Aishameriane Venes Schmidt

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

Florianópolis

2017

Aishameriane Venes Schmidt

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

Trabalho de Conclusão apresentado ao Curso de Ciências Econômicas do Centro de Ciências da Administração e Socioeconômicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Bacharel em Ciências Econômicas.

Universidade do Estado de Santa Catarina - UDESC

Centro de Ciências da Administração e Socioeconômicas

Bacharelado em Economia

Orientador: Fernando Pozzobon

Florianópolis

2017

Aishameriane Venes Schmidt

Estimação Bayesiana em Modelos Lineares com Aplicações em Economia

Trabalho de Conclusão apresentado ao Curso de Ciências Econômicas do Centro de Ciências da Administração e Socioeconômicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Bacharel em Ciências Econômicas.

Trabalho aprovado. Florianópolis, novembro de 2017:

Fernando Pozzobon
Orientador

Professor
Convidado 1

Professor
Convidado 2

Florianópolis
2017

Este trabalho é dedicado ao Gilles, pelo apoio e suporte;
ao professor Fernando, pela orientação e confiança;
aos professores e professoras, pelos ensinamentos;
aos e às colegas de graduação e mestrado, pela torcida;
aos amigos e amigas, pela paciência e compreensão;
às minhas famílias, pelo amor e oportunidades;
ao Kiiro, Hiro, Bernardo e Léia (em memória), pelas alegrias;
e ao Lilo (em memória), por tudo.

"What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has no predicted observable results that have not occurred."

(Harold Jeffreys, 1961)

RESUMO

Com o advento da computação, a econometria bayesiana tem cada vez mais espaço nos trabalhos acadêmicos, tanto por sua flexibilidade na incorporação de informações a priori como também pela possibilidade de gerar estimativas confiáveis mesmo na ausência ou escassez de eventos observáveis. Neste sentido, o presente trabalho se propõe a apresentar o modelo normal de regressão linear múltipla comparando os resultados obtidos pela estimação de mínimos quadrados ordinários (MQO) com as estimativas bayesianas utilizando uma priori conjugada normal-gama. Os principais teoremas das estimativas bayesianas para o modelo foram demonstrados analiticamente. Adicionalmente, realizaram-se simulações onde foram comparados os resultados obtidos pela abordagem clássica e bayesiana, considerando diferentes especificações dos valores de parâmetros da densidade a priori e processo gerador dos dados. Nas simulações, constatou-se os resultados teóricos desenvolvidos, tanto com relação à convergência das estimativas bayesianas para o estimador de MQO com o aumento do tamanho da amostra, como para os efeitos de variação nas informações a priori. Ao final, utilizando o ferramental bayesiano apresentado, adaptou-se o estudo de Bonini e Pozzobon (2016) sobre discriminação de gênero e prêmio salarial no mercado de Tecnologia da Informação, utilizando os dados da Rais de 2011 para trabalhadores de TI do Estado de Santa Catarina. As estimativas obtidas, utilizando como priori as estimativas reportadas no estudo original, foram consistentes com as de Bonini e Pozzobon (2016) e o desempenho da estimativa bayesiana na previsão de dados para fora da amostra foi superior do que o desempenho das estimativas de MQO.

Palavras-chave: Econometria Bayesiana. Modelos Lineares. Simulação de Monte Carlo. Mercado de Trabalho.

ABSTRACT

The advances in Computer Sciences from the past decades contributed to the development of modern Bayesian Econometric methods, which have gained space in academic research, both for the flexibility in allowing the explicit use of prior information in the models and because of the possibility to generate efficient estimates even in the lack or total absence of observable events. This work presents the normal linear multiple regression model, and compares the estimates obtained trough ordinary least squares method (OLS) with the bayesian estimates using a natural conjugate normal-gamma prior. The main theorems of the bayesian approach were proved. Moreover, simulation experiments, comparing the classical and bayesian estimates, changing parameters and other model specifications such as the prior hyperparameters, were made. In the simulated data, the theoretical results were observed, like convergence from bayesian estimates to OLS values with increase in the sample size as well as with changes in the hyperparameters. An empirical evaluation, based on the work of Bonini and Pozzobon (2016), was performed for a subsample of the original article. Specifically, the labor data for the year of 2011 of IT analysts from Santa Catarina State were used to verify gender wage discrimination and another factors that could influence wages in this sector. The findings were consistent with those from Bonini and Pozzobon (2016), and the out of sample estimates using bayesian methods were superior with those obtained using OLS.

SUMÁRIO

1 INTRODUÇÃO 13

2 REVISÃO BIBLIOGRÁFICA 18

2.1 Conceitos básicos de inferência bayesiana 19

2.2 Econometria Bayesiana 28

2.3 O modelo linear normal de regressão clássico com priori normal-gama conjugada 30

2.3.1 A função de verossimilhança 30

2.3.2 A densidade a priori 32

2.3.3 A densidade a posteriori 33

2.3.4 A densidade preditiva a posteriori 34

3 SIMULAÇÃO 36

3.1 Simulações sem anomalias nos dados 36

3.1.1 Especificações utilizadas 36

3.1.2 Análise dos resultados 38

3.2 Simulação com erro de medida 41

4 APLICAÇÃO 45

4.1 Prêmio salarial e discriminação de gênero no mercado de trabalho de TI 45

4.2 Descrição da base de dados 47

4.3 Determinação da priori 51

4.4 Resultados do modelo 51

4.5 Previsão fora da amostra 55

5 CONSIDERAÇÕES FINAIS 57

REFERÊNCIAS 59

APÊNDICES 63

APÊNDICE A – DEMONSTRAÇÕES 64

APÊNDICE B – SIMULAÇÕES 77

APÊNDICE C – CÓDIGOS 115

1 INTRODUÇÃO

A econometria, enquanto uma área distinta dentro das ciências econômicas, tem sua gênese no início do século XX com o uso intensivo de métodos estatísticos para validação de hipóteses a respeito de fenômenos observados na economia (MORGAN, 2003). De fato, o caráter subjetivo das incertezas que norteiam os problemas econômicos desenvolve um papel fundamental na teoria econômica e economia aplicada (GEWEKE, 1999), pois a maioria dos modelos econométricos surge a partir da observação da realidade, de forma que são necessárias metodologias flexíveis que incorporem os fatos observados.

Métodos econométricos permitem que pesquisadores, de posse de dados diversos, possam testar hipóteses e investigar fenômenos, dando respostas quantitativas a perguntas econômicas. Por exemplo, na área de macroeconomia é de interesse prever quais serão as flutuações do Produto Interno Bruto (PIB) para os próximos semestres, investigar quais possíveis determinantes dos ciclos econômicos ou ainda prever o comportamento de agentes a uma nova política econômica (DEL NEGRO; SCHORFHEIDE, 2012). Ferramentas econométricas também são utilizadas na área de microeconomia para avaliação de impacto de políticas públicas, determinação de curvas de oferta e demanda de certos mercados, etc.

De maneira geral, a abordagem econométrica tenta relacionar uma variável y, dita *variável dependente* ou *explicada*, com outras variáveis, chamadas de *variáveis independentes* ou *explicativas* e usualmente denotadas por X_i (GREENBERG, 2014). Por exemplo, y pode representar a decisão de compra de determinado pacote de televisão a cabo e as variáveis explicativas X_i podem incluir valores, opções de canais, facilidade na instalação, entre outros fatores que influenciam na decisão dos consumidores.

Métodos de inferência estatística, tanto na sua versão clássica quanto bayesiana, buscam inferir, a partir de dados coletados, conclusões a respeito dos parâmetros populacionais. A estatística clássica, que se baseia em p-valores e na noção de repetição de experimentos, foi desenvolvida em grande parte no início do século XIX e teve a formalização das ideias a partir dos trabalhos de Karl Pearson, Ronald Fisher e Jerzy Neyman (PAULINO; AMARAL.; MURTEIRA, 2003). Métodos clássicos de inferência se baseiam na ideia de que existem parâmetros populacionais desconhecidos, mas fixos, que influenciam no modelo gerador dos dados. Os dados, por sua vez, serão modelados como variáveis aleatórias que seguem, por hipótese, a forma funcional de uma distribuição de probabilidade conhecida[1] cujos parâmetros serão estimados. Sob este prisma, não é possível calcular a probabilidade de um parâmetro θ ser maior do que um determinado

[1] Os métodos que não assumem nenhuma distribuição adjacente são conhecidos como não paramétricos e não serão objeto de estudo deste trabalho.

valor a, uma vez que θ é constante. Também não é possível fazer inferências ou atribuir probabilidades a eventos que não foram observados.

A estatística bayesiana, por outro lado, dá um tratamento diferenciado à probabilidade: tudo aquilo que é desconhecido pode ter uma probabilidade associada (no sentido da probabilidade de Kolmogorov) e portanto, nesta abordagem, parâmetros desconhecidos podem ser tratados como variáveis aleatórias (GREENBERG, 2014). De fato, os métodos bayesianos se ocupam de encontrar as distribuições de probabilidade para os parâmetros e a partir delas fazer os cálculos de interesse.

No enfoque bayesiano, a probabilidade é interpretada como o grau de crença na ocorrência de determinado evento. Esta interpretação é chamada de *interpretação subjetiva de probabilidade*. Isso possibilita pensar não apenas nos eventos usualmente trabalhados na estatística frequentista - os quais precisam ser eventos aleatórios que podem ser reproduzidos infinitas vezes sob mesmas condições mantendo as distribuições de probabilidade subjacentes - mas também em eventos raros (que podem até mesmo ter ocorrência única) ou ainda contrafactuais (que nunca ocorrem). É importante ressalvar que o mesmo procedimento não pode ser feito utilizando a abordagem frequentista. Por exemplo, o estimador de máxima verossimilhança (EMV) é afetado pela não ocorrência dos eventos. Se, em três lançamentos de uma moeda, forem observadas três faces cara, o EMV para o resultado "sair coroa" será zero.

Na estatística clássica, também não faz sentido o questionamento sobre qual a probabilidade de uma pessoa ser "alta" ou "baixa", pois ou a pessoa é alta ou não, no sentido que a altura é uma característica intrínseca ao indivíduo. Isto evidencia a limitação da modelagem clássica para tratamento de eventos únicos. Na abordagem Bayesiana, como a probabilidade é subjetiva, essa probabilidade será um valor entre 0 e 1.

