_t, \beta_t, c_t)$ :

$$\frac{\partial}{\partial \alpha_t}(\ln \mathbf{L}) = (1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} [u_{jt} - P_t(\theta_j)] W_{jt}(\theta_j - \beta_t) [P(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] d\theta_j = 0 \quad (2.73)$$

$$\frac{\partial}{\partial \beta_t}(\ln \mathbf{L}) = -\alpha_t(1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} [u_{jt} - P_t(\theta_j)] W_{jt} [P(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] d\theta_j = 0 \quad (2.74)$$

$$\frac{\partial}{\partial c_t}(\ln \mathbf{L}) = (1 - c_t)^{-1} \sum_{j=1}^N \int_{\mathfrak{R}} \left[ \frac{u_{jt} - P_t(\theta_j)}{P_t(\theta_j)} \right] [P(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] d\theta_j = 0, \quad (2.75)$$

as quais não possuem solução analítica.

### 2.3.2.5 Método da Quadratura

Um problema importante encontrado na implementação da estimação dos parâmetros dos itens, relaciona-se com as equações (2.73) a (2.75), pois envolvem integrais que não apresentam solução analítica. Por conta disso, algum meio deve ser encontrado para a solução (aproximação) numérica de uma integral.

Embora existam muitos métodos de aproximações de integrais, na TRI têm sido frequente a aplicação do método *Gauss-Hermite*, usualmente denominado de *método de quadratura gaussiana*.

Segundo Baker e Kim (2004), se  $g(\theta | \boldsymbol{\nu})$  é uma função contínua com integral finita, ela pode ser aproximada, para qualquer grau de precisão, por uma outra função que assume um número finito de pontos. Dessa forma, o problema de obter a integral de uma função contínua é substituído pela obtenção da soma das áreas de um número finito, em  $q$  de retângulos. Os pontos médios de cada retângulo,  $X_k$  ( $k = 1, \dots, q$ ), são denominados de *nós* (ou *pontos de quadratura*). Cada nó tem um peso  $A(X_k)$  associado que leva em conta a altura da função de densidade  $g(\theta | \boldsymbol{\nu})$  na vizinhança de  $X_k$  e a largura dos retângulos. Os valores  $X_k$  e  $A(X_k)$  são obtidos resolvendo um conjunto de equações que envolvem a função  $g(\theta | \boldsymbol{\nu})$  e o número de nós (Hildebrand, 1956, pág. 327-330). Uma tabela para  $X_k$  e  $A(X_k)$  relativa a função gaussiana pode ser encontrada em (Stroud e Secrest, 1966).

Bock e Lieberman (1970) indicam que para aproximar os valores de  $X_k$  e  $A(X_k)$  a uma distribuição gaussiana de  $g(\theta | \boldsymbol{\nu})$ , basta multiplicar os valores tabelados de  $X_k$  por  $\sqrt{2}$  e dividir os valores tabelados dos pesos dos nós de quadratura  $A(X_k)$  por  $\sqrt{\pi}$ .

Como exemplo de aplicação do método de quadratura, considera-se um indivíduo aleatoriamente amostrado (neste caso sem o índice  $j$ ) de uma população com  $g(\theta | \boldsymbol{\nu})$ , e, ao substituir a aproximação por quadratura, por uma integração na equação (2.73), resulta em

$$\alpha_t : (1 - c_t) \sum_{j=1}^N \sum_{k=1}^q [u_{jt} - P_t(X_k)] W_{tk}(X_k - \beta_t) [P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] = 0, \quad (2.76)$$

$$\beta_t : -\alpha_t(1 - c_t) \sum_{j=1}^N \sum_{k=1}^q [u_{jt} - P_t(X_k)] W_{tk} [P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] = 0, \quad (2.77)$$

$$c_t : (1 - c_t)^{-1} \sum_{j=1}^N \sum_{k=1}^q \left[ \frac{u_{jt} - P_t(X_k)}{P_t(X_k)} \right] [P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \nu)] = 0, \quad (2.78)$$

onde o termo de peso é dado por,

$$W_{tk} = \frac{P_t^*(X_k) Q_t^*(X_k)}{P_t(X_k) Q_t(X_k)}. \quad (2.79)$$

E assim, as equações (2.76) a (2.78) representam a forma de quadratura das equações de log-verossimilhança marginal para os modelos de 3 parâmetros. De forma que sobre a aproximação de Bock e Lieberman (1970), a técnica de Newton-Rapson é empregada para estimar  $3n$  parâmetros dos itens de um teste simultaneamente. Tal como no JMLE, os valores observados dos  $u_{jt}$  nas derivadas segundas da função de log-verossimilhança foram trocados pelos respectivos valores esperados, permitindo assim que o método de Fisher fosse utilizado.

### 2.3.3 Estimação Bayesiana

Nesta seção, apresenta-se uma visão geral dos procedimentos utilizados para estimação das habilidades, uma vez que estas serão o foco do próximo capítulo. Lá serão considerados, entre outras coisas, o processo de calibração dos itens, já conhecido e realizado, para então se discutir a prática dos métodos de seleção de itens propostos no CAT. Assim, o material tratado nesta seção faz considerações sobre os métodos apresentados na seção anterior, além de apresentar procedimentos de estimação reconhecidamente bayesianos, em particular os métodos EAP e MAP, empregados para estimação das habilidades dos examinandos em diferentes métodos de seleção de itens do CAT. Neste sentido, a importância dos métodos abordados nas seções (2.3.2.4) e (2.3.2.5), além do material adicional disponibilizado no Apêndice (A), que compõe os aspectos matemáticos e metodológicos necessários para a compreensão do contexto seguinte.

Neste sentido, a experiência com o procedimento JMLE mostrou que há um número inerente de problemas. Entre eles, as estimativas dos parâmetros de habilidades que podem assumir valores não razoáveis e uma incapacidade de calcular estimativas de habilidades para padrões não usuais de respostas dos examinandos, por exemplo, quando todos os itens são respondidos corretamente ou incorretamente. Do ponto de vista teórico, o problema principal é que o uso do JMLE para testes de tamanho finito podem resultar em estimativas que não são estatisticamente consistentes.

O procedimento de aproximação apresentado por Bock e Aitkin (1981) utilizando algoritmo EM resolve o problema de inconsistência para as estimativas de parâmetros dos itens. A estimação de máxima verossimilhança marginal se distingue do JMLE pela suposição de que a habilidade dos examinandos tem uma distribuição em uma população. E assim, quando esta distribuição é conhecida ou pode ser estimada, permite a integridade da função de verossimilhança. Embora o procedimento MMLE/EM resolva o problema da inconsistência das estimativas dos parâmetros dos itens, os problemas de desvios dos valores das estimativas dos parâmetros dos itens em alguns conjuntos de dados e a falta de um significado da estimação das habilidades dos examinandos para padrões não usuais de respostas dos itens, permanece.

Neste sentido, uma forma de prevenir a ocorrência de tais desvios é o uso de procedimentos de estimação baseados na aproximação bayesiana. Um aproximação para aplicar estes métodos, em particular para estimação em TRI focada em modelos de 3 parâmetros, é devida ao trabalho de Swaminathan e Gifford (1986) que utilizam o modelo logístico para CCI. Apesar das diferenças entre as escolas “frequentistas” e “bayesianas” quanto ao emprego de métodos estatísticos, os pesquisadores de TRI tiveram que adotar uma postura mais pragmática em relação aos métodos bayesianos, uma vez que estes começavam a ser vistos como significantes melhorias na estimação dos parâmetros. Desta forma, os métodos apresentados nesta seção podem ser considerados uma extensão dos procedimentos MMLE, em particular os implementados via algoritmo EM.



### 2.3.3.1 Estimação Bayesiana de Parâmetros dos Itens

Para ilustrar o uso do teorema de Bayes em TRI, suponha que o objetivo seja estimar parâmetros de itens quando parâmetros de habilidades sejam conhecidos. Assim, considere uma função contínua  $g(\zeta|\boldsymbol{\eta})$  representar a distribuição de probabilidade que reflita a crença à priori quanto à distribuição dos possíveis valores do  $i$ -ésimo parâmetro dos itens, condicional aos parâmetros em  $\boldsymbol{\eta}$ . Assuma que cada um dos  $n$  itens têm distribuições à priori de mesma forma. Considere  $U$  uma matriz  $N \times n$  de resposta de itens e  $\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta})$  a função de verossimilhança da amostra de resposta de itens condicionais aos parâmetros em  $\boldsymbol{\zeta}$  e  $\boldsymbol{\theta}$ . De forma que a distribuição à posteriori ao longo dos itens e examinandos pode ser expressa por

$$g(\zeta|\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta})g(\zeta|\boldsymbol{\eta}). \quad (2.80)$$

É de interesse, neste momento, contrastar a equação (2.80) com a equação bayesiana utilizada no procedimento MMLE, apresentada em (2.61). No procedimento anterior nota-se que foi utilizada uma “*igualdade*”, onde agora é considerada uma “*proporcionalidade*”. O termo do denominador em (2.61) é agora tratado como uma constante de proporcionalidade. Adicionalmente, a equação (2.80) define uma distribuição à posteriori para os parâmetros dos itens, diferentemente da situação anterior, onde a distribuição à posterior foi pensada para os parâmetros de habilidades. Assim, segundo Baker e Kim (2004), função de verossimilhança  $\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta})$  é definida como

$$\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta}) = \prod_{j=1}^N \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}} = \prod_{j=1}^N P(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta}), \quad (2.81)$$

onde  $P(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta})$  é a probabilidade de um vetor de respostas do  $j$ -ésimo indivíduo, condicional ao valor conhecido de  $\theta_j$  e aos parâmetros dos itens em  $\boldsymbol{\zeta}$ . E esta é a mesma verossimilhança que serviu como base para os procedimentos JMLE e MMLE, descritos anteriormente. Inferências sobre os parâmetros desconhecidos dos itens em  $\boldsymbol{\zeta}$  tipicamente tomam a forma de estimativas pontuais que maximizam a probabilidade à posteriori  $g(\zeta|\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\eta})$  com respeito aos parâmetros desconhecidos. Em estimação bayesiana, isto é frequentemente a moda estimada da distribuição de probabilidade a posteriori, sendo o resultado estimado tipicamente conhecido como *estimador bayesiano modal* ou BME (Mislevy e Stocking, 1989; Swaminathan e Gifford, 1986).

No contexto bayesiano, a quantidade de “contribuição” de cada um dos termos da equação (2.80) determina as características da distribuição à posteriori. Assim, a contribuição relativa da distribuição a priori  $g(\zeta|\boldsymbol{\eta})$  e da verossimilhança  $\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta})$  para  $g(\zeta|\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\eta})$ , e assim para as estimativas, são uma característica importante. Se o número de examinandos é grande e a contribuição da distribuição de probabilidade a priori é pequena, a verossimilhança irá dominar a distribuição de probabilidade a posteriori. Isto é, as estimativas dos parâmetros dos itens obtidas de  $g(\zeta|\mathbf{U}, \boldsymbol{\theta}, \boldsymbol{\eta})$  na equação (2.80) dependerão quase totalmente dos dados das respostas dos itens observados, representados por  $\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\theta})$ , de forma que  $g(\zeta|\boldsymbol{\eta})$  terá pouco efeito sobre as estimativas. Neste caso, os BME’s são provavelmente quase idênticos às estimativas obtidas pela estimação via máxima verossimilhança. Se a contribuição da distribuição de probabilidade a priori é substancial, os BME’s diferem dos EMV’s.



### 2.3.3.2 Estimação Bayesiana Marginalizada

Como uma extensão da solução de Bock & Aitkin, apresentada no Apêndice (B), Mislevy (1986) apresentou um procedimento para estimação de parâmetros dos itens que empregou procedimento bayesiano clássico em dois estágios, atribuído à Lindley e Smith (1972) no qual a informação a priori é especificada de modo hierárquico. Inicialmente têm-se a densidade conjunta de todos os parâmetros antes da coleta dos dados, de modo que estes são assumidos serem independentes e provenientes de variáveis aleatórias contínuas com distribuição de probabilidade especificada por

$$g(\theta, \nu, \zeta, \eta) = \prod_{j=1}^N g(\theta_j | \nu) g(\nu) \prod_{t=1}^n g(\zeta_t | \eta) g(\eta), \quad (2.82)$$

onde o termo  $g(\theta_j | \nu)$  representa a distribuição de probabilidade do parâmetro de habilidade ( $\theta_j$ ) de um examinando e é condicional aos parâmetros populacionais da distribuição da habilidade contida no vetor  $\nu$ . Uma vez que as habilidades são assumidas serem independentes e identicamente distribuídas,  $\nu$  contém a média ( $\mu_\theta$ ) e a variância ( $\sigma_\theta^2$ ) desta distribuição de habilidade a priori.

Segundo Lindley e Smith (1972), no modelo hierárquico os  $\theta_j$  são parâmetros e os parâmetros populacionais  $\mu_\theta$  e  $\sigma_\theta^2$  são conhecidos como *hiperparâmetros*. De modo que estes podem ser tratados como variáveis aleatórias com distribuição de probabilidade, que neste caso é denotada por  $g(\nu)$ .

Ainda sobre a equação (2.82), o termo  $g(\zeta_t | \eta)$  representa a distribuição de probabilidade para os parâmetros do  $t$ -ésimo item contido no vetor  $\zeta_t$ , condicional aos parâmetros populacionais no vetor  $\eta$ . Tal como no caso das habilidades, os parâmetros  $\alpha_t$ ,  $\beta_t$  e  $c_t$ , contidos em  $\zeta_t$  serão referenciados como parâmetros, e  $\eta$  como o vetor contendo os parâmetros populacionais, ou hiperparâmetros, para o item.

Assim, para fazer inferência sobre todos os parâmetros desconhecidos, no caso,  $\theta$ ,  $\nu$ ,  $\zeta$  e  $\eta$ , após os dados serem coletados, a distribuição a posteriori ao longo de todos os itens e examinandos obtida pela aplicação do teorema de Bayes é dada por

$$g(\theta, \nu, \zeta, \eta | \mathbf{U}) \propto L(\mathbf{U} | \theta, \zeta) g(\theta | \nu) g(\nu) g(\zeta | \eta) g(\eta), \quad (2.83)$$

onde a especificação da informação à priori é realizada através dos parâmetros, neste caso hiperparâmetros, das distribuições de probabilidade  $g(\theta | \nu)$  e  $g(\zeta | \eta)$  que constituem o primeiro estágio no modelo de Lindley e Smith (1972). O segundo estágio é constituído pela especificação dos parâmetros das distribuições de probabilidade dos hiperparâmetros, neste caso, pelos parâmetros que caracterizam  $g(\nu)$  e  $g(\eta)$ .

Segundo Mislevy (1986), a equação (2.83) contém toda a informação disponível sobre os parâmetros desconhecidos. Para estimar os parâmetros dos itens em (2.83), a marginalização aproximada de Bock e Lieberman (1970), apresentada na seção (2.3.2.4) pode ser utilizada. Sobre esta aproximação, a função de verossimilhança é marginalizada com respeito à habilidade. Em geral, a escolha das variáveis para marginalizar é ditada pela distinção entre os parâmetros que se deseja estimar e aqueles de nenhum interesse.

Baker e Kim (2004) destacam que se a CCI e distribuição de probabilidade à priori das habilidades forem corretamente especificadas, as estimativas dos parâmetros (para testes de comprimento finito) serão aproximadamente próximas de seus valores verdadeiros à medida que o número de examinandos cresce. Mislevy (1986) sugere que a distribuição dos  $\eta$ , em muitas avaliações educacionais, não são especialmente de interesse, de modo que é apropriado que se considere os hiperparâmetros de  $\eta$  como parâmetros de perturbação. Na equação (2.83)

a distribuição dos parâmetros de perturbação podem ser removidos integrando sobre suas distribuições de probabilidades.

Para encontrar a distribuição de probabilidade a posteriori que é apropriada para fazer inferências sobre os parâmetros dos itens, e os parâmetros da distribuição da habilidade na população de examinandos, integra-se o termo à direita em (2.83), sobre as distribuições de probabilidade das habilidades  $g(\theta|\boldsymbol{\nu})$  com respeito a  $\theta$  e os parâmetros populacionais dos itens  $g(\eta)$  com respeito a  $\eta$ , resultando em uma distribuição marginalizada a posteriori da forma

$$\begin{aligned} g(\zeta, \boldsymbol{\nu}|\mathbf{U}) &\propto \int \int \mathbf{L}(U|\boldsymbol{\theta}, \boldsymbol{\zeta})g(\theta|\boldsymbol{\nu})g(\nu)g(\zeta|\boldsymbol{\eta})g(\eta)d\theta d\eta \\ &\propto \mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\nu})g(\zeta)g(\nu), \end{aligned} \quad (2.84)$$

onde  $\mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\nu})$  é a *verossimilhança marginal* resultante da integração de  $\mathbf{L}(U|\boldsymbol{\theta}, \boldsymbol{\zeta})$  com respeito à habilidade. É possível notar que a integração sobre a distribuição populacional das habilidades eliminou a dependência de estimativas dos parâmetros dos itens sobre as estimativas das habilidades dos examinandos individuais. Entretanto, a verossimilhança marginal está ainda condicionada aos hiperparâmetros  $\mu_\theta$  e  $\sigma_\theta^2$  da distribuição populacional das habilidades e assim,  $g(\nu)$  e estes valores de hiperparâmetros devem ser especificados. De qualquer forma, integrar sobre a distribuição populacional dos parâmetros dos itens em (2.84) não elimina a necessidade de especificar valores para os hiperparâmetros  $\eta$  em  $g(\zeta)$ .

Sob a suposição de independência entre os itens, a estimação pode ser feita um item por vez. Dessa maneira, para estimar os parâmetros de um item, deve-se derivar a equação (2.84) com respeito aos parâmetros dos itens e igualar o conjunto de equações a zero. Por conveniência, pode-se trabalhar com o logaritmo da distribuição de probabilidade a posteriori marginal. Logo, o sistema de equações para estimação bayesiana dos parâmetros dos itens é dado por:

$$\frac{\partial}{\partial v_t}[\ln \mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\nu})] + \frac{\partial}{\partial v_t}[\ln g(\zeta)] + \frac{\partial}{\partial v_t}[\ln g(\nu)] = 0, \quad (2.85)$$

onde  $v_t$  representa um parâmetro associado com o  $t$ -ésimo item, por exemplo,  $\alpha_t$ . O sistema de equações de estimativa representado pela equação (2.85) constitui as equações modais bayesianas marginalizadas nas quais os BME's representam a moda conjunta da distribuição a posteriori. Uma vez que  $g(\nu)$  não contém parâmetros de itens, sua derivada com respeito à  $v_i$  é zero. Assim,  $g(\nu)$  pode ser retirada da equação (2.85), resultando no seguinte sistema de equações para obtenção dos BME's dos parâmetros dos itens:

$$\underbrace{\frac{\partial}{\partial v_t}[\ln \mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\nu})]}_{\text{Verossimilhança}} + \underbrace{\frac{\partial}{\partial v_t}[\ln g(\zeta)]}_{\text{Distribuição à priori}} = 0. \quad (2.86)$$

As parcelas envolvidas na equação (2.86) merecem atenção especial. Enquanto a primeira refere-se à função de verossimilhança, a segunda está associada às distribuições a priori dos parâmetros dos itens. Como cada componente dessa equação envolve elementos distintos, de forma que serão descritos separadamente, e conectados mais adiante para obter o conjunto final de solução das equações.

### 2.3.3.3 O Componente da Verossimilhança

Nesta seção, trata-se o primeiro termo da equação (2.86), onde será considerado a CCI do modelo logístico de 3 parâmetros. Sendo assim, considera-se

$$\mathbf{L} = \mathbf{L}(U|\boldsymbol{\zeta}, \boldsymbol{\nu}) = \prod_{j=1}^N P(U_j|\boldsymbol{\zeta}, \boldsymbol{\nu}), \quad (2.87)$$

onde  $P(U_j|\boldsymbol{\zeta}, \boldsymbol{\nu}) = \int P(U_j|\boldsymbol{\theta}, \boldsymbol{\zeta})g(\boldsymbol{\theta}|\boldsymbol{\nu})d\boldsymbol{\theta}$  é a probabilidade marginalizada de um vetor de resposta de itens  $U_j$  com respeito aos parâmetros  $\boldsymbol{\zeta}$ . Substituindo (2.87) em (2.86), e tomando o logaritmo do primeiro termo, tem-se

$$\frac{\partial}{\partial v_t} \left[ \sum_{j=1}^N \ln P(U_j|\boldsymbol{\zeta}, \boldsymbol{\nu}) \right]. \quad (2.88)$$

Por conveniência, em função dos cálculos envolvidos para as três equações de solução serem similares, apresenta-se apenas o cálculo para o parâmetro de discriminação ( $\alpha_t$ ). Inicialmente considere a transformação<sup>2</sup>  $a_t = \ln \alpha_t$ , da qual segue  $\alpha_t = e^{a_t}$ . Utilizando a transformação para reescrever a CCI do modelo de 3 parâmetros, tem-se

$$\begin{aligned} P_t(\theta_j) &= c_t + (1 - c_t) \left\{ \frac{\exp[e^{a_t}(\theta_{jt} - \beta_t)]}{1 + \exp[e^{a_t}(\theta_{jt} - \beta_t)]} \right\} \\ &= c_t + (1 - c_t)P_t^*(\theta_j). \end{aligned} \quad (2.89)$$

A derivada parcial da equação (2.88) com respeito ao parâmetro de discriminação transformado é obtido inicialmente e é dado por:

$$\begin{aligned} \frac{\partial}{\partial a_t} [\ln \mathbf{L}] &= e^{a_t}(1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} \frac{u_{jt} - P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \times \\ &= (\theta_{jt} - \beta_t)P_t^*(\theta_j)Q_t^*(\theta_j)[P(\theta_j|\mathbf{U}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})]d\theta. \end{aligned} \quad (2.90)$$

De modo que considerando o peso ( $W_{jt}$ ), definido em (2.39), a componente de verossimilhança marginalizada produz as seguintes equações:

$$a_t : e^{a_t}(1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} [u_{jt} - P_t(\theta_j)]W_{jt}(\theta_{jt} - \beta_t)[P(\theta_j|\mathbf{U}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})]d\theta \quad (2.91)$$

$$b_t : -e^{a_t}(1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} [u_{jt} - P_t(\theta_j)]W_{jt}[P(\theta_j|\mathbf{U}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})]d\theta \quad (2.92)$$

$$c_t : (1 - c_t)^{-1} \sum_{j=1}^N \int_{\mathfrak{R}} \left[ \frac{u_{jt} - P_t(\theta_j)}{P_t(\theta_j)} \right] [P(\theta_j|\mathbf{U}_j, \boldsymbol{\zeta}, \boldsymbol{\nu})]d\theta. \quad (2.93)$$

Exceto pelo multiplicador  $e^{a_t}$ , as equações (2.91) a (2.93) são iguais às equações (2.73) a (2.75). Em razão da marginalização, todas as três equações acima envolvem uma integração com respeito à habilidade, de modo que podem ser reescritas no formato de quadratura gaussiana. Considerando os  $q$  nós de quadratura, o número esperado de examinandos  $\bar{f}_{tk}$

<sup>2</sup>A transformação resulta em cada  $a_t$  ter uma distribuição normal a priori com densidade que é proporcional a  $\exp[-0.5(a_t - \mu_a)^2/\sigma_a^2]$ . A distribuição normal a priori de cada  $a_t$  é definida por seus hiperparâmetros,  $\mu_a$  e  $\sigma_a$ . Desde que a suposição de intercambialidade seja empregada, estas distribuições a priori são todas assumidas ter a mesma forma e os mesmos valores de hiperparâmetros

e o número esperado de respostas corretas  $\bar{r}_{tk}$ , os argumentos para o uso da quadratura numérica são os mesmos utilizados no apêndice (B), lembrando apenas que a quadratura é tomada agora para a verossimilhança de  $U_j$  condicional sobre  $\theta_j = X_k$  e os parâmetros dos itens, de forma que reescrevendo as equações (2.91) a (2.93) em sua forma de quadratura numérica, têm-se

$$a_t : e^{a_t}(1 - c_t) \sum_{k=1}^q [\bar{r}_{tk} - \bar{f}_{tk}P_t(X_k)]W_{tk}(X_k - \beta_t) \quad (2.94)$$

$$\beta_t : -e^{a_t}(1 - c_t) \sum_{k=1}^q [\bar{r}_{tk} - \bar{f}_{tk}P_t(X_k)]W_{tk} \quad (2.95)$$

$$c_t : (1 - c_t)^{-1} \sum_{k=1}^q \left[ \frac{\bar{r}_{tk} - \bar{f}_{tk}P_t(X_k)}{P_t(X_k)} \right]. \quad (2.96)$$

Entretanto, em uma solução bayesiana, as componentes correspondentes à distribuição de probabilidade a priori, o segundo termo na equação (2.86), necessitam ser acrescentados aos componentes da verossimilhança marginal. Estes termos serão apresentados na próxima seção.

### 2.3.3.4 A Componente da Distribuição à Priori

Segundo Araujo *et al.* (2009), uma das principais características da Estatística Bayesiana consiste na idéia de que os parâmetros possuem distribuições, ao contrário da análise frequentista em que os parâmetros são constantes. Como o parâmetro  $\alpha_t$  deve ser positivo,  $\beta_t$  pode assumir qualquer valor real e  $c_i$  deve estar no intervalo  $[0; 1]$ , assume-se distribuições para esses parâmetros levando em conta cada limitação. A escolha da distribuição a priori dos parâmetros dos itens é arbitrária, embora um número limitado de funções seja empregado na prática. Além disso, faz-se necessário especificar valores numéricos para os hiperparâmetros das distribuições o que, em geral, é realizado de maneira subjetiva (Baker e Kim, 2004). A seguir, serão consideradas as suposições mais frequentes na literatura.

As distribuições mais adotadas para o parâmetro  $\alpha_t$  são: Log-Normal e Qui-Quadrado. A justificativa teórica para o uso dessas distribuições residem no fato de que valores de  $\alpha_t$  são tipicamente maiores que zero, sugerindo que a distribuição de  $\alpha_t$  pode ser modelada por uma distribuição unimodal e positivamente assimétrica (Mislevy, 1986). Em geral, como o parâmetro  $\alpha_t$  deve ser positivo, assume-se a distribuição Log-Normal com parâmetros  $\nu_\alpha = (\mu_\alpha, \sigma_\alpha^2)$ .

A densidade do parâmetro  $\alpha_t$ , sob suposição de que esse pertence à distribuição Log-Normal, é expressa por:

$$f(\alpha_t | \mu_\alpha, \sigma_\alpha^2) = \frac{1}{\alpha_t \sqrt{2\pi\sigma_\alpha^2}} \exp \left[ -\frac{(\ln \alpha_t - \mu_\alpha)^2}{2\sigma_\alpha^2} \right].$$

Assim, a derivada do logaritmo da função associada ao parâmetro  $\alpha_t$  será igual a:

$$\frac{\partial \ln f(\alpha_t | \mu_\alpha, \sigma_\alpha^2)}{\partial \alpha_t} = -\frac{1}{\alpha_t} \left[ 1 + \frac{(\ln \alpha_t - \mu_\alpha)}{\sigma_\alpha^2} \right]. \quad (2.97)$$

Como o parâmetro de dificuldade do item pertence ao intervalo  $-\infty < \beta_t < +\infty$  e este está medido na mesma escala de distribuição das habilidades, pode-se adotar a distribuição Normal com vetor de parâmetros  $\nu_\beta = (\mu_\beta, \sigma_\beta^2)$ .

Sabe-se que a densidade da Normal para  $\beta_t$  é expressa por:

$$f(\beta_t | \boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta^2) = \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp \left[ -\frac{(\beta_t - \mu_\beta)^2}{2\sigma_\beta^2} \right].$$

Logo, a derivada do logaritmo da função associada ao parâmetro  $\beta_t$  será igual a:

$$\frac{\partial \ln f(\beta_t | \boldsymbol{\mu}_\beta, \boldsymbol{\sigma}_\beta^2)}{\partial \beta_t} = -\frac{(\beta_t - \mu_\beta)}{\sigma_\beta^2}. \quad (2.98)$$

Como  $c_t$  representa a probabilidade de acerto ao acaso, seu valor só pode pertencer ao intervalo  $[0; 1]$ . Dessa maneira, [Swaminathan e Gifford \(1986\)](#) sugeriram uma priori *Beta* para esse parâmetro. A função densidade da Beta com parâmetros  $\nu_c = (\gamma_c - 1, \delta_c - 1)$  é dada por:

$$f(c_t | \gamma_c, \delta_c) = \frac{\Gamma(\gamma_c + \delta_c - 2)}{\Gamma(\gamma_c - 1)\Gamma(\delta_c - 1)} c_t^{\gamma_c - 2} (1 - c_t)^{\delta_c - 2}.$$

Logo, a derivada do logaritmo da função associada ao parâmetro  $c$  será igual a:

$$\frac{\partial \ln f(c_t | \gamma_c, \delta_c)}{\partial c_t} = \frac{(\gamma_c - 2)}{c_t} - \frac{(\delta_c - 2)}{1 - c_t}. \quad (2.99)$$

### 2.3.4 Estimação Bayesiana via Algoritmo EM

Com as componentes derivadas das equações de verossimilhança em forma de quadratura, disponíveis nas equações (2.94) a (2.96), em conjunto com as componentes das distribuições à priori dos parâmetros, apresentadas nas equações (2.97) a (2.99), as equações de estimação da *Moda a Posteriori* são definidas por:

$$a_t : D(1 - c_t) \sum_{k=1}^K \sum_{i=1}^q [(\bar{r}_{kit} - P_{it} \bar{f}_{kit})(X_i - b_t) W_{it}] - \frac{1}{a_t} \left[ 1 + \frac{(\ln a_t - \mu_a)}{\sigma_a^2} \right] = 0 \quad (2.100)$$

$$b_t : -D a_t (1 - c_t) \sum_{k=1}^K \sum_{i=1}^q [(\bar{r}_{kit} - P_{it} \bar{f}_{kit}) W_{it}] - \frac{(b_t - \mu_b)}{\sigma_b^2} = 0; \quad (2.101)$$

$$c_t : \sum_{k=1}^K \sum_{i=1}^q \left[ (\bar{r}_{kit} - P_{it} \bar{f}_{kit}) \frac{W_{it}}{P_{it}^*} \right] + \frac{(\alpha_c - 2)}{c_t} - \frac{(\beta_c - 2)}{1 - c_t} = 0. \quad (2.102)$$

Como os dados artificiais ( $\bar{f}_{kit}$  e  $\bar{r}_{kit}$ ) são desconhecidos, faz-se necessário estimar essas quantidades de uma maneira iterativa, juntamente com os parâmetros dos itens. Dessa maneira [Bock e Aitkin \(1981\)](#) propuseram uma adaptação do algoritmo EM de ([Dempster et al., 1977](#)).

#### 2.3.4.1 Algoritmo EM

O algoritmo EM aplicado à TRI consiste em um processo iterativo para se obter estimativas dos parâmetros dos itens na presença de variáveis não-observáveis (proficiência dos examinandos). Para melhor entendimento sobre o algoritmo EM, esse procedimento será descrito brevemente.

Suponha que as proficiências estejam restritas a um conjunto de  $i = 1, \dots, q$  valores  $\bar{\theta}_i$ , com probabilidades  $\pi_1, \dots, \pi_q$ . Essa suposição pode ser feita porque as aproximações de

integrais são realizadas por métodos de quadratura e os valores  $\bar{\theta}_i$  corresponderão aos pontos de quadraturas (anteriormente representados por  $X_i$ ). Defina também  $f_k = (f_{k1t}, \dots, f_{kqt})'$ , como o número de examinandos do grupo  $k$  de cada nível  $q$  de proficiência para o  $t$ -ésimo item e  $r_k = (r_{k1t}, \dots, r_{kqt})'$ , o número de respostas corretas do grupo  $k$  para cada valor  $q$  de proficiência para o item  $t$ . Se as proficiências dos  $N_k$  examinandos da população  $k$  que responderam ao item  $t$  são selecionadas aleatoriamente da população, a probabilidade conjunta que os  $f_{kit}$  indivíduos tenham proficiência  $\bar{\theta}_i$ , segue uma distribuição *Produto-Multinomial* dada por:

$$P(F_t = f_t | \boldsymbol{\pi}) \equiv P(f_t | \boldsymbol{\pi}) = \prod_{k=1}^K \left\{ \frac{N_k!}{\prod_{i=1}^q f_{kit}!} \prod_{i=1}^q \pi_i^{f_{kit}} \right\}, \quad t = 1, \dots, n_k. \quad (2.103)$$

Dados  $f_{kit}$  e  $\bar{\theta}_i$ , a probabilidade de se obter  $r_{kit}$  acertos ao item  $t$  dentre as  $f_{kit}$  respostas dos examinados do grupo  $k$  com habilidade  $\bar{\theta}_i$  será:

$$P(R_{kit} = r_{kit} | \mathbf{f}_{kit}, \bar{\boldsymbol{\theta}}_i) \equiv P(r_{kit} | \mathbf{f}_{kit}, \bar{\boldsymbol{\theta}}_i) = \prod_{k=1}^K \left\{ \binom{f_{kit}}{r_{kit}} P_{kit}^{r_{kit}} Q_{kit}^{f_{kit}-r_{kit}} \right\}, \quad (2.104)$$

onde:  $P_{kit} = c_t + \frac{(1 - c_t)}{1 + \exp[-Da_t(\bar{\theta}_{it} - b_t)]}$  e  $Q_{kit} = 1 - P_{kit}$ .

A probabilidade conjunta dos vetores  $f$  e  $r$ , dados  $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \dots, \bar{\theta}_q)'$  e  $\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_q$ , é:

$$\begin{aligned} P(f, r | \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) &= P(r | \mathbf{f}, \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) P(f | \bar{\boldsymbol{\theta}}, \boldsymbol{\pi}) = P(r | \mathbf{f}, \bar{\boldsymbol{\theta}}) P(f | \boldsymbol{\pi}) = \\ &= \prod_{k=1}^K \left\{ \left[ \prod_{i=1}^q \prod_{t=1}^n \binom{f_{kit}}{r_{kit}} P_{kit}^{r_{kit}} Q_{kit}^{f_{kit}-r_{kit}} \right] \times \left[ \frac{N_k!}{\prod_{i=1}^q f_{kit}!} \prod_{i=1}^q \pi_i^{f_{kit}} \right] \right\}. \end{aligned} \quad (2.105)$$

Pelo critério da fatoração, tem-se que  $(f, r)$  é uma estatística suficiente para os dados completos  $(U, \theta)$  (Baker e Kim, 2004).

Ignorando alguns termos constantes em relação ao parâmetro dos itens,  $b$ , a função logaritmo da Verossimilhança para os dados completos pode ser escrita da seguinte forma:

$$\ln L(b) \propto \sum_{k=1}^K \sum_{i=1}^q \sum_{t=1}^n [r_{kit} \ln P_{kit} + (f_{kit} - r_{kit}) \ln Q_{kit}]. \quad (2.106)$$

Embora  $(f, r)$  sejam quantidades não-observáveis, pode-se estimar os parâmetros dos itens por meio da esperança do logaritmo da verossimilhança condicionada a  $u_{\dots}$  e  $b$  e das seguintes definições (Baker e Kim, 2004):

$$\bar{r}_{kit} = E[r_{kit} | u_{\dots}, b] = \sum_{j=1}^{N_k} \left[ \frac{P(U_{kj} | \mathbf{b}, \mathbf{X}_i) A_i}{\sum_{i=1}^q P(U_{kj} | \mathbf{b}, \mathbf{X}_i) A_i} \right]. \quad (2.107)$$

$$\bar{f}_{kit} = E[f_{kit} | u_{\dots}, b] = \sum_{j=1}^{N_k} \left[ u_{kit} \frac{P(U_{kj} | \mathbf{b}, \mathbf{X}_i) A_i}{\sum_{i=1}^q P(U_{kj} | \mathbf{b}, \mathbf{X}_i) A_i} \right]. \quad (2.108)$$

Assim, obtém-se:

$$E[\ln L(b)] \propto \sum_{k=1}^K \sum_{i=1}^q \sum_{t=1}^n [\bar{r}_{kit} \ln P_{kit} + (f_{kit} - \bar{r}_{kit}) \ln Q_{kit}]. \quad (2.109)$$

Maximizar a equação (2.109) é equivalente a resolução das equações de Máxima Verossimilhança Marginal para os parâmetros dos itens apresentadas em (2.94) a (2.96). Contudo, deseja-se estimar a moda da distribuição *a posteriori* marginalizada. Logo, basta acrescentar as quantidades referentes às distribuições *priori* dos parâmetros dos itens que estão apresentadas nas equações (2.100) a (2.102).

O primeiro passo do algoritmo EM consiste no cálculo dos dados “artificiais” utilizando os valores *a priori* dos parâmetros dos itens. Esse passo é conhecido como Esperança (E). A próxima etapa consiste no cálculo dos parâmetros dos itens (passo de Maximização, M). Esses parâmetros são calculados um item por vez tendo por base os dados artificiais do passo E. Como as equações (2.100), (2.101) e (2.102) são não-lineares nos parâmetros, algum método iterativo como *Newton-Raphson* ou *Scoring de Fisher* deve ser usado no passo M. Para maiores detalhes sobre os métodos iterativos recomenda-se [Andrade et al. \(2000\)](#); [Azevedo \(2003\)](#).

Pode-se resumir os procedimentos da estimação Bayesiana pela *Moda a posteriori* para a adaptação do algoritmo EM de [Bock e Aitkin \(1981\)](#) da seguinte maneira ([Baker e Kim, 2004](#)):

### 1. Passo da Esperança (E):

- (a) Utilize a forma da quadratura  $\bar{\theta}_i$  e estimativas iniciais dos parâmetros dos itens para calcular a verossimilhança do vetor de respostas dos itens de cada examinando do grupo  $k$  para os  $q$  pontos de quadratura.
- (b) Utilize as equações (2.107) e (2.108) para gerar  $r_{kit}$  e  $f_{kit}$ , o número esperado de examinandos do grupo  $k$  submetidos ao item  $t$  e o número de respostas corretas (desse mesmo grupo) para o  $t$ -ésimo item a cada ponto  $q$  de quadratura.

### 2. Passo da Maximização (M):

- (a) Resolva as equações (2.100) a (2.102) utilizando as quantidades  $r_{kit}$  e  $f_{kit}$ , que são os dados artificiais do passo E. Como esses valores dependem dos valores desconhecidos dos parâmetros dos itens, essas equações devem ser resolvidas iterativamente (por exemplo, métodos de Newton-Raphson/Fisher).

Os passos que compõem cada iteração do algoritmo EM são repetidos até que algum critério de parada seja alcançado.

#### 2.3.4.2 Estimação Máxima a Posteriori (MAP)

A estimativa Bayesiana Modal ou Máxima a Posteriori (MAP) das habilidades de um examinando é baseada na seguinte forma do teorema de Bayes:

$$g(\theta_j | \mathbf{U}_j, \boldsymbol{\zeta}) \propto L(\mathbf{U}_j | \boldsymbol{\theta}_j, \boldsymbol{\zeta}) g(\theta_j). \quad (2.110)$$

Sendo assumido que a distribuição à priori  $g(\theta)$  é normal com hiperparâmetros  $\mu_\theta$  e  $\sigma_\theta^2$  que são especificados com antecedência. Tal como no caso da estimação dos parâmetros dos itens, o logaritmo é tomado, resultando em

$$\ln g(\theta_j | \mathbf{U}_j, \boldsymbol{\zeta}) \propto \ln L(\mathbf{U}_j | \boldsymbol{\theta}_j, \boldsymbol{\zeta}) + \ln g(\theta_j). \quad (2.111)$$

O componente da verossimilhança é dado por

$$L(\mathbf{U}_j | \boldsymbol{\theta}_j, \boldsymbol{\zeta}) = \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}. \quad (2.112)$$



Para obter um estimador MAP dos examinandos, o procedimento de escoring de Fisher é utilizado. Assim, a primeira e a segunda derivada da equação (2.111) com respeito a  $\theta_j$  são necessárias. Desde que os parâmetros dos itens sejam conhecidos, não há necessidade para transformação do parâmetro de discriminação. Portanto, a parametrização  $\alpha_t$ ,  $\beta_t$  e  $c_t$  pode ser empregada. Desta forma, a primeira derivada da equação (2.111) é dada por:

$$L_\theta = \ln L(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta}) = \sum_{t=1}^n \alpha_t \left[ \frac{P_t(\theta_j) - c_t}{P_t(\theta_j) - (1 - c_t)} \right] [u_{jt} - P_t(\theta_j)] - \left( \frac{\theta_{jt} - \mu_{\theta_j}}{\sigma_{\theta_j}^2} \right). \quad (2.113)$$

A segunda derivada é dada por:

$$L_{\theta\theta} = \ln g(\theta_j) = \sum_{t=1}^n \alpha_t^2 \frac{(P_t(\theta_j) - c_t) Q_t(\theta_j)}{(1 - c_t)^2 P_t(\theta_j)} \left[ \frac{u_{jt}c_t - P_t(\theta_j)^2}{P_t(\theta_j)} \right] - \frac{1}{\sigma_{\theta_j}^2}. \quad (2.114)$$

Para utilizar o método de escoring de Fisher, basta trocar  $u_{jt}$  por sua esperança, resultando em

$$\Lambda_{\theta\theta} = - \sum_{t=1}^n \alpha_t^2 \left[ \frac{P_t(\theta_j) - c_t}{1 - c_t} \right]^2 \frac{Q_t(\theta_j)}{P_t(\theta_j)} - \frac{1}{\sigma_{\theta_j}^2}. \quad (2.115)$$

A estimativa de  $\theta_j$  é obtida ao resolver a seguinte equação iterativamente, começando com um valor inicial de  $\theta_j = \theta_j^{(0)}$ :

$$[\hat{\theta}_j]_{(t+1)} = [\hat{\theta}_j]_{(t)} - [\Lambda_{\theta\theta}]_{(t)}^{-1} \times [L_\theta]_{(t)}, \quad (2.116)$$

onde  $t$  indexa a iteração. Dado que os examinandos são independentes, haverá uma única equação para cada examinando. O procedimento de estimação Modal de Bayes sempre convergirá para todos os possíveis padrões de resposta de itens. Assim, estimativas das habilidades podem ser obtidas mesmo quando um examinando responde todos os itens corretamente ou incorretamente (Mislevy e Bock, 1990).

### 2.3.4.3 Estimação da Esperança a Posteriori (EAP)

O procedimento de estimação da esperança a posteriori de Bayes (EAP) é baseado na seguinte forma do teorema de Bayes:

$$g(\theta_j|\mathbf{U}_j, \boldsymbol{\zeta}) = \frac{P(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta})g(\theta_j)}{P(U_j)}. \quad (2.117)$$

Sob a suposição de independência local, a probabilidade de ocorrer o vetor de respostas  $U_j = [u_{1j}, u_{2j}, \dots, u_{nj}]'$  do  $j$ -ésimo examinando é

$$P(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta}) = \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}},$$

que é a verossimilhança condicional sob o valor de  $\theta_j$ . Para um indivíduo aleatoriamente escolhido de uma população com distribuição de habilidade  $g(\theta_j)$ , o denominador da equação (2.117) é dado por

$$P(U_j) = \int_{\mathfrak{R}} P(U_j|\boldsymbol{\theta}_j, \boldsymbol{\zeta})g(\theta_j)d\theta_j,$$



que é a probabilidade incondicional. Da equação (2.117), a esperança não condicional de  $\theta_j$ , dado  $U_j$  é

$$E(\theta_j | \mathbf{U}_j, \boldsymbol{\zeta}) = \frac{\int_{\mathfrak{R}} \theta_j g(\theta_j) \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}} d\theta_j}{\int_{\mathfrak{R}} g(\theta_j) \prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}} d\theta_j}. \quad (2.118)$$

Uma vez que esta equação envolve integrais, a aproximação da quadratura de *Gauss-Hermite* é utilizada para aproximar a distribuição normal  $g(\theta)$ . Segundo [Mislevy e Bock \(1982\)](#), a forma da quadratura dada é

$$E(\theta_j | \mathbf{U}_j, \boldsymbol{\zeta}) = \bar{\theta}_j = \frac{\sum_{k=1}^q X_k(X_k) A(X_k)}{\sum_{k=1}^q (X_k) A(X_k)}. \quad (2.119)$$

A equação (2.119) possui algumas particularidades, a saber:

- Não é iterativa, resultando na estimativa EAP da habilidade diretamente.
- A distribuição a priori da habilidade empregada em seu cálculo é a distribuição final da quadratura ajustada.
- Os valores de  $L(X_k)$  estão também disponíveis no primeiro estágio.

Em função da facilidade de obtenção de suas estimativas, [Mislevy e Stocking \(1989\)](#) recomenda seu uso como escolha para estimação de habilidades.



## Capítulo 3

# Teste Adaptativo Computadorizado (CAT)

De maneira geral, algoritmos em testes adaptativos contemplam a seguinte estrutura cíclica:

- Escolha de um valor de  $\theta$  inicial.
- Aplicação de métodos de seleção de itens.
- Teste da regra de parada.

Com base no resultado do ciclo, o algoritmo adaptativo estima a habilidade do examinando. Após o cálculo dessa estimativa, um novo item é selecionado. A incorporação da Teoria de Resposta ao Item ao teste adaptativo tornou possível estimar as habilidades dos examinandos com base na resposta a um único item ou a uma série de itens. Na prática, a estimação das proficiências dos examinandos em um teste convencional, estilo papel-e-caneta é feita por meio de algum software comercial, à exemplo do BILOG, MULTILOG, PARSCALE, etc. Embora tenham surgido algumas aplicações comerciais ou não (Magis e Barrada, 2017; Magis *et al.*, 2012), os programas no CAT são desenvolvidos especificamente para um teste particular, utilizando ferramentas próprias, que na maioria das vezes não são aplicáveis a outros testes.

Em relação à aplicação de métodos de seleção de itens, destaca-se que TRI é a técnica que fornece grandes contribuições nessa área. Uma vez que diferentes examinandos podem receber diferentes conjuntos de itens calibrados em uma mesma escala, por meio da TRI obtém-se estimativas que são comparáveis entre si. Birnbaum (1968) foi o primeiro a mostrar que estimativas de Máxima Verossimilhança para as proficiências dos examinandos podem ser obtidas de um conjunto arbitrário de itens na TRI. Ele também introduziu os conceitos de FII e da FIT comumente utilizados na seleção de itens em CAT. Bock e Aitkin (1981), por sua vez, apresentaram fórmulas baseadas em pontos de quadratura que tornam mais eficientes o cálculo numérico da média e da variância da distribuição da posteriori. Como na estimação dos parâmetros dos itens, eles propuseram o uso prático dessa fórmulas para a estimação das proficiências, introduzindo os termos MAP e EAP, apresentados anteriormente.

Ponto importante no desenvolvimento de testes adaptativos são os critérios de seleção dos itens, onde alguns dos mais utilizados na prática são apresentados na seção (3.2). Geralmente, esses critérios baseiam-se em funções que dependem tanto dos parâmetros dos itens como das estimativas iterativas das proficiências. Por meio de alguma característica dessa função, faz-se possível a seleção iterativa dos itens no CAT.

### 3.1 Componentes do CAT

Um valor  $\theta$  inicial é necessário para um examinando no início do teste, quando nenhuma informação do examinando é conhecida. Uma opção simples é o uso antecipado da média da distribuição de habilidades como valor inicial para todos os examinandos. Alguma aleatoriedade também podem ser incluída para evitar a atribuição de itens iniciais similares aos examinandos.

Os métodos de seleção de itens, são sem dúvida, o componente mais importante no sistema CAT. Eles precisam não apenas servir aos propósitos da otimização da eficiência estatística ao estimar os parâmetros de habilidade do examinando, como também satisfazer múltiplas restrições não estatísticas (por exemplo, balanceamento de conteúdo, contagem de palavras, balanceamento de chave de resposta) e controle a exposição de cada item (Zheng, 2014).

Prática comum no CAT é assumir que os valores dos parâmetros dos itens estimados na construção do banco de itens já foram calibrados com grande precisão a tal ponto que se pode tratar as estimativas como valores verdadeiros para esses parâmetros (van der Linden e Glas, 2000). Esses autores ainda destacam que ignorar erros nos valores das estimativas dos parâmetros dos itens é uma estratégia sem sérias conseqüências quando o tamanho da amostra na fase de calibração do banco é grande.

Muitos programas operacionais de testes adaptativos encontram necessariamente uma base para seleção de itens não somente nos procedimentos estatísticos mas também impondo restrições no procedimento de seleção de itens para controlar certos atributos como balanceamento do conteúdo ou frequência de exposição do item, podendo assim influenciar a estimativa da habilidade ( $\theta$ ). Para se obter uma medida para tal habilidade, existem na TRI diversos estimadores, sendo os de Máxima Verossimilhança e os Bayesianos, os mais difundidos.

De modo que, os principais componentes do CAT são tratados com as técnicas já discutidas na seção (2.3) do capítulo anterior. Assim, antes de apresentar os métodos de seleção de itens, retoma-se alguns conceitos de estimação aplicados ao CAT.

Considere um banco de itens com  $n$  itens, onde o  $i$ -ésimo item do banco será administrado no CAT. Por conveniência, um teste adaptativo de tamanho fixo  $n$  será assumido, embora os resultados apresentados também possam ser estendidos a um teste com critério de parada definido por um nível fixo de precisão na estimação da habilidade. Suponha que  $p - 1$  itens foram selecionados.

Assim, a expressão (2.35) da função de Verossimilhança que relaciona a habilidade e as respostas do  $j$ -ésimo examinando aos primeiros  $p - 1$  itens, para o modelo de 3 parâmetros, é dada por:

$$L(\theta_j | u_1, \dots, u_{p-1}) = \prod_{t=1}^{p-1} P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}. \quad (3.1)$$

Vale ressaltar que a suposição de que as habilidades dos examinandos são estocasticamente independentes entre si é fundamental neste processo de estimação. Dessa maneira, como examinandos diferentes não possuem informação de outros examinandos, pode-se estimar cada habilidade separadamente.

No CAT, o procedimento para se obter estimativas por esse método consiste na maximização da função de verossimilhança em (3.1) em relação a todos os possíveis valores de  $\theta_j$ ,

de modo que:

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EMV} \equiv \arg \max_{\theta_j} \{L(\theta_j | u_1, \dots, u_{p-1}) : \theta_j \in (-\infty, \infty)\}. \quad (3.2)$$

Entretanto, os procedimentos de maximização da verossimilhança apresentam problemas tais como: problemas de estimação dos parâmetros de itens respondidos corretamente (ou incorretamente) por todos os examinandos; problemas de estimação das habilidades de examinandos que responderam corretamente (ou incorretamente) a todos os itens e estimativas fora do esperado ( $\alpha_t < 0$  ou  $c_t \notin [0, 1]$ ). Dessa maneira, os métodos Bayesianos apresentam-se como uma alternativa plausível.

Segundo van der Linden e Pashley (2000), na abordagem Bayesiana uma distribuição *a priori* para os valores desconhecidos da habilidade,  $g(\theta_j)$ , é assumida. Combinando a função de verossimilhança e a distribuição *a priori*, tem-se a distribuição *a posteriori* para  $\theta_j$ :

$$g(\theta_j | u_1, \dots, u_{p-1}) = \frac{L(\theta_j; u_1, \dots, u_{p-1})g(\theta_j)}{\int_{\mathfrak{R}} L(\theta_j; u_1, \dots, u_{p-1})g(\theta_j)d\theta_j}. \quad (3.3)$$

Os autores ainda destacam que no CAT a densidade  $g(\theta_j)$  é, tipicamente, assumida ser Uniforme em um intervalo fechado ou, se os examinandos forem permutáveis, uma estimativa empírica da distribuição das habilidades na população dos examinandos pode ser considerada. Geralmente, modela-se a distribuição da população dos examinandos pela Normal. No entanto, para o modelo de 3 parâmetros, uma *a priori* Normal não fornecerá uma distribuição *a posteriori* Normal quando o teste for pequeno.

De acordo com Lord (1980), um examinando é avaliado mais eficientemente quando os itens dos testes não são muito difíceis nem muito fáceis para este candidato. Contudo, os métodos de seleção adaptativa não só avaliam o nível de dificuldade dos itens, mas procuram encontrar uma medida de informação (que é uma combinação dos parâmetros dos itens e da estimativa da habilidade) em busca de uma melhor escolha dos itens para a estimação das habilidades.

Neste sentido, os estimadores pontuais para  $\theta_j$  pelo método Bayesiano baseiam-se na distribuição  $g(\theta_j | u_1, \dots, u_{p-1})$ , apresentada em (3.3). Frequentemente, adota-se a moda (MAP) ou a média (EAP) como estimadores das habilidades. O primeiro, também conhecido como estimador de Bayes Modal (BME) é definido como maximizador da posteriori de  $\theta_j$ :

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{MAP} \equiv \arg \max_{\theta_j} \{g(\theta_j; u_1, \dots, u_{p-1}) : \theta_j \in (-\infty, \infty)\}. \quad (3.4)$$

Diferentemente do estimador EMV, o procedimento de estimação MAP sempre converge independentemente do padrão de resposta dos indivíduos. Uma alternativa está na estimação pela *Média a Posteriori* (EAP). Esse estimador pode ser representado por:

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{EAP} \equiv \int_{\mathfrak{R}} \theta_j g(\theta_j | u_1, \dots, u_{p-1}) d\theta_j. \quad (3.5)$$

O estimador MAP foi inicialmente proposto por Lord (1984) e Mislevy (1986). O uso do estimador EAP em testes adaptativos foi extensivamente testado em Bock e Mislevy (1982).

Segundo van der Linden e Pashley (2000), uma abordagem bayesiana mais focada em princípios é abster-se de todas as estimativas pontuais, e usar a posteriori completa de  $\theta_j$  como o estimador de habilidade para o examinando. Este estimador não só revela o valor mais plausível de  $\theta_j$ , mas também mostra a plausibilidade de qualquer outro valor. Os autores

ressaltam que é comum resumir esta incerteza sobre  $\theta_j$  na forma da variância da distribuição à posteriori de  $\theta_j$ , dada por:

$$Var(\theta_j|u_1, \dots, u_{p-1}) \equiv \int_{\mathfrak{R}} [\theta_j - E(\theta_j|u_1, \dots, u_{p-1})]^2 g(\theta_j|u_1, \dots, u_{p-1}) d\theta_j. \quad (3.6)$$

Sob a abordagem da estimação pelo método da quadratura aplicado aos estimadores bayesianos apresentada na seção (2.3.3.2), o estimador EAP pode ser reescrito como:

$$\hat{\theta}_{j|u_1, \dots, u_{p-1}}^{EAP} = \frac{\int_{\mathfrak{R}} \theta_j L(\theta_j; u_1, \dots, u_{p-1}) g(\theta_j) d\theta_j}{\int_{\mathfrak{R}} L(\theta_j; u_1, \dots, u_{p-1}) g(\theta_j) d\theta_j} \approx \frac{\sum_{i=1}^q X_i L(X_i; u_1, \dots, u_{p-1}) A(X_i) \Delta_i^{-1}}{\sum_{i=1}^q L(X_i; u_1, \dots, u_{p-1}) A(X_i) \Delta_i^{-1}}, \quad (3.7)$$

onde  $X_i$  representa os pontos de quadratura em que  $i = 1, \dots, q$ ;  $A(X_i)$  o peso associado ao ponto  $X_i$  e  $\Delta_i$  o comprimento do intervalo.

No contexto da estimação pelo método EAP, os pesos são as probabilidades correspondentes aos pontos de uma distribuição *a priori* discretizada. Em certos casos, por exemplo, quando uma distribuição *a priori* Normal é assumida, os pontos e os pesos escolhidos podem ser os que possuem melhor precisão para a aproximação numérica da integral.

O Método de *Gauss-Hermite* descrito na seção (2.3.2.5) fornece pontos de quadratura e pesos para uma aproximação exata da integral de qualquer função, ponderada pela função erro gaussiana, que pode ser expressa como uma função polinomial de  $q$  graus (Stroud e Secrest, 1966, pág. 60-66).

Segundo Bock e Mislevy (1982), o que geralmente se faz em CAT consiste na determinação de pontos igualmente espaçados dentro de um intervalo (-3 a +3 desvios-padrão, por exemplo) que serão os pontos de quadraturas e os pesos serão iguais à probabilidade da priori discretizada nestes pontos.

Dados os pontos de quadratura, os cálculos para se obter estimativas pelo EAP não exigem nenhum processo iterativo como no caso do MAP ou EMV, o que o torna bastante vantajoso do ponto de vista computacional. Utilizando o mesmo raciocínio empregado na expressão (3.7), pode-se obter a Variância a Posteriori (VAP) associada à estimativa EAP:

$$\begin{aligned} Var(\theta_j|u_1, \dots, u_{p-1}) &= \frac{\int_{\mathfrak{R}} [\theta_j - \hat{\theta}_{j|u_1, \dots, u_{p-1}}^{EAP}]^2 L(\theta_j|u_1, \dots, u_{p-1}) g(\theta_j) d\theta_j}{\int_{\mathfrak{R}} L(\theta_j|u_1, \dots, u_{p-1}) g(\theta_j) d\theta_j} \\ &\approx \frac{\sum_{i=1}^q [X_i - \hat{\theta}_{j|u_1, \dots, u_{p-1}}^{EAP}]^2 L(X_i; u_1, \dots, u_{p-1}) A(X_i) \Delta_i^{-1}}{\sum_{i=1}^q L(X_i; u_1, \dots, u_{p-1}) A(X_i) \Delta_i^{-1}}. \end{aligned} \quad (3.8)$$

Em aplicações a populações reais, cerca de 80 pontos de quadratura entre 4 desvios-padrão devem ser usados para assegurar um nível de precisão menor que 0,20 (Bock e Mislevy, 1982). O nível de precisão em EAP é dado pela raiz da VAP, conhecido como desvio-padrão à posteriori (*Posteriori Standard Deviation*, PSD). Em qualquer quadratura dada, cerca de aproximadamente 10 pontos destes pontos são necessários ser utilizados.

Segundo Bock e Mislevy (1982), se um teste adaptativo apresentar itens aos examinandos até que um certo nível de precisão (PSD) seja obtido, a confiabilidade das estimativas das habilidades será a mesma para todos os examinandos. É assumido nestes cálculos que

as habilidades dos examinandos na população têm distribuição  $N(0, 1)$ , de forma que o coeficiente de confiabilidade para a estimativa EAP é dado pela correlação intraclasse expressa por:

$$\rho = 1 - [\text{PSD}(\theta_j)]^2. \quad (3.9)$$

Dessa maneira, se as estimativas das habilidades forem utilizadas para seleção e classificação dos examinandos em diversos níveis da escala de habilidade, uma medida de precisão constante para todos os indivíduos assegura que os erros cometidos na classificação serão uniformes para todas as decisões. Esse tipo de equidade nos procedimentos de seleção e classificação dos indivíduos não pode ser alcançado nos testes convencionais de tamanho fixo.

## 3.2 Métodos de Seleção de Itens

Segundo [van der Linden e Glas \(2010\)](#), responder à questão sobre qual dos métodos de estimação de habilidade, dentre os apresentados na seção anterior, é melhor, está intrinsicamente relacionado a outros aspectos do CAT. Em primeiro lugar, a escolha do critério de seleção do item. Aspectos conhecidos que têm impacto nas estimativas de habilidade são:

- a composição do conjunto de itens, independentemente de o procedimento de estimação usar ou não informações colaterais sobre os examinandos,
- a escolha do método para controlar a taxa de exposição dos itens,
- a presença de restrições de conteúdo na seleção de itens.

Alguns destes aspectos são discutidos nas seções deste capítulo. Maiores detalhes sobre os aspectos não tratados nesta dissertação, assim como detalhes sobre outros critérios de seleção de itens tais como *Critério Completamente Bayesiano*, *Critério Bayesiano com Informação Colateral* e *Critério Bayesiano com Parâmetros de Itens Aleatórios*, além de uma criteriosa avaliação dos critérios de seleção itens e dos estimadores de habilidade comumente utilizados no CAT, podem ser encontrados em ([van der Linden e Glas, 2000, 2010](#)).

### 3.2.1 Critério da Máxima Informação (CMI)

[Birnbaum \(1968\)](#) introduziu a FIT como o principal critério para a composição do teste linear. [Lord \(1977\)](#) adaptou as idéias de Birnbaum e utilizou o estimador de máxima verossimilhança como critério para estimar a habilidade dos examinandos em um teste adaptativo de habilidade verbal, tornando-se um dos procedimentos mais utilizados na seleção de itens no CAT.

A FIT é a medida de informação esperada, descrita na seção (2.3.1), também conhecida por *Informação de Fisher*, que quando esperada de  $\theta$  através de  $U$  é dada por:

$$I_U(\theta) = E_{U|\theta} \left[ -\frac{\partial^2 \ln P(u|\theta)}{\partial \theta^2} \right], \quad (3.10)$$

considerada como uma função da habilidade do parâmetro. A motivação de Birnbaum para esta função era o fato de que, aumentando o tamanho o teste, a variância do estimador de máxima verossimilhança era conhecida convergir para a recíproca da expressão (2.24).

A Informação de Fisher é o valor médio da curvatura da verossimilhança. Quanto maior esta curvatura, maior será a informação sumarizada na função de verossimilhança e, conseqüentemente, maior o valor de  $I_U(\theta)$ .

Em testes adaptativos, o critério de informação máxima foi rapidamente adotado como escolha popular. O critério seleciona o  $p$ -ésimo item para maximizar (3.10) no ponto em que  $\theta = \hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}$ . Segundo van der Linden e Glas (2010), isto pode ser representado por:

$$t_p \equiv \arg \max_j \{I_{U_1, \dots, U_{p-1}, U_j}(\hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}) : j \in R_p\}, \quad (3.11)$$

onde  $t$  é o  $t$ -ésimo item do banco,  $t = 1, \dots, n$ , e  $p$ , a posição do  $t$ -ésimo item no teste adaptativo. Suponha que  $p - 1$  itens foram administrados no CAT. Os índices dos itens administrados formam o conjunto  $S_{p-1} = \{t_1, \dots, t_{p-1}\}$ , enquanto os itens restantes formam o conjunto  $R_p = \{1, \dots, n\} \setminus S_{p-1}$ , onde a função de verossimilhança para o conjunto de respostas  $U_{t_1} = u_{t_1}, \dots, U_{t_{p-1}} = u_{t_{p-1}}$  obtido sobre os primeiros  $p - 1$  itens é expressa por:

$$L(\theta|U_{t_1}, \dots, U_{t_{p-1}}) \equiv \prod_{k=1}^{p-1} \frac{\{exp[\alpha_{t_k}(\theta - \beta_{t_k})]\}^{u_{t_k}}}{1 + exp[\alpha_{t_k}(\theta - \beta_{t_k})]}. \quad (3.12)$$

Em função da aditividade da função de informação, o critério resume-se a:

$$t_p \equiv \arg \max_j \{I_{U_j}(\hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}) : j \in R_p\}. \quad (3.13)$$

Segundo Costa (2009), maximizar a Informação de Fisher significa intuitivamente selecionar um item de dificuldade que corresponda exatamente ao nível de habilidade do examinando. Além do mais, itens com maior discriminação, ou equivalentemente, alto valor do parâmetro  $a$ , serão preferencialmente selecionados pelo algoritmo. Para melhor avaliar o impacto dos parâmetros dos itens no cálculo da informação, Costa (2009) realizou uma análise de sensibilidade.

Cabe observar que embora o estimador de máxima log-verossimilhança seja frequentemente defendido como escolha natural, a escolha do estimador de  $\theta$  em (3.13) é livre. Chang e Ying (2009) mostraram que para este critério, o estimador de máxima log-verossimilhança converge para o valor verdadeiro com uma variância amostral que se aproxima do recíproco de (2.24). O resultado é válido apenas para um grupo de itens (infinito) com todos os valores possíveis para o parâmetro de discriminação no item agrupado, distantes de 0 e 1, e valores para o parâmetro de incerteza distantes de 1. Além disso, para o modelo de 3 parâmetros, uma pequena modificação da equação de verossimilhança é necessária para evitar múltiplas raízes. Como essas condições são suaves, acredita-se que os resultados proporcionem uma aproximação útil aos testes adaptativos de um agrupamento de itens bem planejados.

### 3.2.2 Aproximação de Owen

Como uma alternativa ao Critério da Máxima Informação, Owen (1975) propôs uma aproximação empírica do procedimento de Bayes para testes adaptativos, baseado no modelo ogiva normal, expresso por:

$$P_t(\theta_j) = c_t + (1 - c_t)\Phi[\alpha_t(\theta_{jt} - \beta_t)], \quad (3.14)$$



onde  $\Phi(\cdot)$  é a função acumulada da distribuição normal e  $c_t$  uma assíntota inferior para modelar a probabilidade da incerteza quanto à resposta correta ao item  $t$ .

