

BOLETIM



Nº. 02, Volume 1

Setembro 2004

BOLETIM *ISBRA*

Volume 1 Número 2

Setembro 2004

Boletim Oficial do Capítulo Brasileiro da *International Society of Bayesian Analysis*

ISBRA

Presidente:	Josemar Rodrigues
Secretário:	Luís A. Milan
Tesoureiro:	José Galvão Leite

CARTA DO PRESIDENTE

Josemar Rodrigues

vjosemar@power.ufscar.br

Finalmente a ISBRA existe oficialmente. Após varias idas e vindas e muito paciência conseguimos registrar a ISBRA no Registro Civil das Pessoas Jurídicas de São Carlos, sob o número 007251. O próximo passo será conseguir o CGC para abrimos uma conta exclusiva para a ISBRA. Em fevereiro deste ano realizamos o 7EBEB em São Carlos, cujo relatório foi divulgado no boletim da ABE. Para os interessados neste evento temos a disposição o CD das atividades programada durante o mesmo. Estivemos participando do ISBA2004, realizado em Vina Del Mar – Chile, onde fizemos um relatório a ISBA sobre as atividades da ISBRA e enviamos uma carta ao prof. Berger, presidente da ISBA, oficializando a ISBRA como capítulo brasileiro da ISBA. Nesta segunda edição do boletim contamos com quatro artigos selecionados durante 7EBEB, depoimentos de vários pesquisadores sobre o congresso da ISBA no Chile, um documento sobre a criação da ISBRA e uma divulgação do COBALII a ser realizado no período de 6-10/02/2005 no México. Gostaria de agradecer aos colegas que colaboraram com este boletim e em especial ao aluno Milton Y. Saito.

Cordiais saudações,

Josemar Rodrigues ■

CONTEÚDO

1 – ISBRA - Sua História Sobre Nosso Ponto de Vista ...	1
2 – ISBA 2004 World Meeting: Um Encontro Científico Para Se Ter Na Memória	2
3 – Artigos.....	6
Seleção de Modelos de Fronteira de Produção Estocástica: Abordagem Bayesiana.....	6
Inferência Bayesiana Em Modelos De Volatilidade Estocástica Usando Processos Autoregressivos Com Transição Suave	14
Funções de Transferência em Modelos Dinâmicos Lineares Generalizados Bayesianos	20
Bayesian Estimation Via Mcmc For Probit-Normal Model In Item Response Theory	33
4 – Convites.....	43
Carta-convite enviada a Lista da ABE.....	43
II COBAL – Congresso Bayesiano da América Latina	43

1 – ISBRA - Sua História Sobre Nosso Ponto de Vista

Rosangela H. Loschi
(EST–UFMG)

Márcia D’Elia Branco
(IME –USP)

Josemar Rodrigues
(UFSCAR)

La Pégola da Plaza, Plaza Mulato Gil, Santiago, Chile. O ano era 2000 e o mês, Janeiro. Estas são as coordenadas do local e época em que a primeira semente para a criação do Capítulo Brasileiro da ISBA (International Society for Bayesian Analysis) foi lançada.

Estávamos reunidos Pilar Iglesias, Heleno Bolfarine e nós, em torno de uma mesa de bar, comentando o sucesso do IV Seminário Chileno de Estadística Bayesiana que havia recém terminado. Pilar dizia, animadamente, que no primeiro encontro Chileno o número de participantes era apenas quatro ou cinco e que estava feliz por ter tido uma participação de cerca de 30 pessoas nesta quinta edição do seminário. Dizia Pilar que a criação do Capítulo Chileno da ISBA tinha contribuído significativamente para este crescimento. Depois deste depoimento entusiasmado, nos perguntamos: “por que não criamos um Capítulo Brasileiro? Somos um grupo maior e desenvolvemos pesquisa em Estatística Bayesiana há muito mais tempo!”

Voltamos ao Brasil animados com esta idéia e decidimos escrever uma carta a toda comunidade estatística brasileira propondo a criação do capítulo. Queríamos, antes de mais nada, saber se nossa comunidade achava relevante a criação do Capítulo e se estavam interessados em participar. Esta carta esta anexada abaixo do texto e ainda nos lembramos vivamente da emoção de escrevê-la.

Poucos se manifestaram diante da nossa proposta, o que nos levou a propor uma reunião com o grupo de interessados durante o 14º SINAPE, realizado em julho de 2000. Queríamos deixar claro, nesta reunião, que não pretendíamos criar uma associação para fazer frente a ABE ou a qualquer outra associação de Estatística. A meta do capítulo sempre foi aglutinar pessoas e não separá-las. Pretendíamos agrupar pessoas que tinham como propósito ajudar a divulgar ainda mais a Estatística Bayesiana no Brasil. Queríamos com a criação do capítulo criar uma ponte entre nós e a ISBA, pois acreditávamos que este seria um passo importante para nós.

Participaram dessa reunião 13 pessoas: José Galvão Leite, Emanuel Barbosa, Luis Milan, Carlos Pereira, Carlos Diniz, Hélio Migon, Hedibert Lopes, Marlos Viana, Pilar Iglesias, Sergio Wechsler e nós. Nesta reunião decidiu-se pela criação do capítulo e decidiu-se também que a

primeira diretoria do capítulo deveria ser composta por pessoas de diversos departamentos pois assim estaríamos maximizando a nossa chance de cumprir uma de nossas mais preciosas metas – divulgar a Estatística Bayesiana. Ficou definido, com a criação do capítulo, que a organização de todos os Encontros Brasileiros de Estatística Bayesiana (EBEB) estariam sob a responsabilidade da sua diretoria a partir daquela data.

Para a primeira diretoria do capítulo foram eleitos: Sergio Wechsler (Presidente – USP); Hedibert Lopes (Vice-Presidente - UFRJ); Josemar Rodrigues (Tesoureiro – UFSCAR); Márcia Branco (USP) e Rosangela H. Loschi (UFMG) como secretárias.

O ISBRA (nome posteriormente dado ao Capítulo Brasileiro pelo Luis Gustavo Esteves) era ainda um recém-nascido quando assumiu a organização do I congresso Bayesiano da América Latina (I COBAL). O COBAL nasceu da necessidade que tínhamos que conhecer o que estava sendo desenvolvido em Estatística Bayesiana em toda a América Latina e de integrar os grupos de pesquisa dos países latino americanos. O I COBAL, realizado em Ubatuba de 03 a 07 de fevereiro de 2002, incorporou a 6ª edição do Encontro Brasileiro de Estatística Bayesiana (EBEB) e, como fruto da interação entre a ISBRA, através do Sergio, e a então Presidente da ISBA, Alicia Carriquiry, o I COBAL foi considerado como um Encontro Regional da ISBA. Este apoio da ISBA foi um reconhecimento do importante do trabalho que nós, latino americanos, estávamos fazendo e um importante apoio ao nosso capítulo. Cabe ressaltar a vital contribuição do Nelson Tanaka na organização do I COBAL. Embora Nelson não fizesse parte da Diretoria do ISBRA e, portanto, não teria a responsabilidade de ajudar na organização, seu trabalho e contribuição foram imprescindíveis para o sucesso do encontro.

No 15º SINAPE, realizado em julho de 2002, foi eleita a nova diretoria do ISBRA. As metas principais da nova diretoria era organizar o 7º EBEB e encarregar-se da oficialização junto a ISBA da existência do ISBRA e também de elaborar de seu estatuto e fazer seu registro como uma associação científica. Da nova diretoria, faziam parte Josemar Rodrigues (presidente), José Galvão Leite (Tesoureiro) e Luis Milan (Secretário). O dinheiro que foi passado ao ISBRA pela ISBA na época do I COBAL foi importante, posteriormente, para o início da organização do 7º EBEB que foi realizado de 8 a 11 de fevereiro de 2004 em São Carlos-SP.

Durante o 7º EBEB, o ISBRA teve seu estatuto aprovado e decidiu-se pela reeleição da diretoria. Em 6 de Maio de 2004, o ISBRA foi oficialmente registrado como associação científica no cartório de Registro Civil das Pessoas Jurídicas da cidade de São Carlos-SP, sob o número 007251. O último passo para a real criação do ISBRA era enviar uma carta à ISBA oficializando a sua existência. Este passo foi, finalmente, dado em 25 de Julho de 2004. Depois de 4 anos e, curiosamente, depois

de tantas realizações importantes, já existimos. O Capítulo Brasileiro da ISBA, carinhosamente apelidado

de ISBRA, tem hoje 55 filiados. Portanto, agora só nos cabe crescer e frutificar. ■

2 – ISBA 2004 World Meeting: Um Encontro Científico Para Se Ter Na Memória

Rosângela H. Loschi
(EST–UFMG)

Alexandra M. Schmidt
(IM – UFRJ)

Depois de, por algumas vezes, ocupar espaço neste Boletim divulgando o **ISBA 2004 World Meeting** é com enorme satisfação que agora fazemos um pequeno relato de seu sucesso e da importante participação brasileira no encontro.

Depois de seis edições, o 7º Encontro Mundial da Sociedade Internacional de Análise Bayesiana (ISBA) foi, finalmente, sediado por um país Sul Americano. O ISBA 2004 World Meeting ocorreu de 23 a 27 de maio, em Vinã Del Mar (Chile) e foi organizado pela ISBA e pelas universidades Católica de Chile, de Talca, de Valparaíso e las Américas. O encontro teve um nível científico altíssimo e contou com a participação de cerca de 300 pessoas de vários países do mundo das quais cerca de 100 eram estudantes. Da programação científica do ISBA 2004 constaram conferências, tutoriais, debates, sessões técnicas, sessões pôsteres e sessões de estudantes.

O ISBA 2004 World Meeting teve uma organização primorosa e impecável e veio para a América do Sul graças à - o que para muitos foi um ato insano - coragem de uma grande amiga, Pilar Iglesias, que se candidatou a organizar um encontro de tamanho porte bem aqui em solos sul-americanos. A programação científica do ISBA 2004 ficou sob a responsabilidade do Fabrizio Ruggeri que, graças a sua já notória eficiência, nos forneceu uma programação bem diversificada e com qualidade científica de altíssimo nível. O sucesso do ISBA 2004 deveu-se também à colaboração e participação ativa de várias outras pessoas: Ed George (Conference Chairman), Alicia Carriquiry (Finance Committee Chair), Peter Müller (the ISBA treasurer) e Carlos Heriquez (Chair of the Social Events Committee). O comitê local, presidido pela Pilar, contou também com a participação de alguns outros amigos queridos: Fernando Quintana, Glória Icaza, Manuel Galea, Anita Sepúlveda, a Rosita e muitos outros. Além de ter tido um apoio entusiasmado e eficiente de muitos (mas muitos mesmo) estudantes de algumas universidades chilenas.

O ISBA 2004 foi marcado por uma diversidade de “idiomas”: do inglês amplamente falado ao redor do mundo ao pouco falado italiano (mas não em congressos Bayesianos, não é mesmo Fabrizio?); das distribuições de referência ao fundamentalismo de Finetti, passando

pelas últimas novidades na área computacional; das apresentações de grandes pesquisadores na área à de alunos iniciando a carreira acadêmica. Acima de tudo, a partir da vasta gama de assuntos cobertos pelo programa científico do encontro, ficou clara a grande capacidade do paradigma de Bayes em fornecer soluções coerentes para problemas altamente complexos.

A sétima edição do ISBA World Meeting foi aberta com a conferência de um Brasileiro, Dani Gamerman. E contou ainda com a participação significativa de pesquisadores de quase todas as instituições Brasileiras em que há pesquisa em Estatística Bayesiana e, também, com a participação expressiva de vários pesquisadores Latino Americanos. O que significa isto? A nosso ver, significa que a semente plantada há alguns anos atrás pelo Carlinhos, Dani, Hélio e Pericchi (“*unos tipos que llegaron hablado un lenguaje raro*”, como disse o Pericchi na conferência que ministrou durante o I COBAL) brotou, cresceu e deu frutos. Em grande parte, é devido ao pioneirismo e dedicação destes meninos que desenvolvemos, hoje, na América Latina e, especialmente no Brasil, pesquisa em Estatística Bayesiana de excelente qualidade e estamos deixando nossa marca no mundo. Graças a eles, hoje somos muitos.

A qualidade da pesquisa que estamos desenvolvendo no Brasil também é atestada pelo sucesso da Esther Salazar e do Juan Vivar, ambos alunos do Programa de Pós-Graduação em Estatística do IM-UFRJ, que tiveram seus trabalhos de mestrado selecionados para apresentação oral no ISBA 2004. Esther foi orientada pelo Hedibert Lopes e, Juan, pelo Marco Ferreira. Grande feito! Ainda mais se lembrarmos que concorreram com trabalhos de doutorado desenvolvidos em todo o mundo. Atualmente, Esther e Juan seguem cursando o Doutorado no IM-UFRJ.

É tempo de fazer um agradecimento público a todos que estiveram envolvidos na organização do encontro; a todos que embarcaram neste sonho de plantar ainda mais fundo em solos Latino Americanos, essas idéias interessantes baseadas no Paradigma de Bayes. “*Deu orgulho de ser Latino Americano*”, palavras de Alicia que fazemos nossas. É tempo também de dar parabéns aos Brasileiros que lá estiveram, de uma forma ou de outra, representando bem o nosso país.

Infelizmente, também aproveitamos para ressaltar que o ISBA 2004 foi marcado pela falta de apoio das nossas agências financiadoras para estudantes e jovens pesquisadores que participaram do encontro apresentando trabalhos. Em quase sua totalidade, estudantes e jovens pesquisadores brasileiros utilizaram seus próprios recursos para participarem do encontro (coisa não percebida entre os participantes de outros países). É como se tivesse caído no esquecimento de nossos governantes

que a Universidade Brasileira, responsável pelo desenvolvimento de grande parte da pesquisa do país, precisa renovar-se para que não retrocedamos. É como se tivesse caído no esquecimento que a exposição a esta diversidade de idéias é que alimenta a criatividade, vital para o desenvolvimento e a prosperidade. É cuidando dos jovens que garantimos um futuro melhor (frase feita, que cabe bem aqui). O pouco apoio recebido por alguns de nós foi dado pelo IMS (The Institute of Mathematical Statistics). Este apoio cobriu apenas parte do gasto com a nossa inscrição mas nos fez pensar que, pelo menos, em algum canto do mundo, tem alguém acreditando que é importante apoiar a quem está iniciando.

Lamentamos que mais brasileiros não tenham podido participar. Lamentamos, sinceramente, não ter tido recursos para levar para o ISBA 2004 todos os estudantes com quem trabalhamos, para que participassem deste fórum maravilhoso de discussões, em que a diversidade de idéias era a máxima constante e que nos fez voltar ainda mais motivadas para realizarmos nosso trabalho por aqui. Esperamos também que o país acorde e perceba a importância de continuar investindo na pesquisa feita pelas universidades públicas. Perceba que, apesar do bom desempenho de nós brasileiros, que pôde mais uma vez ser testemunhado no ISBA 2004, é preciso continuar investindo no envio de jovens Mestres para grandes centros do exterior, para que façam seus doutorados e vivam a experiência de desenvolver ciência da mais alta qualidade, convivendo com renomados pesquisadores internacionais para que, finalmente, voltem ao Brasil e tragam esta vivência para nossa sociedade. Acreditamos que este intercâmbio de idéias e experiências contribui muito para a renovação e o crescimento que desejamos para o nosso país. Temos sempre em mente que vivemos num mundo globalizado, com fronteiras muito flexíveis, onde a ciência vem se desenvolvendo muito rapidamente, com várias pessoas trabalhando nos mais variados assuntos e que precisamos sempre estar atentos para não ficarmos à margem desse desenvolvimento.

Para corroborar o que dissemos acima sobre o ISBA 2004, anexamos o testemunho de alguns Brasileiros que lá estiveram e também o depoimento de alguns membros da Comissão Organizadora do evento. Iniciamos apresentando a opinião dos participantes Brasileiros.

Dani Gamerman (DME-UFRJ) Foi com muita satisfação que pudemos constatar que, dentre todas as instituições participantes (algumas de renome mundial) que tiveram trabalhos convidados ou aceitos para apresentação oral, nenhuma teve mais representantes que a UFRJ. Nossa participação foi bem variada envolvendo pesquisadores seniores, pesquisadores mais jovens e alunos de pós-graduação. Esse dado confirma, a nível internacional, o que já havia sido observado no Encontro Brasileiro de Estatística Bayesiana realizado em fevereiro em São Carlos, onde a UFRJ teve os maiores números de professores conferencistas e de alunos apresentando pôsteres. Esses dados estão

relacionados com a recente abertura do nosso Doutorado e mostram um sinal muito positivo. Outras instituições brasileiras também tiveram participação com destaque no evento. Isso mostra que, apesar do cenário adverso que enfrentamos nas universidades brasileiras com dificuldades de toda ordem, o esforço que tem sido feito pela comunidade científica de Estatística da UFRJ e do Brasil começa a mostrar resultados. A Estatística Brasileira está de parabéns.

Hedibert F. Lopes (GSB-University of Chicago) O Encontro Mundial de 2004 da Sociedade Internacional para a Análise Bayesiana, ou simplesmente *ISBA2004*, realizado em maio último em Viña del Mar no Chile se desenvolveu numa atmosfera comparável aos maiores encontros Bayesianos e mostra a maturidade dos grupos de pesquisa Bayesianos na América Latina, com destaques para México, Chile e, principalmente, Brasil. O impacto desse encontro é de extrema importância pois possibilita uma maior disseminação e expansão da Estatística Bayesiana na América Latina.

Heleno Bolfarine (IME-USP) Achei muito bom e importante o ISBA 2004 ter sido realizado na América do Sul. Destacaria as seções organizadas e apresentadas por estudantes. Acho que no próximo SINAPE e ISBRA deveríamos ter algo similar.

Márcia D'Élia Branco (IME-USP) Organização germânica e diversão latina. O que mais desejar? O ISBA/2004 foi um sucesso! Os bayesianos chilenos, liderados pela Pilar, mostraram que são capazes de organizar um grande Congresso Internacional de forma impecável. Eles estão de parabéns. (Vale lembrar que boa parte dos organizadores locais do congresso são ex-alunos do IME-USP). Desta vez, o IME participou com dois docentes, eu e o Heleno, e três alunos de doutorado: José Romeo, Víctor Hugo Lachos e Jorge Luis Bazán. A proximidade do local escolhido para o congresso permitiu a participação destes estudantes que voltaram do encontro muito animados com tudo que vivenciaram. E não foram somente eles. Voltei com muitas idéias novas e grande vontade de trabalhar; além de constatar satisfeita a evolução do nível da pesquisa desenvolvida no nosso país, evidenciada pela qualidade das apresentações dos brasileiros no congresso. Agora é começar os preparativos para o próximo encontro latino americano em fevereiro no México (COBAL II), onde espero ver uma participação ainda mais expressiva da comunidade estatística brasileira.

Marco Ferreira (DME-UFRJ) O que eu posso dizer? O ISBA 2004 foi um típico encontro Bayesiano, com um nível científico altíssimo tanto nas conferências plenárias e comunicações orais quanto nas sessões pôsteres. Foi importantíssimo tê-lo aqui na América Latina porque viabilizou a participação de vários pesquisadores seniores, jovens e alunos. Aliás, um motivo de orgulho muito grande foi ter dois alunos do programa de doutorado em estatística da UFRJ selecionados para

apresentação oral. Espero que a cultura dos congressos Bayesianos internacionais que dá incentivo à participação de jovens pesquisadores e alunos seja assimilada aqui no Brasil, afinal de contas, é isso que garante a continuação da obra de nossos ilustres pesquisadores seniores.

Roseli Aparecida Leandro (Depto de Ciências Exatas, ESALQ/USP) A nossa primeira participação em um encontro Bayesiano Internacional foi extremamente interessante, prazerosa e estimulante. Tivemos a oportunidade de desfrutar da companhia de pesquisadores nacionais e internacionais: iniciantes ou consagrados no mundo acadêmico, de ver o envolvimento, a dedicação, o profissionalismo, o brilhantismo e a participação efetiva dos Bayesianos Brasileiros. Sentir o amor à ciência, particularmente à Análise Bayesiana foi fato marcante e gratificante. Parabenizamos a Comissão Organizadora do evento pelo sucesso obtido. Sabemos que organizar encontros acadêmicos na América Latina é um desafio imenso por razões que todos conhecemos. Eles realmente se destacaram. Parabéns a todos pela participação e esperamos que num futuro próximo possa o Brasil ter o prazer de sediar e organizar tal evento. Finalmente, agradecemos a hospitalidade do povo Chileno.

O sucesso do ISBA 2004 deveu-se ao trabalho duro de muitas pessoas. Abaixo vocês podem encontrar os depoimentos de alguns dos envolvidos na organização.

Fabrizio Ruggeri (IMATI –Itália – Scientific Committee Chair) I have had the pleasure of chairing the Scientific Committee of ISBA 2004 and working for more than one year with the Chilean local organizers. I am very impressed for their commitment and the result of their efforts: one of the best organized conference. I have ever been. The organization of an ISBA World Meeting in South America could have been risky since many people are unwilling to spend their scarce research funds for such a long trip. Nonetheless, we have been able to present a scientific and social program which attracted more than 300 people, well above our expectations: a great success! I am very proud of the number of students (more than 100!) who attended the conference: most of them were from South America (and they were there mostly because of the efforts of Pilar Iglesias and Alicia Carriquiry, chairing, respectively, the Organizing and the Financial Committees. Fernando Quintana (the webmaster!), Gloria Icaza and their colleagues and students have been very important for the success of ISBA 2004. That was the scientific part of the conference. There was a social part, which ended for some of us few days later in Santiago with another meeting. I enjoyed a lot the moments I spent with my new South American friends, Pilar and Rosangela first of all! You will see me pretty soon in Chile and Brazil.

Peter Müller (Texas University –MDACC- ISBA Treasurer) The ISBA World meeting in Viña del Mar was an important event for the statistics community of Latin America. Latin American institutions support many outstanding researchers. But due to the limited number of institutions few departments reach the critical mass to achieve the international visibility and reputation required to attract the best graduate students, young researchers and external funding. While top quality research is of primary importance, visibility, reputation and funding are critical hygiene factors to the sustained success of a research group. It is in this respect that the ISBA World meeting provided important opportunities for researchers in Latin American institutions. An overwhelming proportion of invited talks and contributed presentations were by Latin American researchers, thus providing critical opportunities to present top level research carried out in Latin American institutions. Many of the presenters are outstanding scientists who for reasons of budget and geographic constraints do not usually have the opportunity to share their research with peers outside Latin America. The exposure provided at this meeting will translate into increased opportunities for the best researchers to competitive invitations to future international meetings, and to increased awareness and opportunities in the peer review process of top scientific journals. Another important aspect of the meeting was the opportunity for Latin American participants to interact with international peers and to learn about current trends and important research questions. This is especially important in Bayesian Statistics with research traditionally tightly linked to important scientific problems arising in application areas. Being in close touch with important application areas is important for success in our scientific journals. A distinguishing feature of the ISBA World meeting was the high level of student involvement, including several student run sessions scheduled for prime slots in the program. Besides advanced graduate students presenting in these sessions, a large number of graduate students from Latin American institutions had the opportunity to attend. Compared to other long established fields, like mathematics, engineering or biomedical sciences, statistics has the traditional handicap that potential students know very little about research in statistics and possible careers. The ISBA World meeting provided a terrific opportunity for students to learn about the exciting opportunities and challenges in statistics. We were stunned and thrilled by how many students responded to this opportunity, and how involved they became with the meeting. This bodes well for the future of top level statistics research in Latin America.

Para finalizar nosso informe, anexamos as palavras dos presidentes do ISBRA, Josemar Rodrigues, e em seguida o depoimento conjunto do Presidente da ISBA, Jim Berger, e do Coordenador Geral do ISBA 2004, Ed George.

7ISBA2004 – por Josemar Rodrigues

O 7º. Congresso Internacional da ISBA foi realizado em Vinha Del Mar- Valparaiso- Chile durante o período de 23-27/005/2004. É a primeira vez que a ISBA realiza este congresso em um país da América do Sul. Cada congresso realizado pela ISBA é um grande sucesso e uma confirmação das previsões dos professores Zellner e Lindley. A era Bayesiana é uma realidade dentro do cenário estatístico mundial que veio para ficar e contribuir para a solução de problemas complexos nas mais diferentes áreas. No Chile não foi diferente, motivado pelo aspecto multidisciplinar da metodologia bayesiana, os nossos queridos colegas chilenos (Fernando, Pilar, Reinaldo, Gloria e o outros) organizaram com um tremendo sucesso o 7ISBA2004. A programação científica foi organizada por Fabrizio que mais uma vez demonstrou a sua competência, criatividade e seu esforço incansável para dar aos jovens pesquisadores uma oportunidade para expor os seus trabalhos. O congresso reuniu mais de 300 pesquisadores e mais de 100 estudantes de diferentes países para discutir temas atuais envolvendo a metodologia bayesiana, tanto do ponto de vista teórico como aplicado. Para nós brasileiros e como presidente da ISBRA este congresso e o 7º EBEB, recentemente realizado em São Carlos, são fatos concretos que comprovam, definitivamente, que a metodologia bayesiana, como acontece em outros países, é uma realidade que não pode ser ignorada pela nossa comunidade estatística. O 7ISBA2004 foi majestoso, uma conferência de abertura apresentada pelo Prof. Dani Gamermann –UFRJ- e uma conferência de encerramento no Congresso chileno, com muita festa dentro do espírito bayesiano de Valencia e carinho no estilo Pilar. Como disse Fabrizio no recente boletim da ISBA, o 7ISBA2004 é história e sentiremos saudades dos agradáveis momentos que estivemos discutindo temas de interesse com os nossos amigos chilenos e de outros países que lá estiveram. Parabéns a comissão organizadora local pelo sucesso na divulgação da Estatística Bayesiana na América do Sul.

ISBA 2004 – A Smashing Success by Ed George (ISBA2004 Conference Chairman) and Jim Berger (President of ISBA)

It gives us great pleasure to report that our Seventh World Meeting, ISBA 2004 was a smashing success. The meeting took place in Viña del Mar - Valparaiso, Chile on May 23 – May 27, 2004, our first world meeting in South America. Attended by 302 participants from 28 countries, ISBA 2004 brought together Bayesians worldwide to share some of the most exciting current developments in Bayesian theory, methods and applications. The Scientific Committee, chaired by Fabrizio Ruggeri, put together an impressive, innovative program of 36 sessions with 108 talks, 2 poster sessions with 109 posters, opening and closing plenary lectures, 2 sessions devoted to debates of controversial topics and 6 introductory and

advanced tutorial sessions to begin the conference. The Local Organizing Committee, chaired by Pilar Iglesias, put it all together in glorious Chilean venues – a line of fabulous hotels, opening and closing ceremonies with inspiring local musical performances, a banquet dinner at a magnificent casino, surprise musical treats at the poster sessions, a day of excursions and parties, and a spectacular closing gala that lasted to sunrise. The Finance Committee, chaired by Alicia Carriquiry, secured funding for virtually all of the junior participants. These three committees did a superb job, and their heroic and unselfish efforts provided the lifeblood of the meeting. We are all forever grateful for their completely successful efforts.

This meeting will be very hard to beat! The breadth, the depth and the sheer intellectual scope of the presentations was amazing. There was lively and heated discussion throughout, but always with a common devotion for genuine scientific inquiry and a deep respect for the potential of Bayesian Analysis. As always, the meeting provided an environment for experiencing collegial interaction at its best. The camaraderie, the good will, the cooperative spirit was absolutely contagious. When we worked, we worked hard. And of course, when we played, we played just as hard. The ISBA World Meetings just keep getting better and better. The Program Committee is now accepting venue proposals for the next ISBA World Meeting which is set to take place in 2008. Also, ISBA will again be a major organizer of Valencia 8, to be held in Valencia, Spain in 2006. We look forward to seeing you at these and the many other exciting upcoming Bayesian meetings! ■

3 – Artigos

Seleção de Modelos de Fronteira de Produção Estocástica: Abordagem Bayesiana

Luís A. T. Medrano
IM – UFRJ

Hélio S. Migon
IM/COPPE – UFRJ

Resumo

Este artigo trata do problema que a maioria dos estudos a estimação empírica dos modelos de fronteira de produção estocástica utiliza-se, com frequência, a função Coob-Douglas, embora esta possa ser bastante restritiva para descrever certos processos produtivos. Outras especificações de funções de produção encontradas na literatura incluem: a Translog, CES (Constante Elasticity of Substitution) e GPF (Generalized Production Functions). O objetivo deste artigo é comparar essas funções de produção alternativas e modelar a distribuição das ineficiências, como por exemplo: a Gama ou a Lognormal. As comparações serão feitas através do critério DIC (Deviance Information Criterion) proposto por Spiegelhalter et al. 2002.