Para pesquisas na área de economia, considerando que na maioria das vezes não é possível executar experimentos controlados e muito menos replicá-los indefinidamente, métodos que comportem essa natureza não experimental dos dados econômicos passam a ser relevantes (GEWEKE, 1999). Outro aspecto das pesquisas em economia são as "crenças" sobre alguns assuntos, por exemplo, a de que elasticidades devem ser maiores que 1, em módulo. Essas crenças a priori são herdadas da teoria econômica e a abordagem bayesiana possibilita incorporá-las aos modelos, conferindo maior flexibilidade e adaptabilidade dos modelos à situação real. Adicionalmente, o método bayesiano consegue incorporar todo o suporte da função de verossimilhança, não apenas seu máximo (como no EMV), isto é, enquanto o EMV resume toda a informação da amostra em um único valor, os métodos bayesianos utilizam a função de verossimilhança de maneira a ser integralmente incorporada ao modelo. Isso implica que as previsões podem ser a respeito de uma densidade completa de um parâmetro e não apenas um valor pontual.

Métodos de inferência estatística clássica não permitem que sejam incorporadas

na análise informações conhecidas, por exemplo, de estudos pré-existentes. Também não é possível formular hipóteses específicas sobre os parâmetros: uma vez que eles são fixos (mesmo que desconhecidos), não se pode atribuir probabilidade aos mesmos. Outra dificuldade no uso dos métodos frequentistas é em relação à atualização das inferências utilizando novas informações: uma vez que os dados foram coletados e as inferências realizadas, o modelo é imutável no sentido de que novos dados não podem ser incorporados ao modelo criado.

Em contrapartida, a metodologia bayesiana permite tratar os parâmetros como variáveis aleatórias. Desta forma, a inferência a respeito deles passa a ser feita de maneira direta: pode-se atribuir uma distribuição de probabilidade ao parâmetro e com isso fazer perguntas a respeito de probabilidades de ocorrência, por exemplo. Além disso, a natureza dos modelos bayesianos permite que novos dados sejam incorporados de maneira direta ao modelo obtido com dados antigos, em um processo constante de atualização, de forma que nenhum dado é desperdiçado. Na abordagem bayesiana também são incorporadas nas estimativas informações que sejam conhecidas antes da coleta de dados, através da chamada densidade a priori.

Os problemas que a econometria bayesiana se propõe a resolver são o de estimação, o de comparação de modelos e o de previsão. Os métodos de estimação, como o nome indica, buscam, através das informações a priori e dos dados, chegar na distribuição posterior. Este processo pode ser feito tanto analiticamente, o que leva a densidades conhecidas para $f(\theta|y)$, ou as estimativas podem ser feitas com o uso de métodos numéricos, usualmente métodos de Monte Carlo. Já os métodos de comparação de modelos buscam auxiliar na decisão de escolha de qual modelo é o mais adequado em termos de número de variáveis, densidades utilizadas, etc. Por fim, os métodos previsão, assim como os métodos clássicos, tentam, com base em novos dados coletados das variáveis explicativas, prever qual será o valor da variável dependente. Portanto, o enfoque bayesiano permite trabalhar com estimação, comparação de modelos e previsão incorporando probabilidades aos parâmetros, uma vez que possibilita modelar a incerteza a respeito dos parâmetros em termos aleatórios (KOOP; POIRIER; TOBIAS, 2007).

Uma das dificuldades iniciais no uso de métodos bayesianos estava no cálculo dos resultados a posteriori: muitas vezes os modelos acabam não apresentando uma distribuição de probabilidade conhecida e por isso métodos numéricos de integração são necessários. Com o advento da computação e maior acesso a equipamentos, a metodologia bayesiana acabou sendo mais incorporada nas pesquisas e novos métodos foram desenvolvidos, muitos deles possíveis apenas computacionalmente, como métodos de Monte Carlo. Ao mesmo tempo, houve um aumento observável na proporção de artigos científicos utilizando métodos bayesianos em jornais da área de estatística (de cerca de 10% em 1970 para aproximadamente 30% em 2000). Este aumento foi menos intenso nos principais jornais

de economia, mas ainda assim observa-se uma progressão no uso de técnicas bayesianas dentro da econometria (POIRIER, 2006). Uma possível explicação para não ter acontecido uma mudança maior no aumento da proporção de artigos econométricos bayesianos é a falta de livros texto no assunto que sejam de fácil compreensão (POIRIER, 1989).

Dada a natureza dos dados econômicos e a complexidade dos modelos atualmente existentes, a metodologia bayesiana surge como uma alternativa à metodologia clássica de inferência estatística, possibilitando dar um tratamento diferenciado na questão de estimação de parâmetros, comparação de modelos e previsão para novas observações. Porém, ainda existe uma carência de literatura acessível a alunos de graduação, que já era observada por Poirier (1989) e perdura até os dias de hoje. Não há textos introdutórios sobre o assunto que sejam voltados para alunos de cursos de economia, que possa introduzir os métodos bayesianos de econometria de maneira mais simplificada, gerando uma barreira de entrada e a falta de disseminação do assunto. Neste sentido, este trabalho busca fazer uma apresentação inicial do modelo normal de regressão linear multivariado, sob o enfoque bayesiano, mostrando os principais resultados teóricos em comparação com a metodologia clássica e também uma aplicação em dados reais, para responder a seguinte pergunta de pesquisa: como é o comportamento das estimativas bayesianas quando comparadas às estimativas clássicas no modelo de regressão linear, tanto no aspecto analítico como em resultados simulados e empíricos?

O objetivo principal do trabalho é apresentar o modelo de regressão linear com matriz de variância e covariância simples (também chamado de Modelo Normal Clássico de Regressão Linear - MNCRL) com a abordagem bayesiana para os casos de priori conjugada Normal-Gama, apresentando todos os cálculos necessários para obter as posteriori, incluindo a densidade preditiva a posteriori. Como objetivos secundários, os métodos estudados serão comparados, utilizando dados simulados, com a estimação clássica por estimação mínimos quadrados ordinários tanto para os casos onde os dados atendem todas as condições de Gauss-Markov como para o caso onde há erro de medida nos dados. Além disso, uma aplicação baseada no trabalho de Bonini e Pozzobon (2016) será apresentada, adaptando a metodologia proposta no trabalho e comparando os resultados.

A parte teórica do texto compreende o estudo detalhado dos principais resultados já existentes. Serão apresentadas as derivações das equações da verossimilhança, priori, posteriori conjunta e marginais, seguindo a notação e desenvolvimento de Koop (2003). Na etapa de comparação de métodos, serão utilizados dados simulados para comparar as estimativas obtidas pelo estimador de mínimos quadrados ordinários (MQO) com a média da posteriori para quatro especificações diferentes, alterando as médias e variâncias a priori do parâmetro de interesse. Foram elaborados 5 experimentos distintos onde as condições necessárias para validade do Teorema de Gauss Markov são atendidas, alterando os valores do processo gerador dos dados e presença de intercepto no modelo. As estimativas

são comparadas através da média e erro quadrático médio e também são apresentados os histogramas das repetições geradas. Adicionalmente, foram feitos 5 experimentos na presença de erro de medida para comparação de desempenho dos métodos de estimação.

A análise empírica será baseada no trabalho de Bonini e Pozzobon (2016) que analisou dados da Relação Anual de Informações Sociais (RAIS) de salário, produtividade e gênero da força de trabalho dos estados da Região Sul do Brasil para o ano de 2011. As estimativas encontradas por estes autores serão utilizadas como priori para um modelo com estimativas bayesianas e os resultados serão comparados verificando se a abordagem bayesiana para dados mais recentes irá levar a conclusões similares que o estudo original teve.

Esta monografia está dividida em cinco partes e contempla, além desta introdução no capítulo 1, o referencial teórico para apresentação dos principais teoremas que já estão estabelecidos na literatura da área; simulações para comparação de desempenho de modelos e uma aplicação utilizando dados reais, onde são feitas as estimativas para um modelo similar ao de Bonini e Pozzobon (2016). A monografia é encerrada com as considerações finais, que se encontram no capítulo 5. Adicionalmente, fazem parte do trabalho dois apêndices, um com as demonstrações matemáticas dos resultados do capítulo 2 e outro com os resultados das simulações do capítulo 3, organizados em tabelas e gráficos.

2 REVISÃO BIBLIOGRÁFICA

Por volta de 1740, o Reverendo Thomas Bayes (1701-1762) desenvolveu seu trabalho intitulado *"Um Ensaio para Solucionar um Problema na Doutrina das Chances"*[1], que buscava uma solução matemática para o problema que na época era conhecido como problema da probabilidade inversa. Bayes nunca publicou seu ensaio e o mesmo foi descoberto apenas após seu falecimento por Richard Price que, após revisar as notas, o submeteu para publicação.

O ensaio de Bayes era baseado em um experimento mental desenvolvido por ele e que buscava desenvolver um método para chegar em resultados a partir de novas observações de um determinado fenômeno. Seu experimento consistia em imaginar uma mesa totalmente plana e sobre ela uma bola; o objetivo era obter um palpite adequado sobre a posição da bola estando de costas para a mesa. Assumindo que uma bola que fosse lançada na mesa teria probabilidade igual de parar em qualquer posição, outra pessoa lança bolas e informa se ela parou à esquerda ou à direita da primeira bola. Por exemplo, se todas as bolas estiverem à direita da bola inicial, é seguro afirmar que a primeira bola deve estar na extremidade esquerda da mesa. Com este raciocínio, Bayes desenvolveu um método que permitia, com base em dados atuais, fazer afirmações sobre uma condição inicial desconhecida, em um processo que poderia ser constantemente atualizado à medida que novas informações estivessem disponíveis. No caso do experimento, as novas informações seriam as posições das novas bolas lançadas.

Sem conhecer o trabalho de Bayes, Pierre Simon Laplace (1749-1827) desenvolveu uma primeira versão do que hoje é conhecido como o Teorema de Bayes. Sua descoberta surgiu a partir da necessidade de ferramentas matemáticas para seus trabalhos na área de astronomia e, no ano de 1774, através de Richard Price, Laplace incorporou as ideias de Bayes em seu método. Nos dias atuais, o teorema de Bayes está consolidado como parte da teoria das probabilidades e deu origem aos chamados métodos bayesianos de inferência, porém, ao longo da história, surgiram diversas polêmicas e debates teóricos sobre a validade de tais métodos. Para uma história a respeito do desenvolvimento dos métodos Bayesianos, recomenda-se a leitura de Mcgrayne (2015).