Para um vetor de respostas aos primeiros  $p - 1$  itens, a função de verossimilhança é dada por (3.12) com o fator logístico trocado pela ogiva normal. Assim, considerando uma priori  $g(\theta)$ , a seguinte expressão para a distribuição *a priori* de  $\theta$  após  $p - 1$  itens, é dada por:

$$g(\theta|U_{t_1}, \dots, U_{t_{p-1}}) = \frac{L(\theta|U_{t_1}, \dots, U_{t_{p-1}})g(\theta)}{\int_{\mathcal{R}} L(\theta|U_{t_1}, \dots, U_{t_{p-1}})g(\theta)d\theta}. \quad (3.15)$$

O Procedimento de Owen's é baseado em (3.15) como um processo de atualização para a posteriori com a escolha de uma densidade normal para a priori  $g(\theta)$ , onde o item  $p$  é escolhido para satisfazer

$$|\beta_{t_p} - E(\theta|U_{t_1}, \dots, U_{t_{p-1}})| < \delta, \quad (3.16)$$

para um pequeno valor de  $\delta$ , onde  $E(\theta|U_{t_1}, \dots, U_{t_{p-1}})$  é a esperança de  $\theta$  sobre (3.15), agora conhecido como o estimador *Esperado a Priori* (EAP) de  $\theta$ .

Por se tratar de um procedimento recursivo, o mesmo é encerrado tão logo a variância de (3.15) seja menor que um valor limite pré-estabelecido.

Segundo van der Linden (1996), deve ser observado que a verossimilhança (3.15) não tem a família normal como classe conjugada de distribuições. Portanto, se uma normal a priori é escolhida, a posteriori não é normal, e repetida a atualização da posteriori usando (3.15) logo convergiria para uma posteriori que não poderia ser calculada em aplicações de testes adaptativos em tempo real, por computadores dos anos 70. Assim, Owen (1975) propôs trocar a posteriori verdadeira por uma aproximação normal com mesma média e variância, e apresentou uma fórmula fechada para a atualização da posteriori normalizada. A aproximação para a média foi motivada pela apresentação de limites leves sob  $\delta$ , uma vez que o estimador  $\hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}} \equiv E(\theta|U_{t_1}, \dots, U_{t_{p-1}})$  converge para o valor verdadeiro de  $\theta$  em erro quadrático para  $p \rightarrow \infty$  (Owen, 1975, Teorema 2).

É possível notar que  $\beta_{t_p}$  em (3.16) é o único parâmetro que determina a seleção do  $p$ -ésimo item. Nenhuma tentativa adicional é feita para otimizar a seleção do item. Entretanto, Owen fez uma referência ao critério de minimização do risco pré-posteriori sob a função de perda quadrática como uma alternativa para (3.16), mas absteve-se de prosseguir nesta opção por causa de sua complexidade computacional (van der Linden e Glas, 2010).

### 3.2.3 Critério da Máxima Informação Global (CMIG)

Para lidar com a estimação do erro no início do teste, Chang e Ying (1996) sugeriram trocar a Informação de Fisher expressa em (3.11) por uma medida baseada na Informação de Kullback-Leibler (IKL), sendo esta uma medida geral para a “distância” entre duas distribuições. Quanto maior é a IKL, mais fácil é a discriminação entre duas distribuições, ou equivalentemente, entre os valores dos parâmetros que as indexam (Lehmann e Casella, 1998, seção 1.7).

Para o modelo de resposta definido em (2.30), a medida IKL para a distribuição das respostas sobre o  $p$ -ésimo item no teste associado com o valor da habilidade verdadeira ( $\theta_0$ ) do examinando e a corrente estimativa da habilidade ( $\hat{\theta}_{p-1}$ ) é dada por:

$$K_{t_p}(\hat{\theta}_{p-1}, \theta_0) \equiv E_{U_{t_p}} \left[ \ln \frac{L(\theta_0|U_{t_p})}{L(\hat{\theta}_{p-1}|U_{t_p})} \right]. \quad (3.17)$$

Portanto, a medida pode ser calculada como

$$K_{t_p}(\hat{\theta}_{p-1}, \theta_0) = p_{t_p}(\theta_0) \ln \frac{p_{t_p}(\theta_0)}{p_{t_p}(\hat{\theta}_{p-1})} + [1 - p_{t_p}(\theta_0)] \ln \frac{1 - p_{t_p}(\theta_0)}{1 - p_{t_p}(\hat{\theta}_{p-1})}. \quad (3.18)$$

Em função da esperança condicional entre as respostas, a informação nas respostas para os primeiros  $p$  itens do teste podem ser escritas como

$$K_p(\hat{\theta}_{p-1}, \theta_0) \equiv E_{U_{i_p}} \left[ \ln \frac{L(\theta_0|U_{t_1}, \dots, U_{t_p})}{L(\hat{\theta}_{p-1}|U_{t_1}, \dots, U_{t_p})} \right] = \sum_{h=1}^p K_{t_h}(\hat{\theta}_{p-1}, \theta_0). \quad (3.19)$$

A IKL informa sobre quão bem a variável resposta é capaz de discriminar entre a corrente habilidade estimada,  $\hat{\theta}_{p-1}$ , e o valor da habilidade verdadeira  $\theta_0$ . Em razão do valor verdadeiro da habilidade ser desconhecido, [Chang e Ying \(1996\)](#) propuseram trocar (3.17) por sua integral sobre um intervalo para as estimativas de habilidade corrente,  $[\hat{\theta}_{p-1} - \delta_p, \hat{\theta}_{p-1} + \delta_p]$ , onde  $\delta_p$  é uma função decrescente do número de postos dos itens no teste. Assim, o  $p$ -ésimo item é então selecionado de acordo com

$$t_p \equiv \arg \max_j \left\{ \int_{\hat{\theta}_{p-1} - \delta_p}^{\hat{\theta}_{p-1} + \delta_p} K_j(\hat{\theta}_{p-1}, \theta) d\theta : j \in R_p \right\}. \quad (3.20)$$

[Chang e Ying \(1996\)](#) destacam que a implementação do procedimento de seleção apresentado em (3.20) é mais eficiente quando  $\delta_p$  decresce para 0, porém, nestes casos o tamanho do teste se aproxima de  $\infty$ . Quanto à especificação de  $\delta$ , espera-se que o resultado do intervalo  $(\hat{\theta}_{p-1} - \delta_p, \hat{\theta}_{p-1} + \delta_p)$  contenha  $\theta_0$ . Pela teoria assintótica, o EMV de  $\hat{\theta}_{p-1}$  é assintoticamente Normal com média  $\theta_0$  e variância  $1/I_U(\theta_0)$ . Dessa maneira, um intervalo de confiança para  $\theta_0$  deve ser do tipo

$$\left[ \hat{\theta}_{p-1} - \frac{p}{\sqrt{I_U(\hat{\theta}_{p-1})}}, \hat{\theta}_{p-1} + \frac{p}{\sqrt{I_U(\hat{\theta}_{p-1})}} \right],$$

onde a constante  $p$  é selecionada de acordo com uma específica probabilidade de cobertura. Para um teste de tamanho  $n$ , a medida de Informação de Fisher será de ordem  $n$ , de forma que [Chang e Ying \(1996\)](#) concluíram que um valor razoável para  $\delta_p$  seria igual a  $\frac{p}{\sqrt{n}}$ .

### 3.2.4 Critério da Máxima Verossimilhança Ponderada (CMVP)

[Veerkamp e Berger \(1997\)](#) observaram que embora na maioria dos procedimentos de teste adaptativo, o item com o valor mais alto da função de informação para a estimativa do parâmetro de habilidade fosse selecionado, a informação mais alta para a estimativa do parâmetro de habilidade, poderia não ser a mais alta para o valor do parâmetro verdadeiro. E assim, a diferença entre a estimativa da habilidade e o valor verdadeiro pudesse prejudicar os resultados.

Em vez de usar apenas um único valor da função de informação, os valores da função de informação para toda a gama de valores na escala de habilidade podem ser usados. Esse intervalo pode ser transformado em um único valor por uma média ponderada. Segundo [Veerkamp e Berger \(1997\)](#), esta idéia geral de selecionar itens por meio da maior informação

média ponderada em toda a gama de habilidades pode ser formulada como:

$$t_p \equiv \arg \max_j \left\{ \int_{\mathfrak{R}} W_{p-1}(\theta | u_{t_1}, \dots, u_{t_{p-1}}) I_{U_j}(\hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}) d\theta : j \in R_p \right\}. \quad (3.21)$$

Vale ressaltar que tal como em (3.11), o critério seleciona o  $p$ -ésimo item para maximizar (3.10) no ponto em que  $\theta = \hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}$ , porém atribuindo peso em toda a escala de habilidade. De forma que a expressão em (3.21), seja considerada como um procedimento geral, de acordo com a escolha da função de ponderação  $W_{p-1}$  e com o estimador  $\hat{\theta}_{u_{t_1}, \dots, u_{t_{p-1}}}$  utilizado, sendo o caso mais comum o uso do estimador EMV, apresentado em (3.2).

Uma alternativa proposta por Warm (1989) para maximizar (3.1), denominada *estimador de verossimilhança ponderada* (WLE), dada por:

$$\hat{\theta}_{j_{u_1, \dots, u_{p-1}}}^{WLE} \equiv \arg \max_{\theta_j} \{W_{p-1}(\theta_j) L(\theta_j | u_1, \dots, u_{p-1}) : \theta_j \in (-\infty, \infty)\}, \quad (3.22)$$

foi testada por Veerkamp e Berger (1997) em conjunto com outras duas formas de definir a função de ponderação, para avaliar o critério de seleção.

A primeira delas, considera o caso em que um teste é projetado para selecionar pessoas para colocação ou admissão, onde muitas vezes são necessárias maiores informações para alguns valores pré-especificados na escala de habilidades. Uma função de ponderação com picos para estes valores pode ser apropriada nesta situação.

A segunda maneira de abordar a questão, trata do uso de informações auxiliares sobre os examinandos, representadas pela distribuição a priori das habilidades  $g(\theta)$  como uma função ponderadora, isto é,  $W_{p-1}(\theta | u_{t_1}, \dots, u_{t_{p-1}}) = g(\theta)$ .

O foco do trabalho de Veerkamp e Berger (1997) considerou a incerteza da estimativa da habilidade sendo tratada por funções de ponderação especiais, em particular o caso em que  $W_{p-1}(\theta | u_{t_1}, \dots, u_{t_{p-1}}) = L(\theta | u_{t_1}, \dots, u_{t_{p-1}})$ . Esta escolha equivale a dar mais peso ao valor da função de informação  $I_{U_j}(\theta)$  quando é mais provável que o valor de  $\theta$  correspondente na escala de habilidades, seja o verdadeiro valor da habilidade do examinado.

Portanto, o CMVP é definido como a área sobre uma função que é o produto entre a função de verossimilhança e a função de Informação de Fisher. Vale ressaltar que embora a medida de Informação de Fisher tenha sido utilizada, o procedimento pode ser estendido para a IKL (van der Linden e Glas, 2010).

Se o estimador de máxima verossimilhança, apresentado em (3.2), é utilizado, o critério (3.21) no caso em que  $W_{p-1}(\theta | u_{t_1}, \dots, u_{t_{p-1}}) = L(\theta | u_{t_1}, \dots, u_{t_{p-1}})$  coloca mais peso sobre os valores de  $\theta$  próximos à corrente estimativa da habilidade. No início do teste, a função de verossimilhança é uniforme, e os valores distantes de  $\hat{\theta}_{p-1}$  recebem peso substancial. No final do teste, a função de verossimilhança tende a tornar-se pico, e quase todo o peso vai para valores próximos de  $\hat{\theta}_{p-1}$ .

Veerkamp e Berger (1997) também especificaram um *Critério de Informação Pontual* (CIP) e um *Critério de Informação Intervalar* (CII), sendo que este último, tal como (3.20), assume integração sobre um intervalo finito de valores de  $\theta$  sobre a corrente estimativa da habilidade. Entretanto, ao invés de definir um intervalo de tamanho  $\delta_p$ , eles sugeriram um intervalo de confiança para  $\theta$ . Segundo van der Linden e Glas (2010), a mesma sugestão poderia ser possível para o critério em (3.20).

Por fim, Veerkamp e Berger (1997) ressaltam que tanto o CMVP, quanto o CII apresentam resultados melhores que o CIP porque eles levam em consideração a incerteza da estimativa da habilidade. Entretanto, desde que a função de verossimilhança revela a quan-

tidade de peso a ser atribuída ao valor da função de informação em todos os valores possíveis na escala de habilidade, pode inferir-se que o CMVP será melhor do que o CII, em razão da função de ponderação associada ao CII ser constante para todo o intervalo.

# Capítulo 4

## CAT via Filtro de Kalman Não Linear

Até o presente momento, foram apresentados os principais aspectos para a discussão do objetivo principal desta dissertação. Os conceitos que fundamentam a compreensão dos estimadores utilizados no CAT foram apresentados no capítulo 2. Uma breve discussão sobre alguns dos estimadores utilizados no CAT e seus principais aspectos foi tratado no capítulo 3, onde é possível observar a relevância do framework Bayesiano no contexto moderno dos testes adaptativos.

Embora o trabalho pioneiro de Birnbaum (1968), juntamente com as inovações propostas por Bock e Aitkin (1981); Bock e Lieberman (1970); Bock e Mislevy (1982); Lord (1980); Owen (1975) tenham sido de enorme importância no desenvolvimento e aprimoramento das técnicas desenvolvidas ao longo do tempo, merecem destaque, os trabalhos de Lord (1980); Owen (1975), pelo caráter visionário, em particular de Owen, no aprimoramento de métodos Bayesianos para estimação de proficiências em testes adaptativos.

Neste capítulo, são apresentadas duas formulações do *Filtro de Kalman Não Gaussiano* (FKNG) para estimação de proficiências em testes adaptativos computadorizados (CAT). As premissas utilizadas e as formulações estatísticas para o problema são apresentadas, de tal forma, que estabelecem uma relação recursiva entre os parâmetros envolvidos. A estimação destes parâmetros trata, dentro do contexto Bayesiano, do procedimento de atualização proposto pelo Filtro de Kalman, condicionalmente a uma história passada e à observação presente, para fazer previsões de observações futuras. Este processo é desenvolvido, nesta dissertação, pela criação de um algoritmo CAT para a estimação da proficiência de um examinando considerando-se os itens apresentados ao mesmo, e a verificação da convergência do processo recursivo.

A seguir, faz-se uma breve descrição de um Modelo Dinâmico Gaussiano.

### 4.1 O Modelo Dinâmico Gaussiano

Considere que  $Y_t, Y_{t-1}, \dots, Y_1$  são escalares, ou vetores que denotam valores observados de uma variável de interesse nos tempos  $t, t-1, \dots, 1$ . Assuma que  $Y_t$  depende de uma quantidade não observável  $\theta_t$ , conhecida como *Estado Natural*. O objetivo é fazer inferência sobre  $\theta_t$ , que pode ser um escalar ou um vetor, cuja dimensão é independente da dimensão de  $Y_t$ . A relação entre  $Y_t$  e  $\theta_t$  é linear e é especificada pela *equação das observações*, dada por

$$Y_t = F_t \theta_t + v_t, \quad (4.1)$$

onde  $F_t$  é uma quantidade conhecida. Aqui, o *erro de observação*, denotado por  $v_t$ , é assumido ser normalmente distribuído com média zero e variância conhecida  $V_t$ , isto é,  $v_t \sim N(0, V_t)$ .

A principal diferença entre o Filtro de Kalman e a representação do modelo linear convencional, é que no Filtro de Kalman, o *estado natural*, análogo aos coeficientes de regressão no modelo linear convencional, não são assumidos serem constantes, mas podem mudar com o tempo. Esta característica dinâmica é incorporada na formulação do Filtro de Kalman via a *equação de sistema*, em que

$$\theta_t = G_t \theta_{t-1} + w_t, \quad (4.2)$$

onde  $G_t$  é também uma quantidade conhecida, e o *erro da equação de sistema*  $w_t \sim N(0, W_t)$ , com  $W_t$  conhecido. Definidas as equações (4.1) e (4.2), a atratividade do Filtro de Kalman está relacionada à habilidade de inserir um conhecimento do comportamento do sistema em um modelo estatístico. É importante notar que a relação entre as equações (4.1) e (4.2), especificadas através de  $F_t$  e  $G_t$ , podem mudar ou não com o tempo, assim como as variâncias  $V_t$  e  $W_t$ .

Tal como as suposições usuais de um modelo linear acerca dos termos de erros, são também postulados que  $v_t$  e  $w_t$  são independentes, embora extensões para os casos de dependência sejam mais diretas, e não tratadas por [Meinhold e Singpurwalla \(1983\)](#).

Tal como definido por [Harvey \(1989\)](#), o Filtro de Kalman trata-se de um procedimento recursivo, que no exemplo tratado acima, tem como objetivo fazer inferência sobre o estado natural  $\theta_t$ . Isto pode ser conduzido utilizando-se diretamente o Teorema de Bayes, onde se postula

$$\text{Prob}(\text{Estado Natural} \mid \text{Dados}) \propto \frac{\text{Prob}(\text{Dados} \mid \text{Estado Natural}) \times \text{Prob}(\text{Estado Natural})}{\text{Prob}(\text{Estado Natural})}, \quad (4.3)$$

podendo ser reescrito para as equações (4.1) e (4.2), considerando o tempo  $t$ , como

$$\underbrace{P(\theta_t | Y_t)}_{\text{distribuição à posteriori para } \theta} \propto \underbrace{P(Y_t | \theta_t, Y_{t-1})}_{\text{verossimilhança para } \theta} \times \underbrace{P(\theta_t | Y_{t-1})}_{\text{distribuição à priori para } \theta}. \quad (4.4)$$

Assim, o processo recursivo pode ser melhor explicado se o foco da atenção for no ponto do tempo  $t - 1$ ,  $t = 1, 2, \dots$ , e nos dados observados até então,  $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots, Y_1)$ . Ao que se segue utiliza-se manipulações matriciais para permitir que  $Y$  e/ou  $\theta$  sejam vetores, sem observá-los explicitamente como tal. Assim, em  $t - 1$ , o estado de conhecimento de  $\theta_{t-1}$  é incorporado na seguinte declaração de probabilidade para  $\theta_{t-1}$

$$(\theta_{t-1} | Y_{t-1}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1}), \quad (4.5)$$

em que (4.5) representa a distribuição à posteriori de  $\theta_{t-1}$ . É importante notar que o processo recursivo é iniciado no tempo 0, escolhendo-se  $\hat{\theta}_0$  e  $\Sigma_0$  como sendo as melhores alternativas para a média e a variância de  $\theta_0$ , respectivamente.

Neste ponto, o foco se divide em dois estágios:

**Estágio 1:** *Antes de observar  $Y_t$* , a melhor escolha para  $\theta_t$  é governada pela equação de sistema (4.2). Desde que  $\theta_{t-1}$  é descrito por (4.5), o estado de conhecimento sobre  $\theta_t$  é incorporado na seguinte distribuição à priori

$$(\theta_t | Y_{t-1}) \sim N(G_t \hat{\theta}_{t-1}, R_t), \quad (4.6)$$

onde  $R_t = G_t \Sigma_{t-1} G_t' + W_t$ .

Na obtenção de (4.6), que representa a priori para  $\theta_t$ , no próximo ciclo de (4.4), é utilizado

um resultado bem conhecido sobre distribuições normais multivariadas, descritas por  $X \sim N(\mu, \Sigma)$ , que consideram para qualquer constante  $C$ , a seguinte relação

$$CX \sim N(C\mu, C\Sigma C'), \quad (4.7)$$

onde  $C'$  denota a transposta de  $C$ .

**Estágio 2:** Ao observar  $Y_t$ , o objetivo é calcular a posteriori de  $\theta_t$ , usando (4.4). Entretanto, para fazer isto é necessário conhecer a verossimilhança,  $P(Y_t|\theta_t, Y_{t-1})$ , cuja determinação é obtida pelos seguintes argumentos: considere  $e_t$  denotar o erro na predição de  $Y_t$  a partir do ponto  $t-1$ , de modo que

$$e_t = Y_t - \hat{Y}_t = Y_t - F_t G_t \hat{\theta}_{t-1}. \quad (4.8)$$

Desde que  $F_t$ ,  $G_t$  e  $\hat{\theta}_{t-1}$  são todos conhecidos, observar  $Y_t$  é equivalente à observar  $e_t$ . Assim, (4.4) pode ser reescrita como

$$P(\theta_t|Y_t, Y_{t-1}) = \underbrace{P(\theta_t|e_t, Y_{t-1})}_{\text{distribuição à posteriori para } \theta} \propto \underbrace{P(e_t|\theta_t, Y_{t-1})}_{\text{verossimilhança para } \theta} \times \underbrace{P(\theta_t|Y_{t-1})}_{\text{distribuição à priori para } \theta}. \quad (4.9)$$

Substituindo (4.1) em (4.8),  $e_t$  é reescrito como

$$e_t = F_t(\theta_t - G_t \hat{\theta}_{t-1}) + v_t, \quad (4.10)$$

de modo que  $E(e_t|\theta_t, Y_{t-1}) = F_t(\theta_t - G_t \hat{\theta}_{t-1})$ . Como  $v_t \sim N(0, V_t)$ , a verossimilhança é descrita por

$$(e_t|\theta_t, Y_{t-1}) \sim N(F_t(\theta_t - G_t \hat{\theta}_{t-1}), V_t). \quad (4.11)$$

Utilizando agora o Teorema de Bayes na expressão (4.9), obtem-se a atualização de  $\theta_t$  que descreve o conhecimento sobre o estado no tempo  $t$ , tal que

$$P(\theta_t|Y_t, Y_{t-1}) = \frac{P(e_t|\theta_t, Y_{t-1}) \times P(\theta_t|Y_{t-1})}{\int P(e_t, \theta_t|Y_{t-1}) d\theta_t}. \quad (4.12)$$

Uma vez calculada a posteriori atualizada em (4.12), é possível voltar à expressão (4.5) para o próximo ciclo do processo recursivo.

Como o exemplo apresentado supõe normalidade para a expressão (4.5), e o esforço gasto no cálculo da atualização de (4.12) é muito tedioso, Meinhold e Singpurwalla (1983) mostraram que tal distribuição à posteriori é da forma apresentada em (4.5) de acordo com resultados conhecidos em estatística multivariada, e em algumas propriedades da distribuição normal (Anderson, 1958, pág. 28-29). Dessa forma, concluem que a distribuição à posteriori expressa em (4.12) é dada por

$$(\theta_t|e_t, Y_{t-1}) \sim N[G_t \hat{\theta}_{t-1} + R_t F_t' (V_t + F_t R_t F_t')^{-1} e_t, R_t - R_t F_t' (V_t + F_t R_t F_t')^{-1} F_t R_t]. \quad (4.13)$$

## 4.2 O Filtro de Kalman Não Linear (FKNG)

Segundo Chen *et al.* (2003), a teoria de filtros estocásticos foi estabelecida no início da década de 1940, devido ao trabalho pioneiro de Kolmogorov (1941) e Wiener (1949), culminando na publicação do clássico Filtro de Kalman (Kalman, 1960).



A partir do estudo de Kalman (1960), vários outros trabalhos exploraram aplicações iterativas utilizando o framework bayesiano. Porém, tendo em mente que o Filtro de Kalman é limitado por seus pressupostos, inúmeros métodos de filtragem não-linear ao longo de sua linha foram propostos e desenvolvidos para superar sua limitação. Não surpreendentemente, a teoria bayesiana também foi estudada na literatura de filtragem, sendo uma das primeiras pesquisas de estimativa bayesiana iterativa encontrada no trabalho de Ho e Lee (1964), no qual eles especificaram o princípio e o procedimento da filtragem bayesiana, utilizando uma forma fechada do Filtro de Kalman-Wiener para estimação linear em ruído gaussiano. Spragins (1965) discutiu a aplicação iterativa de Bayes para a estimação de parâmetros sequenciais e chamou-a de *aprendizagem bayesiana*.

Harrison e Stevens (1976) utilizaram como modelo básico para previsões o Modelo Linear Dinâmico (DLM) definido juntamente com as relações de recorrência do Filtro Kalman gerando uma série de formulações do modelo. Meinhold e Singpurwalla (1983) mostraram que o filtro de Kalman é uma simples aplicação do Teorema da Bayes em uma sequência de espaço de probabilidades, popularizando assim os modelos dinâmicos Bayesianos na comunidade estatística.

A partir de então, o Filtro de Kalman passou a ser utilizado juntamente com os Modelos Lineares Generalizados Dinâmicos (DLMG), para efeitos de previsões, sendo aplicado por West *et al.* (1985) em uma classe de modelos de resposta univariada e por Fahrmeir (1992) para respostas multivariadas, baseado na idéia dos momentos do Filtro de Kalman extendido.

Harvey (1989) definiu o Filtro de Kalman como um procedimento recursivo para calcular o estimador ideal do vetor de estados no tempo  $t$ , com base nas informações disponíveis no tempo  $t$ . Segundo Meinhold e Singpurwalla (1983), o Filtro de Kalman é melhor compreendido se for lançado como um problema de inferência Bayesiana. Tal característica ficou evidenciada nos trabalhos de Harrison e Stevens (1971, 1976) que primeiramente se interessaram por predição Bayesiana. Entretanto, estes trabalhos tiveram foco na implementação do Filtro de Kalman, e não em sua formulação básica.

Nas próximas seções são apresentados dois modelos dinâmicos para a estimação de proficiências em plataformas CAT. O primeiro deles, possui uma apresentação mais voltada a formulações de Filtros de Kalman Não Gaussianos (Harvey, 1989), onde  $c_t$  é não nulo. Enquanto no segundo, um enfoque direcionado aos modelos dinâmicos de Smith (1979), considera  $c_t = 0$ .

### 4.3 Modelos Dinâmicos para o CAT

Em um processo CAT para a estimação da proficiência em dada área de um certo indivíduo  $j$ , considere que a sua proficiência real,  $\theta_j$ , possa ser modelada de acordo com um processo aleatório, que varia suavemente ao redor de  $\theta_j$ . Considerando um conjunto de  $n$  questões ou itens aplicados ao indivíduo  $j$ , pode-se considerar que  $\{\theta_{j1}, \theta_{j2}, \dots, \theta_{jn}\}$  represente uma série temporal (ou dados de painel) das flutuações em torno de  $\theta_j$  (ponto de convergência) ao longo da sessão CAT. Tais flutuações, assim como  $\theta_j$ , são variáveis latentes (não observadas) e serão modeladas de acordo com Filtros de Kalman Não Lineares.

O processo observável  $\{\mathbf{X}_{jt}\}$ ,  $t = 1, 2, \dots, n$  em estudo, trata-se do conjunto de respostas dadas às  $n$  questões, pelo indivíduo  $j$ . Neste caso, considerando um *processo de Bernoulli*, no tempo  $t$ , um acerto é tal que  $x_{jt} = 1$  e um erro é tal que  $x_{jt} = 0$ .

Portanto, uma dada trajetória de respostas emitidas pelo indivíduo  $j$ , é tratada em conjunto com as flutuações ao redor de  $\theta_j$ , como sendo um processo duplamente estocástico,  $\{\mathbf{X}_{jt}, \Theta_{jt}\}$ ,  $t = 1, 2, \dots, n$ , em  $\{\Theta_{jt}\}$ , ou alguma transformação desta série, será descrita



através de um modelo não linear  $g$ , de tal forma que

$$\theta_{j,t+1} = g(\theta_{j,t}, \epsilon), \quad (4.14)$$

e  $\epsilon$  seja proveniente de um modelo a ser definido.

No caso da série observada, seja

$$P(X_{jt}|p_{jt}) = p_{jt}^{x_{jt}}(1 - p_{jt})^{1-x_{jt}} \quad ; \quad x_{jt} \in \{0, 1\}, \quad (4.15)$$

e

$$p_{jt} = p_{jt}(\theta_{jt}) = g(a_t, b_t, c_t, \theta_{jt}) \quad ; \quad t = 1, 2, \dots, n, \quad (4.16)$$

em que  $p_{jt}$  é a probabilidade de acerto ao item  $t$  de um indivíduo com proficiência  $\theta_{jt}$  considerando-se  $(a_t, b_t, c_t)$  conhecido.

No caso do modelo logístico de 3 parâmetros,

$$p_{jt}(\theta_{jt}) = c_t + \frac{1 - c_t}{1 + \exp[-a_t(\theta_{jt} - b_t)]}. \quad (4.17)$$

Em termos de um modelo dinâmico, pode-se definir as equações de *observação* e do *sistema*, como sendo:

- Equação das observações:

$$\begin{aligned} P(X_{jt}|p_{jt}(\theta_{jt})) &= p_{jt}(\theta_{jt})^{x_{jt}}(1 - p_{jt}(\theta_{jt}))^{1-x_{jt}}, \\ p_{jt}(\theta_{jt}) &= g_1(a_t, b_t, c_t, \theta_{jt}). \end{aligned} \quad (4.18)$$

- Equação do sistema:

$$\theta_{j,t+1} = g_2(\theta_{j,t}; a_t, b_t, c_t, \epsilon). \quad (4.19)$$

## 4.4 Pressupostos diversos

Antes de apresentar os processos utilizados para estimar as proficiência segundo duas formulações que consideram modelos não lineares, discute-se nesta seção, alguns pressupostos e considerações necessários para a aplicação dos modelos a seguir apresentados.

Como dito anteriormente, para que o conjunto de respostas dadas às  $n$  questões aplicadas a um indivíduo  $j$ , considerado como um processo observável  $\{\mathbf{X}_{jt}\}$ ,  $t = 1, 2, \dots, n$ , seja modelado dentro do contexto de modelos dinâmicos não lineares, via estimação bayesiana, alguma transformação desta série será utilizada, de modo que sob a luz da suposição de intercambialidade, as distribuições a priori utilizadas no processo de estimação dos parâmetros do modelo devem ser assumidas ter a mesma forma. O que se espera é que esta suposição, reduza os efeitos de diferentes classificações dos indivíduos ao se atribuir diferentes distribuições a priori. Mas, neste caso, de onde vem os valores de  $\theta$  ?

Segundo [Van Der Linden e Hambleton \(1997\)](#) o fato de que esses valores são de fato transformações estritamente monótona das respostas dos examinandos, é invocado. O processo de estimativa começa, assim como a maioria das técnicas de modelagem de respostas a itens, com uma classificação preliminar dos examinandos induzida por alguma estatística

adequada. Para exames de múltipla escolha, por exemplo, essa estatística pode ser o número de itens corretos e, para escalas psicológicas, pode ser a pontuação da escala. Ou pode vir de um teste completamente diferente ou de algum outro teste.

As classificações assim fornecidas são substituídas pelos quantis correspondentes da distribuição normal padrão, e são esses valores que fornecem o valor de  $\theta$ . Por quê a distribuição normal é usualmente atribuída aos hiperparâmetros? Em parte, porque tradicionalmente a distribuição de valores de traços latentes têm sido considerada como normalmente distribuída. Mas, na verdade, qualquer distribuição alvo pode ser utilizada. E neste sentido, as transformações utilizadas neste estudo, consideram a distribuição beta por razões expostas ao longo do texto e os hiperparâmetros associados à distribuição dos parâmetros dos itens como provenientes de uma  $N(0, 1)$ .

Outra questão a ser considerada é quanto à métrica de avaliação global dos modelos estudados. Muito embora diferentes medidas possam ser utilizadas para avaliar a qualidade dos ajustes, neste estudo elencou-se o critério RMSE (*Raiz Quadrada do Erro Quadrático Médio*) em razão da simplicidade de compreensão de seus resultados. Outras medidas poderiam ser utilizadas para o mesmo propósito, a exemplo do MAE (*Erro Absoluto Médio*). Em uma breve comparação entre as duas métricas, observa-se que o RMSE dá maior peso a erros maiores, o que na prática é desejável, já que enquanto estudo de simulação, é desejado eliminar para resultados de estimadores com grandes erros. Outro ponto é que o RMSE é sempre maior ou igual ao MAE, o que significa que se os tamanhos de amostras utilizadas para compara fossem muito grandes o RMSE poderia não ser adequado. Porém, nesta dissertação, os testes realizados levam em consideração provas de tamanho fixo, sendo considerados CATs com no máximo 100 questões.

## 4.5 Estimação de proficiência via Filtro de Kalman com $c_t$ não nulo

No contexto de um teste adaptativo, considere  $\{\mathbf{X}_t\} = \{X_1, X_2, \dots, X_n\}$  ser um processo aleatório binário, onde  $X_t$  assume o valor 1 se o item  $t$  é corretamente respondido e 0, caso contrário. Em um modelo logístico de 3 parâmetros, tal como apresentado em (2.23), a probabilidade de sucesso de acerto do item,  $p_t$ , é escrita como

$$p_t = P(X_t = 1 | \theta_j, a_t, b_t, c_t) = c_t + \frac{1 - c_t}{1 + e^{-a_t(\theta_j - b_t)}}, \quad (4.20)$$

onde  $\theta_j$  é o fator latente, ou proficiência do  $j$ -ésimo examinando, e  $a_t$ ,  $b_t$  e  $c_t$  são os parâmetros (conhecidos) do  $t$ -ésimo item. Considerando a reparametrização

$$\eta_t = a_t(\theta_j - b_t), \quad (4.21)$$

a probabilidade de sucesso, expressa em (4.20), pode ser reescrita em termos de  $\eta_t \in \mathfrak{R}$ , como,

$$p_t = \frac{1 + c_t e^{-\eta_t}}{1 + e^{-\eta_t}}. \quad (4.22)$$

Assim,  $\{\mathbf{X}_t\}$  descreve um processo de *Bernoulli* com probabilidade de sucesso  $p_t \geq c$ , de

modo que

$$\begin{aligned} P(X_t = x_t | \eta_t) &= p_t^{x_t} (1 - p_t)^{1-x_t} \\ &= \frac{[1 + c_t e^{-\eta_t}]^{x_t} [(1 - c_t) e^{-\eta_t}]^{1-x_t}}{1 + e^{-\eta_t}} \quad ; \quad 0 < c_t < 1, \end{aligned} \quad (4.23)$$

onde  $x_t = 0$  ou  $1$ .

Considerando a função hipergeométrica (Gradshteyn e Ryzhik, 2007, pág. 1005) dada por

$$F(\alpha, \beta, \gamma, z) = \frac{1}{B(\beta, \gamma - \beta)} \int_0^1 q^{\beta-1} (1 - q)^{\gamma-\beta-1} (1 - qz)^{-\alpha} dq \quad \gamma > \beta > 0. \quad (4.24)$$

Considerando  $q = -e^{-\eta_t}$  e substituindo

$$\begin{aligned} \beta &= b_t \\ \alpha &= a_t \\ z &= 1 - Z \\ \beta - \gamma + 1 &= d - a_t, \end{aligned}$$

em (4.24), tem-se

$$\int_{\mathbb{R}} \frac{e^{-\eta_t b_t} [1 + (1 - Z) e^{-\eta_t}]^{-a_t}}{[1 + e^{-\eta_t}]^{d-a_t}} d\eta_t = B(b_t, d - b_t) F(a_t, b_t, d; Z). \quad (4.25)$$

Para calcular a densidade marginal expressa por

$$P(X_t) = \int_{\mathbb{R}} P(x_t | \eta_t) f(\eta_t) d\eta_t, \quad (4.26)$$

como função do produto da probabilidade condicional de  $X_t | \eta_t$ , pela densidade marginal de  $\eta_t$ ,

$$f(\eta_t) = \frac{1}{B(\alpha, \beta)} \frac{e^{-\alpha \eta_t}}{[1 + e^{-\eta_t}]^{\alpha+\beta}}, \quad (4.27)$$

reparametriza-se

$$e^{-\eta_t} = e^{-a_t(\theta_t - b_t)} = h_t, \quad (4.28)$$

de modo que considerando  $h_t \sim BST(\alpha, \beta)$ , obtém-se uma distribuição *Beta do Segundo Tipo* (Aral e Gupta, 2016; Miheesan, 2004; Stancu, 1995), expressa por

$$f(h_t) = \frac{h_t^{\alpha-1} (1 + h_t)^{-\alpha-\beta}}{B(\alpha, \beta)} \quad ; \quad h_t \geq 0. \quad (4.29)$$

Assim, considerando a transformação logaritmica em (4.28),

$$-\log(h_t) = \eta_t, \quad (4.30)$$

segundo Nadarajah (2005), o parâmetro  $\eta_t \sim BetaExp(\alpha, \beta)$  tem distribuição *Beta Exponenciada*. De maneira semelhante, aplicando o logaritmo em (4.28) e diferenciando ambos

os termos, tem-se

$$\begin{aligned} -\ln h_t &= \eta_t \\ -\frac{dh_t}{h_t} &= d\eta_t, \end{aligned} \quad (4.31)$$

que substituindo em (4.29), resulta em

$$\begin{aligned} f(h_t)d(h_t) &= \frac{h_t^{\alpha-1}(1+h_t)^{-\alpha-\beta}}{B(\alpha, \beta)} dh_t \\ &= \frac{h_t^\alpha(1+h_t)^{-\alpha-\beta}}{B(\alpha, \beta)} \frac{dh_t}{h_t} \\ &= \frac{e^{-\alpha\eta_t}(1+e^{-\eta_t})^{-\alpha-\beta}}{B(\alpha, \beta)} d\eta_t, \end{aligned} \quad (4.32)$$

de forma que a densidade marginal apresentada em (4.27) é expressa por

$$\begin{aligned} f(\eta_t) &= \frac{1}{B(\alpha, \beta)} \frac{e^{-\alpha\eta_t}}{[1+e^{-\eta_t}]^{\alpha+\beta}} \\ &\quad \text{Beta Exponenciada} \\ &= \frac{1}{B(\alpha, \beta)} \left( \frac{1}{1+e^{-\eta_t}} \right)^\beta \left( \frac{e^{-\eta_t}}{1+e^{-\eta_t}} \right)^\alpha, \end{aligned} \quad (4.33)$$

e tem distribuição *Beta Exponenciada*, com parâmetros  $\alpha$  e  $\beta$ .

Portanto, substituindo (4.23), (4.25) e (4.33) em (4.26), tem-se um resultado fechado para

$$\begin{aligned} P(X_t = x_t) &= \int_{\mathbb{R}} \frac{[1+c_t e^{-\eta_t}]^{x_t} [(1-c_t)e^{-\eta_t}]^{1-x_t}}{1+e^{-\eta_t}} \frac{1}{B(\alpha, \beta)} \frac{e^{-\alpha\eta_t}}{[1+e^{-\eta_t}]^{\alpha+\beta}} d\eta \\ &= \frac{(1-c_t)^{1-x_t}}{B(\alpha, \beta)} \int_{\mathbb{R}} \frac{e^{-\eta_t(\alpha+1-x_t)} [1+c_t e^{-\eta_t}]^{x_t}}{[1+e^{-\eta_t}]^{\alpha+\beta+1}} d\eta_t \\ &= \frac{(1-c_t)^{1-x_t}}{B(\alpha, \beta)} B(\alpha+1-x_t, \beta) F(-x_t, \alpha+1-x_t, \alpha+\beta+1-x_t, 1-c_t), \end{aligned} \quad (4.34)$$

onde os parâmetros  $a_t, b_t, c_t, d$  e  $Z$  de (4.25) são expressos por

$$\begin{aligned} a_t &= -x_t \\ b_t &= \alpha + 1 - x_t \\ c_t &= 1 - Z \\ d &= \alpha + \beta + 1 - x_t \\ d - a_t &= \alpha + \beta + 1 \\ d - b_t &= \alpha + \beta + 1 - x_t - \alpha - 1 + x_t = \beta. \end{aligned}$$

Considerando-se  $X_t$  uma v.a. com distribuição de Bernoulli, tem-se

$$\begin{aligned} P(X_t = 0) &= \frac{(1 - c_t)}{B(\alpha, \beta)} B(\alpha + 1, \beta) F(0, \alpha + 1, \alpha + \beta + 1, 1 - c_t) \\ &= \frac{\alpha}{\alpha + \beta} - c_t \frac{\alpha}{\alpha + \beta}, \end{aligned} \quad (4.35)$$

$$\begin{aligned} P(X_t = 1) &= F(-1, \alpha, \alpha + \beta, 1 - c_t) \\ &= \frac{\beta}{\alpha + \beta} + c_t \frac{\alpha}{\alpha + \beta}. \end{aligned} \quad (4.36)$$

A partir da densidade condicional de  $\eta_t|x_t$

$$\begin{aligned} f(\eta_t|x_t) &= \frac{P(x_t|\eta_t)f(\eta_t)}{P(X_t)} \\ &= \frac{[1 + c_t e^{-\eta_t}]^{x_t} [(1 - c_t)e^{-\eta_t}]^{1-x_t}}{1 + e^{-\eta_t}} \frac{1}{B(\alpha, \beta)} \frac{e^{-\alpha\eta_t}}{[1 + e^{-\eta_t}]^{\alpha+\beta}} \times \\ &\times \frac{B(\alpha, \beta)}{(1 - c_t)^{1-x_t}} \frac{1}{B(\alpha + 1 - x_t, \beta) F(-x_t, \alpha + 1 - x_t, \alpha + \beta + 1 - x_t, 1 - c_t)} \\ &= \frac{[1 + c_t e^{-\eta_t}]^x e^{-\eta_t(\alpha+1-x_t)}}{B(\alpha + 1 - x_t, \beta) F(-x_t, \alpha + 1 - x_t, \alpha + \beta + 1 - x_t, 1 - c_t)} \times \\ &\times \frac{1}{[1 + e^{-\eta_t}]^{\alpha+\beta+1}}, \end{aligned} \quad (4.37)$$

e considerando a relação entre a função hipergeométrica  $F(0; \beta; \gamma; z) = F(\alpha; 0; \gamma; z) = 1$  (Gradshteyn e Ryzhik, 2007, pág. 1013) avalia-se os valores da v.a. Bernoulli, tal como em (4.35) e (4.36), de modo que

$$\begin{aligned} f(\eta_t|x_t = 0) &= \frac{e^{-\eta_t(\alpha+1)}}{B(\alpha + 1, \beta) F(0, \alpha + 1, \alpha + \beta + 1, 1 - c_t) [1 + e^{-\eta_t}]^{\alpha+\beta+1}} \\ &= \frac{e^{-\eta_t(\alpha+1)}}{B(\alpha + 1, \beta) [1 + e^{-\eta_t}]^{\alpha+\beta+1}}, \end{aligned} \quad (4.38)$$

$$\begin{aligned} f(\eta_t|x = 1) &= \frac{e^{-\eta_t\alpha} [1 + c_t e^{-\eta_t}]}{B(\alpha, \beta) F(-1, \alpha, \alpha + \beta, 1 - c_t) [1 + e^{-\eta_t}]^{\alpha+\beta+1}} \\ &= \frac{e^{-\eta_t\alpha} [1 + c_t e^{-\eta_t}]}{B(\alpha, \beta) [1 + e^{-\eta_t}]^{\alpha+\beta+1}} \frac{\alpha + \beta}{\beta + c_t\alpha} \\ &= \frac{1}{\beta + c_t\alpha} \left\{ \beta \frac{e^{-\eta_t\alpha}}{B(\alpha, \beta + 1) [1 + e^{-\eta_t}]^{\alpha+\beta+1}} + c_t\alpha \frac{e^{-\eta_t(\alpha+1)}}{B(\alpha + 1, \beta) [1 + e^{-\eta_t}]^{\alpha+\beta+1}} \right\}. \end{aligned} \quad (4.39)$$

Segundo as expressões (4.38) e (4.39), e considerando a avaliação da distribuição *Beta Exponenciada* (Nadarajah, 2005), surgem distintas distribuições de modo que para (4.38), têm-se

$$\eta_t|X_t = 0 \sim BetaExp(\alpha + 1, \beta), \quad (4.40)$$

e para (4.39) uma distribuição mista, dada por

$$\eta_t | X_t = 1 \sim \frac{\beta}{\beta + c_t \alpha} \text{BetaExp}(\alpha, \beta + 1) + \frac{c_t \alpha}{\beta + c_t \alpha} \text{BetaExp}(\alpha + 1, \beta). \quad (4.41)$$

A esperança e a variância de (4.40), são dadas por

$$E(\eta_t | X_t = 0) = \Psi(\beta) - \Psi(\alpha + 1) = \Psi(\beta) - \Psi(\alpha) - \frac{1}{\alpha}, \quad (4.42)$$

$$\text{Var}(\eta_t | X_t = 0) = \Psi'(\alpha + 1) + \Psi'(\beta) = \Psi'(\alpha) + \Psi'(\beta) - \frac{1}{\alpha^2}, \quad (4.43)$$

onde a expressão  $\Psi(\alpha + 1) = \Psi(\alpha) + \frac{1}{\alpha}$  pode ser encontrada em Gradshteyn e Ryzhik (2007, pág. 904, 8.365-1) e são válidas as relações

$$\Psi'(Z + 1) = \Psi'(Z) - \frac{1}{Z^2}.$$

$$\Psi'(Z) = \frac{\Gamma''(Z)}{\Gamma(Z)} - \Psi^2(Z).$$

Avaliando os momentos de (4.41), têm-se

$$\bullet \quad E[\eta_t^2 | X_t = 1] = \Psi'(\alpha) + \Psi'(\beta) + \Psi^2(\alpha) + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) + 2\frac{(1 - c_t)}{\beta + c_t \alpha} \{\Psi(\beta) - \Psi(\alpha)\}. \quad (4.44)$$

$$\bullet \quad E^2[\eta_t | X_t = 1] = \Psi^2(\alpha) + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) + 2\frac{(1 - c_t)}{\beta + c_t \alpha} [\Psi(\beta) - \Psi(\alpha)] + \left[ \frac{1 - c_t}{\beta + c_t \alpha} \right]^2. \quad (4.45)$$

Portanto, de (4.44) e (4.45) têm-se a esperança e a variância de (4.41) expressas por

$$E(\eta_t | X_t = 1) = [\Psi(\beta) - \Psi(\alpha)] + \frac{1 - c_t}{\beta + c_t \alpha}. \quad (4.46)$$

$$\text{Var}(\eta_t | X_t = 1) = \Psi'(\alpha) + \Psi'(\beta) - \left[ \frac{1 - c_t}{\beta + c_t \alpha} \right]^2. \quad (4.47)$$

### 4.5.1 Distribuição a Priori

Na busca por uma distribuição à priori para ser utilizada em conjunto com a verossimilhança de  $\{\mathbf{X}_t\}$ , a escolha sugerida em (4.33), expressa por

$$f(\eta_t) = \frac{1}{B(\alpha, \beta)} \left( \frac{1}{1 + e^{-\eta_t}} \right)^\beta \left( \frac{e^{-\eta_t}}{1 + e^{-\eta_t}} \right)^\alpha,$$

é proveniente da convolução obtida através da transformação logito  $\pi = \frac{e^{-\eta}}{1 + e^{-\eta}}$ , de onde segue a relação

$$\eta_t = \ln \left\{ \frac{1 - \pi_t}{\pi_t} \right\}, \quad (4.48)$$

em que  $0 < \pi_t < 1$  segue a clássica distribuição  $Beta(\alpha_t, \beta_t)$ , com  $a_t > 0$  e  $b_t > 0$ . O que resulta na função de densidade conhecida como modelo *Logito-Beta*, isto é,  $\eta_t \sim LogitBeta(\alpha_t, \beta_t)$ , também conhecida por *Beta Exponenciada*, dada por

$$f(\eta_t) = \frac{1}{B(\alpha_t, \beta_t)} \frac{e^{-\alpha_t \eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t}}, \quad (4.49)$$

para  $-\infty < \eta_t < \infty$ .

### 4.5.2 Estimação de $\eta$ pelo Método EAP

No intuito de estimar os parâmetros  $\alpha_t$  e  $\beta_t$  de (4.49), suas propriedades são inicialmente avaliadas através do cálculo de sua FGM (*Função Geradora de Momentos*), dada por

$$M(q) = E[e^{q\eta_t}] = \int_{\mathfrak{R}} \frac{1}{B(\alpha, \beta)} \frac{e^{-(\alpha-q)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha+\beta}} d\eta_t = \frac{B(\alpha - q, \beta + q)}{B(\alpha, \beta)}, \quad q \leq \alpha. \quad (4.50)$$

Utilizando a relação  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$  para calcular a primeira e a segunda derivada de (4.50) em relação à  $q$ , tem-se

- $\frac{\partial M(q)}{\partial q} = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{-\Gamma'(\alpha - q)\Gamma(\beta + q) + \Gamma(\alpha - q)\Gamma'(\beta + q)\}. \quad (4.51)$

- $\frac{\partial^2 M(q)}{\partial q^2} = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{\Gamma''(\alpha - q)\Gamma(\beta + q) - \Gamma'(\alpha - q)\Gamma'(\beta + q) - \Gamma'(\alpha - q)\Gamma'(\beta + q) + \Gamma(\alpha - q)\Gamma''(\beta + q)\}. \quad (4.52)$

Lembrando que a função digamma ( $\Psi$ ) é definida como  $\Psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ , observa-se seu caráter recursivo ao calcular a primeira derivada, expressa por

- $\Psi'(\alpha) = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \left[ \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right]^2 = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - [\Psi(\alpha)]^2. \quad (4.53)$

Avaliando (4.51) e (4.52) em  $q = 0$ , têm-se

- $E(\eta_t) = \left. \frac{\partial M(q)}{\partial q} \right|_{q=0} = \Psi(\beta) - \Psi(\alpha). \quad (4.54)$

- $E(\eta_t^2) = \left. \frac{\partial^2 M(q)}{\partial q^2} \right|_{q=0} = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - 2\Psi(\alpha)\Psi(\beta) + \frac{\Gamma''(\beta)}{\Gamma(\beta)}. \quad (4.55)$

Portanto, de (4.54) e (4.55), têm-se

- $Var(\eta_t) = E(\eta_t^2) - [E(\eta_t)]^2 = \Psi'(\alpha) + \Psi'(\beta). \quad (4.56)$

De forma que a esperança e a variância para a distribuição *Beta Exponenciada* apresentada em (4.49), são dadas por

$$E(\eta_t) = \Psi(\beta_t) - \Psi(\alpha_t). \quad (4.57)$$

$$Var(\eta_t) = \Psi'(\alpha_t) + \Psi'(\beta_t). \quad (4.58)$$

Portanto,

$$\boxed{\eta_t \sim BetaExp(\alpha_t, \beta_t)} \quad (4.59)$$

tem distribuição *Beta Exponenciada* (Nadarajah, 2005).

Embora em um contexto diferente dos testes adaptativos, o trabalho de Das e Das (2009) discute sobre o problema de inferência estatística de proporções binomiais em procedimentos adaptativos no contexto Bayesiano, sugerindo um procedimento heurístico de decisão para conduzir a inferência.

### 4.5.3 Elicitação da priori $BetaExp(\alpha_t, \beta_t)$ de $\eta_t$

No intuito de avaliar os valores dos hiperparâmetros  $\alpha_t$  e  $\beta_t$  da priori *Beta Exponenciada* e sua relação com os parâmetros conhecidos  $a_t$  e  $b_t$  do modelo 3PL, são apresentados os detalhes da reparametrização utilizada, além de demonstrar de que maneira se relacionam os parâmetros do modelo 3PL com os parâmetros da *Beta Exponenciada*.

Considerando-se que em (4.57):

$$Se \begin{cases} E(\eta_t) > 0, & \Psi(\beta_t) > \Psi(\alpha_t) \\ E(\eta_t) < 0, & \Psi(\beta_t) < \Psi(\alpha_t) \end{cases} \quad (4.60)$$

Sabe-se que

$$\Psi(Z + 1) = \Psi(Z) + \frac{1}{Z}.$$

(I) Supondo que  $\alpha_0 = \beta_0 + 1 > 0$

$$\begin{aligned} E(\eta_0) &= \Psi(\beta_0) - \Psi(\beta_0 + 1) \\ &= \cancel{\Psi(\beta_0)} - \cancel{\Psi(\beta_0)} - \frac{1}{\beta_0}. \end{aligned}$$

Logo,

$$\boxed{Se E(\eta_0) < 0 \begin{cases} \beta_0 = -\frac{1}{E(\eta_0)} \\ \alpha_0 = \beta_0 + 1. \end{cases}} \quad (4.61)$$

(II) Supondo que  $\beta_0 = \alpha_0 + 1 > 0$

$$\begin{aligned} E(\eta_0) &= \Psi(\beta_0) - \Psi(\alpha_0) \\ &= \cancel{\Psi(\alpha_0)} + \frac{1}{\beta_0} - \cancel{\Psi(\alpha_0)}. \end{aligned}$$

Logo,

$$\boxed{Se E(\eta_0) > 0 \begin{cases} \alpha_0 = \frac{1}{E(\eta_0)} \\ \beta_0 = \alpha_0 + 1. \end{cases}} \quad (4.62)$$



### 4.5.4 Distribuição a Posteriori

Considerando a aplicação do Teorema de Bayes e as seguintes densidades:

- **Verossimilhança**: Processo *Bernoulli*  $\{X_t\}$  com reparametrização  $\eta_t = a_t(\theta_j - b_t)$

$$P(X_t = x_t | \eta_t) = \frac{[1 + c_t e^{-\eta_t}]^{x_t} [(1 - c_t) e^{-\eta_t}]^{1-x_t}}{1 + e^{-\eta_t}} \quad ; \quad 0 < c_t < 1, \quad (4.63)$$

- **Priori**:  $\eta_t \sim BetaExp(\alpha_t, \beta_t)$

$$f(\eta_t) = \frac{1}{B(\alpha_t, \beta_t)} \frac{e^{-\alpha_t \eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t}} \quad ; \quad -\infty < \eta_t < \infty, \quad (4.64)$$

obtêm-se a preditiva de  $X_t$  como distribuição incondicional de Bernoulli, dada por

$$\begin{aligned} P(X_t) &= \int_{\mathfrak{R}} P(X_t | \eta_t) f(\eta_t) d\eta_t \\ &= \frac{[(1 - c_t)\alpha_t]^{1-x_t} [\beta_t + c_t\alpha_t]^{x_t}}{\alpha_t + \beta_t}. \end{aligned} \quad (4.65)$$

Portanto, a distribuição a posteriori resultante para atualização do processo, é dada por

$$\begin{aligned} f(\eta_t | X_t) &= \frac{P(X_t | \eta_t) f(\eta_t)}{P(X_t)} \\ &= \frac{\alpha_t + \beta_t}{\alpha_t^{1-x_t} B(\alpha_t, \beta_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{\beta_t + c_t \alpha_t} \right]^{x_t} \frac{e^{-(\alpha_t + 1 - x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}} \\ &= \frac{1}{B(\alpha_t + 1 - x_t, \beta_t + x_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{1 + \frac{c_t \alpha_t}{\beta_t}} \right]^{x_t} \frac{e^{-(\alpha_t + 1 - x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}}. \end{aligned} \quad (4.66)$$

Assim, uma vez que a  $t$ -ésima observação estiver disponível, a distribuição a posteriori expressa em (4.66) segue uma variação da distribuição *Beta Exponenciada de Libby e Novick* (Nadarajah, 2005), com parâmetros  $\alpha_t + 1 - x_t$  e  $\beta_t + x_t$ , podendo ser reescrita tal como apresentada pela densidade condicional calculada em (4.37), e expressa por

$$\begin{aligned} f(\eta_t | x_t) &= \frac{[1 + c_t e^{-\eta_t}]^x e^{-\eta_t(\alpha + 1 - x_t)}}{B(\alpha + 1 - x_t, \beta) F(-x_t, \alpha + 1 - x_t, \alpha + \beta + 1 - x_t, 1 - c_t)} \times \\ &\times \frac{1}{[1 + e^{-\eta_t}]^{\alpha + \beta + 1}}. \end{aligned}$$

Em particular, dado  $X_t = 0$ , obtém-se

$$f(\eta_t | X_t = 0) = \frac{e^{-(\alpha + 1)\eta_t}}{B(\alpha + 1, \beta_t) [1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}}, \quad (4.67)$$

onde  $\eta_t | X_t \sim LogitBeta(\alpha_t + 1, \beta_t)$ . Quando  $X_t = 1$ , (4.66) é reescrita como

$$f(\eta_t | X_t = 1) = \frac{1}{\beta_t + c_t \alpha_t} \left[ \beta_t \frac{e^{-\alpha_t \eta_t}}{B(\alpha_t, \beta_t + 1) [1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}} + c_t \alpha_t \frac{e^{-(\alpha_t + 1)\eta_t}}{B(\alpha_t + 1, \beta_t) [1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}} \right], \quad (4.68)$$

representando uma mistura entre a  $LogitBeta(\alpha_t, \beta_t + 1)$  e a  $LogitBeta(\alpha_t + 1, \beta_t)$ . Assim, avaliando a esperança e a variância a posteriori, para as respostas incorretas ( $X_t = 0$ ) ao item  $t$ , têm-se

$$E[\eta_t | X_t = 0] = \Psi(\beta_t) - \Psi(\alpha_t) - \frac{1}{\alpha_t}, \quad (4.69)$$

$$Var[\eta_t | X_t = 0] = \Psi'(\alpha_t) + \Psi'(\beta_t) - \frac{1}{\alpha_t^2}. \quad (4.70)$$

E para as respostas corretas ( $X_t = 1$ ) ao item  $t$ , têm-se

$$E[\eta_t | X_t = 1] = \Psi(\beta_t) - \Psi(\alpha_t) + \frac{1 - c_t}{\beta_t + c_t \alpha_t}, \quad (4.71)$$

$$Var[\eta_t | X_t = 1] = \Psi'(\alpha_t) + \Psi'(\beta_t) - \left[ \frac{1 - c_t}{\beta_t + c_t \alpha_t} \right]^2. \quad (4.72)$$

Portanto, o valor esperado de  $\eta_t$ , tal como obtido em (4.57), e expresso por

$$E(\eta_t) = \Psi(\beta_t) - \Psi(\alpha_t),$$

é reduzido por uma resposta incorreta (4.69), ou aumentado por uma resposta correta (4.71).

### 4.5.5 Distribuição Preditiva

Obtida a  $t$ -ésima observação, a distribuição  $(\eta_t | X_t) \sim Logit-LibbyNovick(\alpha_t + 1 - x_t, \beta_t + x_t)$  apresenta forma fechada distinta para as respostas corretas e incorretas, expressas respectivamente em (4.67) e (4.68).