Palavras-chaves: Seleção de modelos, modelo de fronteira estocástica Bayesiano, MCMC, função de produção.

1 – Introdução

Economicamente é razoável admitir-se que as unidades produtivas operam próximas da fronteira de produção em razão, por exemplo, da alta competitividade dos mercados. Assim, seriam completamente eficientes, desviando-se da fronteira somente por acidentalidades incontroláveis. Todavia, devido a vários fatores, as firmas podem se distanciar da sua fronteira. Embora, considerações econômicas possam nos fornecer alguma idéia relativamente a eficiência das firmas, não se tem uma forma funcional exata para distribuição dessas quantidades. Esta questão será empiricamente abordada neste artigo, comparando-se os modelos de fronteira com especificações alternativas das distribuições das eficiências.

Na estimação empírica dos modelos de fronteira estocástica, a função Coob-Douglas, dada suas características simples e de fácil estimação. No entanto, ela impõe algumas restrições como elasticidades e retornos de escala constante, além de apresentar elasticidades de substituição iguais a um. Dentre outras formas postuladas na literatura para funções de produção destacamos: a Translog, CES (Constant Elasticity of Substitution) e GPF's (Generalized Production Functions).

O objetivo deste artigo é comparar essas funções de produção alternativas e selecionar a distribuição das

ineficiências, como por exemplo: a Gama ou a Lognormal. As comparações serão feitas através do critério de seleção DIC (Spiegelhalter et al. 2002). Além disto, os modelos serão estudados usando simulação estocástica. Em particular usaremos o software WinBugs (Bayesian analysis using Gibbs sampling), recém desenvolvido (Spiegelhalter, 1996), para ilustrar a implementação das análises Bayesianas.

O restante do artigo é organizado da seguinte forma. Na segunda seção são apresentadas as funções de produção e suas propriedades. Os modelos de fronteira de produção, que serão comparados neste estudo, são apresentados na seção 3. Os critérios de seleção de modelos são discutidos na quarta seção e os resultados das comparações entre os modelos de fronteira estocástica utilizando-se dados simulados e reais são apresentados na quinta seção. Na última seção são apresentados as conclusões e extensões deste estudo.

2 – Aspectos Básicos de Função de Produção

Dá-se o nome de função de produção à relação entre as quantidades produzidas e os insumos. O conhecimento desta relação é fundamental para o planejamento empresarial. A função de produção indica o máximo de produto que se pode obter a partir de diversas combinações dos insumos. Assim, a função de produção pode ser convenientemente visualizada como a **fronteira** entre as regiões viáveis e não viáveis no espaço dos *inputs* e *outputs*. De uma forma geral, a função de produção é o lugar geométrico de todas as possíveis combinações eficientes dos diversos insumos e das diversas variáveis tecnológicas e operacionais de um determinado sistema produtivo. Tradicionalmente, a função de produção pode ser representada por:

$$y = f(x_1, x_2) \quad (1)$$

onde y representa o produto e x_1 e x_2 são os insumos. Na próxima seção serão apresentadas algumas propriedades técnicas da função de produção.

2.1 – Retorno de Escala

Uma forma de analisar a função de produção é observar a taxa de variação do produto, quando todos os insumos variam na mesma proporção. O retorno de escala está ligado ao incremento obtido na produção quando todos os inputs são igualmente incrementados. Uma função de produção é dita homogênea de grau k , se pode ser expressa na forma expressada na:

$$y = f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^k f(x_1, x_2, \dots, x_n) \quad (2)$$

De acordo com a variação da quantidade produzida em função da variação da quantidade utilizada dos insumos é possível identificar-se três tipos de ganhos de escala: (i) retornos crescentes de escala ($k > 1$); (ii) retornos constantes de escala ($k = 1$) e (iii) retornos decrescentes de escala ($k < 1$).

2.2 – Elasticidade de Substituição

Outro conceito relevante para caracterizar as funções de produção é a “elasticidade de substituição”. Esta mede a variação percentual na proporção dos fatores devido a variação na taxa marginal de substituição técnica, ou seja, é uma medida da facilidade com que os fatores podem ser substituídos entre si. Pode ser vista também como uma medida da curvatura da isoquanta.

$$\varepsilon = \frac{d \left[\ln \left(\frac{x_2}{x_1} \right) \right]}{d \left[\ln \left(\frac{f_{x_2}}{f_{x_1}} \right) \right]} = \frac{d \left(\frac{x_2}{x_1} \right) \frac{f_{x_2}}{x_2} \frac{x_1}{f_{x_1}}}{d \left(\frac{f_{x_2}}{f_{x_1}} \right) \frac{f_{x_2}}{x_2} \frac{x_1}{f_{x_1}}} \quad (3)$$

$$= \frac{f_{x_2} f_{x_1} \left[f_{x_1} x_1 + f_{x_2} x_2 \right] x_1}{x_1 x_2 \det(B)} \quad (4)$$

onde $B = \begin{bmatrix} 0 & f_{x_1} & f_{x_1} f_{x_2} f_{x_1} & x_1 f_{x_2} & x_2 f_{x_2} f_{x_1} & x_2 f_{x_2} & x_2 \end{bmatrix}$ ou ainda como:

$$\varepsilon = \frac{r(x_2 R + x_1)}{x_1 x_2 \left(r \frac{\partial r}{\partial x_1} - \frac{\partial r}{\partial x_2} \right)} \quad (5)$$

onde $r = \frac{\partial y}{\partial x_2} / \frac{\partial y}{\partial x_1} = \frac{f_{x_2}}{f_{x_1}}$ e $0 \leq \varepsilon < \infty$.

Na elasticidade de substituição, quanto maior o valor melhor será a substituição de um input pelo outro. Se $\varepsilon = 0$, então a substituição não é possível. Já quando $\varepsilon = \infty$, os inputs são substitutos perfeitamente.

2.3 – Principais Tipos de Funções de Produção

A seguir serão descritas quatro classes de funções de produção. As duas primeiras apresentam elasticidade de substituição constante. Coob-Douglas e a CES. Para evitar este tipo de restrição, foi introduzido na literatura a classe das funções de produção Generalizada (GPF), que generalizam as anteriores em dois sentidos: na elasticidade de substituição e no retorno de escala.

2.3.1 – Função de Coob-Douglas

A função de produção de Coob-Douglas foi introduzida por Charles W. Coob e Paul H. Douglas (1928), tem como característica de ser simples e fácil estimar, tem sido bastante usada em análise econométrica. No entanto, ela impõe algumas restrições com elasticidades e retornos de escala constantes, além de apresentar elasticidades de substituição iguais a um. Em geral, a forma desta função para um produto Q , que está relacionado aos dosi fatores; capital e trabalho (K e L), é expressa por:

$$Q = \gamma K^{\beta_1} L^{\beta_2} \quad (6)$$

Ou, também pode ser escrita como

$$\log Q = \log \gamma + \beta_1 \log K + \beta_2 \log L \quad (7)$$

onde γ , β_1 , e β_2 são parâmetros tais que $\beta_1 > 0$ e $\beta_2 > 0$.

A elasticidade de substituição é expressa por:

$$\varepsilon = \frac{r(Kr + L)}{LK \left(r \frac{\partial r}{\partial L} - \frac{\partial r}{\partial K} \right)} = 1, \quad \text{onde } r = \frac{\beta_2 L}{\beta_1 K}, \quad \frac{\partial r}{\partial L} = \frac{\beta_2}{\beta_1 K} \text{ e}$$

$\frac{\partial r}{\partial K} = \frac{\beta_2 L}{\beta_1 K^2}$. O retorno de escala desta função é dada pela soma de β_1 e β_2 . Neste caso se $\beta_1 + \beta_2 = 1$ temos um retorno de escala constante.

2.3.2 - Função C.E.S (Constant elasticity of Substitution)

Uma extensão natural da família Coob-douglas seria considerar funções de produção com elasticidade de substituição constante, mas não necessariamente unitária para todos os pontos da curva. Uma classe especial de função de produção conhecido com Elasticidade Constante de Substituição (CES) foi introduzido por Arrow, Cherney, Minhas e Solow (1961). A função de produção C.E.S para dois insumos é dada pela seguinte expressão:

$$Q = \gamma \left[(1 - \delta) K^{-\rho} + \delta L^{-\rho} \right]^{-1/\rho} \quad (8)$$

ou, tomando o logaritmo pode ser escrito como

$$\log Q = \log \gamma + \nu \log \left\{ \left[(1 - \delta) K^{-\rho} + \delta L^{-\rho} \right]^{-1/\rho} \right\}$$

Onde γ, ν, δ e ρ são parâmetros tais que $0 < \gamma < \infty$, $0 < \delta < 1$, $-\infty < \nu < \infty$ e $-1 < \rho < \infty$. Fica clara a forma não linear da função e, também, uma estreita relação com a família de transformações Box-Cox (ver Zellner, pg. 173).

Embora não seja absolutamente necessário, para procedermos as inferências, é pratica comum se fazer uma expansão de Taylor em torno de $\rho = 0$ obtendo:

$$\log Q \approx \log \gamma + \nu(1 - \delta) \log K + \nu \delta \log L +$$

$$\rho \nu \delta (1 - \delta) \left\{ -\frac{1}{2} \log^2 (K/L) \right\} \quad (9)$$

$$= \beta_1 + \beta_2 \log K + \beta_3 \log L + \beta_4 x_4$$

onde $x_4 = -1/2 \log^2 (K/L)$, $\beta_1 = \log \gamma$, $\beta_2 = \nu(1 - \delta)$, $\beta_3 = \nu \delta$ e $\beta_4 = \rho \nu \delta (1 - \delta)$. Portanto, pela Eq. 9 podemos observar que quando $\rho = 0$, a função C.E.S tem a forma da função Cobb-Douglas.

A elasticidade de substituição desta função é dada por $\varepsilon = \frac{r(Kr + L)}{KL \left(r \frac{\partial r}{\partial L} - \frac{\partial r}{\partial K} \right)} = \frac{1}{1 + \rho}$, onde:

$$r = \left(\frac{\delta}{1 - \delta} \right) \left(\frac{K}{L} \right)^{1 + \rho}, \quad \frac{\partial r}{\partial K} = \frac{(1 + \rho) R}{K} \text{ e } \frac{\partial r}{\partial L} = \frac{(1 + \rho) R}{L}.$$

A função de produção CES tem o retorno de escala caracterizado pelo parâmetro ν . Por exemplo, se $\nu > 0$ temos um retorno de escala crescente.

2.3.3 – Função de Produção Translog

Muitas vezes deseja-se elaborar os modelos em formas funcionais mais flexíveis, permitindo-se que efeitos de segunda ordem sejam levados em consideração. Um exemplo desses efeitos são elasticidade de substituição, as quais são funções das derivadas segundas. Essas parcelas são obviamente desprezadas em modelos como os anteriores. A classe de funções que descrevemos agora serve como uma aproximação local de segunda ordem para uma função de produção geral, pelo menos duas vezes diferenciável. Fazendo-se uma aproximação de Taylor em torno de \mathbf{x}_0 , isto é:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)' \frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}_0) +$$

$(\mathbf{x} - \mathbf{x}_0)' \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}} f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)$ e considerando as derivadas parciais, avaliadas neste ponto, como sendo nossos parâmetros (constantes) obtemos a especificação denominada de Translog. Suponha que $\log Q = f(x_1, x_2)$, onde x_1 e x_2 são os logaritmos dos insumos e f uma forma funcional duas vezes diferenciável e que $x_0 = 0$. Neste caso, a função de produção Translog é dada pela seguinte expressão:

$$\log Q = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2} \beta_3 x_1^2 + \frac{1}{2} \beta_4 x_2^2 + \beta_5 x_1 x_2 \quad (10)$$

onde $x_1 = \log L$, $x_2 = \log K$, $\beta_0 = \log A$ e $0 < \beta_0 < 1$, $\beta_i > 0$ para $i = 1, \dots, 5$. Pela Eq. 10 podemos observar que quando $\beta_3 = \beta_4 = \beta_5 = 0$, a função Translog recai na forma funcional da Cobb-Douglas.

3 – Função de Produção Generalizada (GPF)

Zellner e Revankar (1970) propõem a função de produção generalizada (GPF's). Esta função foi introduzida com o objetivo de permitir uma generalização em duas direções. Primeiro, permite que a elasticidade de substituição possa ser constante, mas desconhecida, ou variável. A segunda generalização é permitir que os retornos de escala variem com o nível do produto.

Em termos determinísticos, considere a seguinte equação diferencial:

$$\frac{dQ}{df} = \frac{Q}{f} \frac{\alpha(Q)}{\alpha_f} \quad (11)$$

com solução $Q = g(f)$ onde $\alpha(Q)$ é o retorno de escala, função do produto Q , $f(L, K)$ é uma função de produção qualquer, onde L e K são os inputs capital e trabalho, respectivamente, α_f é parâmetro do retorno de escala associado a função de produção f . Suponha que g seja uma transformação monotônica de f de forma que a elasticidade de substituição associada a $Q = g(f)$ é a mesma da função de f .

Uma particular função GPF é determinada quando $\alpha(Q)$ tem a seguinte forma:

$$\alpha(Q) = \frac{\alpha}{1 + \lambda Q} \quad (12)$$

onde $\lambda > 0$ e $\alpha > 0$. É claro que para $\lambda = 0$ o retorno de escala não será função do produto (Q). Substituindo-se a expressão de $\alpha(Q)$ na Eq. (11), temos a seguinte

equação diferencial $\frac{dQ}{df} = \frac{Q}{f} \frac{\alpha}{\alpha_f (1 + \lambda Q)}$ cuja solução é

dada por

$$Q e^{\lambda Q} = C f^{\alpha/\alpha_f}$$

onde C é uma constante de integração.

Se, por exemplo, f tem a forma da função de produção Cobb-Douglas, ou seja, $f = AK^{\alpha_f(1-\delta)}L^{\alpha_f\delta}$, então obtemos a função de produção generalizada:

$$Q e^{\lambda Q} = \gamma K^{\alpha(1-\delta)} L^{\alpha\delta} \quad (13)$$

Com $\gamma = CA$, $0 < \delta < 1$ e γ . Ou, tomando-se o logaritmo de ambos os lados, tem-se:

$$z = y + \lambda \exp(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (14)$$

onde $y = \log Q$, $x_1 = \log K$, $x_2 = \log L$, $\beta_1 = \log \gamma$, $\beta_2 = \alpha(1-\delta)$ e $\beta_3 = \alpha\delta$. Pela Eq. (14) observa-se que, se $\lambda = 0$, a função GPF's, temos a forma da função Cobb-Douglas.

3 – Modelo de Fronteira de Produção Estocástica

Nesta seção, apresentaremos os modelos de fronteira de produção estocástica que serão comparados neste artigo. Nestes modelos, além das diferentes formas funcionais apresentadas na seção anterior, consideraremos duas distribuições alternativas para a representação do termo de ineficiência.

Inicialmente, os modelos de fronteira de produção estocástica foram propostos por Aigner, Lovell e Schmidt (1977) e Meeusen e Van Den Broecker (1977). Em sua forma geral tem-se:

$$y_i = f(x_i, \beta) - u_i + v_i \quad i = 1, \dots, N \quad (15)$$

onde y_i é, usualmente, o logaritmo do output, x_i denota o logaritmo do vetor de inputs, $f(\cdot, \cdot)$ é a função de produção determinística, v_i tem distribuição normal com média 0 e variância σ^2 , representando as flutuações aleatórias, que não depende da firma e u_i modela a eficiência da i -ésima firma, tendo uma distribuição assimétrica. Assume-se que v_i e u_i sejam independentes. A eficiência que corresponde a firma i é definida por $r_i = \exp(-u_i)$.

As distribuições assumidas para a componente de ineficiência foram, neste artigo, as distribuições gama e a log-normal. As distribuições log-normal e gamma têm comportamento bastante similar se seus parâmetros forem escolhidos adequadamente. Detalhes sobre equivalências dessas distribuições são encontradas em Migon (2002).

Por questões de parcimônia, no caso lognormal, iremos considerar somente a sub-família com parâmetro de locação igual a zero, destacando, entretanto que mesmo assim ela terá moda deslocada da origem o que é altamente desejável na modelagem das ineficiências. Um argumento bem conhecido na literatura da área destaca que a importância de se modelar as ineficiências através de distribuições com moda diferente de zero advém da compreensão de que fatores relacionados com a eficiência gerencial, provavelmente, não se distribuem segundo densidades que decaem monotonicamente na população em geral.

Os modelos comparados neste artigo seguem as quatro especificações de função de produção discutidas na seção anterior agregadas com um termo de erro composto. Na parcela do erro que modelara a ineficiência usaremos as duas distribuições discutidas acima. Assim ao todo findaremos com oito especificações de fronteira de produção estocástica, a saber

Modelo 1: Modelo usando a função de produção Cobb-Douglas considerando a distribuição do termo de ineficiência *Gama*.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - u_i + v_i \quad (16)$$

Modelo 2: Modelo com a função de produção CES e o termo de ineficiência com distribuição *Gama*.

$$y_i = \beta_0 + \nu \times \log \left\{ \left[(1-\delta) \exp(x_1)^{-\rho} + \delta \exp(x_2)^{-\rho} \right]^{-1/\rho} \right\} - u_i + v_i \quad (17)$$

com $v_i \sim N(0, \sigma^2)$ e $u_i \sim G(P, \theta)$.

Modelo 3: Usando a função de produção Translog e assumindo-se que o termo de ineficiência tenha distribuição *Gama*.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2} \beta_3 x_1^2 + \frac{1}{2} \beta_4 x_2^2 + \beta_5 x_1 x_2 - u_i + v_i \quad (18)$$

com $v_i \sim N(0, \sigma^2)$ e $u_i \sim G(P, \theta)$.

Modelo 4: Modelo utilizando a função de produção GPF com distribuição *Gama* no termo de ineficiência.

$$z_i = y_i + \lambda \exp(y_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - u_i + v_i \quad (19)$$

com $v_i \sim N(0, \sigma^2)$ e $u_i \sim G(P, \theta)$.

Como já mencionamos os Modelos de 5 a 8 têm as especificações acima, todavia com erro Lognormal com locação zero. Através da Eq. (19) podemos observar que a função de produção generalizada (GPF) usada nos Modelos 4 e 8, não explicita uma relação direta entre Y e X, sendo necessário definir o Jacobiano da transformação de Z para Y, que é dada pela seguinte expressão:

$$J = \prod_{i=1}^N (\lambda + 1/y_i) \quad (20)$$

A função de verossimilhança, para os modelo 4 e 8, é dada pela seguinte forma:

$$f(y_i | \beta, \sigma^2, \lambda, u_i) = f_N(z | x_i' \beta - u_i, \sigma^2) |J_i| \quad (21)$$

onde $z = y + \lambda \exp(y)$. A seguir, são descritas as distribuições a priori assumidas para os oito modelos.

3.1 – distribuição a Priori

A implementação da inferência segundo o paradigma Bayesiano requer que as distribuições a priori sejam especificadas para os parâmetros do modelo. Uma estratégia poderia ser eliciar a distribuição desses parâmetros consultando especialistas. Por exemplo, poderíamos arguir especialistas do setor que estivessemos modelando sobre sua opinião com respeito a ineficiências média das firmas e, também a variabilidade associada a esta informação. Baseado nessas informações obteríamos os parâmetros da distribuição da ineficiência, por exemplo da Gama, e retornaríamos a êle com um gráfico da densidade resultante e alguma característica, como por exemplo particulares percentis, a moda etc. Assim, ele teria a oportunidade de revisar as informações inicialmente fornecidas e, como fartamento documentado na literatura, em poucas iterações teria acessado a distribuição a priori que melhor descrevesse seu conjunto inicial de informações. Alternativamente, poderíamos simplesmente utilizar conhecimentos teóricos sobre prioris de referência, por exemplo utilizando prioris uniformes na reta toda para os parâmetros de locação e, também, para os parâmetros de escala, porém transformados pelo log. Neste artigo utilizaremos uma via alternativa. Usaremos prioris próprias, porém pouco informativas. Assim, para os parâmetros definidos na reta toda, proporemos uma normal centrada em zero e com precisão muito pequena, enquanto que, para aqueles definido nos reais positivos, utilizaremos uma distribuição gama com média em 1 e variância muito grande.

A priori definida para σ^2 foi: $\sigma^2 \sim GI(n_0, a_0)$.

Esta distribuição a priori é a mesma para todos os modelos analisados. A seguir, serão definidas as distribuições a priori para os parâmetros da componente de ineficiência:

Gama: $P \sim G(d_0, \varepsilon_0)$ e $\theta \sim G(\nu_0, \omega_0)$,

Lognormal: $\mu \sim N(a_0, A_0)$ e $\psi \sim G(d_0, D_0)$.

As distribuições a priori definidas para os demais parâmetros foram:

MODELO 1 e 5: $\beta_j \sim N(b_0, H_0), j = 1, \dots, 3$

MODELO 2 e 6: $\beta_0 \sim N(b_0, H_0)$, $\nu \sim N(m_0, s_0)$,
 $\delta \sim Beta(\omega_0, d_0)$ e $\rho \sim NT_{[-1, \infty]}(r_0, g_0)$

MODELO 3 e 7: $\beta_j \sim N(b_0, H_0), j = 1, \dots, 5$

MODELO 4 e 8: $\beta_0 \sim N(b_0, H_0)$ e $\lambda \sim U(0, 1)$, para $j = 1, \dots, 3$

Diferente de todas as distribuições a priori, a priori definida para o parâmetro ρ do modelo 2, não faz parte da lista de distribuições do WinBUGS. Logo, esta distribuição somente poderá ser implementada na versão 1.4 do WinBugs da seguinte forma:

Suponha que $y \sim f(y|\theta)$, onde f representa uma densidade não disponível no pacote. Seja $\Lambda \sim Po(\lambda)$ $\lambda > 0$. Logo com é sabido $P(\Lambda = 0|\theta) = \exp(-\lambda)$. Assim podemos definir $\lambda = -\log f(y)$ e então $P[\Lambda = 0] = f(y)$. Deste forma, podemos implementar qualquer distribuição que não está incluída na lista de distribuições do WinBUGS. Para mais detalhes, ver o manual de WinBUGS versão 1.4 (Spiegelhalter et al., 1999).

Os valores escolhidos para os hiperparâmetros que levaram a uma distribuição a priori relativamente vaga nos oito modelos foram:

$$\begin{bmatrix} d_0 & \varepsilon_0 \\ \nu_0 & \omega_0 \\ b_0 & H_0^{-1} \\ m_0 & \varepsilon_0^{-1} \\ \omega_0 & \varphi_0 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 \\ 0.01 & 0.01 \\ 0 & 0.001 \\ 0 & 0.001 \\ 1 & 1 \end{bmatrix}$$

3.2 – Distribuições condicionais a posteriori

Para aplicar o método de simulação estocástica na estimação desses modelos econométricos precisamos obter, sempre que possível, as distribuições condicionais completas. Se estas tiverem formas reconhecidas e forem disponíveis para amostragem serão utilizadas na implementação do método denominado Gibbs sampling. Caso essas distribuições não estejam disponíveis para amostragem deveremos utilizar alguma forma de proposta para gerar pontos no espaço dos parâmetros e aceitá-los ou não. Esses métodos são denominados de Metropolis-Hasting. No pacote que estaremos utilizando, WinBugs, estão disponíveis diferentes alternativas de geração de dados para implementação da simulação estocástica. Embora o pacote seja inteligente para escolha dessas alternativas é sempre bom analisarmos, analiticamente, a natureza dessas distribuições condicionais completas para verificarmos, por exemplo, se é log-concava, se estão definida numa região limitada etc.

A seguir vamos obter as a distribuição condicional completa para o termo de ineficiência u_i e os seus parâmetros associados:

- Gama

$$\begin{aligned} - (\theta | Y < X, resto) &\sim G\left(NP + \nu_0, \sum_{i=1}^N u_i + \omega_0\right) \\ - p(P | Y, X, resto) &\propto P^{d_0-1} \Gamma(P)^{-N} \exp(QP) I_{(0,\infty)} \end{aligned}$$

onde $Q = \sum_{i=1}^N \log \nu_i - \xi + N \log \theta$

$$\begin{aligned} - p(u_i | Y, X, resto) &\propto u_i^{p-1} \exp\left[-\frac{(-u_i + e_i)^2}{2\sigma^2} - \theta u_i\right], \\ &1, \dots, N \end{aligned}$$

Na fronteira de produção Normal-Gama as distribuições condicionais de P e u_i não estão disponíveis para amostragem. Infelizmente o método ARS (Adaptive Rejection Sampling) proposto por Gilks (1992) não pode ser aplicado, pois quando

$P < 1$ a distribuição condicional de u não é log-côncava. A estratégia adotada no WinBugs é gerar u univariadamente componente a componente através do método Slice Sampling (Neal, 1997). Como também a distribuição condicional de P não é log-côncava quando $d_0 < 1$, o método usado pelo WinBugs foi o algoritmo de Slice Sampling.

- Lognormal

$$\begin{aligned} - (\mu | Y, X, resto) &\sim N\left[\frac{\sum_{i=1}^N N \log u_i / \psi + A^{-1}}{a(N\psi + A^{-1})^{-1}}\right] \\ - (\psi | Y, X, resto) &\sim G\left[d + N, D + \sum_{i=1}^N (\log(u_i - \mu)^2) - 1\right] \\ - p(u_i | Y, X, resto) &\propto N[y_i - x_i' \beta, \sigma^2 LN[\mu, \psi]], \\ &1, \dots, N \end{aligned}$$

Na fronteira de produção Normal-Lognormal, as distribuições condicionais de e e σ estão disponíveis na forma fechada. Já a distribuição condicional de u_i não está disponível para amostragem, porem a distribuição é log-concava. Neste caso, o método utilizado pelo WinBugs para gerar valores de u foi o algoritmo de Slice Sampling.

As outras distribuições condicionais a posteriori para os oito modelos, estão brevemente descritos a seguir:

- MODELO 1 E 5

$$\begin{aligned} - (\beta | Y, X, resto) &\sim N_k\left[(H_0 + \sigma^2 X'X)^{-1}(H_0 b_0 + X'(Y + u)\sigma^2), H_0 + \sigma^2 X'X\right] \\ - (\sigma^2 | Y, X, resto) &\sim G\left[\frac{(N - N_0)}{2}, \frac{(y + u - X\beta)'(y + u - X\beta) + a_0}{2}\right] \end{aligned}$$

Nos modelos 1 e 5, onde é usada a função de produção Cobb-Douglas, as distribuições condicionais de e e σ^2 são conhecidas e disponíveis para amostragem.

- MODELO 2 E 6

$$\begin{aligned} - (\beta_0 | Y, X, resto) &\sim N\left[(H_0 + \sigma^2)^{-1}(H_0 b_0 + 1'(Y + u)\sigma^2), H_0 + \sigma^2\right] \\ - (\sigma^2 | Y, X, resto) &\sim G\left[\frac{(N - N_0)}{2}, \frac{(y + u - X^* \pi)'(y + u - X^* \pi) + a_0}{2}\right] \\ - p(\delta | Y, X, resto) &\propto \exp\left[-\frac{(y + u - X^* \pi)'(y + u - X^* \pi)}{2\sigma^2}\right] \\ - p(\nu | Y, X, resto) &\propto \exp\left[-\frac{(y + u - X^* \pi)'(y + u - X^* \pi)}{2\sigma^2} - \frac{s_0(\nu - m_0)^2}{2}\right] \\ - p(\rho | resto) &\propto \exp\left[-\frac{(y + u - X^* \pi)'(y + u - X^* \pi)}{2\sigma^2} - \frac{(\rho - \tau_0)^2}{2g_0}\right] I_{(-1,\infty)}(\rho) \end{aligned}$$

onde $X^* = \left[1 \log\left\{\left[(1 - \delta)\exp(X_1)^{-\rho} + \delta \exp(X_2)^{-\rho}\right]^{-1/\rho}\right\}\right]$ e $\pi = (\beta_0, \nu)$

Como a função de produção CES tem uma forma mais complexa, os parâmetros relacionados a esta função tem sua distribuição condicional não disponível para amostragem. Assim, o método utilizado pelo WinBugs para amostrar da distribuição condicional de δ e ν foi o algoritmo

do Slice Sampling e da distribuição condicional de o algoritmo de Metropolis-Hastings (Metropolis et al., 1953 e Hastings, 1970). Já os parâmetros relacionados a fronteira (β e σ^2) tem a sua distribuição condicional conhecida.