A respeito do uso de métodos bayesianos em econometria, Geweke (2001) e Zellner (1985) citam autores da metade do século XIX como precursores no desenvolvimento do assunto, sendo que um dos primeiros livros textos de econometria bayesiana é a obra de Zellner (1971), intitulada "*An introduction to bayesian inference in econometrics*". Daquela época até os dias atuais, com o advento da computação, métodos bayesianos se tornaram

[1] Do original em inglês *"An Essay toward Solving a Problem in the Doctrine of Chances"*.

cada vez mais acessíveis e inúmeros trabalhos foram desenvolvidos na área. O artigo de Zellner (1983) traz algumas aplicações de econometria bayesiana desenvolvidas até a década de 80.

Atualmente, existem diversos modelos macroeconométricos onde os métodos bayesianos são amplamente utilizados, como por exemplo os modelos de vetores aleatórios (VAR), desenvolvidos por Christopher Sims no início da década de 80. Doan, Litterman e Sims (1984) estimaram pela primeira vez um VAR bayesiano (BVAR) enquanto que a generalização do modelo considerando parâmetros variando no tempo (TVP-VAR), desenvolvida por Cogley e Sargent em 2001, já foi estimada utilizando o amostrador de Gibbs, um método bayesiano. Mais tarde, em 2005, os mesmos autores incluíram volatilidade estocástica multivariada no TVP-VAR e novamente estimaram o modelo utilizando o amostrador de Gibbs. De acordo com Geweke, Koop e Van Dijk (2011), os modelos dinâmicos estocásticos de equilíbrio geral (DSGE), utilizados amplamente em Bancos Centrais, são predominantemente estimados utilizando inferência bayesiana. De fato, desde 2011 o Banco Central do Brasil (BACEN) utiliza o SAMBA (modelo analítico estocástico com abordagem bayesiana[2]) para auxiliar na condução da política macroeconômica no país. Caldeira, Moura e Santos (2015) fazem uma comparação de modelos, incluindo um TVP-VAR bayesiano com uso de priori de Minnesota para previsão de dados macroeconômicos do Brasil.

Na área de finanças, modelos bayesianos também tem sido utilizados, em particular pela capacidade de lidar com modelos altamente parametrizados além da possibilidade de incorporar conhecimentos prévios de mercado. Por exemplo, Philipov e Glickman (2006) propõem um modelo com volatilidade estocástica Wishart para otimização de portfólios. Kastner, Frühwirth-Schnatter e Lopes (2017) fazem a estimação bayesiana eficiente de matrizes de covariância dinâmicas em modelos de series temporais multivariadas, aplicando o modelo para dados de taxas de câmbio.

Recomenda-se a leitura de Geweke, Koop e Van Dijk (2011) para métodos bayesianos nas áreas de microeconomia, macroeconomia, finanças e marketing, além das demais obras citadas neste texto.

2.1 CONCEITOS BÁSICOS DE INFERÊNCIA BAYESIANA

Nesta seção serão apresentados os principais conceitos que serão abordados ao longo do trabalho. Os resultados são baseados principalmente nas obras de Davidson e MacKinnon (1993), Casella e Berger (2002), Robert (2007), Koop (2003), Koop, Poirier e Tobias (2007) e Bauwens, Lubrano e Richard (2003).

[2] Tradução livre de *Stochastic Analytical Model with a Bayesian Approach.*

Existem três principais definições de probabilidade: a empírica, axiomática e a subjetiva. A probabilidade empírica (ou clássica), como o nome sugere, é relacionada a observações de fenômenos físicos e é calculada a partir do número de ocorrências de um determinado fenômeno em n repetições de um experimento. Esta definição, embora de fácil entendimento e correspondência direta com a realidade, não pode ser generalizada para espaços amostrais infinitos e pode se tornar de difícil uso para aplicações mais abstratas.

A definição axiomática foi desenvolvida pelo matemático soviético Andrey Kolmogorov (1903-1987) no século XX e define como probabilidade uma função que atende a três axiomas. Esta formalização da probabilidade permitiu que seu uso fosse incorporado em aplicações mais teóricas e é a definição mais utilizada até os dias atuais.

Por fim, a probabilidade subjetiva está relacionada com a crença associada a um fato e não está diretamente relacionada com espaços amostrais e funções de probabilidade. Essa noção é caracterizada por ser pessoal, de forma que a probabilidade, neste caso, é uma representação daquilo que a pessoa acredita. A probabilidade subjetiva pode ser atribuída a tudo que é desconhecido, no sentido de que mesmo fatos estabelecidos como o valor do PIB de um país podem ser associado a uma probabilidade, caso seu valor exato seja desconhecido. Recomenda-se os textos de Casella e Berger (2002) e Mittelhammer (2013) para os conceitos e teoremas fundamentais de probabilidade.

Mittelhammer (2013) afirma que enquanto a probabilidade se ocupa de questões dedutivas como *"dado um espaço de probabilidade, o que pode ser deduzido sobre os resultados possíveis de um experimento?"*, as questões de inferência estatística se ocupam de responder a pergunta contrária: *"dadas as características do resultado do experimento, o que pode ser dito a respeito do espaço amostral?"*. O processo de inferência estatística, de maneira geral, é fazer afirmações que possam ser generalizadas para a população, a partir de informações obtidas em amostras. Neste ponto, a inferência clássica e inferência bayesiana se diferenciam: enquanto a primeira trata parâmetros populacionais como quantidades fixas e os dados como variáveis aleatórias, a inferência bayesiana utiliza a noção de probabilidade subjetiva para tudo que não é conhecido, incluindo parâmetros.

Métodos de inferência bayesiana mesclam as definições subjetiva e axiomática de probabilidade. O enfoque bayesiano considera que os parâmetros populacionais são variáveis aleatórias e o método, baseado no princípio da verossimilhança, consiste em combinar as informações contidas na amostra com as informações a priori. A base da econometria Bayesiana é o Teorema de Bayes, apresentado na Equação (2.1), que relaciona a probabilidade condicional de um evento A dado a ocorrência de outro evento B com a probabilidade inversa, isto é, a probabilidade de B ocorrer dado que o evento A ocorreu.

Teorema 2.1. ***Teorema de Bayes*** *(Retirado de Greene (2012))*
Se A e B são eventos em Ω com $\mathbb{P}(B) \neq 0$, então:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{2.1}$$

A Equação (2.1) sugere que três elementos são essenciais ao trabalhar com econometria bayesiana: a função de verossimilhança dos dados amostrais ($\mathbb{P}(B|A)$), a distribuição a priori para os parâmetros ($\mathbb{P}(A)$) e a distribuição a posteriori ($\mathbb{P}(A|B)$), também referente aos parâmetros. Os dois primeiros são combinados, utilizando a fórmula de Bayes do teorema (2.1), para chegar na distribuição (no sentido de densidade de probabilidade) a posteriori, ou posterior.

A função de verossimilhança é a mesma que é estudada em cursos de econometria tradicionais: dada uma amostra de dados de uma população, obtêm-se a função densidade de probabilidade conjunta (f.d.p.) e ela é escrita como uma função dos parâmetros. Esta componente irá trazer ao modelo bayesiano a informação que os dados têm sobre os parâmetros, seguindo o princípio da verossimilhança, descrito em Robert (2007) e atribuído a Fisher, Barnard e Birnbaum, que diz que toda a informação sobre um parâmetro θ em uma observação X está contida na função de verossimilhança $\ell(\theta|X)$. De maneira geral, se houver um escalar α tal que, para duas observações de uma distribuição com parâmetro θ, denotadas por X_1 e X_2, tal que é possível escrever $\ell(\theta|X_1) = \alpha \cdot \ell(\theta|X_2)$ para todo θ, então X_1 e X_2 levam às mesmas conclusões no processo de inferência. Formalmente, tem-se a definição enunciada em (2.2).

Definição 2.2. Função de Verossimilhança (Adaptado de Davidson e MacKinnon (1993))
Seja $f(\boldsymbol{x}|\boldsymbol{\theta})$ a densidade conjunta de uma amostra $\boldsymbol{X} = (X_1, \ldots, X_n)$. Então, dado que a amostra $\boldsymbol{X} = \boldsymbol{x}$ foi observada, a função de $\boldsymbol{\theta}$ definida como

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) \tag{2.2}$$

é chamada de *função de verossimilhança*. Em particular, se $X_1, \ldots, X_n$ é uma amostra aleatória, então:

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{2.3}$$

A priori é a informação a respeito dos parâmetros que incorpora o conhecimento prévio aos dados, isto é, através do conhecimento do pesquisador, define-se função associada ao vetor de parâmetros que se deseja estimar. Algumas vezes esta função será uma densidade de probabilidade (f.d.p.) bem definida ou, no caso de prioris impróprias, será uma função

que não define uma densidade. O processo de determinação da priori é de fundamental importância ao método bayesiano, pois ela, junto com a verossimilhança dos dados, determina a posteriori, afetando os resultados como médias, medianas e intervalos de credibilidade a posteriori.

Em uma aplicação onde há um único parâmetro desconhecido que represente uma probabilidade (como p do modelo binomial), por exemplo, pode-se utilizar como priori uma distribuição Beta(α, β) pois o suporte da mesma varia no intervalo $[0, 1]$. É importante notar que neste caso não seria adequado o uso de alguma outra função que assumisse valores fora deste intervalo por se tratar de uma probabilidade.

A densidade a posteriori do parâmetro será proporcional ao produto da priori com a verossimilhança, utilizando o teorema de Bayes:

$$\underbrace{\mathbb{P}(\theta|y)}_{\text{posteriori}} = \frac{\overbrace{\mathbb{P}(y|\theta)}^{\text{verossimilhança}} \quad \overbrace{\mathbb{P}(\theta)}^{\text{priori}}}{\underbrace{\mathbb{P}(y)}_{\substack{\text{f.d.p. marginal} \\ \text{(constante)}}}} \tag{2.4}$$

Com $\mathbb{P}(y) = \int_\Theta \mathbb{P}(y|\theta)d\theta$, onde Θ é o espaço de eventos para a variável aleatória θ.

Uma vez que o termo do denominador em (2.4) é uma constante, ele pode ser suprimido e pode-se reescrever a posteriori como proporcional ao produto da verossimilhança pela priori:

$$\mathbb{P}(\theta|y) \propto \mathbb{P}(y|\theta)\mathbb{P}(\theta) \tag{2.5}$$

onde $\propto$ é o símbolo de proporcionalidade.

Uma aplicação da metodologia bayesiana é na inferência de parâmetros, similar ao procedimento clássico, conforme ilustrado no exemplo a seguir.