Para a predição de uma observação um passo a frente, postula-se que

$$c_t = 0, \quad (4.73)$$

$$h_t = e^{-a_t(\theta - b_t)} = e^{-\eta_t}, \quad (4.74)$$

onde  $h_t \sim BST(\alpha_t, \beta_t)$  é conhecida na literatura por distribuição *Beta do Segundo Tipo*, apresentada em (4.29), e expressa por

$$f(h_t) = \frac{h_t^{\alpha-1} (1+h_t)^{-\alpha-\beta}}{B(\alpha, \beta)} \quad ; \quad h_t \geq 0.$$

A esperança e a variância desta distribuição são dadas por

$$E[h_t] = \frac{\alpha_t}{\beta_t - 1}, \quad (4.75)$$

$$Var[h_t] = \frac{\alpha_t(\alpha_t + \beta_t - 1)}{(\alpha_t - 1)^2(\beta_t - 2)}. \quad (4.76)$$

Portanto, de (4.73) e (4.74) deriva-se a distribuição conjugada a priori, *Beta Exponenciada* (Barreto-Souza *et al.*, 2010; Nadarajah, 2005),  $\eta_t \sim BetaExp(\alpha_t, \beta_t)$ , apresentada como a distribuição a priori em (4.64), sendo uma generalização da distribuição *Beta do Segundo*

Tipo, resultante da convolução logito  $\eta_t = \ln \left\{ \frac{1-\pi_t}{\pi_t} \right\}$ , e expressa por

$$f(\eta_t) = \frac{1}{B(\alpha_t, \beta_t)} \left( \frac{1}{1 + e^{-\eta_t}} \right)^{\beta_t} \left( \frac{e^{-\eta_t}}{1 + e^{-\eta_t}} \right)^{\alpha_t}. \quad (4.77)$$

Assim, a verossimilhança do modelo quando a  $t$ -ésima observação está disponível, é dada por

$$\begin{aligned} P(X_t = x_t | h_t) &= \left( \frac{1}{1 + h_t} \right)^{x_t} \left( \frac{h_t}{1 + h_t} \right)^{1-x_t}, \\ &= \left( \frac{1}{1 + e^{-\eta_t}} \right)^{x_t} \left( \frac{e^{-\eta_t}}{1 + e^{-\eta_t}} \right)^{1-x_t}, \\ &= P(X_t = x_t | \eta_t). \end{aligned} \quad (4.78)$$

Considerando como verossimilhança a expressão apresentada em (4.78) e como priori a expressão apresentada em (4.64), têm-se a preditiva de  $X_t$ , dada por

$$\begin{aligned} P(X_t) &= \int_{\mathfrak{R}} P(X_t = x_t | \eta_t) f(\eta_t) d\eta_t, \\ &= \frac{1}{B(\alpha_t, \beta_t)} \int_{\mathfrak{R}} \frac{e^{-[\alpha_t+1-x_t]\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t+\beta_t+1}} d\eta_t, \\ &= \frac{B(\alpha_t + 1 - x_t, \beta_t + x_t)}{B(\alpha_t, \beta_t)}, \end{aligned} \quad (4.79)$$

e como  $X_t$  assume valores 0 ou 1, tem-se

$$\bullet \quad P(X_t = 1) = \frac{B(\alpha_t, \beta_t + 1)}{B(\alpha_t, \beta_t)} = \frac{\beta_t}{\alpha_t + \beta_t}. \quad (4.80a)$$

$$\bullet \quad P(X_t = 0) = \frac{B(\alpha_t + 1, \beta_t)}{B(\alpha_t, \beta_t)} = \frac{\alpha_t}{\alpha_t + \beta_t}. \quad (4.80b)$$

### 4.5.6 Um modelo dinâmico para o CAT via Filtro de Kalman

Definida a priori de  $\eta_t$ , e determinada a distribuição preditiva da série observada  $x_t \in \{0, 1\}$ , a densidade da distribuição à posteriori quando é utilizada a reparametrização da *Beta Exponenciada*, é dada por

$$f(\eta_t | x_t) = \frac{1}{B(\alpha_t + 1 - x_t, \beta_t + x_t)} \frac{e^{-[\alpha_t+1-x_t]\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t+\beta_t+1}}. \quad (4.81)$$

De forma que, considerando a atualização dos parâmetros da *Beta Exponenciada* para mudança de estado, tem-se

$$\begin{cases} \dot{\alpha}_t = \alpha_{t+1|t} \\ \dot{\beta}_t = \beta_{t+1|t} \end{cases}.$$

Assim, a série observada é atualizada passando por um estado transiente, indicado pela mudança da variável de interesse, podendo ser entendida como

$$\begin{array}{ccccc} & \curvearrowright & & & \\ x_t & & \eta_{t+1} & & x_{t+1}, \\ & & \curvearrowleft & & \end{array} \quad (4.82)$$

onde a função de densidade dada em (4.81) atualiza os parâmetros expressos em (4.82) através de

$$f(\eta_{t+1}|x_t) = \frac{1}{B(\dot{\alpha}_t, \dot{\beta}_t)} \frac{e^{-\dot{\alpha}_t \eta_t}}{[1 + e^{-\eta_t}]^{\dot{\alpha}_t + \dot{\beta}_t}}, \quad (4.83)$$

de forma que (4.83) passa agora a ser a priori, de “*parâmetros atualizados*”, para o novo ciclo do processo aleatório representado pela variável  $X_t$ , atualizado pela posteriori

$$\begin{aligned} P(x_{t+1}|x_t) &= \int_{\mathbb{R}} f(\eta_{t+1}|x_t) P(x_{t+1}|\eta_{t+1}) d\eta_{t+1}, \\ &= \int_{\mathbb{R}} \frac{1}{B(\dot{\alpha}_t, \dot{\beta}_t)} \frac{e^{-\dot{\alpha}_t \eta_t}}{[1 + e^{-\eta_t}]^{\dot{\alpha}_t + \dot{\beta}_t}} \left[ \frac{1}{1 + e^{-\eta_t}} \right]^{x_{t+1}} \left[ \frac{e^{-\eta_t}}{1 + e^{-\eta_t}} \right]^{1-x_{t+1}} d\eta_t, \\ &= \frac{1}{B(\dot{\alpha}_t, \dot{\beta}_t)} \int_{\mathbb{R}} \frac{e^{-[\dot{\alpha}_t + 1 - x_{t+1}]\eta_t}}{[1 + e^{-\eta_t}]^{\dot{\alpha}_t + \dot{\beta}_t + 1}} d\eta_t, \\ &= \frac{B(\dot{\alpha}_t + 1 - x_{t+1}, \dot{\beta}_t + x_{t+1})}{B(\dot{\alpha}_t, \dot{\beta}_t)}. \end{aligned} \quad (4.84)$$

O processo aleatório binário  $\{\mathbf{X}_t\}$  tem então um suposto sistema de “*equações de transição*”, dado por

$$\begin{cases} \alpha_{t+1} = \phi \alpha_t + 1 - x_{t+1} \\ \beta_{t+1} = \phi \beta_t + x_{t+1} \end{cases}. \quad (4.85)$$

Lembrando que,

$$B(\alpha + 1, \beta) = B(\alpha, \beta) \cdot \frac{\alpha}{\alpha + \beta},$$

e que de (4.80) tem-se

$$\bullet \quad E(X_t) = \frac{\beta_t}{\alpha_t + \beta_t},$$

sendo similar ao resultado encontrado em (4.84)

$$\bullet \quad E(X_{t+1}|X_t) = \frac{\dot{\beta}_t}{\dot{\alpha}_t + \dot{\beta}_t} = \frac{\phi \beta_t}{\phi(\alpha_t + \beta_t)} = E(X_t),$$

justifica as *equações de transição*

$$\begin{cases} \dot{\alpha}_t = \phi \alpha_t \\ \dot{\beta}_t = \phi \beta_t \end{cases}. \quad (4.86)$$

Portanto, considerando a posteriori apresentada em (4.66), expressa por

$$f(\eta_t|X_t) = \frac{1}{B(\alpha_t + 1 - x_t, \beta_t + x_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{1 + \frac{c_t \alpha_t}{\beta_t}} \right]^{x_t} \frac{e^{-(\alpha_t + 1 - x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}}, \quad (4.87)$$

e supondo agora que o processo é considerado suave,

$$(\eta_{t+1}|X_t) \approx (\eta_t|X_t), \quad (4.88)$$

de modo que

$$f(\eta_{t+1}|x_t) = \frac{e^{-\eta_{t+1}\dot{\alpha}_t}}{B(\dot{\alpha}_t, \dot{\beta}_t)[1 + e^{-\eta_{t+1}}]^{\dot{\alpha}_t + \dot{\beta}_t}}, \quad (4.89)$$

juntamente com a verossimilhança de  $X_t|\eta_t$ , apresentada em (4.23), e expressa por

$$P(X_t = x_t|\eta_t) = \frac{[1 + c_t e^{-\eta_t}]^{x_t} [(1 - c_t) e^{-\eta_t}]^{1-x_t}}{1 + e^{-\eta_t}} \quad ; \quad 0 < c_t < 1,$$

reescrevem a distribuição um passo a frente, expressa em (4.84), como

$$\begin{aligned} P(x_{t+1}|x_t) &= \int_{\mathbb{R}} f(\eta_{t+1}|x_t) P(x_{t+1}|\eta_{t+1}) d\eta_{t+1} \\ &= \int_{\mathbb{R}} \frac{e^{-\eta_{t+1}\dot{\alpha}_t} [1 + c_{t+1} e^{-\eta_{t+1}}]^{x_{t+1}} [(1 - c_{t+1}) e^{-\eta_{t+1}}]^{1-x_{t+1}}}{B(\dot{\alpha}_t, \dot{\beta}_t) [1 + e^{-\eta_{t+1}}]^{\dot{\alpha}_t + \dot{\beta}_t + 1}} d\eta_{t+1} \\ &= \frac{(1 - c_{t+1})^{1-x_{t+1}}}{B(\dot{\alpha}_t, \dot{\beta}_t)} \int_{\mathbb{R}} \frac{e^{-\eta_{t+1}(\dot{\alpha}_t + 1 - x_{t+1})} [1 + c_{t+1} e^{-\eta_{t+1}}]^{x_{t+1}}}{[1 + e^{-\eta_{t+1}}]^{\dot{\alpha}_t + \dot{\beta}_t + 1}} d\eta_{t+1} \\ &= \frac{(1 - c_{t+1})^{1-x_{t+1}}}{B(\dot{\alpha}_t, \dot{\beta}_t)} B(\dot{\alpha}_t + 1 - x_{t+1}, \dot{\beta}_t) \times \\ &\quad \times F(-x_{t+1}, \dot{\alpha}_t + 1 - x_{t+1}, \dot{\alpha}_t + 1 - x_{t+1} + \dot{\beta}_t, 1 - c_{t+1}). \end{aligned} \quad (4.90)$$

A menos dos parâmetros  $\dot{\alpha}_t$ ,  $\dot{\beta}_t$ , e do tempo no qual é avaliada a resposta ( $x_{t+1}$ ), a expressão (4.90) é idêntica ao resultado fechado encontrado para o processo Bernoulli  $\{X_t\}$  em (4.34), expresso por

$$P(X_t = x_t) = \frac{(1 - c_t)^{1-x_t}}{B(\alpha, \beta)} B(\alpha + 1 - x_t, \beta) F(-x_t, \alpha + 1 - x_t, \alpha + \beta + 1 - x_t, 1 - c_t), \quad (4.91)$$

onde

- $P(X_t = 0|x_t) = P(X_t = 0)$ ,
- $P(X_t = 1|x_t) = P(X_t = 1)$ ,

sendo  $P(X_t = 0)$  e  $P(X_t = 1)$  avaliadas, segundo (4.91) como

$$\begin{aligned} P(X_t = 0) &= \frac{\alpha}{\alpha + \beta} - c_t \frac{\alpha}{\alpha + \beta}, \\ P(X_t = 1) &= \frac{\beta}{\alpha + \beta} + c_t \frac{\alpha}{\alpha + \beta}. \end{aligned}$$

Portanto, (4.90) pode ser reescrita como

$$P(x_{t+1}|x_t) = \left( \frac{\dot{\beta}_t + c_{t+1}\dot{\alpha}_t}{\dot{\alpha}_t + \dot{\beta}_t} \right)^{x_{t+1}} \left( \frac{(1 - c_{t+1})\dot{\alpha}_t}{\dot{\alpha}_t + \dot{\beta}_t} \right)^{1-x_{t+1}}, \quad (4.92)$$

e uma vez desejada a predição de um passo à frente, postula-se que  $\eta_{t+1}$  comporta-se como uma distribuição  $BetaExp(\alpha_{t+1|t}, \beta_{t+1|t})$ , dada por

$$f(\eta_{t+1}|X_t) = \frac{1}{B(\alpha_{t+1|t}, \beta_{t+1|t})} \frac{e^{-\alpha_{t+1|t}\eta_{t+1}}}{[1 + e^{-\eta_{t+1}}]^{\alpha_{t+1|t} + \beta_{t+1|t}}}, \quad (4.93)$$

onde os parâmetros  $\alpha_{t+1|t} > 0$  e  $\beta_{t+1|t} > 0$  são encontrados pelas equações de transição

$$\begin{cases} \alpha_{t+1|t} = \phi\alpha_t \\ \beta_{t+1|t} = \phi\beta_t \end{cases}, \quad (4.94)$$

com  $0 < \phi < 1$  sendo o hiperparâmetro do Filtro de Kalman [Harvey \(1989\)](#). Assim, a distribuição preditiva resultante é dada por

$$\begin{aligned} P(X_{t+1}|X_t) &= \int_{\mathbb{R}} P(X_{t+1}|\eta_{t+1})f(\eta_{t+1}|X_t)d\eta_{t+1}, \\ &= \frac{[(1 - c_{t+1})\alpha_{t+1|t}]^{1-x_{t+1}}[\beta_{t+1|t} + c_{t+1}\alpha_{t+1|t}]^{x_{t+1}}}{\alpha_{t+1|t} + \beta_{t+1|t}}, \end{aligned} \quad (4.95)$$

isto é

$$P(X_{t+1} = 0) = \frac{\alpha_{t+1|t}}{\alpha_{t+1|t} + \beta_{t+1|t}} - c_{t+1} \frac{\alpha_{t+1|t}}{\alpha_{t+1|t} + \beta_{t+1|t}}, \quad (4.96)$$

$$P(X_{t+1} = 1) = \frac{\beta_{t+1|t}}{\alpha_{t+1|t} + \beta_{t+1|t}} + c_{t+1} \frac{\alpha_{t+1|t}}{\alpha_{t+1|t} + \beta_{t+1|t}}. \quad (4.97)$$

Uma vez que a  $(t + 1)$ -ésima informação estiver disponível, a distribuição a posteriori  $f(\eta_{t+1}|X_{t+1})$  é obtida aplicando-se o resultado obtido em (4.66), expresso por

$$f(\eta_t|X_t) = \frac{1}{B(\alpha_t + 1 - x_t, \beta_t + x_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{1 + \frac{c_t \alpha_t}{\beta_t}} \right]^{x_t} \frac{e^{-(\alpha_t + 1 - x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t + \beta_t + 1}},$$

resultando na distribuição *Beta Exponenciada de Libby e Novick*, com parâmetros dados pelas equações de atualização

$$\begin{cases} \alpha_{t+1} = \alpha_{t+1|t} + 1 - x_{t+1} \\ \beta_{t+1} = \beta_{t+1|t} + x_{t+1} \end{cases}. \quad (4.98)$$

Portanto, dada as recursões em (4.94) e (4.98), a função de log-verossimilhança para o hiperparâmetro desconhecido  $\phi$  é calculada como

$$l(\phi) = \sum_{t=1}^{n-1} \ln[P(X_{t+1}|X_t)], \quad (4.99)$$

sendo a maximização de seu resultado utilizado nas atualizações das estimativas da proficiência  $\tilde{\theta}_j$ , descrita a seguir. Assim, considerando a probabilidade de sucesso  $P(X_t = 1|\theta_j, a_t, b_t, c_t)$ , apresentada em (4.22), e expressa por

$$p_t = \frac{1 + c_t e^{-\eta_t}}{1 + e^{-\eta_t}},$$

para obter da proficiência desejada, avalia-se o parâmetro  $\eta_t$ , dado por

$$\eta_t = \ln \frac{p_t - c_t}{1 - p_t}, \tag{4.100}$$

de forma que consideram-se:

- as equações de transição:

$$\begin{cases} \alpha_{t+1} = \phi\alpha_t + 1 - x_{t+1} \\ \beta_{t+1} = \phi\beta_t + x_{t+1} \end{cases}, \tag{4.101}$$

- a esperança da distribuição *Beta Exponenciada*:

$$\hat{\eta}_t = \Psi(\beta_t) - \Psi(\alpha_t),$$

- a reparametrização expressa em (4.21), dada por:

$$\hat{\eta}_t = a_t(\tilde{\theta}_{jt} - b_t), \tag{4.102}$$

onde, nesta última, os parâmetros  $a_t$  e  $b_t$  comportam-se como variáveis aleatórias, pois variam de um item para outro. Igualar (4.102) a (4.100), viabiliza a estimação dos parâmetros  $\beta_t$  e  $\theta_{jt}$ , dados por

$$\alpha_t(\theta_{jt} - \beta_t) = \ln \frac{p_t - c_t}{1 - p_t}$$

$$\beta_t = \theta_{jt} - \frac{1}{\alpha_t} \ln \frac{p_t - c_t}{1 - p_t} \tag{4.103}$$

$$\theta_{jt} = \beta_t + \frac{1}{\alpha_t} \ln \frac{1 - p_t}{p_t - c_t} \tag{4.104}$$

Portanto,  $\tilde{\theta}_{jt}$  é expresso por

$$\tilde{\theta}_{jt} = \frac{\hat{\eta}_t}{\hat{a}_t} + \hat{b}_t, \tag{4.105}$$

de forma que, a esperança condicional  $E(\hat{\eta}_t|\hat{a}_t, \hat{b}_t)$ , permite avaliar  $E(\tilde{\theta}_{jt}|a_t, b_t)$ . Entretanto, o objetivo é encontrar  $E(\tilde{\theta}_{jt})$ , de forma que tomando-se a média de  $\tilde{\theta}_{jt}$ , tem-se o estimador desejado:

$$\hat{\theta}_{jt} = \sum_{r=1}^t \frac{\tilde{\theta}_{jr}}{t}. \tag{4.106}$$

Pode-se aplicar outros filtros em vez de (4.106), ou além de (4.106), como por exemplo o *Alisamento Exponencial* sobre o  $\{\tilde{\theta}_{jt}\}$ . Da mesma forma, um segundo filtro pode ser aplicado para tentar reduzir ainda mais as oscilações provenientes de  $\{\tilde{\theta}_{jt}\}$ .

**Algoritmo 1:** Estimaco de proficincias via *Filtro de Kalman*.

---

**Input:** Defina um valor para  $\epsilon$

- 1 Inicializao de variveis
- 2 Considere  $\{X_1, \dots, X_\tau\}$  ser um conjunto inicial de respostas
- 3 Maximize (4.99) com respeito à  $\phi$ , utilizando as recurses apresentadas em (4.94) e (4.98)
- 4 Calcule  $\hat{\eta}_\tau = E[\eta_\tau | X_\tau]$  por (4.69) e (4.71)
- 5 Calcule  $\hat{\nu}_\tau = Var[\eta_\tau | X_\tau]$  por (4.70) e (4.72)
- 6 Obtenha  $\tilde{\theta}_j$  de (4.21)
- 7  $\nu \leftarrow \hat{\nu}_\tau, t \leftarrow \tau$
- 8 **while**  $\nu > \epsilon$  **do**
- 9      $t \leftarrow t + 1$
- 10     Amostre um item com parmetros  $\{a_t, b_t, c_t\}$  e considere sua resposta  $x_t$
- 11     Tomando  $\{X_1, \dots, X_t\}$ , maximize (4.99) em relao à  $\phi$ , utilizando as recurses apresentadas em (4.94) e (4.98)
- 12     Calcule  $\hat{\eta}_t = E[\eta_t | X_t]$  por (4.69) e (4.71)
- 13     Calcule  $\hat{\nu}_t = Var[\eta_t | X_t]$  por (4.70) e (4.72)
- 14     Obtenha  $\tilde{\theta}_j$  de (4.105)
- 15     Obtenha  $\hat{\theta}_j$  de (4.106)
- 16      $\nu \leftarrow \hat{\nu}_t$
- 17 **end**
- 18 **return**  $\hat{\theta}_j$

---

## 4.6 Estimaco de proficincia via Modelo Dinmico Power Steady com $c_t$ nulo

Nesta seo,  apresentada uma formulao alternativa à dada na seo anterior, com a particularidade que assume-se que  $c_t = 0 \forall t$ . Tal pressuposto, faz-se necessrio para alguns desenvolvimentos tericos.

Na expresso (4.17), considere o modelo 3PL em que  $c_t = 0$ , isto ,

$$p_{jt}(\theta_{jt}) = \frac{1}{1 + \exp[-Da_t(\theta_{jt} - b_t)]}, \quad (4.107)$$

com  $a_t$  e  $b_t$  conhecidos e  $D$  sendo um fator de escala, constante e igual a 1. Quando deseja-se que a funo logstica fornea resultados semelhantes ao da funo ogiva normal, utiliza-se o valor 1.702.

Na expresso (4.107) seja

$$\eta_{jt} = Da_t(\theta_{jt} - b_t), \quad (4.108)$$

de modo que

$$\theta_{jt} = \frac{\eta_{jt} + (Da_t b_t)}{Da_t}. \quad (4.109)$$

Por motivos tericos, so parametrizadas as expresses (4.18) e (4.107), em funo de



$\eta_{jt}$ . Na descrição do modelo dinâmico para a estimação das proficiências, o seguinte roteiro é seguido:

- (1) obter uma distribuição a priori para  $\eta_{jt}$ ;
- (2) elicitar a priori descrita em (4.18);
- (3) descrever um Modelo *Power Steady Smith* (1979) para a estimação da proficiência via CAT.

#### 4.6.1 Distribuição a Priori de $\eta_{jt}$

No estabelecimento de uma priori para  $\eta_{jt}$ , primeiramente estabelece-se uma priori para a transformada

$$h_{jt} = e^{-Dat(\theta_{jt}-b_t)} = e^{-\eta_{jt}}, \quad (4.110)$$

e, então, através do método jacobiano aplicado em uma transformação específica, obtém-se a priori de interesse, ou seja, a priori de  $\eta_{jt}$ .

Considerando a probabilidade de sucesso  $p_{jt}$  expressa por

$$p_{jt} = \frac{1}{1 + e^{-\eta_{jt}}} = \frac{1}{1 + h_{jt}}, \quad (4.111)$$

o Modelo *Bernoulli* para as respostas do indivíduo  $j$  é dado por

$$P(X_{jt}|h_{jt}) = \left(\frac{1}{1 + h_{jt}}\right)^{x_{jt}} \left(\frac{h_{jt}}{1 + h_{jt}}\right)^{1-x_{jt}}. \quad (4.112)$$

Como  $h_{jt} = e^{-\eta_{jt}} \geq 0$ , assume-se que  $h_{jt}$  seja modelado, a priori, através de uma distribuição *Beta do Segundo Tipo* (Gupta, 2000; Miheesan, 2004; Stancu, 1995), tal que,

$$f(h_{jt}) = \frac{\Gamma(\gamma_t)\Gamma(\delta_t)}{\Gamma(\gamma_t + \delta_t)} h_{jt}^{\gamma_t-1} (1 + h_{jt})^{-(\gamma_t+\delta_t)} \quad 0 < h_{jt} < \infty. \quad (4.113)$$

Dessa forma,  $h_{jt} \sim BST(\gamma_t, \delta_t)$ , e segundo DasGupta (2011),

$$E(h_{jt}) = \frac{\gamma_t}{\delta_t - 1}, \quad (4.114a)$$

$$V(h_{jt}) = \frac{\gamma_t(\gamma_t + \delta_t - 1)}{(\gamma_t - 1)^2(\delta_t - 2)}. \quad (4.114b)$$

Para obter a priori de  $\eta_{jt}$  considera-se a transformação (Nadarajah, 2005)

$$\eta_{jt} = -\log(h_{jt}), \quad (4.115)$$

de forma que a correspondente distribuição *Beta Exponenciada*, é expressa por

$$\begin{aligned} F_{\eta_{jt}} &= P(\eta_{jt} \leq \eta_{jt}) = P(-\log(h_{jt}) \leq \eta_{jt}) = P(\log(h_{jt}) \geq -\eta_{jt}) = P(h_{jt} \geq e^{-\eta_{jt}}), \\ &= 1 - P(h_{jt} \leq e^{-\eta_{jt}}) \end{aligned} \quad (4.116)$$

$$\begin{aligned} f_{\eta_{jt}} &= \frac{\partial}{\partial \eta_{jt}} F_{\eta_{jt}} = \frac{\partial}{\partial \eta_{jt}} [1 - P(h_{jt} \leq e^{-\eta_{jt}})] = \frac{\partial}{\partial \eta_{jt}} [1 - F_{h_{jt}}(e^{-\eta_{jt}})], \\ &= \frac{\partial}{\partial h_{jt}} [-F_{h_{jt}}(e^{-\eta_{jt}})] \frac{\partial h_{jt}}{\partial \eta_{jt}} = -f_{h_{jt}}(e^{-\eta_{jt}}) \frac{\partial}{\partial \eta_{jt}} [e^{-\eta_{jt}}] = f_{h_{jt}}(e^{-\eta_{jt}}) \cdot e^{-\eta_{jt}}. \end{aligned} \quad (4.117)$$

Logo

$$\boxed{f_{\eta_{jt}} = f_{h_{jt}}(e^{-\eta_{jt}}) \cdot e^{-\eta_{jt}}} \quad (4.118)$$

onde  $f_{h_{jt}}$  é um *Beta do Segundo Tipo* com parâmetros  $\gamma_t$  e  $\delta_t$ , apresentada em (4.113). Assim, avaliando (4.118) tem-se:

$$\begin{aligned} f_{\eta_{jt}} &= [e^{-\eta_{jt}}]^{\gamma_t-1} \cdot [1 + e^{-\eta_{jt}}]^{-(\gamma_t+\delta_t)} \cdot \frac{\Gamma(\gamma_t)\Gamma(\delta_t)}{\Gamma(\gamma_t + \delta_t)} \cdot e^{-\eta_{jt}}, \\ &= \left[ e^{-\eta_{jt}(\gamma_t-1)} \cdot \frac{1}{[1 + e^{-\eta_{jt}}]^{(\gamma_t+\delta_t)}} \right] \cdot \frac{\Gamma(\gamma_t)\Gamma(\delta_t)}{\Gamma(\gamma_t + \delta_t)} \cdot e^{-\eta_{jt}} \\ &= e^{-\eta_{jt}\gamma_t} \cdot \frac{1}{[1 + e^{-\eta_{jt}}]^{(\gamma_t+\delta_t)}} \cdot \frac{\Gamma(\gamma_t)\Gamma(\delta_t)}{\Gamma(\gamma_t + \delta_t)}. \end{aligned} \quad (4.119)$$

Portanto,

$$\boxed{\eta_{jt} \sim BetaExp(\gamma_t, \delta_t)} \quad (4.120)$$

tem distribuição *Beta Exponenciada* (Nadarajah, 2005).

#### 4.6.2 Elicitação da priori $BetaExp(\gamma_t, \delta_t)$ de $\eta_{jt}$

No intuito de avaliar os valores dos hiperparâmetros  $\gamma_t$  e  $\delta_t$  da priori *Beta Exponenciada* apresentada em (4.120) e sua relação com os parâmetros conhecidos  $a_t$  e  $b_t$  do modelo 3PL apresentado em (4.107), são apresentados os detalhes da reparametrização utilizada, além de demonstrar de que maneira se relacionam os parâmetros do modelo 3PL com os parâmetros de (4.120).

Reescrevendo (4.119) na forma da família exponencial, e considerando o  $j$ -ésimo indivíduo cujo índice  $t$  indica o tempo do processo ( $t$ -ésimo item administrado ao indivíduo  $j$ )  $\{\eta_{jt}\}$  em que os parâmetros são avaliados, tem-se

$$\begin{aligned} f(\eta_{jt}) &= \frac{\Gamma(\gamma_t + \delta_t)}{\Gamma(\gamma_t)\Gamma(\delta_t)} \frac{e^{-\gamma_t\eta_{jt}}}{[1 + e^{-\eta_{jt}}]^{\gamma_t+\delta_t}} \\ &\propto \exp[-\gamma_t\eta_{jt} - (\gamma_t + \delta_t) \cdot \ln[1 + \exp(-\eta_{jt})] - b(\gamma_t, \delta_t)], \end{aligned} \quad (4.121)$$

onde

$$b(\gamma_t, \delta_t) = - \left[ \ln(\Gamma(\gamma_t + \delta_t)) - \ln(\Gamma(\gamma_t)) - \ln(\Gamma(\delta_t)) \right].$$

Portanto,

$$f(\eta_{jt}) \propto \exp \left\{ \gamma_t \left[ -\eta_{jt} - \ln[1 + \exp(-\eta_{jt})] \right] + \delta_t \left[ -\ln[1 + \exp(-\eta_{jt})] \right] - b(\gamma_t, \delta_t) \right\} \quad (4.122)$$

e utilizando a transformação expressa em (4.108):

$$\boxed{\eta_{jt} = Da_t(\theta_{jt} - b_t)} \quad (4.123)$$

reescreve-se (4.122) em termos das estatísticas suficientes  $T_1$  e  $T_2$ , a saber:

$$f(\eta_{jt}) \propto \exp \left\{ \gamma_t T_1(\eta_{jt}) + \delta_t T_2(\eta_{jt}) - b(\gamma_t, \delta_t) \right\} \quad (4.124)$$

onde

$$T_1(\eta_{jt}) = - \left( \eta_{jt} + \ln [1 + \exp(-\eta_{jt})] \right) \quad (4.125)$$

$$T_2(\eta_{jt}) = - \ln [1 + \exp(-\eta_{jt})]. \quad (4.126)$$

No entanto, de acordo com as propriedades da família exponencial (DasGupta, 2011),

$$E(T_1) = \frac{\partial b(\gamma_t, \delta_t)}{\partial \gamma_t} = \psi(\gamma_t) - \psi(\gamma_t + \delta_t),$$

$$E(T_2) = \frac{\partial b(\gamma_t, \delta_t)}{\partial \delta_t} = \psi(\delta_t) - \psi(\gamma_t + \delta_t),$$

$$Var(T_1) = \frac{\partial^2 b(\gamma_t, \delta_t)}{\partial \gamma_t^2} = \psi'(\gamma_t) - \psi'(\gamma_t + \delta_t),$$

$$Var(T_2) = \frac{\partial^2 b(\gamma_t, \delta_t)}{\partial \delta_t^2} = \psi'(\delta_t) - \psi'(\gamma_t + \delta_t),$$

$$Cov(T_1, T_2) = \frac{\partial^2 b(\gamma_t, \delta_t)}{\partial \gamma_t \partial \delta_t} = -\psi'(\gamma_t + \delta_t),$$

$$E(\eta_{jt}) = E[T_2(\eta_{jt}) - T_1(\eta_{jt})] = E[T_2(\eta_{jt})] - E[T_1(\eta_{jt})],$$

$$= \psi(\gamma_t) - \psi(\gamma_t + \delta_t) - [\psi(\delta_t) - \psi(\gamma_t + \delta_t)]$$

$$= \psi(\delta_t) - \psi(\gamma_t),$$

$$Var(\eta_{jt}) = Var[T_2(\eta_{jt}) - T_1(\eta_{jt})] = Var[T_2(\eta_{jt})] + Var[T_1(\eta_{jt})] - 2 \times Cov[T_1(\eta_{jt}), T_2(\eta_{jt})]$$

$$= \psi'(\delta_t) - \psi'(\gamma_t + \delta_t) + \psi'(\gamma_t) - \psi'(\gamma_t + \delta_t) + 2 \times \psi'(\gamma_t + \delta_t)$$

$$= \psi'(\delta_t) + \psi'(\gamma_t).$$

Assim, utilizando as propriedades de primeiro e segundo momentos para elicitar os valores de  $\alpha_t$  e  $\beta_t$ , conclui-se que

$$\boxed{f_{jt} = E(\eta_{jt}) = \psi(\delta_t) - \psi(\gamma_t)} \quad (4.127)$$

onde substituindo (4.123) no lado esquerdo da igualdade (4.127), tem-se

$$f_{jt} = E(\eta_{jt}) = E \left( Da_t(\theta_{jt} - b_t) \right) = Da_t E(\theta_{jt}) \stackrel{0}{=} Da_t b_t,$$

com  $\theta_{jt} \sim N(0, 1)$ . Logo,

$$f_{jt} = -Da_t b_t = \psi(\delta_t) - \psi(\gamma_t), \quad (4.128)$$

com  $\psi(z) \approx \ln(z)$ . Portanto,

$$\boxed{f_{jt} = -Da_t b_t \approx \ln \left[ \frac{\delta_t}{\gamma_t} \right]}. \quad (4.129)$$

Avaliando a variância de  $\eta_{jt}$  tem-se

$$q_{jt} = \text{Var}(\eta_{jt}) = \psi'(\delta_t) + \psi'(\gamma_t) \approx \frac{1}{\delta_t} + \frac{1}{\gamma_t},$$

onde  $\psi'(z) \approx z^{-1}$ . Da mesma maneira, tem-se

$$q_{jt} = \text{Var}(\eta_{jt}) = \text{Var}(Da_t(\theta_{jt} - b_t)) = (Da_t)^2 \text{Var}(\theta_{jt}),$$

com  $\theta_{jt} \sim N(0, 1)$ . Então,

$$\boxed{q_{jt} = (Da_t)^2 \approx \frac{1}{\delta_t} + \frac{1}{\gamma_t}} \quad (4.130)$$

De (4.129) e (4.130) têm-se

$$f_{jt} = -Da_t b_t \approx \ln \left( \frac{\delta_t}{\gamma_t} \right) \Rightarrow -f_{jt} = Da_t b_t \approx \ln \left( \frac{\delta_t}{\gamma_t} \right) \quad (4.131a)$$

$$q_{jt} = (Da_t)^2 \approx \frac{1}{\delta_t} + \frac{1}{\gamma_t}. \quad (4.131b)$$

Portanto, conclui-se que

- $f_{jt} \approx \ln \left( \frac{\delta_t}{\gamma_t} \right) \Rightarrow \exp(f_{jt}) \approx \frac{\delta_t}{\gamma_t} \Rightarrow$

$$\boxed{\gamma_t \approx \frac{\delta_t}{\exp(f_{jt})} = \delta_t \exp(-f_{jt})}. \quad (4.132)$$

- $q_{jt} = \frac{1}{\delta_t} + \frac{1}{\gamma_t} = \frac{1}{\delta_t} + \frac{\exp(f_{jt})}{\delta_t} \Rightarrow$

- $q_{jt} = \frac{1}{\delta_t} \left[ 1 + \exp(f_{jt}) \right] \Rightarrow$

$$\boxed{\delta_t = \frac{1 + \exp(f_{jt})}{q_{jt}}}. \quad (4.133)$$

Logo,

$$\begin{aligned}\gamma_t &= \left[ \frac{1 + \exp(f_{jt})}{q_{jt}} \right] \exp(-f_{jt}) = \frac{\exp(-f_{jt}) + \exp(0)}{q_{jt}} = \frac{\exp(-f_{jt}) + 1}{q_{jt}}, \\ \delta_t &= \frac{1 + \exp(f_{jt})}{q_{jt}},\end{aligned}$$

e considerando (4.129) e (4.130)

$$\begin{aligned}f_{jt} &= -Da_t b_t, \\ q_{jt} &= (Da_t)^2,\end{aligned}$$

conclui-se que a relação entre os parâmetros do modelo 3PL e os hiperparâmetros da distribuição à priori apresentada em (4.120) é expressa por:

$$\gamma_t = \frac{1 + \exp(Da_t b_t)}{(Da_t)^2}, \quad (4.134)$$

$$\delta_t = \frac{1 + \exp(-Da_t b_t)}{(Da_t)^2}. \quad (4.135)$$

### 4.6.3 Um Modelo *Power Steady* para o CAT

Nesta seção, é formalizada a estimação das proficiências segundo a metodologia proposta por Smith (1979). Ao final da exposição, é apresentado um algoritmo que descreve o processo a ser implementado em linguagem R.

Seja

- **Série observada**:  $x_{jt} \in \{0, 1\}$  com

$$P(X_{jt}|\eta_{jt}) = \left[ \frac{1}{1 + e^{-\eta_{jt}}} \right]^{x_{jt}} \left[ \frac{e^{-\eta_{jt}}}{1 + e^{-\eta_{jt}}} \right]^{1-x_{jt}}, \quad (4.136)$$

isto é,

$$(X_{jt}|\eta_{jt}) \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-\eta_{jt}}}\right) = \text{Bern}(p_t),$$

de modo que  $P(X_{jt}|\eta_{jt})$  pertence à família exponencial.

- **Priori para  $\eta_{jt}$** : Considere uma priori *Beta Exponenciada* com hiperparâmetros  $\gamma_t$  e  $\delta_t$ , isto é,

$$f(\eta_{jt}) = \frac{1}{B(\gamma_t, \delta_t)} \frac{e^{-\gamma_t \eta_{jt}}}{[1 + e^{-\eta_t}]^{\gamma_t + \delta_t}}, \quad (4.137)$$

onde  $\eta_{jt} \sim \text{BetaExp}(\gamma_t, \delta_t)$ , em que

$$E(\eta_{jt}) = \psi(\delta_t) - \psi(\gamma_t) \quad (4.138a)$$

$$V(\eta_{jt}) = \psi'(\delta_t) + \psi'(\gamma_t), \quad (4.138b)$$

onde  $\gamma_t$  e  $\delta_t$  são descritos por:

$$\gamma_t = \frac{1 + \exp(Da_t b_t)}{(Da_t)^2}, \quad (4.139a)$$

$$\delta_t = \frac{1 + \exp(-Da_t b_t)}{(Da_t)^2}. \quad (4.139b)$$

- **Posteriori para  $\eta_{jt}$** :

$$\begin{aligned} P(\eta_{jt}|X_{jt}) &\propto P(X_{jt}|\eta_{jt})f(\eta_{jt}) \\ &\propto \frac{e^{-\eta_{jt}(1-x_{jt})}}{1 + e^{-\eta_{jt}}} \frac{e^{-\gamma_t \eta_{jt}}}{[1 + e^{-\eta_{jt}}]^{\gamma_t + \delta_t}} \\ &\propto \frac{e^{-\eta_{jt}(\gamma_t - x_{jt} + 1)}}{[1 + e^{-\eta_{jt}}]^{\gamma_t + \delta_t + 1}}. \end{aligned} \quad (4.140)$$

Logo,

$$(\eta_{jt}|X_{jt}) \sim \text{BetaExp}(\gamma_t - x_{jt} + 1, \delta_t + x_{jt}), \quad (4.141)$$

implicando que  $f(\eta_{jt})$  é uma *priori conjugada* (CP).

Portanto,

- (1)  $P(X_{jt}|\eta_{jt})$  pertence à família exponencial;
- (2)  $f(\eta_{jt})$  é priori conjugada.

Assim, de acordo com [Smith \(1979\)](#), é possível escrever a série temporal Bernoulli como um modelo dinâmico “*Power Steady*”, de modo que:

$$(\eta_{jt}|X_j^t) \sim CP(X_{jt}), \quad (4.142a)$$

$$P(\eta_{j,t+1}|X_j^t) \propto [P(\eta_{jt}|X_j^t)]^k \quad ; \quad 0 < k < 1, \quad (4.142b)$$

em que, CP denota a priori conjugada para a família exponencial de  $(X_{jt}|\eta_{jt})$ , e  $\mathbf{X}_j^t$  denota a série  $\mathbf{X}_j^t = \{X_{j1}, X_{j2}, \dots, X_{jt}\}$ , de forma que os estados  $\{\mathbf{H}_{jt}\}$  evoluem de acordo com o modelo (4.142).

Portanto, o modelo dinâmico não-gaussiano em estudo é composto por

### Equação das observações

$$(X_{jt}|\eta_{jt}) \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-\eta_{jt}}}\right), \quad (4.143)$$

### Equação do sistema

$P(\eta_{j,t+1}|\eta_{jt})$  não definida, mas os estados  $\{\mathbf{H}_{jt}\}$  evoluem de acordo com o modelo *Power Steady* descrito pelas expressões (4.142a) e (4.142b).

Sendo assim, de (4.140) e (4.142), têm-se:

$$\begin{aligned} P(\eta_{j,t+1}|X^t) &\propto \left\{ \frac{e^{-\eta_{jt}(\alpha_{t|t}-x_{jt}+1)}}{[1+e^{-\eta_{jt}}]^{\alpha_{t|t}+\beta_{t|t}+1}} \right\}^k \\ &\propto \frac{e^{-\eta_{jt}[k(\alpha_{t|t}-x_{jt}+1)]}}{[1+e^{-\eta_{jt}}]^{k(\alpha_{t|t}+\beta_{t|t}+1)}} \end{aligned} \quad (4.144)$$

$$\propto \frac{e^{-\eta_{j,t+1}\alpha_{t+1|t}}}{[1+e^{-\eta_{j,t+1}}]^{\alpha_{t+1|t}+\beta_{t+1|t}}}. \quad (4.145)$$

Para tornar a notação compatível ao modelo Power Steady, substitui-se o parâmetro  $\gamma_t$  por  $\alpha_{t|t}$  e o parâmetro  $\delta_t$  por  $\beta_{t|t}$ .

Portanto,

$$(\eta_{j,t+1}|X_j^t) \sim BetaExp\left(k(\alpha_{t|t}-x_{jt}+1), k(\beta_{t|t}+x_{jt})\right),$$

isto é,

$$(\eta_{j,t+1}|X_j^t) \sim BetaExp(\alpha_{t+1|t}, \beta_{t+1|t}),$$

de modo que  $\alpha_{t+1|t} = k(\alpha_{t|t} - x_{jt} + 1)$  e  $\beta_{t+1|t} = k(\beta_{t|t} + x_{jt})$ .

E assim o processo recursivo fica descrito da seguinte maneira:

- **Estágio de predição**: Reduz-se em obter os parâmetros da distribuição  $P(\eta_{j,t+1}|X_j^t)$ , que consiste em calcular os valores dos parâmetros

$$\begin{cases} \alpha_{t+1|t} = k(\alpha_{t|t} - x_{jt} + 1), \\ \beta_{t+1|t} = k(\beta_{t|t} + x_{jt}). \end{cases} \quad (4.146)$$

- **Estágio de atualização**: Consiste em obter a distribuição de  $P(\eta_{j,t+1}|X_j^{t+1})$ :

$$\begin{aligned} P(\eta_{j,t+1}|X_j^{t+1}) &= P(\eta_{j,t+1}|X_j^t, X_{j,t+1}) \\ &= \frac{P(\eta_{j,t+1}, X_j^t, X_{j,t+1})}{P(X_j^t, X_{j,t+1})} \\ &\propto P(X_{j,t+1}|\eta_{j,t+1}, X_j^t)P(\eta_{j,t+1}|X_j^t) \\ &\propto P(X_{j,t+1}|\eta_{j,t+1})P(\eta_{j,t+1}|X_j^t). \end{aligned} \quad (4.147)$$

De (4.136) e (4.137) têm-se que

$$\begin{aligned} P(\eta_{j,t+1}|X_j^{t+1}) &\propto \frac{e^{-\eta_{j,t+1}(1-x_{j,t+1})}}{1+e^{-\eta_{j,t+1}}} \frac{e^{-\eta_{j,t+1}\alpha_{t+1|t}}}{[1+e^{-\eta_{j,t+1}}]^{\alpha_{t+1|t}+\beta_{t+1|t}}} \\ &\propto \frac{e^{-\eta_{j,t+1}[\alpha_{t+1|t}+1-x_{j,t+1}]}}{[1+e^{-\eta_{j,t+1}}]^{\alpha_{t+1|t}+\beta_{t+1|t}+1}} \\ &\propto \frac{e^{-\eta_{j,t+1}\alpha_{t+1|t+1}}}{[1+e^{-\eta_{j,t+1}}]^{\alpha_{t+1|t+1}+\beta_{t+1|t+1}}}, \end{aligned} \quad (4.148)$$

isto é,  $(\eta_{j,t+1}|X_j^{t+1}) \sim \text{BetaExp}(\alpha_{t+1|t+1}; \beta_{t+1|t+1})$ .

Portanto, o estágio de atualização consiste em calcular os valores dos parâmetros

$$\begin{cases} \alpha_{t+1|t+1} = \alpha_{t+1|t} + 1 - x_{j,t+1}, \\ \beta_{t+1|t+1} = \beta_{t+1|t} + x_{j,t+1}. \end{cases} \quad (4.149)$$

Ou seja,

$$\begin{cases} \alpha_{t+1|t+1} = k(\alpha_{t|t} - x_{jt} + 1) + (1 - x_{j,t+1}), \\ \beta_{t+1|t+1} = k(\beta_{t|t} + x_{jt}) + x_{j,t+1}. \end{cases} \quad (4.150)$$

#### 4.6.4 Estimação de $\eta$ pelo Método EAP

Para estimar  $\eta_{jt}$  através da *Esperança a Posteriori* (EAP), utilizam-se as propriedades do modelo descrito em (4.138), e conhecendo os valores dos parâmetros, expressos em (4.139), avaliam-se as propriedades do modelo, através das expressões (4.138a) e (4.138b):

$$\begin{aligned} \hat{\eta}_{j,t+1} &= E(\eta_{j,t+1}|X_j^{t+1}) \\ &= \psi(\beta_{t+1|t+1}) - \psi(\alpha_{t+1|t+1}), \end{aligned} \quad (4.151a)$$

$$\begin{aligned} \text{Var}(\hat{\eta}_{j,t+1}) &= \text{Var}(\eta_{j,t+1}|X_j^{t+1}) \\ &= \psi'(\beta_{t+1|t+1}) + \psi'(\alpha_{t+1|t+1}), \end{aligned} \quad (4.151b)$$

de modo que, utilizando a transformação (4.123) e a expressão (4.151a), é avaliada, a cada estágio de atualização, um passo à frente, o valor da proficiência dada por:

$$\hat{\theta}_{j,t+1} = \frac{\hat{\eta}_{j,t+1} + Da_t b_t}{Da_t}, \quad (4.152)$$

$$s.d.(\hat{\theta}_{j,t+1}) = \frac{1}{Da_t} \sqrt{\psi'(\beta_{t+1|t+1}) + \psi'(\alpha_{t+1|t+1})}. \quad (4.153)$$

#### 4.6.5 Estimação de $\eta$ pelo Método MAP

Para estimar  $\eta$  através da *Moda a Posteriori* (MAP), calcula-se o estimador de máxima verossimilhança a ser avaliado através da distribuição a posteriori. obtida em (4.141), de forma que

- $P(\eta_{j,t}|X_j^t) \propto \frac{e^{-\eta_{j,t}\alpha_{t|t}}}{[1+e^{-\eta_{j,t}}]^{\alpha_{t|t}+\beta_{t|t}}}$  .
- $\ell = \log P(\eta_{j,t}|X_j^t) = -\eta_{j,t}\alpha_{t|t} - (\alpha_{t|t} + \beta_{t|t}) \log[1 + e^{-\eta_{j,t}}]$  .
- $\dot{\ell} = \frac{\partial \ell}{\partial \eta_{j,t}} = -\alpha_{t|t} + \frac{(\alpha_{t|t} + \beta_{t|t})e^{-\eta_{j,t}}}{1+e^{-\eta_{j,t}}}$  .
- $\dot{\ell} = 0 \Rightarrow -\alpha_{t|t}(1 + e^{-\eta_{j,t}}) + (\alpha_{t|t} + \beta_{t|t})e^{-\eta_{j,t}} = 0 \Rightarrow e^{-\eta_{j,t}} = \frac{\alpha_{t|t}}{\beta_{t|t}}$  .
- $\tilde{\eta}_{j,t} = \log\left(\frac{\beta_{t|t}}{\alpha_{t|t}}\right)$  .
- $\ddot{\ell} = -\frac{(\alpha_{t|t} + \beta_{t|t})e^{-\eta_{j,t}}}{(1+e^{-\eta_{j,t}})^2}$  .



A variância de  $\tilde{\eta}_{j,t}$  pode ser encontrada por  $Var(\tilde{\eta}_{j,t}) = (\mathfrak{F})^{-1}$ , em que  $\mathfrak{F} = -E(\ddot{\ell})$ , é a Informação de Fisher. Devido à complexidade do cálculo do valor para  $E(\ddot{\ell})$ , utilizou-se a aproximação  $Var(\tilde{\eta}_{j,t}) \approx (\ddot{\ell})^{-1} \Big|_{\eta=\hat{\eta}}$ .

Em resumo,

$$\tilde{\eta}_{j,t} = \log \left( \frac{\beta_{t|t}}{\alpha_{t|t}} \right), \quad (4.154a)$$

$$Var(\tilde{\eta}_{j,t}) = \frac{(1 + e^{-\tilde{\eta}_{j,t}})^2}{(\alpha_{t|t} + \beta_{t|t})e^{-\tilde{\eta}_{j,t}}}, \quad (4.154b)$$

de modo que a cada estágio de atualização, um passo à frente, o valor da proficiência estimada pela *Moda a Posteriori* (MAP) é dada por:

$$\tilde{\theta}_{j,t+1} = \frac{\tilde{\eta}_{j,t+1} + Da_t b_t}{Da_t}, \quad (4.155)$$

$$s.d.(\tilde{\theta}_{j,t+1}) = \frac{1}{Da_t} \sqrt{\frac{(1 + e^{-\tilde{\eta}_{j,t+1}})^2}{(\alpha_{t+1|t+1} + \beta_{t+1|t+1})e^{-\tilde{\eta}_{j,t+1}}}}. \quad (4.156)$$

A seguir, o algoritmo que descreve os estágios de predição e atualização apresentados na seção 4.6, para estimação das proficiências dos indivíduos submetidos ao CAT. O procedimento baseia-se no modelo dinâmico desenvolvido segundo a metodologia proposta por Smith (1979). Vale lembrar que o algoritmo 2 reflete apenas o processo de estimação das proficiências, não contemplando o fluxo completo do CAT. Assim, o índice  $t$  reflete apenas o conjunto de itens *administrados*, e não todos os itens *disponíveis* no CAT.

---

**Algoritmo 2:** Estimação de proficiências via Modelo Dinâmico *Power Steady*.

---

**Input:** Defina valores iniciais para  $\epsilon$  e  $\nu$

- 1 Seja  $\{X_{j1}, \dots, X_{jI}\}$  um conjunto de respostas, onde  $j = 1 \dots N$
  - 2 Obtenha valores iniciais de  $\alpha_{t|t}$  e  $\beta_{t|t}$
  - 3 **while**  $\nu > \epsilon$  **do**
  - 4     Considere o  $j$ -ésimo indivíduo a responder um CAT
  - 5     **for**  $t \leftarrow 1$  **to**  $n_I$  **do**
  - 6         Amostre um item com parâmetros  $(a_t, b_t)$  e considere sua resposta  $x_{jt}$
  - 7         Calcule  $\gamma_t$  e  $\delta_t$  por (4.139)
  - 8         **if**  $t = 1$  **then**
  - 9             | Avalie  $\alpha_{t+1|t}$  e  $\beta_{t+1|t}$  por (4.146) // Estágio de predição
  - 10         **end**
  - 11         **else**
  - 12             | Avalie  $\alpha_{t+1|t+1}$  e  $\beta_{t+1|t+1}$  por (4.149) e (4.150) // Estágio de atualização
  - 13         **end**
  - 14         Calcule  $\hat{\eta}_{j,t+1} = E(\eta_{j,t+1}|X_j^{t+1})$  por (4.151a)
  - 15         Calcule  $\hat{\nu}_{j,t+1} = Var(\eta_{j,t+1}|X_j^{t+1})$  por (4.151b)
  - 16         Obtenha  $\hat{\theta}_{j,t+1}$  de (4.152)
  - 17     **end**
  - 18      $\nu \leftarrow \hat{\nu}_{j,t+1}$
  - 19 **end**
  - 20 **return**  $\hat{\theta}_{j,t+1}$
-



# Capítulo 5

## Simulações e Resultados

No capítulo 4 foram apresentados três estimadores para a avaliação da proficiência em estudos do CAT, através de modelos dinâmicos, a saber:

- $\hat{\theta}$ : estimador para a *Esperança a Posteriori* (EAP), avaliado segundo o modelo *Filtro de Kalman*, discutido na Seção 4.5.6 e apresentado pela expressão (4.106).
- $\tilde{\theta}$ : estimador para a *Moda a Posteriori* (MAP), avaliado segundo o modelo *Power Steady*, discutido na Seção 4.6.5 e apresentado pela expressão (4.155).
- $\hat{\theta}$ : estimador para a *Esperança a Posteriori* (EAP), avaliado segundo o modelo *Power Steady*, discutido na Seção 4.6.5 e apresentado pela expressão (4.152).

Para o presente estudo, os dados apresentados neste capítulo consideram como parâmetros, informações obtidas de acordo como o banco de itens fornecido pelo CEBRASPE, referente à Prova de Proficiência em Inglês Instrumental I, da Universidade de Brasília (UnB), que desde o primeiro semestre de 2010 é aplicada por meio do CAT.

Assim, ao longo do capítulo, são apresentadas simulações de CATs, considerando-se provas de tamanho fixo ( $n = 45$ ), para diferentes cenários, onde os estimadores acima descritos são avaliados. Em todas as simulações realizadas, para cada um dos estimadores acima relacionados, foram considerados conjuntos de  $\theta$ 's verdadeiros pré-fixados, de maneira que o estudo visa observar o comportamento dos estimadores em níveis de proficiência específicos, testados no intervalo  $[-3, 3]$ , em geral com diferenças de 0.25 entre cada valor. Assim, quanto menor o valor do  $\theta$  considerado como verdadeiro, menor se espera ser a proficiência estimado de um determinado indivíduo, em Inglês Instrumental I.

### 5.1 Simulações com o Modelo *Power Steady*

A seguir, apresenta-se o processo utilizado para construção das simulações realizadas com o modelo *Power Steady* (vide Seção 4.6.3). Importante lembrar que nas simulações tratou-se inicialmente, do modelo *Power Steady*, em função deste ser um caso mais simples do que o modelo *Filtro de Kalman*, uma vez que no modelo desconsidera-se o parâmetro de acerto casual, e utiliza-se o modelo 2PL para cálculo das probabilidades de acerto. Vale ressaltar que no estudo em questão, as simulações realizadas consideram os parâmetros já calibrados segundo o modelo 3PL, com ajuste para uma ogiva normal. Portanto, consideram a constante  $D = 1.702$ , no cálculo da probabilidade de acerto, pela expressão (4.17), utilizada para estimação da proficiência  $\theta$ .

O processo de construção dos estimadores  $\tilde{\theta}$  e  $\hat{\theta}$ , que tratam, respectivamente, do método *Moda a Posteriori* (MAP) e *Esperança a Posteriori* (EAP), apresentou algumas restrições, a saber:

- a) Tanto para o método EAP, quanto para o método MAP, o processo de estimação de  $\theta$  coletou um número mínimo de 2 a 5 respostas iniciais, antes de gerar a primeira estimação.
- b) Para as primeiras 2 a 5 respostas iniciais foram consideradas a estimação do  $\theta$  pelo método EAP do modelo 3PL de TRI tradicional, para se obter um ponto de partida para os algoritmos propostos.
- c) Com o método MAP simulou-se dados a partir do terceiro cenário, em função dos testes realizados quanto ao comportamento dos parâmetros do modelo dinâmico utilizado, em ambos os modelos considerados.

Por se tratar de simulações, e considerando o fato de que todos os estimadores estudados são provenientes de modelos dinâmicos, a cada item aplicado, uma atualização dos parâmetros  $\alpha$  e  $\beta$  é realizada, sendo necessárias estimativas iniciais destes parâmetros. Para tanto, as primeiras 2 a 5 respostas iniciais foram desconsideradas no cômputo do desvio-padrão do estimador, até que se atinja um ponto de estabilização e melhoria do processo de estimação. Assim, a construção gradativa do processo de estimação, levou em consideração cinco diferentes cenários para os valores iniciais das 2 a 5 respostas utilizadas na atribuição dos parâmetros iniciais dos modelos dinâmicos, a saber:

**Cenário 1:**  $\alpha_0$  e  $\beta_0$  definidos de acordo com a relação de inter-dependência entre os parâmetros  $\alpha$  e  $\beta$  definidos pelo modelo dinâmico, avaliados para cada uma das respostas iniciais, utilizadas para estimar a proficiência pelo método EAP do modelo 3PL de TRI tradicional.

**Cenário 2:**  $\alpha_0$  e  $\beta_0$  definidos respectivamente de acordo com a média dos parâmetros  $a$  e  $b$  dos itens.

**Cenário 3:** após a execução e observação dos resultados dos cenários anteriores, utilizou-se diferentes fatores de desconto ( $k$ ). De acordo com as observações realizadas, foram geradas simulações para casos onde o fator de desconto  $\alpha_0 = \beta_0 = 7$ .

**Cenário 4:** De acordo com as observações realizadas anteriormente,  $\alpha_0 = \beta_0 = 50$ .

**Cenário 5:**  $\alpha_0$  e  $\beta_0$  definidos de acordo com resultados das estimativas de  $\hat{\theta}$  e  $Var(\hat{\theta})$ , avaliadas inicialmente pelo método EAP/3PL de um CAT convencional, com posterior atualização da relação de inter-dependência entre os parâmetros  $\alpha$  e  $\beta$  definidos pelo modelo dinâmico, utilizado para estimar a proficiência segundo o modelo *Power Steady* descrito no capítulo anterior.

O processo de simulação dos estimadores  $\hat{\theta}$  (EAP) e  $\tilde{\theta}$  (MAP) só passaram a ser simultâneos após os estudos realizados com os hiperparâmetros  $\alpha$  e  $\beta$ , utilizados nos modelos em questão, nas “*equações de transição*”. Os Cenários 1 e 2 podem ser considerados como estágios de ajuste do modelo, uma vez que, durante as simulações realizadas, muitas alterações e correções foram realizadas até que se obtivesse resultados razoavelmente satisfatórios sobre os valores iniciais a serem considerados.

### 5.1.1 Avaliando o CAT tradicional

No intuito de avaliar os estimadores propostos pelos modelos de *Filtro de Kalman* e *Power Steady*, simulações de um CAT tradicional, considerando um conjunto de  $\theta$ 's fixos  $\{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$ , foram considerados. As execuções evidenciaram resultados conhecidos da literatura (Bock e Mislevy, 1982; Costa, 2009).

A Figura (5.1) apresenta os boxplots comparativos das distribuições dos desvios-padrão obtidos na estimação, via CAT, dos  $\theta$ 's relativos a “respondentes” (simulados) com proficiências fixadas no conjunto das nove possibilidades elencadas acima. Como se observa, o perfil médio (em vermelho) aproxima-se do valor 0.2 ao longo das questões, com lenta estabilização em torno da 15ª questão. Este é um comportamento esperado dos desvios-padrão no CAT convencional (Bock e Mislevy, 1982; Costa, 2009). A Figura (5.2) traz as trajetórias das estimativas dos  $\theta$ 's de três indivíduos ao longo das questões. Também, apresentam-se as trajetórias individuais dos desvios-padrão (ao lado). Na Figura (5.3) apresenta-se a trajetória dos  $\theta$ 's estimados de cada um dos indivíduos com uma das nove proficiências fixadas. Observa-se convergência e intervalos de confiança cada vez mais precisos ao longo das questões.

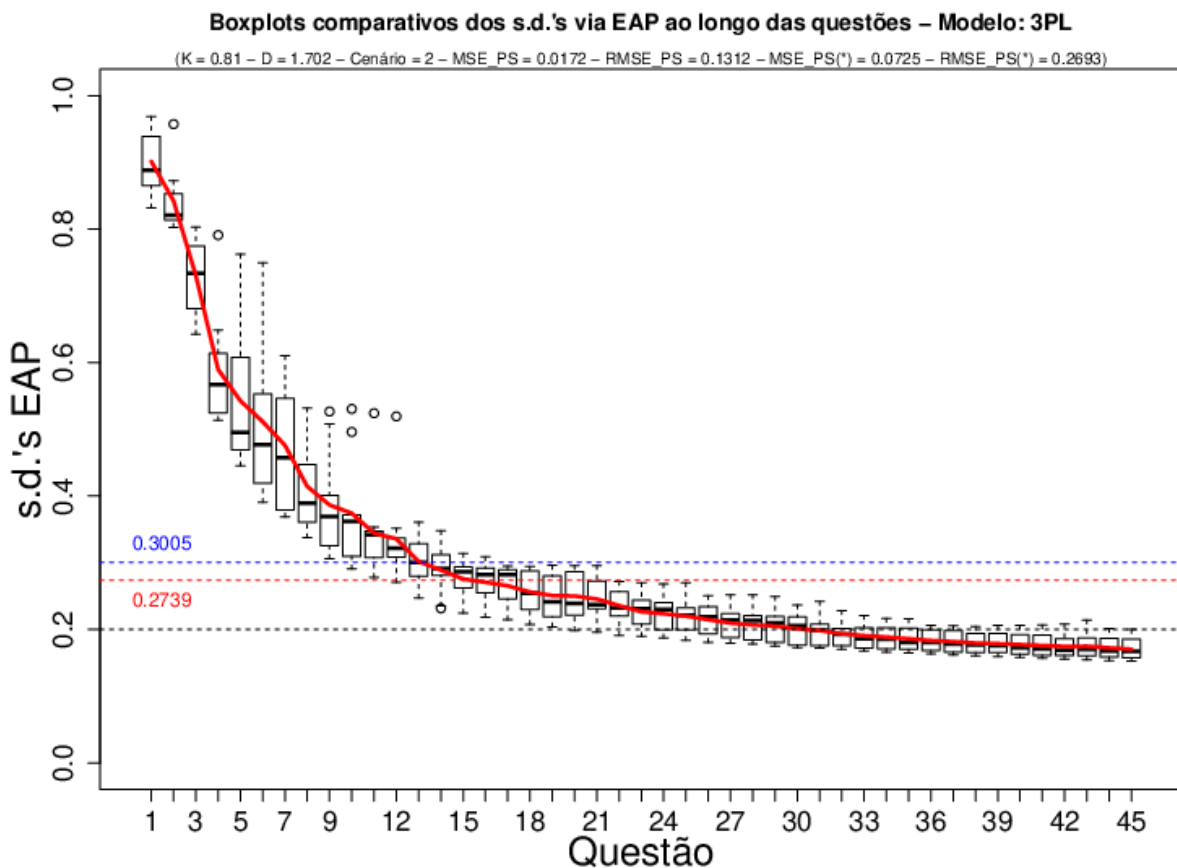


Figura 5.1: CAT EAP/3PL: Trajetória dos  $sd(\theta)$

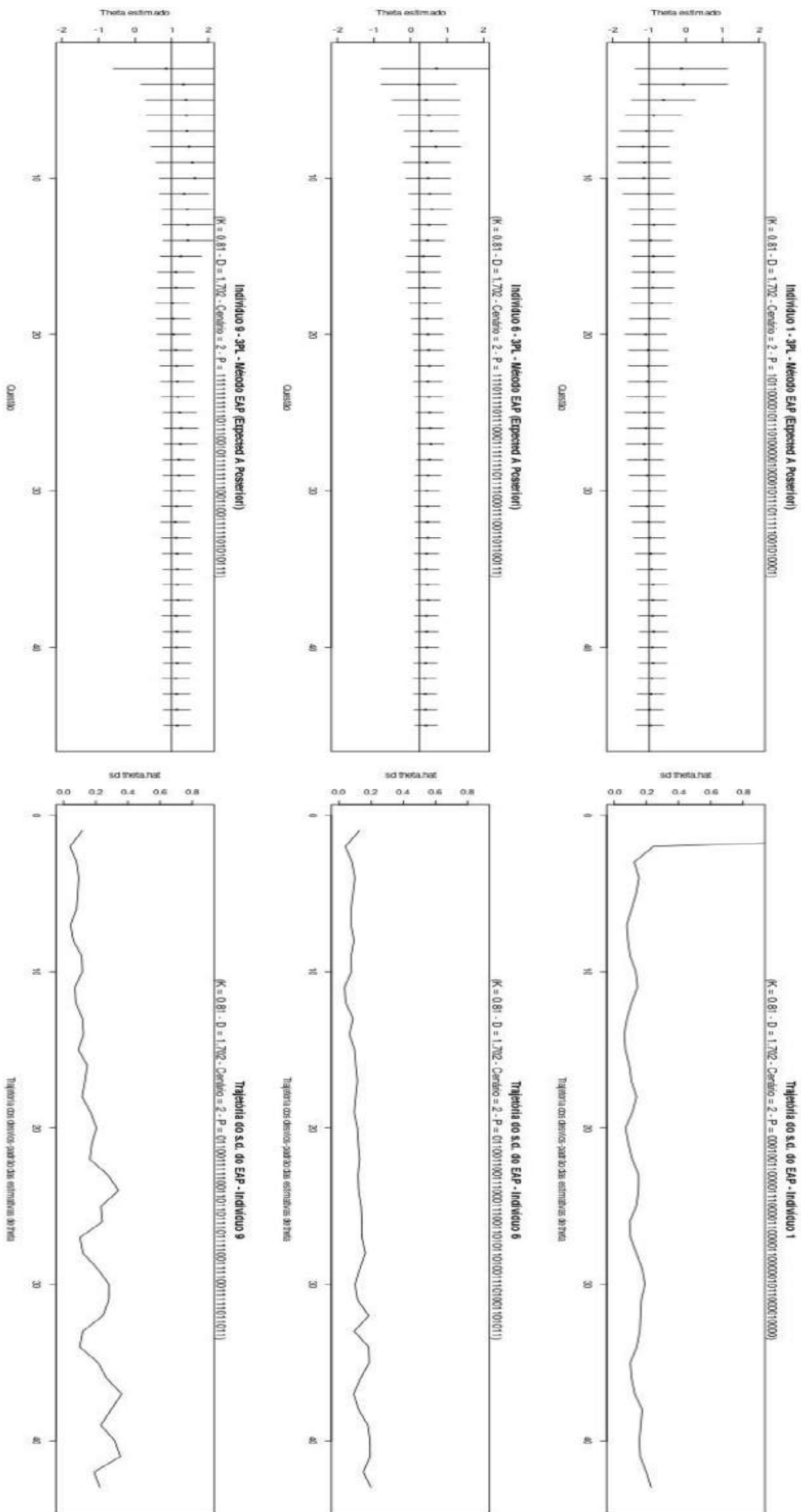


Figura 5.2: CAT EAP/3PL: Estimativas de Proficiência

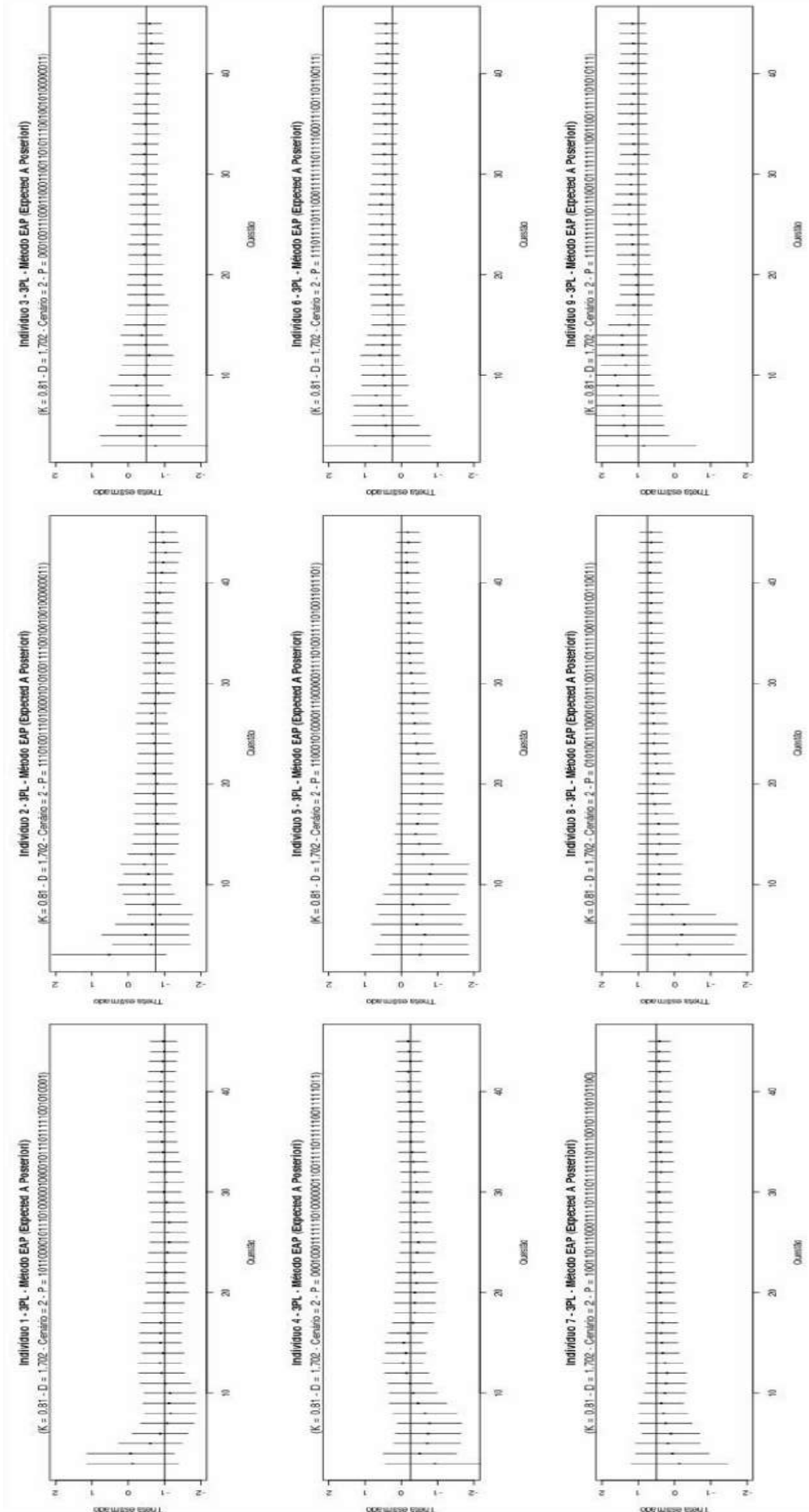


Figura 5.3: CAT EAP/3PL: Trajetória das Proficiências Estimadas

### 5.1.2 Avaliando o comportamento de $\alpha_t$ e $\beta_t$ do modelo *Power Steady*

Como mencionado na Seção 4.5.6, os parâmetros  $\alpha_t$  e  $\beta_t$ , são parte integrante das *equações de transição* apresentadas em (4.94), e expressas por

$$\begin{cases} \alpha_{t+1} = \phi\alpha_t + 1 - x_{t+1} \\ \beta_{t+1} = \phi\beta_t + x_{t+1} \end{cases},$$

sendo, portanto, parte integrante do estágio de atualização da probabilidade de acerto, modelada utilizando-se o modelo 2PL.

Na simulação de um CAT, via modelo *Power Steady*, os parâmetros  $\alpha_t$  e  $\beta_t$  fazem parte do núcleo do processo recursivo, impactando, diretamente, nos resultados obtidos das proficiências estimadas a cada nova questão ofertada a um respondente em um procedimento CAT. Assim, o que se espera obter desta relação dos  $\alpha$ 's e  $\beta$ 's é uma suavização das trajetórias das estimativas da proficiência ao longo das questões, tornando assim, os estimadores  $\hat{\theta}$  (EAP) e  $\tilde{\theta}$  (MAP) boas alternativas ao CAT tradicional com EAP/3PL.

### 5.1.3 Cenários 1 e 2

São tratados nesta seção, os Cenários 1 e 2, descritos no início da Seção 5.1. Em função da dinâmica de dependência do modelo *Power Steady* em relação às *equações de transição*, em que a cada questão atribuída a um respondente, uma resposta é avaliada, possibilitando assim, a estimação atualizada da proficiência do respondente, dado os parâmetros da questão respondida e a proficiência avaliada no estágio anterior. Duas escolhas distintas foram utilizadas como valores iniciais para  $\alpha_t$  e  $\beta_t$ , a fim de dar início ao processo de atualização das estimativas. Em ambas, considerou-se a proficiência inicial próxima ao valor da suposta proficiência média dos indivíduos, isto é,  $\theta = 0$ , para selecionar a primeira e a segunda questões, de forma que o banco de itens foi filtrado da seguinte forma:

$$\text{se } t = 1 \begin{cases} 0 < a \leq 1 \\ -1.5 \leq b \leq 0.5 \end{cases}, \quad (5.1)$$

$$\text{se } t = 2 \begin{cases} 0 < a \leq 1 \\ 0 < b \leq 1.5 \end{cases}. \quad (5.2)$$

Assim, nas primeiras simulações concentraram-se esforços em avaliar o comportamento destes parâmetros  $(\alpha_t, \beta_t)$ , estabelecendo uma relação entre a variabilidade observada a cada nova questão escolhida pela simulação do CAT, com a respectiva proficiência estimada, considerando  $\theta \in [-1, 1]$ .