- **MODELO 3 E 7**

$$(\beta | Y, X, resto) \sim N^k \left[(H_0 + \sigma^2 X'X)^{-1} (H_0 b_0 + X'(Y+1)\sigma^{-2}), H_0 + \sigma^2 X'X \right]$$

$$- p(\sigma^2 | Y, X, resto) \propto \exp \left[\frac{(N - N_0)(y+u - X'\beta)'(y+u - X'\beta) + a_0}{2\sigma^2} \right]$$

onde $X^{*i} = [1 \quad X_i \quad X_2 \quad X_3 \quad 1/2X_2 \quad 1/2X_3 \quad X_2 X_3]$

X_i é a coluna i da matrix X .

Nos modelos 3 e 7, onde é utilizado a forma funcional Translog, as distribuições condicionais completas de β e σ^2 são conhecidas.

- **MODELO 4 E 8**

$$- (\beta | Y, X, resto) \sim N_k \left[(H_0 + \sigma^2 X'X)^{-1} (H_0 b_0 + X'(Z+u)\sigma^{-2}), H_0 + \sigma^2 X'X \right]$$

$$- (\sigma^2 | Y, X, resto) \sim G \left[\frac{(N - N_0)(z+u - X\beta)'(z+u - X\beta) + a_0}{2}, \frac{1}{2\sigma^2} \right]$$

$$- p(\lambda | resto) \propto \exp \left[\frac{1}{2\sigma^2} (z+u - X\beta)'(z+u - X\beta) + \sum_{i=1}^N \log(y_i \lambda + 1) \right]$$

onde $z = Y + \lambda \exp Y$

Já nos modelos 4 e 8 onde é usada a função de produção generalizada (GPF), a distribuição condicional de λ não se encontra na forma fechada. Neste caso, o método utilizado pelo WinBUGS para gerar valores de λ foi o algoritmo de Slice Sampling. Neste modelo, as distribuições condicionais completas para os parâmetros β e σ são conhecidas.

4 – Critérios de Seleção

Esta seção introduz duas medidas de diagnósticos que foram utilizados para medir o desempenho dos diferentes modelos estudados.

A escolha de modelos é uma atividade fundamental que vem tornando-se cada vez mais importante nas análises estatísticas, uma vez que, devido aos avanços computacionais, é possível construir modelos cada vez mais complexos. Tal complexidade normalmente aumenta de acordo com a estrutura imposta pelos modelos que requerem especificações em cada um de seus níveis

A seguir será apresentado o critério DIC (Deviance Information Criterion) para a seleção de modelos.

4.1 – Deviance Information Criterion (DIC)

Do ponto de vista frequentista a avaliação do modelo é baseado na deviance. Analogamente, Dempster (1974) sugere examinar a distribuição a posteriori da deviance clássica definido por:

$$D(\theta) = -2 \ln f(y | \theta) + 2 \ln g(y) \quad (22)$$

onde $f(y | \theta)$ é a função de verossimilhança e $\ln g(y)$ é um termo que só depende dos dados. Dempster (1974) propõe comparar a média à posteriori de $D(\theta)$ e Spiegelhalter et al. (2002) segue essa sugestão no desenvolvimento do DIC como um critério de escolha de modelos. Baseado na distribuição à posteriori de $D(\theta)$, o DIC consiste em duas componentes: um termo que mede a bondade do ajuste e outro termo de penalidade para o crescimento da complexidade do modelo.

$$DIC = \bar{D} + P_D \quad (23)$$

1. O primeiro termo é uma medida Bayesiana de ajuste do modelo, que é definido como a média a posteriori da deviance

$$\bar{D} = E_{\theta|y} [y] = E_{\theta|y} [-2 \ln f(y | \theta)] \quad (24)$$

2. A segunda componente mede a complexidade do modelo através do número efetivo de parâmetros p_D , definida como a diferença entre a média a posteriori da deviance e a deviance avaliada em $\bar{\theta}$, que é a média a posteriori do parâmetro θ :

$$P_D = \bar{D} - D(\bar{\theta}) = E_{\theta|y} [D] - D(E_{\theta|y} [\theta]) \quad (25)$$

5 – Aplicação

Nesta seção apresentaremos dois exemplos de estimação e seleção de modelos de fronteira de produção estocástica. O primeiro envolve um exercício de simulação com um conjunto de dados artificialmente gerados, tendo como objetivo verificar como se comportam os modelos descritos e a capacidade de discriminação do método proposto. No segundo exemplo utilizaremos um conjunto de dados reais analisados em alguns artigos da literatura. Neste exemplo, além do uso dos método de seleção de modelos propostos, mostraremos a distribuição dos postos das eficiências para algumas firmas selecionadas.

5.1 – Dados Artificiais

O objetivo desta simulação é ver se os critérios de seleção que serão utilizados, conseguem identificar o verdadeiro modelo do qual os dados foram gerados. Foram simulados um conjunto de dados de 100 observações do modelo de fronteira de produção Normal-Lognormal usando a função de produção Cobb-Douglas (MODELO 5). Este conjunto de dados artificiais contém 3 regressores com os coeficientes iguais a $\beta = (1,1,1)$. A matriz X foi gerada de uma $N(0,1)$ para cada uma das N firmas. Os restante dos parâmetros foram $\sigma^2 = 0.05$, $\psi = 1$ e $\mu = 0$.

Os resultados obtidos são baseados em 10.000 iterações das quais foram descartadas as 5.000 primeiras. Na Tabela 1 são apresentados os resultados da estatística DIC juntamente com \bar{D} e p_D para cada um dos oito modelos.

Tabela 1: Resultado do Deviance Information Criterion para os oito modelos de fronteira de produção para o conjunto de dados artificiais

Modelo	\bar{D}	$D(\theta)$	p_D	DIC
MODELO 1: Coob-Douglas e Gama	114,412	42,712	71,700	186,112
MODELO 2: CES e Gama	115,443	41,723	73,719	189,162
MODELO 3: Translog e Gama	124,268	50,555	73,713	197,981
MODELO 4: GPF e Gama	137,400	67,902	69,498	206,898
MODELO 5: Cobb-douglas e Lognormal	80,919	-2,617	83,536	164,454
MODELO 6: CES e Lognormal	80,804	-3,959	84,763	165,567
MODELO 7: Translog e Lognormal	89,46	4,575	84,885	174,345
MODELO 8: GPF e Lognormal	96,040	13,436	82,604	178,644

Pela Tabela 1 observamos que o modelo de fronteira de produção mais adequado para descrever o conjunto de dados artificiais, baseando-se no DIC, foi o modelo usando forma funcional Cobb-Douglas com termo de ineficiência Lognormal (MODELO 5), ou seja, o critério de seleção DIC conseguiu identificar o modelo verdadeiro.

5.2 – Dados Reais

Para ilustrar uma comparação com dados reais, serão utilizados os dados usados por Zellner e Revankar(1970) para estudar a produção da indústria de equipamentos de transportes de 25 estados, nos Estados Unidos, no ano de 1957. Neste conjunto de dados foram definidas as seguintes variáveis: $y = \log(\text{produção}/m)$, $x_1 = \log(\text{capital}/m)$ e $x_2 = \log(\text{trabalho}/m)$ onde m é o número de estabelecimentos em cada estado.

Todos os resultados encontrados foram baseados em amostras nas quais a convergência foi testada através do critério de Brooks e Gelman(1998) que está implementada na versão 1.4 do WinBugs.

Os resultados para os MODELOS 1-8 são baseados em 10.000 iterações. Foram realizadas 20.000 iterações das quais foram descartadas as 10.000 primeiras para o período de “burn-in”.

Na Tabela 2 são apresentados os resultados da estatística DIC juntamente com \bar{D} e p_D para cada um dos oito modelos.

Tabela 2: Resultado do Deviance Information Criterion para os oito modelos de fronteira de produção

Modelo	\bar{D}	$D(\theta)$	p_D	DIC
MODELO 1: Coob-Douglas e Gama	-20,485	-31,626	11,141	-9,344
MODELO 2: CES e Gama	-20,999	-31,035	10,037	-10,962
MODELO 3: Translog e Gama	-23,448	-37,120	13,672	-9,776
MODELO 4: GPF e Gama	-22,450	-33,671	11,221	-11,229
MODELO 5: Cobb-douglas e Lognormal	-37,6881	-60,218	22,530	-15,158
MODELO 6: CES e Lognormal	-38,881	-57,957	19,077	-19,804
MODELO 7: Translog e Lognormal	-40,291	-64,352	24,061	-16,231
MODELO 8: GPF e Lognormal	-38,9	-60,214	21,314	-17,586

Pela Tabela 2 podemos observar que os modelos de fronteira de produção mais adequado para descrever dados baseando-se no DIC foram os modelos usando forma funcional Translog e a função de produção Generalizada com termo de ineficiência Lognormal (Modelo 7 e 8).

Para concluir a comparação dos modelos, as distribuições dos ranks associada às medidas de eficiência (Goldstain e Spiegelhalter, 1996) são apresentadas. Essas distribuições são diretamente obtidas usando as cadeias de

MCMC. Se denotarmos por $e_j^{(m)} = \exp(-\mu_j^{(m)})$ como a medida de eficiência para a j^{th} firma na m^{th} iteração é a posição ocupada por $e_j^{(m)}$ na ordenação.

Figura 1-4 apresenta a distribuição a posteriori dos postos relativos às medidas de eficiência a melhor firma - rho[10] e para a pior firma - (rho[4]) segundo os modelos de fronteira Normal-Gama e Normal-Lognormal usando a função de produção: Cobb-Douglas, CES, Translog e GPF.

O objetivo desta última análise é comparar a capacidade dos modelos de classificar as firmas. Pelas Figuras 1-4 nota-se que o modelo de fronteira normal-lognormal com apenas um parâmetro e utilizando qualquer forma funcional, conseguiu diferenciar melhor as firma mais eficientes das menos eficientes. Também, podemos observar que a variabilidade da distribuição dos postos para uma considerada firma eficiente (rho[10]) utilizando o modelo de fronteira Normal-Lognormal com a função de produção CES é bem menor do que utilizando o modelo de fronteira Normal-Gama com a mesma forma funcional.

Figura 1: Distribuição a posteriori do rank para a melhor (rho[10]) e o pior (rho[4]) firma segundo o modelo de fronteira Normal-Lognormal usnado a função de produção Coob-douglas e Translog

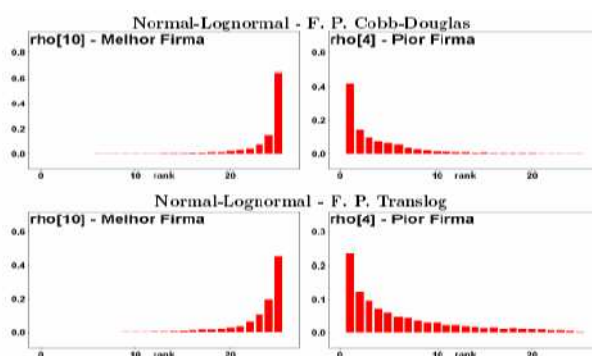


Figura 2: Distribuição a posteriori do rank para a melhor (rho[10]) e o pior (rho[4]) firma segundo o modelo de fronteira Normal- Lognormal usnado a função de produção CES e GPF

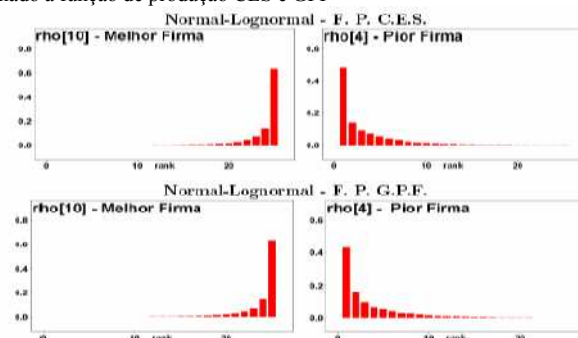


Figura 3: distribuição a posteriori do rank para a melhor ($\rho[10]$) e pior ($\rho[4]$) firma segundo o modelo de fronteira Normal-Gama usando a função de produção Cobb-Douglas e Translog

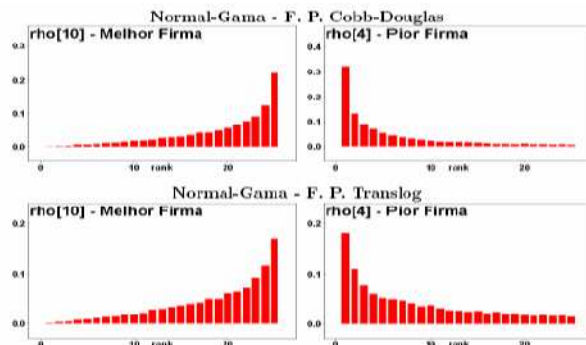
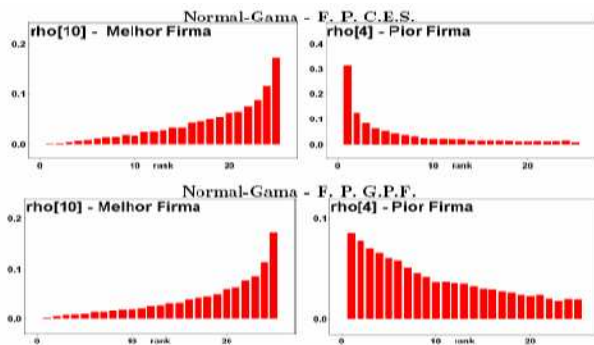


Figura 4: distribuição a posteriori do rank para a melhor ($\rho[10]$) e pior ($\rho[4]$) firma segundo o modelo de fronteira Normal-Gama usando a função de produção CES e GPF



6 – Conclusão e Extensões

Discutimos neste artigo duas questões centrais relacionadas com modelos de fronteira de produção estocástica, a especificação da parte determinística - função de produção - e da estrutura estocástica da eficiência. Inicialmente mostramos que o DIC é capaz de discriminar eficientemente os modelos propostos, selecionando tanto a função de produção quanto a componente de erro assimétrico. Baseado em dados reais, pudemos verificar a importância da componente de erro log-normal quando comparada com a gama. Além de mais parcimoniosa, pois fixamos $\mu = 0$, produz valores de DIC favoráveis a ela para todos os modelos propostos de função de produção.

Em resumo, a aplicação com dados artificiais mostrou que o critério de seleção DIC conseguiu identificar o modelo onde foram gerados os dados. Já a ilustração empírica envolvendo dados reais identificou que o modelo que se mostrou mais apropriado, foi o fronteira de produção com termo de ineficiência lognormal e usando a forma funcional CES (Constant Elasticity of Substitution) ou a GPF (Generalized Production Functions). Notou-se, também, que o modelo de fronteira normal-lognormal com apenas um parâmetro e utilizando qualquer forma funcional, conseguiu diferenciar melhor as firma mais eficientes das menos eficientes.

Em investigações futuras iremos comparar esses procedimentos em dados de painel e comparar os algoritmos utilizados no Gibbs sampling.

Referências

- [1] Aigner, D., C. Lowell, e P. Schmidt (1977) - Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, **6**, 21-37.
- [2] Arrow, K.J., Chenery, H.B., Minhas, B.S. and Solow, R.M. (1961) Capital-Labor Substitution and Economic Efficiency. *Review of Economics and Statistics*, Vol. **43**, p.225-50.
- [3] Brooks S P and Gelman A (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. **7**, 434-455.
- [4] Cobb, C.W. and Douglas, P.H. (1928) A Theory of Production. *American Economic Review*, Vol. **18**, p.139-65. B. Cornet (1982)
- [5] Geisser, S. e Eddy, W. (1979) A predictive approach to model selection. *J. Am. Statist. Ass.*, **74**, 153-160.
- [6] Gelfand, A. E. e Ghosh, S. K. (1998) - Model choice: a minimum posterior predictive loss approach, *Biometrika*, **85**, 1-11.
- [7] Gilks W. (1992) Derivative-free adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, **41** 337-348.
- [8] Hasting, W. K. (1970) Monte Carlo sampling methods using Markov chain Monte Carlo with applications. *J. Am. Statist. Ass.*, **93**, 585-595.
- [9] Kass, R. E. e Raftery, A. E. (1995) - Bayes Factor, *JASA*, **90**, n. 430. 773-7796.
- [10] Meeusen, W. e J. van den Broeck (1977) Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* **8**, 435-444.
- [11] Metropolis, N. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087-1091.
- [12] Neal R (1997) - *Markov chain Monte Carlo methods based on 'slicing' the density function*, Department of Statistics, University of Toronto.

- [13] Spiegelhalter, D. J., Thomas, A., Best, N. G. e Lunn, D. (1999). WinBUGS Version 1.4 User Manual, MRC Biostatistics Unit
- [14] Spiegelhalter, D. J., Best, N. G. Carlin, B. P. e van der Linde, A. (2002), "Bayesian measure of model complexity and fit", *Journal of the Royal Statistical Society, Series B*, **64**, Parte 3, forthcoming.
- [15] Stevenson, R. E. (1980). Generalised Stochastic Frontier Estimation. *Journal of Econometrics*, **13**, 57-66.
- [16] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, **36**, 111-147.
- [17] Zellner, A. e N. Revankar (1970), Generalized Production Functions," *The Review of Economic Studies*, **36**, 241-250. ■

Inferência Bayesiana Em Modelos De Volatilidade Estocástica Usando Processos Autoregressivos Com Transição Suave

Esther Salazar

Departamento de Métodos Estatísticos
Universidade Federal do Rio de Janeiro

Hedibert F. Lopes

Graduate School of Business
University of Chicago

Resumo

Este trabalho tem dois objetivos principais. O primeiro é fazer inferência Bayesiana para todos os parâmetros envolvidos no modelo logístico auto regressivo de transição suavizada de ordem k , LSTAR(k). O segundo objetivo é modelar a volatilidade estocástica de séries temporais econômicas e financeiras supondo que o logaritmo da variância condicional segue um modelo LSTAR(k). Por tanto, é proposto um modelo dinâmico Bayesiano não-linear e não-normal, cujos estados (log-volatilidades) evoluem segundo uma estrutura LSTAR(k). Estudos simulados e aplicações à séries financeiras testam, respectivamente, a modelagem proposta e sua aplicabilidade a situações reais.

Palavras-chave: Monte Carlo via Cadeias de Markov, Inferência Bayesiana, MCMC com saltos reversíveis, Seleção de Modelos, Modelo Dinâmico.

1 – Introdução

Os modelos autoregressivos de transição suavizada (STAR), proposto inicialmente na forma univariada por Chan e Thong (1986) e desenvolvido posteriormente por Luukkonen et al. (1988) e Teräsvirta (1994), têm sido Nesse trabalho, consideramos uma importante classe dos modelos STAR, chamado modelo logístico STAR de ordem k , LSTAR(k). A inferência para este tipo de modelo não é trivial pois a estimação dos parâmetros de suavização, γ , e de locação, c , que definem a transição logística, requer cuidado especial. Na literatura existem várias propostas para fazer inferência nos modelos LSTAR (Luukkonen, 1990) pro exemplo, o uso do método de mínimos quadrados ponderados.

A primeira, e talvez principal, contribuição desse trabalho é a proposta de um algoritmo Monte Carlo via Cadeias de Markov (MCMC) para fazer inferência exata dos modelos LSTAR(p) sob o enfoque Bayesiano. Adicionalmente, estendemos o tradicional modelo de volatilidade estocástica, onde o log-volatilidade evolui segundo um processo autoregressivo de ordem 1, para um processo LSTAR(k).

O resto do artigo está organizado como segue. Na Seção 2 é apresentado o modelo LSTAR(k) incluindo a especificação do modelo. Na Seção 3 são apresentadas as priors associadas aos parâmetros do modelo e descrevemos o algoritmo MCMC para simulação da posteriori dos parâmetros. Na Seção 4 propomos o modelo de volatilidade estocástica com LSTAR(k) e apresentamos o processo de estimação. Na Seção 5 aplicamos os modelos a séries simuladas e reais. Conclusões e extensões são apresentados na Seção 6.

2 – Modelos LSTAR de ordem k

Os modelos LSTAR de ordem k denotado por LSTAR(k), são modelos auto-regressivos que incorporam na sua estrutura um termo de suavização. Esse termo tem, em geral, a forma de um modelo AR(k), ponderado por uma função, F , cuja forma determina a transição. Nesse estudo consideramos a função logística, por ser a mais comumente encontrada na literatura. Mais especificamente, o LSTAR(k) é escrito da seguinte forma,

$$y_t = x_t' \theta_1 + x_t' F(\gamma, c, y_{t-d}) \theta_2 + \varepsilon_t \quad (1)$$

onde $x_t' = (1, y_{t-1}, \dots, y_{t-k})$, $\theta_i' = (\theta_{0i}, \theta_{1i}, \dots, \theta_{ki})$ para $i = 1, 2$,

$$\varepsilon_t \sim N[0, \sigma^2] \quad e$$

$F(\gamma, c, y_{t-d}) = \{1 + \exp(-\gamma(y_{t-d} - c))\}^{-1}$. O parâmetro $\gamma > 0$ é responsável pela suavidade da função F , enquanto o escalar c é o parâmetro de locação (o parâmetro threshold). Finalmente, d é um parâmetro de retardo (*delay*) e y_{t-1} é a variável de transição. Assuma-se ainda, e sem perda de generalidade, que $d \leq k$ e que $y_{-k+1}, \dots, y_0, y_1, \dots, y_T$ são observações disponíveis.

O modelo LSTAR(k) pode ser visto como um modelo de dois regimes, mais especificamente temos que $E(y_t | x_t) = x_t' \theta_1$ e $E(y_t | x_t) = x_t' \theta_2$ representam o

primeiro e segundo regime respectivamente. Assim, podemos expressar o modelo da seguinte forma,

$$y_t = (1 - F(\gamma, c, y_{t-d}))x'_t\theta_1 + (1 - F(\gamma, c, y_{t-d}))x'_t\theta_2 + \varepsilon_t \quad (2)$$

onde ε_t é normalmente distribuído com média zero e variância constante σ^2 sobre os dois regimes. Além disso, a escolha entre os dois regimes é representada pela função logística $F(\gamma, c, y_{t-d})$. O modelo (2) pode ser expressado como em (1) onde $\theta_2 = \delta_2 - \theta_1$ representa a contribuição de considerar um segundo regime.

Finalmente, consideramos a seguinte reparametrização que será muito útil no cálculo das distribuições a posteriori dos parâmetros envolvidos no modelo,

$$y_t = x'_t(\gamma, c, d)\theta + \varepsilon_t \quad (3)$$

onde $\theta' = (\theta'_1, \theta'_2)$ representa o vetor de parâmetros lineares, (γ, c) são parâmetros não lineares e $x'_t(\gamma, c, d) = [x'_t, F(\gamma, c, y_{t-d})x'_t]$.

3 – Inferência a Posteriori

Nesta seção apresentamos duas propostas para fazer inferência dos parâmetros. A primeira, considerando prioris subjetivas e a segunda usando prioris não informativas. Num primeiro análise, apresentamos propostas para os parâmetros lineares θ , para os parâmetros não lineares γ e c e para a variância σ^2 , ou seja temos que estimar $2k+5$ parâmetros. Em seguida, propomos a estimação do parâmetro de retardo d e por último fazemos uma proposta para a estimação da ordem de auto-regressão k , a traves de um algoritmo Monte Carlos via Cadeias de Markov com saltos reversíveis.

3.1 – Distribuições a priori

As distribuições a priori consideradas neste trabalho são de dois tipos. A primeira, de tipo subjetivo, assumindo que temos algum tipo de informação ao respeito dos dados. A segunda considerando prioris Cauchy e não informativas.

Prioris subjetivas

Por simplicidade e sem perda de generalidade, adotaremos prioris condicionalmente conjugadas. Ou seja,

$$\theta \sim N(m_0, \sigma_0^2 I_{2k+2}), \quad \gamma \sim G(a, b), \quad c \sim N(m_c, \sigma_c^2)$$

$$\text{e } \tau^2 \sim GI\left(\frac{\nu}{2}, \frac{\nu s^2}{2}\right). \quad \text{Os hiperparâmetros}$$

$m_0, a, b, m_c, \sigma_c^2, \nu$ e s^2 são escolhidos de forma que as prioris sejam pouco informativas, como ilustrado nas aplicações da seção 4.

Prioris não informativas

Lubrano (2000) propõe as seguintes prioris para os parâmetros lineares e não lineares:

$(\theta_2 | \sigma^2, \gamma) \sim N_{k+1}(0, \sigma^2 \exp(\gamma) N_0^{-1})$, a priori condicional de θ_2 é informativa quando $\gamma \rightarrow 0$, nesse caso temos que θ_2 é zero, quando $\gamma = 0$ a precisão é igual a N_0 / σ^2 , por

outro lado a precisão é zero quando γ é positiva. A distribuição a priori para γ é a Cauchy truncada dada pela seguinte expressão,

$$\pi(\gamma) = \begin{cases} (1 + \gamma^2)^{-1} & \gamma > 0, \\ 0 & \text{caso contrário.} \end{cases} \quad (4)$$

A amplitude de variação do parâmetro threshold c é determinado pelo mínimo e máximo de y_t assim definimos uma priori uniforme tal que $\pi(c) \sim U[c_a, c_b]$ onde $D(c_a) = 0.15$, $D(c_b) = 0.85$ e D representa a distribuição empírica de y_t . Finalmente, considera-se uma priori não informativa para σ^2 tal que $\pi(\sigma^2) \propto \sigma^{-2}$.

MCMC em modelos LSTAR(k)

Assumiremos nesta seção que k é conhecido. Dessa forma, a simulação da posteriori de θ, γ, c e σ^2 é feita a partir de um algoritmo MCMC, que simula das distribuições condicionais completas a posteriori. Mais especificamente, as distribuições condicionais completas de θ e σ^2 (independentemente do tipo de priori utilizada), são respectivamente, normal e gama inversa, possibilitando o uso do Amostrador de Gibbs (Gelfand e Smith, 1990). Por outro lado, a distribuição condicional completa conjunta de γ e c não é conhecida, o que sugere o uso de um algoritmo do tipo Metropolis-Hastings (Metropolis, Rosenbluth, Rosenbluth, Teller e Teller, 1953) - (Hasting, 1970). Para mais detalhes sobre o Amostrador de Gibbs o algoritmo de Metropolis-Hastings e outros métodos MCMC consultar, por exemplo o compêndio editado por Gilks, Richardson e Spiegelhalter (1996) ou os livros de Gamerman (1997) e Robert e Casella (1999).

Abaixo, $[\theta]$ denota a distribuição condicional completa a posteriori do parâmetro θ e $N_p(\mu, \Sigma)$ denota a distribuição normal p -variada com vetor de média μ e matriz de covariância Σ . Finalmente, todos os somatórios sem índices são subentendidos entre 1 a T .