Exemplo 2.3. O espaço amostral do lançamento de uma moeda é $\Omega = \{\text{cara, coroa}\}$. Define-se uma variável aleatória $Y(\cdot)$, que é igual a 1 se o resultado foi cara, cuja função densidade de probabilidade é dada por:

$$\mathbb{P}(Y = y|\theta) = \begin{cases} \theta, \text{ se } y = 1. \\ 1 - \theta, \text{ se } y = 0, \end{cases} \tag{2.6}$$

onde θ representa o parâmetro associado à probabilidade de sair cara. A notação $\mathbb{P}(Y = y|\theta)$ denota que Y depende do parâmetro θ.

A f.d.p. de Y pode ser reescrita como:

$$\mathbb{P}(Y = y|\theta) = \theta^y(1 - \theta)^{1-y} \tag{2.7}$$

Se forem realizados n lançamentos independentes da moeda, então a função densidade de probabilidade conjunta dos n lançamentos é dada por:

$$\begin{aligned}
\mathbb{P}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n|\theta) = & \\
&= \mathbb{P}(\{Y_1 = y_1\} \cap \{Y_2 = y_2\} \cap \ldots \cap \{Y_n = y_n\}|\theta) \\
&= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n|\theta) \\
&= \mathbb{P}(Y_1 = y_1|\theta) \cdot \mathbb{P}(Y_2 = y_2|\theta) \ldots \cdot \mathbb{P}(Y_n = y_n|\theta) \\
&= \theta^{y_1}(1-\theta)^{1-y_1} \ldots \theta^{y_n}(1-\theta)^{1-y_n} \\
&= \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} \qquad (2.8)
\end{aligned}$$

De forma que se obtém:

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n|\theta) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i} \qquad (2.9)$$

Após serem observados os n lançamentos, a Equação (2.9) é uma função apenas do parâmetro desconhecido θ. Assim, $\mathbb{P}(Y_1, Y_2, \ldots, Y_n|\theta)$, quando vista como função de θ, é a função de verossimilhança, definida em (2.2). Uma diferença crucial entre a f.d.p e a verossimilhança está no fato de que a função de verossimilhança **não** é uma função densidade de probabilidade para θ, logo, sua integral ou somatório em relação a θ é diferente de 1.

Assim, depois da observação de todos os n lançamentos, pode-se usar a amostra $\underset{\sim}{\boldsymbol{y}} = \{y_i\}_{i=1}^{n}$ para estimar θ. Na abordagem clássica utilizando estimação por máxima verossimilhança, a expressão em (2.9) será maximizada em relação a θ, de forma a encontrar $\hat{\theta}_{MV}$ que maximiza a chance de ocorrência de $\underset{\sim}{\boldsymbol{y}}$. Tomando o logaritmo de (2.9), obtêm-se:

$$l(\theta|\boldsymbol{y}) = log(\mathbb{P}(\boldsymbol{y}|\theta)) = \left(\sum_{i=1}^{n} y_i\right) log(\theta) + \left(n - \sum_{i=1}^{n} y_i\right) log(1-\theta) \qquad (2.10)$$

Tomando a derivada de (2.10) em relação a θ e igualando a 0 é obtida a seguinte expressão[3]:

$$\begin{aligned}
\left.\frac{dl(\theta|\boldsymbol{y})}{d\theta}\right|_{\hat{\theta}_{MV}} &= \frac{\sum y_i}{\hat{\theta}_{MV}} - \frac{n - \sum y_i}{1 - \hat{\theta}_{ML}} = 0 \\
\Rightarrow \sum y_i(1 - \hat{\theta}_{MV}) &= (n - \sum y_i)\hat{\theta}_{MV} \qquad (2.11)
\end{aligned}$$

[3] Deste ponto em diante serão omitidos os limites no somatório até o final do exemplo. Todas as somas, a menos que indicado o contrário, vão de i igual a 1 até n.

De forma que o valor de θ que maximiza a função de verossimilhança é dado por:

$$\hat{\theta}_{MV} = \frac{\sum y_i}{n} \tag{2.12}$$

A Equação (2.12) diz que o valor de $\hat{\theta}$ que maximiza a verossimilhança é a média amostral. Intuitivamente, isso é adequado, uma vez que se forem observadas 5 caras em 10 lançamentos, um palpite empírico para a probabilidade de sair cara seria justamente 1/2.

Para a inferência bayesiana, é necessária uma distribuição a priori para θ, que caracteriza o conhecimento prévio que a pessoa conduzindo o estudo tem a respeito desse valor desconhecido. Como já mencionado anteriormente neste texto, pode-se usar $\theta \sim \text{Beta}(\alpha, \beta)$, pois neste caso, $0 \leq \theta \leq 1$. A distribuição Beta tem dois parâmetros, $\underline{\alpha}$ e $\underline{\beta}$, ambos positivos e sua densidade é dada por:

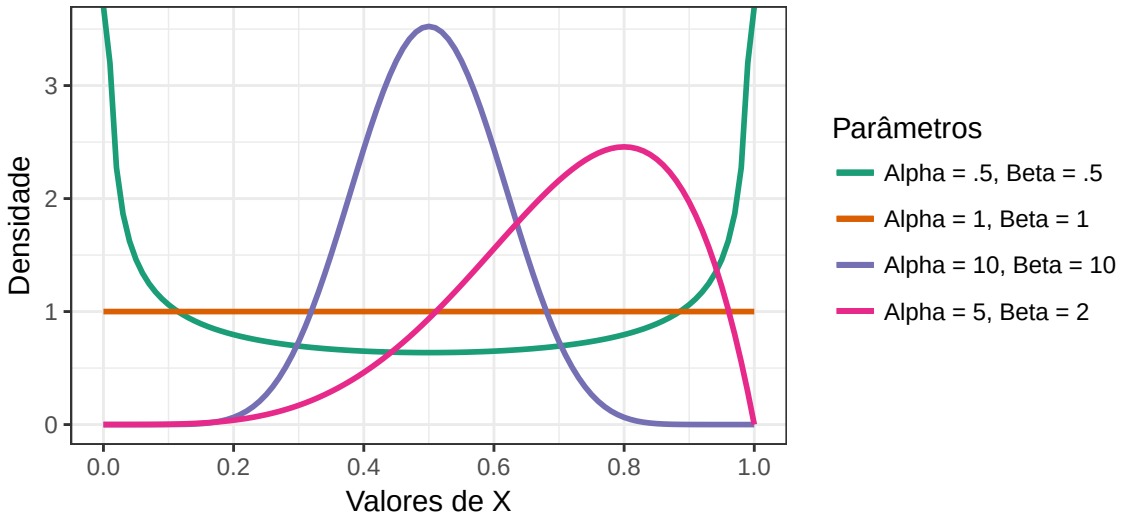
$$\mathbb{P}(\theta|\underline{\alpha}, \underline{\beta}) = \frac{\Gamma(\underline{\alpha} + \underline{\beta})}{\Gamma(\underline{\alpha})\Gamma(\underline{\beta})}\theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\beta}-1} \tag{2.13}$$

onde $\Gamma(n) = (n-1)!$ e $0 \leq \theta \leq 1$. Os valores $\underline{\alpha}$ e $\underline{\beta}$ são chamados de *hiperparâmetros*. A denominação de *hiperparâmetros* é utilizada para diferenciar os parâmetros da distribuição à priori, que são conhecidos e pré-determinados, daqueles parâmetros que são desconhecidos cujas características estão sendo investigadas. Neste trabalho, será utilizada a notação de uma barra inferior para parâmetros da priori (os hiperparâmetros) e barra superior para parâmetros da posteriori.

A Figura (1) tem quatro representações diferentes para a distribuição $Beta(\alpha, \beta)$, onde é possível notar o efeito da variação dos parâmetros. Sempre que $\alpha = \beta$, a forma da densidade é simétrica e a média da variável aleatória será o valor 0.5. Isso decorre do fato de que se $\theta \sim \text{Beta}(\alpha, \beta)$, então $\mathbb{E}[\theta] = \alpha/(\alpha + \beta)$. Valores maiores de α em relação a β irão resultar em uma curva com cauda à esquerda e o contrário (quando $\beta > \alpha$) resulta em uma distribuição com cauda para a direita. Um caso particular da distribuição está representada pela linha alaranjada, onde observa-se que $\text{Beta}(1, 1)$ é igual a uma densidade $\text{Uniforme}(0, 1)$.

Outra vantagem da escolha de uma priori Beta, neste exemplo, é que ela é uma priori *conjugada* ao modelo Binomial. Um modelo é dito conjugado quando dada uma priori e uma verossimilhança a posteriori sempre pertence à mesma família de distribuições que a priori (KOOP, 2003). No exemplo apresentado, tem-se que a verossimilhança Binomial da Equação (2.9) quando combinada com uma priori Beta, como a da Equação (2.13), tem como resultado uma posteriori também de distribuição Beta. Além de possibilitar o uso de fórmulas analíticas fechadas para o processo de inferência, o uso de modelos conjugados facilita o mecanismo de atualização da posteriori quando houverem mais dados disponíveis.

Figura 1 – Diferentes formas da distribuição $Beta(\alpha, \beta)$



Fonte – Elaboração própria com base em valores simulados.

Utilizando (2.4), (2.9) e (2.13) tem-se:

$$\mathbb{P}(\theta|\boldsymbol{y}) = \frac{\overbrace{\theta^{\sum y_i}(1-\theta)^{n-\sum y_i}}^{\text{eq.(2.9)}}\overbrace{\theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\beta}-1}\frac{\Gamma(\underline{\alpha}+\underline{\beta})}{\Gamma(\underline{\alpha})\Gamma(\underline{\beta})}}^{\text{eq. (2.13)}}}{\int_0^1 \theta^{\sum y_i}(1-\theta)^{n-\sum y_i}\theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\beta}-1}\frac{\Gamma(\underline{\alpha}+\underline{\beta})}{\Gamma(\underline{\alpha})\Gamma(\underline{\beta})}d\theta} = \frac{\overbrace{\mathbb{P}(Y|\theta)\mathbb{P}(\theta)}^{\text{eq (2.4)}}}{\int_0^1 \mathbb{P}(Y|\theta)\mathbb{P}(\theta)d\theta} \tag{2.14}$$

De forma que:

$$\mathbb{P}(\theta|y) = \frac{\theta^{\overbrace{(\underline{\alpha}+\sum y_i)}^{\bar{\alpha}}-1}(1-\theta)^{\overbrace{(\underline{\beta}+n-\sum y_i)}^{\bar{\beta}}-1}}{\int_0^1 \theta^{(\underline{\alpha}+\sum y_i)-1}(1-\theta)^{(\underline{\beta}+n-\sum y_i)-1}d\theta} \tag{2.15}$$

A expressão em (2.15) é a f.d.p. de uma variável aleatória que segue uma distribuição Beta com parâmetros $\bar{\alpha} = (\underline{\alpha} + \sum y_i)$ e $\bar{\beta} = (\underline{\beta} + n - \sum y_i)$. A constante de integração é obtida através da integral que encontra-se no denominador.