Para o Cenário 1,  $\alpha_0$  e  $\beta_0$  foram calculados, respectivamente, segundo a relação entre os parâmetros do modelo 3PL e os hiperparâmetros da distribuição a priori *Beta Exponenciada*, expressos em (4.134) e (4.135), por

$$\begin{aligned} \alpha_t &= \frac{1 + \exp(Da_t b_t)}{(Da_t)^2}, \\ \beta_t &= \frac{1 + \exp(-Da_t b_t)}{(Da_t)^2}. \end{aligned} \quad (5.3)$$



Para o Cenário 2,  $\alpha_0$  e  $\beta_0$  tomaram, respectivamente, a média dos parâmetros  $a$  e  $b$  do banco de itens filtrado onde os registros tivessem  $-0.5 \leq b \leq 0.5$ . Para os dados de Inglês Instrumental do CEBRASPE, obteve-se as estimativas apresentadas no Quadro (5.2), de sorte que fixou-se  $\alpha_0 = 0.7361$  e  $\beta_0 = -0.0419$ , de acordo com a expressão (5.3).

**Quadro 5.1:** *Estatísticas descritivas dos parâmetros do banco de itens, sem filtros.*

1	<b>a</b>	<b>b</b>	<b>c</b>
2 N . . . . .	383	383	383
3 Mínimo . . . . .	0.4304	-3.7601	0.0357
4 1o. Quartil . . . . .	0.6756	-1.1085	0.1134
5 Mediana . . . . .	0.9369	-0.4753	0.1392
6 Média . . . . .	0.9468	-0.4381	0.1383
7 3o. Quartil . . . . .	1.1120	0.2280	0.1653
8 Máximo . . . . .	2.8560	2.3313	0.3599
9 Variância . . . . .	0.1239	1.0297	0.0024
10 Desvio Padrão . . . . .	0.3520	1.0147	0.0493
11 Erro Padrão . . . . .	0.0180	0.0519	0.0025

**Quadro 5.2:** *Estatísticas descritivas dos parâmetros do banco de itens filtrado.*

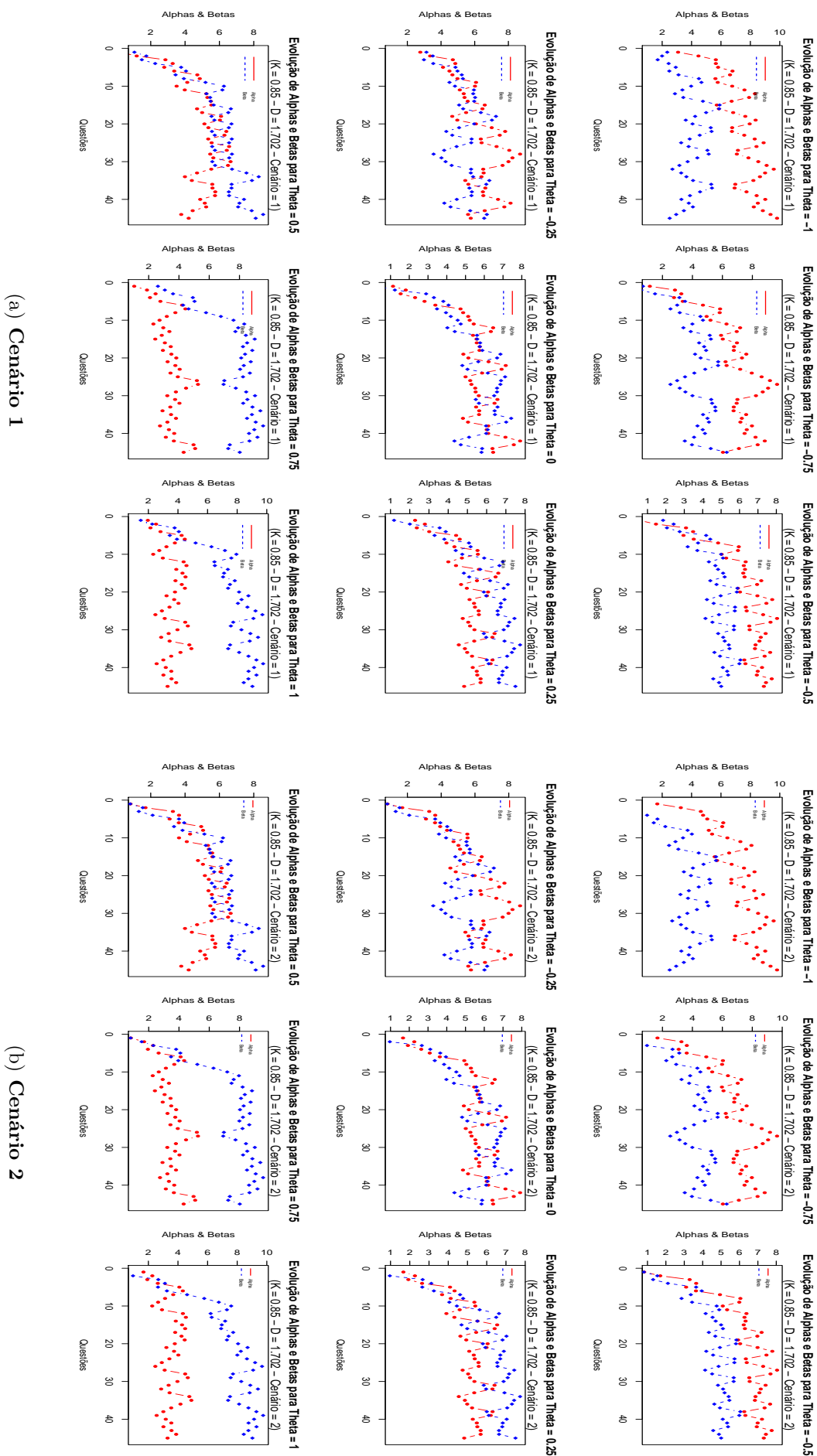
1	<b>a</b>	<b>b</b>	<b>c</b>
2 N . . . . .	78	78	78
3 Mínimo . . . . .	0.4613	-0.4927	0.0452
4 1o. Quartil . . . . .	0.5837	-0.2971	0.1172
5 Mediana . . . . .	0.7377	-0.0508	0.1440
6 Média . . . . .	0.7361	-0.0419	0.1434
7 3o. Quartil . . . . .	0.9094	0.2182	0.1721
8 Máximo . . . . .	0.9963	0.4911	0.2622
9 Variância . . . . .	0.0308	0.0857	0.0026
10 Desvio Padrão . . . . .	0.1754	0.2927	0.0511
11 Erro Padrão . . . . .	0.0199	0.0331	0.0058

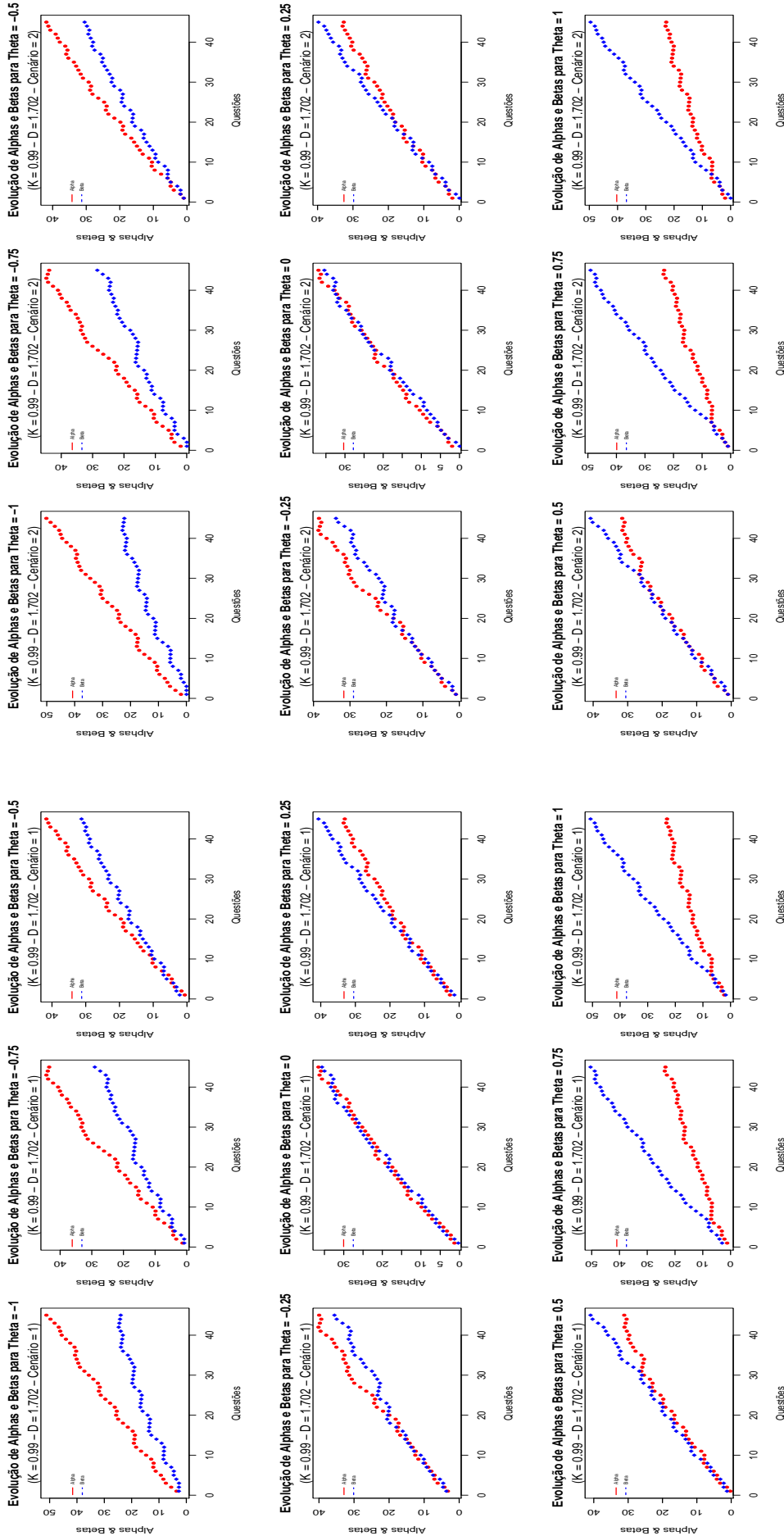
Como descrito na Seção 4.6.3, o modelo *Power Steady* tem forte relação com os fatores de desconto, representados em (4.94) pelo parâmetro  $\phi$  nas equações de transição e pelo parâmetro  $k$  nos resultados das simulações apresentadas a seguir. Tal como postulado por Smith (1979), quanto maior o valor de  $k$ , maior o grau de suavização do processo.

As trajetórias de  $\theta$ , sob os Cenários 1 e 2, são simuladas. Os resultados para  $k \in (0.8, 0.99]$ , quanto ao comportamento dos parâmetros dinâmicos  $\alpha_t$  e  $\beta_t$ , para algumas escolhas nesse intervalo, são mostrados nas Figuras (5.4) e (5.5).

Ao analisar os comportamentos apresentados, algumas observações podem ser feitas:

- Em função de sua natureza dinâmica, os valores de  $\alpha$  e  $\beta$  não foram afetados pela média dos parâmetros  $a$  e  $b$  (vide expressão 5.3) dos itens, que foram utilizados no Cenário 2 para iniciar o processo de atualização. Assim, pouca ou quase nenhuma diferença se observa entre os Cenários 1 e 2.
- As curvas descritas nas Figuras (5.4) e (5.5) parecem crescer indefinidamente com o número de questões.
- Observa-se que, quanto menor o valor do fator de desconto,  $k$ , menor é a amplitude de valores observados de  $\alpha$  e  $\beta$ .
- As curvas descritas por  $\alpha_t$  e  $\beta_t$  se cruzam sempre quando  $\theta \approx 0$ .
- Os perfis de  $\alpha$  (cor vermelha) e os de  $\beta$  (cor azul) apresentam menor tendência de crescimento ao longo das questões, para menores valores de  $k$ , de maneira que quanto mais  $k$  se aproxima de 1, maior é a tendência de crescimento dos valores de  $\alpha$  e  $\beta$ .





(b) Cenário 2

(a) Cenário 1

Figura 5.5: Power Steady - Comportamento dos parâmetros  $\alpha$  e  $\beta$ , quando  $k = 0.99$  (Cenários 1 e 2).

- Independente do valor de  $k$ , há uma notória inversão da ascensão das curvas de  $\alpha$  e  $\beta$ , quando  $|\theta| \approx 1$ , criando assim, um afastamento das mesas, de modo que, na maioria dos casos,

$$\begin{cases} \alpha > \beta & \text{quando } \theta \approx -1 \\ \alpha < \beta & \text{quando } \theta \approx 1 \end{cases}.$$

No intuito de avaliar a precisão das estimativas de  $\theta$ , utilizou-se o critério RMSE (*Root of the Mean Squared Error*) para comparar o desempenho dos estimadores sob o método *Power Steady* (PS) nos Cenários 1 e 2. O RMSE é definido por

$$RMSE = \sqrt{\frac{1}{N_{itens}} \sum_{j=1}^{N_{itens}} (\hat{\theta}_{p,j} - \theta_{p,j})^2}, \quad (5.4)$$

em que  $p$  denota o índice de  $\theta \in \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$ , de modo que

$$\begin{cases} p = 1 & (\theta = -1) \\ p = 2 & (\theta = -0.75) \\ \dots & \dots \\ p = 8 & (\theta = 0.75) \\ p = 9 & (\theta = 1) \end{cases}.$$

**Tabela 5.1:** Power Steady - Comparações por RMSE (Cenários 1 e 2)

Execução	$k$	Modelo/RMSE		
		EAP/3PL	PS (Cenário 1)	PS (Cenário 2)
1	0.81	0.1388	0.7966	0.7966
2	0.82	0.1277	0.7845	0.7845
3	0.83	0.2338	0.7725	0.7726
4	0.84	0.1906	0.7608	0.7608
5	0.85	0.2224	0.7492	0.7492
6	0.86	0.1820	0.7377	0.7378
7	0.87	0.1536	0.7263	0.7265
8	0.88	0.2200	0.7151	0.7152
9	0.89	0.2220	0.7039	0.7041
10	0.90	0.1505	0.6928	0.6931
11	0.91	0.1508	0.6819	0.6823
12	0.92	0.2012	0.6712	0.6717
13	0.93	0.1539	0.6607	0.6615
14	0.94	0.1749	0.6507	0.6516
15	0.95	0.1463	0.6411	0.6423
16	0.96	0.2057	0.6322	0.6337
17	0.97	0.2047	0.6240	0.6259
18	0.98	0.1469	0.6168	0.6191
19	0.99	0.1358	0.6106	0.6134

A Tabela (5.1) apresenta os RMSE's obtidos utilizando o CAT tradicional (EAP/3PL), e o modelo *Power Steady* (PS), executado sob os Cenários 1 e 2 e níveis crescentes de fator de desconto. As informações apresentadas em destaque evidenciam as execuções onde a efetividade dos Cenários 1 e 2 foram praticamente a mesma. Em todos os demais, apesar de se observar uma melhora quanto maior o valor de  $k$ , os resultados mostraram que o modelo *Power Steady* (PS), para ambos os cenários, possui desempenho inferior ao método tradicional EAP/3PL, tal como demonstram os altos valores de RMSE, apresentados pelo modelo CAT *Power Steady* em comparação ao CAT tradicional.

Em resumo, o que se conclui sobre  $\alpha$  e  $\beta$  sob os Cenários 1 e 2 é que:

- Para  $|\theta| \approx 1$  os parâmetros  $\alpha$  e  $\beta$  tendem a se distanciar ao longo das execuções, de forma que, para  $\theta \approx -1$ ,  $\alpha > \beta$  e para  $\theta \approx 1$ ,  $\beta > \alpha$ .
- Para  $\theta \approx 0$ , as curvas dos parâmetros  $\alpha$  e  $\beta$  se cruzam.
- Para fatores de desconto pequenos ( $\approx 0.8$ ),  $\alpha$  e  $\beta$  são pequenos, com valores máximos em torno de 7, apresentando curvas não crescentes, porém mais ciclantes.
- Para fatores de desconto grandes ( $\approx 0.99$ ),  $\alpha$  e  $\beta$  são grandes, com valores máximos de 50, apresentando curvas paralelas e suaves. No entanto, como demonstram as Figuras (5.6) a (5.8), as trajetórias dos valores estimados de  $\theta$  continuam ciclantes.
- O RMSE do modelo *Power Steady* foi sistematicamente superior ao do modelo EAP tradicional.

Complementando as observações realizadas sobre o comportamento dos parâmetros  $\alpha$  e  $\beta$ , a trajetória do estimador da proficiência  $\hat{\theta}$  (EAP), para o mesmo conjunto de indivíduos, considerando  $\theta \in \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$  e valores de fator de desconto  $k \in \{0.81, 0.90, 0.99\}$ , são apresentados nas Figuras (5.6) a (5.8), sendo apenas para o Cenário 1, uma vez que, conforme verificado anteriormente, não parece haver diferença significativa entre os Cenários 1 e 2.

Neste ponto, a principal conclusão é que o perfil da trajetória de estimativas de  $\hat{\theta}$  apresenta comportamento cíclico, variando em torno do valor verdadeiro de  $\theta$ , independente do valor do fator de desconto  $k$ . Dessa maneira, alguma estratégia precisava ser investigada para se tentar reduzir o efeito de tal comportamento. Assim, são propostos os Cenários 3 e 4, para os quais valores de  $\alpha_0$  e  $\beta_0$  são testados utilizando-se os conhecimentos adquiridos com as simulações realizadas nos Cenários 1 e 2.

#### 5.1.4 Cenários 3 e 4

A experiência com os Cenários 1 e 2 evidenciou a necessidade de se fazer mais testes para definir os parâmetros  $\alpha_0$  e  $\beta_0$ , de modo a estabilizar o crescimento dos parâmetros dinâmicos  $\alpha$  e  $\beta$ , principalmente ao se tentar utilizar fatores de desconto altos ( $k > 0.9$ ), acentuando o comportamento ciclante do estimador  $\hat{\theta}$ .

Os Cenários 3 e 4 surgiram da necessidade de se tentar corrigir os problemas apontados nos Cenários 1 e 2, onde os valores  $\alpha_t$  e  $\beta_t$  apresentaram comportamento crescente, impactando, assim, as estimativas dos  $\theta$ 's. Neste contexto, o Cenário 3 tratou as simulações para  $k \leq 0.85$  e considerou como estimativa inicial para os parâmetros do modelo dinâmico,  $\alpha_0 = \beta_0 = 7$ . No Cenário 4 tratou-se de simulações para o caso de  $k > 0.85$ , onde foram considerados  $\alpha_0 = \beta_0 = 50$ . As Figuras (5.9) e (5.10) apresentam, respectivamente, os resultados do comportamento do  $\alpha$  e do  $\beta$  para estas simulações.

A fixação, em 7 ou 50, nos valores  $\alpha_0$  e  $\beta_0$  parece estabilizar o efeito crescente destes parâmetros ao longo da execução do modelo, padrão este observado, nitidamente, na Figura (5.9), tendo impacto direto nos valores das proficiências estimadas. A comparação dos resultados das Tabelas (5.2) e (5.1) mostra que os valores do RMSE para o Cenário 3 são menores do que os avaliados pelos Cenários 1 e 2, mas ainda superiores ao modelo tradicional.

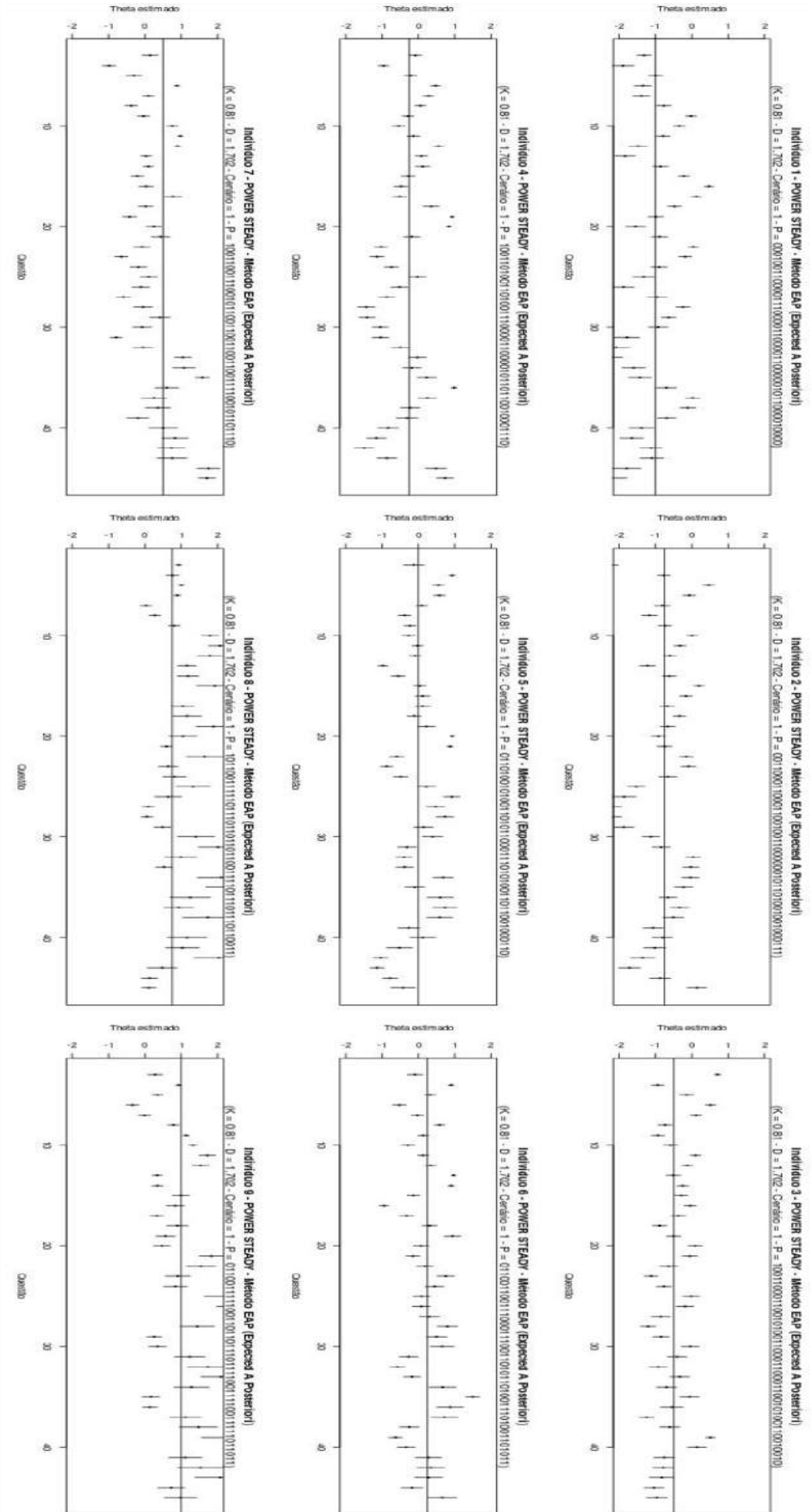


Figura 5.6: Cenário 1: Power Steady - Trajetória das estimativas de proficiência, quando  $k = 0.81$

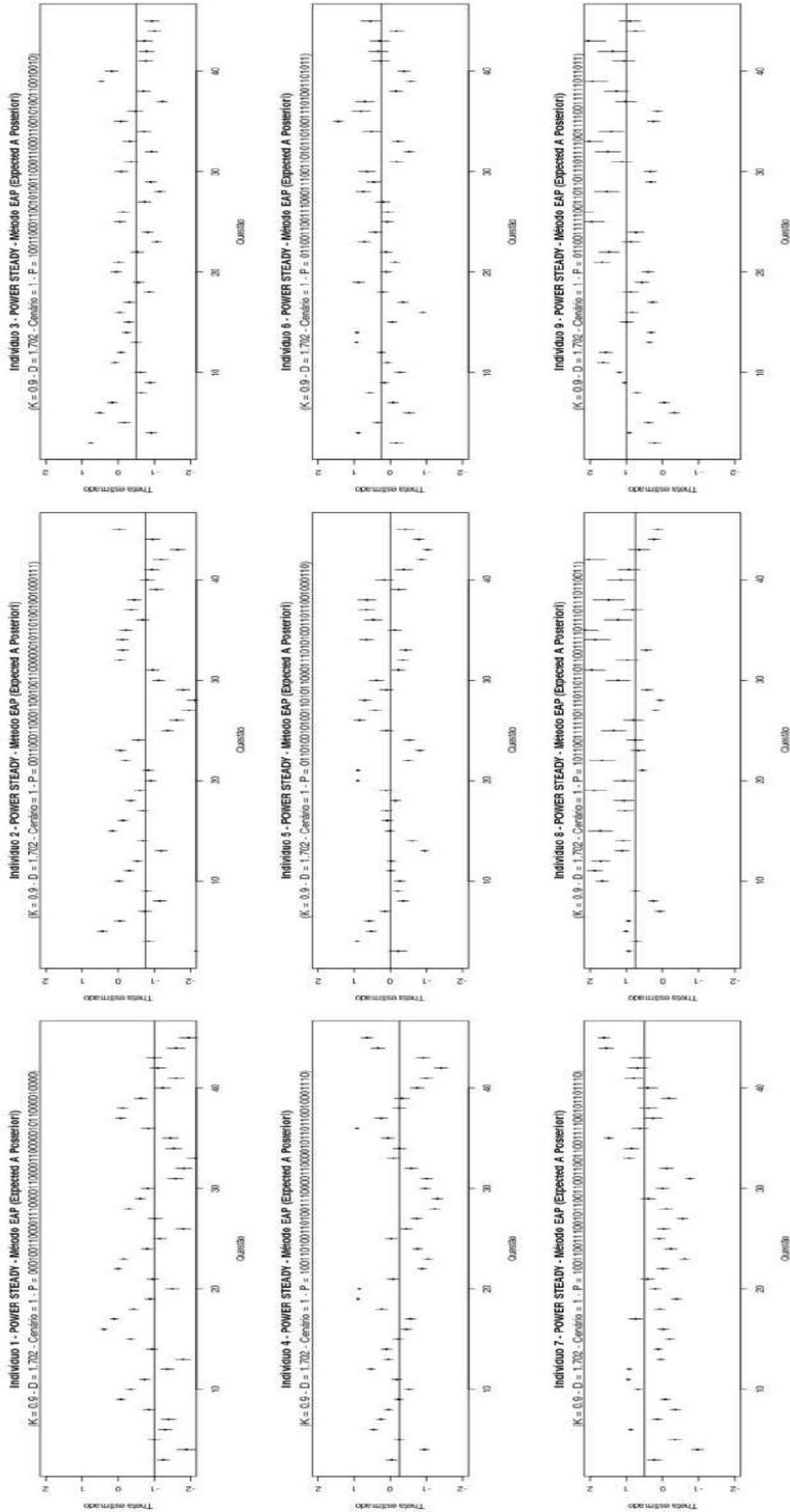


Figura 5.7: Cenário 1: Power Steady - Trajetória das estimativas de proficiência, quando  $k = 0.90$

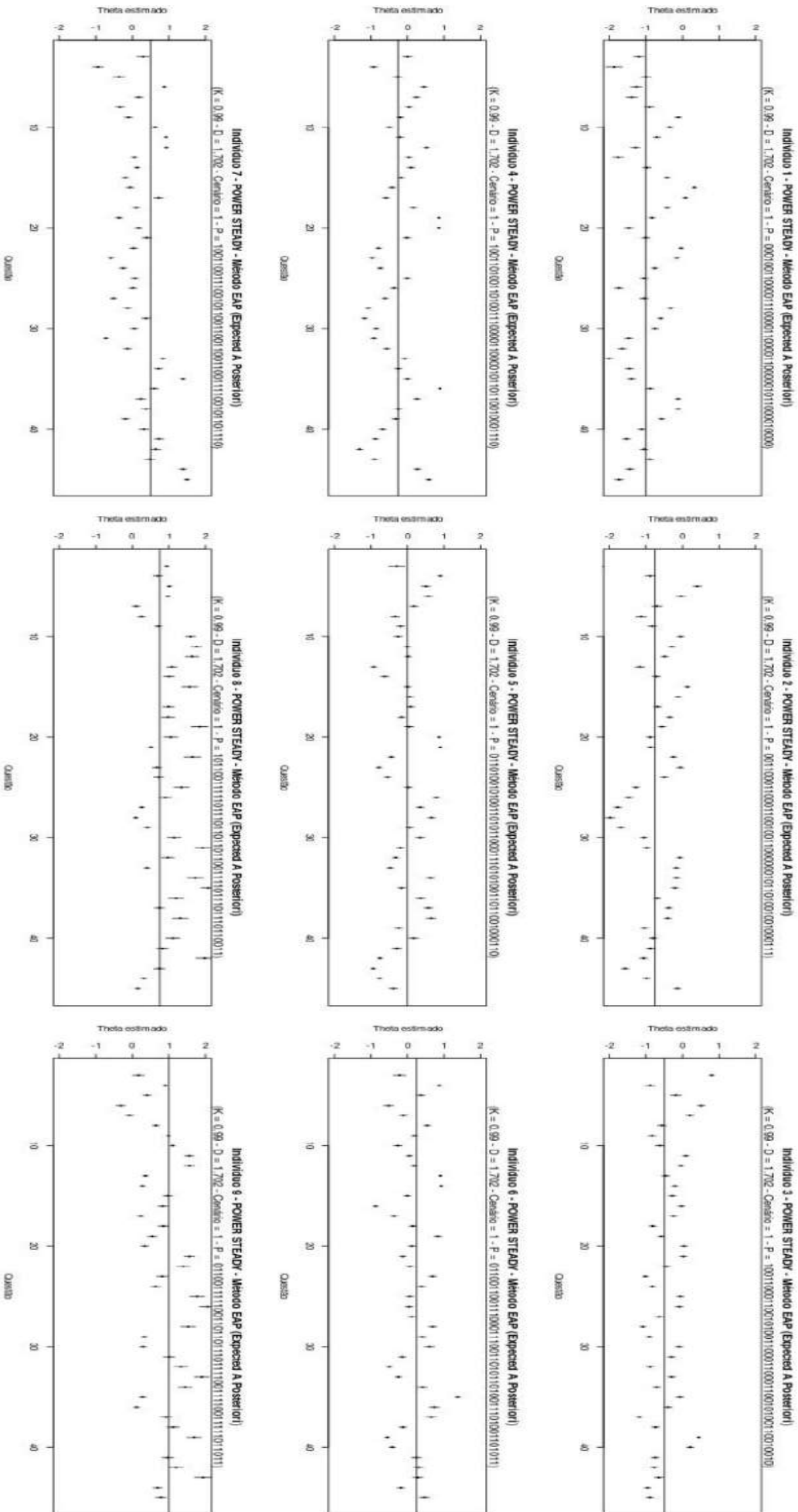
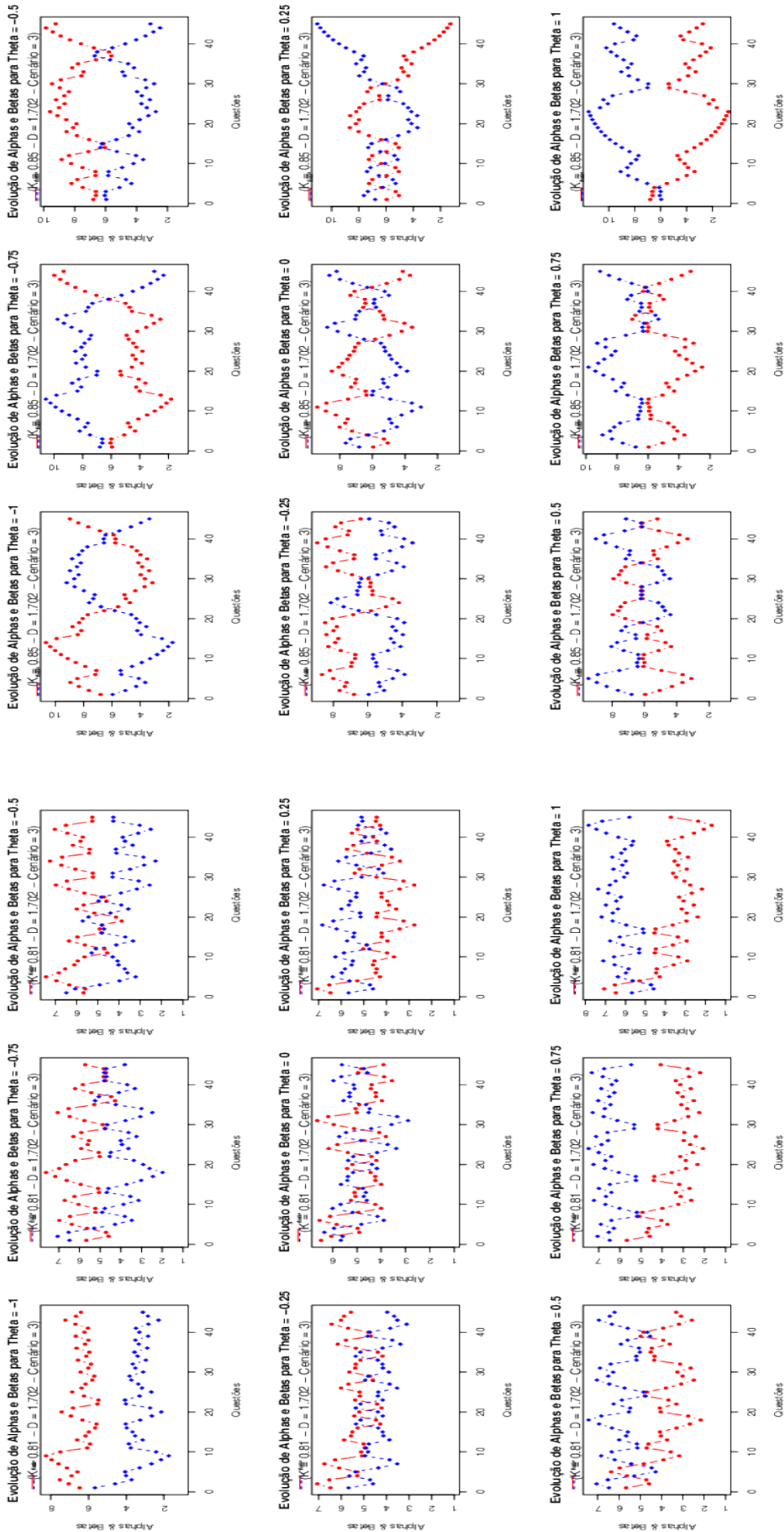


Figura 5.8: Cenário 1: Power Steady - Trajetória das estimativas de proficiência, quando  $k = 0.99$

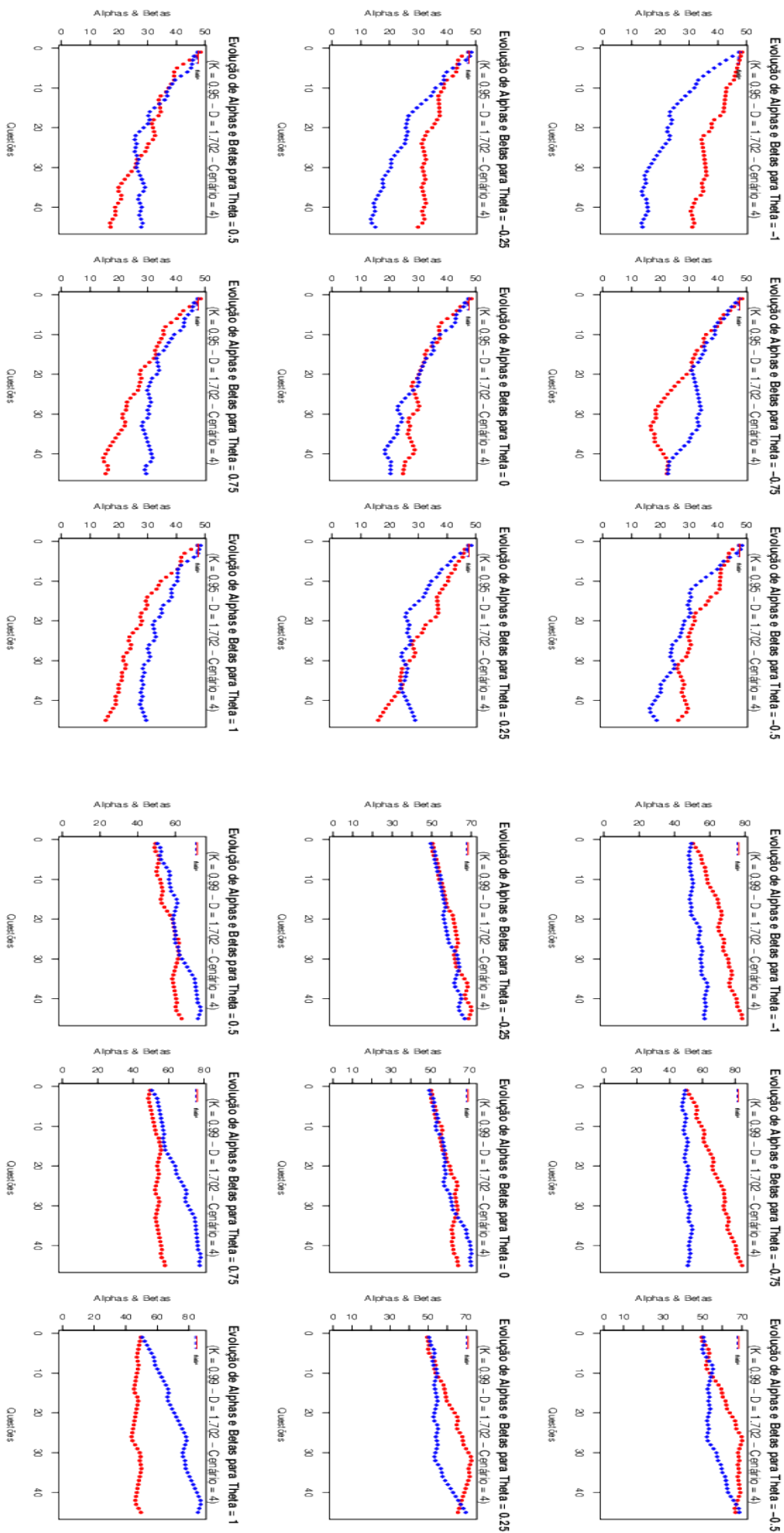




(a)  $k = 0.81$

(b)  $k = 0.85$

**Figura 5.9: Cenário 3: Power Steady - Comportamento dos parâmetros  $\alpha$  e  $\beta$ , quando  $\alpha_0 = \beta_0 = 7$ ,  $k \in \{0.81, 0.85\}$ .**



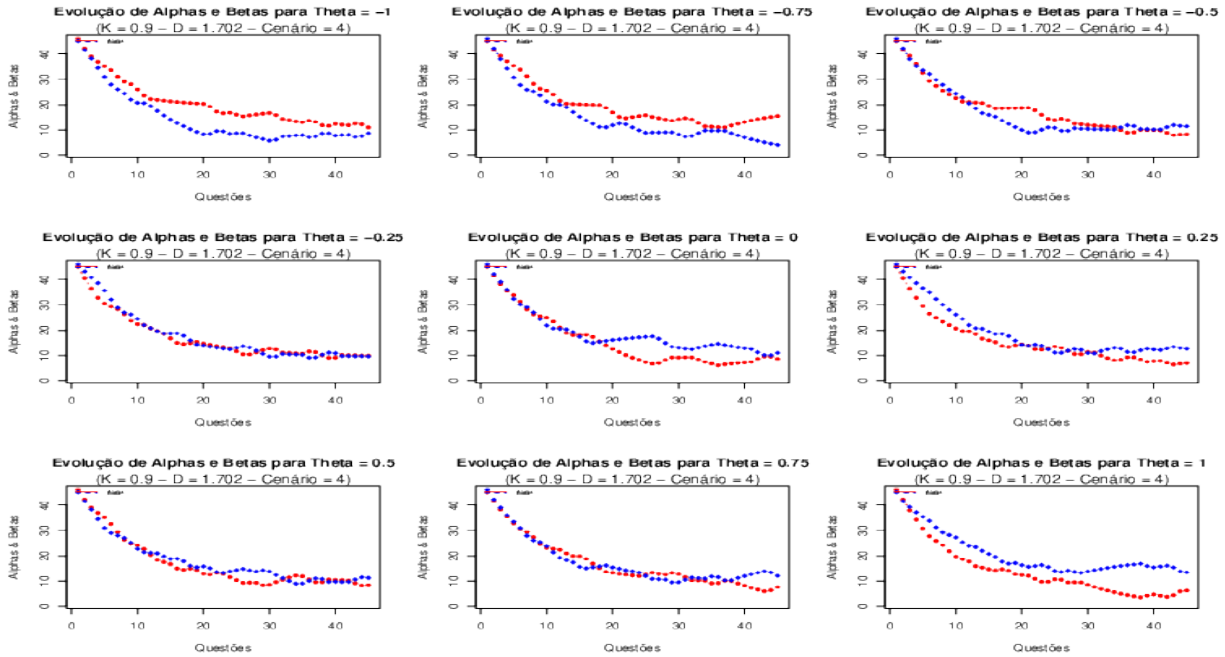
(a)  $k = 0.95$

(b)  $k = 0.99$

Figura 5.10: Cenário 4: Power Steady - Comportamento dos parâmetros  $\alpha$  e  $\beta$ , quando  $\alpha_0 = \beta_0 = 50, k \in \{0.95, 0.99\}$ .

**Tabela 5.2:** Power Steady - *Comparações por RMSE (Cenários 3 e 4)*

Execução	$k$	Modelo/RMSE		
		EAP/3PL	PS (Cenário 3)	PS (Cenário 4)
1	0.81	0.1873	0.5140	-
2	0.85	0.3094	0.5246	-
3	0.90	0.1816	-	0.4390
4	0.95	0.1433	-	0.4225
5	0.99	0.1401	-	0.4194

**Figura 5.11:** Cenário 4: Power Steady - *Comportamento dos parâmetros  $\alpha$  e  $\beta$ , quando  $\alpha_0 = \beta_0 = 50, k = 0.90$* 

Para melhor compreender o impacto dos Cenários 3 e 4 no comportamento das proficiências estimadas, foram observadas as trajetórias das proficiências estimadas de alguns indivíduos, juntamente com a respectiva trajetória dos desvios-padrão das estimativas de proficiência. As Figuras (5.12) e (5.13) revelam, no entanto, que ao fixar  $\alpha_0$  e  $\beta_0$ , em 7 ou 50, não há alteração no comportamento do estimador  $\hat{\theta}$ , que continua ciclante, apresentando, agora, ciclos mais longos e, desvios-padrão distantes do comportamento ideal apresentado pela Figura (5.1).

Para o Cenário 4 ( $\alpha_0 = \beta_0 = 50$ ) houve uma tendência de decrescimento nas curvas de  $\alpha$  e  $\beta$  para valores de  $k \leq 0.95$ . Os resultados da Figura (5.10) mostram que, para valores de  $k > 0.95$ , o comportamento crescente reaparece. O fato curioso é que tal inversão parece indicar uma suave melhora no efeito dos parâmetros  $\alpha$  e  $\beta$  sobre as proficiências estimadas, como pode ser observado nas últimas 3 linhas da Tabela (5.2), onde à medida que  $k$  cresce, os valores dos RMSE obtidos vão ficando gradativamente menores, embora ainda muito maiores do que os apresentados pelo modelo 3AP/3PL.

Apesar das mudanças propostas nos parâmetros dinâmicos nos Cenários 3 e 4, ainda não surtiram efeito para atenuar o comportamento ciclante do estimador  $\hat{\theta}$ , um efeito sistemático é observado quanto à variabilidade do mesmo, de maneira que, quanto maior o valor

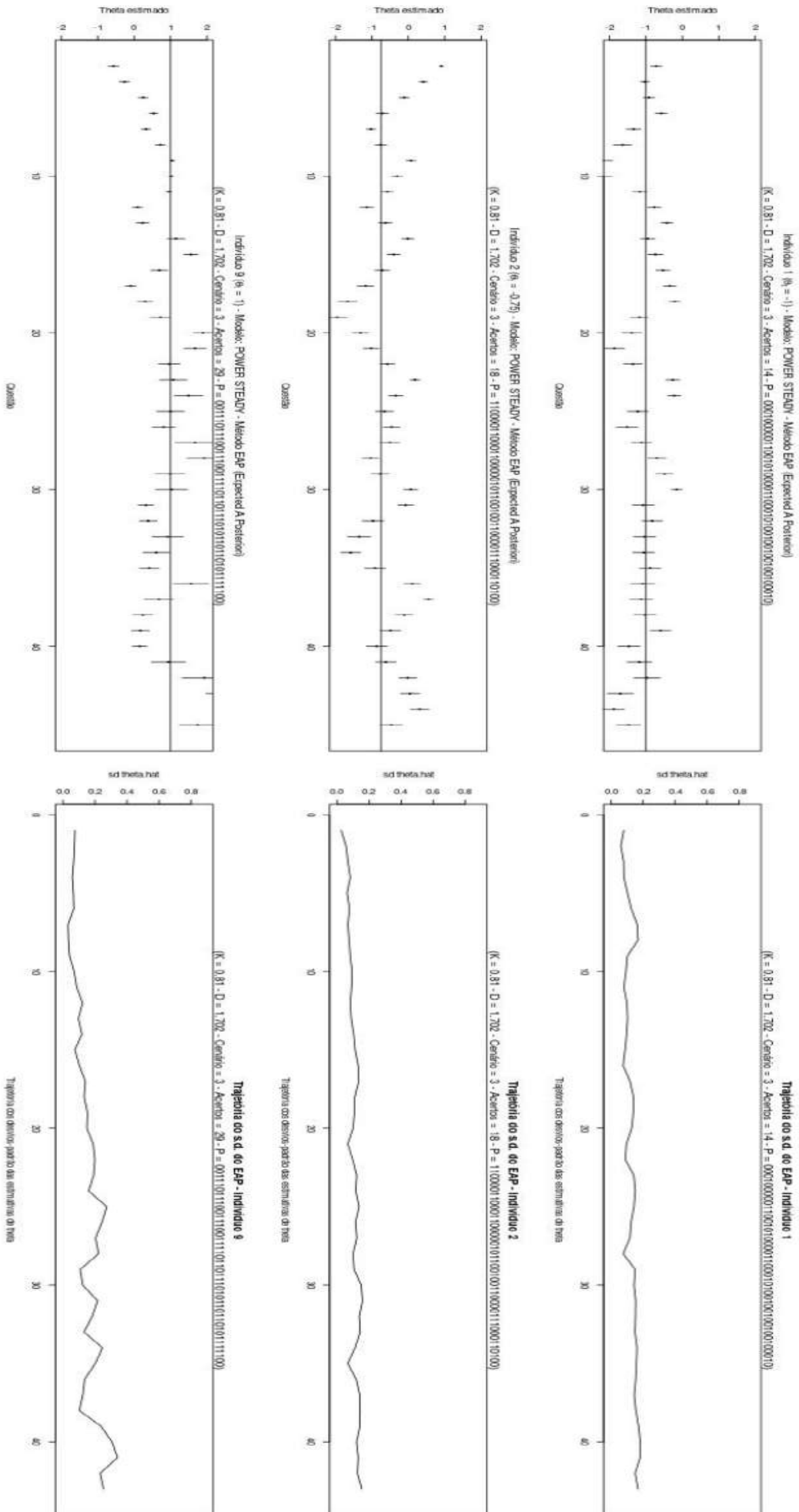
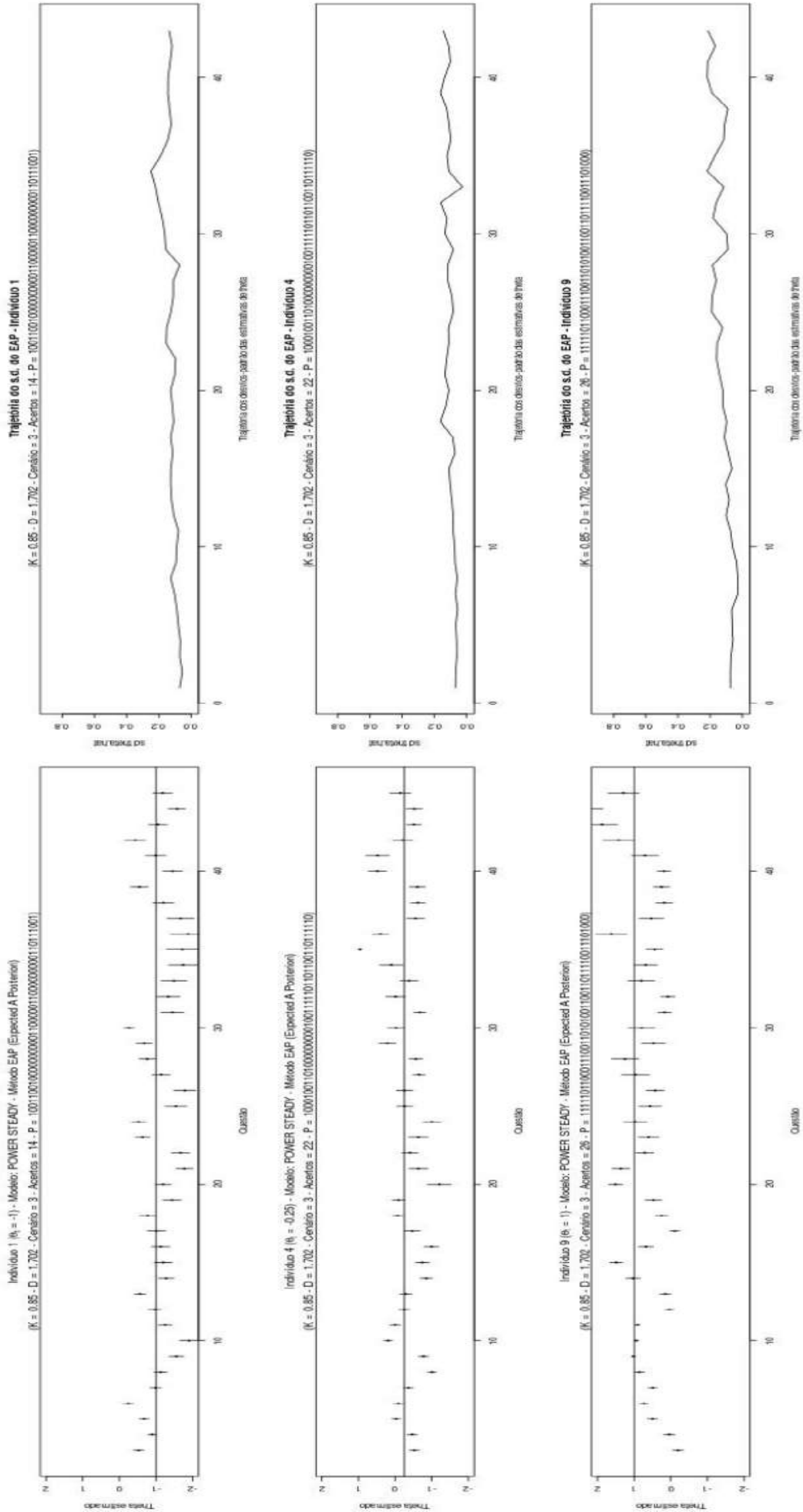


Figura 5.12: Cenário 3: Power Steady - Estimativas de proficiência quando  $\theta \in [-1, 1]$ ,  $k = 0.81$



**Figura 5.13:** Cenário 3: Power Steady - Estimativas de proficiência, quando  $\theta \in [-1, 1]$ ,  $k = 0.85$

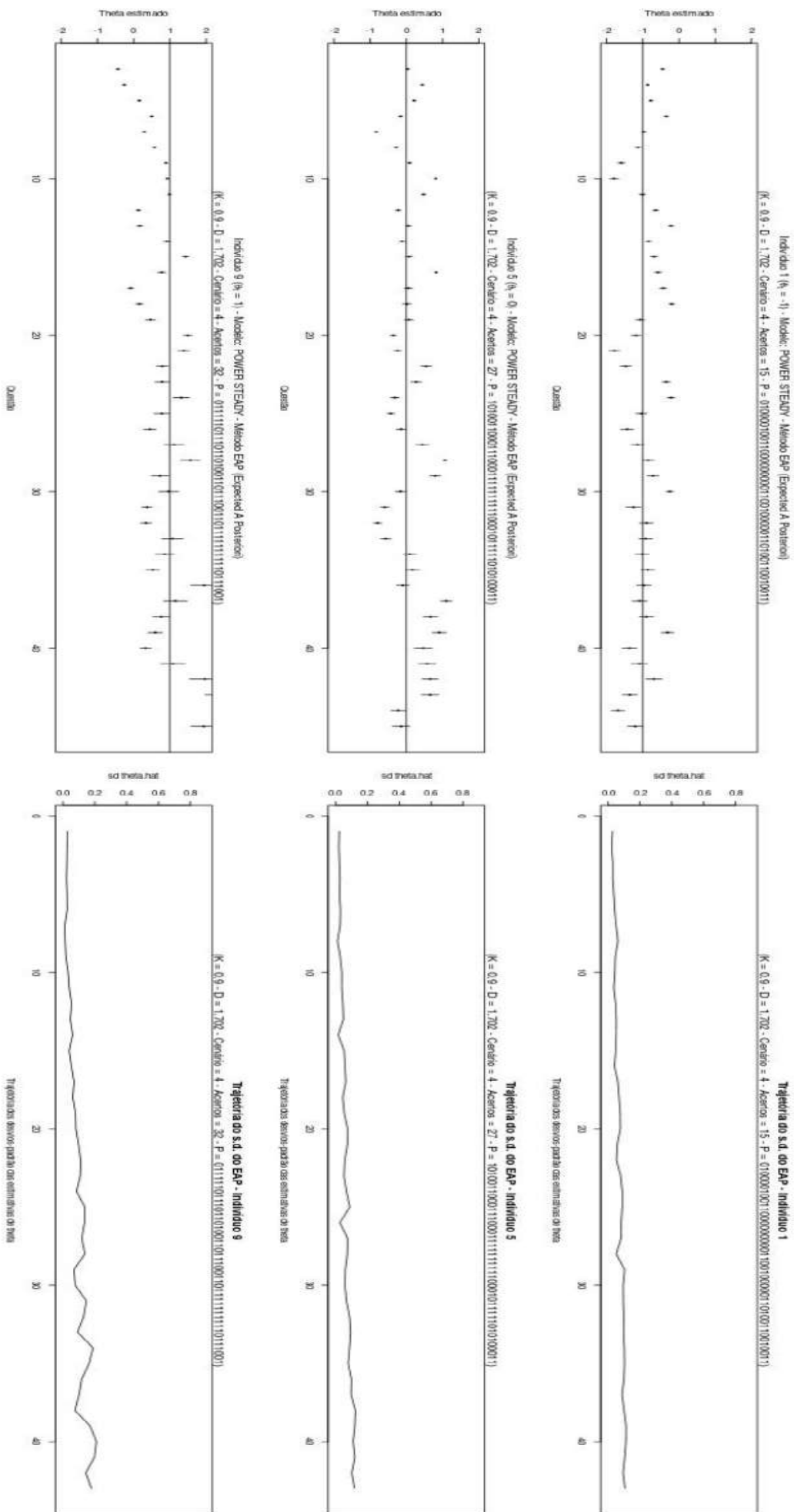
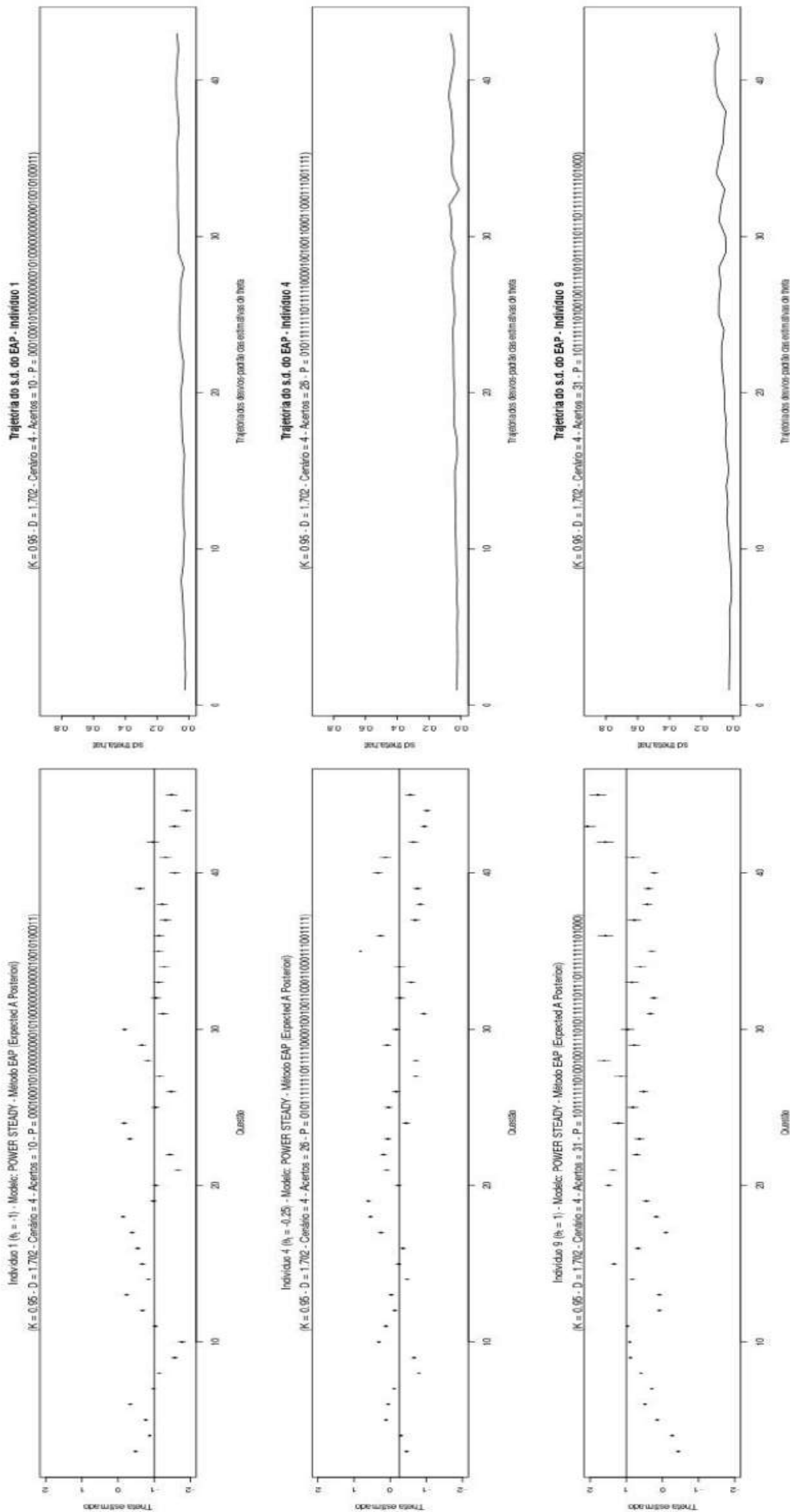


Figura 5.14: Cenário 4: Power Steady - Estimativas de proficiência, quando  $\theta \in [-1, 1]$ ,  $k = 0.90$



**Figura 5.15:** Cenário 4: Power Steady - Estimativas de proficiência, quando  $\theta \in [-1, 1]$ ,  $k = 0.95$

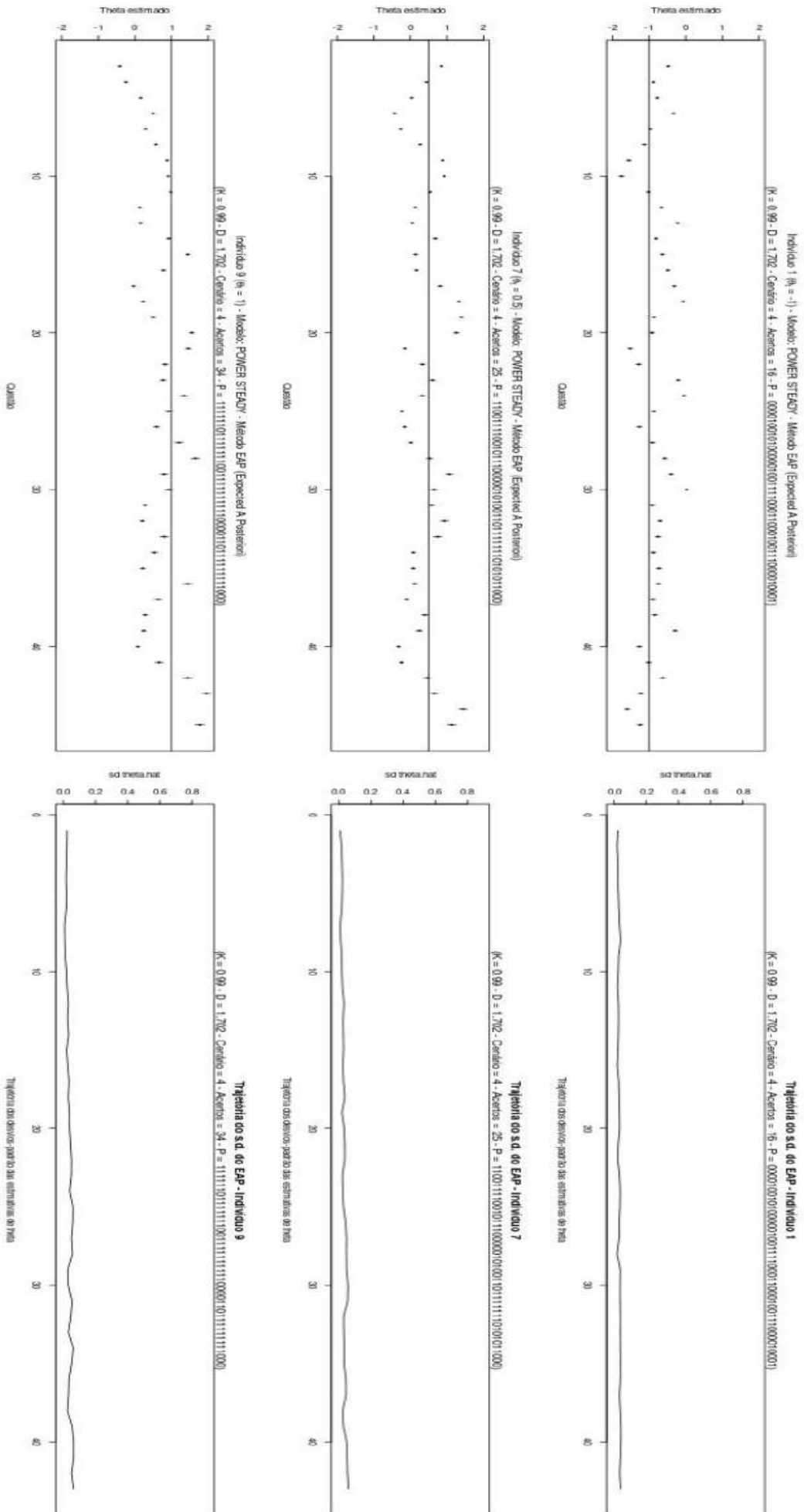


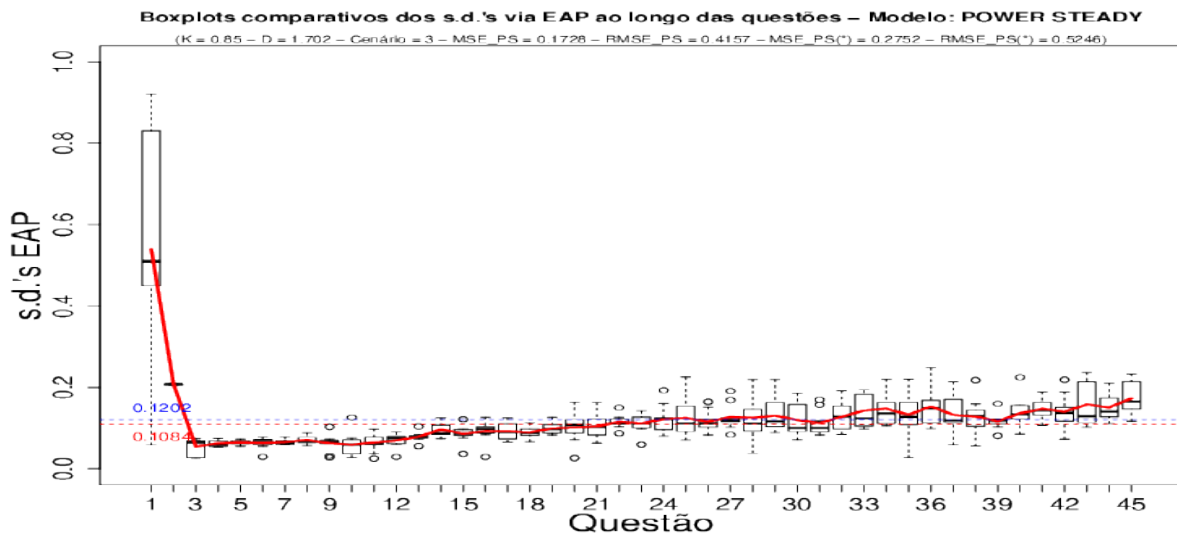
Figura 5.16: Cenário 4: Power Steady - Estimativas de proficiência quando  $\theta \in [-1, 1]$ ,  $k = 0.99$



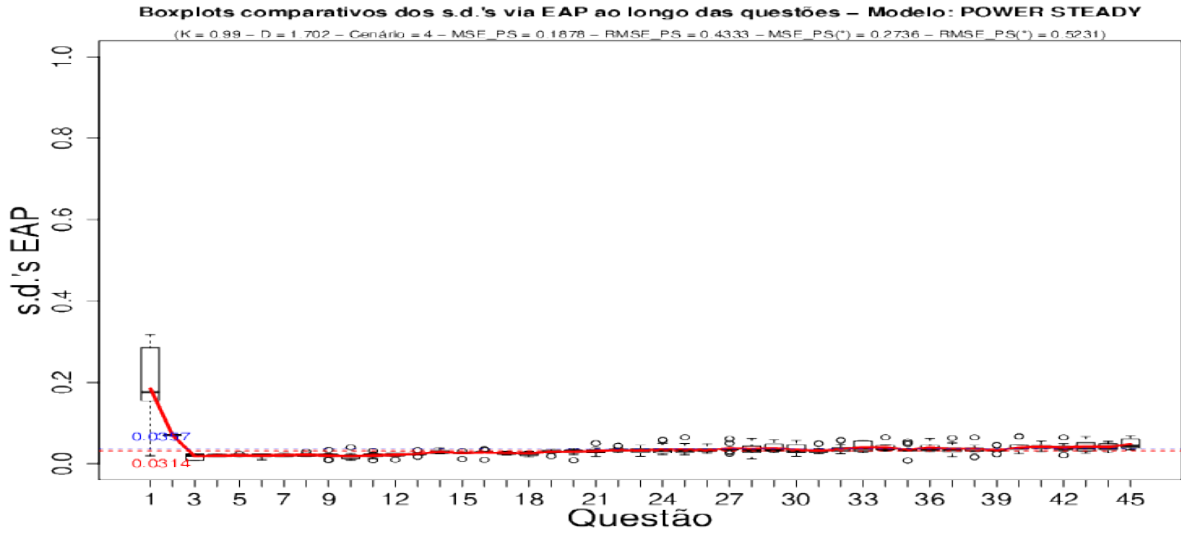
de  $k$ , menor são os desvios-padrão observados ao longo da trajetória das estimativas de  $\theta$  mascarando, assim, a precisão na estimação de tal parâmetro.