As condicionais completas obtidas a partir das prioris subjetivas são as seguintes:

$[\theta]$ Combinando o modelo

$$y_t \sim N_{2k+2}(x'_t(\gamma, c, y_{t-d})\theta, \sigma^2) \text{ com a priori}$$

$$\theta \sim N(m_0, \sigma_0^2 I_{2k+2}), \text{ é fácil verificar que}$$

$$[\theta] \sim N(u_0, C_0) \text{ onde}$$

$$C_0^{-1} = \{\sigma_0^{-2} I_{2k+2} + \sigma^{-2} \sum x_t(\gamma, c, d)x'_t(\gamma, c, d)\}^{-1} \text{ e}$$

$$u_0 = C_0 \{\sigma^{-2} \sum x_t(\gamma, c, d)y_t + \sigma_0^{-2} m_0\}.$$

$[\tau^2]$ Amostrar τ^2 é mais simples. Primeiro definimos

$$\varepsilon_t = y_t - x'_t(\gamma, c, d)\theta, \text{ de modo que}$$

$$\varepsilon_t \sim N(0, \sigma^2) \text{ e portanto,}$$

$$[\sigma^2] \sim GI\left[(T + \nu)/2, (\nu s^2 + \sum \varepsilon_i^2)/2\right].$$

$[\gamma, c]$ As condicionais completas de γ e c não têm forma fechada. Como sugerido na seção 2 amostramos γ e c conjuntamente usando o algoritmo de Metropolis-Hastings com a seguinte proposta de transição: $\gamma^* \sim G\left[(\gamma^{(i)})^2 / \Delta_\gamma, \gamma^{(i)} / \Delta_\gamma\right]$ e $c^* \sim N(c^{(i)}, \Delta_c)$, onde $\gamma^{(i)}$ e $c^{(i)}$ são valores correntes de γ e c . Denota-se $f_G(x|a, b)$ a densidade da $G(a, b)$ no ponto x e $f_N(x|a, b)$ a densidade da $N(a, b)$ no ponto x . O par (γ^*, c^*) é aceito com probabilidade igual ao mínimo entre a unidade e α , onde

$$\alpha = \frac{\prod_{t=1}^T f_N(y_t | x'_t(\gamma^*, c^*, d)\theta, \sigma^2) f_G(\gamma^* | a, b) f_N(c^* | m_c, \sigma_c^2)}{\prod_{t=1}^T f_N(y_t | x'_t(\gamma^{(i)}, c^{(i)}, d)\theta, \sigma^2) f_G(\gamma^{(i)} | a, b) f_N(c^{(i)} | m_c, \sigma_c^2)} \times \frac{f_G(\gamma^{(i)} | (\gamma^*)^2 / \Delta_\gamma, \gamma^* / \Delta_\gamma)}{f_G(\gamma^* | (\gamma^{(i)})^2 / \Delta_\gamma, \gamma^{(i)} / \Delta_\gamma)}$$

Analogamente, as condicionais completas a posteriori calculadas a partir das priors não informativas são:

$[\theta]$ Combinando o modelo

$$y_t \sim N_{2k+2}\left(x'_t(\gamma, c, y_{t-d})\theta, \sigma^2\right) \text{ com a priori } (\theta_2 | \sigma^2, \gamma) \sim N_{k+1}(0, \sigma^2 \exp(\gamma) N_0^{-1}) \text{ temos que } [\theta] \sim N_{2k+2}(u_0^*, C_0^*) \text{ onde}$$

$$C_0^* = \left(\sum x_t(\gamma, c, y_{t-d})x'_t(\gamma, c, y_{t-d})\sigma^{-2} + \Sigma^{-1}\right), u_0^* = \Sigma_0 \left(\sum x_t(\gamma, c, y_{t-d})y_t\sigma^{-2}\right) \text{ e } \Sigma \text{ é uma matriz de dimensão } (2k + 2 \times 2k + 2) \text{ tal que}$$

$$\Sigma^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & N_0/\sigma^2 \exp(\gamma) \end{pmatrix} \quad (5)$$

A matriz N_0 pode ser a matriz identidade o um múltiplo dela. Neste artigo nos consideramos a primeira opção.

$[\sigma^2]$ Do modelo (3) tem se que $\varepsilon_t = y_t - x'_t(\gamma, c, y_{t-d})\theta$.

Assim, combinando com a priori não informativa $\pi(\sigma^2) = \sigma^{-2}$ temos que

$$[\sigma^2] \sim GI\left[(T + k + 1)/2, (\exp(-\gamma)\theta_2^2\theta_2 + \sum \varepsilon_i^2)/2\right].$$

$[\gamma, c]$ De novo, amostramos γ e c conjuntamente usando o algoritmo de Metropolis-Hastings usando as seguintes propostas de transição:

$$\gamma^* \sim G\left[(\gamma^{(i)})^2 / \Delta_\gamma, \gamma^{(i)} / \Delta_\gamma\right] \text{ e } c^* \sim Nrr_{[c_a, c_b]}(c^{(i)}, \Delta_c),$$

onde $\gamma^{(i)}$ e $c^{(i)}$ são os valores correntes de γ e c . A probabilidade de aceitação do par (γ^*, c^*) é o mínimo entre a unidade e α tal que:

$$\alpha = \frac{\prod_{t=1}^T f_N(\varepsilon_t^* | 0, \sigma^2) f_N(\theta_2 | 0, \sigma^2 \exp(\gamma^*) N_0^{-1}) \pi(\gamma^*) \pi(c^*)}{\prod_{t=1}^T f_N(\varepsilon_t^{(i)} | 0, \sigma^2) f_N(\theta_2 | 0, \sigma^2 \exp(\gamma^{(i)}) N_0^{-1}) \pi(\gamma^{(i)}) \pi(c^{(i)})} \times \frac{\left[\phi\left(\frac{c_b - c^{(i)}}{\sqrt{\Delta_c}}\right) - \phi\left(\frac{c_a - c^{(i)}}{\sqrt{\Delta_c}}\right)\right] f_G(\gamma^{(i)} | (\gamma^*)^2 / \Delta_\gamma, \gamma^* / \Delta_\gamma)}{\left[\phi\left(\frac{c_b - c^*}{\sqrt{\Delta_c}}\right) - \phi\left(\frac{c_a - c^*}{\sqrt{\Delta_c}}\right)\right] f_G(\gamma^* | (\gamma^{(i)})^2 / \Delta_\gamma, \gamma^{(i)} / \Delta_\gamma)}$$

onde $\varepsilon_t^* = y_t - x'_t(\gamma^*, c^*, y_{t-d})\theta$, $\varepsilon_t^{(i)} = y_t - x'_t(\gamma^{(i)}, c^{(i)}, y_{t-d})\theta$ e $\phi(\cdot)$ é a função de probabilidade da distribuição normal padrão.

Inferência sobre o parâmetro de retardo d é bem simples. Seja $p(d)$ a probabilidade a priori de um particular valor de d , entre 1 e d_m . A distribuição a posteriori condicional para d é $p(d | y, \theta) \propto p(y | d, \theta) p(d)$, que pode ser trivialmente calculado para todos os valores de d . Dessa forma, (θ, d) formam o novo vetor de parâmetros. Após a convergência das cadeias de Markov tenham sido alcançadas, a seqüência $d^{(1)}, d^{(2)}, \dots, d^{(M)}$ representa uma amostra da posteriori marginal de d , i. e. $\Pr(d | y)$.

3.2 – RJMCMC em modelos LSTAR(k)

Propostas para a incorporação de k dentro de uma modelagem completamente Bayesiana já existe na literatura. Huertas e Lopes (2000), por exemplo, analisam a produção industrial brasileira através de um modelo AR(k) onde $k = 2c + s$, isto é, c pares de raízes complexas conjugadas e s raízes reais, para c e s , e por tanto k , desconhecidos. Troughton e Godsill (1997) sugere um algoritmo RJMCMC para modelos autoregressivos de ordem k , AR(k). Generalizamos o trabalho de Troughton e Godsill (1997) ao propor um algoritmo RJMCMC (Green, 1995) que permita a incorporação de k no vetor de parâmetros em modelos LSTAR(k).

No caso do modelo LSTAR(k) o vetor de parâmetros θ de dimensão $(2k + 2 \times 1)$, definido em (1), é amostrado de acordo à ordem de autoregressão k . Porem, amostrar a ordem k implica câmbios no espaço paramétrico do modelo. A idéia é usar movimentos de saltos reversíveis, onde $\phi = (\gamma, c, \sigma^2, d)$ permanece invariante durante o movimento. Supondo que num determinado passo do algoritmo RJMCMC tem-se (ϕ, k) , o algoritmo cicla de acordo com os seguintes passos: Passo I: Amostra-se uma nova ordem k' a partir da proposta $q(k' | k)$,

Passo II: Amostra-se θ da condicional completa a posteriori $p(\theta^{(k)} | y, k', \phi)$ (Seção 3),

Passo III: Calcula-se a probabilidade de aceitação $\alpha\left\{\left(k', \theta^{(k)} \rightarrow \left(k', \theta^{(k')}\right)\right)\right\}$,

Passo IV: Move-se para $\left(k', \theta^{(k')}\right)$ com probabilidade α .

A probabilidade de aceitação do movimento é obtida por

$$\alpha\left\{\left(k, \theta^{(k)} \rightarrow \left(k', \theta^{(k')}\right)\right)\right\} = \min\left\{1, \frac{p\left(k', \theta^{(k')}\right) q\left(k | k'\right) p\left(\theta^{(k)} | y, k, \phi\right)}{p\left(k, \theta^{(k)} | y, \phi\right) q\left(k' | k\right) p\left(\theta^{(k')} | y, k', \phi\right)}\right\} \quad (6)$$

que pode ser simplificada a partir da fórmula do candidato (Besag, 1989)

$$\alpha\left\{\left(k, \theta^{(k)} \rightarrow \left(k', \theta^{(k')}\right)\right)\right\} = \min\left\{1, \frac{p\left(k' | y, \phi\right) q\left(k | k'\right)}{p\left(k | y, \phi\right) q\left(k' | k\right)}\right\} \quad (7)$$

Condicional a ϕ , a probabilidade a posteriori para a ordem do modelo $p(k | y, \phi)$ pode ser reescrita por

$$p(k | y, \phi) \propto p(k) \int p(y | k, \theta, \phi) p(\theta | k) d\theta \quad (8)$$

Conseqüentemente, de acordo ao tipo de priori utilizada (subjetivas e objetivas, nesse ordem), encontramos as seguintes expressões.

$$p(k | y, \phi) \propto (2\pi\sigma)^{-\frac{T}{2}} \sigma_0^{-2k+k_0} |C_0| \exp\left\{-\frac{1}{2}(\sigma^{-2} y^T y + \sigma_0^{-2} m_0^T m_0 - u_0^T C_0^{-1} u_0)\right\}$$

$$p(k | y, \phi) \propto (2\pi\sigma)^{-\frac{T}{2}} (2\pi)^{-\frac{k+1}{2}} [\sigma^2 \exp(\gamma)]^{\frac{k+1}{2}} |C_0^*|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\sigma^{-2} y^T y + -u_0^{*T} C_0^{*-1} u_0^*)\right\}$$

onde u_0^* , u_0^* , C_0 e C_0^* são definidos na seção 3.

4 – Volatilidade Estocástica

Um dos modelos mais simples para a volatilidade estocástica de séries financeiras assume que o retorno, y_t , são tais que $y_t | h_t \sim N(0, h_t)$ e a variância evolue temporalmente segundo o processo AR(1), $\lambda_t | \lambda_{t-1} \sim N(\phi_{01} + \phi_{11}(\lambda_{t-1} - \phi_{01}), \tau^2)$, para $\lambda_t = \log h_t$.

Nessa seção generalizamos esse modelo básico ao permitir que $\lambda_t \sim LSTAR(p)$, ou seja que o log-volatilidade evolua segundo um modelo logístico autoregressivo de transição suavizada de ordem p . O objetivo é fazer inferência sobre o logaritmo da variância das observações para cada tempo e os parâmetros do modelo LSTAR(p). Este problema pode ser facilmente tratado se observarmos que (i) $y_t | h_t \sim N(0, h_t)$ e (ii)

$$\lambda_t | \lambda_{t-1} \sim N(g(\theta, \lambda_{t-1}), \tau^2),$$

formam um modelo dinâmico não-normal e não linear (West e Harrison, 1997), onde $\lambda_{t-1} = (\lambda_{t-1}, \dots, \lambda_{t-p}, \lambda_{t-d})$ e a evolução do sistema é dada por

$$g(\theta, \lambda_{t-1}) = \phi_{01} + \sum_{i=1}^p \phi_{1i} \lambda_{t-i} + \left(\phi_{02} + \sum_{i=1}^p \phi_{2i} \lambda_{t-i}\right) F(\gamma(\lambda_{t-d} - c)).$$

Inferência sobre h_1, \dots, h_T é obtida através do algoritmo MCMC proposto por Jarquier, Polson e Rossi (1994) para o caso do modelo de volatilidade estocástica simples. A inferência sobre θ (do modelo LSTAR) é feita através do algoritmo proposto na Seção 2. Mais especificamente, a condicional completa para cada um dos λ_t é tal que

$$p(\lambda_t | \lambda_{-t}, \theta, y) \propto p(y_t | \lambda_t) \prod_{i=0}^p p(\lambda_{t+i-1}, \theta).$$

Utilizamos, como proposta, $\lambda_t^* \sim N(\lambda_t^{(i)}, \Delta)$, para Δ escolhido convenientemente.

5 – Aplicações

Nesta seção apresentamos alguns estudos e aplicações a séries financeiras para testar a modelagem proposta e sua aplicabilidade a situações reais.

5.1 – Exemplo 1: Serie Simulada

Simulamos uma série de tamanho 1000 segundo a seguinte estrutura LSTAR(2):

$$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})F(y_{t-2}) + \varepsilon_t \quad (9)$$

onde $F(y_{t-d}) = [1 + \exp\{-100(y_{t-2} - 0.02)\}]^{-1}$ e $\varepsilon_t \sim N(0, 0.02^2)$. O gráfico da série simulada é mostrado na Figura (1). Num primeiro análise, consideramos prioris subjetivas para os parâmetros do modelo. Fazendo 5000 iterações, escolhemos as estimativas condicionadas a $k = 2$ e $d = 2$, as cadeias simuladas para esses parâmetros foram de tamanho 4997. O parâmetro de retardo d como mais alta probabilidade a posteriori foi $d = 2$, cuja probabilidade a posteriori, $\Pr(d = 2 | y)$ foi estimada em 99.94%. A tabela (1) apresenta médias e desvios-padrões das distribuições marginais a posteriori. As médias a posteriori foram calculadas com as últimas 2498 iterações.

Figura 1

Num segundo análise usando prioris não informativas, definidas na Seção 3, testamos o algoritmo RJMCMC como um método alternativo para a seleção de

Par	Valores	Estimativa	Par.	Valores	Estimativa
θ_{01}	0	-0.0028 (0.0021)	θ_{02}	0.02	0.023 (0.0036)
θ_{11}	1.8	1.7932 (0.0525)	θ_{12}	-0.9	-0.8735 (0.0637)
θ_{21}	-106	-1.0809 (0.0654)	θ_{22}	0.795	0.7861 (0.0746)
γ	100	100.87 (4.9407)	c	0.02	0.0169 (0.0034)
σ^2	0.0001	0.00037 (0.000016)			

Tabela 1: Médias a posteriori para os parâmetros do modelo LSTAR (2) obtidos via MCMC, como em (9), com cadeias de tamanho 2498. As quantidades entre parênteses são os desvios-padrões.

modelos, a tabela (2) apresenta a probabilidade a posteriori da ordem do modelo. Segundo esses resultados o algoritmo teve uma ótima performance na escolha do ordem da auto-regressão k .

k	d			$p(k y)$
	1	2	3	
1	3	0	0	0.0003
2	1	9996	0	0.9997
3	0	0	0	0
4	0	0	0	0

Tabela 2: Resultados do algoritmo RJMCMC. A coluna $p(k|y)$ representa as medias a posteriori da ordem do modelo. Tamanho das cadeias 10000.

5.2 – Exemplo 2. Industrial Production Index (IPI) US

Analisamos a serie Índice de Produção Industrial (IPI) dos Estados Unidos¹, entre os anos 1960-2001. A série original é re-parametriza com a transformação logaritmo, além disso transformamos a serie em dados trimestrais usando a media. Por último pegamos a quarta diferenca dos dados a fim de ter uma série estacionaria e remover algum tipo de sazonalidade. A série resultante é mostrada na Figura (2)

Figura 2

Ajustamos os dados usando o algoritmo RJMCMC, considerando prioris não informativas. A figura (3) mostra a densidade marginal a posteriori $p(k|y)$, o máximo é estimado em $k = 5$, por tanto escolhemos um modelo LSTAR(5) para modelar os dados. O parâmetro d estimado é 3, com probabilidade a posteriori 0.994.

Figura 3

$$y_t = -0.07 + 0.909 y_{t-1} - 0.411 y_{t-2} + 0.51 y_{t-3} - 1.265 y_{t-4} + 0.541 y_{t-5} +$$

$$\left(\begin{matrix} (0.005) & (0.182) & (0.214) & (0.247) & (0.225) & (0.185) \\ 0.008 + 0.555 y_{t-1} - 0.153 y_{t-2} - 0.412 y_{t-3} + 1.124 y_{t-4} - 0.488 y_{t-5} \end{matrix} \right)$$

$$\times \left(1 + \exp \left\{ -960.58 \left(y_{t-3} + 0.0076 \right) \right\} \right) + \varepsilon_t, \quad \hat{\sigma}^2 = 0.00019$$

k	d				$p(k y)$
	1	2	3	4	
1	0	0	0	0	0
2	0	0	1	2	0.00006
3	5	0	0	1	0.00012
4	134	1	9328	31	0.18988
5	81	7	27679	77	0.55688
6	246	3	9528	141	0.19836
7	8	1	531	50	0.01180
8	1	0	70	5	0.00152
9	19	0	422	7	0.00896
10	25	0	1591	5	0.03240

Tabela 3: Resultados do algoritmo RJMCMC para a serie IPI-US. A coluna $p(k|y)$ representa as medias a posteriori da ordem do modelo. Tamanho das cadeias 50000.

Usamos o modelo estimado com $k = 5$ e $d = 3$ para fazer previsão no período 2002:1-2004:2. Da tabela (3) observamos que os modelos com $k = 4$ e $k = 6$ tem também alta probabilidade a posteriori, assim uma alternativa para fazer previsão é usar uma mistura dos

modelos com $k = \{4,5,6\}$ e $d = 3$. Neste caso os valores previstos serão ponderados pela probabilidade a posteriori de k ($P(k|y)$). Na literatura este método é conhecido como *Mistura Bayesiana de Modelos* (MBM). Para mais detalhes sobre este método ver por exemplo Raftery, Madigan e Hoeting (1997), Clyde (1999) e Hoeting, Madigan, Raftery e Volinsky (1999).

A figura 4 mostra os valores previstos usando o modelo com $k = 5$ e $d = 3$, e para a mistura de modelos com alta probabilidade a posteriori de k . De acordo com a figura (4) a mistura de modelos produz previsões mais acuradas para o curto prazo para a série de IPI-US. Uma possível explicação para essa melhora na previsão quando combinamos os vários modelos provem do fato que todos os modelos estudados são de uma forma ou de outra superparametrizados e que individualmente sofrem de "overfitting" que, em geral, produz baixa capacidade preditiva, apesar de ajudar razoavelmente bem os modelos dentro da amostra.

Figura 4

5.3 – Exemplo 3. IBOVESPA

A série IBOVESPA, índice da bolsa de valores do estado de São Paulo, contém 1678 observações registradas diariamente entre 02/01/1995 e 11/10/2001. Utilizamos esses dados para modelar a volatilidade estocástica usando a transformação logaritmo das primeiras diferenças, ou seja, os log-retornos. Neste caso, modelamos a log-volatilidade usando um modelo LSTAR com duas defasagens. A inferência se baseou em 1500 iterações obtidas a partir do algoritmo MCMC, após descartar as 1500 iterações iniciais. Usando prioris objetivas no processo inferencial obtivemos que a média a posteriori para $d = 2$ foi a maior ($P(d = 2|y) = 0.617$), a tabela (4) mostra os parâmetros estimados do modelo SV-LSTAR(2). As estimativas das volatilidades dos log-retornos são apresentados na Figura (5) juntamente com a série dos log-retornos. Finalmente, a Figura (6) mostra o comportamento da função de transição estimada das log-volatilidades ao longo do tempo.

Par.	Med. Pos.	Par.	Med. Pos.
θ_{01}	0.457 (2.001)	θ_{02}	-0.637(2.002)
θ_{11}	0.255 (0.185)	θ_{12}	1.020(0.217)
θ_{21}	0.778 (0.290)	θ_{22}	-1.075(0.299)
γ	89.191 (12.071)	c	-8428(0.034)
τ^2	0.045 (0.005)		

Tabela 4: IBOVESPA: Médias a posteriori dos parâmetros do modelo SV-LSTAR(2) dos log-retornos quando $d = 2$ ($P(d = 2|y) = 0.617$) As quantidades entre parênteses são os desvios-padrões.

Figura 5

¹ A série IPI-US pode ser obtida no site: <http://www.economagic.com/sub-info/>

Figura 6

6 – Conclusões

O método de estimação proposto para o modelo $LSTAR(k)$, em geral, apresenta resultados satisfatórios. O algoritmo RJMCMC mostra-se como uma alternativa eficiente para a seleção de modelos não lineares. A convergência dos algoritmos MCMC é rápida. Os estudos simulados mostram que o método de estimação funciona e pode ser automatizado para uso em situações práticas. O algoritmo RJMCMC proposto nesse artigo possibilita a estimação conjunta de todas as quantidades desconhecidas do modelo, inclusive sua própria dimensão. Sua implementação é relativamente simples e bons resultados de convergência foram observados. Entretanto, o algoritmo não deve ser visto somente como uma alternativa para a seleção de modelos, mas como uma ferramenta adicional a ser incluída na vasta caixa de ferramentas dos métodos de seleção e comparação de modelos já existentes.

Por outro lado a modelagem proposta para a volatilidade estocástica, oferece uma generalização para os modelos para o log-volatilidade atualmente utilizados. O processo de estimação pode ser, neste caso, um pouco mais complexo, dado que a cada iteração do MCMC, tem que se calcular, além dos parâmetros do modelo $LSTAR$, os valores da log-volatilidade (estados) para cada instante no tempo. Todos os programas foram feitos usando o pacote Ox^2

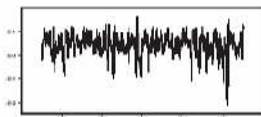


Figura 1: Série $LSTAR(k)$ simulada de tamanho 1000

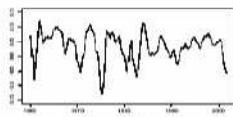


Figura 2: Série IPI-US

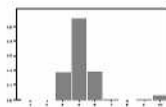


Figura 3: Probabilidade a posteriori estimada de $p(k|y)$ para a série IPI-US obtida do algoritmo RJMCMC

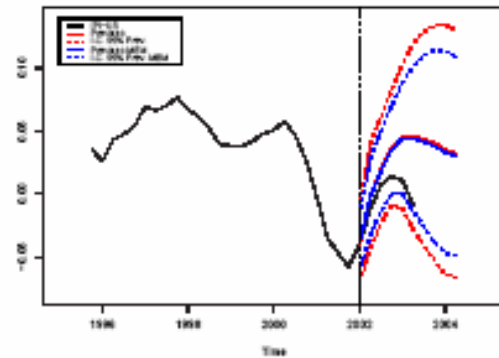


Figura 4: Previsão da série IPI-US, período 2002:1 - 2004:2. A linha vertical denota o tempo de início da previsão. A linha vermelha denota os valores previstos. A linha azul denota os valores previstos usando o método Mistura Bayesiana de Modelos. As linhas pontilhadas denotam os intervalos de confiança para cada tipo de previsão.

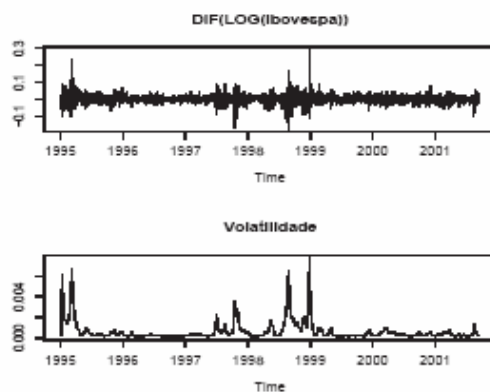


Figura 5: IBOVESPA: Acima, log-retornos entre 02/01/1995 e 11/10/2001 da série índice da bolsa de valores do estado do São Paulo. Embaixo, médias a posteriori da volatilidade ao longo do tempo estimados via MCMC para o modelo SV-LSTAR(2).

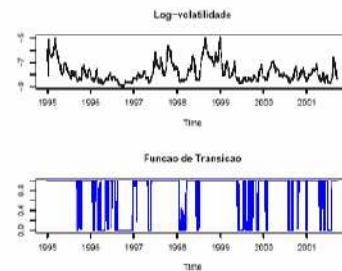


Figura 6: IBOVESPA: Acima, log-volatilidade ajustada ao longo do tempo. Embaixo, função de transição do modelo SV-LSTAR(2) para a log-volatilidade.

Referências

Besag, J. (1989). A candidate's formula – a curious result in Bayesian prediction. *Biometrika*, 76(1), 183.

Chan, K. S. e Tong, H. (1986). On estimating Thresholds in autoregressive models. *Journal of Time Series Analysis*, 7, 179–190.

Clyde, M. (1999). Bayesian model averaging and model search strategies (with discussion). Em J. Bernardo, A. Dawid, J. Berger e A. Smith (Eds.), *Bayesian Statistics 6* (pp. 157–185). Oxford University Press.

² O pacote pode ser obtido livremente no site: <http://www.nuff.ox.ac.uk/Users/Doornik/index.html>

- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. London: Chapman & Hall.
- Gelfand, A. E. e Smith, A. M. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.
- Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (Eds.). (1996). *Markov Chain Monte Carlo in Practice* (3rd ed.). London: Chapman & Hall.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation Bayesian model determination. *Biometrika*, 57(1), 97–109.
- Hastings, W. R. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 97–109.
- Hoeting, J. A., Madigan, D., Raftery, A. E. e Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14, 401–404.
- Huerta, G. e Lopes, H. (2000). Bayesian forecasting and inference in latent structure for the Brazilian industrial production index. *Brazilian Review of Econometrics*, 20, 1–26.
- Jarquier, E., Polson, N. e Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal Business of Econometrics Statistics*, 12(4), 371–417.
- Lubrano, M. (2000). Nonlinear econometric modeling in time series. Helsinki: Cambridge University Press. (Proceedings of the Eleventh International Symposium in Economic Theory)
- Luukkonen, R. (1990). On linearity testing and model estimation in nonlinear time series analysis. Helsinki: Finish Statistical Society.
- Luukkonen, R., Saikkonen, P. e Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491–499.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087–1091.
- Raftery, A. E., Madigan, D. e Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92, 179–191.
- Robert, C. e Casella, G. (1999). *Monte Carlo statistical methods*. Springer-Verlag.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425), 208–218.
- Troughton, P. T. e Godsill, S. J. (1997). *A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves* (Tech. Rep. No. CUED/F-INFENG/TR.304). Department of Engineering, University of Cambridge.
- West, M. e Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). New York: Springer Verlag.

Funções de Transferência em Modelos Dinâmicos Lineares Generalizados Bayesianos

Mariane Branco Alves
IME-UERJ / IM-UFRJ

Dani Gamerman
IM-UFRJ

Marco Antonio Rosa Ferreira
IM-UFRJ

Resumo

O presente trabalho tem por objetivo a modelagem da dinâmica do efeito de covariáveis X_t sobre uma variável resposta Y_t , com distribuição na família

exponencial. Supõe-se que as variáveis envolvidas no problema sejam observadas temporalmente.

A metodologia empregada é modelagem dinâmica bayesiana, com estimação baseada em médias e intervalos de credibilidade a posteriori. Como a variável resposta tem distribuição pertencente à família exponencial, somos levados à classe dos modelos dinâmicos lineares generalizados. Nesta classe de modelos, não se obtém forma analítica fechada para a distribuição a posteriori dos parâmetros de interesse. Para aproximação da informação a posteriori são utilizados métodos de Monte Carlo via cadeias de Markov.

Em particular, objetiva-se avaliar a significância e a forma do efeito defasado de regressoras sobre Y_t , propondo-se, para este fim, o uso de funções de transferência.

Os modelos propostos são aplicados a dados simulados para respostas Poisson, Binomial e Gama. Finalmente, como ilustração da metodologia adotada, um modelo

Poisson é aplicado a dados reais de poluição em São Paulo, de forma a avaliar o impacto ao longo do tempo do poluente Monóxido de Carbono sobre a mortalidade infantil por causas respiratórias naquela metrópole.

1 – Introdução

Objetiva-se, neste trabalho, modelar o impacto de uma variável regressora, X_t , sobre uma variável resposta, Y_t , cuja distribuição pertença à família exponencial. Supõe-se que todas as variáveis envolvidas no problema sejam observadas por T períodos de tempo. O foco primordial de atenção recai sobre a análise dos efeitos cumulativos da regressora (ou de um conjunto de regressoras) sobre Y_t , através do tempo. A motivação para tanto é a constatação de que, muitas vezes, o efeito de variáveis explicativas sobre a resposta não se faz sentir apenas imediatamente, levando-nos a avaliar o período e a forma de persistência das mesmas.