A média da posteriori pode ser calculada utilizando:

$$\begin{aligned}
\mathbb{E}[\theta|y] &= \int_0^1 \theta\mathbb{P}(\theta|y)d\theta \\
&= \frac{\bar{\alpha}}{\bar{\alpha}+\bar{\beta}} \\
&= \frac{\overbrace{\underline{\alpha}+\sum y_i}^{\bar{\alpha}}}{\underbrace{\underline{\alpha}+\sum y_i}_{\bar{\alpha}}+\underbrace{\underline{\beta}+n-\sum y_i}_{\bar{\beta}}}
\end{aligned}$$

$$= \frac{\underline{\alpha} + \sum y_i}{\underline{\alpha} + \underline{\beta} + n} = \frac{\underline{\alpha}}{\underline{\alpha} + \underline{\beta} + n} + \frac{\sum y_i}{\underline{\alpha} + \underline{\beta} + n} \tag{2.16}$$

Que pode ser reescrita como:

$$\mathbb{E}[\theta|y] = \underbrace{\frac{\underline{\alpha}}{\underline{\alpha} + \underline{\beta}}}_{\mathbb{E}[\theta]} \left[\frac{\underline{\alpha} + \underline{\beta}}{\underline{\alpha} + \underline{\beta} + n}\right] + \underbrace{\frac{\sum y_i}{n}}_{\hat{\theta}_{MV}} \left[\frac{n}{\underline{\alpha} + \underline{\beta} + n}\right] \tag{2.17}$$

É importante observar que da forma como $\mathbb{E}[\theta|y]$ está escrita na Equação (2.1), a média de θ, após observar os dados, nada mais é que uma média ponderada entre a média da priori, $\mathbb{E}[\theta]$, e o estimador de máxima verossimilhança, $\hat{\theta}_{ML}$. Os pesos desta média são $\underline{\alpha} + \underline{\beta}$ para a priori e o tamanho amostral n para a componente dos dados. Se o tamanho amostral for muito superior à soma dos hiperparâmetros, então a média da posteriori será praticamente igual ao estimador de máxima verossimilhança, pois $\lim_{n\to\infty}[\mathbb{E}(\theta|y)] = \hat{\theta}_{ML}$.

Agora, supondo que seja observada a seguinte sequência: $\underset{\sim}{Y} = \{1, 1, 1\}$, então, utilizando o estimador de máxima verossimilhança, o seguinte resultado é obtido:

$$\hat{\theta}_{MV} = \frac{\sum y_i}{n} = \frac{1+1+1}{3} = \frac{3}{3} = 1 \tag{2.18}$$

O resultado acima evidencia que a probabilidade de ocorrência de coroa estimada por máxima verossimilhança é 0. Há, então, um *excesso de ajuste* do EMV: ele tenta escolher os parâmetros de forma a maximizar o evento que foi encontrado nos dados, logo, quando algo não acontece, sua probabilidade será zero e isto está refletido nos dados. No exemplo das moedas isso fica claro: se forem observadas 3 caras, o valor de θ que maximiza que essa observação tenha vindo de uma moeda é igual a 1, sendo que o evento coroa será considerado como tendo probabilidade zero.

Sob o enfoque bayesiano, usando uma priori Beta$(1, 1)$ = Uniforme$(0, 1)$, tem-se que a estimativa bayesiana para a probabilidade de ocorrência de coroa é 0.2. Isso decorre do cálculo da média à posteriori (Equação 2.19), utilizando as propriedades da distribuição Beta.

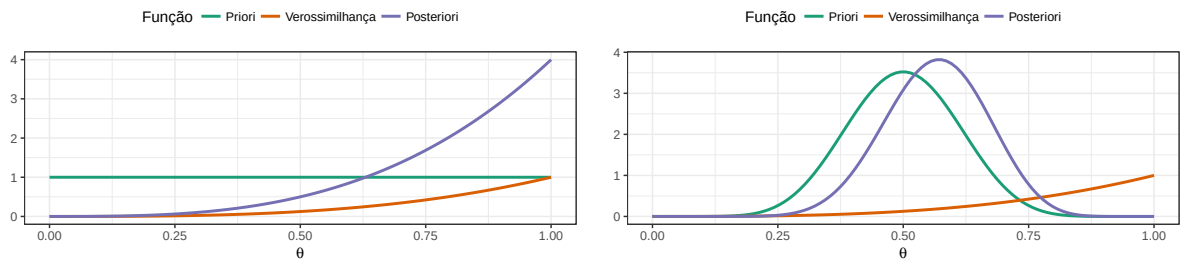
$$\begin{aligned} \bar{\alpha} = \underline{\alpha} + \sum y_i = 1 + 3 = 4, \qquad & \bar{\beta} = \underline{\beta} + n - \sum y_i = 1 + 3 - 3 = 1 \\ \mathbb{E}[\theta|y] = \frac{\bar{\alpha}}{\bar{\alpha} + \bar{\beta}} \qquad & = \tfrac{4}{4+1} = \tfrac{4}{5} = 0.8 \end{aligned} \tag{2.19}$$

O exemplo (2.3) reforça que se um evento nunca é observado, o estimador de máxima verossimilhança atribui probabilidade 0 para a sua ocorrência. Já para a estimativa Bayesiana, por mais que não tenha sido observada nenhuma coroa, considera-se que não houveram repetições suficientes para que a estimativa de sua probabilidade seja considerada igual a zero. Além disso, a metodologia bayesiana permite incorporar crenças à priori.

No exemplo (2.3), utilizou-se uma priori dita não informativa para o cálculo da média a posteriori, isto é, ao assumir que $\theta \sim \text{Beta}(1, 1)$, está se supondo que qualquer valor de θ entre 0 e 1 é igualmente provável. Neste caso, a informação a priori terá "pouco peso", quando confrontada com a verossimilhança dos dados, na determinação da posteriori. Caso houvesse alguma crença anterior que sugerisse algo diferente, como por exemplo, que a moeda é não viesada ($p(\text{cara}) = p(\text{coroa}) = 1/2$), poderia ter sido utilizada uma priori como $\text{Beta}(10, 10)$, que concentra valores em torno de $1/2$.

Uma ferramenta útil para comparar priori, posteriori e verossimilhança dos dados é o *triplot*, que é um gráfico que contém as três funções plotadas no mesmo plano[4]. A Figura (2) contém os triplots para duas situações: na primeira (Figura 2a) são reproduzidos os valores do exemplo (2.3) onde a priori é uma distribuição Beta com parâmetros $\underline{\alpha} = \underline{\beta} = 1$, os valores para a verossimilhança Binomial são $n = \sum y = 3$ e a posteriori é uma densidade Beta com parâmetros $\bar{\alpha} = 4$ e $\bar{\beta} = 1$. Na Figura (2b) usa-se a mesma configuração da verossimilhança, porém a priori utilizada é mais informativa, com valores $\underline{\alpha} = \underline{\beta} = 10$ (caracterizando o que seria uma confiança maior de que a moeda é não viciada), o que produz uma posteriori Beta com parâmetros $\bar{\alpha} = 23$ e $\bar{\beta} = 10$.

Figura 2 – Triplots para duas prioris diferentes no exemplo (2.3)



(a) Priori Beta(1,1) e posteriori Beta(4,1) (b) Priori Beta(10,10) e posteriori Beta(23,10)

Fonte – Elaboração própria com base em valores simulados.

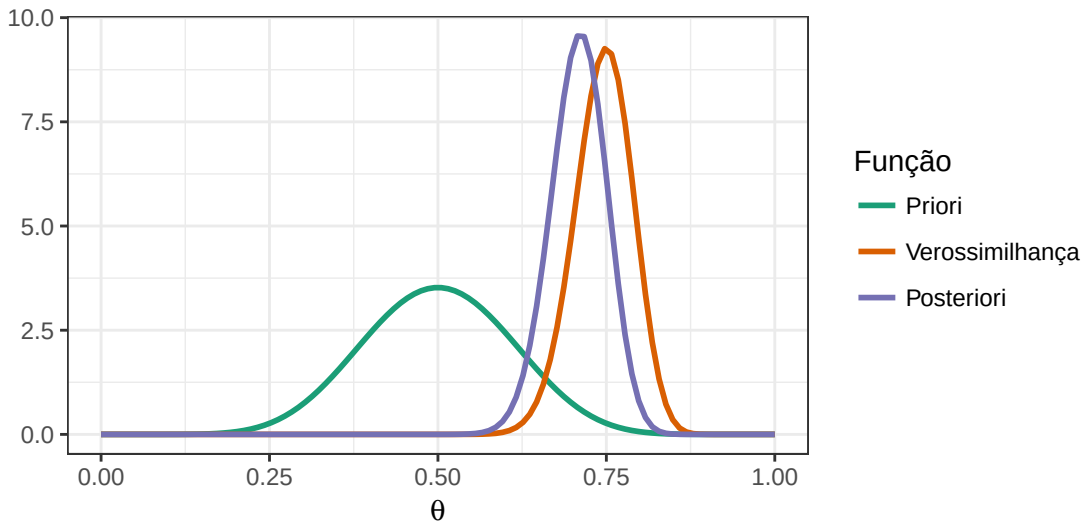
É possível observar que na Figura (2a), utilizando uma priori não informativa, o formato da posteriori é o mesmo da verossimilhança. Embora a distribuição Beta(1, 1) tenha valor esperado igual a $1/2$, o fato desta densidade ser uniforme para todos valores entre 0 e 1 implica que a crença à priori na moeda ser não viesada é "fraca". Com isso, os dados da verossimilhança irão ter um peso maior na posteriori. Por outro lado, ao utilizar uma priori mais informativa, cuja média ainda é $1/2$ mas os valores estão mais concentrados em torno desta média, a crença à priori passa a ter mais "peso" na determinação da posteriori, o que é acentuado pelo baixo tamanho amostral. De fato, a

[4] Uma vez que a função de verossimilhança não é uma densidade de probabilidade (e portanto não integra 1), é comum multiplicá-la por uma constante de maneira que fique na mesma escala que a priori e a posteriori no triplot. Este procedimento não afeta a análise dos resultados, uma vez que as estimativas obtidas não são afetadas por transformações monotônicas da verossimilhança, pela propriedade de invariância do EMV (CASELLA; BERGER, 2002).

média da distribuição Beta(23, 10) é aproximadamente 0.56, um valor bastante próximo da média a priori.