As Figuras (5.17) e (5.18) apresentam as trajetórias dos desvios-padrão das estimativas ao longo das questões. Observa-se que, apesar das mesmas encontrarem-se abaixo do limite 0.2, variam muito pouco e com tendência de crescimento ao longo do CAT, sendo um padrão totalmente indesejável para um estimador no contexto.

Em função do modelo *Power Steady* depender dos valores iniciais de seus parâmetros, a sugestão alternativa, nesta dissertação, é calcular  $\alpha_0$  e  $\beta_0$  ajustados de acordo com uma estimativa inicial de  $\theta$ , proveniente do método EAP/3PL do CAT convencional, até 3 ou 5 iterações, tal como explicado na Seção 5.1.6, deixando, a partir de então, o processo de atualização seguir normalmente, como proposto a seguir, pelo modelo *Power Steady*.



**Figura 5.17: Cenário 3: Power Steady - Trajetória das estimativas de desvio-padrão de  $\hat{\theta}$ , quando  $k = 0.85$**



**Figura 5.18: Cenário 4: Power Steady - Trajetória das estimativas de desvio-padrão de  $\hat{\theta}$ , quando  $k = 0.99$**

### 5.1.5 Cenário 5

Os resultados das simulações com os cenários anteriores mostraram que a questão crucial para a obtenção de estimativas plausíveis do modelo *Power Steady* é altamente dependente da forma com que se escolhem os parâmetros  $\alpha_0$  e  $\beta_0$ , de modo que, a sugestão discutida na seção anterior, é materializada no Cenário 5, onde é aplicado um método sistemático para obtenção de valores iniciais para os parâmetros iniciais do modelo dinâmico.

### 5.1.6 Obtenção de valores iniciais de $\alpha_{t|t}$ e $\beta_{t|t}$

Uma questão crucial, é definir valores iniciais de  $\alpha_{t|t}$  e  $\beta_{t|t}$ , discutidos a seguir.

Considerando os desenvolvimentos da Seção 4.6.2, têm-se que

$$f_{jt} = E(\eta_{jt}) \approx \log \left( \frac{\beta_{t|t}}{\alpha_{t|t}} \right), \quad (5.5)$$

$$q_{jt} = \text{Var}(\eta_{jt}) \approx \frac{1}{\beta_{t|t}} + \frac{1}{\alpha_{t|t}}, \quad (5.6)$$

De (5.5) e (5.6),

$$\alpha_{t|t} = \beta_{t|t} \exp(-f_{jt}), \quad (5.7)$$

$$\beta_{t|t} = \frac{1 + \exp(f_{jt})}{q_{jt}}. \quad (5.8)$$

Em (5.5) e (5.6), tomando-se as seguintes aproximações:

$$\begin{aligned}\hat{\eta}_{jt} &\approx E(\eta_{jt}), \\ \text{Var}(\hat{\eta}_{jt}) &\approx \text{Var}(\eta_{jt}),\end{aligned}$$

e recordando que  $\eta_{jt} = Da_t(\theta_{jt} - b_t)$ , então tem-se

$$\hat{\eta}_{jt} = Da_t(\hat{\theta}_{jt} - b_t) \approx E(\eta_{jt}), \quad (5.9)$$

$$\text{Var}(\hat{\eta}_{jt}) = (Da_t)^2 \text{Var}(\hat{\theta}_{jt}) \approx \text{Var}(\eta_{jt}). \quad (5.10)$$

Dessa forma,

$$\alpha_{t|t} = \beta_{t|t} \exp(-Da_t(\hat{\theta}_{jt} - b_t)), \quad (5.11)$$

$$\beta_{t|t} = \frac{1 + \exp(Da_t(\hat{\theta}_{jt} - b_t))}{(Da_t)^2 \text{Var}(\hat{\theta}_{jt})}. \quad (5.12)$$

Nas expressões (5.11) e (5.12) a idéia é, com base em alguma estimativa de  $\theta_{jt}$ , calcular valores iniciais bem ajustados (“tuned”) para  $\alpha_{t|t}$  e  $\beta_{t|t}$ , uma vez que o método Power Steady é bastante dependente dos valores iniciais desses referidos parâmetros.

Nossa sugestão, nesta dissertação, é utilizar uma estimativa inicial de  $\theta_{jt}$  proporcionada pelo método EAP do CAT convencional nas primeiras 3 a 5 iterações (estimativas provisórias) do CAT. Dessa forma,  $\hat{\theta}_{jt}$  e  $\text{Var}(\hat{\theta}_{jt})$  são, então, substituídos pelos respectivos valores provisórios do CAT via método EAP convencional e, posteriormente, calcula-se  $\alpha_{t|t}$  e  $\beta_{t|t}$ . A partir desse ponto, utilizam-se as equações de atualização, apresentadas em (4.149) e (4.150), e expressas por

$$\begin{cases} \alpha_{t+1|t+1} = \alpha_{t+1|t} + 1 - x_{j,t+1}, \\ \beta_{t+1|t+1} = \beta_{t+1|t} + x_{j,t+1}. \end{cases}$$

Ou seja,

$$\begin{cases} \alpha_{t+1|t+1} = k(\alpha_{t|t} - x_{jt} + 1) + (1 - x_{j,t+1}), \\ \beta_{t+1|t+1} = k(\beta_{t|t} + x_{jt}) + x_{j,t+1}. \end{cases}$$

Para o Cenário 5, foram consideradas simulações onde  $\theta \in [-3, 3]$ , com saltos (Gaps) de 0.25 entre cada valor. Para avaliar o comportamento dos parâmetros iniciais de  $\alpha$  e  $\beta$ , são apresentadas nas Tabelas (5.3) a (5.6), estatísticas descritivas de  $\alpha_t$  e  $\beta_t$ , avaliados segundo o método EAP, sendo considerados casos para  $\theta = 0.85$  e  $\theta = 0.99$ . Nas Tabelas (5.3) a (5.10) os parâmetros  $\lambda$  e  $\tau$  serão discutidos ao final desta seção.

**Tabela 5.3:** Estatísticas de caracterização para  $\alpha_t$ , quando  $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	7.91500	8.89000	1.26300	10.80500	5.7370	9.90600
-2.75	7.50550	8.53870	0.75990	11.11770	6.0640	9.59170
-2.50	6.75700	7.67800	1.02900	10.36800	4.3570	8.94800
-2.25	7.08700	7.45500	1.04200	10.94400	5.4390	8.78700
-2.00	7.05000	7.68100	1.32600	10.04900	6.0610	8.54000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.63600	2.24530	0.54890	5.09300	1.6222	3.65610
2.25	2.06730	2.14080	0.41520	4.15810	1.3147	2.89000
2.50	1.88230	1.82990	0.28500	4.22760	0.9552	2.49840
2.75	1.32470	1.15810	0.18600	3.47810	0.6381	1.81560
3.00	0.94864	0.75859	0.07796	2.44527	0.3821	1.45321

**Tabela 5.4:** Estatísticas de caracterização para  $\beta_t$ , quando  $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	2.1235	2.0946	0.3568	4.4005	1.3358	2.8991
-2.75	2.5353	2.3681	0.7907	4.9990	1.6721	3.2777
-2.50	3.3102	3.1754	0.8032	6.4711	2.2803	4.1004
-2.25	2.9756	3.1050	0.5605	5.5234	1.7597	4.0566
-2.00	2.9877	2.9811	0.3707	5.7187	2.1409	4.2562
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	7.3610	7.9340	1.1040	11.3470	6.2110	9.6320
2.25	7.9096	8.4961	0.8859	11.4292	5.9903	9.9768
2.50	8.0821	8.6980	0.8626	11.4847	7.2437	10.1008
2.75	8.6590	10.3480	0.9570	12.0510	6.7550	11.3390
3.00	9.0127	10.1098	0.8689	12.2391	7.4298	11.7701

**Tabela 5.5:** Estatísticas de caracterização para  $\alpha_t$ , quando  $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	27.7060	28.6120	1.2630	54.1750	11.7550	42.6760
-2.75	25.6452	25.5575	0.7599	53.5738	10.3458	40.7591
-2.50	25.3980	26.7330	1.0290	49.9290	8.6360	41.1080
-2.25	25.4000	28.5360	1.0420	51.4860	8.6100	36.9960
-2.00	22.3600	20.1600	1.3260	44.5710	10.9320	36.1550
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	7.0755	8.0182	0.5489	13.7648	2.5180	9.0284
2.25	6.2770	7.0740	0.6560	12.4170	4.4410	7.9010
2.50	6.0572	7.4802	0.6108	10.2872	2.5786	8.5615
2.75	6.0853	5.6651	0.6318	12.0148	4.3738	8.1877
3.00	3.0470	3.5281	0.6037	4.4012	2.5205	3.9406

**Tabela 5.6:** Estatísticas de caracterização para  $\beta_t$ , quando  $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	6.0671	6.2855	0.5638	12.9330	3.4821	8.6591
-2.75	8.1370	9.1770	1.0730	14.2390	4.9510	11.5030
-2.50	8.4846	8.5033	0.9355	17.4410	6.6557	11.0063
-2.25	8.4665	6.7316	0.8857	16.2109	6.0270	13.6359
-2.00	11.4097	14.3067	0.5079	22.5340	4.3887	16.1688
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	26.5380	26.6860	1.1040	53.4900	12.6650	42.8750
2.25	27.2602	26.3344	0.8859	54.9923	10.4439	44.4059
2.50	27.4323	26.3219	0.8627	56.8997	12.3430	42.4345
2.75	27.4800	28.4240	0.9570	54.9270	10.5880	42.8840
3.00	30.4309	30.0739	0.8689	63.5836	12.3883	47.7540

De acordo com os resultados do método EAP apresentados pelas Tabelas (5.3) a (5.6), observa-se uma tendência de crescimento do parâmetro  $\beta_t$  à medida em que  $k$  cresce. Tal comportamento, considerando-se  $k = 0.99$ , atingiu o valor máximo de 63.5836 para  $\beta_t$ , quando  $\theta = 3$  (vide Tabela 5.6). No parâmetro  $\alpha_t$ , observa-se uma tendência de decréscimo à medida em que  $k$  cresce, atingindo valor mínimo de 0.6037, ao longo das questões respondidas quando  $\theta = 3$  (vide Tabela 5.5).

Nas Tabelas (5.7) a (5.10) apresentam-se estatísticas de sumarização dos parâmetros  $\alpha$  e  $\beta$  sob o cenário 5 e método MAP. Verifica-se que  $\alpha_t$  e  $\beta_t$  têm comportamento similar aos obtidos com o método EAP (Tabelas 5.3 a 5.6). Além disso, similares aos Cenários 1 e 2.

**Tabela 5.7:** Estatísticas de caracterização para  $\alpha_t$ , quando  $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	8.0730	8.7390	1.3060	11.1810	7.4400	9.7380
-2.75	8.1770	9.1210	1.3160	11.3000	6.5820	10.0750
-2.50	7.3100	7.7740	1.4510	11.0100	5.6500	8.8350
-2.25	7.1840	7.4900	1.3740	11.1590	5.6030	8.8340
-2.00	7.1766	8.1686	0.5718	10.5181	6.4923	9.1261
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.4952	2.4594	0.4217	5.8359	1.5833	3.0905
2.25	1.6504	1.5938	0.2988	3.5921	1.0001	2.2060
2.50	2.3113	2.2469	0.5137	4.3043	1.4366	3.0738
2.75	1.8650	1.8010	0.4650	4.3950	1.3000	2.3810
3.00	1.7125	1.7312	0.4783	3.4000	1.0900	2.2373

**Tabela 5.8:** Estatísticas de caracterização para  $\beta_t$ , quando  $k = 0.85, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	2.0911	2.0895	0.4191	4.4594	1.1113	2.7761
-2.75	1.9620	1.8990	0.5150	3.8260	1.1100	2.6660
-2.50	2.7697	2.9582	0.3037	5.4750	1.2681	4.0944
-2.25	2.9090	2.8270	0.4410	7.1780	1.4760	3.9710
-2.00	2.7750	2.7510	0.4990	5.6850	1.8310	3.8600
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	7.5000	8.0760	0.9610	10.9690	6.0140	9.4200
2.25	8.3234	9.3068	0.8746	11.1937	7.2491	10.1551
2.50	7.6660	8.6139	0.9597	11.5562	6.2080	10.2056
2.75	8.0967	9.5857	0.8494	11.3600	4.7928	10.4687
3.00	8.2773	9.2516	0.9588	11.5297	7.6429	10.3471

**Tabela 5.9:** Estatísticas de caracterização para  $\alpha_t$ , quando  $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	26.5180	24.7540	1.3060	53.7610	13.5990	41.0110
-2.75	26.9670	26.6790	1.3160	51.5570	10.8060	44.3560
-2.50	24.4980	25.0240	1.4510	50.8110	12.9110	35.5080
-2.25	23.6830	26.9300	1.3740	45.0700	8.9190	35.6170
-2.00	22.6570	25.1890	0.6660	45.7310	11.3310	33.0210
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	7.9495	6.3221	0.6663	15.8219	4.5408	12.9396
2.25	5.9549	6.0544	0.6404	11.5809	2.4798	10.4735
2.50	7.4137	9.1600	0.5923	11.6927	4.4447	10.3164
2.75	7.0280	7.7062	0.6308	13.8477	6.0832	8.9418
3.00	4.9985	5.6997	0.6489	9.2868	2.4877	7.4551

**Tabela 5.10:** Estatísticas de caracterização para  $\beta_t$ , quando  $k = 0.99, \lambda = 1, \tau = 1$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	7.7360	9.6140	1.0460	14.1510	2.0360	10.7780
-2.75	7.1914	7.4677	0.9477	16.1799	4.6683	8.2573
-2.50	9.4336	9.6048	0.5589	17.9787	2.5278	16.4239
-2.25	10.2984	8.0653	0.6968	22.4284	6.2920	16.4225
-2.00	10.7835	8.9499	0.7742	21.7583	3.6085	18.3730
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	25.6600	28.4140	0.9610	51.7780	10.5140	38.7340
2.25	27.5707	28.0463	0.8747	56.8494	12.3542	40.2325
2.50	26.1253	24.6050	0.9597	55.5755	10.5321	40.9555
2.75	26.4515	26.4715	0.8494	53.3013	8.4493	42.6645
3.00	28.5881	28.1692	0.9588	58.4753	12.4915	44.0686

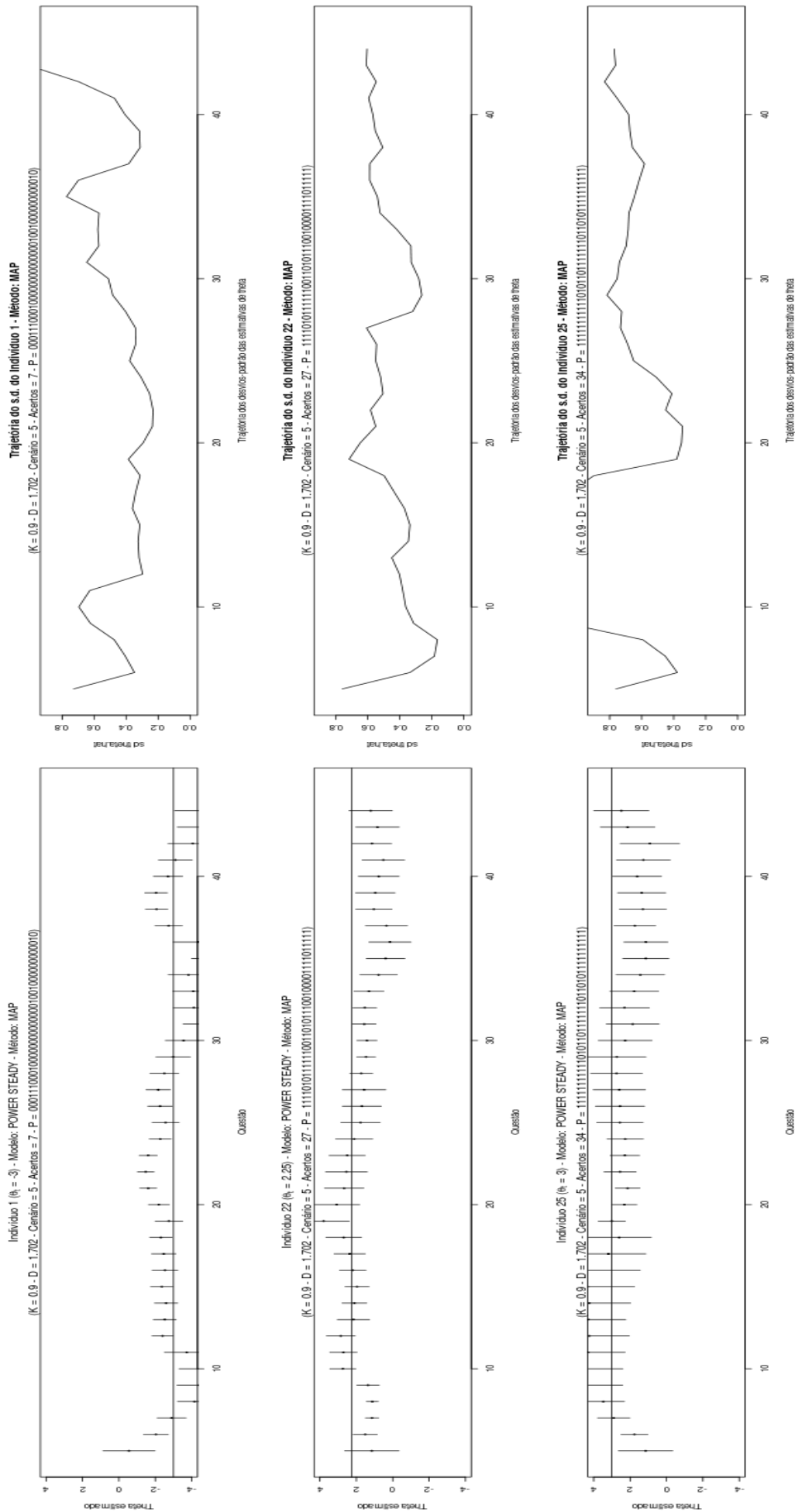
Analisando-se, agora, o comportamento da trajetória do estimador  $\tilde{\theta}$  (MAP) ao longo do CAT, apresentado pelas Figuras (5.19) a (5.21), onde considerou-se apenas o método MAP, uma vez que houve grande similaridade entre os resultados obtidos pelos métodos MAP e EAP, observou-se que o efeito cíclico dos estimadores  $\hat{\theta}$  e  $\tilde{\theta}$  se manteve.

Como os resultados obtidos com o método MAP para os parâmetros  $\alpha_t$  e  $\beta_t$  do modelo dinâmico permaneciam indicando trajetórias cíclicas, para ambos os estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , ao longo do CAT, duas estratégias foram testadas no intuito de se investigar a causa dos ciclos, a saber:

- 1) Inclusão de duas constantes nas equações de transição, na tentativa de corrigir respectivamente, o padrão cíclico das estimativas de  $\theta$  ao longo das questões CAT e a variância do estimador, para a qual se observou comportamento constante ao longo da trajetória do CAT (ao invés de observar-se variâncias decrescentes ao longo do processo). Assim,







**Figura 5.20:** Cenário 5: Power Steady - Estimativas de proficiência quando  $\theta \in [-3, 3]$ ,  $k = 0.90$



em função da forte relação observada entre os parâmetros  $\alpha_t$  e  $\beta_t$  e os estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , as “*equações de transição*” apresentadas em (4.150), expressas por

$$\begin{cases} \alpha_{t+1} = k(\alpha_t - x_t + 1) + (1 - x_{t+1}), \\ \beta_{t+1} = k(\beta_t + x_t) + x_{t+1}. \end{cases}$$

foram acrescentadas de dois parâmetros,  $\lambda$  e  $\tau$ , para tentar atenuar o comportamento cíclico das trajetórias de estimativas de  $\theta$ :

$$\begin{cases} \alpha_{t+1}^* = \{\lambda[k(\alpha_t^* - x_t + 1) + (1 - x_{t+1})]\}^\tau \\ \beta_{t+1}^* = \{\lambda[k(\beta_t^* + x_t) + x_{t+1}]\}^\tau \end{cases}. \quad (5.13)$$

Considerou-se  $\lambda$ , uma constante para suavizar as oscilações de  $\alpha_t$  e  $\beta_t$  e  $\tau$  uma constante para reduzir a variância dos estimadores.

- 2) Adaptação dos estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , baseado nos estudos realizados com o *Filtro de Kalman*, desenvolvidos na Seção 4.5.6, onde o estimador  $\hat{\theta}$  foi então apresentado em (4.106), como solução para encontrar um estimador suavizado de  $\theta$ , calculado progressivamente, com as últimas estimativas de  $\theta$  e definido por:

$$\hat{\theta}_{jt} = \sum_{r=1}^t \frac{\tilde{\theta}_{jr}}{t}, \quad (5.14)$$

sendo apresentado na Seção 5.2 os estudos de simulação realizados. Utilizando o mesmo raciocínio, são propostos, nesta dissertação, duas suavizações, para cada um dos estimadores  $\tilde{\theta}$  (estimador para a *Moda a Posteriori*) e  $\hat{\theta}$  (estimador para a *Esperança a Posteriori*), a saber:

- Estimador suavizado de  $\tilde{\theta}$  (PS\_MAP\_AVG): consiste em tomar a média das estimativas correntes via *Power Steady* para a *Moda a Posteriori* (MAP) de  $\theta$ , até a *t-ésima* questão, sendo expresso pela média atualizada dos  $\tilde{\theta}$  até o “*tempo atual*”, isto é, caso o indivíduo tenha respondido à *t-ésima* questão, tem-se:

$$\bar{\tilde{\theta}}_{jt} = \sum_{r=1}^t \frac{\tilde{\theta}_{jr}}{t}. \quad (5.15)$$

- Estimador suavizado de  $\tilde{\theta}$  (PS\_MAP\_MEDIAN): consiste em tomar a mediana das estimativas correntes via *Power Steady* para a *Moda a Posteriori* (MAP) de  $\theta$ , até a *t-ésima* questão, sendo expresso pela mediana atualizada dos  $\tilde{\theta}$  até o “*tempo atual*”, isto é, caso o indivíduo tenha respondido à *t-ésima* questão, tem-se:

$$\dot{\tilde{\theta}}_{jt} = \text{Mediana}\{\tilde{\theta}_{j1}, \dots, \tilde{\theta}_{jt}\}. \quad (5.16)$$

- Estimador suavizado de  $\hat{\theta}$  (PS\_EAP\_AVG): consiste em tomar a média das estimativas correntes via *Power Steady* para a *Esperança a Posteriori* (EAP) de  $\theta$ , até a *t-ésima* questão, sendo expresso pela média atualizada dos  $\hat{\theta}$  até o “*tempo atual*”, isto é, caso o indivíduo tenha respondido à *t-ésima* questão, tem-se:

$$\bar{\hat{\theta}}_{jt} = \sum_{r=1}^t \frac{\hat{\theta}_{jr}}{t}. \quad (5.17)$$

- Estimador suavizado de  $\hat{\theta}$  (PS\_EAP\_MEDIAN): consiste em tomar a mediana das estimativas correntes via *Power Steady* para a *Esperança a Posteriori* (EAP) de  $\theta$ , até a  $t$ -ésima questão, sendo expresso pela mediana atualizada dos  $\hat{\theta}$  até o “tempo atual”, isto é, caso o indivíduo tenha respondido à  $t$ -ésima questão, tem-se:

$$\hat{\theta}_{jt} = \text{Mediana}\{\hat{\theta}_{j1}, \dots, \hat{\theta}_{jt}\}. \quad (5.18)$$

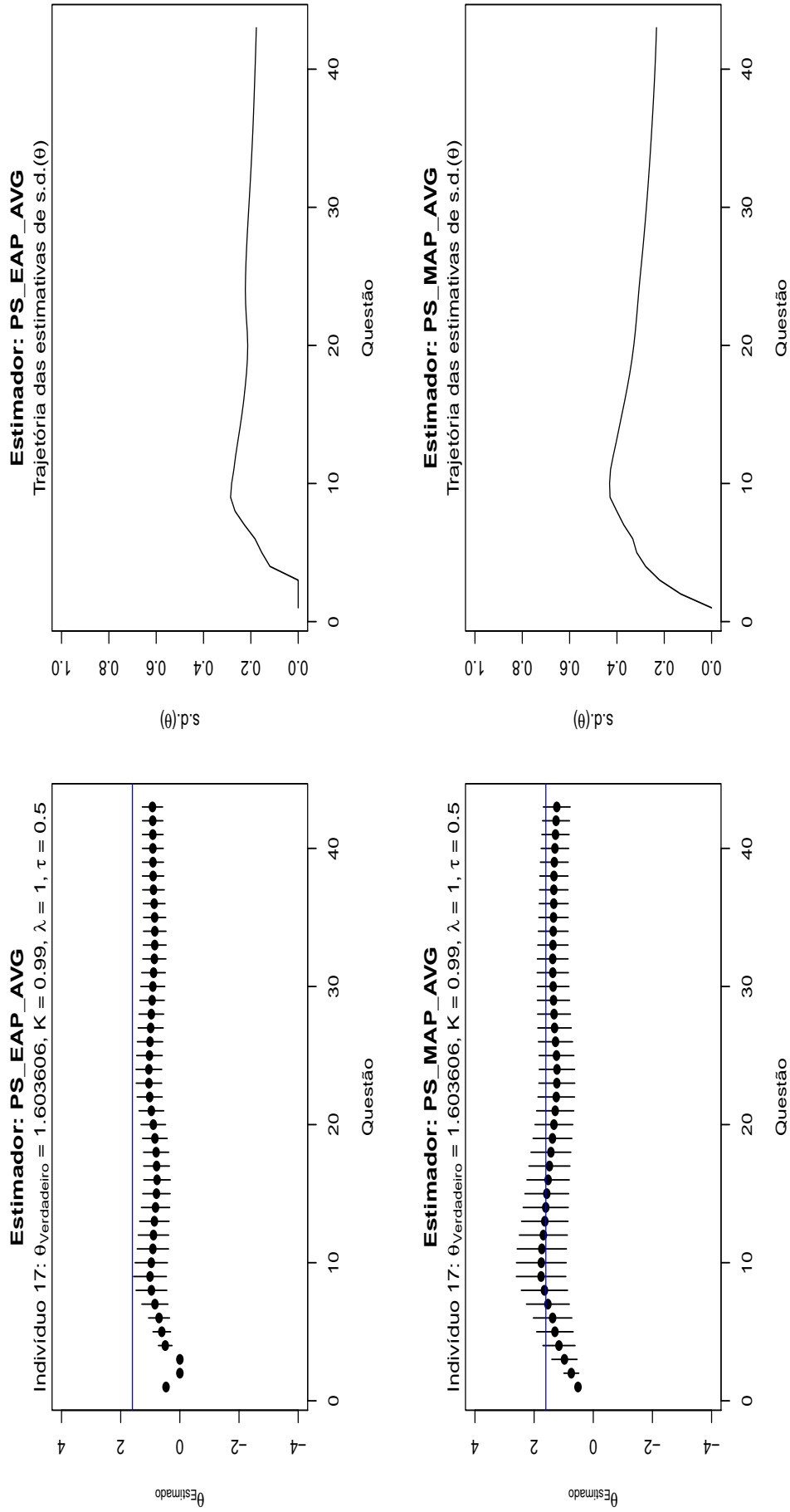
Através de simulações, as trajetórias suavizadas de estimativas de proficiência estimadas, são obtidas por  $\hat{\theta}$  e  $\tilde{\theta}$ , tal como definido nas expressões (5.15) a (5.18), apresentadas nas Figuras (5.22) e (5.23), sendo estas, exemplos dos principais resultados a serem discutidos na seção a seguir.

### 5.1.7 Obtenção de valores iniciais de $\alpha_t^*$ e $\beta_t^*$

Para estudar o impacto das constantes  $\lambda$  e  $\tau$  no modelo dinâmico do *Power Steady* (vide expressão (5.13)), foram simulados 8 cenários de CATs, tomando-se  $\theta \sim N(0, 1)$ , de forma aleatória, ou fixando-se alguns valores, para os quais alguns resultados são discutidos, em que  $k \in \{0.5, 0.65, 0.75, 0.8, 0.85, 0.9, 0.95, 0.99\}$ , a saber:

- **Cenário 5a:**  $\theta$  (fixo)  $\in [-3, 3]$ ,  $\lambda = 1$ ,  $\tau = 0.5$
- **Cenário 5b:**  $\theta$  (aleatório)  $\in [-3, 3]$ ,  $\lambda = 1$ ,  $\tau = 0.5$
- **Cenário 5c:**  $\theta$  (fixo)  $\in [-3, 3]$ ,  $\lambda = 2$ ,  $\tau = 0.5$
- **Cenário 5d:**  $\theta$  (fixo)  $\in [-3, 3]$ ,  $\lambda = 2$ ,  $\tau = 1$
- **Cenário 5e:**  $\theta$  (aleatório)  $\in [-3, 3]$ ,  $\lambda = 2$ ,  $\tau = 0.5$
- **Cenário 5f:**  $\theta$  (fixo)  $\in [-3, 3]$ ,  $\lambda = 3$ ,  $\tau = 0.5$
- **Cenário 5g:**  $\theta$  (fixo)  $\in [-3, 3]$ ,  $\lambda = 3$ ,  $\tau = 1$
- **Cenário 5h:**  $\theta$  (aleatório)  $\in [-3, 3]$ ,  $\lambda = 3$ ,  $\tau = 0.5$

Os melhores resultados obtidos com os oito cenários, acima descrito, representados pelo Cenário 5, tal como se observa na Tabela (5.11), são apresentados nas Figuras (5.26) e (5.27). Os demais resultados, encontram-se no Apêndice C.



**Figura 5.22: Cenário 5b:** Power Steady - Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{\sigma}$ , quando  $k = 0.99, \lambda = 1, \tau = 0.5, \theta(\text{aleatório})$

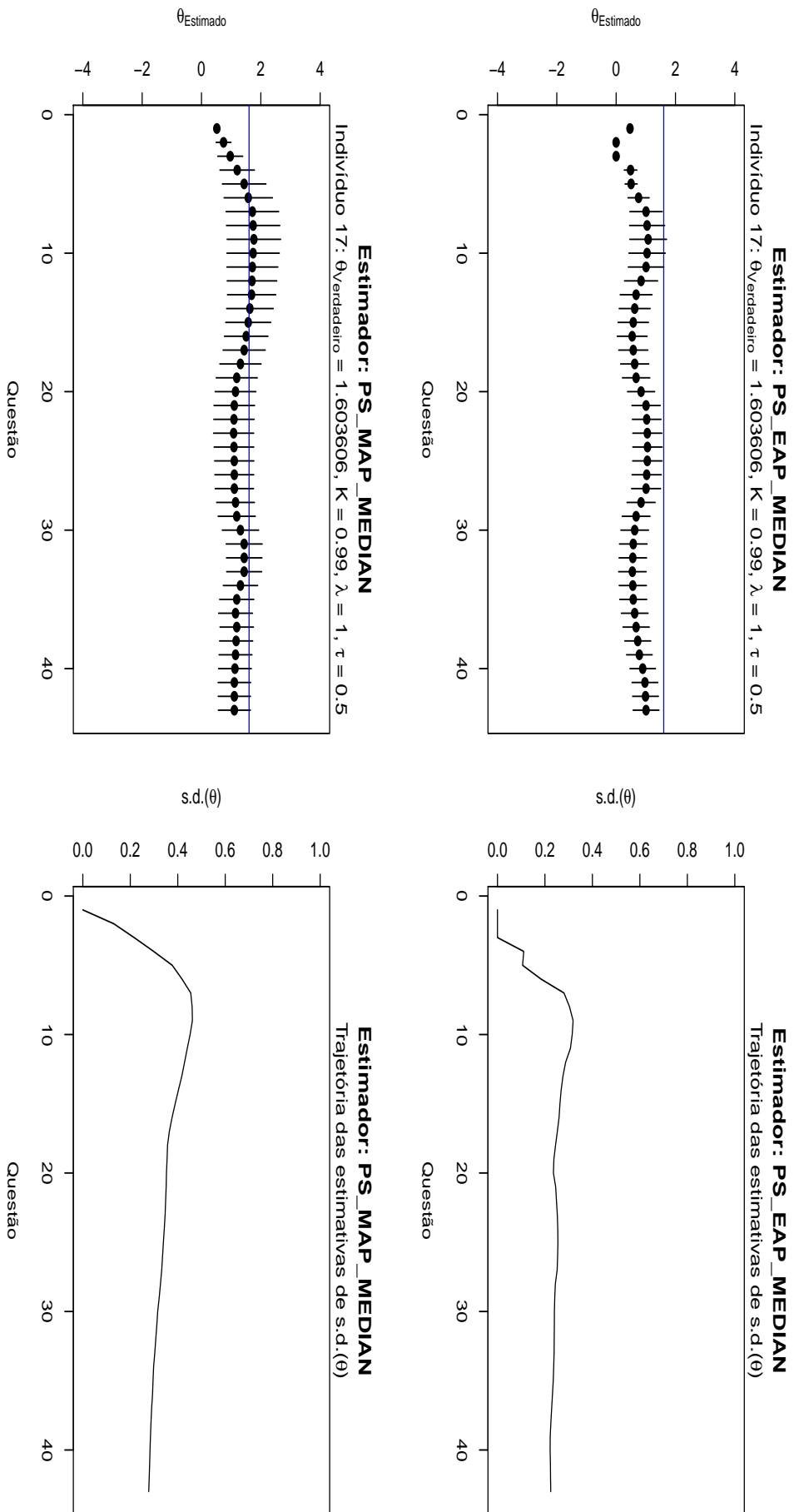


Figura 5.23: Cenário 5b: Power Steady - Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{s.d.}(\theta)$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$ ,  $\theta$  (aleatório)

**Tabela 5.11:** Power Steady - *Comparações por RMSE, quando  $k = 0.99$* 

Estimador	$\theta \sim N(0, 1)$	Cenário	Power Steady		CAT Tradicional
			RMSE <sub>EAP</sub>	RMSE <sub>MAP</sub>	RMSE <sub>EAP/3PL</sub>
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5 ( $\lambda = 1, \tau = 1$ )	0.608637	0.956001	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.275657	0.312662	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			0.321442	0.387185	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5a ( $\lambda = 1, \tau = 0.5$ )	1.133575	1.438154	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.750067	1.304804	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			0.789450	1.318761	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Aleatório	5b ( $\lambda = 1, \tau = 0.5$ )	0.806453	0.834599	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.372060	0.678351	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			0.370497	0.657118	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5c ( $\lambda = 2, \tau = 0.5$ )	1.729390	1.621681	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			1.204145	1.354660	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			1.218218	1.382375	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5d ( $\lambda = 2, \tau = 1$ )	1.342763	1.102676	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.967549	0.980812	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			1.016493	0.978668	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Aleatório	5e ( $\lambda = 2, \tau = 0.5$ )	1.053790	0.790084	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.645747	0.623794	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			0.654511	0.620763	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5f ( $\lambda = 3, \tau = 0.5$ )	1.855128	2.026129	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			1.312158	1.580721	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			1.330259	1.600668	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Fixo	5g ( $\lambda = 3, \tau = 1$ )	1.188951	1.168527	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			1.052088	1.017658	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			1.093236	1.018696	-
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )   EAP_3PL	Aleatório	5h ( $\lambda = 3, \tau = 0.5$ )	0.990353	1.157380	0.1358
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )   PS_MAP_AVG ( $\bar{\hat{\theta}}$ )			0.638180	0.769242	-
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )			0.652242	0.781539	-

Pela Tabela (5.11), observa-se que, com exceção dos Cenários 5e e 5g, em todos os demais cenários realizados, o valor do RMSE obtido com o método EAP é inferior ao obtido com o método MAP, quer sejam considerados  $\alpha_t^*$  e  $\beta_t^*$  (Cenários 5a a 5h), ou  $\alpha_t$  e  $\beta_t$  (Cenário 5). Destacam-se, entretanto, os Cenários 5, 5b, 5e e 5h, sendo estes, os resultados mais favoráveis dentre os oito cenários.

As correções realizadas com a inclusão das constantes  $\lambda$  e  $\tau$ , parecem não ter solucionado o comportamento cíclico dos estimadores PS\_EAP ( $\hat{\theta}$ ) e PS\_MAP ( $\hat{\theta}$ ), para quaisquer dos cenários estudados, tal como ocorre nos Cenários 5 e 5b, como mostram as Figuras (5.24) e (5.25). Quando avaliados em termos dos estimadores PS\_MAP\_AVG ( $\bar{\hat{\theta}}$ ) e PS\_MAP\_MEDIAN ( $\hat{\theta}$ ), para o Cenário 5, tal como apresentado nas Figuras (5.26) e (5.27), produziram melhoras significativas nas estimativas de  $\theta$ , onde uma boa suavização da trajetória das estimativas de proficiência, além de aderência ao valor verdadeiro de  $\theta$ , foram observados a partir da sexta questão, evidenciando, assim, a superioridade dos estimadores PS\_MAP\_AVG ( $\bar{\hat{\theta}}$ ) e PS\_MAP\_MEDIAN ( $\hat{\theta}$ ), em relação ao estimador PS\_MAP ( $\hat{\theta}$ ). O mesmo padrão ocorre com os estimadores PS\_EAP\_AVG ( $\bar{\hat{\theta}}$ ) e PS\_EAP\_MEDIAN ( $\hat{\theta}$ ), em relação ao estimador PS\_EAP ( $\hat{\theta}$ ).

Curioso, no caso dos estimadores PS\_MAP\_AVG ( $\bar{\hat{\theta}}$ ) e PS\_MAP\_MEDIAN ( $\hat{\theta}$ ), é observar que seus desvios-padrão tenham sido expressivamente pequenos, fato este, que embora corrigido, ao se utilizar a constante  $\tau = 0.5$ , faz com que os estimadores apresentem estimativas majoritariamente subestimadas do valor de  $\theta$  verdadeiro, como mostram as Figuras

(5.22) e (5.23) para o Cenário 5b, evidenciando, assim, a superioridade dos estimadores utilizando o método MAP, para o Cenário 5b.

Embora os ajustes dos Cenários 5b, 5e e 5h tenham apresentado bons resultados, todos apresentaram valores de RMSE maiores que os do CAT tradicional. Comparando os resultados da Tabela (5.11) com as Figuras (5.28) a (5.33), onde utilizou-se os pares  $\{(\theta, \hat{\theta}), (\theta, \tilde{\theta}), (\theta, \bar{\theta}), (\theta, \check{\theta}), (\theta, \dot{\theta}), (\theta, \ddot{\theta})\}$  para verificar o quanto os estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , e suas respectivas versões dadas pelas expressões (5.15) a (5.18), se ajustaram ao  $\theta$  verdadeiro, observou-se que, embora o valor de RMSE obtido no Cenário 5b tenha sido menor do que os respectivos valores obtidos nos Cenários 5e e 5h, nenhum dos três cenários obteve melhor desempenho em relação ao Cenário 5, que apresentou a maior correlação entre os valores de  $\theta$  e  $\hat{\theta}$  ( $\rho = 0.942$ ) e entre os valores de  $\theta$  e  $\tilde{\theta}$  ( $\rho = 0.874$ ), dentre todos os cenários estudados.

Portanto, a preferência pelo Cenário 5 é sustentada pelas seguintes conclusões:

- Melhor suavização da trajetória das estimativas de proficiência (Figuras (5.26) e (5.27)).
- Melhor aderência ao valor verdadeiro de  $\theta$  (Figuras (5.26) e (5.27)).
- Baixo valor de RMSE em relação os demais cenários (Tabela (5.11)).
- Alta correlação entre:
  - $\theta$  e  $\check{\theta}$ ,  $R_{\text{Ajustado}}^2 = 0.982$ , (Figura (5.33)).
  - $\theta$  e  $\dot{\theta}$ ,  $R_{\text{Ajustado}}^2 = 0.972$ , (Figura (5.32)).
  - $\theta$  e  $\bar{\theta}$ ,  $R_{\text{Ajustado}}^2 = 0.986$ , (Figura (5.31)).
  - $\theta$  e  $\tilde{\theta}$ ,  $R_{\text{Ajustado}}^2 = 0.979$ , (Figura (5.30)).

Entre os oito cenários simulados, além do destaque para os Cenários 5b, 5e e 5h, discutidos acima, observou-se, entre os cenários restantes, que o desempenho dos Cenários 5c e 5f foram os piores entre todos os demais. Na Tabela (5.11), apresentam os RMSE mais altos de todas as simulações utilizando-se as expressões (5.15) a (5.18). Entre as Figuras (5.28) a (5.33) sempre observa-se pontos muito distantes da reta do ajuste, implicando em estimativas de proficiência díspares de seus valores verdadeiros

De maneira geral, o modelo *Power Steady* apresenta estimativas de proficiência cíclicas, tal como apresentado pelas Figuras (5.8), (5.16), (5.21), (5.24) e (5.25) devido à forte dependência de seus parâmetros  $\alpha$  e  $\beta$ , apresentado nas figuras (5.5), (5.10), (5.11), e também pelas tabelas (5.5), (5.6), (5.9) e (5.10), onde, claramente, observa-se que estes parâmetros tendem a assumir valores altos, a depender das escolhas iniciais dos mesmos.

As tentativas de corrigir tais comportamentos, discutidas nesta dissertação, evidenciaram casos com ajustes razoáveis para alguns dos cenários simulados, apresentados nas figuras (5.22), (5.23) e (5.32) e Tabela (5.11). Apesar destes resultados, alguns cenários apresentaram situações que necessitam maiores investigações (vide figuras (5.26) e (5.27)) pois, apesar da boa aderência ao valor verdadeiro de  $\theta$  pelas estimativas de  $\theta$  obtidas com o uso das expressões (5.15) a (5.18), para o Cenário 5, com boa estabilização das trajetórias, as amplitudes dos intervalos de confiança mostraram-se pouco variáveis, quando o ideal é que esses intervalos contivessem  $\theta$  e fossem sucessivamente mais estreitos. Assim, os ajustes realizados utilizando-se os parâmetros  $\alpha^*$  e  $\beta^*$ , definidos pela expressão (5.13), juntamente com as médias e medianas dos estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , suavizados tal como indicado pelas expressões



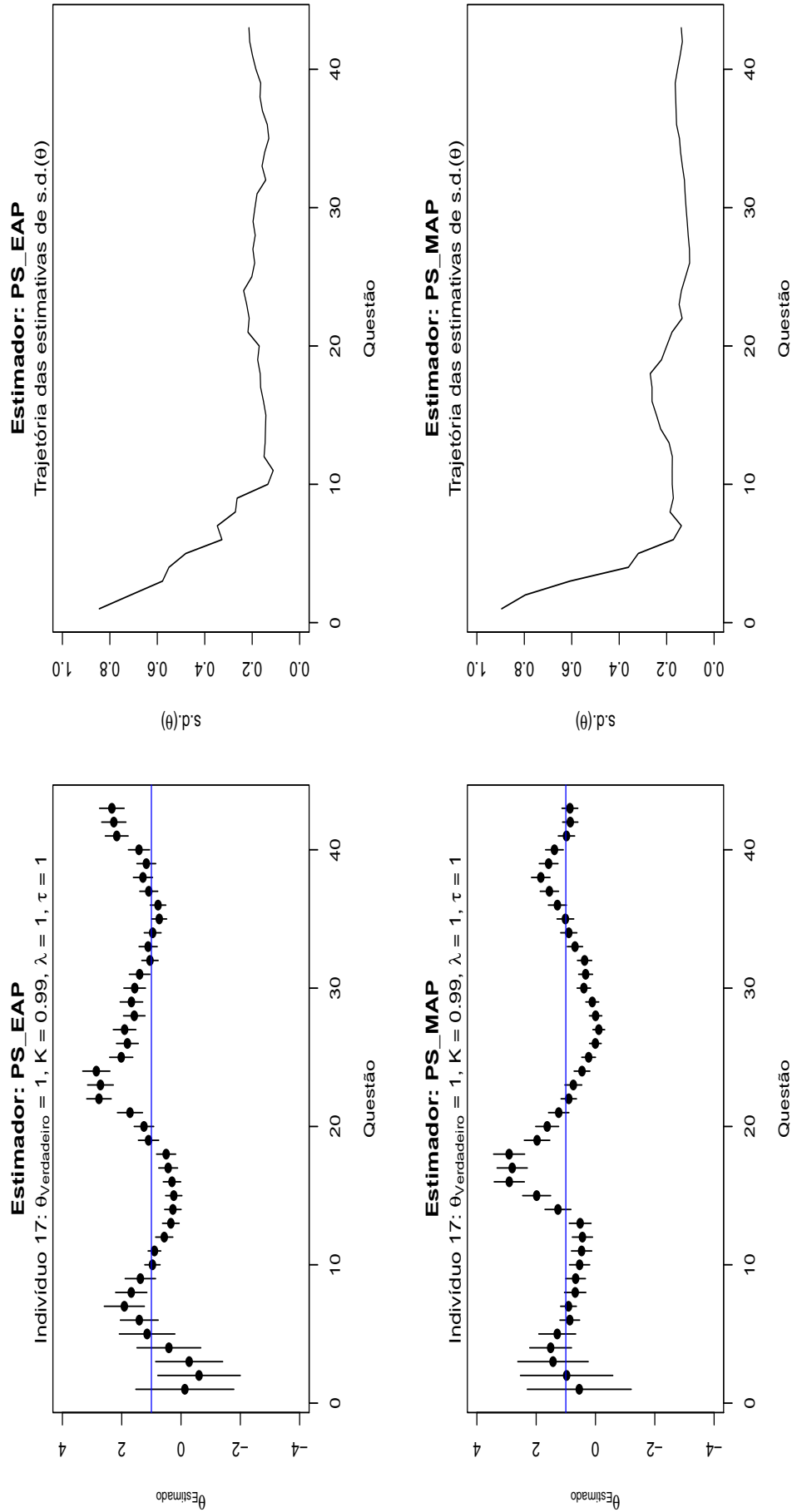


Figura 5.24: *Cenário 5*: Power Steady - Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{\theta}$ , quando  $k = 0.99, \lambda = 1, \tau = 1, \theta(fixo)$

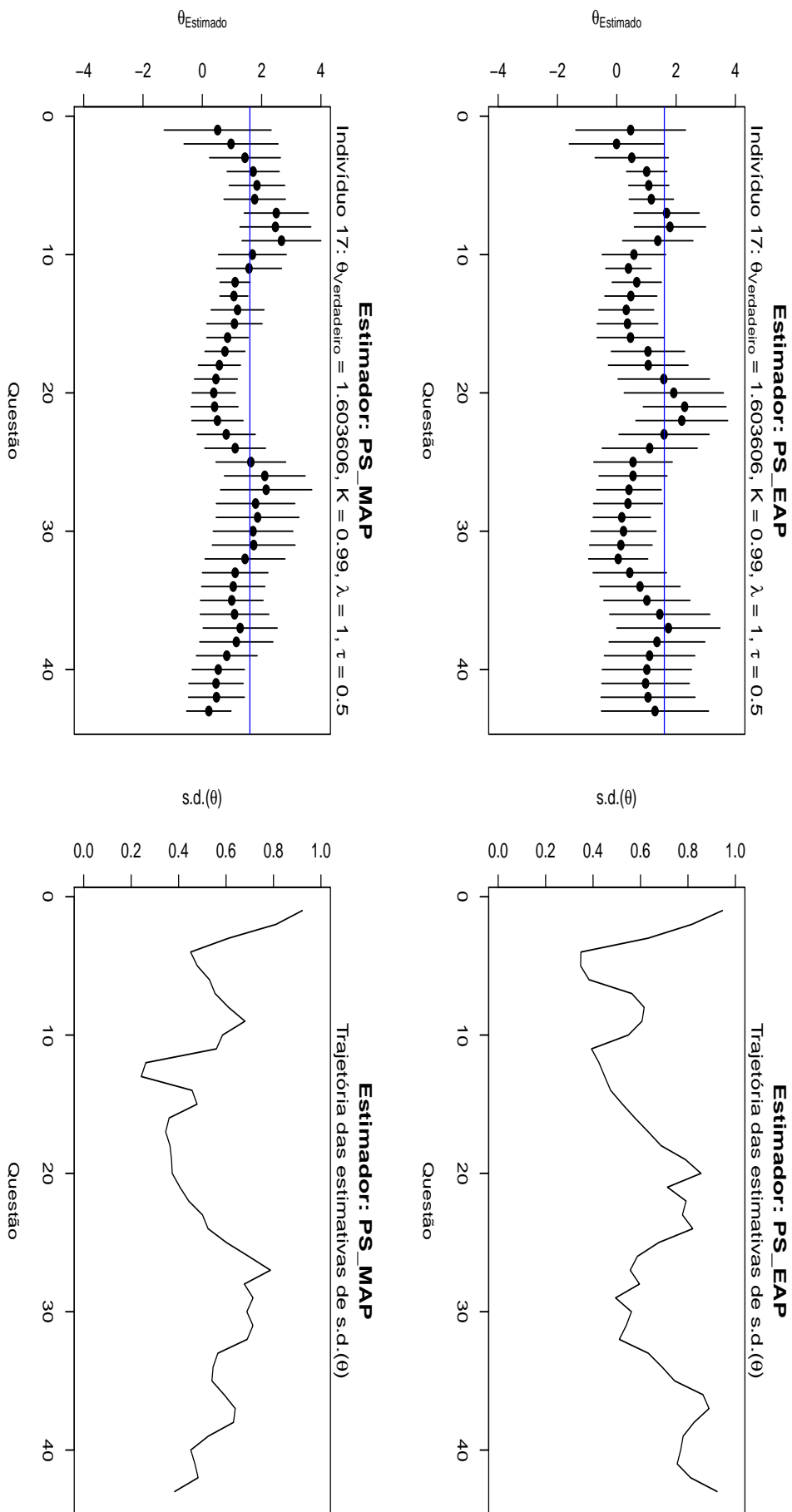


Figura 5.25: Cenário 5b: Power Steady - Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{\sigma}$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$ ,  $\theta$  (aleatório)

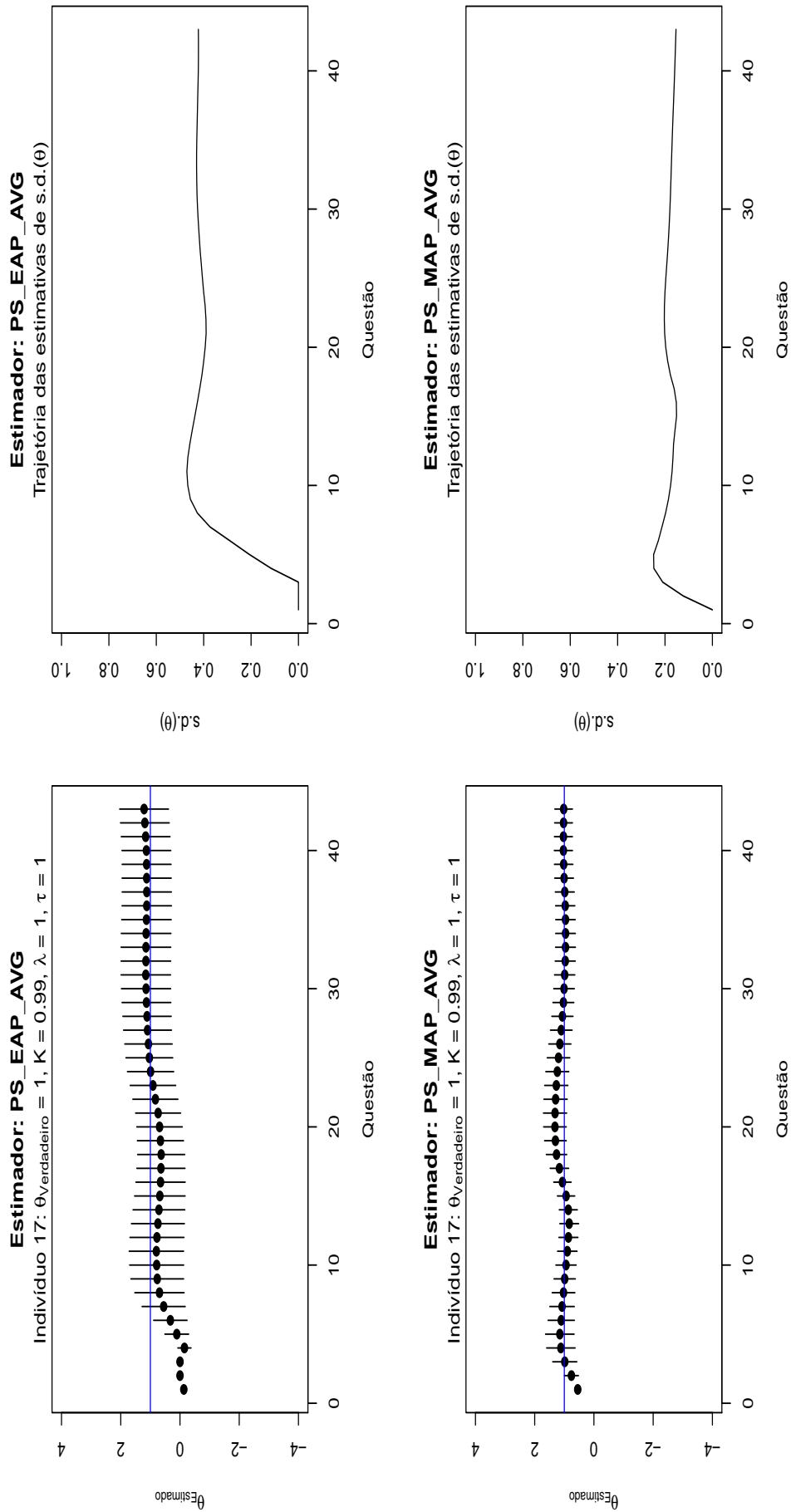


Figura 5.26: Cenário 5: Power Steady Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{\sigma}$ , quando  $k = 0.99, \lambda = 1, \tau = 1, \theta(\text{fixo})$

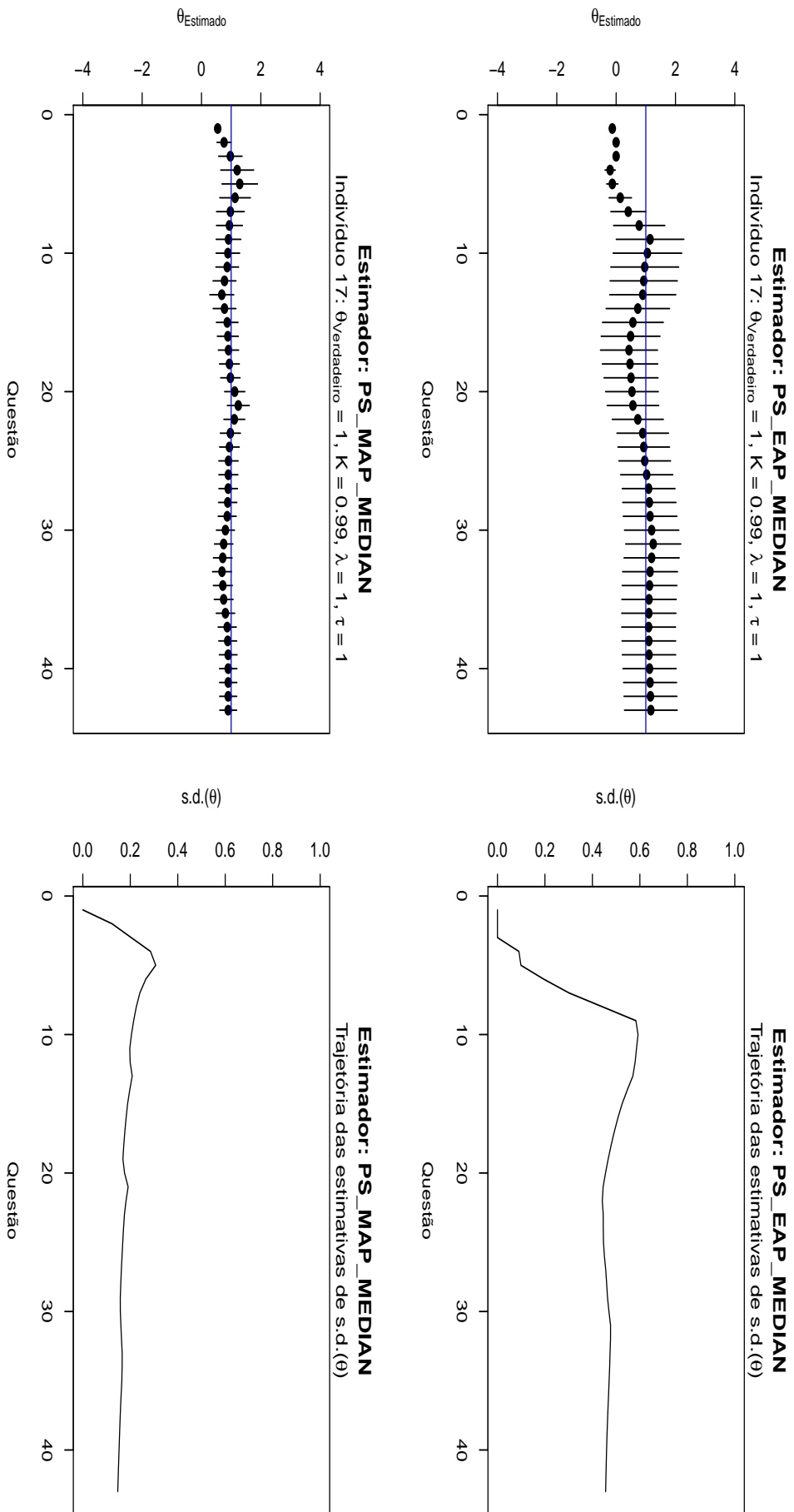
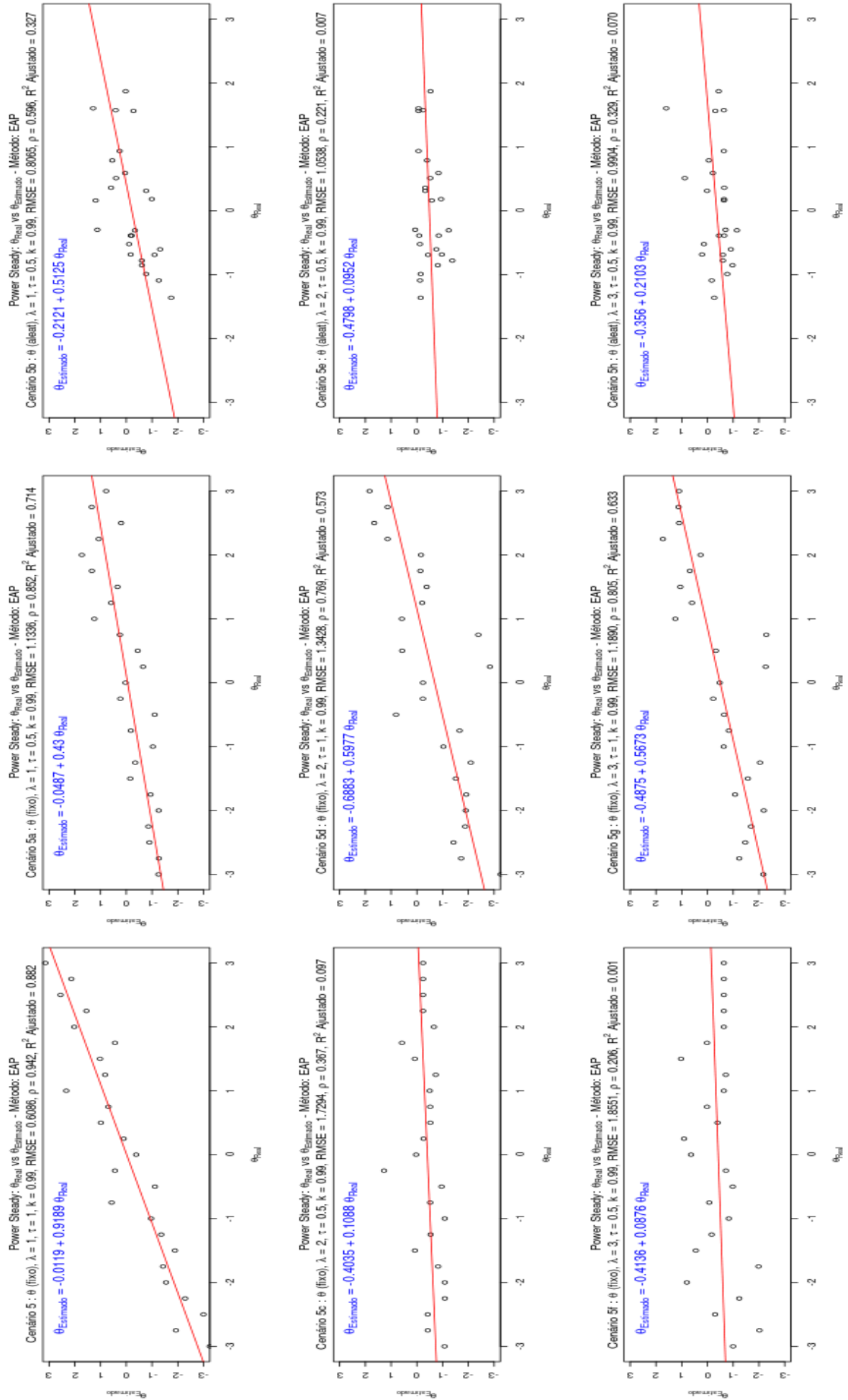


Figura 5.27: Cenário 5: Power Steady - Trajetória dos estimadores  $\hat{\theta}$  e  $\hat{s.d.}(\theta)$ , quando  $k = 0.99, \lambda = 1, \tau = 1, \theta(\text{fixo})$