Na maioria das aplicações que consideram efeitos passados de covariáveis sobre a resposta, um número máximo de defasagens é fixado para a regressora, com base em alguma análise prévia. Se o efeito da variável persiste por um período razoável, tem-se, algumas vezes, modelos com muitos parâmetros para medir tal impacto. Fazemos uso de funções de transferência para este fim. A utilização dessas funções permite, de forma parcimoniosa e bastante flexível, modelar diversas formas de impacto de X_t sobre Y_t . Como poderá ser visto adiante, a adoção de funções de transferência geradas a partir de formas autoregressivas permite que, com poucos parâmetros, avalie-se impactos tantos instantes futuros quanto se deseje e, assim, o período de persistência da regressora deixa de ser prefixado, passando a ser estimado no próprio modelo proposto. As funções de transferência são inseridas como blocos estruturais nos modelos propostos, que pertencem à classe dos modelos dinâmicos lineares generalizados.

Este trabalho está organizado da seguinte forma: na seção 2, é feita uma breve descrição da classe de modelos lineares generalizados e ressaltada sua utilidade para a modelagem de dados cuja distribuição seja membro da família exponencial. Na seção 3, descrevemos sucintamente os modelos dinâmicos lineares, destacando-os como uma boa opção para a modelagem de séries temporais. A teoria de funções de transferência também é revisada nesta seção. Em seguida, na seção 4, os modelos dinâmicos lineares generalizados são apontados como uma extensão natural tanto para os modelos lineares generalizados (no sentido de tratar a correlação serial e permitir a evolução dos parâmetros) quanto para os modelos dinâmicos lineares (no sentido de ajustar variáveis com distribuição não Gaussiana). A abordagem inferencial aqui utilizada é bayesiana e envolve o cálculo de integrais que nem sempre podem ser obtidas analiticamente. Aproximações para estas

integrais são obtidas por meio de métodos de simulação de Monte Carlo via Cadeias de Markov (MCMC), discutidos na seção 5. Na seção 6, os modelos descritos são aplicados a exercícios com dados simulados para respostas com distribuição Poisson, Binomial e Gama. Finalmente, na seção 7, o modelo Poisson é ajustado a dados reais de mortalidade infantil por doenças respiratórias em São Paulo, no período de 1994 a 1997.

2 – Modelos Lineares Generalizados

Um modelo linear generalizado [19] é definido por observações cuja distribuição de probabilidade seja membro da família exponencial, por um conjunto de covariáveis independentes, às quais é aplicada uma estrutura linear e, ainda, por uma função de ligação, que estabelece a relação entre a média da variável resposta e as covariáveis. Se Y_t é a variável resposta e F_t é o vetor de covariáveis no instante t , tem-se então a seguinte estrutura para o modelo linear generalizado:

$$p(y | \eta_t) \propto \exp\left(\frac{y\eta_t - b(\eta_t)}{\phi_t}\right), t = 1, \dots, T \quad (1)$$

$$\mu_t = E[Y_t | \eta_t] = \dot{b}(\eta_t).$$

η_t é o parâmetro natural da distribuição, \dot{b} denota a primeira derivada de b e a relação entre a média μ_t e o vetor de regressoras ($n \times 1$) no instante t , F_t , é parametrizada pelo vetor desconhecido θ e dada por:

$$g(\mu_t) = F_t \theta$$

onde g é uma função não linear especificada.

Embora essa classe de modelos seja bastante ampla, existem limitações inerentes a esta, principalmente no que tange à dependência entre as observações. Supõe-se, nos modelos lineares generalizados, que toda a estrutura de correlação entre as observações Y_1, Y_2, \dots, Y_T esteja guardada no parâmetro (ou vetor paramétrico) θ ; assim, condicionalmente a θ , as observações são supostas independentes. Portanto, em situações nas quais haja correlação entre as observações, como ocorre em aplicações de séries temporais, o uso destes modelos pode comprometer a validade das inferências realizadas.

3 – Modelos Dinâmicos Lineares

Os Modelos Dinâmicos Lineares [23] constituem uma boa opção para se considerar formalmente a correlação temporal entre as observações. Tais modelos são expressos de forma hierárquica. No primeiro nível da hierarquia, tem-se a equação observacional, que relaciona a variável resposta ao vetor F_t de regressoras no instante t , por meio de uma forma linear que pode incorporar

componentes estruturais, tais como tendência, sazonalidade e as funções de transferência adotadas neste trabalho:

$$Y_t = F_t \theta + v_t, v_t \sim N(0, \Sigma_t) \quad (2)$$

Na estrutura acima, tem-se o vetor de parâmetros θ_t evoluindo ao longo do tempo, ao contrário do que ocorre nos modelos lineares generalizados, em que estes são mantidos fixos. Em um segundo estágio na hierarquia do modelo, tem-se a equação de evolução, que descreve o comportamento temporal e estocástico dos coeficientes de regressão:

$$\theta_t = G_t \theta_{t-1} + u_t, u_t \sim N(0, W_t) \quad (3)$$

onde G_t é uma matriz de transição ($n \times n$) conhecida e W_t é uma matriz de covariâncias ($n \times n$).

No contexto bayesiano, o modelo completa-se com uma densidade a priori $\theta \sim N(a_1, R_1)$ e com a hipótese de que os erros u_t e v_t sejam serial e mutuamente independentes. Pode-se, facilmente, acrescentar outros níveis à hierarquia do modelo e atribuir densidades a priori a hiperparâmetros.

Tradicionalmente, a inferência sobre os modelos dinâmicos lineares é feita seqüencialmente. Primeiramente, a informação inicial sobre os parâmetros de interesse é resumida em sua densidade a priori. Uma hipótese usual é a de que o modelo seja fechado a informações externas, ou seja, excetuando-se a informação contida na densidade a priori, toda a informação relevante é trazida somente pela verossimilhança. Desta forma, a informação disponível no instante t é $D_t = \{D_{t-1}, y_t\}$. Combinando-se, por meio do teorema de Bayes, a densidade a priori à verossimilhança obtida através do modelo postulado para os dados, obtém-se, a cada instante, a densidade atualizada para os parâmetros de interesse, chamada densidade a posteriori. Esta, por sua vez, é usada para a construção da densidade a priori no instante seguinte. Esse processo de atualização é resumido a seguir. Inicialmente, obtém-se a densidade a priori

$$\pi(\theta_t | D_{t-1}) = \int \pi(\theta_t | \theta_{t-1}) \pi(\theta_{t-1} | D_{t-1}) d\theta_{t-1} \quad (4)$$

onde $\pi(\theta_t | \theta_{t-1})$ é obtida aplicando-se a equação de evolução e $\pi(\theta_{t-1} | D_{t-1})$ é a densidade a posteriori do instante anterior. O processo de atualização dos parâmetros completa-se ao se obter a densidade a posteriori a partir do teorema de Bayes:

$$\pi(\theta_t | D_t) \propto p(y_t | \theta_t) \pi(\theta_t | D_{t-1}) \quad (5)$$

O objetivo primordial em diversas análises e, em particular, em análises envolvendo séries temporais, é a previsão de valores futuros do processo modelado. Uma vez que se determine a densidade a priori para o instante t , a distribuição preditiva é obtida facilmente, da seguinte forma:

$$p(y_t | D_{t-1}) = \int p(y_t | \theta_t) \pi(\theta_t | D_{t-1}) d\theta_t \quad (6)$$

Se Σ_t , W_t ou elementos de F_t e G_t forem desconhecidos, ou mesmo se a distribuição de $Y_{\{t\}}$ não for Gaussiana, as equações (4), (5) e (6) não têm solução analítica, devendo-se recorrer a métodos computacionais de aproximação para obtenção destas.

3.1 - Funções de Transferência

O vetor de parâmetros de estado, θ_t contém, para cada instante t , todas as quantidades relevantes para a explicação da função de resposta média por meio de uma regressão. Pode-se ter, inseridos em θ_t , por exemplo, blocos de tendência, sazonalidade e covariáveis.

Neste trabalho deseja-se avaliar, em particular, de que forma o efeito dos níveis passados de covariáveis é refletido sobre o processo de interesse em instantes futuros. Alguns exemplos são os efeitos de: uma campanha de vacinação sobre o número de casos de uma doença, elevações nos níveis de determinados poluentes sobre o número de óbitos ou internações hospitalares, uma campanha publicitária sobre o volume de vendas de certo produto, níveis de chuva sobre o fluxo de um rio e etc. Para tanto, um dos blocos estruturais a inserir nos modelos dinâmicos lineares generalizados é o de efeitos passados e presentes de covariáveis. Tais efeitos são modelados, aqui, por meio de funções de transferência [2]. A abordagem aqui adotada é a escolha de formas paramétricas que reflitam a dinâmica do impacto das covariáveis parcimoniosamente. Estas funções podem ter as mais variadas formas, indo desde um efeito de impacto instantâneo, passando por formas de decaimento exponencial, até formas de efeitos que crescem até atingir valor máximo, voltando a decrescer posteriormente. Passamos então a descrevê-las.

Suponha-se que a relação entre Y_t e X_t possa ser representada da seguinte forma:

$$\begin{aligned} (1 - \delta_1 B - \dots - \delta_r B^r) Y_t &= \\ &= (\omega_0 - \omega_1 B - \dots - \omega_s B^s) X_{t-b} \\ &= (\omega_0 B^b - \omega_1 B^{b+1} - \dots - \omega_s B^{b+s}) X_t \end{aligned}$$

ou

$$\delta(B) Y_t = \omega(B) B^b X_t = \Omega(B) X_t, \quad (7)$$

onde B é o operador defasagem, ou seja, $BY_t = Y_{t-1}$.

Alternativamente, pode-se representar a relação entre Y_t e X_t através de um filtro linear:

$$\begin{aligned} Y_t &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots \\ &= \frac{\Omega(B)}{\delta(B)} \\ &= v(B) X_t \end{aligned} \quad (8)$$

O polinômio $v(B) = v_0 + v_1B + v_2B^2 + \dots$ é a função de transferência entre Y_t e X_t . Os pesos v_0, v_1, v_2, \dots ue aparecem na função de transferência são denominados função de resposta ao impulso. O filtro (8) é dito estável se, para $B \leq 1$, $v(B)$ é convergente.

Consideremos, por exemplo, a função de transferência obtida da estrutura AR(1), que passaremos a denominar função de primeira ordem:

$$Y_t = \rho Y_{t-1} + BX_t \quad (9)$$

Resolvendo a equação a diferenças acima por substituições recursivas, tem-se:

$$Y_t = \beta X_t + \rho\beta X_{t-1} + \rho^2\beta X_{t-2} + \dots \quad (10)$$

Suponhamos $\beta > 0$. O formato das funções de resposta ao impulso e de transferência dependerá dos valores do parâmetro autoregressivo. Então teremos os seguintes comportamentos para a função de resposta ao impulso:

- $0 < \rho < 1$: decaimento geométrico
- $-1 < \rho < 0$: decaimento geométrico com alternância de sinal
- $\rho > 1$: crescimento geométrico
- $\rho < -1$: crescimento geométrico com alternância de sinal

Claramente, a forma gerada para tais funções depende do comportamento das soluções da equação a diferenças que as define. Detalhes sobre a solução de equações a diferenças podem ser obtidos em Hamilton [15]. Ao se substituir a equação (8) em (7), obtém-se a seguinte identidade

$$\begin{aligned} (1 - \delta_1 B - \dots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \dots) &= \\ = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) B^b & \quad (11) \end{aligned}$$

De acordo com Box e Jenkins [2], equacionando-se (11) em B tem-se:

$$v_j = \begin{cases} 0 & j < b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_0 & j = b \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_{j-b} & j = b+1, \dots, b+s \\ \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} & j > b+s \end{cases}$$

Assim, de forma geral, o comportamento da função de resposta ao impulso gerada por polinômios no operador defasagem de ordens r e s aplicados a Y_t e $B^b X_t$, respectivamente, consiste em:

- B valores nulos v_0, v_1, v_{b-1}

- $s - r + 1$ valores sem padrão determinado (se $s < r$ tais valores não ocorrem)
- Valores v_j , com $j \geq b + s - r + 1$ seguindo o padrão ditado pela equação a diferenças de ordem r .

Raízes distintas e reais do polinômio autoregressivo de ordem r geram comportamentos de decaimento exponencial, raízes distintas e complexas fornecem termos senoidais e raízes iguais fornecem termos polinomiais. Portanto, a função de resposta ao impulso será uma mistura de termos exponenciais, polinomiais e senoidais, dependendo do estudo das raízes do polinômio autoregressivo. Estas diferentes formas preconizam modelos diferentes para os efeitos de impacto sobre a série estudada, que podem ser comparados. A figura (1) ilustra os comportamentos de funções de transferência geradas de estruturas com $r = 2, s = 2$

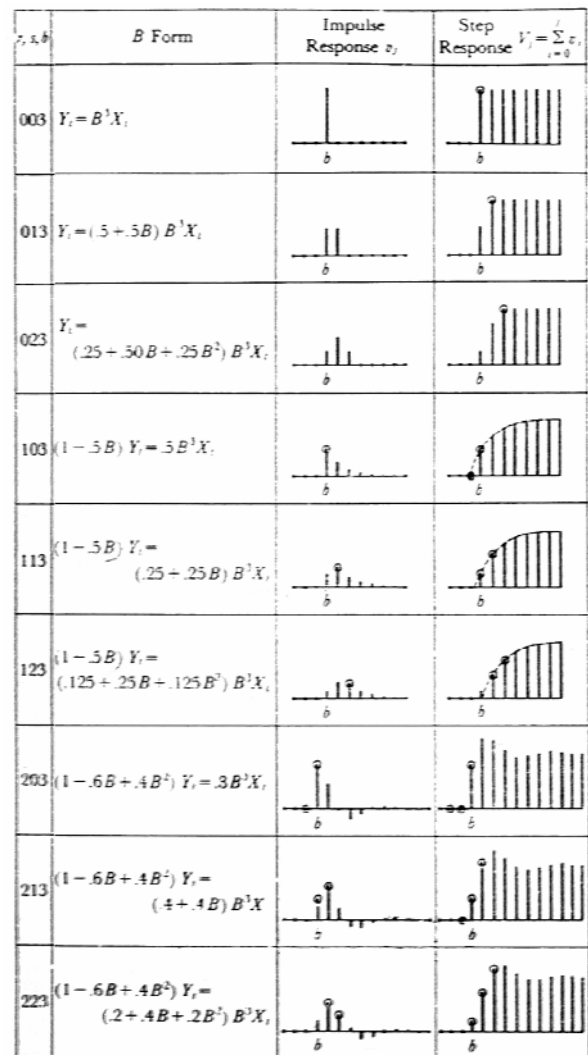


Figura 1: Exemplos de funções de resposta ao impulso e de transferência (Fonte: Box e Jenkins, 1994)

Box e Jenkins [2] propõem um procedimento de identificação de funções de transferência baseado na função de correlação cruzada entre a regressora X_t e a resposta Y_t para o caso em que há apenas uma regressora. Edlund [9] faz uma revisão de métodos de identificação das funções de transferência e estende o procedimento de identificação para o caso em que há mais que uma regressora envolvida no problema. Neste trabalho, não será feita identificação prévia de funções de transferência. Ao invés disso, pretende-se, em aplicações futuras, ajustar diversas funções de transferência, propondo variadas formas de impacto de X_t sobre Y_t e, com base em algum procedimento de seleção entre modelos, como, por exemplo, os propostos por Spiegelhalter et al [22] e por Gelfand e Ghosh [12], escolher aquela que melhor reflete a relação entre os dados analisados, tendo como critério não só a qualidade do ajuste, mas também a parcimônia do modelo escolhido.

4 – Modelos Dinâmicos Lineares Generalizados

Embora os Modelos Dinâmicos Lineares sejam aplicados com sucesso a diversas situações em que se deseja tratar a dependência temporal entre observações, estes ainda apresentam algumas limitações. Talvez, sua maior restrição resida na hipótese de que o processo a ser modelado seja Gaussiano, como pode ser visto na equação (2).

Em algumas situações, faz-se necessária, então, uma classe de modelos que estenda os modelos lineares generalizados no sentido de permitir a evolução temporal dos parâmetros de estado e que, concomitantemente, estenda os modelos dinâmicos lineares no sentido de permitir a modelagem de processos não Gaussianos. Tal extensão, formalizada por West, Harrison e Migon [24] e denominada modelos dinâmicos lineares generalizados é obtida combinando-se a estrutura observacional dos modelos lineares generalizados à estrutura evolutiva dos modelos dinâmicos lineares: a equação de observação (2) é substituída pela equação (1), que compreende observações em toda a família exponencial e a relação entre a média do processo (transformada via função de ligação) e o efeito das covariáveis passa a ser descrita por uma regressão dinâmica:

$$g(\mu_t) = F_t' \theta_t$$

onde $\mu_t = E[y_t | \theta_t] = \dot{b}(\eta_t)$. O modelo completa-se com a equação (3) que descreve a evolução dos parâmetros, junto a uma priori para θ_1 .

Na classe dos modelos dinâmicos lineares generalizados, como mencionado na seção 3, encontram-se dificuldades para a obtenção das equações (4), (5) e (6) de forma exata, devido à verossimilhança não ser, necessariamente, Gaussiana. West, Harrison e Migon [24] aproveitam a idéia descrita em Migon [18] e utilizam Linear Bayes, trabalhando apenas com o

primeiro e o segundo momentos da distribuição de $\theta_{t-1} | D_{t-1}$ para aproximação destas integrais. Para contornar o problema de estimação da variância de evolução, utilizam fatores de desconto. Singh e Roberts [21] fazem uma aproximação através da equação observacional modificada

$$\tilde{y}_t = F_t' \theta_t + \tilde{v}_t, \tilde{v}_t \sim N(0, \tilde{V}_t) \quad (12)$$

onde

$$\tilde{y}_t = \tilde{y}_t(\theta_t) = y(\mu_t) + (y_t - \mu_t) \dot{g}(\mu_t) \quad (13)$$

e

$$\tilde{V}_t(\theta_t) = \ddot{b}(\eta_t) [\dot{g}(\mu_t)]^2. \quad (14)$$

Os valores das observações e variâncias modificadas \tilde{y}_t e \tilde{V}_t são definidos seqüencialmente e dependem da distribuição original de Y_t , através da função de ligação do modelo adotado, constituindo-se, em última instância, em uma aproximação linear para as observações y_t em torno da média μ_t . Ferreira e Gamerman [7] fazem uma revisão bastante detalhada de diversos métodos para aproximação da densidade a posteriori nos modelos dinâmicos lineares generalizados e propõem o uso de métodos de Monte Carlo via Cadeias de Markov para este fim.

Além dos problemas envolvidos na estimação de modelos dinâmicos com distribuição na família exponencial, o uso de funções de transferência acarreta não-linearidade. Pole [20] considera modelos de função de transferência de primeira ordem, com resposta Gaussiana. Nesse caso, a não-linearidade no parâmetro autoregressivo da função de transferência torna (4), (5) e (6) intratáveis analiticamente. A solução adotada por Pole, então, é uma variação dinâmica do método de quadratura Gaussiana, trabalhando-se com grades atualizadas seqüencialmente para o parâmetro autoregressivo e usando o fato de que, condicionalmente a este parâmetro, o modelo é linear. A abordagem adotada no presente trabalho é a obtenção de amostras da posteriori utilizando métodos de simulação estocástica via Cadeias de Markov. Por meio do uso destes métodos, pode-se facilmente contornar o problema de não normalidade da resposta e de não-linearidade nos parâmetros autoregressivos das funções de transferência. Alguns aspectos sobre a aplicação desses métodos são discutidos a seguir.

5 – Monte Carlo via Cadeias de Markov

Um dos focos principais do processo de inferência Bayesiana é a determinação da densidade a posteriori dos parâmetros de interesse. Nos modelos dinâmicos lineares generalizados, esta densidade é dada por:

$$\pi(\theta_1, \dots, \theta_T, W) \propto \left\{ \prod_{t=1}^T p(y_t | \theta_t) \right\} \cdot \left\{ \prod_{t=1}^T \pi(\theta_t | \theta_{t-1}) \right\}$$

$$\cdot \pi(\theta_t) \pi(W), \quad (15)$$

onde assumimos $W_t = W$, constante, e F_t e G_t , conhecidos, $\forall t$.

Aqui serão utilizados os métodos de Monte Carlo via Cadeias de Markov (MCMC) [10] para aproximação da distribuição acima.

Os métodos MCMC consistem na construção de uma cadeia de Markov que, sob determinadas condições, tenha como distribuição estacionária a densidade da qual se deseja gerar valores; neste caso, em particular, a densidade a posteriori (15). Assim, uma vez que se verifique que a convergência da cadeia gerada tenha sido atingida, pode-se utilizar os valores desta (obtidos após convergência) como amostra da distribuição estacionária. Toda a inferência sobre essa densidade é feita, então, de forma empírica. Por exemplo, para obtenção da estimativa do valor esperado a posteriori, toma-se a média da amostra gerada ou para obtenção de intervalos de credibilidade, tomam-se os percentis adequados na amostra gerada.

Os algoritmos MCMC mais usuais em aplicações de inferência bayesiana são o amostrador de Gibbs e o algoritmo Metropolis-Hastings [10]. O amostrador de Gibbs é útil quando não se consegue gerar valores da densidade conjunta dos parâmetros de interesse, mas as densidades condicionais completas¹ destes parâmetros são conhecidas e a geração de valores das mesmas é possível. O algoritmo de Gibbs procede da seguinte forma:

- Propõem-se valores iniciais para todos os parâmetros de interesse.
- Geram-se valores das densidades condicionais completas de cada um dos parâmetros, sempre condicionando-se ao último valor gerado de cada um dos demais parâmetros.
- Repete-se o passo 2 até que a convergência da cadeia tenha sido atingida.

Em geral, nos modelos dinâmicos lineares generalizados, não se consegue amostrar das densidades condicionais completas de todos os parâmetros de interesse. Nestes casos, pode-se inserir passos Metropolis-Hastings no algoritmo de Gibbs. O algoritmo Metropolis-Hastings é útil para a geração de valores daqueles parâmetros cujas densidades condicionais completas não tenham forma conhecida. Este algoritmo procede da seguinte forma para geração de um valor θ_{ij} da densidade condicional completa

$\pi(\theta_{ij} | \theta_{t-j})$ de um parâmetro qualquer θ_j no instante t :

1. Gera-se um valor θ_t^* e um núcleo de transição

$$q(\theta_t, \theta_t^*)$$

¹ Se a densidade de interesse é $\pi(\theta_1, \dots, \theta_k)$, então a densidade

condicional completa para θ_j é $\pi(\theta_j | \theta_{-j}) =$

$\pi(\theta_j | \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k)$

2. Aceita-se o valor gerado com probabilidade

$$\min \left\{ 1, \frac{\pi_t(\theta_t^*) / q_t(\theta_t, \theta_t^*)}{\pi_t(\theta_t) / q_t(\theta_t^*, \theta_t)} \right\} \quad (16)$$

onde π_t é a densidade condicional completa de θ_{ij} para $t = 1, \dots, T$.

Inserir-se este procedimento no passo 2 do amostrador de Gibbs, para geração dos parâmetros que não apresentem densidades condicionais completas conhecidas.

5.1 – Estratégia de Amostragem e Escolha de Densidades Propostas

Ao se utilizar os algoritmos MCMC para a estimação de modelos de espaço de estados, surgem algumas questões. A primeira delas é a forma de atualização dos parâmetros estruturais. A estratégia mais natural para atualização dos parâmetros estruturais é a geração individual de valores destes, como, por exemplo, em Carlin e Stoffer [3]. Entretanto, a correlação entre valores vizinhos dos parâmetros faz com que a convergência neste tipo de estratégia torne-se lenta. Existem diversas propostas para se contornar este problema. Carter e Kohn [4] e Frühwirth-Schnatter [8] propõem a geração simultânea de todos os parâmetros de estado, utilizando o amostrador de Gibbs e obtêm resultados eficientes. Knorr-Held [17], propõe uma solução intermediária entre a geração individual e a geração simultânea de todo o vetor de parâmetros estruturais, particionando-o em blocos cujos tamanhos podem variar (determinística ou estocasticamente) de 2 a T. A metodologia empregada faz uso de gerações dos parâmetros estruturais via algoritmo Metropolis-Hastings e Knorr-Held [17] relata bons resultados ao permitir variação estocástica dos tamanho dos blocos a cada iteração do algoritmo.

A abordagem aqui adotada é aquela proposta por Gamerman [11] Para contornar o problema de correlação inerente aos parâmetros de estado e otimizar a convergência da cadeia, Gamerman faz uso da relação bem definida entre estes parâmetros e os erros de evolução, descrita pela equação (3). A amostragem é feita sobre estes últimos (μ_t), que, a priori, não são correlacionados, usando movimentos individuais. O custo computacional envolvido reside justamente na necessidade de, a cada iteração, calcular os valores dos parâmetros de estado a partir do acúmulo dos valores gerados para os erros de evolução.

Como mencionado anteriormente, em geral não se obtém forma fechada para as densidades condicionais completas de todos os parâmetros envolvidos no problema, o que leva à adoção de passos Metropolis-Hastings dentro do amostrador de Gibbs e, portanto, à escolha de densidades propostas para geração. Em Knorr-Held [17], por exemplo, cada bloco de parâmetros é gerado de uma densidade a priori autoregressiva Gaussiana, com momentos conhecidos. Segundo Knorr-Held, a escolha dessa

densidade proposta é bastante vantajosa, visto que esta reflete a estrutura de autocorrelação existente entre os parâmetros; além disso, como as gerações são feitas de densidades a priori, as razões de aceitação reduzem-se a razões de verossimilhança, o que facilita os cálculos. Na abordagem adotada por Gamerman [11], as densidades propostas utilizadas nos passos Metropolis-Hastings são as densidades condicionais completas obtidas do modelo de trabalho proposto por Singh e Roberts [21] e descrito em (12), (13) e (14). No presente trabalho, alternou-se entre o uso desta última proposta, à qual passaremos a nos referir como "proposta baseada na verossimilhança", e de propostas definidas por um passeio aleatório centrado no valor corrente do parâmetro de interesse (ou de alguma transformação do parâmetro que o leve à reta real). Ao se utilizar a primeira abordagem, pretende-se aproximar a proposta da verossimilhança. Em contrapartida, como a proposta passeio aleatório é simétrica, a adoção desta faz com que as razões de aceitação do algoritmo Metropolis-Hastings simplifiquem-se, tornando o algoritmo mais rápido.