À medida que o tamanho amostral aumenta, a verossimilhança começa a afetar mais a posteriori. Na Figura (3) está o triplot para a situação onde a priori é uma distribuição Beta(10, 10) e a verossimilhança é a de uma distribuição binomial com parâmetros $n = 100$ e $p = 0.75$ (isto é, $\sum y_i = 75$). A densidade a posteriori agora terá parâmetros $\bar{\alpha} = 85$ e $\bar{\beta} = 35$ e seu valor esperado é igual a $^{85}/_{120} \approx 0.71$, bastante próximo do valor esperado dos dados.

Figura 3 – Triplot para priori Beta(10,10) e verossimilhança Binomial com $n = 100$ e $\sum y = 75$.



Fonte – Elaboração própria com base em valores simulados.

2.2 ECONOMETRIA BAYESIANA

A ideia da econometria é compreender fenômenos econômicos observáveis, tais como ciclos econômicos, comportamento dos agentes, impacto de políticas monetárias, etc., utilizando, para tal, técnicas estatísticas. Embora as aplicações econométricas mais visíveis para o público em geral sejam na área de macroeconomia, onde variáveis como taxas de inflação, taxas de juros e PIB são previstas, existe uma infinidade de outros problemas não relacionados à previsões macro onde o uso da econometria é empregado: avaliação de impacto de políticas públicas, efeitos da educação em salários dos trabalhadores, entre outros. A econometria se tornou uma área separada da estatística matemática pois os dados econômicos, em sua maioria, são *não-experimentais* e por isso métodos específicos para lidar com a natureza destes problemas precisaram ser desenvolvidos (WOOLDRIDGE, 2015).

Usualmente, para verificar os relacionamentos entre variáveis e testar teorias econômicas, formula-se um modelo econômico, que é uma Equação matemática descrevendo as relações de interesse. Somente após a especificação deste modelo econômico o modelo econométrico é definido. A diferença entre os dois tipos de modelos é que o primeiro é uma equação exata, determinística, enquanto o modelo econométrico permite que fatores não observáveis (e considerados estocásticos) sejam incluídos na modelagem.

Modelos de regressão são o cerne da econometria. Por exemplo, no modelo linear de regressão, é possível modelar o relacionamento de uma variável dependente y com k variáveis explicativas, x_1, x_2, ..., x_k através da seguinte forma funcional:

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x_1} + \beta_2 \boldsymbol{x_2} + \ldots + \beta_k \boldsymbol{x_k} + \boldsymbol{\varepsilon} \tag{2.20}$$

Onde:

1. $\boldsymbol{y}$ é um vetor com N entradas da variável dependente, isto é, para cada unidade amostral é observado um valor de y, em que N é o tamanho total da amostra;

2. $\boldsymbol{x_i},\ i \in \{1, \ldots, k\}$ são vetores $N \times 1$ de variáveis independentes, também observáveis;

3. $\beta_i,\ i \in \{1, \ldots, k\}$ são os parâmetros do modelo, que deseja-se estimar;

4. $\boldsymbol{\varepsilon}$ é um vetor $N \times 1$ de erros não observáveis.

A solução clássica de modelos como o descrito na Equação (2.20) envolve encontrar estimativas $\hat{\beta}_1, \ldots, \hat{\beta}_k$ para os parâmetros populacionais $\beta_1, \ldots, \beta_k$, de forma que novas observações das variáveis independentes (que usualmente serão de mais fácil obtenção do que a variável dependente) possam ser utilizadas para uma aproximação de quais seriam os valores de y. Além disso, os coeficientes $\hat{\beta}_1, \ldots, \hat{\beta}_k$ estimados fornecem uma interpretação direta sobre a magnitude e a direção dos efeitos que os regressores x_i têm sobre y.

Empiricamente, utilizando uma forma reduzida de (2.20), poderia-se montar um modelo para explicar o consumo das famílias y_t em função da renda disponível (x_t), para cada ano no universo de n anos. Neste caso, o modelo econômico adjacente ao modelo econométrico seria da função consumo onde o parâmetro multiplicando a renda disponível é a propensão marginal a consumir e o parâmetro de intercepto representa o consumo autônomo (DAVIDSON; MACKINNON, 2004). Existem diferentes formas para estimação neste tipo de modelo, incluindo os métodos como estimação por máxima verossimilhança, estimador de mínimos quadrados, estimador pelo método dos momentos, entre outros, que, por serem métodos clássicos, tratam os parâmetros como quantidades fixas, porém desconhecidas e podem ser encontrados, por exemplo, em Wooldridge (2015) e Davidson e MacKinnon (2004). As próximas subseções, onde serão apresentados dois métodos bayesianos para estimação de (2.20), são baseadas no texto de Koop (2003).

2.3 O MODELO LINEAR NORMAL DE REGRESSÃO CLÁSSICO COM PRIORI NORMAL-GAMA CONJUGADA

Nesta seção são apresentados os resultados obtidos para o MLRNC utilizando a priori conjugada normal-gama. Embora os resultados analíticos sejam conhecidos e possam ser vistos, por exemplo, nas obras de Koop (2003), Koop, Poirier e Tobias (2007) e Bauwens, Lubrano e Richard (2003), não há uma fonte onde todos os passos dos cálculos estejam explicitados. Dessa maneira, as derivações que constam no apêndice (A) foram elaboradas pela autora para este tcc, com base nas obras citadas.

O modelo (2.20) pode ser escrito matricialmente como:

$$Y = X\beta + \varepsilon \tag{2.21}$$

Onde:

- Y é um vetor de tamanho $N \times 1$ da variável explicada (endógena ou dependente);
- X é uma matriz $N \times k$ cuja primeira coluna é toda formada de números 1 e as demais colunas representam entradas de variáveis explicativas (exógenas ou independentes);
- β é um vetor $k \times 1$ de parâmetros;
- ε é um vetor $N \times 1$ de erros.

De acordo com Koop (2003), as hipóteses usuais para este modelo são:

1. Os erros tem distribuição normal multivariada com média zero e matriz de variância e covariância homocedástica, isso é, $\varepsilon \sim N(0_N, \mathbb{I}h^{-1})$, onde $\mathbb{I}$ é a matriz identidade $n \times n$, 0_N denota que a média é um vetor de zeros com N entradas e $h^{-1} = \sigma^2$. O parâmetro h é chamado de *precisão* e representa o inverso da variância;
2. Os elementos de X são fixos ou, caso sejam variáveis aleatórias, são independentes de ε com f.d.p. $p(X|\lambda)$, onde λ é um vetor de parâmetros que não depende de β ou h.

2.3.1 A função de verossimilhança

A função de verossimilhança será determinada pela função densidade de probabilidade conjunta de Y e X condicional aos parâmetros desconhecidos, ou seja $p(Y, X|\beta, h, \lambda)$.

Como os X são independentes dos erros ε, pode-se reescrever a verossimilhança da seguinte forma:

$$p(Y, X|\beta, h, \lambda) = \underbrace{p(Y|X, \beta, h)}_{\substack{\text{só esta parte é} \\ \text{relevante para} \\ \text{inferência a respeito} \\ \text{de } \beta \text{ e } h}} \underbrace{p(X|\lambda)}_{\substack{\text{independente} \\ \text{de } \beta \text{ e } \lambda}} \tag{2.22}$$

Uma vez que $p(X|\lambda)$ não depende nem de β nem de h, então, para fins de cálculo da posteriori, será apenas uma constante que poderá ser omitida. Por simplificação de notação, será suprimido também o X de $p(Y|X, \beta, h)$.

Utilizando as suposições a respeito dos erros, pode-se derivar a forma analítica da verossimilhança. Os seguintes resultados podem ser encontrados em Koop (2003) e serão utilizados ao longo deste trabalho:

1. $p(Y|\beta, h)$ segue uma distribuição normal;
2. $\mathbb{E}[Y|\beta, h] = X\beta$;
3. $Var[Y|\beta, h] = h^{-1}I$.

Utilizando os resultados acima com a definição da distribuição normal multivariada[5], chega-se na função de verossimilhança:

$$p(y \mid \beta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{h}{2}(Y - X\beta)'(Y - X\beta)\right] \tag{2.24}$$

É interessante reescrever a função de verossimilhança de maneira que em sua equação apareçam as quantidades de mínimos quadrados ordinários (MQO), o que, ao final do processo, possibilitará uma interpretação analítica mais clara da posteriori. As seguintes quantidades de MQO serão utilizadas para reescrever (2.24):

$$\nu = N - k \quad \text{(graus de liberdade)} \tag{2.25}$$

$$\hat{\beta} = (X'X)^{-1}(X'Y) \tag{2.26}$$

[5] Um vetor Y de entradas contínuas e dimensão $k \times 1$ segue uma distribuição normal multivariada com média μ (vetor com k entradas) e matriz de covariância Ω, de tamanho $k \times k$ positiva definida, se sua f.d.p. é dada por:

$$f_N(y|\mu, \Omega) = \frac{1}{2\pi^{\frac{k}{2}}}|\Omega|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(y - \mu)'\Omega^{-1}(y - \mu)\right] \tag{2.23}$$

$$s^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\nu} = \frac{SQR}{\nu} \quad \text{(estimador não viesado de } \sigma^2\text{)} \tag{2.27}$$

Utilizando (2.25), (2.26) e (2.27), a verossimilhança descrita em (2.24) pode ser escrita como produto do núcleo[6] de uma densidade Normal para $\beta|h$ vezes o núcleo de uma densidade Gama para h (a demonstração está no apêndice A):

$$p(Y|\beta, h) = \underbrace{\left[\frac{1}{(2\pi)^{\frac{N}{2}}} h^{\frac{k}{2}} \exp\left[-\frac{h}{2}(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right]\right]}_{\text{Núcleo de uma Normal para } \beta|h} \underbrace{\left[h^{\frac{\nu}{2}} \exp\left[-\frac{h}{2}\frac{\nu}{s^{-2}}\right]\right]}_{\text{Núcleo de uma Gama para } h} \tag{2.28}$$

2.3.2 A densidade a priori

A priori deve ser estabelecida de forma a representar a crença que se tem a respeito dos parâmetros *antes* (ou *a priori*) de observar os dados. Em geral, prioris podem assumir qualquer forma funcional, podendo ser, inclusive, constantes. A priori escolhida, usualmente, além de refletir a crença sobre o que se deseja inferir, terá também como característica desejável a facilidade de incorporação ao modelo, simplificando os cálculos e idealmente levando a soluções em fórmula analítica fechada[7].