**Figura 5.28:** Power Steady - Método EAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$

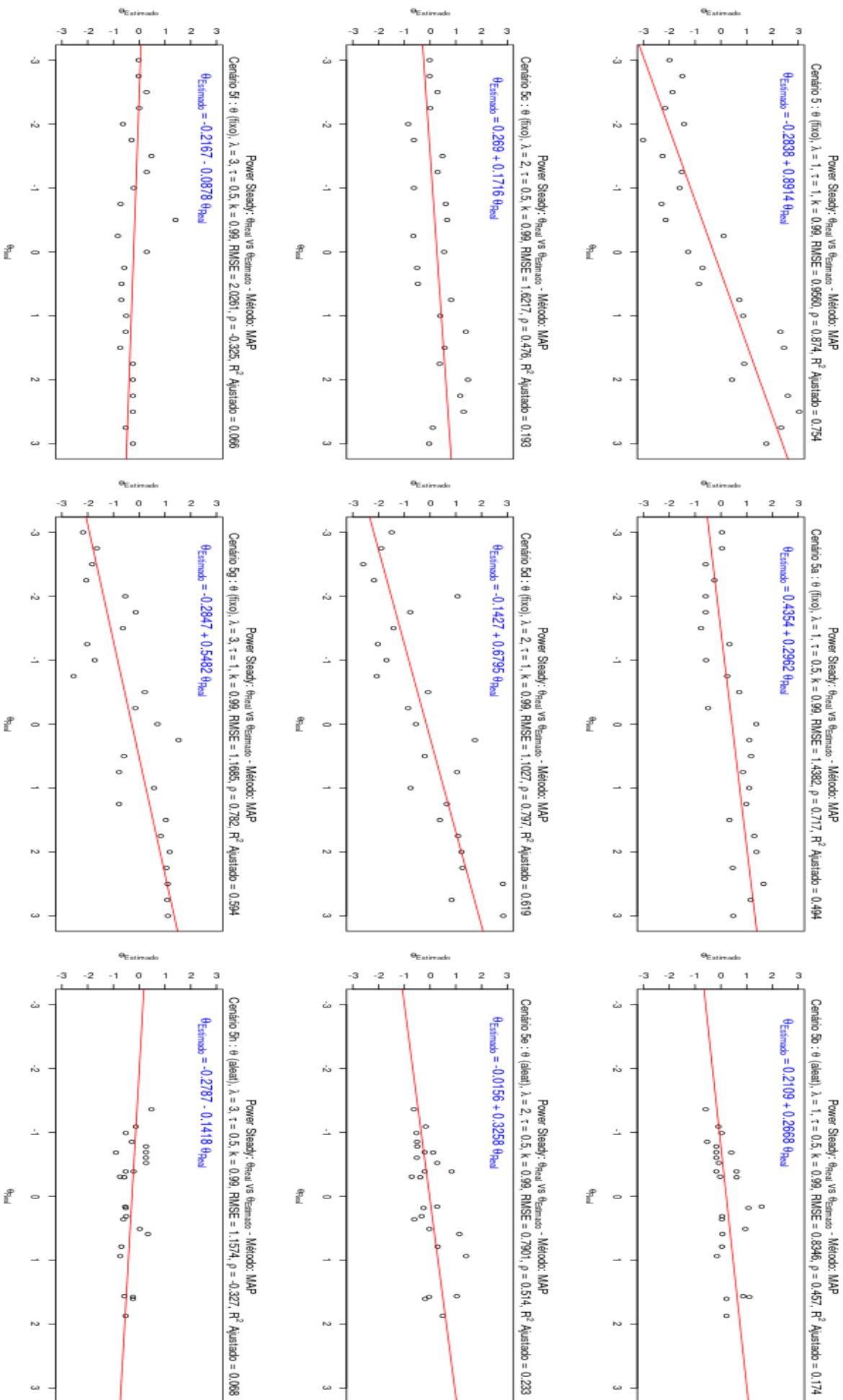
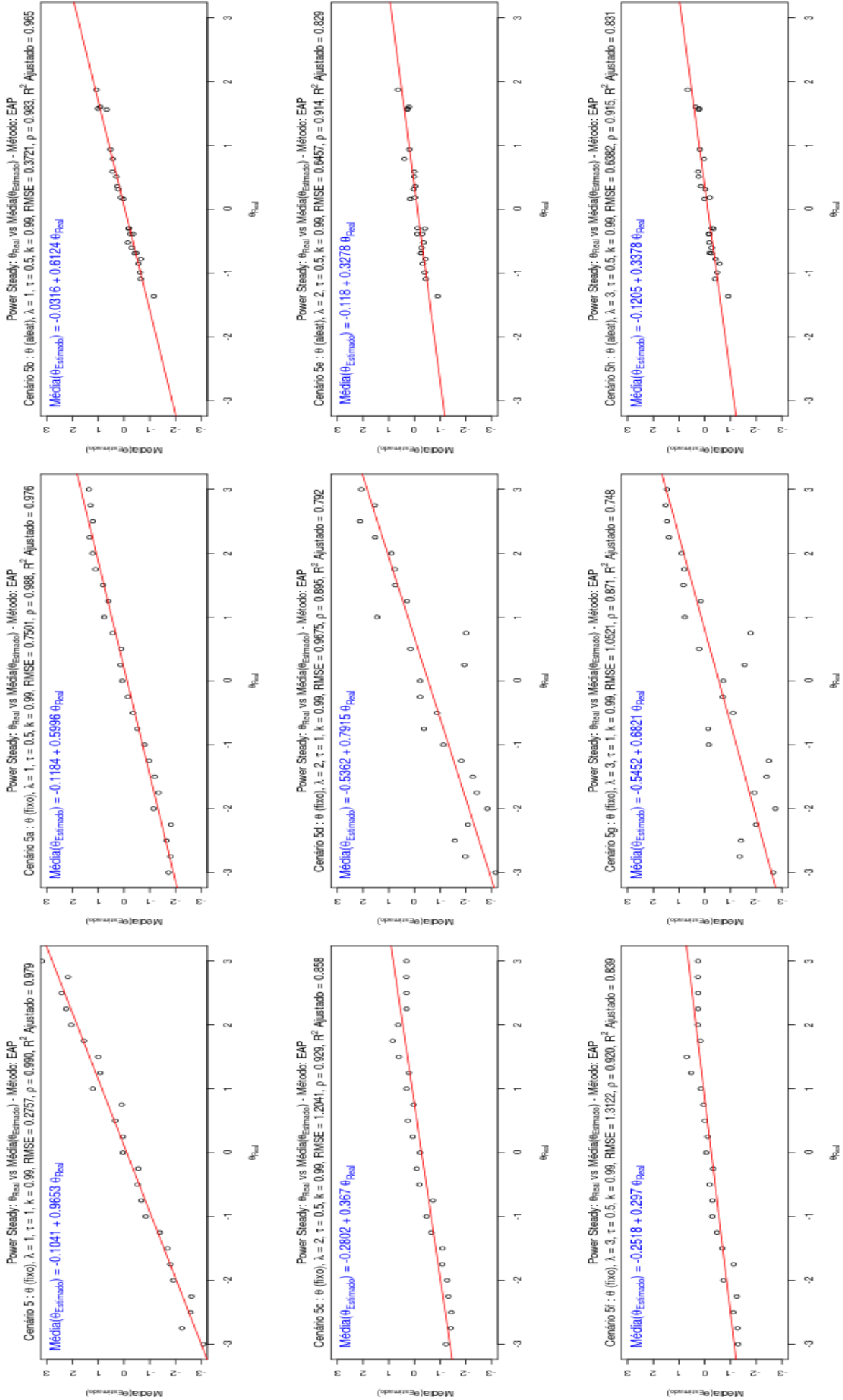


Figura 5.29: Power Steady - Método MAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$



**Figura 5.30:** Power Steady - Método EAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$ .

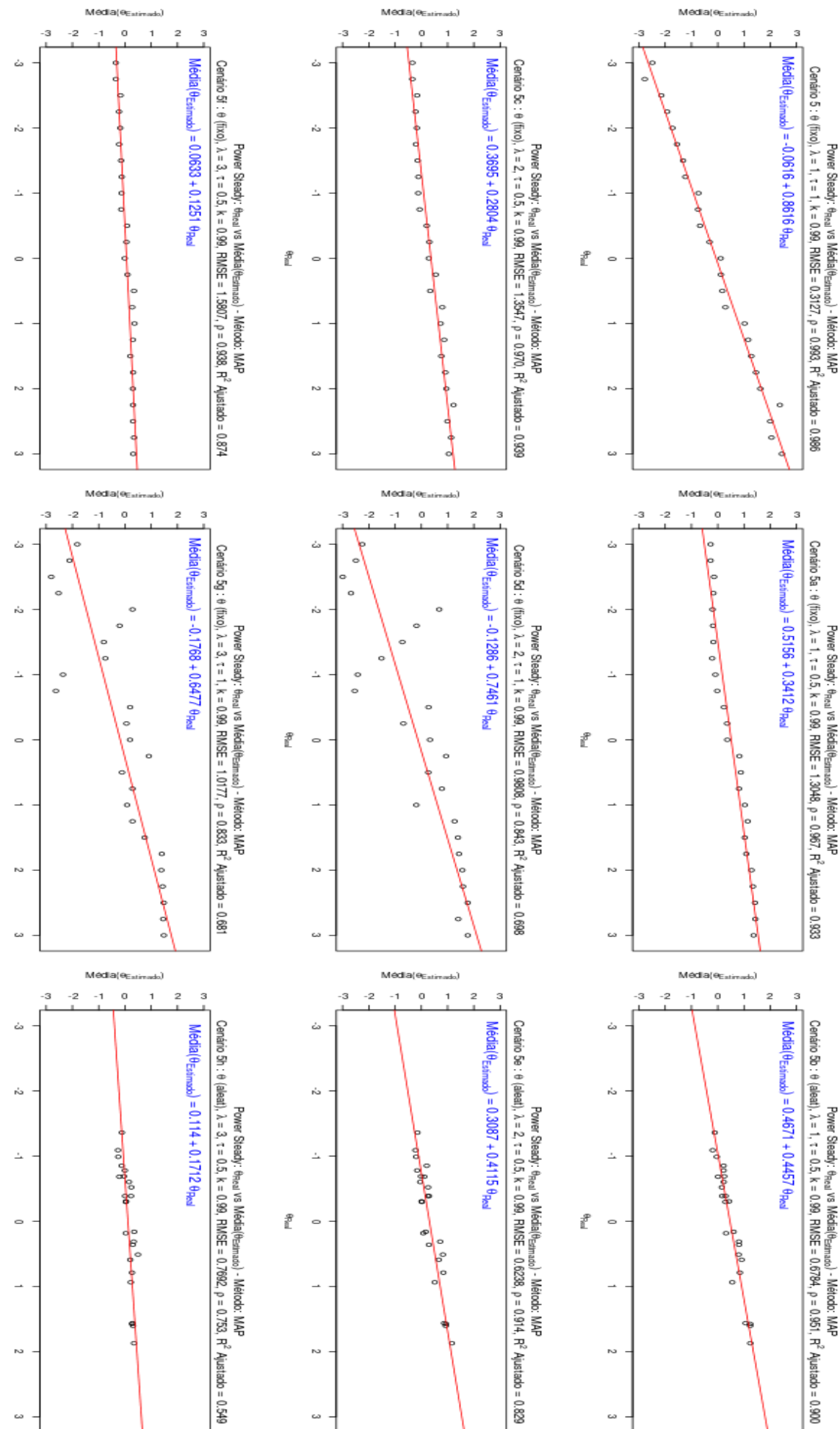
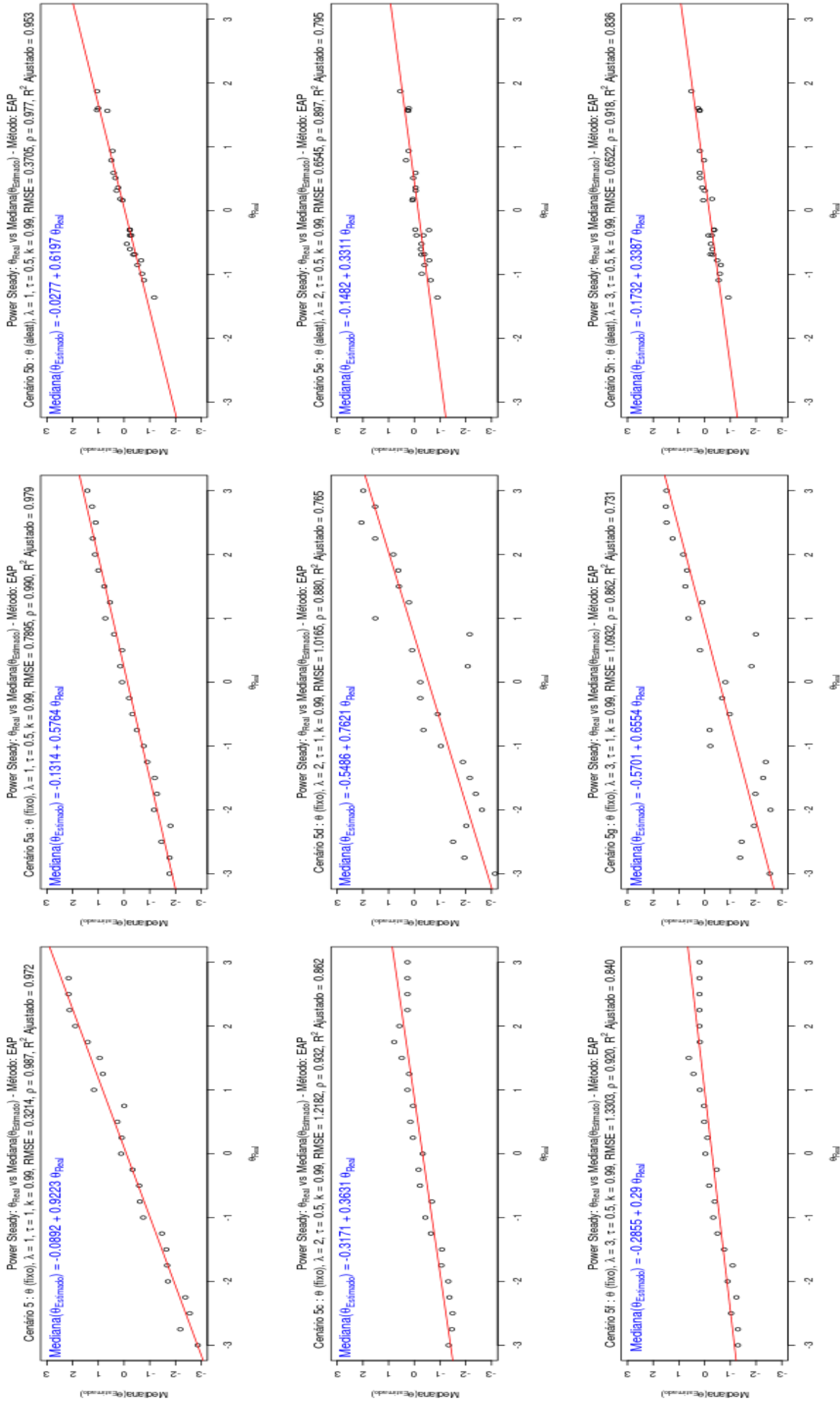


Figura 5.31: Power Steady - Método MAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$





**Figura 5.32:** Power Steady - Método EAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$

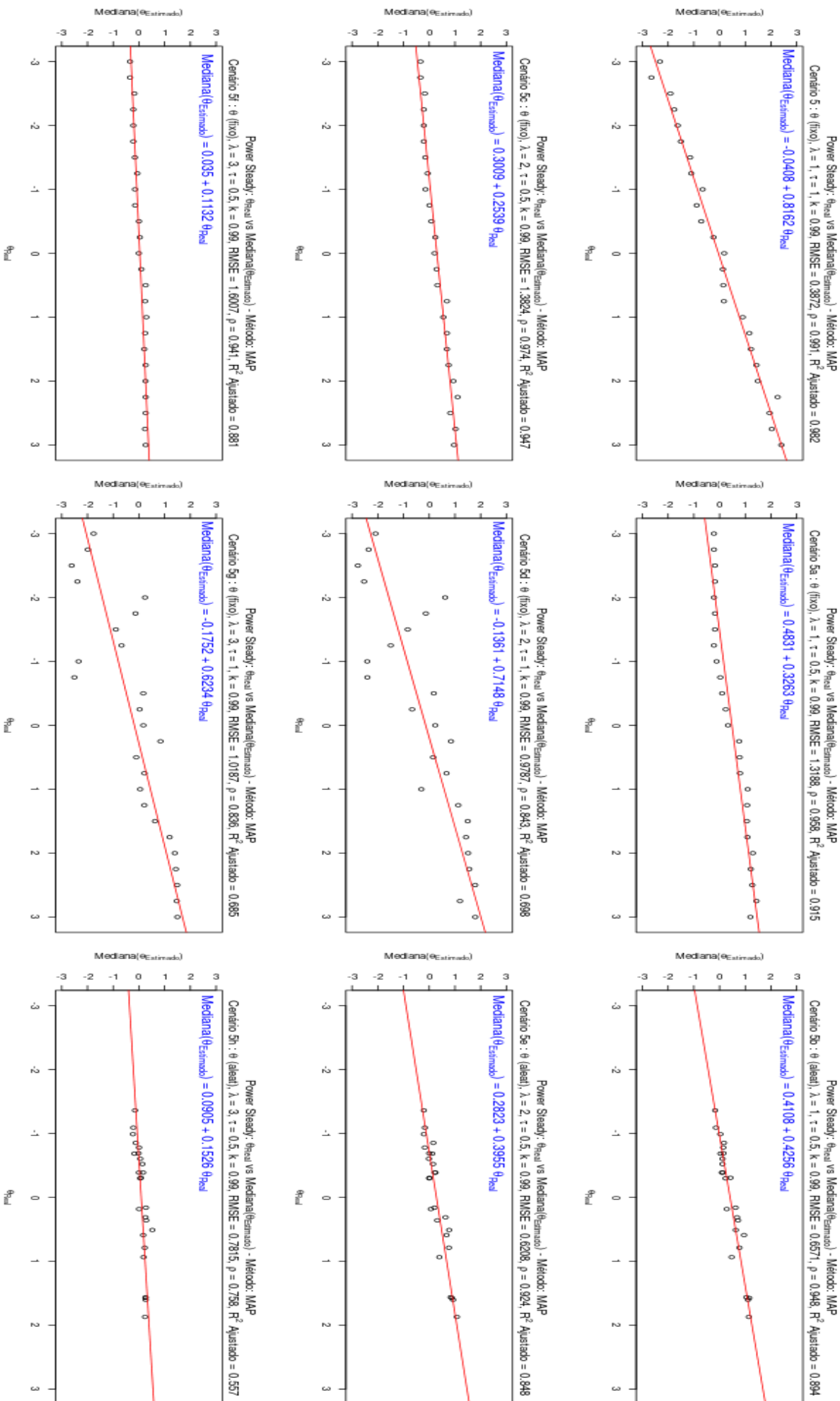


Figura 5.33: Power Steady - Método MAP: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $k = 0.99$

(5.15) a (5.18), permitiram encontrar no Cenário 5, um equilíbrio para adequação dos dados ao modelo *Power Steady*, culminando nos resultados apresentados pelas Figuras (5.26), (5.27) e (5.30).

Portanto, conclui-se que as tentativas de suavizar a trajetória das estimativas de proficiência, utilizando-se o modelo *Power Steady*, ainda que nem sempre satisfatórias para a estimação das proficiências da Prova de Inglês Instrumental I, sejam válidas enquanto metodologia a ser mais amplamente pesquisada. Talvez, até mesmo em outras provas de proficiência da língua Inglesa que possuam um maior número de itens disponíveis, onde a distribuição dos parâmetros dos itens, em particular a do parâmetro  $b$ , apresente maior variabilidade, permitindo, assim, que o modelo escolha itens mais aderentes à proficiência dos indivíduos submetidos ao CAT. De maneira que, o impacto da escolha inicial dos parâmetros  $\alpha$  e  $\beta$ , dada a relação direta dos mesmos com os parâmetros dos itens, permita um estudo mais aprofundado e preciso da aplicação do modelo e do alcance de sua acurácia.

## 5.2 Simulações com o Modelo *Filtro de Kalman*

Nesta seção são tratadas as simulações geradas para o modelo *Filtro de Kalman*, cuja construção metodológica, apresentada na Seção 4.5.6, é utilizada como um caso mais geral dos modelos dinâmicos tratados nesta dissertação, uma vez que o modelo *Power Steady*, para os quais foram discutidos os resultados na Seção 5.1, não incorpora no modelo, o parâmetro  $c_t$  de acerto casual.

No modelo *Filtro de Kalman* considera-se, com base nas expressões (4.87) e (4.88), que o processo *Bernoulli*  $\{\mathbf{X}_t\}$ , no contexto do CAT, é estacionário, de maneira que tem-se

$$f(\eta_{t+1}|X_{t+1}) = \frac{1}{B(\alpha_{t+1|t} + 1 - x_{t+1}, \beta_{t+1|t} + x_{t+1})} \left[ \frac{1 + c_t e^{-\eta_{t+1}}}{1 + \frac{c_t \alpha_{t+1|t}}{\beta_{t+1|t}}} \right]^{x_{t+1}} \times \\ \times \frac{e^{-(\alpha_{t+1|t} + 1 - x_{t+1})\eta_{t+1}}}{[1 + e^{-\eta_{t+1}}]^{\alpha_{t+1|t} + \beta_{t+1|t} + 1}}, \quad (5.19)$$

$$P(\eta_{t+1}|X_t) \approx P(\eta_t|X_t), \quad (5.20)$$

sendo as equações de transição, apresentadas em (4.101), dadas por

$$\begin{cases} \alpha_{t+1} = \phi \alpha_t + 1 - x_{t+1} \\ \beta_{t+1} = \phi \beta_t + x_{t+1} \end{cases}, \quad (5.21)$$

onde  $\phi \in (0, 1)$  funciona como um fator de desconto, sendo parâmetro implícito da função de log-verossimilhança, apresentada em (4.99), e expressa por

$$l(\phi) = \sum_{t=1}^{n-1} \ln[P(X_{t+1}|X_t)].$$

Embora a escolha de  $\phi$  seja feita através de um processo de maximização, nesta dissertação, as simulações executadas calculam a proficiência  $\hat{\theta}$  utilizando o Algoritmo (1), da Seção 4.5.6, onde são considerados valores fixos de  $\phi$  para manter a comparabilidade com os resultados do modelo *Power Steady*. Também são considerados os mesmos cenários apresentados na Seção 5.1.7.

As simulações realizadas para o modelo *Filtro de Kalman* diferenciam-se das simulações do modelo *Power Steady*, além dos aspectos metodológicos, principalmente no uso do modelo 3PL e das equações de atualização (vide expressões (4.101) e (4.150)), para cálculo das probabilidades de acerto, com ajuste para uma ogiva normal ( $D = 1.702$ ), mantendo a comunalidade quanto ao uso dos parâmetros já calibrados provenientes do banco de itens da Prova de Inglês Instrumental I.

### 5.2.1 Avaliando o comportamento de $\alpha_t$ $\beta_t$

Os resultados das simulações com o modelo *Power Steady* evidenciaram o alto impacto dos valores de  $\alpha_t$  e  $\beta_t$  na estimação da proficiência. Portanto, considerando os conhecimentos adquiridos com o modelo *Power Steady*, as simulações executadas levaram em consideração os seguintes aspectos:

- Utilizou-se 5 estimativas iniciais de  $\theta_{jt}$ , ao invés de 3.

- Para a obtenção das estimativas iniciais de  $\theta_{jt}$  utilizou-se a função “*fscorés*” do pacote “*mirtCAT*”.
- Considerou-se as modificações nas “*equações de atualização*” descritas em (4.101), e expressas por

$$\begin{cases} \alpha_{t+1}^* = [\lambda(\phi\alpha_t + 1 - x_{t+1})]^\tau \\ \beta_{t+1}^* = [\lambda(\phi\beta_t + x_{t+1})]^\tau \end{cases}, \quad (5.22)$$

a fim de se manter a comparabilidade com as simulações realizadas com o modelo *Power Steady*.

Como dito anteriormente, para o modelo *Filtro de Kalman* não foram estudados, nesta dissertação, métodos de otimização. Porém, foram realizadas simulações para o conjunto de  $\phi \in \{0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.99\}$ , no intuito de observar o comportamento do modelo para diferentes fatores de desconto, sendo resultados apresentados, preferencialmente para  $\phi = 0.99$ , em razão da comparabilidade com os dados apresentados para o modelo *Power Steady*.

As simulações utilizando o Cenário 5 ( $\lambda = 1, \tau = 1$ ), mostraram que o modelo *Filtro de Kalman* apresenta comportamento sistemático em relação aos parâmetros  $\alpha_t^*$  e  $\beta_t^*$ . Apesar da diferença entre os padrões observados nos valores das estimativas de  $\alpha_t^*$  e  $\beta_t^*$ , tal como apresentadas, respectivamente, pelas tabelas (5.12) e (5.13), para  $\phi = 0.99$ , e pelas tabelas (5.14) e (5.15), para  $\phi = 0.65$ , o comportamento das estimativas de proficiência tendem a ser muito semelhantes, de acordo com o valor do  $\theta$  verdadeiro. As figuras (5.34), (5.35), (5.37) e (5.38) apresentam resultados que evidenciam tal característica, revelando, ainda, que o modelo *Filtro de Kalman* tende à superestimação das proficiências quando  $\theta < 0$ , e à subestimação das mesmas quando  $\theta > 0$ . Em ambos os casos, houve uma suavização da trajetória das estimativas de proficiência, sendo observados intervalos de confiança maiores do que, em geral, foram apresentados pelo modelo *Power Steady*, ao longo da trajetória do CAT. O vício sistemático das estimativas de proficiência é observado para todos os valores de  $\phi$  do grid estudado.

**Tabela 5.12:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $\phi = 0.99$ ,  $\lambda = 1$ ,  $\tau = 1$ , com 5 estimativas provisórias de  $\theta_{jt}$  (Filtro de Kalman - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	11.8070	12.3580	1.4640	23.4430	6.2270	17.6820
-2.75	11.8960	12.2840	1.4420	23.3800	6.1440	17.7890
-2.50	11.2920	11.5570	1.4730	21.6670	5.3310	15.9520
-2.25	10.8630	9.7340	1.4810	22.5400	4.3120	17.0960
-2.00	11.0800	11.2740	1.4810	21.7720	4.3250	16.9520
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	5.9330	5.9610	2.9440	9.6130	4.8750	6.7840
2.25	4.4320	3.3090	2.3660	8.9040	2.6170	5.8350
2.50	4.9600	4.0780	2.5180	11.8100	2.8120	6.6160
2.75	4.2850	4.0760	2.7120	6.4790	3.3320	5.0360
3.00	2.6070	2.5590	2.1570	3.1360	2.4100	2.8080

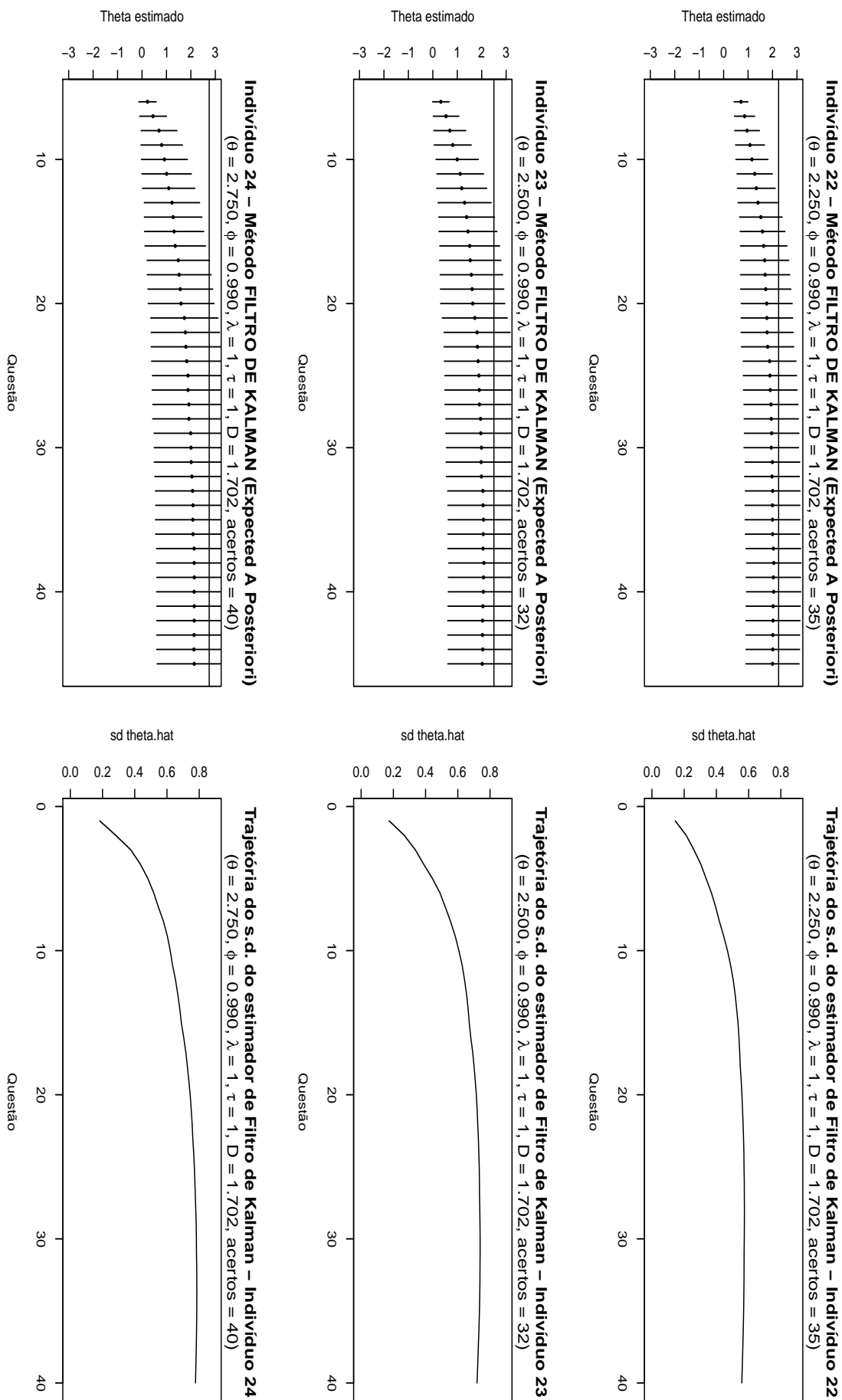


Figura 5.34: Cenário 5: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.99, \lambda = 1, \tau = 1, \theta > 0$

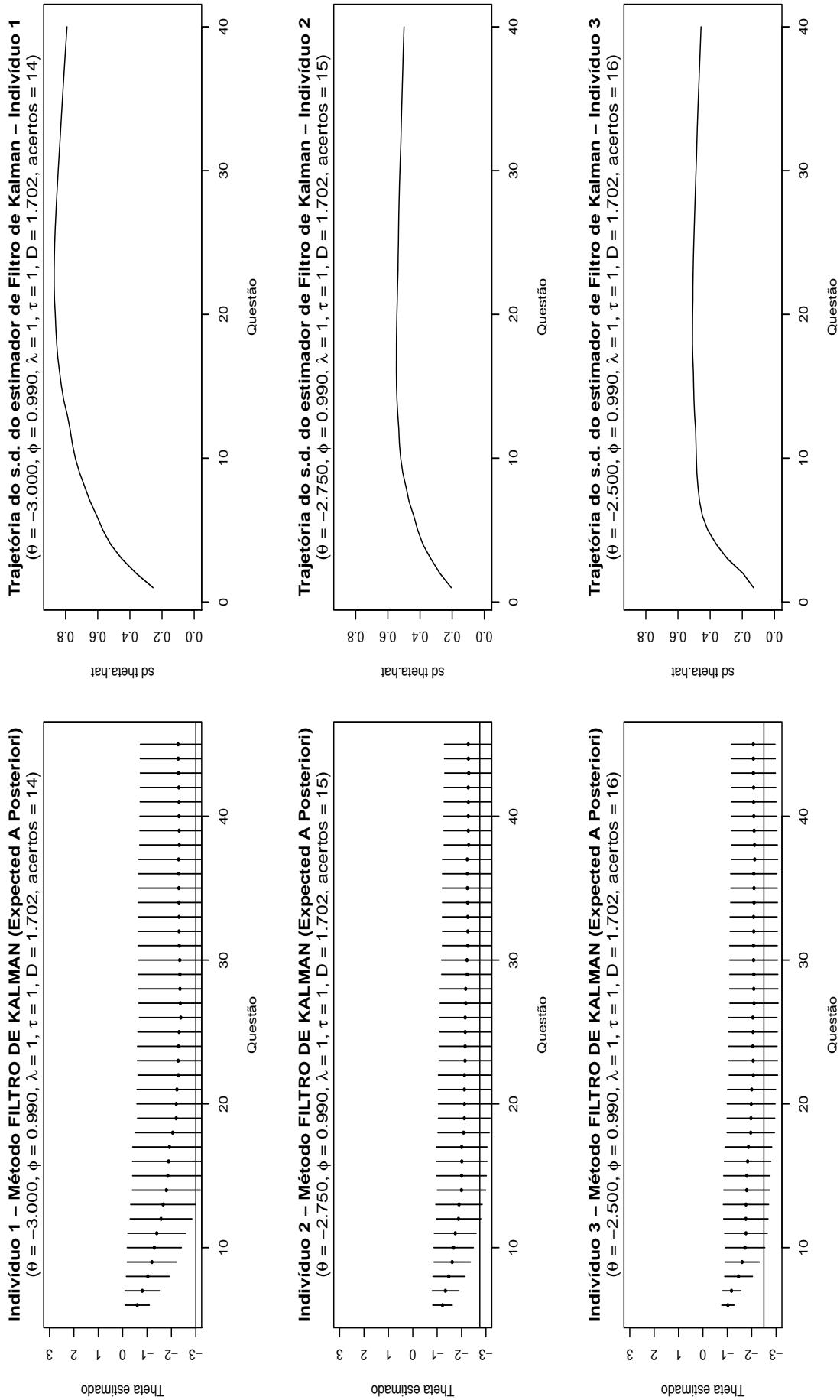


Figura 5.35: *Cenário 5: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.99, \lambda = 1, \tau = 1, \theta < 0$*

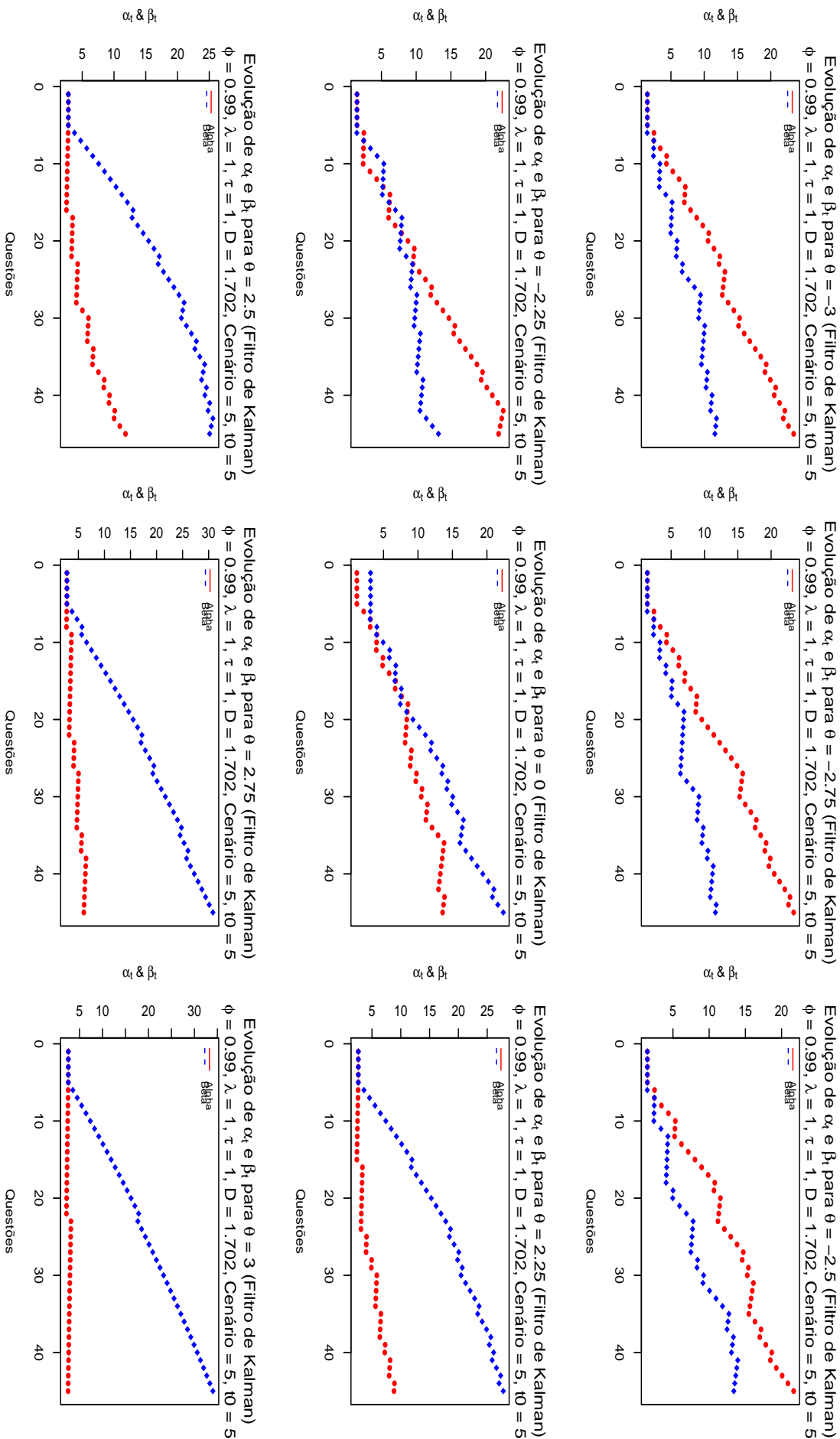


Figura 5.36: Cenário 5: Filtro de Kalman - Trajetória dos  $\alpha_i^*$  e  $\beta_i^*$ , quando  $\phi = 0.99, \lambda = 1, \tau = 1, \theta$  (fixo)



**Tabela 5.13:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $\phi = 0.99$ ,  $\lambda = 1$ ,  $\tau = 1$ , com 5 estimativas provisórias de  $\theta_{jt}$  (Filtro de Kalman - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	6.704	6.681	1.439	11.847	3.319	9.880
-2.75	6.579	6.662	1.417	11.762	3.332	9.623
-2.50	7.235	7.478	1.448	13.949	4.053	11.842
-2.25	7.677	9.190	1.456	13.207	5.233	10.288
-2.00	7.460	7.624	1.456	13.440	5.055	10.536
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	15.282	15.466	3.074	27.875	7.612	22.888
2.25	16.032	17.809	2.627	27.795	9.242	23.359
2.50	15.831	17.133	2.822	25.570	9.424	22.971
2.75	16.478	17.193	2.805	30.810	8.438	24.526
3.00	17.761	17.872	2.569	34.020	9.188	26.307

De modo geral, as estimativas geradas pelo modelo *Power Steady* tiveram menor vício para valores altos do fator de desconto, enquanto o modelo *Filtro de Kalman*, apesar do vício sistemático observado, apresentou melhores resultados com valores menores do fator de desconto, tal como pode ser observado nas figuras (5.37) e (5.38). No caso do modelo *Filtro de Kalman*, esta situação precisa ser melhor investigada, tanto quanto à correção do vício, quanto à otimização do fator de desconto, utilizando-se algoritmos apropriados, a serem desenvolvidos, para testar cenários que considerem calibrações do fator de desconto durante o processamento do CAT, bem como cenários que considerem a otimização a partir do resultado obtido com as proficiências estimadas pelo CAT, segundo os modelos *Power Steady* e *Filtro de Kalman*.

**Tabela 5.14:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $\phi = 0.65$ ,  $\lambda = 1$ ,  $\tau = 1$ , com 5 estimativas provisórias de  $\theta_{jt}$  (Filtro de Kalman - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	1.692500	1.664000	0.752400	2.481400	1.304500	2.021600
-2.75	1.646400	1.631300	0.823700	2.400700	1.290000	1.949500
-2.50	1.597800	1.508000	0.738500	2.439900	1.310600	1.912000
-2.25	1.586000	1.562100	0.576600	2.410500	1.312000	1.893600
-2.00	1.519000	1.457000	0.761000	2.241000	1.232000	1.823000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	0.991300	0.850100	0.172100	2.012000	0.489400	1.307800
2.25	0.858800	0.779600	0.139200	1.718000	0.471800	1.199300
2.50	0.957220	0.883460	0.058830	2.083960	0.507010	1.359160
2.75	0.721480	0.560900	0.060280	1.835080	0.243590	1.043480
3.00	0.492185	0.194982	0.005873	1.680454	0.065336	0.690586

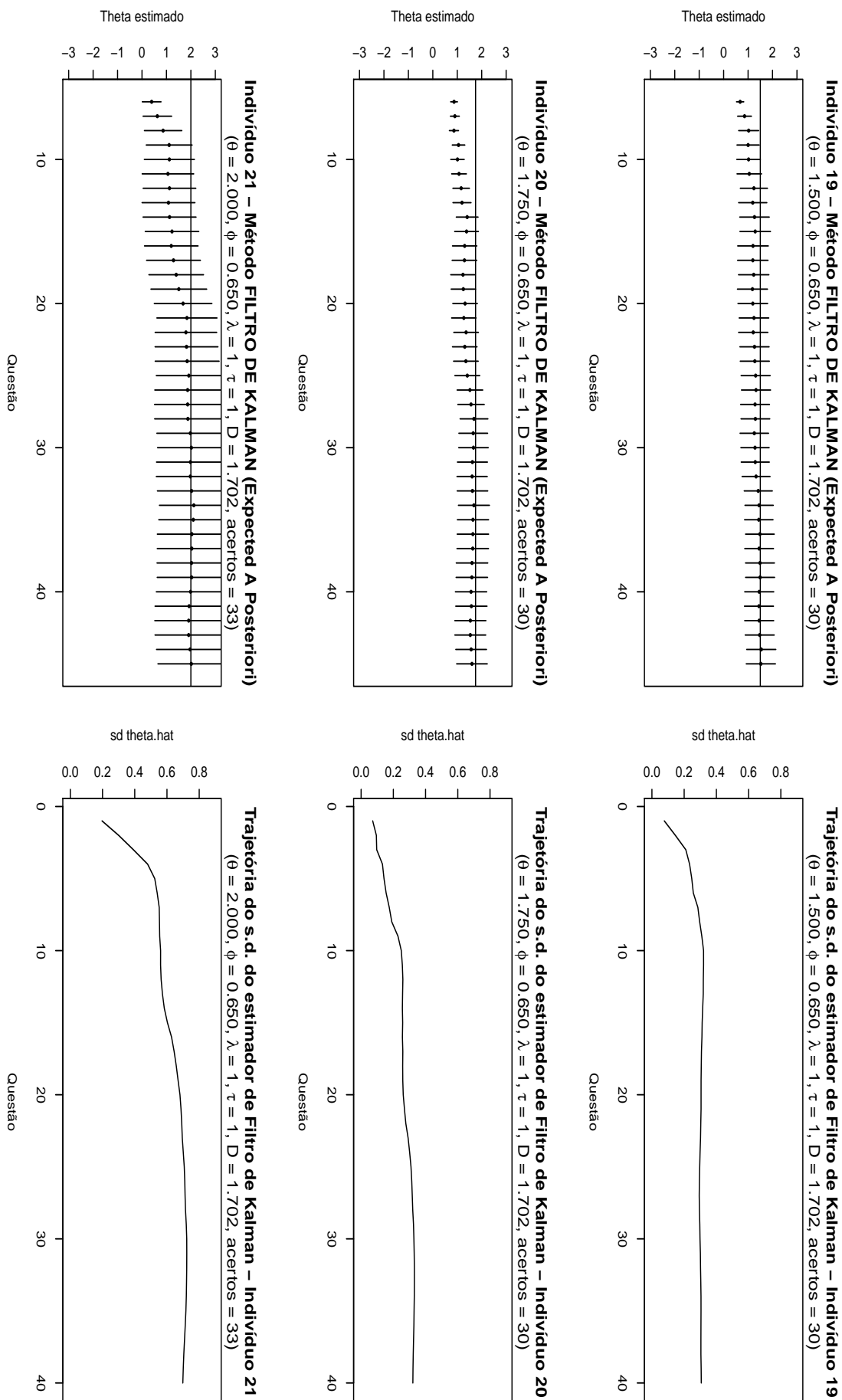


Figura 5.37: Cenário 5: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65, \lambda = 1, \tau = 1, \theta > 0$

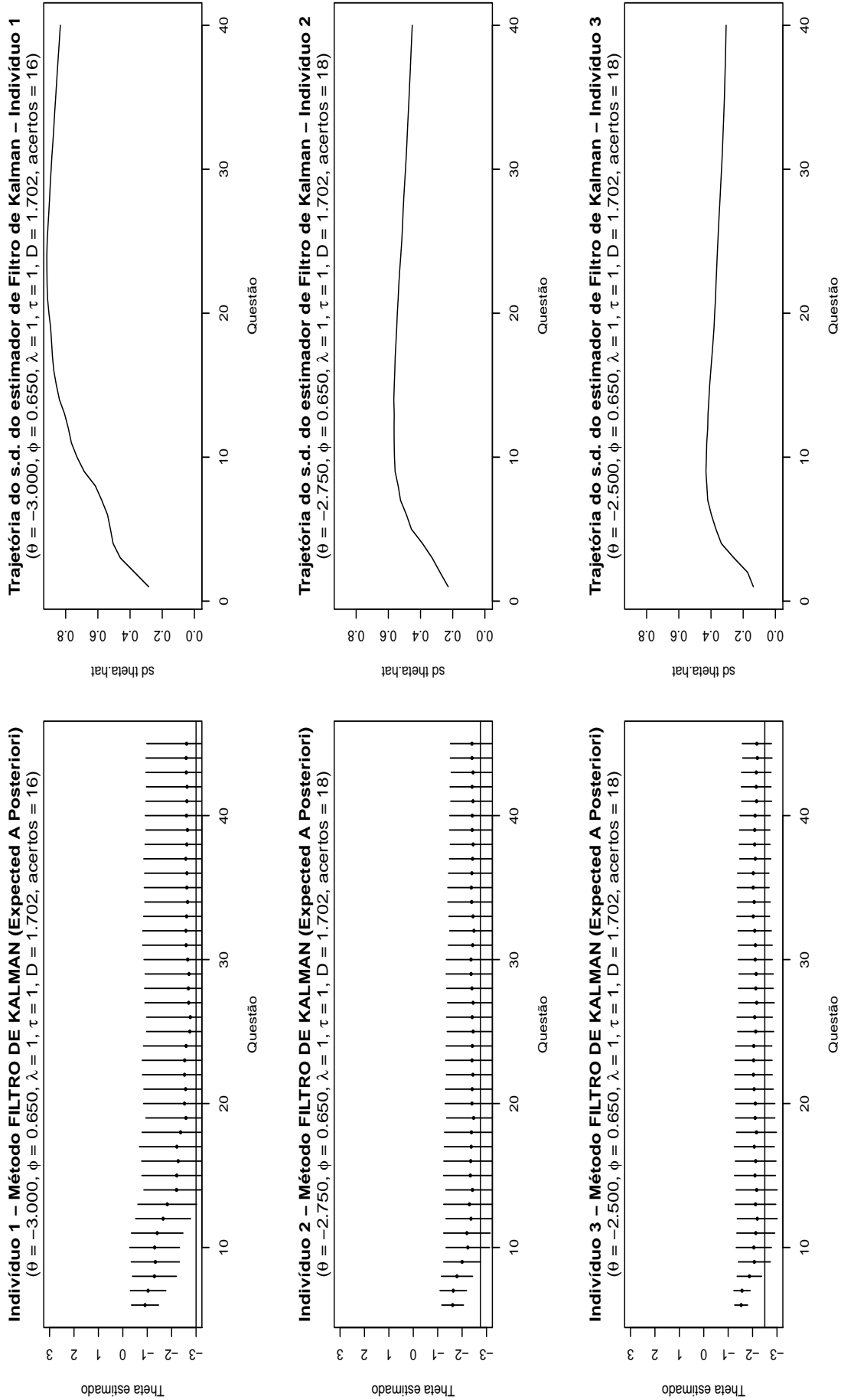


Figura 5.38: *Cenário 5: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65, \lambda = 1, \tau = 1, \theta < 0$*

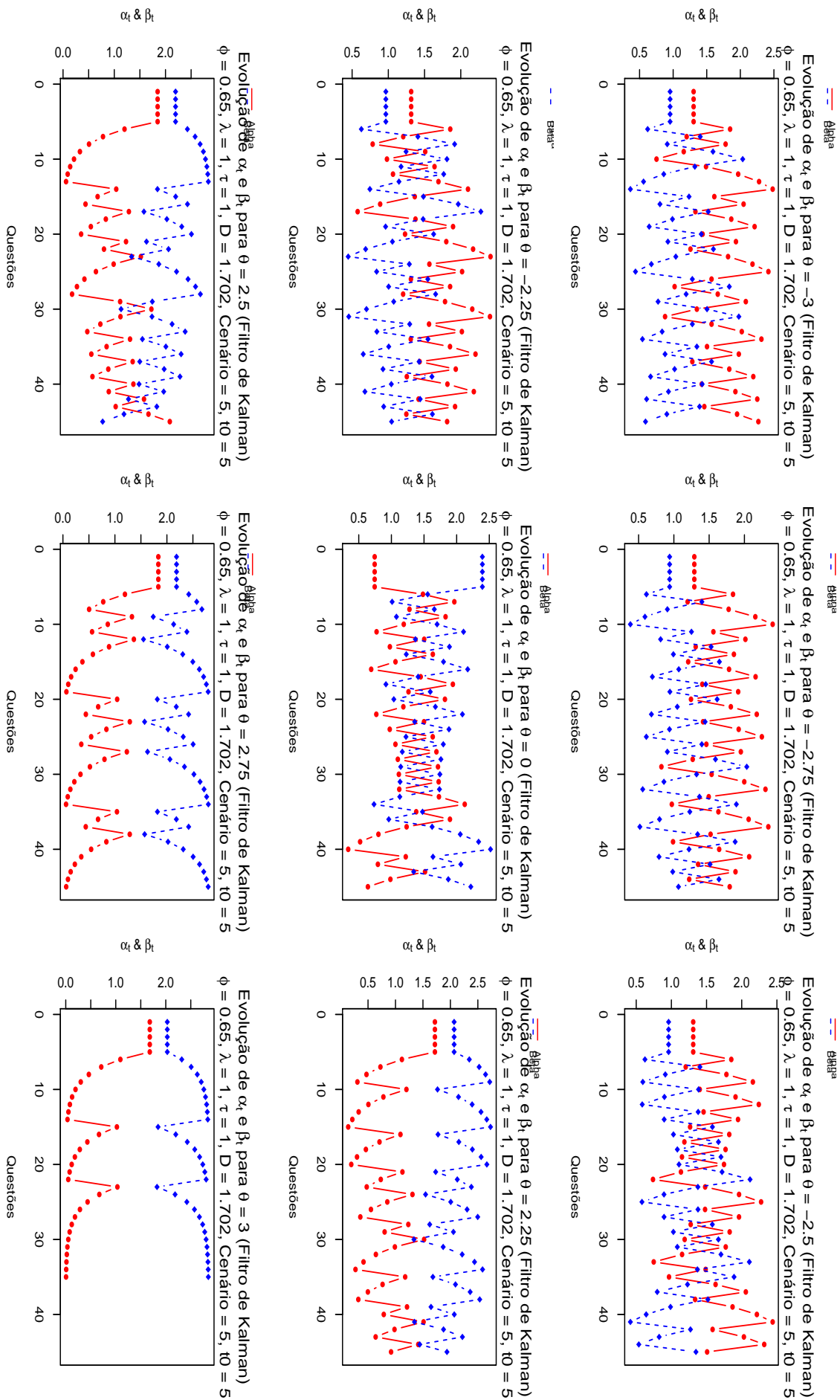


Figura 5.39: Cenário 5: Filtro de Kalman - Trajetória dos  $\alpha_i^*$  e  $\beta_i^*$ , quando  $\phi = 0.65$ ,  $\lambda = 1$ ,  $\tau = 1$ ,  $\theta$  (fixo)

**Tabela 5.15:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $\phi = 0.65$ ,  $\lambda = 1$ ,  $\tau = 1$ , com 5 estimativas provisórias de  $\theta_{jt}$  (Filtro de Kalman - Cenário 5)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	1.0735	0.9545	0.3633	2.0354	0.8035	1.3932
-2.75	1.1151	0.9931	0.3837	2.0335	0.8591	1.3972
-2.50	1.1701	1.0845	0.4172	2.1183	0.8974	1.4058
-2.25	1.1834	1.0565	0.4464	2.2772	0.9360	1.4845
-2.00	1.2506	1.1345	0.6158	2.0960	0.9657	1.5916
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.0970	2.2360	1.0000	2.7690	1.7440	2.4530
2.25	2.1400	2.1260	1.3360	2.7300	1.8750	2.4390
2.50	2.0805	2.1962	0.7732	2.8361	1.7426	2.4289
2.75	2.3130	2.3380	1.5500	2.7970	2.1290	2.6280
3.00	2.5320	2.6770	1.8250	2.8510	2.2600	2.8070

Ao avaliar o Cenário 5a para  $\phi \in \{0.99, 0.65\}$ , apresentados, respectivamente, nas figuras (5.40) e (5.42), para valores de  $\theta < 0$ , e nas figuras (5.41) e (5.43), para valores de  $\theta > 0$ , observou-se que o padrão sistemático das estimativas de proficiências obtidas com o modelo *Filtro de Kalman* pouco se altera, constatando-se, assim, a necessidade de um método de otimização para encontrar o  $\phi$  ideal, que promova melhoria na acurácia das estimativas de proficiência.

Pela Tabela (5.16) observa-se que, tal como no modelo *Power Steady*, dentre os oito cenários analisados com o modelo *Filtro de Kalman*, destacaram-se os Cenários 5b, 5e e 5h, onde foram obtidos os menores valores de RMSE, dentre todos os cenários estudados. Além deste resultado, constata-se, também, que quanto menor o fator de desconto, melhor é o ajuste do modelo.

Neste ponto, observa-se que os cenários criados de acordo com as “*equações de atualização*” expressas em (5.21), evidenciaram uma maior sensibilidade do modelo *Filtro de Kalman*, quanto à modificações nos parâmetros  $\alpha_t^*$  e  $\beta_t^*$ , uma vez que o comportamento dos valores de RMSE apresentados na Tabela (5.16) demonstram, de maneira clara, a direção da acurácia do modelo, em todos os cenários estudados, sendo possível afirmar que o fator de desconto ideal deva ser obtido para valores de  $\phi \leq 0.65$ , destacando, assim, o Cenário 5b como a escolha ideal para testar tais resultados. Entretanto, para os Cenários 5b, 5e e 5h, a maioria dos casos simulados tiveram  $\theta \approx 0$ , tal como se observa no Quadro (5.3), de forma que o *Critério da Máxima Informação*, utilizado no algoritmo implementado para selecionar itens, selecionou itens com valores do parâmetro  $b$  próximos ao valor do  $\theta$  verdadeiro, o que explicaria os baixos valores de RMSE para todos os três cenários.

**Quadro 5.3:** Estatísticas descritivas dos  $\theta$  verdadeiros para os Cenários 5b, 5e e 5h.

	5b	5e	5h
1 N.....	25	25	25
3 Mínimo.....	-2.1106	-2.1106	-2.1106
4 1o. Quartil....	-0.4196	-0.4196	-0.4196
5 Mediana.....	0.1139	0.1139	0.1139
6 Média.....	0.0585	0.0585	0.0585
7 3o. Quartil....	0.5811	0.5811	0.5811
8 Máximo.....	1.6503	1.6503	1.6503
9 Variância.....	0.7768	0.7768	0.7768
10 Desvio Padrão...	0.8813	0.8813	0.8813
11 Erro Padrão....	0.1763	0.1763	0.1763

**Tabela 5.16:** Filtro de Kalman - *Comparações por RMSE*

$\theta$	Cenário	Fator de Desconto ( $\phi$ )							
		0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.99
Fixo	5 ( $\lambda = 1, \tau = 1$ )	2.63738	3.25470	2.38276	2.21391	0.69365	0.36058	0.34477	0.35724
Fixo	5a ( $\lambda = 1, \tau = 0.5$ )	0.41486	0.43524	0.44584	0.46295	0.47724	0.49165	0.50534	0.51349
Aleatório	5b ( $\lambda = 1, \tau = 0.5$ )	<b>0.19919</b>	<b>0.20283</b>	<b>0.20689</b>	<b>0.20434</b>	<b>0.20980</b>	<b>0.21204</b>	<b>0.20599</b>	<b>0.20875</b>
Fixo	5c ( $\lambda = 2, \tau = 0.5$ )	0.54253	0.55575	0.57230	0.58659	0.59519	0.60677	0.61856	0.62748
Fixo	5d ( $\lambda = 2, \tau = 1$ )	0.51295	0.56066	0.59957	0.61889	0.64992	0.67691	0.70112	0.71846
Aleatório	5e ( $\lambda = 2, \tau = 0.5$ )	<b>0.21069</b>	<b>0.21635</b>	<b>0.22013</b>	<b>0.22170</b>	<b>0.22614</b>	<b>0.23138</b>	<b>0.23264</b>	<b>0.23684</b>
Fixo	5f ( $\lambda = 3, \tau = 0.5$ )	0.60135	0.61539	0.62576	0.63572	0.64739	0.65854	0.66794	0.67417
Fixo	5g ( $\lambda = 3, \tau = 1$ )	0.61031	0.63647	0.66597	0.69137	0.71392	0.72350	0.74344	0.75781
Aleatório	5h ( $\lambda = 3, \tau = 0.5$ )	<b>0.22667</b>	<b>0.23280</b>	<b>0.23659</b>	<b>0.23742</b>	<b>0.24225</b>	<b>0.24584</b>	<b>0.24992</b>	<b>0.25188</b>

A trajetória das estimativas de proficiência para os Cenários 5b, 5e e 5h, destacados pela Tabela (5.16), por obter os melhores resultados de RMSE, são apresentadas nas figuras (5.44) e (5.45), onde se observa que, quanto mais  $\theta \approx 0$ , mais os intervalos de confiança tornam-se expressivamente pequenos, comprometendo assim, as estimativas da acurácia das estimativas de proficiência, uma vez que são excessivamente pequenas, embora estas estejam muito próximas do valor de  $\theta$  verdadeiro.

No modelo *Filtro de Kalman*, assim como no modelo *Power Steady*, utilizou-se, neste trabalho, o *Critério da Máxima Informação*, descrito na Seção (3.2.1), como método para seleção de itens. De forma que, quando o valor de  $\theta$  verdadeiro é baixo ( $\theta = 0.65$ ), o critério seleciona itens com baixos valores do parâmetro  $b$ . Sendo este mensurado na mesma escala da proficiência, as estimativas de proficiência tendem a acompanhar este comportamento, fazendo com que sejam muito próximas ao  $\theta$  verdadeiro, o que explica os baixos valores de RMSE apresentados na tabela (5.16).

Para uma melhor avaliação das simulações realizadas com os oito cenários descritos na Seção 5.1.7, diagramas de dispersão são apresentados nas Figuras (5.46) a (5.53), no intuito de se estudar o ajuste das estimativas de proficiência geradas pelo modelo *Filtro de Kalman*, em relação ao  $\theta$  verdadeiro. Comparando os resultados dos valores de RMSE, apresentados na Tabela (5.16), observa-se que, embora a tabela destaque o Cenário 5b como o cenário mais promissor, dentre os oito, examinando-se os resultados apresentados pelas Figuras (5.46) a (5.53), tem-se uma impressão diferente. Notoriamente, para todos os  $\phi$ 's do grid estudado, o valor da correlação entre  $\theta$  e  $\hat{\theta}$  é sempre maior para o Cenário 5a. Pela Tabela (5.16), observa-se que o Cenário 5a é o que apresenta os melhores resultados se desconsiderados os Cenários 5b, 5e e 5h. Tal escolha se sustenta, ao observar, nos resultados das comparações de ajuste entre  $\theta$  e  $\hat{\theta}$ , que em todos os gráficos apresentados pelas figuras (5.46) a (5.53), observam-se os valores de  $\rho$  e  $R_{\text{Ajustado}}^2$  superiores à 0.99.

Portanto, conclui-se que, para o modelo *Filtro de Kalman*, embora as simulações realizadas para o Cenário 5b tenham apresentado bons resultados, deve-se considerar que, como foram escolhidos  $\theta$ 's aleatórios para se ajustar o modelo, sendo observados apenas 25 casos, estudos mais aprofundados devam ser feitos, com mais observações, a fim de se confirmar, de fato, que o Cenário 5b seja melhor que o Cenário 5a. A mesma conclusão aplica-se aos Cenários 5e e 5h, já que ambos, compartilham, com o Cenário 5b, a característica da escolha do  $\theta$  verdadeiro de forma aleatória.