6 – Exercícios com Dados Simulados

Os modelos trabalhados até o presente momento seguem a seguinte estrutura geral:

$$\begin{aligned} Y_t &\sim FE \\ g(\mu_t) &= \alpha_t + \varepsilon_t + \delta' Z_t \\ \alpha_t &= \alpha_{t-1} + u_t, u_t \sim N(0, w) \\ \varepsilon_t &= \rho \varepsilon_{t-1} + \beta' X_t \end{aligned} \quad (17)$$

onde FE denota distribuição na família exponencial, $\mu_t = E[Y_t | \eta_t]$, X_t é um vetor k -dimensional de regressoras cujo efeito cumulativo sobre Y_t se deseja estudar, Z_t é um vetor m -dimensional de regressoras para as quais considera-se apenas o efeito instantâneo sobre Y_t e ρ é um parâmetro autoregressivo assumindo valores em $(0,1)$, de forma a garantir a estabilidade do filtro. O modelo completa-se com priors:

$$\begin{aligned} \alpha &\sim N(m_\alpha, C_\alpha), \beta \sim NM(m_\beta, C_\beta), \delta \sim NM(m_\delta, C_\delta) \\ \rho &\sim U(0,1), \varepsilon_0 \sim N(m_\varepsilon, C_\varepsilon), W \sim GI\left(\frac{n}{2}, \frac{ns}{2}\right). \end{aligned}$$

Essa estrutura tem como casos particulares:

- a função de transferência de primeira ordem, bastando que se faça β e X_t escalares;
- a função de transferência gerada pela estrutura com polinômios de ordem $r=1, s=k-1$, ao se fazer $X_t = (X_t, X_{t-1}, \dots, X_{t-k+1})$
- uma função de transferência conjunta para k regressoras, com um único parâmetro de decaimento ρ para o efeito destas, fazendo-se $X_t = (X_t, X_{t-1}, \dots, X_{t-k+1})$

Nosso objetivo é obter uma aproximação computacional, via MCMC, para a densidade a posteriori conjunta do vetor paramétrico envolvido no modelo acima. Seguindo a proposta de Gamerman [11] substituiu-se a geração dos parâmetros estruturais α_t pela geração de u_t que, a priori, não são correlacionados. Espera-se, com tal procedimento, acelerar a convergência das cadeias geradas. Deve-se, então, aproximar a densidade a posteriori do vetor paramétrico $\Theta = (u_1, \dots, \mu_T, \beta', \delta', \rho, \varepsilon_0, W)$:

$$\begin{aligned} \pi(\Theta | D_r) &\propto \prod_{t=1}^T p(y_t | \Theta) \pi(\Theta) \\ &\propto \prod_{t=1}^T p(y_t | \Theta) \prod_{t=2}^T \frac{1}{\sqrt{W}} \exp\left\{-\frac{1}{2W} u_t^2\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2R_1} (u_t^2 - 2u_t a_1)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2} (\beta - m_\beta)' C_\beta^{-1} (\beta - m_\beta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2} (\delta - m_\delta)' C_\delta^{-1} (\delta - m_\delta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2C_\varepsilon} (\varepsilon_0^2 - 2\varepsilon_0 m_\varepsilon)\right\} \\ &\quad \cdot W^{-\left(\frac{n}{2}+1\right)} \exp\left\{-\frac{ns}{2W}\right\}. \end{aligned} \quad (18)$$

Diretamente de (18) tem-se que a condicional completa para W é:

$$(W | D_r, \Theta_{-W}) \sim IG\left(\frac{T+n-1}{2}, \frac{\sum_{t=2}^T u_t^2 + ns}{2}\right).$$

Os demais parâmetros não possuem condicionais completas com formas conhecidas, sendo necessário, portanto, gerá-los seguindo o algoritmo Metropolis-Hasting.

Inicialmente, a estrutura acima foi ajustada para um conjunto de dados simulados, com distribuição de Poisson e utilizando apenas uma regressora X_t . Ainda não tivemos sucesso na aplicação deste modelo. Pareceu-nos não haver informação suficiente para estimação dos parâmetros estruturais α_t , pois obtivemos intervalos de credibilidade muito amplos para estes. O ajuste deste modelo demanda bastante tempo computacional, devido ao cálculo de α_t em função de u_1, u_2, \dots, u_t a cada iteração. Decidimos, então, simplificar o problema, tornando o nível α constante no tempo, de forma a verificar o desempenho da geração dos demais parâmetros para, posteriormente, retornar ao modelo (19). Observe-se que, com isso, não nos afastamos do objetivo primordial que é o estudo das formas de transferência de X sobre Y . Posteriormente, pretendemos contemplar a evolução temporal dos parâmetros estruturais.

Passamos, então, a trabalhar com o modelo:

$$\begin{aligned} Y &\sim FE(\eta_t) \\ g(\mu_t) &= \alpha + \varepsilon_t + \delta' \mathbf{Z}_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + \beta' \mathbf{X}_t \end{aligned} \quad (19)$$

com priors $\alpha \sim N(m_\alpha C_\alpha)$,
 $\beta \sim NM(\mathbf{m}_\beta \mathbf{C}_\beta)$, $\delta \sim NM(\mathbf{m}_\delta \mathbf{C}_\delta)$, $\rho \sim U(0,1)$,
 $\varepsilon_0 \sim N(m_\varepsilon, C_\varepsilon)$.

Tem-se então o vetor paramétrico $\Theta = (\alpha, \beta', \delta', \rho, \varepsilon_0)$, cuja densidade a posteriori é dada por:

$$\begin{aligned} \pi(\Theta | D_T) &\propto \prod_{t=1}^T p(y_t | \Theta) \pi(\Theta) \\ &\propto \prod_{t=1}^T p(y_t | \Theta) \exp\left\{-\frac{1}{2C_\alpha}(\alpha^2 - 2\alpha m_\alpha)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\beta - \mathbf{m}_\beta)' \mathbf{C}_\beta^{-1}(\beta - \mathbf{m}_\beta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\delta - \mathbf{m}_\delta)' \mathbf{C}_\delta^{-1}(\delta - \mathbf{m}_\delta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2C_\varepsilon}(\varepsilon_0^2 - 2\varepsilon_0 m_\varepsilon)\right\}. \end{aligned} \quad (20)$$

Todo o vetor paramétrico é gerado pelo algoritmo Metropolis-Hastings no amostrador de Gibbs. Para avaliar o desempenho do modelo (19), foram feitas aplicações com dados simulados. Toda a implementação do procedimento de Monte Carlo via Cadeias de Markov foi feita em linguagem FORTRAN 90. Passamos a descrever tais aplicações.

6.1 - Resposta Poisson

Para dados de contagem Poisson, o modelo (19) assume a seguinte forma:

$$\begin{aligned} Y_t &\sim Poisson(\lambda_t) \\ \log(\lambda_t) &= \alpha + \varepsilon_t + \delta' \mathbf{Z}_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + \beta' \mathbf{X}_t \end{aligned} \quad (21)$$

com priors $\alpha \sim N(m_\alpha C_\alpha)$,
 $\beta \sim NM(\mathbf{m}_\beta \mathbf{C}_\beta)$, $\delta \sim NM(\mathbf{m}_\delta \mathbf{C}_\delta)$, $\rho \sim U(0,1)$,
 $\varepsilon_0 \sim N(m_\varepsilon, C_\varepsilon)$.

A densidade a posteriori para $\Theta = (\alpha, \beta', \delta', \rho, \varepsilon_0)$ dada por:

$$\begin{aligned} \pi(\Theta | D_T) &\propto \prod_{t=1}^T p(y_t | \Theta) \pi(\Theta) \\ &\propto \prod_{t=1}^T e^{-\lambda_t} \lambda_t^{y_t} \exp\left\{-\frac{1}{2C_\alpha}(\alpha^2 - 2\alpha m_\alpha)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\beta - \mathbf{m}_\beta)' \mathbf{C}_\beta^{-1}(\beta - \mathbf{m}_\beta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\delta - \mathbf{m}_\delta)' \mathbf{C}_\delta^{-1}(\delta - \mathbf{m}_\delta)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2C_\varepsilon}(\varepsilon_0^2 - 2\varepsilon_0 m_\varepsilon)\right\}. \end{aligned} \quad (22)$$

Foram ajustados, para conjuntos de dados simulados, os seguintes casos particulares deste modelo:

1. Função de transferência de primeira ordem

$$\begin{aligned} Y_t &\sim Poisson(\lambda_t) \\ \log(\lambda_t) &= \alpha + \varepsilon_t + \delta' \mathbf{Z}_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + \beta' \mathbf{X}_t \end{aligned} \quad (23)$$

No modelo acima, X é a regressora cujo impacto sobre Y através do tempo se deseja estudar. O modelo de transferência é de primeira ordem e, como $0 < \rho < 1$, ostula o decaimento do efeito de X sobre Y , sendo a "persistência" desse efeito medida pelo parâmetro ρ . Z é uma regressora (ou um conjunto de regressoras) com impacto imediato sobre Y .

Simulamos um conjunto de dados composto por 4 anos de observações diárias, usando os seguintes valores para o vetor paramétrico: $\alpha = 1.0$, $\beta = 0.03$, $\delta_1 = -0.10$, $\delta_2 = 0.13$, $\rho = 0.9$, $\varepsilon_0 = 0.5$.

O esquema MCMC para amostragem foi o seguinte:

- α, β e ε_0 : Gerados de propostas passeio aleatório centradas nos valores correntes de cada um destes parâmetros, com variâncias calibradas de forma a gerar taxas de aceitação em torno de 50%.
- δ : Vetor gerado da proposta baseada na verossimilhança, definida pela condicional completa obtida do modelo de trabalho (12)
- ρ : Aplicou-se a transformação $\psi = \text{logit}(\rho)$, gerando-se ψ de um passeio aleatório centrado no valor corrente da cadeia e com variância calibrada de acordo com a taxa de aceitação.

A figura 2 exhibe os histogramas das 20.000 últimas iterações da cadeia de Markov referente à amostra da posteriori de cada parâmetro, quando se supôs que estas já tivessem convergido. As retas

traçadas nos gráficos indicam os verdadeiros valores dos parâmetros. Na figura 2 observe-se que, embora o verdadeiro valor de α oculte-se na cauda do histograma da amostra final, este está de acordo com a densidade a posteriori de α , que é a curva tracejada exibida no gráfico. A figura 3 exibe a comparação entre os reais valores de λ_t e os estimados e algumas estatísticas sumárias das amostras da posteriori encontram-se na tabela 1. A partir das amostras da posteriori, é possível calcular estimativas dos valores esperados a posteriori para a função de resposta ao impulso, junto a intervalos de credibilidade. A figura 4 exibe o aumento percentual em λ , provocado por aumentos no valor de X em relação à sua média histórica observada, para lags de 0 a 30.

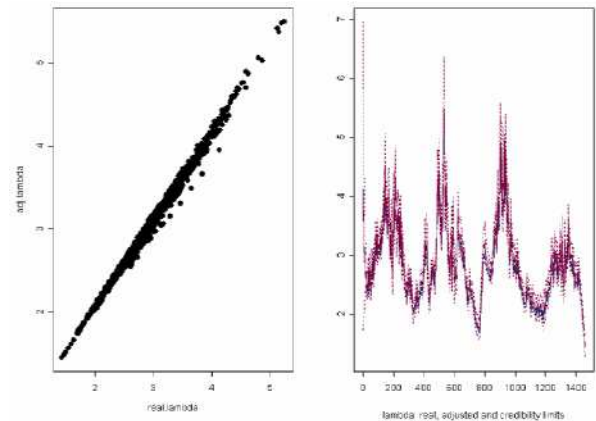


Figura 3: Comparação entre λ_t 's estimados e os reais, segundo o modelo (23).

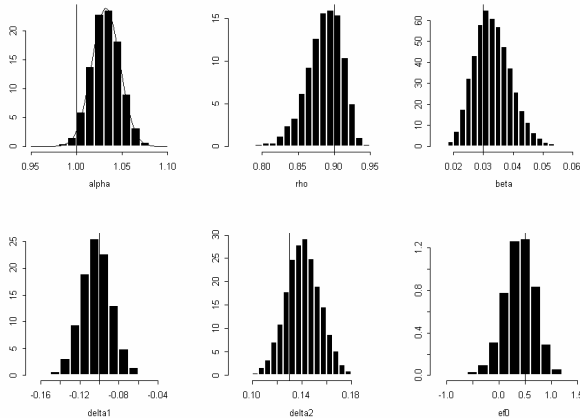


Figura 2: Resposta Poisson: Histogramas das iterações finais das cadeias geradas usando os dados simulados segundo o modelo (23)

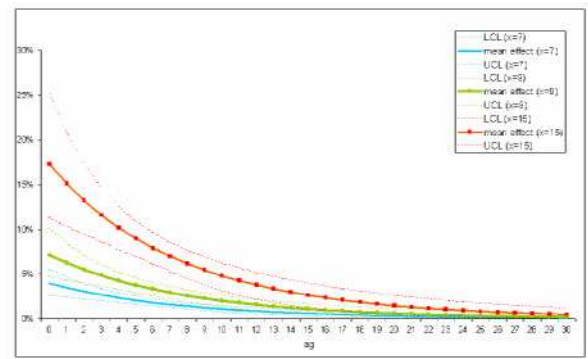


Figura 4: Aumentos percentuais em λ provocados por incrementos em X em relação à sua média histórica, segundo o modelo (23).

Tabela 1: Resposta Poisson: Estatísticas sumárias da posteriori obtidas da estimação do modelo (23)

	α	ρ	β	δ_1	δ_2	ϵ_0
Mínimo	0.9516	0.7668	0.01487	-0.16620	0.08656	-1.1380
1° Quartil	1.0220	0.8713	0.02877	-0.11320	0.13170	0.2000
Mediana	1.0320	0.8890	0.03275	-0.10300	0.14090	0.3972
Média	1.0320	0.8861	0.03320	-0.10290	0.14110	0.3862
3° Quartil	1.0430	0.9044	0.03720	-0.09267	0.15040	0.5909
Máximo	1.0970	0.9494	0.05864	-0.03581	0.19360	1.3620

2. Função de transferência de ordem $r = 1, s = 1$

$$\begin{aligned}
 Y_t &\sim \text{Poisson}(\lambda_t) \\
 \log(\lambda_t) &= \alpha + \varepsilon_t + \delta'Z_t \\
 \varepsilon_t &= \rho\varepsilon_{t-1} + \beta_0 X_t + \beta_1 X_{t-1},
 \end{aligned}
 \tag{24}$$

Simulamos um conjunto de dados a partir desta estrutura, usando: $\alpha = 1.0$, $\beta = 0.03$, $\delta_1 = -0.10$, $\delta_2 = 0.13$, $\rho = 0.9$, $\varepsilon_0 = 0.5$.

A função de resposta ao impulso gerada por essa estrutura é exibida na figura 5. A figura 6 exibe os histogramas das 20.000 últimas iterações de cada cadeia de Markov gerada como aproximação para a densidade a posteriori dos parâmetros envolvidos.

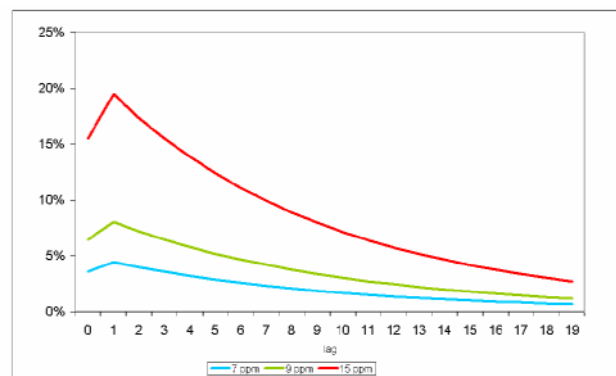


Figura 5: Resposta Poisson: Aumentos percentuais em λ provocados por incrementos em X em relação à sua média histórica, segundo o modelo (24)

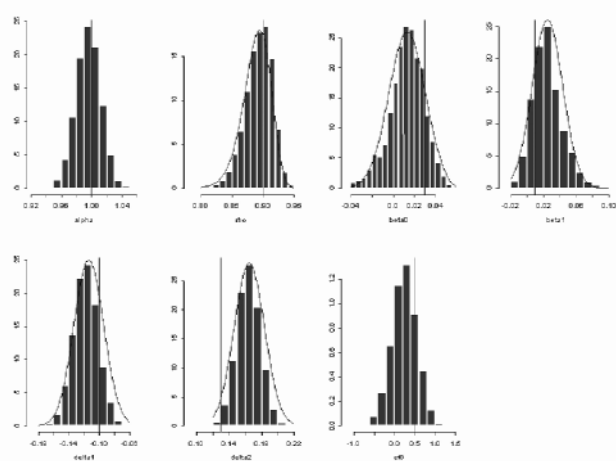


Figura 6: Resposta Poisson: Histogramas das iterações finais das cadeias geradas usando os dados simulados segundo o modelo (24)

Tabela 2: Resposta Poisson: Estatísticas sumárias da posteriori obtidas da estimação do modelo (24)

	α	ρ	β_0	β_1	δ_1	δ_2	ε_0
Mínimo	0.9469	0.7571	0.005458	-0.066900	-0.16450	0.03549	-1.449000
1º Quartil	0.9930	0.8773	0.041910	-0.022130	-0.11620	0.09276	-0.005901
Mediana	1.0030	0.8940	0.052580	-0.008745	-0.10630	0.10210	0.201400
Média	1.0030	0.8903	0.053720	-0.009420	-0.10620	0.10210	0.191700
3º Quartil	1.0140	0.9081	0.064600	0.002985	-0.09581	0.11150	0.411800
Máximo	1.0590	0.9533	0.103000	0.041910	-0.04164	0.15670	1.314000

6.2 – Resposta Binomial

Para dados de contagem Binomial, o modelo (19) assume a seguinte forma:

$$Y_t \sim Bin(n, p_t)$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha + \varepsilon_t + \delta'Z_t \quad (25)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + \beta'X_t$$

com prioris $\alpha \sim N(m_\alpha, C_\alpha)$, $\beta \sim NM(m_\beta, C_\beta)$,

$\delta \sim NM(m_\delta, C_\delta)$, $\rho \sim U(0,1)$, $\varepsilon_0 \sim N(m_\varepsilon, C_\varepsilon)$.

A densidade a posteriori para $\Theta = (\alpha, \beta', \delta', \rho, \varepsilon_0)$ dada por:

$$\begin{aligned} \pi(\Theta | D_T) &\propto \prod_{t=1}^T p(y_t | \Theta) \pi(\Theta) \\ &\propto \prod_{t=1}^T p_t^{y_t} (1-p_t)^{n-y_t} \\ &\bullet \exp\left\{-\frac{1}{2C_\alpha}(\alpha^2 - 2\alpha m_\alpha)\right\} \\ &\bullet \exp\left\{-\frac{1}{2}(\beta - m_\beta)' C_\beta^{-1} (\beta - m_\beta)\right\} \\ &\bullet \exp\left\{-\frac{1}{2}(\delta - m_\delta)' C_\delta^{-1} (\delta - m_\delta)\right\} \\ &\bullet \exp\left\{-\frac{1}{2C_\varepsilon}(\varepsilon_0^2 - 2\varepsilon_0 m_\varepsilon)\right\}. \end{aligned} \quad (26)$$

6.3 – Resposta Gama

Para dados com distribuição Gama, o modelo (19) assume a seguinte forma:

$$Y_t \sim Gama(\varphi, \lambda_t)$$

$$\log\left(\frac{\varphi}{\lambda_t}\right) = \alpha + \varepsilon_t + \delta'Z_t \quad (27)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + \beta'X_t$$

com priors $\varphi \sim G(a, l)$ $\alpha \sim N(m_\alpha C_\alpha)$,
 $\beta \sim NM(m_\beta C_\beta)$, $\delta \sim NM(m_\delta C_\delta)$, $\rho \sim U(0, 1)$,
 $\varepsilon_0 \sim N(m_\varepsilon, C_\varepsilon)$.

A densidade a posteriori para $\Theta = (\varphi, \alpha, \beta', \delta', \rho, \varepsilon_0)$ dada por:

$$\begin{aligned} \pi(\Theta | D_T) &\propto \prod_{t=1}^T p(y_t | \Theta) \pi(\Theta) \\ &\propto \prod_{t=1}^T \frac{(\lambda_t y_t)^\varphi}{\Gamma(\varphi)} \exp\left\{-\sum_{t=1}^T \lambda_t y_t\right\} \\ &\cdot \exp\{-l\varphi\} \varphi^{a-1} \exp\left\{-\frac{1}{2C_\alpha}(\alpha^2 - 2\alpha m_\alpha)\right\} \\ &\cdot \exp\left\{-\frac{1}{2}(\beta - m_\beta)' C_\beta^{-1} (\beta - m_\beta)\right\} \\ &\cdot \exp\left\{-\frac{1}{2}(\delta - m_\delta)' C_\delta^{-1} (\delta - m_\delta)\right\} \\ &\cdot \exp\left\{-\frac{1}{2C_\varepsilon}(\varepsilon_0^2 - 2\varepsilon_0 m_\varepsilon)\right\}. \end{aligned} \quad (28)$$

Foram realizados exercícios de simulação para as respostas Binomial e Gama, obtendo-se bons resultados, no sentido de se conseguir aproximar bastante bem as distribuições a posteriori dos parâmetros de interesse, por meio da geração de amostras das mesmas via Monte Carlo por Cadeias de Markov. Passamos então a aplicar os modelos propostos a conjuntos de dados reais. Uma destas aplicações, envolvendo resposta Poisson, é apresentada a seguir.

7 – Aplicação a Dados Reais

Em diversas grandes metrópoles ao redor de todo o mundo, especialistas vêm tentando identificar quais são os poluentes mais danosos à saúde, quantificando seu impacto sobre diferentes patologias. Detalhes sobre os efeitos da poluição sobre a saúde podem ser vistos em Wilson e Spengler [25]. Uma questão de interesse em tais estudos é o tempo de persistência dos efeitos de cada poluente, bem como a forma de evolução, ou seja, se o efeito de elevações de determinado poluente, por exemplo, decresce ao longo do tempo ou cresce, atingindo impacto máximo dias após a ocorrência de elevação, para então decair. Na maioria dos estudos desse tipo, um número máximo de defasagens é postulado para a série temporal de níveis do poluente a estudar. Com isso, pode ser difícil identificar processos de longa persistência ou, ao tentar fazê-lo por meio de muitas defasagens, o modelo final pode se tornar sobreparametrizado.

Passamos, então, a descrever os dados de contagens de óbitos por doenças respiratórias de crianças e níveis de poluentes, coletados em São Paulo, no período de 1994 a 1997, aos quais serão aplicados os modelos propostos na seção anterior.

Há algum tempo tem-se percebido que o número de internações por doenças respiratórias em São Paulo

parece estar correlacionado às condições atmosféricas e níveis de poluição na região metropolitana. Essa hipótese tem servido como motivação para o desenvolvimento de diversos estudos que buscam quantificar o efeito de diferentes poluentes sobre o número de internações e/ou óbitos.

A figura 7 exhibe a série de contagens diárias de óbitos por doenças respiratórias para crianças com idades inferiores a 5 anos, no período de 1994 a 1997. Pode-se observar que o número médio de contagens é bastante baixo, o que tornaria a adoção da distribuição normal (e, portanto, de Modelos Dinâmicos Lineares) inadequada.

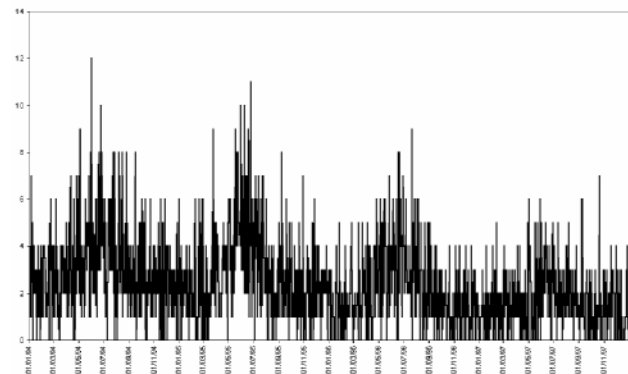


Figura 7: Contagem de óbitos de menores de 5 anos em SP - Jan/1994 a Dez/1997

A seguir, na figura 8, encontra-se a série do poluente Monóxido de Carbono, no mesmo período. Havia, na série original, alguns períodos com dados omissos. Chegamos a gerar as observações omissas via MCMC, mas como o número de "missings" é razoável, isso elevou muito o tempo de execução do algoritmo, o que dificultava a checagem do programa FORTRAN implementado para o algoritmo MCMC. Passamos, então, a preencher previamente as observações omissas utilizando o aplicativo desenvolvido por Junger [16], com metodologia baseada no algoritmo EM. Observe-se que, com isso, subestimamos a incerteza associada ao problema. Pretende-se, posteriormente, gerar todos os dados omissos utilizando o algoritmo MCMC.

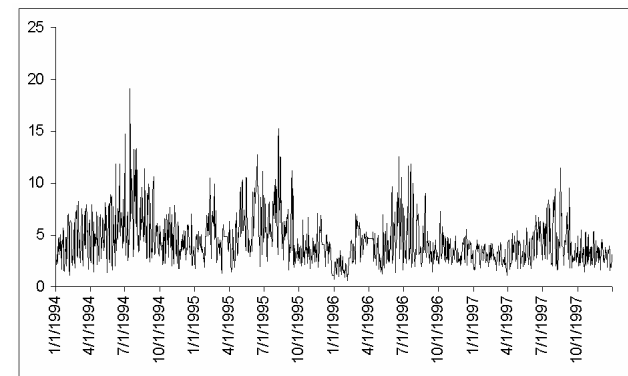


Figura 8: Níveis diários de CO (ppm) em SP - Jan/1994 a Dez/1997.

Ao analisar estes mesmos dados, por meio de modelos aditivos generalizados, Conceição et al. [6] constataram a

existência de associação significativa entre mortalidade infantil por doenças respiratórias e níveis diários de CO, SO₂, PM₁₀. Ao incluir os poluentes simultaneamente no modelo, observaram relação significativa apenas para o poluente CO e significância marginal do poluente SO₂. Botter et al. [1] modelaram as contagens totais de óbitos de maiores de 65 anos, também em São Paulo, de 1991 a 1993, utilizando modelos de espaço de estados em que os poluentes são inseridos através de um processo latente, enquanto variáveis meteorológicas são colocadas na equação de observação. Ferreira e Gamerman [7] avaliaram a associação entre poluentes e mortalidade infantil em São Paulo durante o ano de 1991, utilizando, como variáveis explicativas, os níveis de NO₂ e CO, relacionados às contagens de óbitos por meio de Modelos Dinâmicos Lineares Generalizados. O modelo ali adotado não considera os efeitos de médio e longo prazo das covariáveis, medindo apenas o impacto imediato dos poluentes sobre o número de óbitos. Outra aplicação recente de Modelos Dinâmicos Lineares Generalizados a dados de poluição foi feita por Chiogna e Gaetan [5], que estudaram a relação entre óbitos por causas não acidentais e níveis poluentes na cidade de Birmingham, Alabama. Mais uma vez, no modelo adotado, permite-se a evolução temporal dos coeficientes, mas não se considera o efeito cumulativo de covariáveis.

Analisou-se, aqui, a relação entre níveis de Monóxido de carbono e óbitos por doenças respiratórias de crianças em São Paulo. Para tanto, foi adotado o modelo Poisson com função de transferência gerada a partir de um polinômio autoregressivo de ordem $r = 1$ e o número de defasagens aplicadas à variável poluente foi $s = 2$. Além disso, variáveis climáticas foram inseridas no modelo, mas apenas o efeito imediato destas foi avaliado:

$$\begin{aligned}
 \text{Obitos}_t &\sim \text{Poisson}(\lambda_t) \\
 \log(\lambda_t) &= \alpha + \varepsilon_t + \delta_1 \text{Temp}_t + \delta_2 \text{Um}_t \\
 \varepsilon_t &= \rho \varepsilon_{t-1} + \beta_0 \text{CO}_t + \beta_1 \text{CO}_{t-1} + \beta_2 \text{CO}_{t-2}
 \end{aligned}$$

A figura 9 exibe histogramas das amostras da densidade a posteriori. Estatísticas sumárias de tais amostras podem ser vistas na tabela 3.

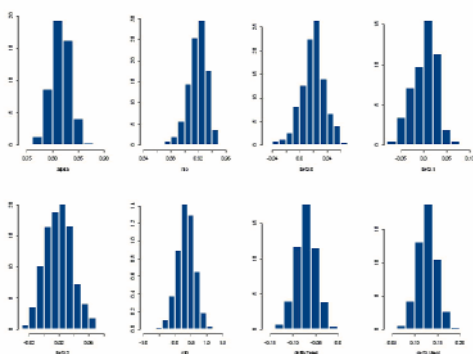


Figura 9: Sumário da amostra da distribuição a posteriori de $\varphi, \alpha, \beta_0, \beta_1, \beta_2, \delta_1, \delta_2, \rho, \varepsilon_0$

Tabela 3: Estatísticas sumárias da amostra da densidade a posteriori

	1º Quartil	Média	3º Quartil
α	0.8031	0.8156	0.8281
ρ	0.9107	0.9188	0.9287
β_0	0.0096	0.0198	0.0301
β_1	-0.0169	0.0023	0.0215
β_2	0.0066	0.0201	0.0323
ε_0	0.1753	0.3504	0.5346
δ_1	-0.0851	-0.0704	-0.0558
δ_2	0.1135	0.1274	0.1411

Como se pode perceber pelas estatísticas associadas ao parâmetro autoregressivo, ρ a memória do processo de contagens de óbitos em relação ao poluente CO é bastante alta. Com base na amostra da distribuição a posteriori de β e ρ , calculou-se a função de resposta ao impulso esperada e intervalos de credibilidade a 90%, exibidos na figura A figura 10. O nível médio de CO observado no período de análise foi 4,52 ppm. Nesta figura, registra-se o efeito (em termos de aumento percentual na contagem esperada de óbitos) de elevações dos níveis deste poluente em relação à média observada. Um aumento para 19 ppm, por exemplo, como o registrado em 14 de julho de 1994, de acordo com o modeo ajustado, não tem impacto imediato significativo sobre o processo de contagens de óbitos, mas representa um aumento em torno de 30% no número esperado de óbitos de crianças por doenças respiratórias após dois dias. Além disso, como se pode ver na figura 10, este efeito é bastante duradouro.