Quando a priori escolhida não é uma constante e sim uma densidade de probabilidade, seus hiperparâmetros serão auxiliares para a "calibragem" da priori. No exemplo (2.3) foi utilizada como priori uma distribuição Beta$(1, 1)$. Apesar desta densidade ter média 0.5, que seria indicativa de uma moeda honesta, a "certeza" atribuída a esta média é baixa: como exemplificado na Figura (1), este é um caso particular da distribuição uniforme e por isso qualquer valor entre 0 e 1 é igualmente provável. Em geral, variâncias maiores na distribuição a priori estão associadas com um grau de incerteza maior e vice-versa.

Uma priori é dita ser *conjugada natural* quando, além de produzir uma posteriori de mesma distribuição (como no caso da priori conjugada), a priori também tem a mesma distribuição que a verossimilhança. Koop (2003) destaca que o uso de uma conjugada natural tem a vantagem adicional de que a informação à priori pode ser interpretada de maneira similar à da verossimilhança, isto é, a priori pode ser vista como proveniente de um conjunto de dados fictícios gerados por um mesmo processo gerador dos dados.

O modelo de regressão descrito em (2.21) tem dois parâmetros desconhecidos, β e h, e portanto a priori pode ser denotada como uma densidade conjunta $p(\beta, h)$. Utilizando a definição de densidade condicional, a priori pode ser escrita como $p(\beta, h) = p(\beta|h)p(h)$. Esta notação permite pensar em uma priori para β condicional a h e outra para a precisão

[6] Informalmente, núcleo de uma distribuição e/ou de uma verossimilhança é tudo que resta da função após remover as contantes. Para uma definição formal, recomenda-se a leitura da seção 2.3 de Bauwens, Lubrano e Richard (2003).

[7] Em aplicações onde não é possível chegar em uma posteriori com fórmula analítica fechada, métodos de Monte Carlo são utilizados para inferência a posteriori.

h. A Equação (2.28) sugere que a f.d.p a priori conjugada natural é o produto de uma normal condicional para $\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V})$ e uma gama para $h \sim G(\underline{\nu}, \underline{s}^{-2})$, que resulta em uma distribuição Normal-Gama:

$$p(\beta, h) \propto h^{\frac{k+\underline{\nu}-2}{2}} \exp\left[-\frac{h}{2}(\underline{\nu}\underline{s}^2 + (\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta}))\right] \tag{2.29}$$

O termo $\underline{\nu}\underline{s}^2$ de (2.29) é uma soma de quadrados dos resíduos a priori e a constante de integração foi omitida, e por isso foi utilizado o sinal de proporcionalidade. Neste caso, é utilizada a notação $p(\beta, h) \sim \mathcal{NG}(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu})$, de forma que os hiperparâmetros $\underline{\beta}, \underline{V}, \underline{s}^{-2}$ e $\underline{\nu}$ são os que devem ser escolhidos pela pessoa que está conduzindo a análise.

2.3.3 A densidade a posteriori

A f.d.p. posterior é proporcional à função de verossimilhança (2.28) multiplicada pela f.d.p. a priori (2.29). Enquanto a notação da priori era apenas $p(\beta, h)$ para mostrar a não influência dos dados, para a posteriori usa-se a notação $p(\beta, h|y)$ para deixar claro que é uma densidade após a inclusão da informação da verossimilhança. Combinando a verossimilhança e a priori definida na seção anterior, tem-se que a posteriori do modelo será dada por (a demonstração está no apêndice A):

$$p(\beta, h|y) \propto h^{\frac{\bar{\nu}+k-2}{2}} \exp\left[-\frac{h}{2}\left(\bar{\nu}\bar{s}^2 + (\beta - \bar{\beta})'\bar{V}^{-1}(\beta - \bar{\beta})\right)\right] \tag{2.30}$$

O lado direito da Equação 2.30 é o núcleo de uma densidade Normal-Gama com parâmetros:

- $\bar{V} = (\underline{V}^{-1} + X'X)^{-1}$

- $\bar{\beta} = \bar{V}(\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta})$. Isso implica que o parâmetro $\bar{\beta}$ da posteriori é uma média ponderada do parâmetro da priori $\underline{\beta}$ com o estimador de MQO $\hat{\beta}$. Se a precisão $\underline{V}^{-1}$ for baixa (ou seja, a variância à priori for alta), a contribuição de $\underline{\beta}$ para a posteriori será pequeno.

- $\bar{\nu} = \underline{\nu} + N$, ou seja, os graus de liberdade da priori são adicionados ao tamanho amostral para fornecer os graus de liberdade da posteriori.

- $\bar{\nu}\bar{s}^2 = \underline{\nu}\underline{s}^2 + \nu s^2 + (\hat{\beta} - \underline{\beta})'[\underline{V} + (X'X)^{-1}]^{-1}(\hat{\beta} - \underline{\beta})$, onde $\underline{\nu}\underline{s}^2$ é a SQR da priori, νs^2 é a SQR da verossimilhança e o último termo penaliza quanto maior a diferença entre o valor de $\underline{\beta}$ e o estimador de MQO, $\hat{\beta}$.

A expressão final para a densidade a posteriori de β e h é dada por:

$$p(\beta, h|y) = \underbrace{\frac{|\bar{V}|^{-\frac{1}{2}}}{\left(\frac{2\bar{s}^{-2}}{\bar{\nu}}\right)^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{\bar{\nu}}{2}\right)(2\pi)^{\frac{k}{2}}}}_{\text{Constante}} \underbrace{h^{\frac{\bar{\nu}+k-2}{2}} \exp\left[-\frac{h}{2}\left(\bar{\nu}\bar{s}^2 + (\beta-\bar{\beta})'V^{-1}(\beta-\bar{\beta})\right)\right]}_{\text{Equação 2.30}} \quad (2.31)$$

Uma vez que foi utilizada a priori conjugada do modelo, é fácil fazer atualização da posteriori após a coleta de novas informações. Sabendo-se que a distribuição a posterior é uma normal-gama, posteriori obtida pode ser utilizada como uma nova priori e é feita a atualização dos seus parâmetros com as informações provenientes da verossimilhança dos novos dados.

A Equação (2.31) é uma f.d.p. conjunta para β e h, porém usualmente é de interesse apenas fazer inferências para β. O termo h, neste caso, é chamado *de parâmetro incômodo* (BAUWENS; LUBRANO; RICHARD, 2003). Da definição da distribuição normal-gama, sabe-se que a marginal para β é dada por uma distribuição t multivariada, isso é, $\beta \sim t(\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})$ (KOOP, 2003). A verificação pode ser feita calculando a integral de (2.31) em relação a h e está desenvolvida no apêndice A.

Como a densidade marginal a posteriori possui infinitos valores, é comum que a média a posteriori seja utilizada como estimativa pontual para β:

$$\mathbb{E}[\beta|h] = \int\int \beta \underbrace{\overbrace{p(\beta, h|y)}^{\text{f.d.p. conjunta de } \beta \text{ e } h} dh}_{\text{marginal posterior de } \beta} d\beta = \int \beta \left[\int p(\beta, h|y) dh\right] d\beta = \int \beta p(\beta|y) d\beta \quad (2.32)$$

Usando as propriedades da distribuição t multivariada, a esperança dada em (2.32) será $\mathbb{E}[\beta|Y] = \bar{\beta}$ e $Var[\beta|Y] = \frac{\bar{\nu}s^2}{\bar{\nu}-2}\bar{V}$. Isto significa que, diferente das estimativas clássicas, não é necessária a estimativa da variância h^{-1} para saber informações do parâmetro β.

2.3.4 A densidade preditiva a posteriori

Supondo que deseja-se fazer previsão para N novas observações da variável dependente (onde não necessariamente N é o tamanho amostral utilizado na verossimilhança original), que são geradas de acordo com o seguinte modelo:

$$y^* = X^*\beta + \varepsilon^*$$

Onde ε^* é independente de ε e segue uma distribuição normal com média zero e matriz de covariância $h^{-1}\mathbb{I}_N$ e X^* é uma matriz $N \times k$, que contém as k variáveis explicativas para cada uma das N novas observações. Como ε^* é independente de ε, então

y^* é independente de y, de maneira que o cálculo da densidade preditiva a posteriori será:

$$p(y^*|y) = \int\int \underbrace{p(y^*|y,\beta,h)}_{=p(y^*|\beta,h)} \cdot \underbrace{p(\beta,h|y)}_{\text{posteriori}} d\beta dh = \int\int p(y^*|\beta,h)p(\beta,h|y)d\beta dh \tag{2.33}$$

O termo $p(y^*|\beta,h)$ é a verossimilhança para N novos dados e $p(\beta,h|y)$ é a posteriori calculada nas seções anteriores. A densidade $y^*|y$ segue uma distribuição t multivariada[8] de parâmetros $X^*\overline{\beta}$, $\bar{s}^2\left\{\mathbb{I}_N + X^*\overline{V}X^{*\prime}\right\}$ e $\bar{\nu}$. A demonstração está no apêndice A.

Além da inferência para os parâmetros e previsão para novas observações, existe ainda o problema de comparação entre modelos, que não será objeto de estudo deste trabalho. Recomenda-se a leitura do capítulo 3 de Koop (2003) e capítulo 2 de Bauwens, Lubrano e Richard (2003) para os resultados de comparação de modelos e teste de restrições no modelo linear normal com priori conjugada natural. Ferramentas de apresentação dos resultados a posteriori, como intervalos de credibilidade, serão apresentados nos próximos capítulos.

[8] Um vetor de dimensão K, $Y = (Y_1, Y_2, \ldots, Y_K)'$ segue uma distribuição t multivariada com parâmetros μ (vetor de tamanho K), Ω (matriz $K \times K$ positiva definida), e ν (um escalar positivo, denominado *graus de liberdade*), denotada por $Y \sim t(\mu, \Omega, \nu)$, se sua f.d.p. é dada por:

$$f_t(y|\mu,\Omega,\nu) = \frac{1}{c_t}|\Omega|^{-\frac{1}{2}}\left[\nu + (y-\mu)'\Omega^{-1}(y-\mu)\right]^{-\frac{\nu+K}{2}},$$

onde $c_t = \frac{\pi^K 2\Gamma\left(\frac{\nu}{2}\right)}{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+K}{2}\right)}$. Para o caso onde $K = 1$, tem-se a distribuição t de Student. Sua média e variância são dadas, respectivamente, por $\mathbb{E}[Y] = \mu$ e $Var[Y] = \frac{\nu}{\nu-2}\Omega$, contanto que $\nu > 2$. No caso onde $\nu = 1$ a distribuição t é igual à distribuição de Cauchy, que não tem seus momentos bem definidos.