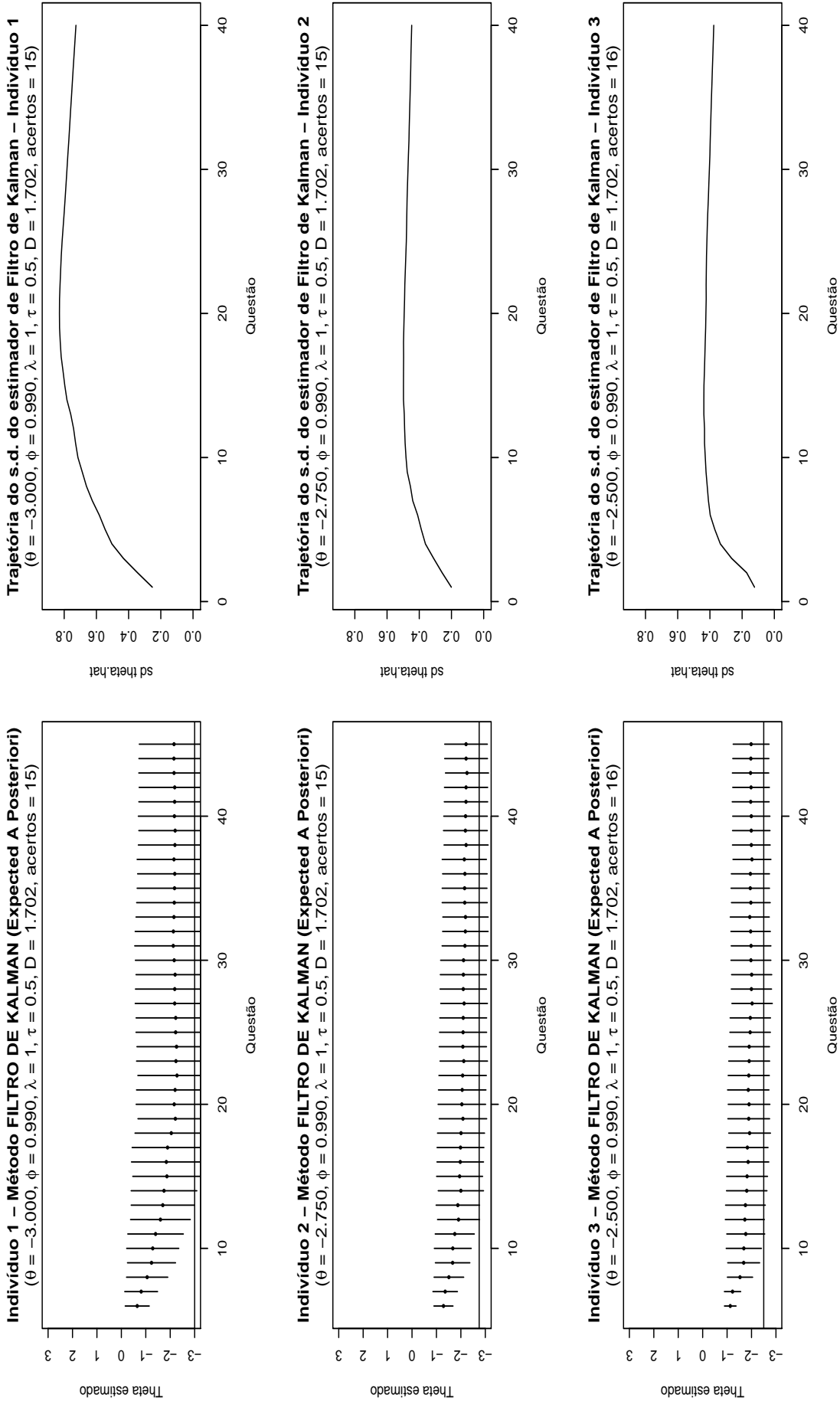


Figura 5.40: *Cenário 5a: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.99, \lambda = 1, \tau = 0.5, \theta < 0$*

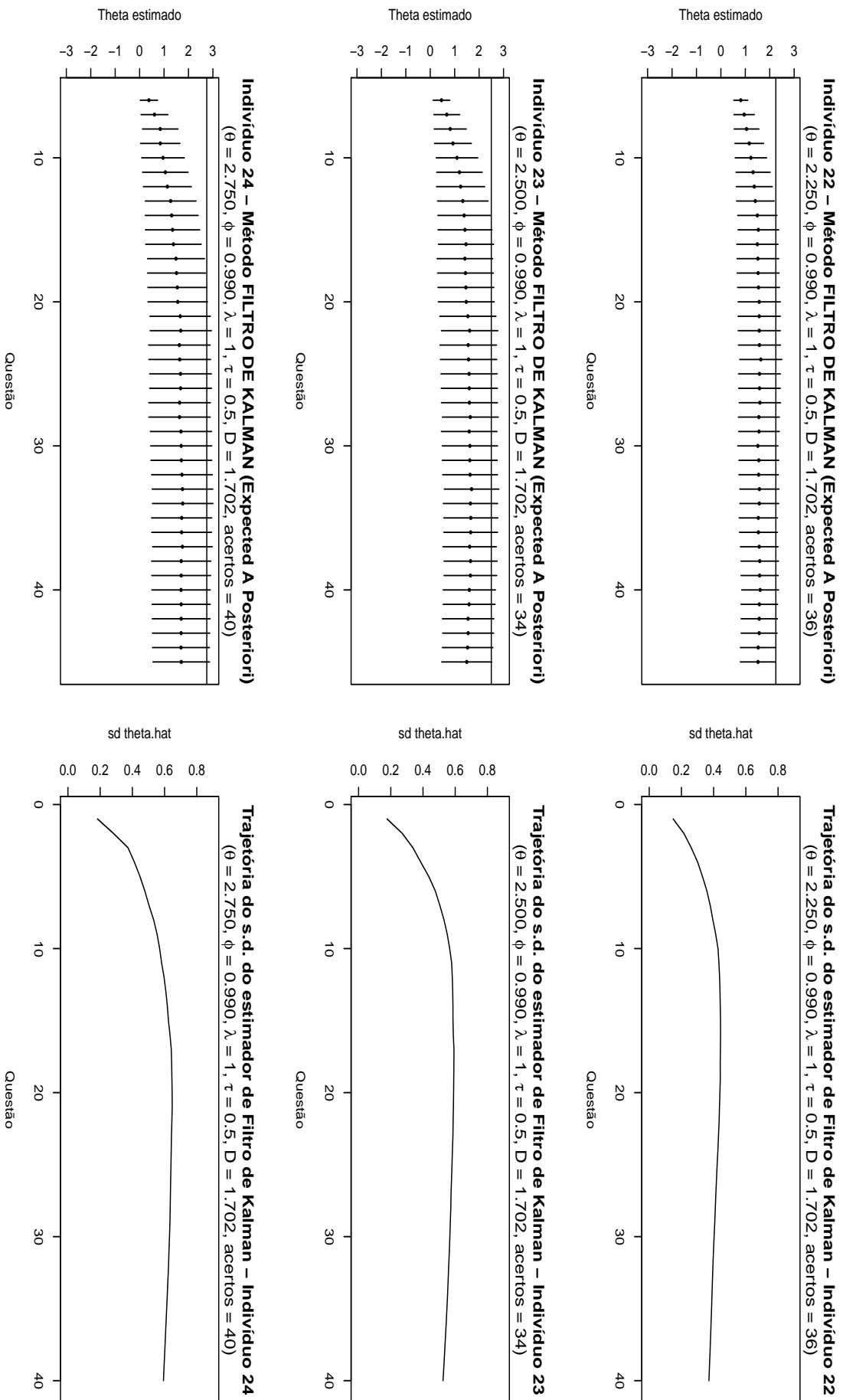


Figura 5.41: Cenário 5a: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.99, \lambda = 1, \tau = 0.5, \theta > 0$



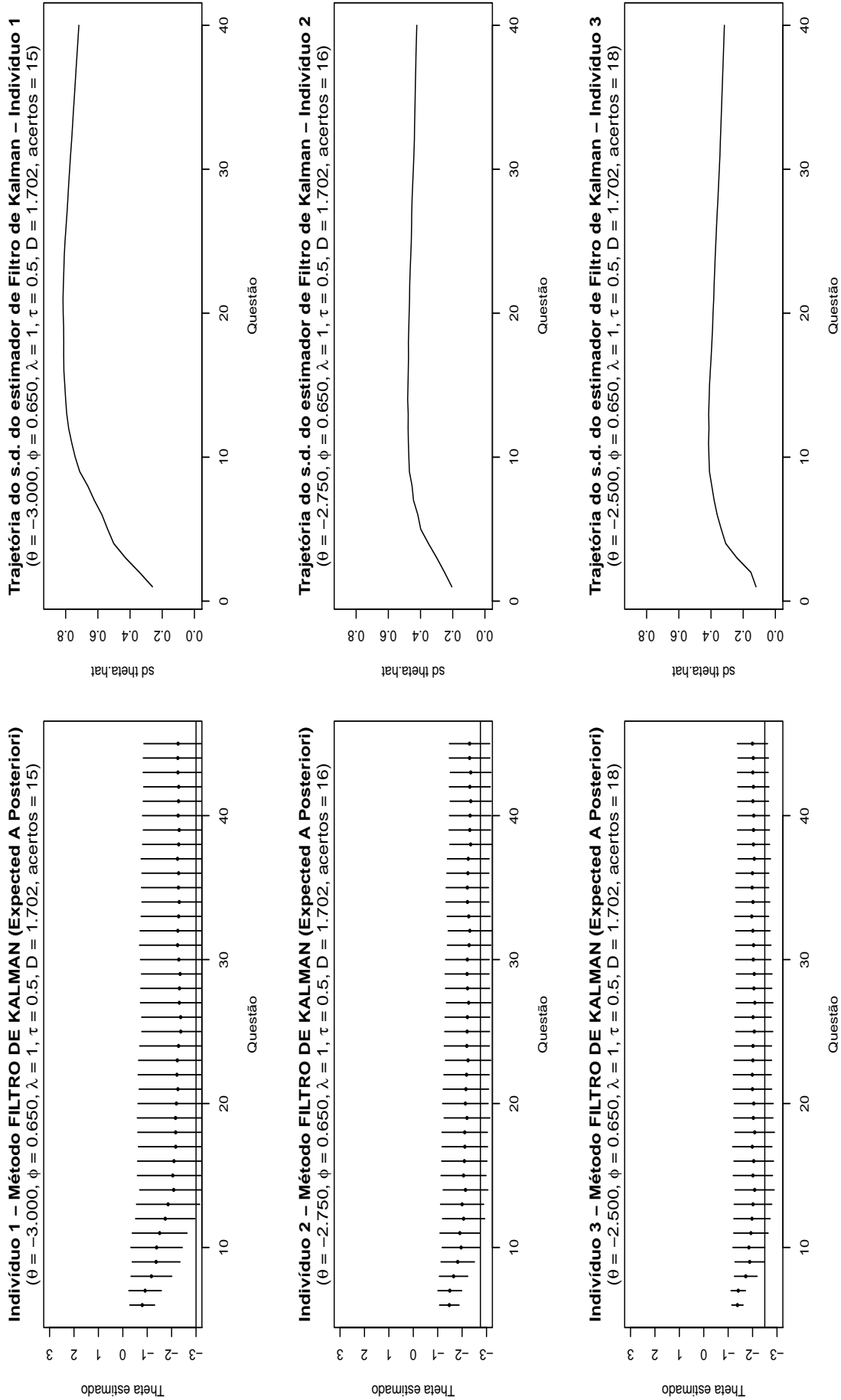


Figura 5.42: Cenário 5a: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta < 0$

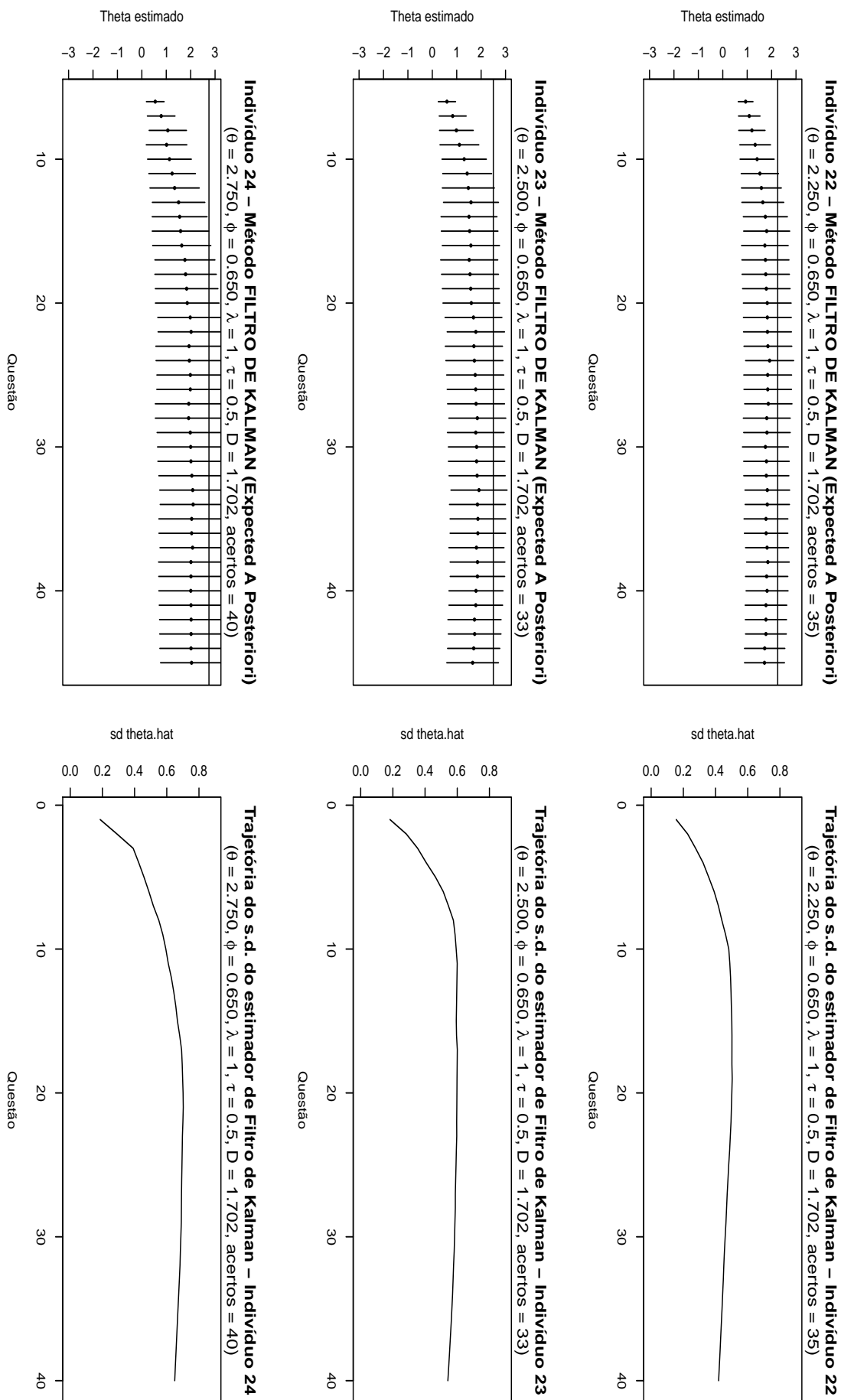


Figura 5.43: Cenário 5a: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta > 0$

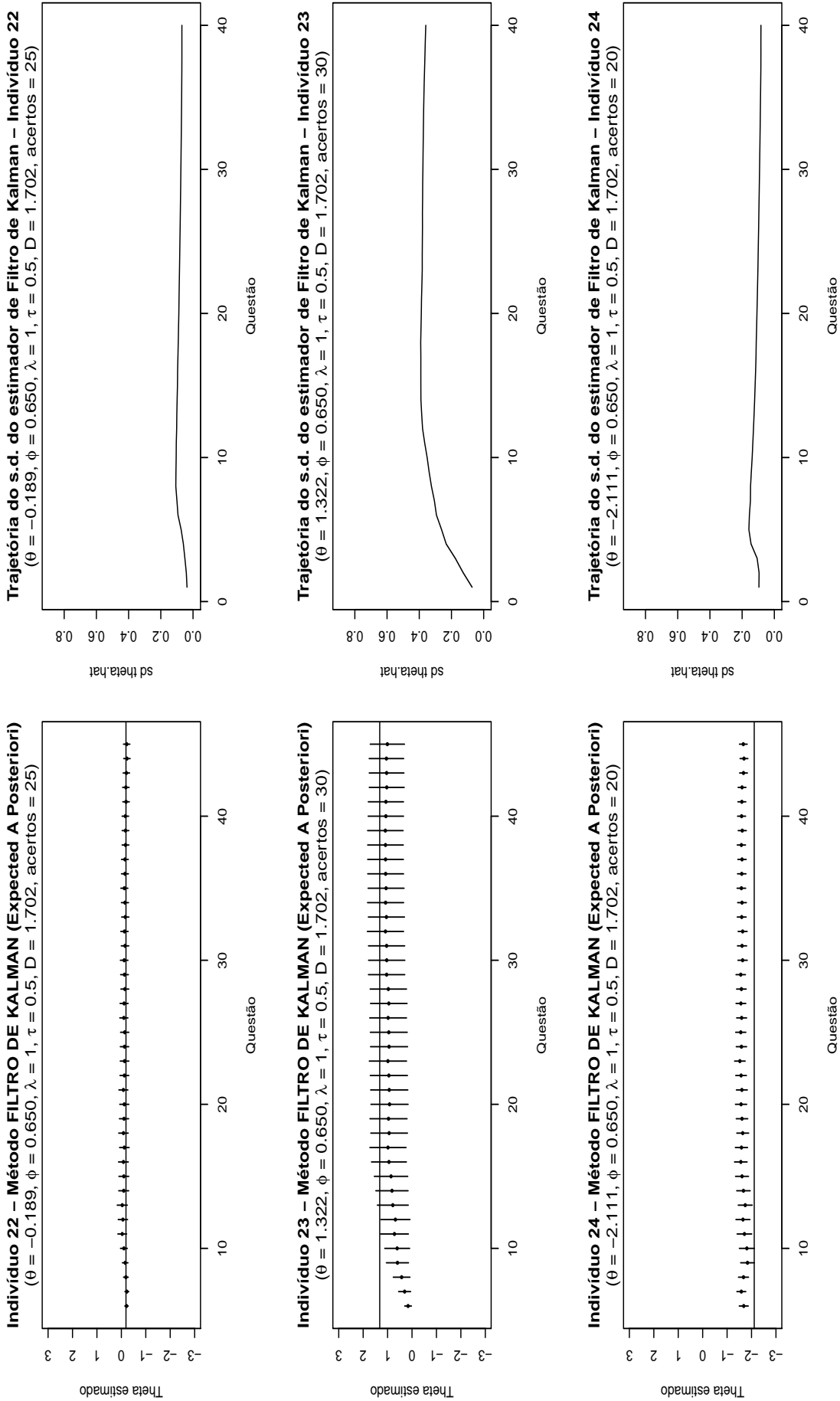


Figura 5.44: *Cenário 5b: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65$ ,  $\lambda = 1$ ,  $\tau = 0.5$ ,  $\theta$  (Aleatório)*

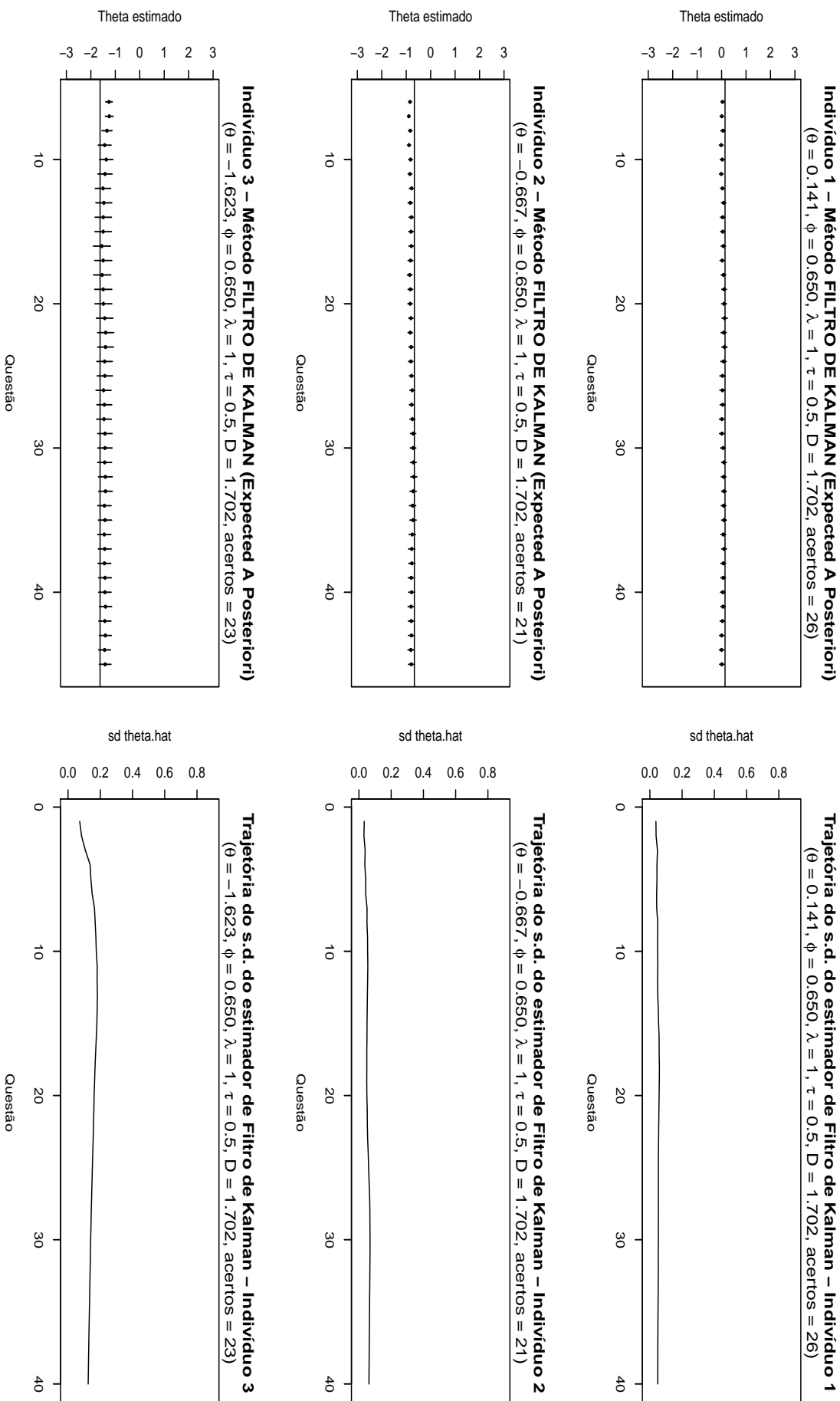
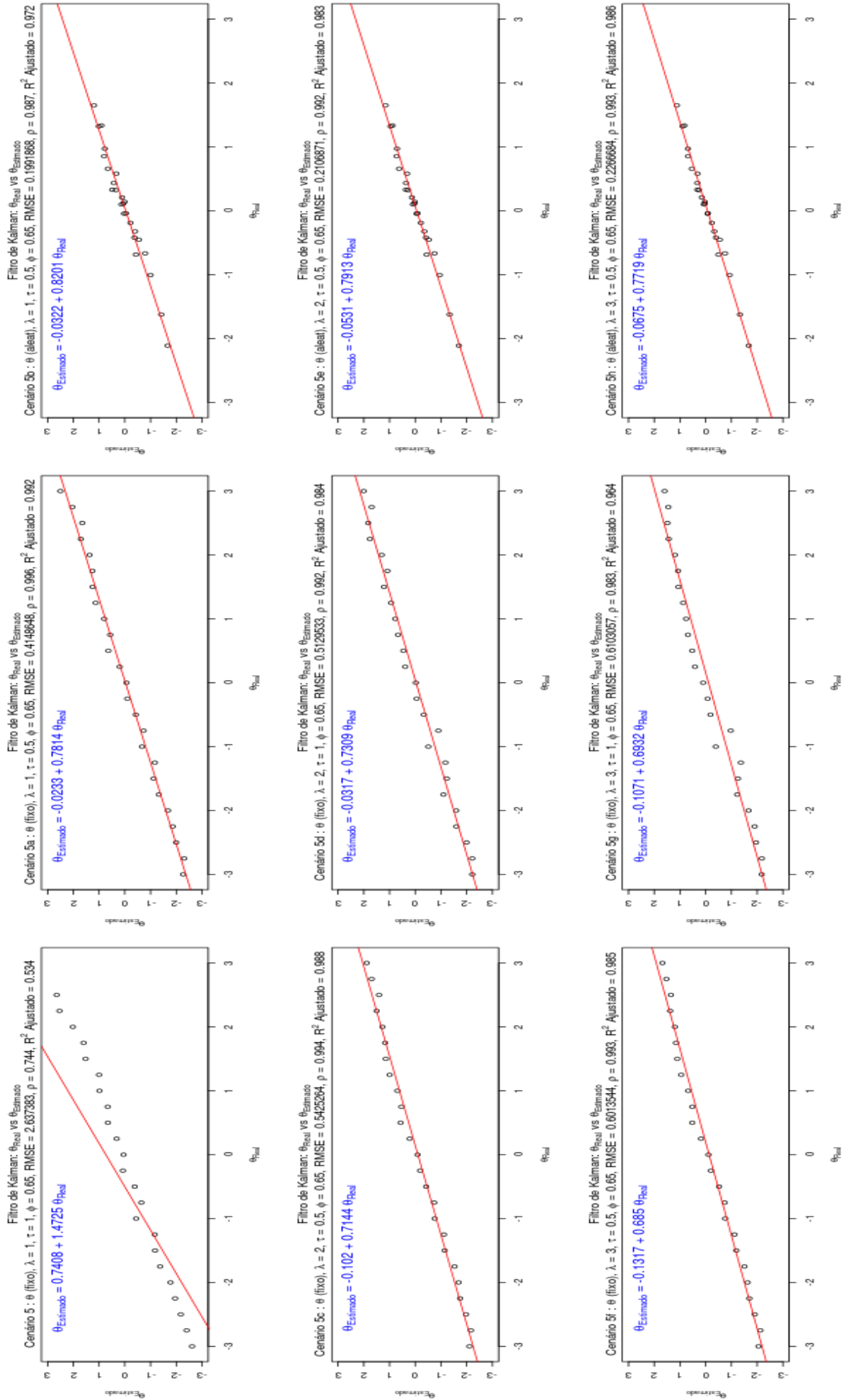


Figura 5.45: Cenário 5b: Filtro de Kalman - Trajetória do estimador  $\hat{\theta}$ , quando  $\phi = 0.65, \lambda = 1, \tau = 0.5, \theta$  (Aleatório)



**Figura 5.46:** Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.65$

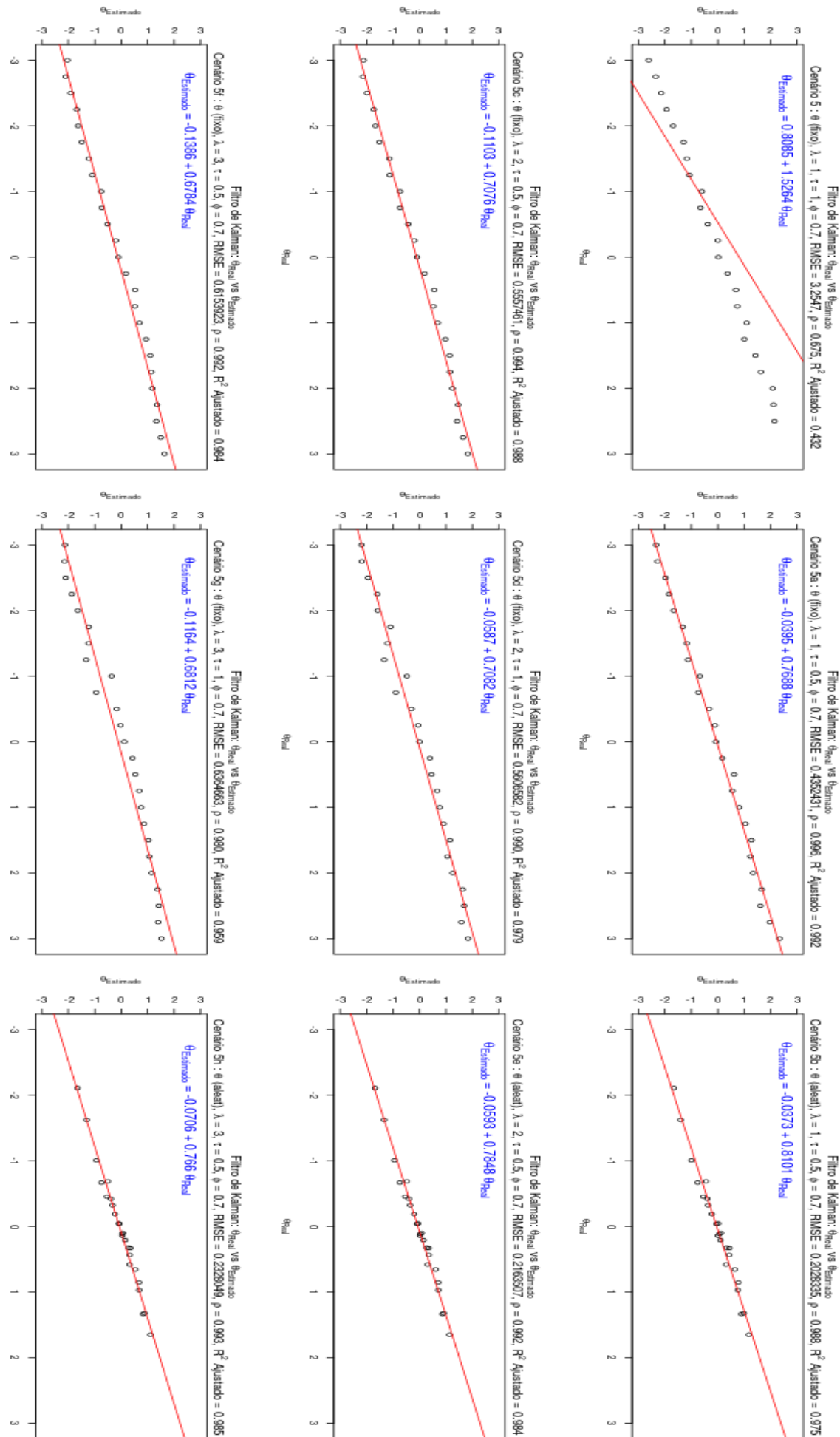
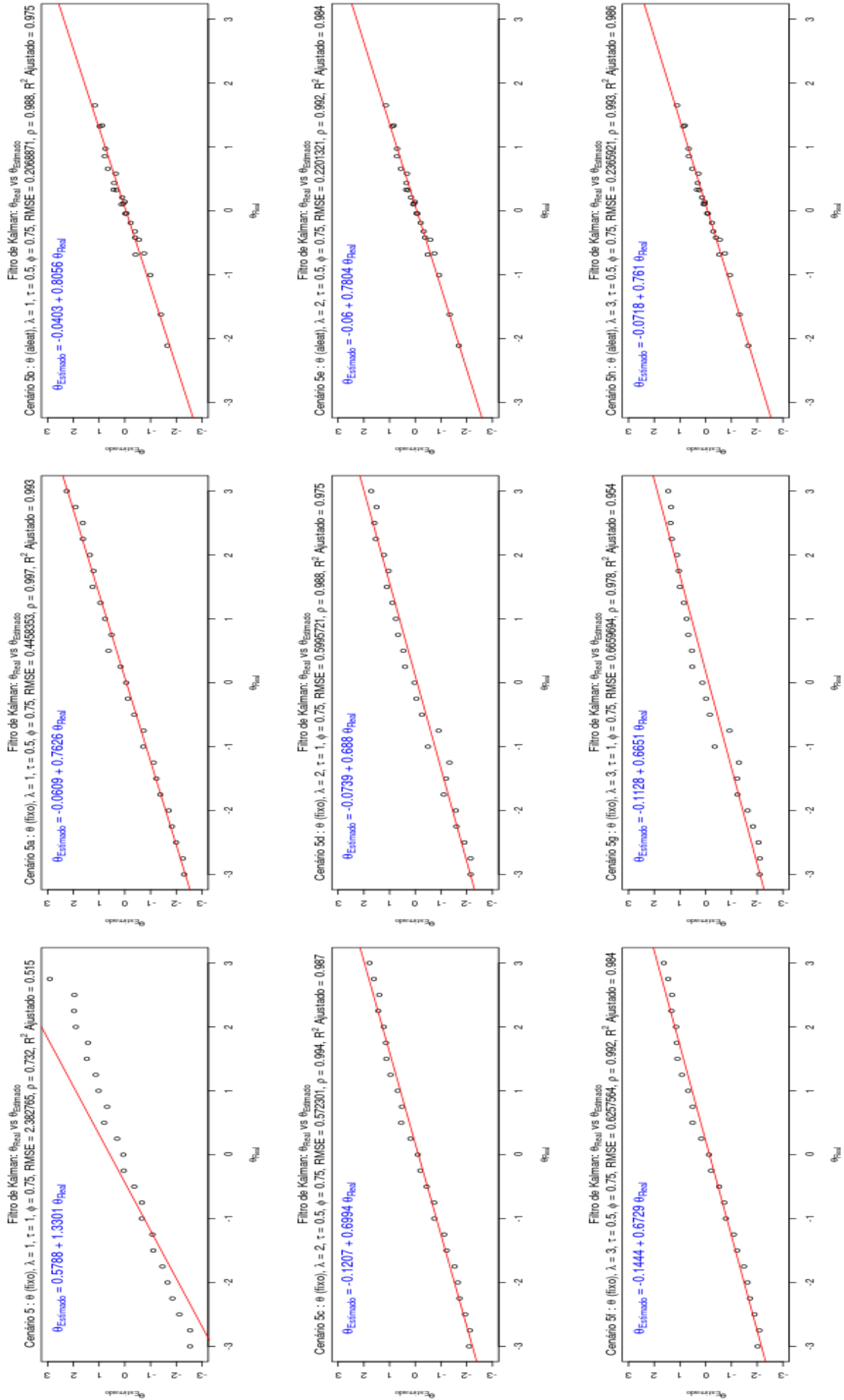


Figura 5.47: Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.70$



**Figura 5.48:** Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.75$

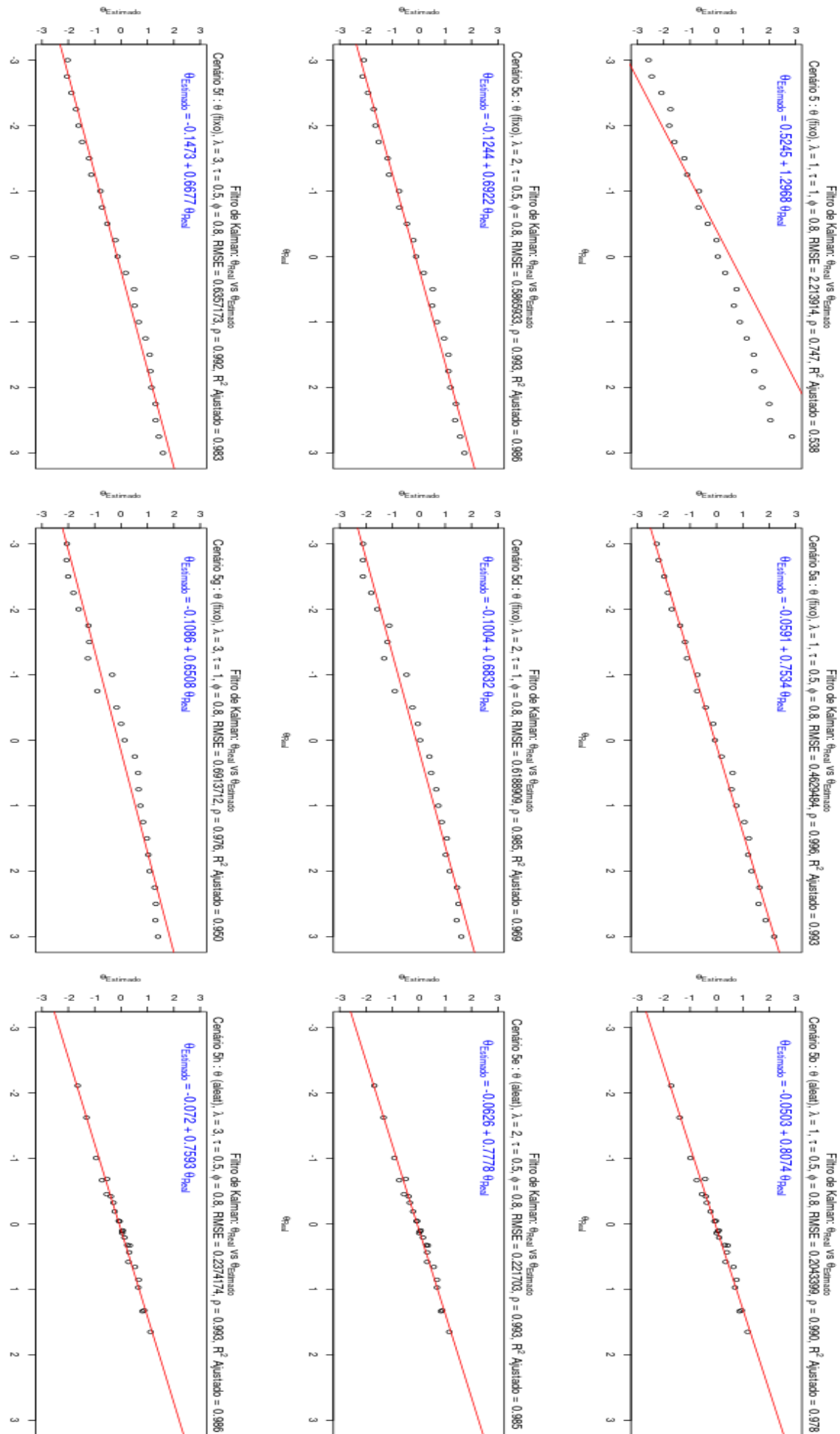
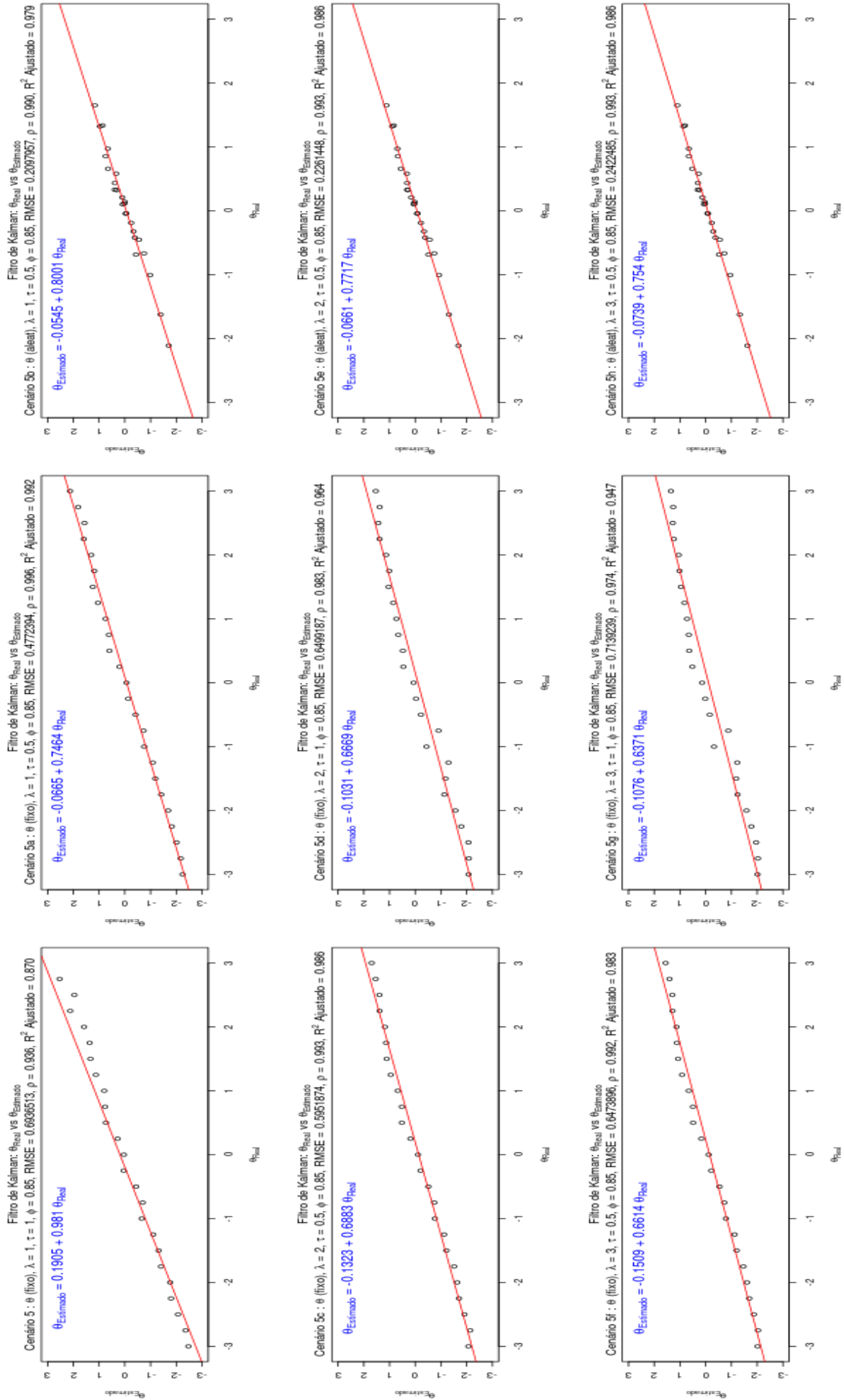


Figura 5.49: Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.80$





**Figura 5.50:** Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.85$

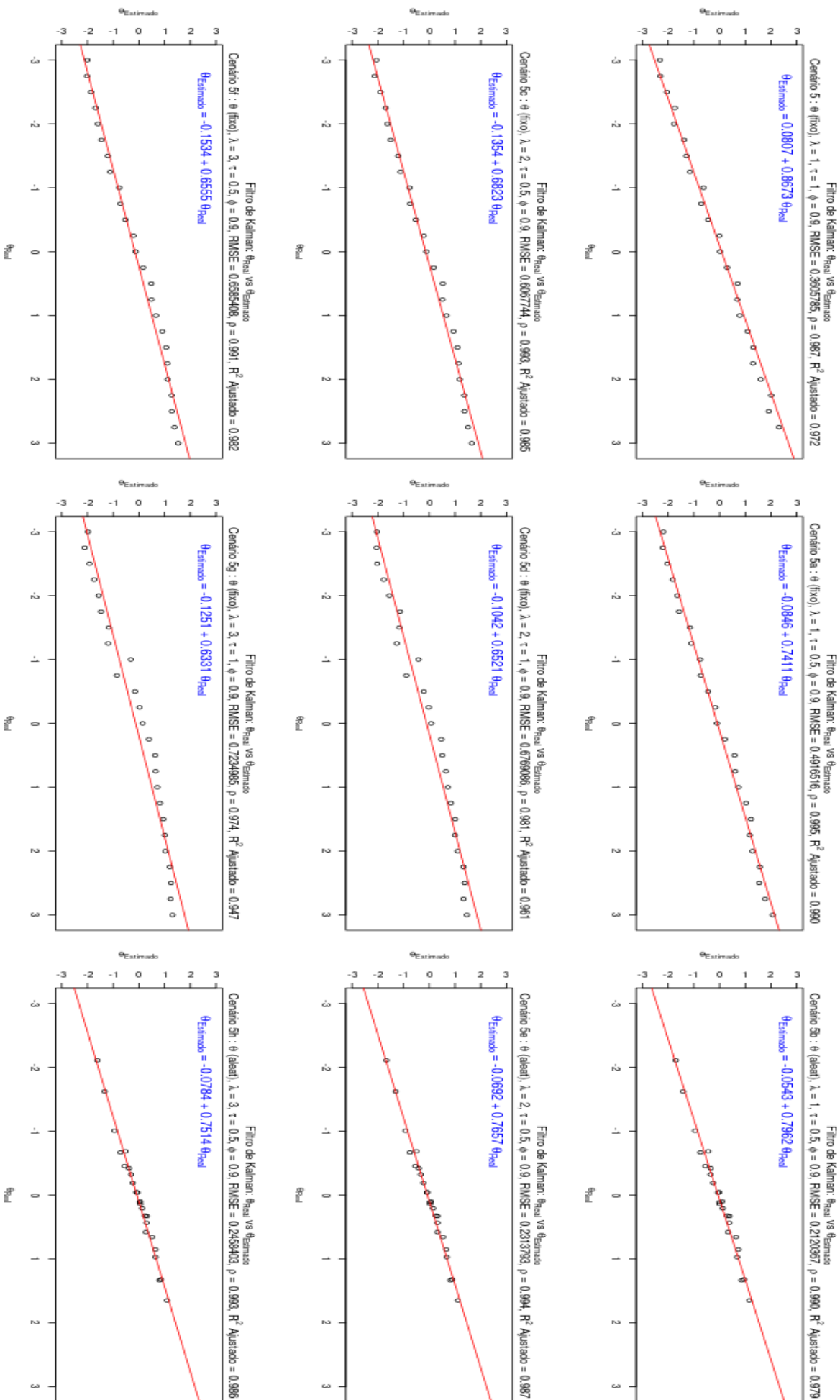
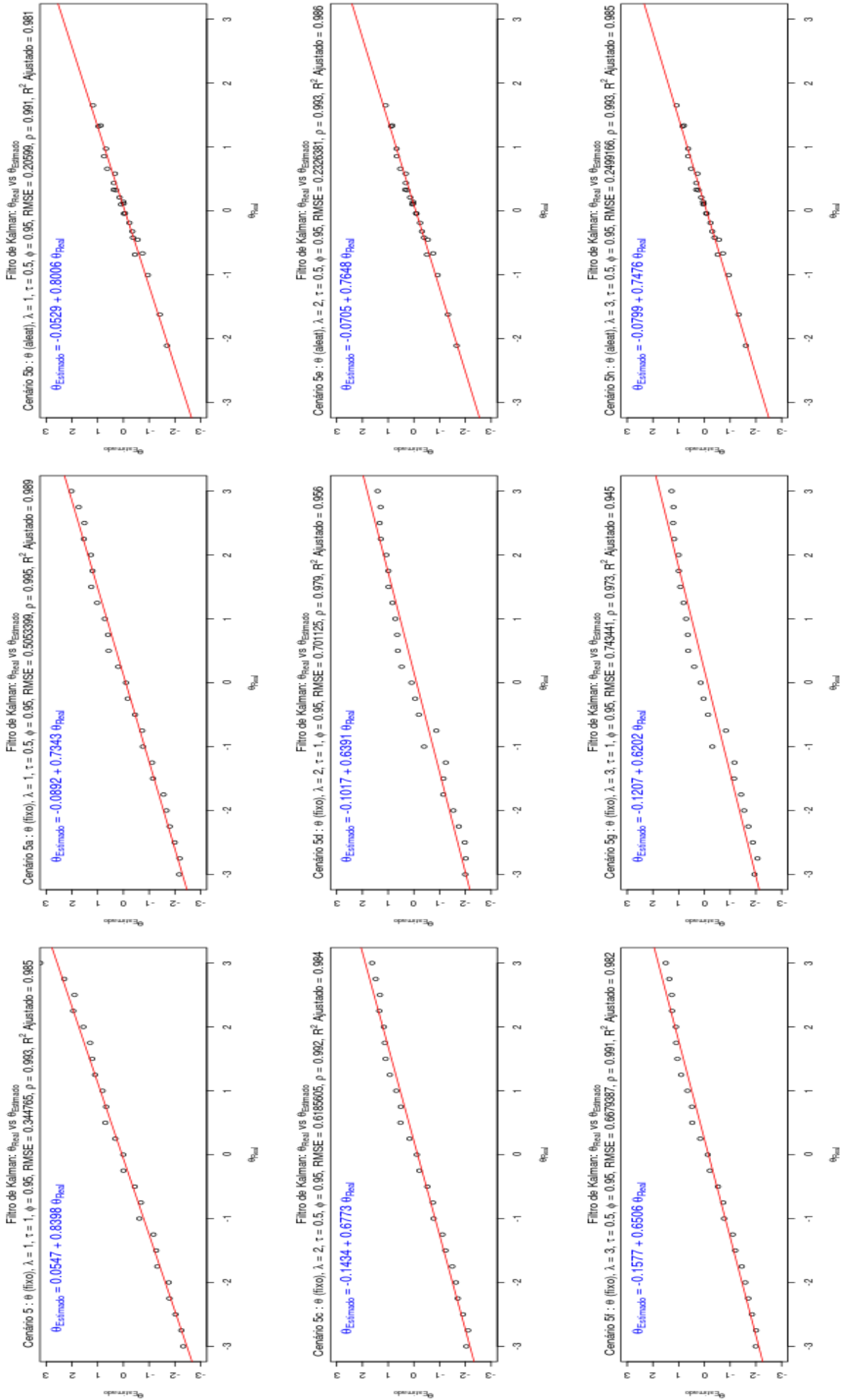


Figura 5.51: Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.90$



**Figura 5.52:** Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.95$

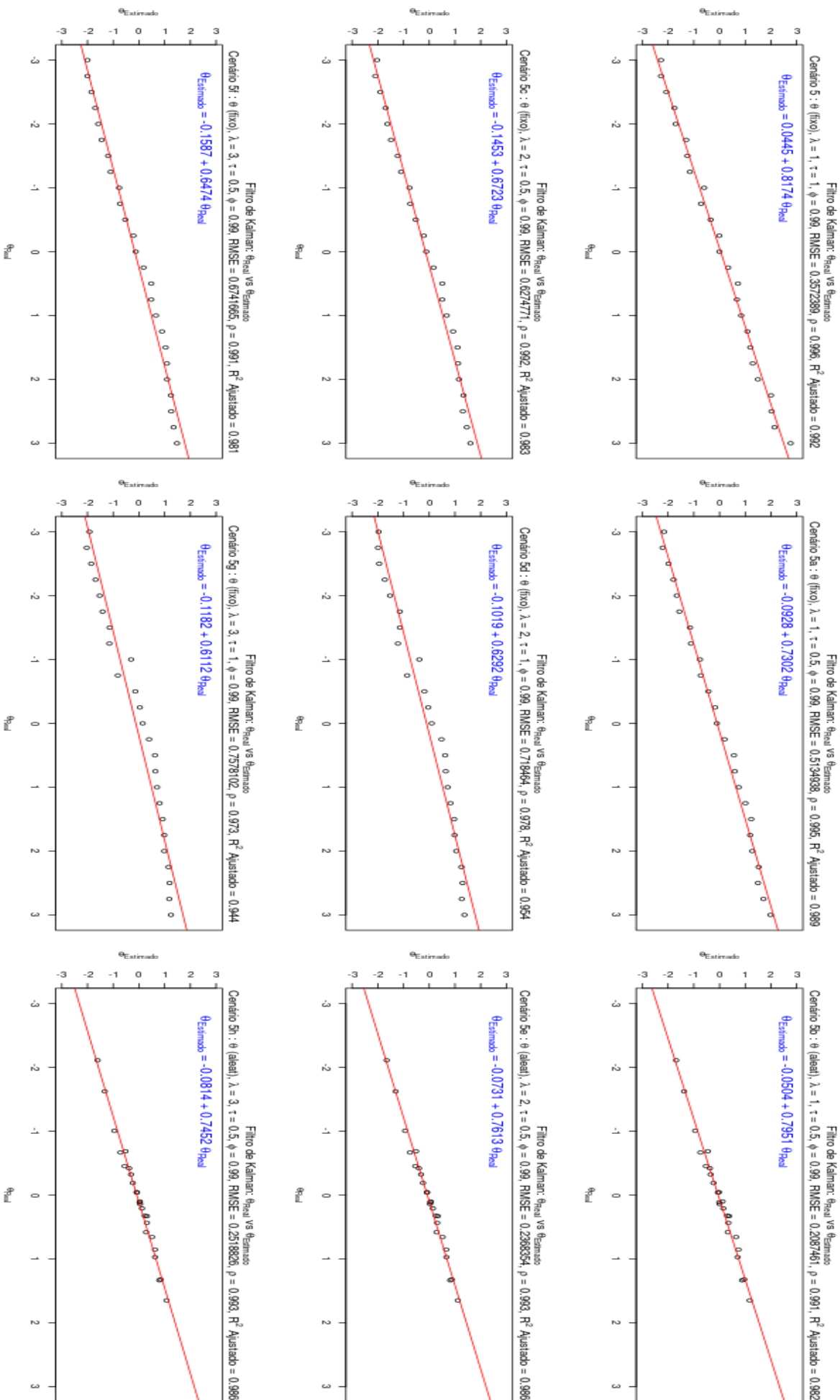


Figura 5.53: Filtro de Kalman: Comparação entre  $\theta$  vs  $\hat{\theta}$ , quando  $\phi = 0.99$

# Capítulo 6

## Conclusões e Trabalhos Futuros

### 6.1 Conclusões

A perspectiva de encontrar métodos que sejam competitivos à aplicação dos *Testes Computadorizados Adaptativos* (CAT), em particular, em sua versão tradicional do modelo EAP/3PL, abre uma ampla possibilidade de pesquisas na área. Tratou-se, nesta dissertação, de uma destas possibilidades, utilizando-se o poder de adequação dos modelos dinâmicos às formulações não-gaussianas (Harvey, 1989). Ponderando o uso de diferentes formulações, o presente trabalho procurou trazer duas formulações para o contexto de estimação de proficiências no CAT. Uma baseada no princípio do *Filtro de Kalman*, apresentado por Harrison e Stevens (1976), e a outra, com enfoque direcionado aos modelos dinâmicos de Smith (1979).

Os estudos de simulação apresentados foram divididos por cenários, de maneira a propiciar a aplicação dos mesmos, em ambas as formulações adotadas. Foram utilizados dados de um teste já consolidado, aplicado aos alunos de graduação da Universidade de Brasília desde 2010, para avaliar a proficiência em Inglês Instrumental I, cujos parâmetros, já se encontravam calibrados segundo o modelo TRI EAP/3PL. A partir destes dados, foram geradas inicialmente, simulações para um CAT tradicional, considerando-se provas de tamanho fixo ( $n = 45$ ), para se obter uma referência para as comparações realizadas com os dois modelos estudados.

Foram simuladas as respostas de 200 indivíduos, para os quais resultados revelaram que a trajetória dos desvios-padrão das estimativas de proficiência apresentavam comportamento estável, convergindo, gradativamente, ao longo do CAT, para o limite de 0.2, resultado este compatível com a literatura sobre testes adaptativos Bock e Mislevy (1982); Costa (2009). Um resultado bastante utilizado nas comparações realizadas com os dois modelos considerados, é o valor de RMSE para a estimativa de proficiência, que para o CAT tradicional foi  $\approx 0.13$ .

As simulações realizadas para os dois modelos propostos revelaram o alto impacto nas estimativas de  $\theta$ , dos hiperparâmetros  $\alpha$  e  $\beta$ , no modelo dinâmico *Power Steady*. Uma vez estabelecida a relação de dependência entre  $\alpha$  e  $\beta$  com os parâmetros  $a$  e  $b$  dos itens, utilizaram-se diferentes cenários para avaliar os resultados dos dois modelos. Ao longo do estudo, consideram-se treze cenários, sendo os quatro primeiros utilizados como ajustes para sistematizar a criação dos demais cenários. Assim, os resultados para os Cenários 1 e 2 revelaram, além de um crescimento exponencial dos parâmetros  $\alpha$  e  $\beta$ , uma forte dependência destes com as estimativas de proficiência.

Para a simulação do modelo dinâmico *Power Steady*, considerou-se o fator de desconto  $k$  tal que  $k \in \{0.81, \dots, 0.99\}$ , com saltos de 0.01. Foram observados comportamentos cíclicos das estimativas de proficiência nesta formulação, fazendo com que os valores de RMSE

avaliados, pertencessem ao intervalo  $\{0.61, \dots, 0.80\}$ , valores estes muito superiores à média de 0.18, observada para as simulações do CAT Tradicional. Na tentativa de corrigir o padrão cíclico observado e reduzir o valor de RMSE obtido nos Cenários 1 e 2, foram gerados os Cenários 3 e 4, com base em restrições aos valores iniciais assumidos pelos parâmetros  $\alpha$  e  $\beta$  anteriormente. As modificações realizadas, estabilizaram um pouco os parâmetros  $\alpha$  e  $\beta$ , obtendo-se, assim, melhorias nos valores de RMSE avaliados, observados no intervalo  $\{0.42, \dots, 0.52\}$ .

Até o cenário 4, os resultados avaliados eram apenas para os estimadores PS\_EAP ( $\hat{\theta}$ ) e PS\_MAP ( $\tilde{\theta}$ ), sem apresentar resultados expressivos. A partir do Cenário 5 em diante, além da inclusão de parâmetros de penalização ( $\lambda$  e  $\tau$ ) às *equações de atualização*, as avaliações passaram a ser feitas também para os estimadores PS\_MAP\_AVG ( $\bar{\theta}$ ), PS\_MAP\_MEDIAN ( $\hat{\theta}$ ), PS\_EAP\_AVG ( $\hat{\theta}$ ) e PS\_EAP\_MEDIAN ( $\hat{\theta}$ ), considerados suavizações da média e da mediana, para os métodos EAP e MAP. Com as simulações realizadas, concluiu-se que, entre os seis estimadores do modelo *Power Steady*, o que melhor se destacou foi o PS\_EAP\_AVG ( $\hat{\theta}$ ), que para o Cenário 5, atingiu RMSE igual a 0.27.

De modo geral, as suavizações estabilizaram os problemas de efeitos cíclicos apresentados pelos estimadores PS\_EAP ( $\hat{\theta}$ ) e PS\_MAP ( $\tilde{\theta}$ ), permitindo que estimativas mais acuradas do  $\theta$  verdadeiro fossem encontradas (vide Cenário 5). Porém, estudos mais aprofundados precisam ser realizados para se verificar as razões pelas quais o comportamento da trajetória dos desvios-padrão das estimativas de proficiência tenderam a apresentar intervalos de confiança tão pequenos, em particular para o método MAP, ao longo do CAT. Além deste fato, outras ocorrências necessitam ser investigadas, pois foram observados casos de subestimação ou superestimação, majoritária para alguns cenários.

Mesmo com problemas estruturais a serem investigados, o modelo *Power Steady* obteve resultados promissores quanto ao objetivo geral que motivou o estudo, uma vez que a convergência das estimativas de proficiência estabilizaram-se, aproximadamente, a partir da 10ª questão, de modo que, se corrigidos os problemas relativos ao desvio-padrão, o modelo permitiria diminuir o número de questões a serem apresentadas aos examinandos do teste, tornando o processo de avaliação mais rápido e eficiente.

O modelo *Filtro de Kalman* mostrou-se menos propenso às oscilações dos parâmetros  $\alpha$  e  $\beta$ , de forma que os resultados obtidos para os diferentes fatores de desconto ( $\phi$ ) apresentaram um padrão sistemático para as estimativas de proficiências, representando um vício de estimativa, de modo que o modelo tende a gerar resultados superestimados quando  $\theta < 0$  e subestimados quando  $\theta > 0$ . Ainda que o modelo *Filtro de Kalman* apresente tal característica, de modo geral, a suavização da trajetória dos desvios-padrão das estimativas de proficiência não é tão expressiva quanto no modelo *Power Steady*, a não ser em casos em que  $\theta \approx 0$ . Sua principal característica, observada com os resultados das simulações realizadas, é quanto à estabilidade quando avaliado para diferentes fatores de desconto.

Sem tratar do tema otimização, o presente estudo mostrou que, tal como o modelo *Power Steady*, o modelo *Filtro de Kalman*, apesar do vício sistemático, é bastante promissor para estudos mais aprofundados, podendo até mesmo se transformar em uma opção bastante competitiva em relação ao CAT tradicional. De maneira geral, o modelo *Filtro de Kalman* apresentou-se mais estável que o modelo *Power Steady*, visto que suas estimativas, suavizadas pelo método EAP, são aparentemente mais comparáveis ao estimador PS\_EAP\_AVG ( $\hat{\theta}$ ). Assim, considerando que, dentre os estimadores apresentados com o modelo *Power Steady* o estimador PS\_EAP\_AVG ( $\hat{\theta}$ ) obteve os melhores resultados, e que o modelo *Filtro de Kalman* apresentou-se mais estável, com menos oscilações, entre diferentes cenários, conclui-se que o estimador FK\_EAP ( $\hat{\theta}$ ) proporcionou os melhores resultados dentre todos os estima-

dores apresentados, obtendo-se um valor de RMSE igual 0.19919, para o Cenário 5b.

Embora os resultados apontem para o FK\_EAP ( $\hat{\theta}$ ), como sendo o melhor estimador dentre os estudados nesta dissertação, considerando-se a discussão realizada na Seção 5.2.1, sobre a qualidade das estimativas dos Cenários 5b, 5e e 5h, levanta-se a suspeita de que talvez as estimativas de proficiência do Cenário 5a, sejam mais robustas, embora tenham valor de RMSE igual a 0.41486. Além do mais, ao se considerar o FK\_EAP ( $\hat{\theta}$ ), como a escolha ideal do estudo realizado, levanta-se também outra importante questão, que diz respeito à calibração prévia dos itens utilizados nas simulações avaiadas, ter sido realizada utilizando-se a estimação EAP/3PL para o CAT tradicional, e neste caso, novas simulações com os modelos propostos deveriam ser testadas, partindo-se de cenários onde tais fatores fossem isolados, justificando assim uma recalibração dos itens disponíveis, para avaliar os resultados de forma comparativa.

Como último ponto de destaque, vale ressaltar que, caso o padrão sistemático de superestimação ou subestimação do estimador FK\_EAP ( $\hat{\theta}$ ), venha a ser corrigido apenas com a simples utilização de outro banco de itens, as evidências favoráveis à afirmação de que o estimador FK\_EAP ( $\hat{\theta}$ ) é melhor do que os demais, tornam-se ainda mais fortes, uma vez que o comportamento estável apresentado pelo modelo *Filtro Kalman*, para todos os cenários estudados, corroboram com tal hipótese.

## 6.2 Trabalhos Futuros

Há diversas adaptações que podem ser feitas no presente trabalho, de maneira a torná-lo mais apurado, na busca por um modelo mais competitivo em relação ao CAT tradicional. A seguir, apresentam-se alguns pontos que, durante a execução do presente trabalho, foram questionamentos ou mesmo melhorias a serem desenvolvidas nesta dissertação.

Vale ressaltar que o código para esta dissertação, foi desenvolvido em linguagem R e está registrado, até o presente momento, em ambiente privado no repositório [GitHub](#), podendo ser compartilhado para outros usuários, desde que haja o compromisso dos mesmos em desenvolvimento acadêmico, assim como o compromisso de colaboração técnica.

Dentre os questionamentos (Q) e melhorias (M), listam-se:

- (M) Desenvolvimento de pacote na linguagem R, que considere a utilização de modelos de dinâmicos como o *Power Steady* e o *Filtro de Kalman*, para a estimação de proficiências.
- (M) Desenvolvimento em linguagem C (pacote Rcpp) das rotinas de estimação de proficiências.
- (M) Desenvolvimento de métodos de seleção de itens como o de Sympon-Hetter (Leung *et al.*, 2002; van der Linden, 2003).
- (Q) Avaliação de estimativas de proficiências através de métodos de estimação como MCMC (*Markov Chain Monte Carlo*) (Evensen, 2003).
- (M) Incorporação de métodos de otimização para a busca do fator de desconto ( $\phi$ ) ideal no modelo *Filtro de Kalman* (Fahrmeir e Kaufmann, 1991; Fang *et al.*, 2018; Julier e Uhlmann, 1997; Orderud, 2005; Wan e Van Der Merwe, 2000).
- (Q) Incorporação de métodos de calibração online de novos itens, utilizando modelos dinâmicos (Zheng, 2014).

- (M) Correção do padrão sistemático das estimativas de proficiência no modelo *Filtro de Kalman*.
- (Q) Estudo das estimativas cíclicas de proficiência do modelo *Power Steady* utilizando-se métodos auto-regressivos.
- (Q) Investigação das causas para desvios-padrão das estimativas de proficiência com valores inexpressivos no modelo *Power Steady*.
- (Q) Investigação de modelo Steady para estimação de proficiências, com base em distribuição de Poisson (Thattai, 2016).