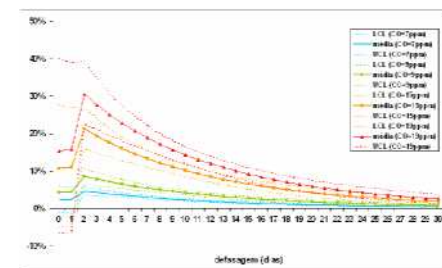


Figura 10: Função de resposta ao impulso obtida com base na amostra da distribuição a posteriori dos parâmetros β e ρ : Risco Relativo de aumentos no nível de CO em relação ao nível base 4,52 ppm.

Agradecimentos

Os autores agradecem ao Projeto PASSO - Poluição Atmosférica e Saúde Social (USP), pela disponibilização do banco de dados de poluentes e a Washington Leite Junger (IMS-UERJ) e Antonio Carlos Ponce de Leon (IMS-UERJ) por terem disponibilizado o aplicativo para preenchimento de dados omissos.

Referências

[1] Botter, D. A., Jørgensen, B. e Peres, A. A. Q. (2000) - *A longitudinal study on mortality and air pollution for São Paulo, Brazil*, Research

- Report **28** 2000, Department of Statistics and Demography, Odense University - SDU.
- [2] Box, G. E. P., Jenkins, G. M. e Reinsel, G. C. (1994) - *Time Series Analysis - Forecasting and Control*, Prentice-Hall, Inc., 3a. ed.
- [3] Carlin, B. P. P. e Stoffer, D. S. (1992) - A Monte Carlo approach to nonnormal and nonlinear state-space-modeling, *J. Am. Statist. Assoc.*, **88**, pp.9-25.
- [4] Carter, C. K. e Kohn, R. (1994) - On Gibbs sampling for state space models, *Biometrika*, **81**, pp.541-553.
- [5] Chiogna, M. e Gaetan, C. (2002) - Dynamic generalized linear models with application to environmental epidemiology, *Journal of the Royal Statistic Society*, série C, **51**, pp.453-68.
- [6] Conceição, G. M. S., Miraglia, S. G. E. K., Kishi, H. S., Saldiva, P. H. N. e Singer, J. M. (2001) - Air pollution and children mortality - A time series study in São Paulo, Brazil, *Environmental Health Perspectives*, **109**, suplemento 3.
- [7] Ferreira, M. A. e Gamerman, D. (2000) - Dynamic Generalized Linear Models, em DEY, D. K., GHOSH, S. K. e MALLICK, B. K., *Generalized Linear Models*, Marcel Dekker Inc., pp 57-72.
- [8] Frühwirth-Schnatter, S. (1992) - Data augmentation and dynamic linear models, *J. Time Series Anal.*, **15**, pp. 183-202.
- [9] Edlund, P.-O. (1987) - *On identification of transfer function models*, Research Report, Departamento de Estatística, Universidade de Estocolmo.
- [10] Gamerman, D. (1997) - *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*, Chapman & Hall.
- [11] Gamerman, D. (1998) - Markov chain Monte Carlo for dynamic generalized linear models, *Biometrika*, **85**, pp. 215-227.
- [12] Gelfand, A. E. e Ghosh, S. K. (1998) - Model choice: a minimum posterior predictive loss approach *Biometrika*, **85**, pp. 1-11.
- [13] Gelman, A. e Rubin, D. R. (1992) - A single series from the Gibbs sampler provides a false sense of security, em *Bayesian Statistics 4* (eds J. M Bernardo et al.), Oxford University Press, pp. 625-631.
- [14] Geweke, J. (1992) - Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion), *Scandinavian Journal of Statistics*, **20**, pp. 147-156.
- [15] Hamilton, J. D. (1994) - *Time Series Analysis*, Princeton University Press.
- [16] Junger, W. L. (2002)- *Imputação de Dados Faltando em Séries Temporais Multivariadas via Algoritmo EM*, monografia de final de graduação, Departamento de Estatística, Univeridade do Estado do Rio de Janeiro.
- [17] Knorr-Held, L. (1999) - Conditional prior proposals in dynamic models, *Scandinavian Journal of Statistics*, **26**, pp. 129-144.
- [18] Migon, H. S. (1984)- *An Approach to Non-linear Bayesian Forecasting Problems with Applications*, tese de doutorado, Departamento de Estatística, Univeridade de Warwick..
- [19] McCullagh, P. e Nelder, J. A. (1983) - *Generalized Linear Models*, Chapman & Hall.
- [20] Pole, A. (1988)- Transfer response models: A numerical approach. Em J M Bernardo, M H DeGroot, D V Lindley, and A F M Smith, editores, *Bayesian Statistics 3*, Oxford University Press.
- [21] Singh, A. C. e Roberts, G. R. (1992) - State space modelling of cross-classified time series of counts , *International Statistics Review*, **60**, pp. 321-36.
- [22] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. e van der Linde, A. (2002) - Bayesian measures of model complexity and fit (with discussion and rejoinder), *J. Roy. Statist. Soc.*, Ser. B, **64**, 583-639.
- [23] West, M. e Harrison, J. (1989) - *Bayesian Forecasting and Dynamic Models*, Springer-Verlag.
- [24] West, M., Harrison, J. e Migon, H. S. (1985) - Dynamic generalized linear models and Bayesian forecasting, *J. Am. Statist. Assoc.*, **80**, pp. 73-96.
- [25] Wilson, R. e Spengler, J. (1996) - *Particles in Our Air - Concentrations and Health Effects*, Harvard University Press. ■

Bayesian Estimation Via Mcmc For Probit-Normal Model In Item Response Theory

Jorge L. Bazán, Heleno Bolfarine
Department of Statistics
University of São Paulo

Roseli Aparecida Leandro
Department of Exact Sciences
University of São Paulo - ESALQ

Summary

Item response theory (IRT) models a set of dichotomous multivariate responses corresponding to I items of a test applied to n subjects. IRT is widely used in several evaluation systems considering frequentist methodology. In this paper we consider the normal ogiveprobit-normal model and present two Bayesian estimation procedures considering MCMC methodology for simulation from the posterior distribution of the latent variables. We illustrate the interpretation of model parameters considering an application to 14 items of a mathematical test for sixth grade students using the WinBUGS package. A sensitivity analysis for the prior distribution is considered and also chain convergence is analyzed.

Keywords: probit-normal model, item response theory, Bayesian estimation, sensitivity analysis.

1 – Introduction

Item response theory (IRT) considers models for describing how the probability of answering correctly an item of a test depend on subjects' ability or latent proficiency and also parameters related to the questions being answered. Several references in recent statistical literature can be associated to such models. Specially we mention Baker (1992), van der Linden and Hambleton (1997) and in Portuguese, the excellent summary by Andrade et al. (2000). Several characterizations for such models have been developed in the last 40 years, specially with respect to the status of the latent variables (Borsboom et al. 2003) -latent parameters in other contexts (for example Baker, 1992)-, and with respect to classical or Bayesian approaches. In this paper, we consider the characterization due to Holland and Rosenbaum (1986) and Bartholomew & Knoop (1999) and the Bayesian approach in Albert (1992).

Bayesian estimation is not new in IRT models and recently can be distinguished in estimation with and without Markov chain Monte Carlo (MCMC). MCMC simulation methods (Chen et al. 2000, Patz & Junker, 1999) is a method of simulating random samples from any theoretical multivariate distribution -in particular, from the multivariate posterior distribution that is the focus of Bayesian inference- so that features of the theoretical distribution can be estimated by corresponding

features of the random sample. The most widely used IRT models are the dichotomous item response models. Two models, logistic-normal model and probit-normal model, described in the following section, are part of these models. Educational evaluation studies conducted in Brazil and other countries typically is based on the logit normal model. This may be due to the fact that this model is implemented in commercial packages and use of the estimation procedures is more complex in probit-normal model and this may explain the very limited use of this model within the classical approach. On the other hand, the use of strict Bayesian approaches in logit-normal model is somewhat more complex (see Rupp et al., 2004). In Brazil, Assunção (1999) implemented the approach by Patz & Junker for MCMC estimation of logistic-normal model. With the use of MCMC techniques, the probit-normal model became more popular and several extensions were proposed in the literature including: Three parameter models (Sahu, 2002, Glas & Meijer, 2003), Multidimensional models (Béguin & Glas, 2001; Jackman, 2001; Linardakis & Dellaportas, 2002), Multilevel models (Fox and Glas, 2001), Testlet (Wang, et al., 2003), Item confirmatory factor analysis model (Segall, 2003) and Measurement error model (Fox & Glas, 2003).

An MCMC implementation for probit-normal model using directly the likelihood function is implemented in Spiegelhalter et al. (1996) via *adaptive rejection sampling (ARS) scheme* (Gilks and Wild, 1992) or in Spiegelhalter et al. (2003) via the *slice sampling* (Neal, 2003). Another MCMC implementation, introduced in Albert (1992), is based on the *data augmented Gibbs sampling (DAGS) scheme* using auxiliary latent variables. Although Bayesian estimation for the probit-normal model has been developed in the literature, very few papers discuss in detail the problem of prior specification for the item parameters. Some results in this direction are found in Albert & Ghosh (2000) and Ghosh et al. (2000) where it is discussed theoretically the consequences of using precise prior distributions for the item parameters. In this paper, we present a review of the two main procedures for Bayes estimation via MCMC of the probit-normal model described above. A sensitivity study of the model with respect to prior specification for the item parameters is conducted. The paper is organized as follows. In section 2, dichotomous item response models is presented and a literature revision of their characteristics and estimation approach is conducted. In Section 3 is presented a Bayesian estimation via MCMC for probit-normal model using the adaptative rejection sampling (ARS) scheme and the data augmented Gibbs sampling (DAGS) scheme. Prior specification and an adequate formulation for the model is presented. Section 4 deals with an application to a real data set coming from a mathematical test and sensitivity analysis for different priors in item parameters is conducted.

2 – Dichotomous item response Model: Characteristics and Estimation

We consider that

$$Y_{ij} | u_i, \eta_j \sim \text{Bernoulli}(p_{ij}), \quad (1)$$

where Y_{ij} are the dichotomous response corresponding to subject i on item j , $\eta_j = (a_j, b_j)'$ are the item parameters, where a_j corresponds to item discrimination and b_j corresponds to item difficulty and u_i is the value corresponding to latent variable U_i associated to subjects i , describing its ability in answering the test of I items and p_{ij} the probability of correct answer for subject i in test j , $i = 1, \dots, n, j = 1, \dots, I$. Further, let

$$p_{ij} = P(Y_{ij} = 1 | u_i, \eta_j) = F_{ij}(m_{ij}) \quad (2)$$

with $m_{ij} = a_j u_i - b_j$, a linear function of u_i , $i = 1, \dots, n, j = 1, \dots, I$. The link function $F_{ij}(\cdot)$ is typically known as the *item response function* or *curve characteristic of the item* and satisfies the property of latent monotonicity (strictly nondecreasing function of U_i) The function F_{ij} typically is the same for all i and j and most used are cumulative distributions of the normal and the logistic distributions. Further, negative values of a_j are not expected and b_j and U_i take real values.

IRT models considered typically satisfies the conditional independence property, that is, for subject i , the responses Y_{ij} corresponding to items $j = 1, \dots, I$ are conditionally independent given the values of latent variables U_i , $i = 1, \dots, n$. Further, it is considered independence between responses from different subjects. Under the above assumptions, the joint distributions of

$\mathbf{Y} = (Y_1', \dots, Y_n)'$ with $Y_i = (Y_{i1}, \dots, Y_{iI})'$ given the vector of latent variables $\mathbf{U} = (U_1, \dots, U_n)'$ and the vector of item parameters $\boldsymbol{\eta} = (\eta_1, \dots, \eta_I)'$ can be written as

$$p(\mathbf{Y} = \mathbf{y} | \mathbf{U} = \mathbf{u}, \boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j=1}^I F_{ij}(m_{ij})^{y_{ij}} (1 - F_{ij}(m_{ij}))^{1-y_{ij}} \quad (2)$$

The first IRT model was formally introduced by Lord (1952) and consider $F_{ij}(\cdot) = \Phi(\cdot)$, $i = 1, \dots, n$ and $j = 1, \dots, I$ with $\Phi(\cdot)$ the cumulative function of the standard normal distribution. The model is known as the normal ogive model. Birnbaum (1968) considered $F_{ij}(\cdot) = L_{ij}(\cdot)$ with $L_{ij}(m_{ij}) = e^{m_{ij}} / (1 + e^{m_{ij}})$ denoting the cumulative function of the logistic distribution. This model is known as the logistic model with two parameters. As a special case of this model $a_j = 1$ we have the Rasch model (Fisher and Molenaar, 1995).

2.2 – Characteristics

The dichotomous item response model presented in (1), (2) and (3) involves a total of $n + 2I + 2$ unknown parameters being thus overparametrized. On the other hand, for a fixed number of items, item parameters are known as structural parameters and the latent variables are known as incidental parameters, because they increase with n , the sample size and because the analysis is generally focused on the item parameters. The model is also unidentifiable, since it is preserved under a special class of transformations of the parameters (see Albert, 1992) so that maximum likelihood estimates may not be unique. One way of contouring such difficulties is to impose restrictions on the item parameters as considered, for example, in Bock and Aitkin (1981). Another way follows by specifying a distribution for the latent variables. Lord and Novik (1968, Chap. 16) consider

$$U_i \stackrel{iid}{\sim} N(\mu, \sigma^2), i = 1, \dots, n \quad (4)$$

This assumption establishes that it is believed that the latent variables are well behaved and that the abilities are a random sample from this distribution. (Albert, 1992). As in Tsutakawa (1984), we consider in this paper that $\mu = 0$ e $\sigma^2 = 1$.

The IRT model by considering (1)-(4) is denominated for Albert (1992) and Lord & Novick (1968) as normal ogive model. In Bartholomew & Knoot (1999), the model with normal link and normal latent variable is denoted probit-probit and the model with logistic link and normal latent variable as logit-probit model. We find it more appropriate using the notation probitnormal and logit-normal models, respectively.

2.3 – Estimation

The estimation problem in IRT models can be grouped into three categories: frequentist (or classical) estimation, Bayesian without MCMC and Bayesian with MCMC.

Classical estimation has been dominated by the likelihood approach. Using the former methodology, several approaches have been proposed such as joint likelihood, marginal likelihood and conditional likelihood (Andrade et al., 2000, Baker, 1992). However, the maximum likelihood estimators for the structural parameters are not consistently estimated in the presence of incidental parameters (Neyman & Scott, 1948). The most used approach is the marginal likelihood approach using the pseudo-EM-algorithm with Gaussian quadrature for approximating the integrals needed for implementing the E step of the algorithm involved in the estimation of the item parameters (Book & Aitkin, 1981). This estimation procedure is implemented in adhoc software, for example program BILOG (Mislevy & Bock, 1990), that uses "heuristic" restriction of the model (Du Toit, 2003). The estimation of ability parameters is performed in a second stage with item parameters replaced by estimates computed previously. Limitations of this methodology are discussed in Patz & Junker (1999) and Sahu (2002). To just name a few, we point out the impossibility of estimating U_i for extreme scores,

limitations on the statement of regularity conditions that justify the asymptotic theory under which inference for model parameters and model checking are based (Froelich, 2000). In practical situations this estimation procedure defines the evaluation paradigm as is the case with the BILOG software which incorporate their restrictions of computation as criterion in evaluation area. Specifically, the fact that estimation is done in two step implies that evaluation is also done in two steps: *the algorithm defines the evaluation scheme and not the evaluation scheme defines the algorithm*. Although separate estimation of part of the parameters considering known the remaining parameters is possible, inference based on asymptotic arguments, specially for no *i.i.d* observations, may results on non consistent estimation (Patz & Junker, 1999).

In the case of non MCMC Bayesian estimation, Bayesian marginal estimation is used with maximum and expected a posteriori estimates for latent variables considering hierarchical models or not (see Baker, 1992; Kim et al. 1994). One characteristic of classical and non MCMC Bayesian approaches is that they are based typically on the logistic link function. A review of classical and non MCMC Bayesian approaches can be found in Baker (1992) and Andrade et al. (2000). A review of Bayesian approaches using MCMC can be found in Patz & Junker (1999). It approach for probit-normal model is describe in the following section.

3 – Bayesian inference via MCMC for the probitnormal model

Let $D_{obs} = Y$, the observed data. Hence, the likelihood function for the probit-normal model is given by

$$L(u, \eta | D_{obs}) = \prod_{i=1}^n \prod_{j=1}^I \Phi_{ij}(m_{ij})^{y_{ij}} (1 - \Phi_{ij}(m_{ij}))^{1-y_{ij}},$$

where $\Phi(\bullet)$ is the cumulative function of the standard normal distribution.

3.1 – Prior specification

We consider the following general class of prior distribution:

$$\pi(u, \eta) = \prod_{i=1}^n g_{1i}(u_i) \prod_{j=1}^I g_{2j}(\eta_j)$$

Where $g_{1i}(u_i) = \phi(u_i; 0, 1)$, $i = 1, \dots, n$, are a standard normal distribution priors for the latent variables u_i , and $g_{2j}(\eta_j) = g_{21j}(a_j)g_{22j}(b_j)$, $j = 1, \dots, I$, are priors for item parameter, in which $g_{21j}(\bullet)$ should be proper to guarantee a proper posterior distributions for (u, η) (see Albert & Ghosh, 2000, Ghosh et al. 2000).

Following proposals usually considered (see Rupp et al. 2004), we take the following priors for item

parameters a_j and b_j , $g_{21j}(a_j) = \phi(a_j; \mu_a, s_a^2)$,

$j = 1, \dots, I$ and $g_{22j}(b_j) = \phi(b_j; 0, s_b^2)$, $j = 1, \dots, I$ normal

distribution, so that $g_{2j}(\eta_j) = \phi_2(\eta_j; \mu_\eta, \Sigma_\eta)$, $j = 1, \dots, I$, is a bivariate normal distribution with mean vector

$$\mu_\eta = (\mu_0, 0)' \text{ and covariance matrix } \Sigma_\eta = \begin{pmatrix} s_a^2 & 1 \\ 1 & s_b^2 \end{pmatrix}.$$

3.2 – MCMC using adaptive rejection sampling (ARS) scheme

By considering the likelihood function and prior specification, a joint posterior distribution is given by:

$$f(u, \eta | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^I \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \cdot \prod_{j=1}^I \phi(u_i; 0, 1) \prod_{i=1}^n \phi(\eta_j; \mu_\eta, \Sigma_\eta)$$

or

$$f(u, \eta | D_{obs}) \propto \prod_{i=1}^n \prod_{j=1}^I \Phi(m_{ij})^{y_{ij}} (1 - \Phi(m_{ij}))^{1-y_{ij}} \times \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n u_i^2 + \frac{1}{s_b^2} \sum_{j=1}^I b_j^2 + \frac{1}{s_a^2} \sum_{j=1}^I (a_j - \mu_a)^2 \right) \right]$$

In this way, the probit-normal model can be fitted using MCMC, for example in WinBUGS (see LSAT example in Spiegelhalter et al, 1996).

In this way, the probit-normal model can be fitted using MCMC, for example in WinBUGS (see LSAT example in Spiegelhalter et al, 1996). As joint posterior distribution above is complex, note that all full conditional distributions are non-standard. Hence straightforward implementation of the Gibbs sampler using standard sampling distributions is not possible.

However, all the full conditional distributions for the probit-normal model are log-concave (log of the density is concave) according to (Sahu, 2002). Exact sampling from one dimensional log-concave distributions can be performed using rejection sampling, even when the normalizing constants are unknown (Gilks & Wild, 1992). These authors also develop an adaptive rejection sampling (ARS) scheme. ARS dynamically constructs two envelopes (one lower and one upper) for the distribution to be sampled from using successive evaluations of the density at the rejected points. The algorithm stops when one proposed point has been accepted. This procedure is default in WinBUGS 1.3 for the probit-normal model. In the case of restricted range, WinBUGS 1.4 uses the *slice sampling* (Neal, 2003). Markov chain sampling methods that adapt to characteristics of the distributions being sampled can be constructed using the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function. A Markov chain that converge to this uniform distribution can be constructed

by alternating uniform sampling in the vertical direction with uniform sampling from the horizontal “slice” defined by the current vertical position, or more generally, with some update that leaves the uniform distribution over this slice invariant. Since that this implementation is broadly used in literature we described the DAGS scheme with more generality in following section.

3.3 – An adequate formulation for the probit-normal model

An alternative way of witting the probit-normal model with I items and n subjects taking the test follows by considering that

$$Z_{ij} = m_{ij} + e_{ij} \tag{5}$$

$$e_{ij} \sim N(0,1) \tag{6}$$

$$y_{ij} = \begin{cases} 1, & Z_{ij} > 0; \\ 0, & Z_{ij} \leq 0, \end{cases} \tag{7}$$

It can be shown after some algebraic manipulations that this expressions recover the model formulated above. It follows basically by noticing that $P_{ij} = P[Y_{ij} = 1] = P[Z_{ij} > 0] = \Phi(m_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, I$, which shows that the linear normal structure for the auxiliar latent variable yields a model equivalent to the probit-normal model.

3.4 – MCMC using data augmented Gibbs sampling (DAGS) scheme

Using the new formulation, considering

$D = (\mathbf{Z}, \mathbf{y})$ as the complete data with $\mathbf{Z} = (Z_{11}, \dots, Z_{nI})'$. the “complete data”, it follows that the likelihood function of the augmented data is given by

$$L(u, \eta | D) = \prod_{i=1}^n \prod_{j=1}^I \phi(Z_{ij}; m_{ij}, 1) I(Z_{ij}, y_{ij})$$

in which $I(Z_{ij}, y_{ij}) = I(Z_{ij} > 0)I(y_{ij} = 1) + I(Z_{ij} \leq 0)I(y_{ij} = 0)$, $i = 1, \dots, n$, $j = 1, \dots, I$, is an indicator variable taking the value one if its argument is true, and the value zero otherwise.

Thus, the complete joint posterior distribution is given by

$$f(u, \eta | D) \propto \exp \left[-\frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^I (Z_{ij} - m_{ij})^2 + \sum_{i=1}^n u_i^2 + \frac{1}{s_b} \sum_{j=1}^I b_j^2 + \frac{1}{s_b} \sum_{j=1}^I a_j^2 \right) \right] I(Z_{ij}, y_{ij})$$

so that the conditional complete distributions for the probit-normal model are given by:

- $\pi(Z_{ij} | u_i, \eta_j, D_{obs}) \propto \phi(Z_{ij}; m_{ij}, 1) I(Z_{ij}, y_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, I$

- $\pi(u_i | \mathbf{Z}_{ij}, \eta_j, D_{obs}) \propto \phi(u_i; m_{u_i}, v_{u_i})$, $i = 1, \dots, n$

with $m_{u_i} = \frac{\sum_{j=1}^I a_j (Z_{ij} + b_j)}{\sum_{j=1}^I a_j^2 + 1}$, $i = 1, \dots, n$,

- $\pi(\eta_j | \mathbf{Z}_{ij}, u_i, D_{obs}) \propto \phi_2(\eta_j; m_{\eta_j}, v_{\eta_j})$,

$j = 1, \dots, I$, with

$$m_{\eta_j} = [W'W + \Sigma_\eta^{-1}] [W'Z_j + \Sigma_\eta^{-1} \mu_\eta]$$

$$v_{\eta_j} = [W'W + \Sigma_\eta^{-1}]^{-1}, \text{ with } \mu_\eta = \begin{pmatrix} \mu_0 \\ 0 \end{pmatrix}$$

$$\Sigma_\eta = \begin{bmatrix} S_a^2 & 0 \\ 0 & S_b^2 \end{bmatrix} \text{ e } W = (u, -1) \text{ with}$$

$$W'_i = (u_i, -1), \quad i = 1, \dots, n.$$

With the above conditional distributions, the DAGS can be easily implemented. Routines in R (see the MCMC pack by Martin & Quinn, 2003), Matlab (see Johnson & Albert, 1999) are available in the Web. In WinBUGS the implementation of the above procedure is not direct since it requires a correct specification for the indicator variables (See Appendix).

4 – An application

We illustrate the Bayesian approaches to probit-normal item response theory model developed in the paper, using the data set corresponding to the Mathematics Test applied in Peruvian schools. The prior specification, starting values to define the initial state of the Markov Chain, and convergence diagnostics for the Markov chain are discussed by implementing MCMC using DAGS scheme.

In this application, 14 item of the Peruvian Primary School Mathematical Test (UMC, 2001) were applied to 131 students of high social-economical status. Item response vectors are available from the authors upon request. The summary statistical scores of Mathematics Test are presented in Table 1.

The scores present a mean of 11 points and a standard deviation of almost 2 points. Of the statistics, the scores present negative asymmetry with presence of high scores. The test presents a medium reliability given by the alpha of Cronbach of 0.48, and present a mean proportion of item of 0.774 indicating a easy Test.

Table 1: Summary statistical scores of 14 items in Mathematics Test applied d to 131 students from Peruvian schools

Statistic	value	Statistic	value
Mean	10.840	Median	11
Variance	3.432	Std. Dev.	1.853
Ske	-0.795	Kurtosis	0.449
Minimum	5	Maximum	14
alpha	0.481	Mean P	0.774

4.1 – Sensitivity Analysis using different priors for item parameters

Model checking, or assessing the fit of a model, is a crucial part of any statistical analysis. Before drawing any conclusions from the application of a statistical model to a data set, an investigator should assess the fit of the model to make sure that the model explains the important features of the data set adequately (Stern & Sinharay, in press).

In Bayesian inference, a researcher can check the fit of the model in one of three ways (Gelman, Meng, & Stern, 1996): (1) checking that the posterior inferences are reasonable, given the substantive context of the model; (2) examining the sensitivity of inferences to reasonable changes in the prior distribution and the likelihood; and (3) checking that the model can explain the data, or in other words that the model is capable of generating data like the observed data. In this paper we made a sensitivity analysis of inferences to different specifications in the prior distribution of the item parameter in the probit-normal model.

In order to evaluate the sensitivity of the Bayesian Estimation for the probit-normal model by considering different priors, we conduct an analysis using the data set described above.

Bayesian estimation procedures based in MCMC was implemented in WinBUGS. Chains with 50000 iterations were generated considering thin=1, 5, 10 and discarding the 500 first iterations as in Albert (1992), so that effective sample sizes were 49500, 9900 and 4950, respectively. When using MCMC, the sampled values for initial iterations of the chain are discarded because of their dependence on the starting state and to guarantee proper convergence. Also, in this IRT model, presence of autocorrelation between chain values is expected when latent variables are introduced (Chen et al. 2000). Due to it, thin values up to 10 are recommended.

Table 2: Prior specification for item parameters $\eta_j = (a_j, b_j)$ in the probitnormal model

Prior	Author	a_j prior	b_j prior
A	Johnson & Albert (2000)	$N(2,1)$	$N(0,1)$
B	Congdon (2001)	$N(1,1)$	$N(0,1)$
C	Albert & Ghosh (2000)	$N(0,1)$	$N(0,1)$
D	Sahu (2002); Albert & Ghosh (2000)	$N(0,1)$	$N(0,10000)$
E	Spiegelhalter et al. (1996)	$N(0,1)I(0,.)$	$N(0,10000)$
F	Sahu (2002); Patz & Junker (1999)	$N(1,0.5)I(0,.)$	$N(1,2)$

As it has been mentioned, proper priors for a_j and b_j guarantee that the complete posterior for the model is proper. Albert & Ghosh (2000) mentions that the choice of a proper prior distribution on the latent trait resolves particular identification problems, and, further, informative prior distributions placed for a_j and b_j can be used to reflect the prior belief that the values of the item parameters are not extreme (in the frontier of the parametric space). In the common situation where little prior information is available about the difficulty parameters, one can chose s_b^2 to be a large value. This

choice will have a modest effect on the posterior distribution for non extreme data, and will result in a proper posterior distribution when extreme data (where students are observed to get correct or incorrect answers to every item) is observed (Albert & Ghosh, 2000), also, Sahu (2002) states that larger values of the variance led to unstable estimates. This priors is denominated as vague priors in Roberts (2001)

In Table 2 it is shown some priors considered in the literature for the item parameter $\eta_j = (a_j, b_j)$ in the probit-normal model. We observed that $N(0,1)I(0,.)$ is the notation for a normal distribution with mean 0 and variance 1 truncated for negative values.