# Referências Bibliográficas

- Andersen(1972)** Erling B Andersen. The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 42–54. Citado na pág. [28](#)
- Anderson(1958)** Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*, volume 2. John Wiley - New York. Citado na pág. [57](#)
- Andrade et al.(2000)** D. F. de Andrade, H. R. Tavares e da C. R. Valle. Teoria da resposta ao item: conceitos e aplicações. *ABE, São Paulo*. Citado na pág. [xiii](#), [12](#), [14](#), [17](#), [18](#), [19](#), [20](#), [22](#), [28](#), [41](#)
- Andrich(1978)** D. Andrich. A rating formulation for ordered response categories. *Psychometrika*, 43(4):561–573. Citado na pág. [20](#)
- Aral e Gupta(2016)** Ali Aral e Vijay Gupta. (p, q)-type beta functions of second kind. *Advances in Operator Theory*, 1(1):134–146. Citado na pág. [61](#)
- Araujo et al.(2009)** E. A. C. de Araujo, D. F. de Andrade e S. L. V. Bortolotti. Item response theory. *Revista da Escola de Enfermagem da USP*, 43(SPE):1000–1008. Citado na pág. [38](#)
- Azevedo(2003)** Caio Lucidius Naberezny Azevedo. Métodos de estimação na teoria de resposta ao item. Dissertação de Mestrado, Instituto de Matemática e Estatística da Universidade de São Paulo, 27/02/2003. Citado na pág. [22](#), [41](#)
- Baker e Kim(2004)** F. B. Baker e Seock-Ho Kim. *Item response theory: Parameter Estimation Techniques*. CRC Press. Citado na pág. [13](#), [14](#), [15](#), [19](#), [26](#), [28](#), [29](#), [32](#), [34](#), [35](#), [38](#), [40](#), [41](#), [175](#), [178](#)
- Barreto-Souza et al.(2010)** Wagner Barreto-Souza, Alessandro HS Santos e Gauss M Cordeiro. The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 80(2):159–172. Citado na pág. [68](#)
- Birnbaum(1968)** A. Birnbaum. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, páginas 395–479. Citado na pág. [27](#), [45](#), [49](#), [55](#)
- Birnbaum(1957)** Allan Birnbaum. Efficient design and use of tests of a mental ability for various decision-making problems. *Randolph Air Force Base, Texas: Air University, School of Aviation Medicine*, 26. Citado na pág. [14](#), [16](#)
- Bock(1972)** R. Darrell Bock. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1):29–51. Citado na pág. [20](#), [28](#)

- Bock e Aitkin(1981)** R. Darrell Bock e Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, 46(4): 443–459. Citado na pág. 29, 33, 39, 41, 45, 55, 175, 177, 178
- Bock e Lieberman(1970)** R. Darrell Bock e Marcus Lieberman. Fitting a response model for n dichotomously scored items. *Psychometrika*, 35(2):179–197. Citado na pág. 28, 30, 32, 33, 35, 55
- Bock e Mislevy(1982)** R. Darrell Bock e Robert J Mislevy. Adaptive eap estimation of ability in a microcomputer environment. *Applied psychological measurement*, 6(4):431–444. Citado na pág. 47, 48, 55, 87, 159
- Chang e Ying(1996)** Hua-Hua Chang e Zhiliang Ying. A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3):213–229. Citado na pág. 51, 52
- Chang e Ying(2009)** Hua-Hua Chang e Zhiliang Ying. Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *The Annals of Statistics*, páginas 1466–1488. Citado na pág. 21, 22, 50
- Chen et al.(2003)** Zhe Chen *et al.* Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69. Citado na pág. 57
- Costa(2009)** Denise Reis Costa. Métodos estatísticos em testes adaptativos informatizados. *Master's thesis, UFRJ-Universidade Federal do Rio de Janeiro*, 9. Citado na pág. 50, 87, 159
- Cuesta(1996)** M Cuesta. *Unidimensionalidade*. Madrid: Editorial Universitas. Citado na pág. 9, 10
- Das e Das(2009)** Sonali Das e Sourish Das. An adaptive bayesian analysis for binomial proportions: theory and methods. *South African Statistical Journal*, 43(2):195–218. Citado na pág. 66
- DasGupta(2011)** Anirban DasGupta. *Probability for statistics and machine learning: fundamentals and advanced topics*. Springer Science & Business Media. Citado na pág. 75, 77
- Dempster et al.(1977)** A. P Dempster, N. M. Laird e D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, páginas 1–38. Citado na pág. 22, 39, 178
- Embretson e Reise(2000)** S. E. Embretson e S. P. Reise. *Item response theory for psychologists*. Lawrence Erlbaum Associates, Inc. Citado na pág. 7, 19
- Evensen(2003)** Geir Evensen. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367. Citado na pág. 161
- Fahrmeir(1992)** Ludwig Fahrmeir. Posterior mode estimation by extended kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87(418):501–509. Citado na pág. 58
- Fahrmeir e Kaufmann(1991)** Ludwig Fahrmeir e Heinz Kaufmann. On kalman filtering, posterior mode estimation and fisher scoring in dynamic exponential family regression. *Metrika*, 38(1):37–60. Citado na pág. 161

- Fang et al.(2018)** Huazhen Fang, Ning Tian, Yebin Wang, MengChu Zhou e Mulugeta A Haile. Nonlinear bayesian estimation: from kalman filtering to a broader horizon. *IEEE-E/CAA Journal of Automatica Sinica*, 5(2):401–417. Citado na pág. 161
- Fernández(1990)** José Muñoz Fernández. *Teoría de la respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Pirámide. Citado na pág. 9
- Gradshteyn e Ryzhik(2007)** I. S. Gradshteyn e I. M. Ryzhik. *Table of integrals, series, and products*. Academic press, seventh edição. Citado na pág. 61, 63, 64
- Gulliksen(1950)** Harold Gulliksen. *Theory of mental tests*. John Wiley & Sons Inc. Citado na pág. 3
- Gupta(2000)** SC Gupta. *Fundamentals of mathematical statistics: A modern approach*. Sultan Chand and Sons. Citado na pág. 75
- Haley(1952)** David C Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Relatório técnico, Standford University - CA - Applied Mathematics and Statistics Labs. Citado na pág. 16
- Hambleton e Swaminathan(1985)** R. K. Hambleton e H. Swaminathan. *Item Response Theory: Principles and Applications*. Springer Netherlands, 1 edição. Citado na pág. 9, 10
- Hambleton et al.(1991)** R. K. Hambleton, H. Swaminathan, H. J. Rogers e R. M. Jaeger. *Fundamentals of item response theory* sage publications. Inc, Newbury Park, CA. Citado na pág. 19
- Harrison e Stevens(1971)** P. Jeff Harrison e C. F. Stevens. A bayesian approach to short-term forecasting. *Operational Research Quarterly*, 22(4):341–362. Citado na pág. 58
- Harrison e Stevens(1976)** P. Jeff Harrison e C. F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society Series B*, 38(3):205–247. Citado na pág. 58, 159
- Harvey(1989)** A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press. Citado na pág. ix, xi, 1, 56, 58, 72, 159
- Hattie(1984)** John Hattie. An empirical study of various indices for determining unidimensionality. *Multivariate behavioral research*, 19(1):49–78. Citado na pág. 10
- Hildebrand(1956)** F. B. Hildebrand. *Introduction To Numerical Analysis*. McGraw-Hill Publishing Company Ltda.; New York. Citado na pág. 32
- Ho e Lee(1964)** Yu-Chi Ho e Robert C. K. Lee. A bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9(4):333–339. Citado na pág. 58
- Julier e Uhlmann(1997)** Simon J Julier e Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. Em *Signal processing, sensor fusion, and target recognition VI*, volume 3068, páginas 182–193. International Society for Optics and Photonics. Citado na pág. 161
- Kalman(1960)** Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45. Citado na pág. 57, 58

- Kolakowski e Bock(1973)** D Kolakowski e RD Bock. Normog: Maximum likelihood item analysis and test scoring: Normal ogive model. *Ann Arbor: National Educational Resources*. Citado na pág. 28
- Kolmogorov(1941)** AN Kolmogorov. Stationary sequences in hilbert space (russia). *Bull. Moscow State Univ. Math*, página 40. Citado na pág. 57
- Kullback(1997)** Solomon Kullback. *Information theory and statistics*. Courier Corporation. Citado na pág. ix, xi, 1
- Larson e Madsen(1985)** Jerry W Larson e Harold S Madsen. Computerized adaptive language testing: Moving beyond computer-assisted testing. *CALICO Journal*, 2(3):32–36. Citado na pág. 1
- Le(2013)** Dai-Trang Le. Applying item response theory modeling in educational research. *ProQuest LLC*. Citado na pág. 8
- Lehmann e Casella(1998)** Erich Leo Lehmann e George Casella. *Theory of Point Estimation*. Springer. Citado na pág. 17, 51
- Leung et al.(2002)** Chi-Keung Leung, Hua-Hua Chang e Kit-Tai Hau. Item selection in computerized adaptive testing: Improving the a-stratified design with the sympson-hetter algorithm. *Applied Psychological Measurement*, 26(4):376–392. Citado na pág. 161
- Lindley e Smith(1972)** Dennis V Lindley e Adrian FM Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, páginas 1–41. Citado na pág. 35
- Lord(1980)** F. M. Lord. *Applications of item response theory to practical testing problems*. Routledge. Citado na pág. 1, 19, 47, 55
- Lord e Novick(1968)** F. M. Lord e M. R. Novick. *Statistical Theories of Mental Test Scores*. Addison-Wesley. Citado na pág. 11, 13, 14, 15
- Lord(1974)** Frederic M Lord. Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2):247–264. Citado na pág. 28
- Lord(1977)** Frederic M Lord. A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 1(1):95–100. Citado na pág. 49
- Lord(1984)** Frederic M Lord. Maximum likelihood and bayesian parameter estimation in item response theory. *ETS Research Report Series*, 1984(2). Citado na pág. 47
- Magis e Barrada(2017)** D. Magis e J. R. Barrada. Computerized adaptive testing with r: Recent updates of the package catr. *Journal of Statistical Software*, 76(1):1–19. Citado na pág. 1, 45
- Magis et al.(2012)** David Magis, Gilles Raïche et al. Random generation of response patterns under computerized adaptive testing with the r package catr. *Journal of Statistical Software*, 48(8):1–31. Citado na pág. 45
- Masters(1982)** G. N. Masters. A rasch model for partial credit scoring. *Psychometrika*, 47(2):149–174. Citado na pág. 19

- Meinhold e Singpurwalla(1983)** Richard J. Meinhold e Nozer D. Singpurwalla. Understanding the kalman filter. *The American Statistician*, 37(2):123–127. Citado na pág. 56, 57, 58
- Mihesan(2004)** V Mihesan. The beta approximating operators of second kind. *Studia Univ, Babes-Bolyai, Mathematica, XLIX*, 2:79–88. Citado na pág. 61, 75
- Mills et al.(2005)** Craig N Mills, Maria T Potenza, John J Fremer e William C Ward. *Computer-based testing: Building the foundation for future assessments*. Routledge. Citado na pág. 1
- Mislevy(1986)** R. J. Mislevy. Bayes modal estimation in item response models. *Psychometrika*, 51(2):177–195. Citado na pág. 35, 38, 47
- Mislevy e Bock(1982)** R. J. Mislevy e RD Bock. Bilog, maximum likelihood item analysis and test scoring: Logistic model [computer software]. mooresville, in: Scientific software, 1982. Citado na pág. 43
- Mislevy e Bock(1990)** Robert J Mislevy e R Darrell Bock. *BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models*. Scientific Software International. Citado na pág. 42
- Mislevy e Stocking(1989)** Robert J. Mislevy e Martha L. Stocking. A consumer's guide to logist and bilog. *ETS Research Report Series*, 1989(2). Citado na pág. 34, 43
- Mislevy e Verhelst(1990)** Robert J Mislevy e Norman Verhelst. Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55(2):195–215. Citado na pág. 3
- Muraki(1992)** E. Muraki. A generalized partial credit model: Application of an em algorithm. *ETS Research Report Series*, 16:159–176. Citado na pág. 19
- Nadarajah(2005)** S Nadarajah. Exponentiated beta distributions. *Computers & Mathematics with Applications*, 49(7-8):1029–1035. Citado na pág. 61, 63, 66, 67, 68, 75, 76
- Neyman e Scott(1948)** Jerzy Neyman e Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, páginas 1–32. Citado na pág. 28
- Orderud(2005)** Fredrik Orderud. Comparison of kalman filter estimation approaches for state space models with nonlinear measurements. Em *Proc. of Scandinavian Conference on Simulation and Modeling*, páginas 1–8. Citado na pág. 161
- Owen(1975)** Roger J Owen. A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70 (350):351–356. Citado na pág. 50, 51, 55
- Pasquali(2009)** Luiz Pasquali. *Psicometria: teoria dos testes na psicologia e na educação*. Vozes. Citado na pág. 9
- Samejima(1969)** F. A. Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17. Citado na pág. 14, 19
- Samejima(1997)** F. A. Samejima. Graded response model. *Handbook of Modern Item Response Theory*. Citado na pág. 19



- Smith(1979)** J. Q. Smith. A generalization of the bayesian steady forecasting model. *Journal of the Royal Statistical Society - Series B (Methodological)*, 41(3):375–387. Citado na pág. [ix](#), [xi](#), [58](#), [75](#), [79](#), [80](#), [83](#), [91](#), [159](#)
- Spragins(1965)** J. Spragins. A note on the iterative application of bayes' rule. *IEEE Transactions on Information Theory*, 11(4):544–549. Citado na pág. [58](#)
- Stancu(1995)** DD Stancu. On the beta approximating operators of second type kind. *Rev. Anal. Numér. Théor. Approx.*, páginas 231–239. Citado na pág. [61](#), [75](#)
- Stroud e Secrest(1966)** Arthur H Stroud e Don Secrest. *Gaussian Quadrature Formulas*. Prentice-Hall. Citado na pág. [32](#), [48](#)
- Swaminathan e Gifford(1986)** Hariharan Swaminathan e Janice Gifford. Bayesian estimation in the three-parameter logistic model. *Psychometrika*, 51(4):589–601. Citado na pág. [33](#), [34](#), [39](#)
- Thattai(2016)** Mukund Thattai. Universal poisson statistics of mrnas with complex decay pathways. *Biophysical journal*, 110(2):301–305. Citado na pág. [162](#)
- Van Der Linden e Hambleton(1997)** W. J. Van Der Linden e R. K. Hambleton. Item response theory: Brief history, common models, and extensions. Em *Handbook of modern item response theory*, páginas 1–28. Springer. Citado na pág. [59](#)
- van der Linden e Pashley(2000)** W. J. van der Linden e P. J. Pashley. Item selection and ability estimation in adaptive testing. Em W. J. van der Linden e C. A. W. Glas, editors, *Computerized Adaptive Testing: Theory and Practice*, chapter 1, páginas 1–25. Kluwer Academic, New York. Citado na pág. [47](#)
- van der Linden(1996)** Wim J van der Linden. Bayesian item selection criteria for adaptive testing. research report 96-01. *Educational Resources Information Center*. Citado na pág. [51](#)
- van der Linden(2003)** Wim J van der Linden. Some alternatives to sympson-hetter item-exposure control in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 28(3):249–265. Citado na pág. [161](#)
- van der Linden e Glas(2000)** Wim J van der Linden e Cees AW Glas. *Computerized Adaptive Testing: Theory and Practice*. Kluwer Academic. Citado na pág. [2](#), [46](#), [49](#)
- van der Linden e Glas(2010)** Wim J van der Linden e Cees AW Glas. *Elements of Adaptive Testing*. Springer. Citado na pág. [49](#), [50](#), [51](#), [53](#)
- Veerkamp e Berger(1997)** Wim J. J. Veerkamp e Martijn P. F. Berger. Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22:203–226. Citado na pág. [52](#), [53](#)
- Vitória et al.(2006)** Florbela Vitória, Leandro S Almeida e Ricardo Primi. Unidimensionalidade em testes psicológicos: conceito, estratégias e dificuldades na sua avaliação. *Psic: revista da Vetor Editora*, 7(1):01–07. Citado na pág. [9](#), [10](#), [11](#)
- Wainer(2000)** Howard Wainer. *Computerized adaptive testing*. Wiley Online Library. Citado na pág. [3](#)

- Wainer e Kiely(1987)** Howard Wainer e Gerard L Kiely. Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational measurement*, 24(3):185–201. Citado na pág. [2](#)
- Wan e Van Der Merwe(2000)** Eric A Wan e Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. Em *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, páginas 153–158. Ieee. Citado na pág. [161](#)
- Wang et al.(2002)** Xiaohui Wang, Eric T Bradlow e Howard Wainer. A general bayesian model for testlets: Theory and applications. *ETS Research Report Series*, 2002(1). Citado na pág. [2](#)
- Warm(1989)** Thomas A. Warm. Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3):427–450. Citado na pág. [53](#)
- West et al.(1985)** Mike West, Jeff P. Harrison e Helio S. Migon. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83. Citado na pág. [58](#)
- Wiener(1949)** Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Time Series with Engineering Applications*. John Wiley & Sons; UK. Citado na pág. [57](#)
- Wood et al.(1976)** Robert L Wood, Marilyn S Wingersky e Frederic M Lord. *Logist: a Computer Program for Estimating Examinee Ability and Item Characteristic Curve Parameters*. Educational Testing Service. Citado na pág. [27](#), [28](#)
- Wright e Panchapakesan(1969)** Benjamin Wright e Nargis Panchapakesan. A procedure for sample-free item analysis. *Educational and Psychological measurement*, 29(1):23–48. Citado na pág. [28](#)
- Wright(1977)** Benjamin D Wright. Solving measurement problems with the rasch model. *Journal of Educational measurement*, 14(2):97–116. Citado na pág. [28](#)
- Wright e Stone(2004)** Benjamin Drake Wright e Mark H Stone. *Making measures*. Ph-neron Press. Citado na pág. [9](#)
- Zheng(2014)** Yi Zheng. *New methods of Online Calibration for Item Bank Replenishment*. Ph.d. dissertation, University of Illinois at Urbana-Champaign. Citado na pág. [9](#), [18](#), [21](#), [46](#), [161](#)





# Apêndices



# Teoria da Resposta ao Item

## A Solução de Bock & Lieberman

Na seção (2.3.2.4) para chegarmos à expressão (2.73), precisamos notar que na expressão (2.72), o produto sobre  $h \neq t$  ocorre em função da retirada dos termos  $P_t(\theta)$  e  $Q_t(\theta)$  quando tomadas as derivadas com respeito à  $\alpha_t$ , de forma que a derivada fica expressa por

$$\frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] = \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] \left[ \frac{\partial P_t(\theta)}{\partial \alpha_t} \right] \left[ \frac{u_{jt} - P_t(\theta)}{P_t(\theta) Q_t(\theta)} \right]. \quad (\text{A.1})$$

Apresenta-se a seguir o desenvolvimento do raciocínio, considerando inicialmente  $v = \alpha_t(\theta - \beta_t)$ , de modo que

$$\begin{aligned} \frac{\partial P_t(\theta)}{\partial \alpha_t} &= (1 - c_t)[e^v(-1)(1 + e^v)^{-2}e^v(\theta - \beta_t) + (1 + e^v)^{-1}e^v(\theta - \beta_t)] \\ &= (1 - c_t)(\theta - \beta_t)P_t^*(\theta)Q_t^*(\theta) \\ &= K \end{aligned} \quad (\text{A.2})$$

Definida a derivada da CCI, considerando um modelo de 3 parâmetros, em relação ao parâmetro do item  $\alpha_t$ , tem-se

$$\begin{aligned} &\frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] \\ &= \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \times \\ &\quad \left[ P_t(\theta)^{u_{jt}} \frac{\partial}{\partial \alpha_t} [Q_t(\theta)^{1-u_{jt}}] + Q_t(\theta)^{1-u_{jt}} \frac{\partial}{\partial \alpha_t} [P_t(\theta)^{u_{jt}}] \right] \\ &= \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \times \\ &\quad [P_t(\theta)^{u_{jt}}(1 - u_{jt})Q_t(\theta)^{1-u_{jt}-1}(-K) + Q_t(\theta)^{1-u_{jt}}u_{jt}P_t(\theta)^{u_{jt}-1}(K)] \\ &= K \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \times \\ &\quad [Q_t(\theta)^{1-u_{jt}}u_{jt}P_t(\theta)^{u_{jt}-1} - P_t(\theta)^{u_{jt}}(1 - u_{jt})Q_t(\theta)^{1-u_{jt}-1}]. \end{aligned} \quad (\text{A.3})$$

Reescrevendo o segundo termo na expressão (A.3) por:

$$(-1)^{u_{jt}+1} = \begin{cases} 1 & \text{se } u_{jt} = 1 \\ -1 & \text{se } u_{jt} = 0. \end{cases} \quad (\text{A.4})$$

Portanto, substituindo (A.2) e (A.4) em (A.3), têm-se

$$\begin{aligned}
& \frac{\partial}{\partial \alpha_t} \left[ \prod_{t=1}^n P_t(\theta)^{u_{jt}} Q_t(\theta)^{1-u_{jt}} \right] \\
&= (-1)^{u_{jt}+1} (1-c_t) \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \times \\
& [e^v (-1) (1+e^v)^{-2} e^v (\theta - \beta_t) + (1+e^v)^{-1} e^v (\theta - \beta_t)] \\
&= (-1)^{u_{jt}+1} (1-c_t) \left[ \prod_{h \neq t}^n P_h(\theta)^{u_{jh}} Q_h(\theta)^{1-u_{jh}} \right] \times \\
& (\theta - \beta_t) P_t^*(\theta) Q_t^*(\theta). \tag{A.5}
\end{aligned}$$

Voltando com o índice  $j$  e substituindo (A.5) em (2.69), têm-se

$$\begin{aligned}
\frac{\partial}{\partial \alpha_t} (\log \mathbf{L}) &= \sum_{j=1}^N (-1)^{u_{jt}+1} (1-c_t) \int_{\mathfrak{R}} (\theta_j - \beta_t) \left[ \frac{1}{\prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] \times \\
& \left[ \prod_{h \neq t}^n P_h(\theta_j)^{u_{jh}} Q_h(\theta_j)^{1-u_{jh}} \right] P_t^*(\theta_j) Q_t^*(\theta_j) [P_t(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \tau)] d\theta \\
&= \sum_{j=1}^N (-1)^{u_{jt}+1} (1-c_t) \int_{\mathfrak{R}} (\theta_j - \beta_t) \left[ \frac{1}{\prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] \left[ \frac{P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}}{P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] \times \\
& \left[ \prod_{h \neq t}^n P_h(\theta_j)^{u_{jh}} Q_h(\theta_j)^{1-u_{jh}} \right] P_t^*(\theta_j) Q_t^*(\theta_j) [P_t(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \tau)] d\theta \\
&= \sum_{j=1}^N (1-c_t) \int_{\mathfrak{R}} (\theta_j - \beta_t) \left[ \frac{\prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}}{\prod_{t=1}^n P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] \left[ \frac{(-1)^{u_{jt}+1}}{P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] \times \\
& P_t^*(\theta_j) Q_t^*(\theta_j) [P_t(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \tau)] d\theta \\
&= (1-c_t) \sum_{j=1}^N \int_{\mathfrak{R}} (\theta_j - \beta_t) \left[ \frac{(-1)^{u_{jt}+1}}{P_t(\theta_j)^{u_{jt}} Q_t(\theta_j)^{1-u_{jt}}} \right] P_t^*(\theta_j) Q_t^*(\theta_j) [P_t(\theta_j | \mathbf{u}_j, \boldsymbol{\zeta}, \tau)] d\theta. \tag{A.6}
\end{aligned}$$

Lembrando da relação entre o segundo termo entre colchetes em (A.3) e a expressão obtida em (A.4), resolve-se o termo entre colchetes em (A.6), obtendo

$$\begin{aligned}
\frac{(-1)^{u_{jt}+1}}{P_t(\theta_j)^{u_{jt}}Q_t(\theta_j)^{1-u_{jt}}} &= \left[ \frac{u_{jt}P_t(\theta_j)^{u_{jt}-1}}{P_t(\theta_j)^{u_{jt}}} - \frac{(1-u_{jt})Q_t(\theta_j)^{1-u_{jt}-1}}{Q_t(\theta_j)^{1-u_{jt}}} \right] \\
&= \left[ \frac{u_{jt}}{P_t(\theta_j)} - \frac{(1-u_{jt})}{Q_t(\theta_j)} \right] \\
&= \left[ \frac{Q_t(\theta_j)u_{jt} - (1-u_{jt})P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right] \\
&= \left[ \frac{Q_t(\theta_j)u_{jt} - P_t(\theta_j) + u_{jt}P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right] \\
&= \left[ \frac{u_{jt}[Q_t(\theta_j) + P_t(\theta_j)] - P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right] \\
&= \left[ \frac{u_{jt} - P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right]. \tag{A.7}
\end{aligned}$$

Substituindo (A.7) em (A.6), resulta em

$$\begin{aligned}
\frac{\partial}{\partial \alpha_t}(\log \mathbf{L}) &= (1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} (\theta_j - \beta_t) \left[ \frac{u_{jt} - P_t(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right] P_t^*(\theta_j)Q_t^*(\theta_j)[P_t(\theta_j|\mathbf{u}_j, \boldsymbol{\zeta}, \tau)]d\theta \\
&= (1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} (\theta_j - \beta_t)[u_{jt} - P_t(\theta_j)] \left[ \frac{P_t^*(\theta_j)Q_t^*(\theta_j)}{P_t(\theta_j)Q_t(\theta_j)} \right] \times \\
&\quad [P_t(\theta_j|\mathbf{u}_j, \boldsymbol{\zeta}, \tau)]d\theta. \tag{A.8}
\end{aligned}$$

Por fim, lembrando que a expressão entre colchetes em (A.8) é a mesma apresentada anteriormente em (2.39), e fazendo a devida substituição, tem-se

$$\frac{\partial}{\partial \alpha_t}(\log \mathbf{L}) = (1 - c_t) \sum_{j=1}^N \int_{\mathfrak{R}} [u_{jt} - P_t(\theta_j)](\theta_j - \beta_t)W_{jt}[P_t(\theta_j|\mathbf{u}_j, \boldsymbol{\zeta}, \tau)]d\theta. \tag{A.9}$$

Finalizando assim, o raciocínio que apresenta as etapas para encontrar a equação de log-verossimilhança do procedimento de estimação MMLLE obtido em (2.73). Desenvolvimento semelhante é realizado para as expressões obtidas em (2.74) e (2.75).

## B Solução de Bock & Aitkin

Segundo Baker e Kim (2004), a reformulação das equações de Bock & Lieberman por (Bock e Aitkin, 1981) resultaram numa solução que é computacionalmente plausível e, sob a suposição de que a distribuição da população é conhecida ou concorrentemente estimada com a especificação correta, produz consistentes estimativas dos parâmetros dos itens. Tal como no caso do paradigma de Birnbaum do JMLE, Bock & Aitkin assumiu-se que os itens e os examinandos são independentes. Consequentemente, os parâmetros dos itens podem ser estimados um por vez, com a habilidade dos examinandos estimada um examinando por vez. A reformulação dos resultados resultou em equações de verossimilhança para um item que teve a mesma forma geral das equações (2.48) a (2.48). Para associar às soluções mais formais do MMLLE, inicialmente redistribuiu-se os termos da equação para o parâmetro  $\alpha_i$ , apresentada em (2.76),

resultando em

$$\alpha_t : (1 - c_t) \sum_{k=1}^q (X_k - \beta_t) \times \left[ \sum_{j=1}^N u_{jt} P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) - P_t(X_k) \sum_{j=1}^N P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) \right] W_{tk} = 0. \quad (\text{B.1})$$

Em MMLE, reescreve-se a distribuição das habilidades a posteriori de um examinando que tem nível de habilidade  $X_k$ , utilizando a equação (2.61) em sua forma de quadratura, resultando em

$$P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) = \frac{\prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)}{\sum_{k=1}^q \prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)}. \quad (\text{B.2})$$

De forma que a CCI em forma de quadratura para o modelo de 3 parâmetros é reescrita como

$$P_t(X_k) = c_t(1 - c_t) \frac{\exp[\alpha_t(X_k - \beta_t)]}{1 + \exp[\alpha_t(X_k - \beta_t)]}. \quad (\text{B.3})$$

Portanto,  $Q_t(X_k) = 1 - P_t(X_k)$ , segue do segundo elemento dentro de colchetes da equação (B.1) que podemos definir o termo

$$\begin{aligned} \bar{f}_{tk} &= \sum_{j=1}^N P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) \\ &= \sum_{j=1}^N \left[ \frac{\prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)}{\sum_{k=1}^q \prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)} \right]. \end{aligned} \quad (\text{B.4})$$

É possível notar que (B.4) simplifica a soma sobre os  $N$  examinandos da probabilidade à posteriori de um examinando que tenha habilidade no intervalo de quadratura  $X_k$ , lembrando que este representa o ponto médio de cada retângulo sobre a escala de habilidade.

Da mesma forma, utiliza-se o primeiro termo dentro de colchetes na equação (B.1) para definir

$$\begin{aligned} \bar{r}_{tk} &= \sum_{j=1}^N u_{jt} P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) \\ &= \sum_{j=1}^N \left[ \frac{\prod_{t=1}^n u_{tk} P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)}{\sum_{k=1}^q \prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}} A(X_k)} \right]. \end{aligned} \quad (\text{B.5})$$

A nomenclatura dada por [Bock e Aitkin \(1981\)](#) para

$$L(X_k) = \prod_{t=1}^n P_t(X_k)^{u_{jt}} Q_t(X_k)^{1-u_{jt}}, \quad (\text{B.6})$$

representava a forma de quadratura da probabilidade condicional de  $\mathbf{u}_j$ , dado  $\theta = X_k$  e os parâmetros dos itens, de forma que a expressão (B.2) pode ser reescrita como

$$P(X_k | \mathbf{u}_j, \boldsymbol{\zeta}, \boldsymbol{\tau}) = \frac{L(X_k)A(X_k)}{\sum_{k=1}^q L(X_k)A(X_k)}. \quad (\text{B.7})$$

Portanto, as expressões simplificadas para  $\bar{f}_{tk}$  e  $\bar{r}_{tk}$  são dadas por

$$\bar{f}_{tk} = \sum_{j=1}^N \left[ \frac{L(X_k)A(X_k)}{\sum_{k=1}^q L(X_k)A(X_k)} \right]. \quad (\text{B.8})$$

e

$$\bar{r}_{tk} = \sum_{j=1}^N \left[ \frac{u_{jt}L(X_k)A(X_k)}{\sum_{k=1}^q L(X_k)A(X_k)} \right]. \quad (\text{B.9})$$

Em função da importância conceitual, os termos  $\bar{f}_{tk}$  e  $\bar{r}_{tk}$  necessitam ser detalhados. Neste sentido, os  $\bar{f}_{tk}$  são os números de examinandos na população de tamanho  $N$  a qual se espera ter nível de habilidade  $X_k$ . Isto se deve ao fato da distribuição de examinandos sob os  $q$  nós de quadratura em proporção à distribuição a posteriori dos examinandos estarem em cada nó de quadratura.

Os  $\bar{r}_{tk}$  são os números de examinandos na população de habilidade  $X_k$  que se espera responder corretamente ao item. Como as equações (B.8) e (B.9) são as mesmas, a menos do termo  $u_{jt}$ , que assume valores 0 ou 1, somente os termos correspondentes às respostas corretas são acumulados. Portanto, fixado o  $k$ -ésimo nó de quadratura, a frequência de repostas corretas é obtida pela equação (B.9).

Assim, a equações (2.76) a (2.78) de log-verossimilhança marginal podem ser reescritas por:

$$\alpha_t : (1 - c_t) \sum_{k=1}^q (X_k - \beta_t) [\bar{r}_{tk} - \bar{f}_{tk} P_t(X_k)] W_{tk} = 0, \quad (\text{B.10})$$

$$\beta_t : -\alpha_t (1 - c_t) \sum_{k=1}^q [\bar{r}_{tk} - \bar{f}_{tk} P_t(X_k)] W_{tk} = 0, \quad (\text{B.11})$$

$$c_t : (1 - c_t)^{-1} \sum_{k=1}^q \left[ \frac{\bar{r}_{tk} - \bar{f}_{tk} P_t(X_k)}{P_t(X_k)} \right] = 0. \quad (\text{B.12})$$

Particularidades desta abordagem em relação ao método apresentado na seção (2.3.2.3) é que, de fato, as equações de log-verossimilhança são idênticas na forma, sendo consideradas equações de verossimilhança para uma análise probit para o item  $t$  no qual  $X_k$  é variável

independente. A diferença é que no modelo MLE o número total de examinandos  $f_{tk}$  correspondem ao nível de *habilidade* ( $\theta_k$ ) e o número de respostas corretas  $r_{tk}$  são quantidades desconhecidas.

Na solução MMLE, ambas as quantidades são desconhecidas, e o número esperado de examinandos respondendo ao item  $t$  e o número esperado de respostas corretas são substituídos por estas quantidades desconhecidas em cada nó de quadratura ( $X_k$ ). Assim, na literatura da TRI, as quantidades  $\bar{f}_{tk}$  e  $\bar{r}_{tk}$  são conhecidas como “*dados artificiais*”, uma vez que são criados pelas equações (B.8) e (B.9).

No estágio de estimação dos parâmetros dos itens do procedimento JMLE, os  $f_{tk}$  e  $r_{tk}$  são conhecidos. Entretanto, sobre a aproximação MMLE, os valores de  $\bar{f}_{tk}$  e  $\bar{r}_{tk}$  dependem dos valores estimados dos parâmetros dos itens. Porém, as equações (B.10) a (B.12) não fazem qualquer provisão para esta dependência. Consequentemente, um paradigma necessita ser estabelecido no qual ambos os “*dados artificiais*” e as estimativas dos parâmetros dos itens possam ser obtidos iterativamente. Neste sentido, Bock e Aitkin (1981) empregaram o algoritmo EM desenvolvido por Dempster *et al.* (1977). Os detalhes da utilização do algoritmo EM na abordagem de Bock & Aitkin podem ser vistos em Baker e Kim (2004, pág. 169-176).



# CAT via Filtro de Kalman Não-Linear

## C Memória de Cálculo

Nesta seção, demonstram-se os detalhes dos cálculos apresentados no capítulo 4.

Na seção 4.5, expressão (4.35), temos:

$$\begin{aligned} P(X = 0) &= \frac{(1-c)}{B(\alpha, \beta)} B(\alpha + 1, \beta) F(0, \alpha + 1, \alpha + \beta + 1, 1 - c) \\ &= (1-c) \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} \\ &= (1-c) \frac{\alpha}{\alpha + \beta} \\ &= \frac{\alpha}{\alpha + \beta} - c \frac{\alpha}{\alpha + \beta}. \end{aligned} \tag{C.1}$$

Na seção 4.5, expressão (4.36), temos:

$$\begin{aligned} P(X = 1) &= F(-1, \alpha, \alpha + \beta, 1 - c) \\ &= 1 - (1-c) \frac{\alpha}{\alpha + \beta} \\ &= \frac{\alpha + \beta - \alpha + c\alpha}{\alpha + \beta} \\ &= \frac{\beta}{\alpha + \beta} + c \frac{\alpha}{\alpha + \beta}. \end{aligned} \tag{C.2}$$

Na seção 4.5, expressão (4.37), temos:

$$\begin{aligned}
f(\eta|x) &= \frac{P(x|\eta)f(\eta)}{P(X)} \\
&= \frac{[1 + ce^{-\eta}]^x [(1-c)e^{-\eta}]^{1-x}}{1 + e^{-\eta}} \frac{1}{B(\alpha, \beta)} \frac{e^{-\alpha\eta}}{[1 + e^{-\eta}]^{\alpha+\beta}} \times \\
&\quad \frac{B(\alpha, \beta)}{1} \\
&= \frac{(1-c)^{1-x} B(\alpha + 1 - x, \beta) F(-x, \alpha + 1 - x, \alpha + \beta + 1 - x, 1 - c)}{[1 + ce^{-\eta}]^x [1 + e^{-\eta}]^{\alpha+\beta}} \times \\
&\quad \frac{1}{1 + e^{-\eta}} \frac{1}{[1 + e^{-\eta}]^{\alpha+\beta}} \times \\
&= \frac{(1-c)^{1-x} B(\alpha + 1 - x, \beta) F(-x, \alpha + 1 - x, \alpha + \beta + 1 - x, 1 - c)}{[1 + ce^{-\eta}]^x e^{-\eta(\alpha+1-x)}} \times \\
&\quad \frac{1}{[1 + e^{-\eta}]^{\alpha+\beta+1}} \times \\
&= \frac{1}{B(\alpha + 1 - x, \beta) F(-x, \alpha + 1 - x, \alpha + \beta + 1 - x, 1 - c)} \\
&\quad \frac{[1 + ce^{-\eta}]^x e^{-\eta(\alpha+1-x)}}{B(\alpha + 1 - x, \beta) F(-x, \alpha + 1 - x, \alpha + \beta + 1 - x, 1 - c)} \times \\
&= \frac{1}{[1 + e^{-\eta}]^{\alpha+\beta+1}}. \tag{C.3}
\end{aligned}$$

Na seção 4.5, expressão (4.39), temos:

$$\begin{aligned}
f(\eta|x=1) &= \frac{e^{-\eta\alpha} [1 + ce^{-\eta}]}{B(\alpha, \beta) F(-1, \alpha, \alpha + \beta, 1 - c) [1 + e^{-\eta}]^{\alpha+\beta+1}} \\
&= \frac{e^{-\eta\alpha} [1 + ce^{-\eta}]}{B(\alpha, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} \frac{\alpha + \beta}{\beta + c\alpha} \\
&= \frac{\alpha + \beta}{\beta + c\alpha} \left\{ \frac{e^{-\eta\alpha}}{B(\alpha, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} + c \frac{e^{-\eta(\alpha+1)}}{B(\alpha, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} \right\} \\
&= \frac{\alpha + \beta}{\beta + c\alpha} \left\{ \frac{B(\alpha, \beta + 1)}{B(\alpha, \beta)} \frac{e^{-\eta\alpha}}{B(\alpha, \beta + 1) [1 + e^{-\eta}]^{\alpha+\beta+1}} + \right. \\
&\quad \left. + c \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} \frac{e^{-\eta(\alpha+1)}}{B(\alpha + 1, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} \right\} \\
&= \frac{\alpha + \beta}{\beta + c\alpha} \left\{ \frac{\beta}{\alpha + \beta} \frac{e^{-\eta\alpha}}{B(\alpha, \beta + 1) [1 + e^{-\eta}]^{\alpha+\beta+1}} + c \frac{\alpha}{\alpha + \beta} \frac{e^{-\eta(\alpha+1)}}{B(\alpha + 1, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} \right\} \\
&= \frac{1}{\beta + c\alpha} \left\{ \beta \frac{e^{-\eta\alpha}}{B(\alpha, \beta + 1) [1 + e^{-\eta}]^{\alpha+\beta+1}} + c\alpha \frac{e^{-\eta(\alpha+1)}}{B(\alpha + 1, \beta) [1 + e^{-\eta}]^{\alpha+\beta+1}} \right\}. \tag{C.4}
\end{aligned}$$

Na seção 4.5, expressão (4.43), temos:

$$\begin{aligned}
\text{Var}(\eta_t | X_t = 0) &= \Psi'(\alpha + 1) + \Psi'(\beta), \\
&= \Psi'(\alpha) + \Psi'(\beta) - \frac{1}{\alpha^2}. \tag{C.5}
\end{aligned}$$

Na seção 4.5, expressão (4.44), temos:

$$\begin{aligned}
E(\eta^2|1) &= \frac{\beta}{\beta + c\alpha} \left\{ \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \frac{\Gamma''(\beta + 1)}{\Gamma(\beta + 1)} - 2\Psi(\alpha)\Psi(\beta + 1) \right\} + \\
&\quad + \frac{c\alpha}{\beta + c\alpha} \left\{ \frac{\Gamma''(\alpha + 1)}{\Gamma(\alpha + 1)} + \frac{\Gamma''(\beta)}{\Gamma(\beta)} - 2\Psi(\alpha + 1)\Psi(\beta) \right\} \\
&= \frac{\beta}{\beta + c\alpha} \{ \Psi'(\alpha)\Psi^2(\alpha) + \Psi'(\beta + 1)\Psi^2(\beta + 1) - 2\Psi(\alpha)\Psi(\beta + 1) \} + \\
&\quad + \frac{c\alpha}{\beta + c\alpha} \{ \Psi'(\alpha + 1)\Psi^2(\alpha + 1) + \Psi'(\beta)\Psi^2(\beta) - 2\Psi(\alpha + 1)\Psi(\beta) \} \\
&= \frac{\beta}{\beta + c\alpha} \left\{ \Psi'(\alpha) + \Psi^2(\alpha) + \Psi'(\beta) - \frac{1}{\beta^2} + \Psi^2(\beta) + \right. \\
&\quad \left. + \frac{2}{\beta}\Psi(\beta) + \frac{1}{\beta^2} - 2\Psi(\alpha)\Psi(\beta) - \frac{2}{\beta}\Psi(\alpha) \right\} + \\
&\quad + \frac{c\alpha}{\beta + c\alpha} \left\{ \Psi'(\alpha) - \frac{1}{\alpha^2} + \Psi^2(\alpha) + \frac{2}{\alpha}\Psi(\alpha) + \frac{1}{\alpha^2} - \Psi'(\beta) + \right. \\
&\quad \left. + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) - \frac{2}{\alpha}\Psi(\beta) \right\} \\
&= \Psi'(\alpha) + \Psi^2(\alpha) + \Psi'(\beta) + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) + \\
&\quad + \frac{1}{\beta + c\alpha} \{ 2\Psi(\beta) - 2\Psi(\alpha) \} - \frac{c}{\beta + c\alpha} [-2\Psi(\alpha) + 2\Psi(\beta)] \\
&= \Psi'(\alpha) + \Psi'(\beta) + \Psi^2(\alpha) + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) + \\
&\quad + 2\frac{(1-c)}{\beta + c\alpha} \{ \Psi(\beta) - \Psi(\alpha) \}. \tag{C.6}
\end{aligned}$$

Na seção 4.5, expressão (4.45), temos:

$$\begin{aligned}
E^2[\eta_t|X_t = 1] &= \left( \Psi(\beta) - \Psi(\alpha) + \frac{(1 - c_t)}{\beta + c_t\alpha} \right)^2, \\
&= \Psi^2(\alpha) + \Psi^2(\beta) - 2\Psi(\alpha)\Psi(\beta) + 2\frac{(1 - c_t)}{\beta + c_t\alpha} [\Psi(\beta) - \Psi(\alpha)] + \left[ \frac{1 - c_t}{\beta + c_t\alpha} \right]^2. \tag{C.7}
\end{aligned}$$

Na seção 4.5, expressão (4.46), temos:

$$\begin{aligned}
E(\eta|1) &= \frac{\beta}{\beta + c\alpha} [\Psi(\beta + 1) - \Psi(\alpha)] + \frac{c\alpha}{\beta + c\alpha} [\Psi(\beta) - \Psi(\alpha + 1)] \\
&= \frac{\beta}{\beta + c\alpha} \left[ \Psi(\beta) - \Psi(\alpha) + \frac{1}{\beta} \right] + \frac{c\alpha}{\beta + c\alpha} \left[ \Psi(\beta) - \Psi(\alpha) - \frac{1}{\alpha} \right] \\
&= \left[ \frac{\beta}{\beta + c\alpha} + \frac{c\alpha}{\beta + c\alpha} \right] [\Psi(\beta) - \Psi(\alpha)] + \frac{1 - c}{\beta + c\alpha} \\
&= [\Psi(\beta) - \Psi(\alpha)] + \frac{1 - c}{\beta + c\alpha}. \tag{C.8}
\end{aligned}$$

Na seção 4.5.1, expressão (4.48), temos:

$$\begin{aligned}
 \pi + \pi e^{-\eta} - e^{-\eta} &= 0 \\
 (\pi - 1)e^{-\eta} &= -\pi \\
 e^{-\eta} &= \frac{\pi}{1 - \pi} \\
 -\eta &= \ln \frac{\pi}{1 - \pi} \\
 \eta_t &= \ln \left\{ \frac{1 - \pi_t}{\pi_t} \right\}.
 \end{aligned} \tag{C.9}$$

Na seção 4.5.1, expressão (4.51), temos:

$$\begin{aligned}
 \frac{\partial M(q)}{\partial q} &= \frac{1}{B(\alpha, \beta)} \frac{\partial}{\partial q} \left[ \frac{\Gamma(\alpha - q)\Gamma(\beta + q)}{\Gamma(\alpha + \beta)} \right] \\
 &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{-\Gamma'(\alpha - q)\Gamma(\beta + q) + \Gamma(\alpha - q)\Gamma'(\beta + q)\}.
 \end{aligned} \tag{C.10}$$

Na seção 4.5.1, expressão (4.52), temos:

$$\begin{aligned}
 \frac{\partial^2 M(q)}{\partial q^2} &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{\Gamma''(\alpha - q)\Gamma(\beta + q) - \Gamma'(\alpha - q)\Gamma'(\beta + q) - \\
 &\quad - \Gamma'(\alpha - q)\Gamma'(\beta + q) + \Gamma(\alpha - q)\Gamma''(\beta + q)\}.
 \end{aligned} \tag{C.11}$$

Na seção 4.5.1, expressão (4.53), temos:

$$\begin{aligned}
 \Psi'(\alpha) &= \frac{\Gamma''(\alpha)\Gamma(\alpha) - \Gamma'(\alpha)\Gamma'(\alpha)}{[\Gamma(\alpha)]^2} \\
 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \left[ \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right]^2 \\
 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - [\Psi(\alpha)]^2.
 \end{aligned} \tag{C.12}$$

Na seção 4.5.1, expressão (4.54), temos:

$$\begin{aligned}
 E(\eta) &= \frac{\partial M(t)}{\partial t} \Big|_{t=0} \\
 &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{\Gamma(\alpha)\Gamma'(\beta) - \Gamma'(\alpha)\Gamma(\beta)\} \\
 &= \frac{\Gamma'(\beta)}{\Gamma(\beta)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \\
 &= \Psi(\beta) - \Psi(\alpha).
 \end{aligned} \tag{C.13}$$

Na seção 4.5.1, expressão (4.55), temos:

$$\begin{aligned}
 E(\eta^2) &= \left. \frac{\partial^2 M(t)}{\partial t^2} \right|_{t=0} \\
 &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \{ \Gamma''(\alpha)\Gamma(\beta) - 2\Gamma'(\alpha)\Gamma'(\beta)\Gamma(\alpha)\Gamma''(\beta) \} \\
 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - 2\Psi(\alpha)\Psi(\beta) + \frac{\Gamma''(\beta)}{\Gamma(\beta)}.
 \end{aligned} \tag{C.14}$$

Na seção 4.5.1, expressão (4.56), temos:

$$\begin{aligned}
 Var(\eta) &= E(\eta^2) - [E(\eta)]^2 \\
 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} + \frac{\Gamma''(\beta)}{\Gamma(\beta)} - 2\Psi(\alpha)\Psi(\beta) - [\Psi(\beta) - \Psi(\alpha)]^2 \\
 &= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \Psi^2(\alpha) + \frac{\Gamma''(\beta)}{\Gamma(\beta)} - \Psi^2(\beta) - \cancel{2\Psi(\alpha)\Psi(\beta)} + \cancel{2\Psi(\alpha)\Psi(\beta)} \\
 &= \Psi'(\alpha) + \Psi'(\beta).
 \end{aligned} \tag{C.15}$$

Na seção 4.5.4, expressão (4.66), temos:

$$\begin{aligned}
 f(\eta_t|X_t) &= \frac{P(X_t|\eta_t)f(\eta_t)}{P(X_t)} \\
 &= \frac{\alpha_t + \beta_t}{\alpha_t^{1-x_t} B(\alpha_t, \beta_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{\beta_t + c_t \alpha_t} \right]^{x_t} \frac{e^{-(\alpha_t+1-x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t+\beta_t+1}} \\
 &= \frac{\alpha_t + \beta_t}{\alpha_t^{1-x_t} \beta_t^{x_t} B(\alpha_t, \beta_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{1 + \frac{c_t \alpha_t}{\beta_t}} \right]^{x_t} \frac{e^{-(\alpha_t+1-x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t+\beta_t+1}} \\
 &= \frac{1}{B(\alpha_t + 1 - x_t, \beta_t + x_t)} \left[ \frac{1 + c_t e^{-\eta_t}}{1 + \frac{c_t \alpha_t}{\beta_t}} \right]^{x_t} \frac{e^{-(\alpha_t+1-x_t)\eta_t}}{[1 + e^{-\eta_t}]^{\alpha_t+\beta_t+1}}.
 \end{aligned} \tag{C.16}$$

Na seção 4.5.5, expressão (4.100), temos:

$$\begin{aligned}
 \frac{p_t - c_t}{1 - c_t} &= \frac{1}{1 + e^{-\eta_t}} \\
 1 + e^{-\eta_t} &= \frac{1 - c_t}{p_t - c_t} \\
 e^{-\eta_t} &= \frac{1 - c_t - p_t + c_t}{p_t - c_t} = \frac{1 - p_t}{p_t - c_t} \\
 -\eta_t &= \ln \frac{1 - p_t}{p_t - c_t} \\
 \eta_t &= \ln \frac{p_t - c_t}{1 - p_t}.
 \end{aligned} \tag{C.17}$$



# Simulações e Resultados

## D Cenário 5a: $\theta$ fixo, $\lambda = 1$ , $\tau = 0.5$

Considerando uma pequena alteração na variabilidade dos estimadores  $\hat{\theta}$  e  $\tilde{\theta}$ , provocada pela constante  $\tau = 0.5$ , observa-se que os valores de  $\alpha_t^*$  e  $\beta_t^*$ , observados nas Tabelas (A.2) e (A.1) são bem mais estáveis que aqueles apresentados observados nas Tabelas (5.9) e (5.10).

**Tabela A.1:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	1.9390	1.9900	1.1430	1.9900	1.9900	1.9900
-2.75	1.9390	1.9900	1.1470	1.9900	1.9900	1.9900
-2.50	1.9270	1.9900	1.2050	1.9900	1.9770	1.9900
-2.25	1.9410	1.9900	1.1720	1.9900	1.9900	1.9900
-2.00	1.8540	1.9900	0.8202	1.9900	1.9894	1.9900
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	1.2207	1.1074	0.8287	1.9644	0.9935	1.4274
2.25	1.0634	0.9905	0.8206	1.8371	0.9900	1.0231
2.50	1.0989	0.9934	0.7774	1.8618	0.9900	1.1305
2.75	1.0616	0.9909	0.8023	1.8381	0.9900	1.0231
3.00	1.0064	0.9900	0.8136	1.5437	0.9885	0.9900

**Tabela A.2:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	0.9955	0.9900	0.9900	1.0542	0.9900	0.9905
-2.75	0.9903	0.9900	0.9900	0.9933	0.9900	0.9900
-2.50	1.0017	0.9900	0.7628	1.5437	0.9880	0.9904
-2.25	0.9774	0.9900	0.8474	0.9900	0.9888	0.9900
-2.00	1.0631	0.9900	0.8799	1.6915	0.9900	0.9939
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	1.7788	1.9012	0.9803	1.9900	1.6442	1.9886
2.25	1.8747	1.9897	0.9352	1.9900	1.9185	1.9900
2.50	1.8548	1.9886	0.9796	1.9900	1.7205	1.9900
2.75	1.8739	1.9895	0.9216	1.9900	1.9174	1.9900
3.00	1.9129	1.9900	0.9792	1.9900	1.9729	1.9900

**Tabela A.3:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador	EAP	MAP	EAP ( $\lambda = 1, \tau = 1$ )	MAP ( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )	1.133575	1.438154	0.608637	0.956001
PS_EAP_AVG ( $\tilde{\theta}$ )   PS_MAP_AVG ( $\tilde{\theta}$ )	0.750067	1.304804	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )	0.789450	1.318761	0.321442	0.387185

## E Cenário 5b: $\theta$ aleatório, $\lambda = 1$ , $\tau = 0.5$

**Tabela A.4:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	1.8510	1.9900	1.1510	1.9900	1.9020	1.9900
-1.0908568	1.8800	1.9900	1.0440	1.9900	1.8960	1.9900
-0.9892490	1.7784	1.9020	0.7225	1.9900	1.6442	1.9886
-0.8523327	1.8240	1.9020	1.0940	1.9900	1.6510	1.9900
-0.7795234	1.8171	1.9886	0.7661	1.9900	1.8278	1.9900
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	1.3832	1.3044	0.7159	1.9839	1.1294	1.6084
1.5626897	1.1550	1.0181	0.8542	1.8901	0.9903	1.2941
1.5748699	1.2429	1.1086	0.7511	1.9523	0.9941	1.4480
1.6036061	1.1111	0.9921	0.7810	1.8876	0.9900	1.1294
1.8711056	1.1446	1.0174	0.7865	1.8618	0.9904	1.2909



**Tabela A.5:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	1.1096	0.9920	0.8102	1.8218	0.9900	1.1063
-1.0908568	1.0808	0.9909	0.9360	1.5437	0.9900	1.0465
-0.9892490	1.1741	1.0750	0.3765	1.8618	0.9969	1.4073
-0.8523327	1.1736	1.0488	0.7647	1.6014	0.9908	1.4072
-0.7795234	1.1297	1.0039	0.9186	1.7028	0.9905	1.2362
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	1.6540	1.7120	1.0520	1.9900	1.4290	1.9540
1.5626897	1.7887	1.9606	0.3596	1.9900	1.6948	1.9899
1.5748699	1.7106	1.8748	0.3868	1.9900	1.6172	1.9821
1.6036061	1.8387	1.9880	0.9406	1.9900	1.7205	1.9900
1.8711056	1.8292	1.9823	0.9635	1.9900	1.7158	1.9900

**Tabela A.6:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	1.7227	1.8678	0.5631	1.9899	1.6130	1.9824
-1.0908568	1.6318	1.7139	0.7462	1.9900	1.3840	1.9673
-0.9892490	1.6658	1.6959	0.5774	1.9900	1.5137	1.9581
-0.8523327	1.6623	1.7183	0.7861	1.9895	1.4555	1.9593
-0.7795234	1.6390	1.6810	1.1350	1.9900	1.4570	1.9560
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	1.2565	1.2363	0.3146	1.9819	1.0247	1.4913
1.5626897	1.2263	1.1064	0.6519	1.9650	0.9982	1.4096
1.5748699	1.1412	1.0244	0.3614	1.9549	0.9905	1.3013
1.6036061	1.1365	1.0067	0.6336	1.8879	0.9903	1.2883
1.8711056	1.2376	1.1295	0.6590	1.9554	0.9942	1.4170

**Tabela A.7:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$ , com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método EAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	1.2698	1.1602	0.7518	1.9523	1.0300	1.4475
-1.0908568	1.3556	1.3049	0.3974	1.9880	1.0608	1.6168
-0.9892490	1.3543	1.4041	0.9143	1.9000	1.0238	1.6329
-0.8523327	1.3304	1.2898	0.5251	1.9813	1.0573	1.6784
-0.7795234	1.4089	1.4101	0.5515	1.9556	1.1308	1.6916
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	1.6803	1.8566	0.1606	1.9900	1.5139	1.9844
1.5626897	1.7680	1.9010	1.0190	1.9900	1.6440	1.9880
1.5748699	1.7582	1.9594	0.1609	1.9900	1.7123	1.9895
1.6036061	1.7757	1.9605	0.1273	1.9900	1.7200	1.9895
1.8711056	1.7510	1.8680	0.8740	1.9900	1.6370	1.9890

**Tabela A.8:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 1$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador		EAP	MAP	EAP	MAP
				( $\lambda = 1, \tau = 1$ )	( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )	PS_MAP ( $\hat{\theta}$ )	0.806453	0.834599	0.608637	0.956001
PS_EAP_AVG ( $\tilde{\theta}$ )	PS_MAP_AVG ( $\tilde{\theta}$ )	0.372060	0.678351	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )	PS_MAP_MEDIAN ( $\hat{\theta}$ )	0.370497	0.657118	0.321442	0.387185

## F Cenário 5c: $\theta$ fixo, $\lambda = 2$ , $\tau = 0.5$

**Tabela A.9:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	3.1120	3.2170	1.6160	3.2170	3.2170	3.2170
-2.75	3.1120	3.2170	1.6220	3.2170	3.2170	3.2170
-2.50	3.0960	3.2170	1.7040	3.2170	3.1740	3.2170
-2.25	3.1140	3.2170	1.6570	3.2170	3.2170	3.2170
-2.00	2.9890	3.2170	1.1600	3.2170	3.2140	3.2170
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.1140	1.9860	1.1720	3.2100	1.9800	2.2550
2.25	2.0710	1.9810	1.1600	2.9850	1.9720	2.2720
2.50	1.9830	1.9810	1.0990	2.6070	1.9800	2.0140
2.75	2.0040	1.9800	1.1350	2.9660	1.9710	2.0230
3.00	1.9390	1.9800	1.1510	2.6070	1.9720	1.9810

**Tabela A.10:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	1.9360	1.9800	1.4910	1.9800	1.976	1.9800
-2.75	1.9280	1.9800	1.4050	1.9800	1.975	1.9800
-2.50	1.9310	1.9800	1.0790	2.6070	1.971	1.9810
-2.25	1.9070	1.9800	1.1980	1.9800	1.972	1.9800
-2.00	2.0220	1.9800	1.2440	2.6470	1.980	1.9840
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.9340	3.2100	1.3860	3.2170	2.821	3.2170
2.25	2.9800	3.1740	1.3230	3.2170	2.890	3.2170
2.50	3.0450	3.2170	1.3850	3.2170	3.071	3.2170
2.75	3.0230	3.2130	1.3030	3.2170	3.020	3.2170
3.00	3.0740	3.2170	1.3850	3.2170	3.174	3.2170

**Tabela A.11:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador		EAP	MAP	EAP	MAP
				( $\lambda = 1, \tau = 1$ )	( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )	PS_MAP ( $\hat{\theta}$ )	1.729390	1.621681	0.608637	0.956001
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )	PS_MAP_AVG ( $\bar{\hat{\theta}}$ )	1.204145	1.354660	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}^\dagger$ )	PS_MAP_MEDIAN ( $\hat{\theta}^\dagger$ )	1.218218	1.382375	0.321442	0.387185

## G Cenário 5d: $\theta$ fixo, $\lambda = 2$ , $\tau = 1$

**Tabela A.12:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 1$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	2.062e+11	1366000	3	4.592e+12	744	2.504e+09
-2.75	2.147e+11	1422000	3	4.781e+12	775	2.607e+09
-2.50	2.251e+11	1491000	3	5.013e+12	813	2.734e+09
-2.25	2.168e+11	1436000	3	4.829e+12	783	2.633e+09
-2.00	9.549e+10	632600	1	2.127e+12	341	1.160e+09
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	5.531e+10	366400	1	1.232e+12	199	6.718e+08
2.25	5.093e+10	337400	1	1.134e+12	184	6.187e+08
2.50	5.615e+10	371900	1	1.251e+12	202	6.820e+08
2.75	6.017e+10	398600	1	1.340e+12	217	7.309e+08
3.00	5.649e+10	374200	1	1.258e+12	204	6.862e+08

**Tabela A.13:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 1$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	8.762e+10	580400	2	1.952e+12	313	1.064e+09
-2.75	7.156e+10	474000	2	1.594e+12	255	8.692e+08
-2.50	4.333e+10	287000	1	9.650e+11	153	5.263e+08
-2.25	5.547e+10	367400	1	1.235e+12	197	6.738e+08
-2.00	1.342e+11	889000	2	2.989e+12	485	1.630e+09
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	1.876e+11	1243000	2	4.179e+12	675	2.279e+09
2.25	1.854e+11	1228000	2	4.130e+12	666	2.252e+09
2.50	1.813e+11	1201000	2	4.037e+12	652	2.202e+09
2.75	1.726e+11	1143000	2	3.843e+12	619	2.096e+09
3.00	1.847e+11	1223000	2	4.113e+12	663	2.243e+09

**Tabela A.14:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 1$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador	EAP	MAP	EAP ( $\lambda = 1, \tau = 1$ )	MAP ( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )	2.038332	1.102676	0.608637	0.956001
PS_EAP_AVG ( $\tilde{\theta}$ )   PS_MAP_AVG ( $\tilde{\theta}$ )	2.000198	0.980812	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )	2.043400	0.978668	0.321442	0.387185

## H Cenário 5e: $\theta$ aleatório, $\lambda = 2$ , $\tau = 0.5$

**Tabela A.15:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	2.9950	3.2170	1.6280	3.2170	3.0790	3.2170
-1.0908568	3.0330	3.2170	1.4770	3.2170	3.0550	3.2170
-0.9892490	2.9140	3.0920	1.0220	3.2170	2.8200	3.2160
-0.8523327	2.9820	3.1730	1.5470	3.2170	2.7790	3.2170
-0.7795234	2.8760	3.1740	1.0830	3.2170	2.8420	3.2170
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	2.4060	2.4620	1.0120	3.1420	2.1260	2.7140
1.5626897	2.2600	2.1710	1.2080	3.1670	1.9850	2.6070
1.5748699	2.1580	2.0580	1.0620	3.1930	1.9820	2.4330
1.6036061	2.1220	1.9820	1.1050	3.2010	1.9710	2.3900
1.8711056	2.1270	2.0410	1.1120	3.0050	1.9810	2.3610

**Tabela A.16:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	2.0720	1.9830	1.1460	2.8410	1.9800	2.0660
-1.0908568	2.0320	1.9810	1.3240	2.6070	1.9800	2.0490
-0.9892490	2.0909	1.9886	0.5324	3.0046	1.9760	2.3612
-0.8523327	2.1010	2.0410	1.0810	2.6520	1.9810	2.2770
-0.7795234	2.1870	2.0540	1.2990	2.9660	1.9820	2.4370
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	2.7480	2.7730	1.4890	3.2160	2.6990	3.0590
1.5626897	2.7776	2.8985	0.5085	3.2171	2.5876	3.1980
1.5748699	2.8510	3.0790	0.5470	3.2170	2.7760	3.2130
1.6036061	2.9270	3.1710	1.3300	3.2170	2.7790	3.2170
1.8711056	2.9470	3.1660	1.3630	3.2170	2.7790	3.2170

**Tabela A.17:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 2$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador		EAP	MAP	EAP ( $\lambda = 1, \tau = 1$ )	MAP ( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )	PS_MAP ( $\hat{\theta}$ )	1.053790	0.790084	0.608637	0.956001
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )	PS_MAP_AVG ( $\bar{\hat{\theta}}$ )	0.645747	0.623794	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )	PS_MAP_MEDIAN ( $\hat{\theta}$ )	0.654511	0.620763	0.321442	0.387185

## I Cenário 5f: $\theta$ fixo, $\lambda = 3$ , $\tau = 0.5$

**Tabela A.18:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	4.179	4.344	1.979	4.344	4.343	4.344
-2.75	4.179	4.344	1.987	4.344	4.343	4.344
-2.50	4.162	4.344	2.087	4.344	4.285	4.344
-2.25	4.182	4.344	2.030	4.344	4.343	4.344
-2.00	4.026	4.344	1.421	4.344	4.337	4.344
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	2.914	2.971	1.435	3.631	2.970	3.008
2.25	2.821	2.970	1.421	2.970	2.953	2.970
2.50	2.865	2.970	1.346	3.532	2.970	2.971
2.75	2.856	2.970	1.390	3.631	2.952	2.971
3.00	2.819	2.970	1.409	2.970	2.953	2.970

**Tabela A.19:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	2.8640	2.9700	1.8260	2.9700	2.9590	2.9700
-2.75	2.8530	2.9700	1.7200	2.9700	2.9570	2.9700
-2.50	2.8480	2.9700	1.3210	3.6310	2.9510	2.9710
-2.25	2.8260	2.9700	1.4680	2.9700	2.9540	2.9700
-2.00	2.9620	2.9700	1.5240	3.5490	2.9700	2.9740
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	4.0990	4.3430	1.6980	4.3440	4.1530	4.3440
2.25	4.1520	4.3440	1.6200	4.3440	4.3420	4.3440
2.50	4.1240	4.3440	1.6970	4.3440	4.3220	4.3440
2.75	4.1250	4.3440	1.5960	4.3440	4.2850	4.3440
3.00	4.1580	4.3440	1.6960	4.3440	4.3420	4.3440

**Tabela A.20:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador	EAP	MAP	EAP ( $\lambda = 1, \tau = 1$ )	MAP ( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )   PS_MAP ( $\hat{\theta}$ )	1.855128	2.026129	0.608637	0.956001
PS_EAP_AVG ( $\tilde{\theta}$ )   PS_MAP_AVG ( $\tilde{\theta}$ )	1.312158	1.580721	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )   PS_MAP_MEDIAN ( $\hat{\theta}$ )	1.330259	1.600668	0.321442	0.387185

## J Cenário 5g: $\theta$ fixo, $\lambda = 3$ , $\tau = 1$

**Tabela A.21:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	1.850e+18	2.195e+09	4	5.522e+19	13840	3.482e+14
-2.75	1.895e+18	2.249e+09	4	5.657e+19	14180	3.567e+14
-2.50	2.007e+18	2.381e+09	4	5.989e+19	15010	3.776e+14
-2.25	1.927e+18	2.287e+09	4	5.753e+19	14420	3.627e+14
-2.00	7.362e+17	8.735e+08	2	2.197e+19	5505	1.385e+14
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	6.233e+17	7.396e+08	2	1.860e+19	4662	1.173e+14
2.25	5.692e+17	6.753e+08	2	1.699e+19	4258	1.071e+14
2.50	5.357e+17	6.356e+08	2	1.599e+19	4007	1.008e+14
2.75	5.324e+17	6.318e+08	2	1.589e+19	3983	1.002e+14
3.00	5.412e+17	6.421e+08	2	1.615e+19	4048	1.018e+14

**Tabela A.22:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 1$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-3.00	9.516e+17	1.129e+09	3	2.840e+19	7119	1.791e+14
-2.75	8.128e+17	9.644e+08	3	2.426e+19	6080	1.530e+14
-2.50	4.811e+17	5.709e+08	2	1.436e+19	3596	9.055e+13
-2.25	6.081e+17	7.215e+08	2	1.815e+19	4546	1.144e+14
-2.00	1.272e+18	1.510e+09	2	3.798e+19	9518	2.394e+14
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2.00	1.549e+18	1.838e+09	3	4.624e+19	11590	2.916e+14
2.25	1.522e+18	1.806e+09	3	4.543e+19	11380	2.864e+14
2.50	1.569e+18	1.861e+09	3	4.682e+19	11730	2.952e+14
2.75	1.514e+18	1.796e+09	3	4.518e+19	11320	2.849e+14
3.00	1.609e+18	1.910e+09	3	4.804e+19	12040	3.029e+14

**Tabela A.23:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 1$  com 3 estimativas provisórias de  $\theta_{jt}$ 

Estimador		EAP	MAP	EAP	MAP
				( $\lambda = 1, \tau = 1$ )	( $\lambda = 1, \tau = 1$ )
PS_EAP ( $\hat{\theta}$ )	PS_MAP ( $\hat{\theta}$ )	1.188951	1.168527	0.608637	0.956001
PS_EAP_AVG ( $\tilde{\theta}$ )	PS_MAP_AVG ( $\tilde{\theta}$ )	1.052088	1.017658	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )	PS_MAP_MEDIAN ( $\hat{\theta}$ )	1.093236	1.018696	0.321442	0.387185

## K Cenário 5h: $\theta$ aleatório, $\lambda = 3$ , $\tau = 0.5$

**Tabela A.24:** Estatísticas de caracterização para  $\alpha_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0,1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	4.0430	4.3440	1.9940	4.3440	4.1720	4.3440
-1.0908568	4.0860	4.3430	1.8090	4.3440	4.1180	4.3440
-0.9892490	3.7710	3.9840	1.2510	4.3440	3.4860	4.2850
-0.8523327	3.9400	4.1710	1.8940	4.3440	3.7070	4.3410
-0.7795234	3.8580	4.2800	1.3270	4.3440	3.8090	4.3440
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	3.1800	3.1290	1.2400	4.3060	2.9750	3.4550
1.5626897	2.9430	2.9720	1.4790	3.6810	2.9700	3.0510
1.5748699	2.8830	2.9700	1.3010	3.7330	2.9700	2.9730
1.6036061	2.8130	2.9700	1.3530	2.9700	2.9520	2.9700
1.8711056	2.9270	2.9710	1.3620	3.6310	2.9700	2.9980

**Tabela A.25:** Estatísticas de caracterização para  $\beta_t^*$ , quando  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$  (Power Steady - Método MAP)

$\theta \sim N(0, 1)$	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
-1.3608653	3.0110	2.9710	1.4030	3.7490	2.9700	3.0460
-1.0908568	2.9630	2.9710	1.6210	3.6310	2.9700	3.0460
-0.9892490	3.1994	3.1939	0.6521	4.3119	2.9564	3.6446
-0.8523327	3.1300	3.0800	1.3240	4.3120	2.9710	3.4390
-0.7795234	3.1770	3.0530	1.5910	4.2330	2.9700	3.4220
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9353406	3.8960	4.1610	1.8230	4.3440	3.8190	4.3080
1.5626897	3.9988	4.3373	0.6228	4.3442	3.9840	4.3442
1.5748699	4.0268	4.3442	0.6699	4.3442	4.1498	4.3442
1.6036061	4.1520	4.3440	1.6290	4.3440	4.3420	4.3440
1.8711056	4.0830	4.3440	1.6690	4.3440	4.1720	4.3440

**Tabela A.26:** Power Steady - RMSE para  $k = 0.99$ ,  $\lambda = 3$ ,  $\tau = 0.5$  com 3 estimativas provisórias de  $\theta_{jt}$

Estimador		EAP	MAP	EAP	MAP
				$(\lambda = 1, \tau = 1)$	$(\lambda = 1, \tau = 1)$
PS_EAP ( $\hat{\theta}$ )	PS_MAP ( $\hat{\theta}$ )	0.990353	1.157380	0.608637	0.956001
PS_EAP_AVG ( $\bar{\hat{\theta}}$ )	PS_MAP_AVG ( $\bar{\hat{\theta}}$ )	0.638180	0.769242	0.275657	0.312662
PS_EAP_MEDIAN ( $\hat{\theta}$ )	PS_MAP_MEDIAN ( $\hat{\theta}$ )	0.652242	0.781539	0.321442	0.387185