Priors A, B and C are precise and priors D and E are vague priors assigned to the difficulty parameters. Prior E and F assigned to the discrimination parameters are truncated priors. Estimates of item discrimination (a_j) and difficulty parameters (b_j) based in posterior mean for different priors given in Table 2 are depicted in Figure 1. According to the Figure 1, the probit-normal model is insensitive to the prior specifications for the difficulty and discrimination parameters in the literature. A one-way variance analysis for means of posterior coefficient of variation for discrimination and difficulty parameters in the test (see Table 3), are not significant in relation to the different priors used (a : $F(5, 78) = 1.014$; $p < 0.415$, b : $F(5, 78) = 0.631$; $p < 0.676$).

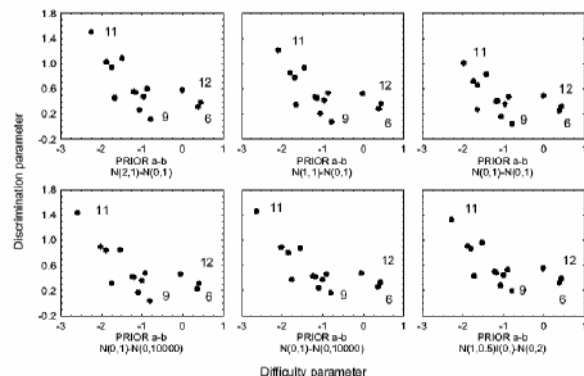


Figure 1: Posterior mean of discrimination and difficulty parameters for different priors in the probit-normal model (N=131,I=14)

In sensitivity analysis we to fit several probability models to the same data. To compare the models corresponding to the six different prior distributions, we computed the posterior expected deviance ($Dbar$), the deviance of the posterior means ($Dhat$), the effective number of parameters ρ_D and deviance information criterion (DIC), as presented by Spiegelhalter et al. (2002). $Dbar$, is the posterior mean of the deviance that is defined as $-2 * \log(\text{likelihood})$. $Dhat$ is a point estimate of the deviance obtained by substituting in the posterior means of estimates of parameter of model. ρ_D is given by $\rho_D = Dbar - Dhat$. DIC is given by $DIC = Dbar + \rho_D = Dhat + 2 * \rho_D$. The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed. They claim that the

DIC as implemented in the WinBUGS software can be used to compare complex models and large differences in the criterion can be attributed to real predictive differences in the models, although the approach has received critics from several authors (see discussion in Spiegelhalter et al. 2002). In hierarchical modelling with auxiliary latent variables as in the probit-normal model, the “likelihood” or “model complexity” is not unique so that the deviance (or *DIC* and *D* which are based on it) of a model with latent variables is not unique and can be calculated in several ways (Delorio and Roberts, 2002). With auxiliary latent variables, WinBUGS uses a complete likelihood of the observed variable and the auxiliary latent variable introduced (as fixed effects and random effects in hierarchical modelling) to obtain posterior distributions for the parameters of interest. When this is the case, marginal *DIC*s for the observed variable (fixed effect) and auxiliary latent variable (random effect) are presented. For a proper comparison of the proposed models, we consider marginal *DIC* for the observed variables because the focus of the analysis is in $p(y|u, \eta)$ and although auxiliary random variables are introduced (in two steps) they are not the focus of the analysis.

Table 3: Summary of posterior coefficient of variation (c.v.) for item parameters and Criterio of Fit for probit-normal models with priors different

	Prior a_j posterior c.v.		b_j posterior c.v.		Criterion of Fit			
	mean	sd	mean	sd	Dbar	Dhar	p_D	DIC
A	0.590	0.344	-1.895	6.747	1445.00	1352.90	92.11	1537.11
B	0.698	0.549	-0.999	1.827	1453.16	1368.36	84.80	1537.96
C	0.890	1.033	-0.979	1.286	1461.09	1383.61	77.48	1538.57
D	0.901	1.090	-1.099	0.646	1453.22	1375.48	77.75	1530.97
E	0.532	0.114	-1.101	0.634	1450.14	1369.86	80.28	1530.43
F	0.481	0.111	-1.048	1.101	1446.60	1358.26	88.34	1534.94

Figure 2 presents comparisons for different priors by considering plotting the means vs standard deviations for posterior coefficient of variation for item parameters. For the discrimination and difficulty parameter, the estimates with different priors are organized in the x-axis (horizontal) from the smaller to the greater mean for the posterior coefficient of variation and the y-axis (vertical) from the smaller standard deviations to the greater standard deviation for the posterior coefficient of variation). Clearly, priors with great precision in discrimination parameter, as it is the case of the truncated priors *F* and *E*, take to more precise estimates. This result suggest that priors with hyperparameter that admits negative values for discrimination parameter (mean equal to zero in prior *C* and *D*) to take more imprecise estimates.

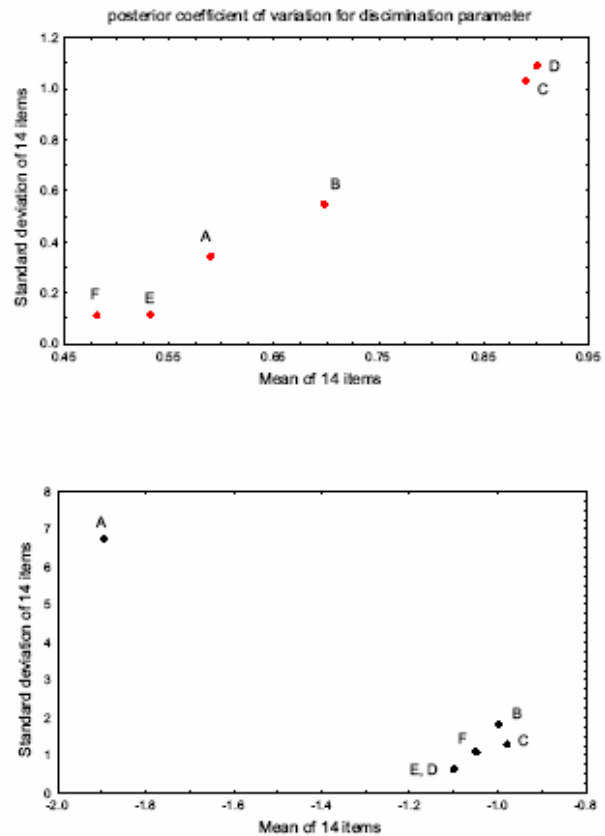


Figure 2: Comparison of the probit-normal model with different priors

With respect to the difficulty parameter, the estimates with different priors are organized from the least (in order, prior *D* and *E*) to the more precise (least vague) prior (in order, prior *A*, *B* and *C*). A large mean and standard deviation are obtained for Prior *A* because the estimates of coefficient of variation for the difficulty parameter of item 3 is large. The imprecision for item 3 follows because the estimates are closed to zero and smaller standard deviations make great coefficients of variation. This fact implies that for prior *A* estimates for difficulty parameter are closed to zero.

Although the priors considered lead to similar posterior analyzes (fit), we find it more appropriate to use priors *E* and *F* because it results in great precision for discrimination estimates and difficulty parameters. Since priors with large variance for the difficulty parameter can lead to posteriors imprecise we considered prior *F* more appropriated. This fact is noted in Sahu (2002), where other choices for the hyper parameters were also investigated, and similar conclusions are observed in our paper.

4.2 - Inference and parameter interpretation

Further studies were performed with the priors $a \sim N(1; 0.5)I(0,)$, and $b \sim N(0, 2)$ as proposed in Sahu (2002). We consider burn-in of 1000 and the chain run for 10000 more iterations after the bur-in with *thin* = 10 to reduce chain dependence. In this work we decided to use five parallel chains. Convergence was diagnosed via

visualization of trace plot using Gelman and Rubin’s statistics (1992) present in the software R at package CODA (Best et al. 1997).

The average bur-in time takes about 90 seconds and to generate about 10000 further iterations for each chain, takes about 957 seconds on a AMD Duron(tm) Processor with 112 MB RAM). Figure 3 depicts chain history and the empirical density for item 12 and the mean and standard deviations for the latent variable describing subject’s ability. Estimates of item discrimination and difficulty parameters for the probitnormal model is presented in table 4. Item 11 is the most discriminative while item 9 is the least. Also, item 11 is the easiest while item 12 is the most difficult.

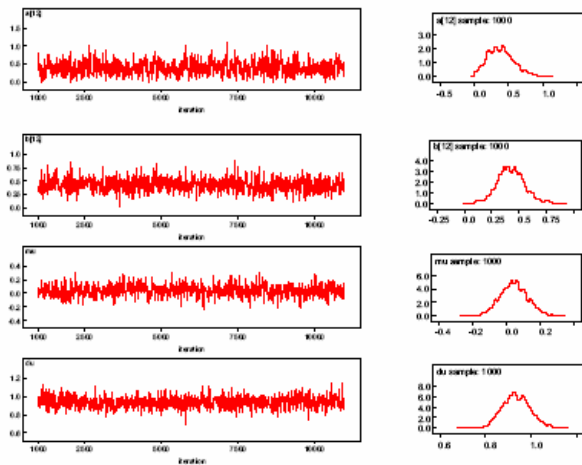


Figure 3: Chain history and density for item parameters of item 12 and mean and deviation standard of latent variable

Estimates of item discrimination and difficulty parameters for the probitnormal model is presented in table 4. Item 11 is the most discriminative while item 9 is the least. Also, item 11 is the easiest while item 12 is the most difficult.

Item 11 says: “Luisa, Dora and Mary bought some cloth. Luisa bought half of a meter, Dora bought 75 centimeters and Mary bought fifty centimeters. Which ones did buy the same quantity of cloth?”. Item 12 say: “A recipient receive 4.5 liters of water by minute. How many liters of water will have the recipient after of one and half hour?”. On the other hand, item 6 say: “It solves the following operations decimal: $0.75 - 0.2 + 1.2 - 0.30$ ”. Clearly, form the formulated items, (latent) ability required to correctly solving item 11 is smaller than that required for solving item 12. For a given ability value, a student has greater probability of success with item 11 than with item 12. Hence, according to the probit-normal model, item 11 is easier than 12.

Table 4: Summary of posterior inference for the parameters under the probitnormal model

		mean	sd	$P_{2.5}$	median	$P_{97.5}$
discrimination Parameter	a_1	0.54	0.24	0.13	0.52	1.09
	a_2	0.29	0.18	0.03	0.27	0.70
	a_3	0.55	0.24	0.16	0.53	1.09
	a_4	0.91	0.33	0.33	0.88	1.61
	a_5	0.49	0.24	0.09	0.48	1.02
	a_6	0.32	0.18	0.03	0.30	0.73
	a_7	0.88	0.34	0.29	0.84	1.60
	a_8	0.97	0.35	0.39	0.92	1.80
	a_9	0.20	0.14	0.01	0.18	0.51
	a_{10}	0.49	0.23	0.08	0.48	1.02
	a_{11}	1.35	0.41	0.64	1.32	2.18
	a_{12}	0.39	0.19	0.06	0.38	0.80
	a_{13}	0.45	0.23	0.05	0.43	0.93
	a_{14}	0.41	0.26	0.03	0.37	0.97
difficulty parameter	b_1	-0.90	0.16	-1.25	-0.90	-0.60
	b_2	-1.09	0.15	-1.37	-1.08	-0.80
	b_3	-0.03	0.13	-0.27	-0.03	0.23
	b_4	-1.91	0.33	-2.62	-1.88	-1.36
	b_5	-1.23	0.18	-1.62	-1.22	-0.90
	b_6	0.37	0.11	0.16	0.37	0.61
	b_7	-1.82	0.33	-2.63	-1.78	-1.28
	b_8	-1.53	0.31	-2.30	-1.50	-1.03
	b_9	-0.80	0.13	-1.06	-0.80	-0.57
	b_{10}	-1.19	0.17	-1.55	-1.18	-0.87
	b_{11}	-2.31	0.46	-3.32	-2.25	-1.53
	b_{12}	0.43	0.12	0.21	0.42	0.67
	b_{13}	-1.01	0.16	-1.35	-1.00	-0.71
	b_{14}	-1.71	0.23	-2.20	-1.70	-1.32
latent variable	u mean	0.94	0.06	0.82	0.93	1.06
	u s.d.	0.04	0.08	-0.12	0.04	0.21

It is also worth noticing with respect to item 11 that a small change in student’s ability (knowing the meaning of half meter, for instance) results in greater probability of success for this item. On the other hand, with respect to item 9, it can be depicted that small changes in student’s ability (knowledge of decimal places, for instance) is not directly translated into a greater increase in the probability of correctly answering the item. Thus, item 11 seems to present greater discriminatory power because it allows to better distinguish between students that know or don’t know some specific knowledge about the items. Box-plots for the item parameters based in posterior inference are presented in Figure 4 indicating the variation in posterior distribution for item parameters. This results is impossible

with frequentist approaches and clearly shows the advantage in Bayesian estimation. For further details on the interpretation of item parameters see Johnson and Albert (1999).

With respect to the latent variable ability, we have that its posterior mean is 0.94 which indicates that the group of student present negative asymmetry also found with the observed scores. Hence, estimated abilities are connected with the observed scores.

5 – Final Discussion

One of the objectives of this paper is to contribute with some literature connected to the probit-normal model in the item response theory literature to the Brazilian statistical community. This model was chosen because it allows extensions to three parameter models, multidimensional models, multilevel models, inclusion of predictor variables and confirmatory factor analysis model. The paper considers two procedures using MCMC

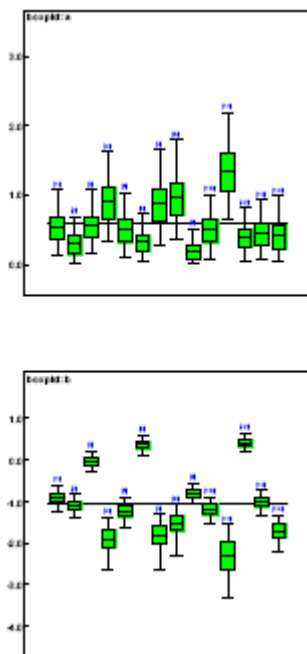


Figure 4: Box-plots for discrimination (a) and difficulty (b) item parameters in probit-normal model

methodology to make inference on item and ability model parameters. The methodology was used to a mathematical test applied to 131 sixth grade students from Peruvian school. A sensitivity analysis by checking model adequacy that follows by using a series of prior distributions is also conducted. It includes the specification of vague prior distributions for the difficulty parameters and precise parameters for the discrimination parameters. The priors specified in Table 1 were suggested in previous studies by other authors. The different priors considered lead to similar estimates of the DIC (described in Spiegelhalter et al., 2002) leading to the conclusion that the Bayesian analysis for the data set under consideration is not sensible to the priors considered. However, the prior considered in Sahu

(2002), which specifies that $a_j \sim N(1; 0.5)I(0,)$ and $b_j \sim N(0, 2)$, $j = 1, \dots$, seems to us the most adequate one. Bazan et al. (2004) proposed the use of a more general family of link functions $F_{ij}(\bullet)$ which can accommodate also asymmetric distributions for the abilities in the population which can be applied in this data set since asymmetry was found in the score.

References

- [1] Albert, J. H. (1992). Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics*, 17, 251 - 269.
- [2] Albert, J.H. & Ghosh, M. (2000). Item response modeling. *Generalized Linear Models: A Bayesian Perspective*(D. Dey, S. Ghosh & Mallick, eds.), Marcel-Dekker, New York, 173-193.
- [3] Andrade, D. F. , Tavares, H. R., & Valle, R. C. (2000). Introdução à Teoria da resposta ao Item : Conceitos e Aplicações. 14o SINAPE : Caxambu, MG.
- [4] Assunção, C. (1999). Estimação dos parâmetros de modelos da teoria de resposta ao item e aplicações. Mestrado em Estatística - UFMG.
- [5] Baker, F.B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- [6] Bartholomew, D.J., & Knoot, M. (1999). *Latent variable models and factor analysis*. (2nd ed.). London: Arnold. (Kendall's Library of Statistics 7.
- [7] Bazán, J., Bolfarine, H., & Branco, M. (2004). A skew item response model. ISBA 2004 World Meeting. Viña del Mar. Chile, May 23-27, 2004. ISBA (International Society for Bayesian Analysis).
- [8] Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation of multidimensional IRT models. *Psychometrika*, 66, 541-562.
- [9] Best N. G., Cowles M. K. & Vines, S, K, (1997) *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.4*. [Computer program]. MRC Biostatistics Unit, Cambridge University.
- [10] Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord & M. R. Novick. *Statistical*

- Theories of Mental Test Scores*. Reading, MA : Addison-Wesley.
- [11] Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* 64,153-168.
- [12] Borsboom, D., Mellenbergh, G. J., & van Heerden, J.(2003). The Theoretical Status of Latent variables. *Psychological Review*, 110, 203-219.
- [13] Congdon P (2001). *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, UK.
- [14] Chen, M-H, Shao, Q. M, & Ibrahim, J. G (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer Verlag.
- [15] Delorio, M. & Robert, C. P. (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin and van der linde (2002). *Journal Royal Statistics Society. Serie B.* 64, 629-630.
- [16] Du Toit, M. (2003). *IRT from SSI*. Scientific Software International.
- [17] Fischer, G. & Molenaar, I. (1995). *Rasch Models. Foundations, recent development, and applications*. The Nerthelands: Springer-Verlag.
- [18] Fox, J. P., & Glas, C. A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 271-288.
- [19] Glas, C. A. W. & Meijer, R. (2003). A Bayesian Approach to Person Fit Analysis in Item Response Theory Models. *Applied Psychological Measurement*, 27,217-233.
- [20] Fox, J. P., & Glas, C. A.W. (2003). Bayesian Modeling of measurement error in predictor variables using item response theory. *Psychometrika*, 68, 169-191.
- [21] Froelich, A. G. (2000). *Assesing unidimensionality of fit item and some asymptotics of parametric item response theory*. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, Departament of Statistics.
- [22] Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733-807.
- [23] Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
- [24] Ghosh, M., Ghosh, A., & Chen, Ming-Hui & Agresti, A. (2000). Noninformative priors for one parameter item response models. *Journal of Statistical Planning and Inference*. 88, 99-115.
- [25] Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- [26] Holland, P., & Rosenbaum, P. (1986). Conditional Association and Unidimensionality in Monotone Latent variable models. *The Annals of Statistics*. 14 1523-1543.
- [27] Jackman, S. (2001). Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking. *Political Analysis*, 9 , 227-242.
- [28] Johnson, V., & Albert, J. (1999). *Ordinal Data Modeling*. New York, MA: Springer-Verlag.
- [29] Kim, S., Cohen, A. S., Baker, F. B., Subkoviak, M. J., & Leonard, T. (1994). An investigation of hierarchical Bayes procedures in item response theory. *Psychometrika*, 59, 405-421.
- [30] Linardakis, M. & Dellaportas, P. (2002). An approach to multidimensional item response modeling. E. I. George. (Ed.). *Bayesian Methods with applications to science policy and oficial statistics*. 331-340.
- [31] Lord, F. M. (1952). *A theory of test scores*. New York: Psychometric Society.
- [32] Lord, F., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Adisson-Wesley.
- [33] Martin, A.D., e Quinn, K. M. (2002). MCMCpack. scythe.wustl.edu/mcmcpack.html.
- [34] Mislevy, R., J. & Bock, R.D. (1990). *BILOG: Maximum likelihood item analysis and test scoring with logistic models for binary items* [Computer program]. Chicago: Scientific Software International.
- [35] Neal, Radford. (2003). Slice Sampling. *The Annals of Statistics*, Vol. 31, No. 3, 705-767.
- [36] Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models.

- Journal of Educational and Behavioral Statistics*, 24, 146-178.
- [37] Robert, C., P. (2001) *The Bayesian Choice: from decision-theoretic foundations to computational implementation*. 2nd ed. New York: Springer-Verlag.
- [38] Rupp, A., Dey, D. K., & Zumbo, B. (2004) (in press). To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian Methodology To Item Response Modeling. *Structural Equations Modeling*.
- [39] Sahu, S. K. (2002). Bayesian Estimation and Model Choice in Item Response Models. *Journal of Statistical Computation and Simulation*, 72, 217-232.
- [40] Segall, D. (2003). Confirmatory Item Factor Analysis using Markov Chain Monte Carlo Estimation with Applications to Online Calibration in CAT. Pre-print.
- [41] Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W.R. (1996). *BUGS 0.5 examples* (Vol. 1 Version i). Cambridge, UK: University of Cambridge.
- [42] Spiegelhalter D., Thomas, A. Best N. & (2003). *WinBUGS version 1.4* [Computer program]. Imperial College & Medical Research Council Biostatistics Unit. Institute of Public Health, Cambridge University.
- [43] Spiegelhalter D J, Best N G, Carlin B P and van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc. B.* 64, 583-640.
- [44] Stern, H. , & Sinharay, S. (in press). Model Checking and Model Diagnostics.
- [45] Tsutakawa, R. M. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics*, 9, 263-276.
- [46] UMC (2001). *Análisis de ítems de las pruebas Crecer 1998. Resultados de lógico matemática en sexto grado de primaria*. Boletín 13. Unit of Mensuration of Educational Quality. Ministry of Education. Perú. Available from <http://www.minedu.gob.pe/umc>
- [47] van der Linden, W.J. & Hambleton, R. K. (1997). *Handbook of Modern Item response Theory*. New York: Springer-Verlag.
- [48] Wang, X, Bradlow, E. T, & Wainer, H. (2003). A General Bayesian Model for Testlets: Theory and Applications. *Applied Psychological Measurement*, 26, 109-128.

Appendix

```

model
{
  #likelihood function
  for (i in 1 : n) {
    for (j in 1 : I) {
      m[i,j] <- a[j]*u[i] - b[j]
      z[i,j] ~ dnorm(m[i,j],1)I(lo[y[i,j]+1],up[y[i,j]+1])
    }
  }
  #priors F for item parameters
  for (j in 1:I) {
    b[j] ~ dnorm(0,2);
    a[j] ~ dnorm(1,0.5)I(0,);
  }
  #prior for latent variable
  for (i in 1:n) { u[i] ~ dnorm(0,1) }
  # auxiliary latent variable
  lo[1] <- -50; lo[2] <- 0; \# i.e., z| y=0 ~ N(m,1)I(-50,0)
  up[1] <- 0; up[2] <- 50; \# i.e., z| y=1 ~ N(m,1)I(0,50)
  # mean and standard deviation for latent variable
  mu<-mean(u[])
  du<-sd(u[])
}
list(n=131, I=14, y= structure(.Data =c(
1,1,0,1,1,0,1,1,1,1,1,0,0,1,
1,1,1,1,1,1,1,1,1,1,1,0,1,1,
...
...
1,0,0,1,1,1,1,1,1,1,1,1,1,1,
1,1,0,1,0,0,1,1,0,0,1,0,1,1),
.Dim=c(131,14))) )
Inits list(b = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0),
a=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1)) ■

```

4 – Convites

Carta-convite enviada a Lista da ABE

Caros Colegas,

Estivemos no Chile em meados de janeiro participando do IV Seminário Chileno de Estatística Bayesiana e ficamos surpreendidos com a qualidade científica do encontro e com a motivação do grupo chileno, apesar dos poucos recursos que tem. Chamou-nos a atenção a rapidez com que o grupo de chilenos, simpatizantes da escola bayesiana, vem crescendo e mais ainda a sua organização enquanto grupo. Para se ter uma idéia, apesar de serem ainda pouco numerosos, eles já se organizaram em um Capítulo Chileno do ISBA, o que lhes permite (por exemplo) conseguir recursos para a vinda de pesquisadores externos para seus encontros.

Achamos que um capítulo brasileiro do ISBA pode ser uma boa maneira de nos fortalecermos ainda mais como grupo, ampliando nossas chances de captar recursos para a organização dos nossos eventos futuros e, também, (para aqueles que não estão associados ao ISBA) e uma boa oportunidade de interagir ainda mais com o que rola no universo bayesiano. Ao associar-se a algum capítulo do ISBA automaticamente associa-se ao ISBA tendo, portanto, todas as regalias que este status confere.

Algumas das vantagens do capítulo bayesiano brasileiro seriam

1. Teríamos contrapartidas externas para organizar os nossos encontros bayesianos.
2. Todas as nossas atividades seriam divulgadas na página e no boletim do ISBA.
3. Os sócios do ISBA, via capítulo, receberiam os boletins e teriam descontos nos encontros bayesianos internacionais.

Estamos dispostos a organizar o capítulo brasileiro, caso vocês concordem com isto, e também julguem que este é um passo realmente importante para a nossa comunidade.

Aqueles que se interessarem em fazer parte do Capítulo Brasileiro do ISBA pedimos que nos enviem um e-mail contendo NOME COMPLETO, ENDERECO FUNCIONAL COM FAX E TELEFONE, E-MAIL e DOCUMENTO DE IDENTIDADE, para que possamos mantê-los informados sobre a viabilização da criação do capítulo. Agradecemos a atenção de todos vocês e ficamos aguardando um e-mail de vocês.

Rosângela, Josemar e Márcia

II COBAL – Congresso Bayesiano da América Latina

Seguindo o sucesso do I COBAL (Ubatuba, Brasil, 2002), o Segundo Congresso Bayesiano da América Latina será realizado de 6 a 10 de fevereiro de 2005, no Hotel Presidente Intercontinental Los Cabos em San José del Cabo, Baja Califórnia, México.

Da programação científica do II COBAL constarão conferências convidadas e sessões Pôster, cobrindo aspectos teóricos e aplicados dos Métodos Bayesianos.

Contribuições para a sessão pôster devem ser enviadas antes de **1º de setembro de 2004** para o email cobal2@sigma.iimas.unam.mx, colocando no *subject* "Program Committee". Os resumos deverão conter título, autores e no máximo 200 palavras. A aceitação dos trabalhos será comunicada no final de outubro de 2004.

A acomodação no Presidente Intercontinental Hotel no período do II COBAL (inclui 4 noites) estarão incluídas na taxa de inscrição da conferência (que é de 650 dólares

antes de 30 de novembro de 2004 e de 750 dólares após esta data). A taxa de inscrição para estudantes é de 570 dólares até 30 de Novembro de 2004 e de 660 dólares após esta data.

O comitê organizador do II COBAL é composto por: Manuel Mendoza (Presidente) e Luiz Enrique Niedo do ITAM-Mx; Eduardo Gutiérrez, Ramsés Mena e Raúl Rueda do IIMAS-UNAM-Mx; Andrés Christen do CIMAT-Mx e Gabriel Huerta da Universidad Nuevo México-EUA. Participam do Comitê Científico, Manuel Mendoza, Eduardo Gutierrez do México, Dani Gamerman do Brasil, Pilar Iglesias do Chile e Gabriel Huerta e Peter Mueller dos Estados Unidos.

San José del Cabo é uma pequena cidade localizada ao sul da península de Baja Califórnia. É uma autêntica cidade mexicana, situada próxima do mar e de colinas. Um espetáculo a parte que pode ser apreciado em fevereiro de cada ano, é a migração de baleias das regiões frias do ártico para as águas mornas e calmas desta região.

O Aeroporto de San José del Cabo é um terminal internacional e pode-se voar para lá diretamente de algumas cidades nos Estados Unidos como, por exemplo, Houston, Denver e Atlanta. Também há conexões para a Cidade do México e outros aeroportos mexicanos.

Maiores informações podem ser obtidas através do email: cobal2@sigma.iimas.unam.mx ou no URL www.dpye.iimas.unam.mx/cobal2.

SUGESTÕES

QUALQUER TIPO DE SUGESTÃO, RECLAMAÇÃO, DOAÇÃO, ETC QUE POSSA SER UTILIZADA PARA MELHORAR A QUALIDADE DO BOLETIM É MUITO BEM VINDA.

vjosemar@power.ufscar.br