3 SIMULAÇÃO

Neste capítulo serão apresentados resultados obtidos através de simulação, utilizando o modelo apresentado no capítulo 2. Foram desenvolvidos scripts[1] na linguagem R, criados com base nos códigos em Matlab referentes ao capítulo 3 de Koop (2003) e disponíveis no site <http://www.wiley.com/legacy/wileychi/koopbayesian/>.

O objetivo nesta parte da monografia é apresentar compararações das estimativas bayesianas para diferentes especificações de priori com as estimativas de MQO, utilizando, para tanto, a média e o erro quadrático médio[2] (EQM) para 10000 (dez mil) repetições considerando diferentes tamanhos amostrais. Foram criados 5 experimentos utilizando dados gerados sem anomalias, detalhados na seção 3.1, e um experimento adicional simulando erro de medida dos dados, que será apresentado na seção 3.2.

3.1 SIMULAÇÕES SEM ANOMALIAS NOS DADOS

Nesta seção são apresentados 5 experimentos, cada um com diferentes especificações no processo gerador dos dados, no valor real dos parâmetros, presença ou não de intercepto e mudanças nas especificações a priori dos parâmetros β (média e variância).

3.1.1 Especificações utilizadas

Os seguintes passos são comuns a todos os experimentos e estão sintetizados na tabela (1):

1. Fixa-se um tamanho amostral, n, $n \in \{30, 100, 500, 1000, 5000\}$, sendo que cada experimento é realizado para cada valor de n;

2. São gerados i vetores $X_1, \ldots, X_i$ (onde i é igual a 4 nos experimentos 1 e 5 e 3 nos demais experimentos) cada um de tamanho n amostrados de maneira independente de distribuições uniformes indicadas na Tabela (1). Estes valores se mantém fixos (isto é, não se alteram nas 10000 repetições do experimento);

3. Especificam-se os valores verdadeiros de $\beta_1, \ldots, \beta_j$, com j igual a 5 no experimento 1, j igual a 4 nos experimentos 2, 4 e 5 e j igual a 3 no experimento 3;

[1] Todos os códigos desenvolvidos para esta monografia estão disponíveis em <https://github.com/aishameriane/TCC>.

[2] O erro quadrático médio é utilizado para comparar estimadores quando não necessariamente os mesmos são não-viesados. De acordo com Gujarati (2006), o erro quadrático médio de um estimador $\hat{\theta}$ é definido como $EQM(\hat{\theta}) = \mathbb{E}[\hat{\theta} - \theta]^2$, que pode ser escrito como a variância de $\hat{\theta}$ mais o seu viés ao quadrado. Obviamente se o estimador é não viesado, seu EQM é a própria variância.

4. É calculado um vetor de pesos que tem papel de inflar a variância dos erros, sendo que cada experimento é repetido para duas especificações diferentes de pesos. A primeira especificação, com pesos menores, utiliza $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$ e a segunda utiliza $peso = \sum \bar{X}_i \cdot \beta_j$, isto é, o peso é resultado da soma de todas as médias dos vetores X_i multiplicados pelo coeficiente β_j correspondente (e, no caso da primeira especificação, ainda divide-se cada termo pelo desvio padrão de X_i). No caso de modelos com intercepto, a componente de peso correspondente é o próprio β_1;

5. Para cada repetição k, $k \in \{1, 2, \ldots, 10000\}$, é simulado um vetor de erros com comprimento n, ε_k, proveniente de uma distribuição normal com média zero e variância unitária;

6. Para cada repetição k é criado um vetor y_k, dado por $y_k = X\beta + peso \cdot \varepsilon_k$;

7. São especificadas 4 prioris diferentes:

   i. A primeira priori utiliza como média os valores reais dos β_j e como variância dos β_j é utilizado 10% do seu valor verdadeiro;

   ii. A segunda priori utiliza como média os valores reais dos β_j e como variância dos β_j é utilizado 40000 (quarenta mil) vezes o seu valor verdadeiro;

   iii. A terceira priori utiliza como média valores quaisquer, obtidos através da geração de números aleatórios[3], e para variância dos β_j é utilizado 10% da média a priori;

   iv. A quarta priori utiliza como média os mesmos valores da Priori 3, enquanto que para variância dos β_j é utilizado o valor de 40000 (quarenta mil) vezes a média a priori.

8. Em todas as prioris o valor especificado para $\underline{s}^2$ foi igual a 20% da média de todos os valores y_k calculados considerando todas as 10000 repetições, isto é, foi tomada a média de cada um dos 10000 vetores e então tomada a média novamente e multiplicou-se esse valor por 0.2;

9. Estimou-se por MQO e pelo método bayesiano (para cada priori) os valores de β_j para cada uma das k repetições, sendo que para cada repetição foi calculado o EQM. Após, calculou-se a média e a média dos EQM.

A Tabela (1) contém o resumo das especificações dadas. Como foram calculados os resultados considerando dois tipos de pesos da variância dos erros diferentes, cada

[3] Gerados a partir de distribuições uniformes cujos parâmetros a e b foram especificados previamente, de acordo com a magnitude do parâmetro a ser estimado. Para valores de β_j entre 0 e 1, utilizou-se $a = 0$ e $b = 1$; para β_j entre 0 e 10, $a = 0$ e $b = 10$ e para valores maiores, $a = 0$ e $b = 1000$.

experimento foi realizado duas vezes. Os resultados obtidos nos 5 experimentos estão nas tabelas (9) a (19) do apêndice B. Para cada experimento, considerando as duas especificações de peso dos erros e dois tamanhos amostrais (o menor e o maior), foram plotados os histogramas das estimativas de cada parâmetro, obtidas utilizando as quatro prioris e por MQO estes gráficos estão nas figuras (11) a (37a) do apêndice B.

3.1.2 Análise dos resultados

As tabelas de resultados bem como os histogramas obtidos para todos os experimentos estão no apêndice B. Nesta seção serão ressaltados os aspectos considerados mais relevantes de forma a resumir os resultados obtidos.

Com relação ao experimento 1 (apêndice B), o processo gerador dos dados (PGD) tinha quatro vetores de variáveis independentes, bastante heterogêneos entre si e isso reflete nas estimativas dos coeficientes associados: coeficientes que multiplicam vetores com valores maiores são melhores estimados em relação àqueles que multiplicam vetores de X que tem menor magnitude. Observa-se que há uma incorreta estimação do coeficiente do intercepto do modelo, de maneira geral, tanto considerando o modelo com ponderação dos erros menores (Tabela 9) como para os pesos maiores (Tabela 10). Uma explicação possível é o fato do coeficiente do intercepto ser uma combinação linear dos demais coeficientes, de maneira que ele irá "absorver" o viés existente em todas as estimativas. Esse problema diminui com o aumento do tamanho amostral, para ambas especificações de pesos dos erros, mas ainda assim o coeficiente β_1 está acompanhado dos maiores valores de EQM.

Ainda em relação ao primeiro experimento, nota-se que em ambas especificações de pesos dos erros as melhores estimativas, em geral, são obtidas pela Priori 1. De maneira intuitiva esta priori representaria uma situação onde o pesquisador ou pesquisadora tem muita certeza sobre a informação a priori (o que se traduz em um valor baixo de variância) e também tem um palpite muito bom sobre a média da distribuição (neste caso, o palpite muito bom utilizado foi o próprio valor verdadeiro, o que não é muito realista em aplicações práticas). Sendo assim, as estimativas obtidas com esta priori funcionam como o "melhor caso possível" e é coerente que sejam obtidas com ela melhores estimativas.

É importante observar que o efeito do tamanho amostral, abordado no capítulo 2, aparece nas simulações: para repetições com tamanho de amostra 5000, os valores das estimativas bayesianas se aproximam de fato das estimativas de MQO. Isso não acontece para o coeficiente do intercepto utilizando a Priori 3, o que é condizente com a teoria estudada, pois esta priori apresenta valores muito diferentes dos valores usados no PGD e também uma baixa variância. Fazendo uma analogia com a realidade, é como se a pessoa por detrás do experimento tivesse muita certeza sobre seu palpite, mas ele é muito destoante do verdadeiro valor dos parâmetros. A priori 2 representa, então, uma espécie de "pior cenário possível".

Tabela 1 – Especificações dos 5 experimentos simulados com dados sem anomalias

Experimento 1						
Dados gerados	**Intercepto**	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	
	Sim	$\mathcal{U}(0,10)$	$\mathcal{U}(500,1000)$	$\mathcal{U}(3500,5000)$	$\mathcal{U}(50000,600000)$	
	β_1	β_2	β_3	β_4	β_5	**Variância** β_i
Valores reais	**2**	**100**	**0.6**	**5**	**0.01**	-
Priori 1	2	100	0.6	0.5	0.01	$0.1 \cdot \beta_i$
Priori 2	2	100	0.6	0.5	0.01	$40000 \cdot \beta_i$
Priori 3	5.1	2.7	15	98.5	37.7	$0.1 \cdot \beta_i$
Priori 4	5.1	2.7	15	98.5	37.7	$40000 \cdot \beta_i$
Experimento 2						
Dados gerados	**Intercepto**	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	
	Sim	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	-	
	β_1	β_2	β_3	β_4	β_5	**Variância** β_i
Valores reais	**1**	**10**	**50**	**100**	**-**	-
Priori 1	1	10	50	100	-	$0.1 \cdot \beta_i$
Priori 2	1	10	50	100	-	$40000 \cdot \beta_i$
Priori 3	0.27	6.22	73.10	720.29	-	$0.1 \cdot \beta_i$
Priori 4	0.27	6.22	73.10	720.29	-	$40000 \cdot \beta_i$
Experimento 3						
Dados gerados	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	**Intercepto**	
	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	-	Não	
	β_1	β_2	β_3	β_4	β_5	**Variância** β_i
Valores reais	**10**	**50**	**100**	**-**	**-**	-
Priori 1	10	50	100	-	-	$0.1 \cdot \beta_i$
Priori 2	10	50	100	-	-	$40000 \cdot \beta_i$
Priori 3	48.17	0.57	61.09	-	-	$0.1 \cdot \beta_i$
Priori 4	48.17	0.57	61.09	-	-	$40000 \cdot \beta_i$
Experimento 4						
Dados gerados	**Intercepto**	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	
	Sim	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	-	
	β_1	β_2	β_3	β_4	β_5	**Variância** β_i
Valores reais	**0.1**	**0.1**	**100**	**100**	**-**	-
Priori 1	0.1	0.1	100	100	-	$0.1 \cdot \beta_i$
Priori 2	0.1	0.1	100	100	-	$40000 \cdot \beta_i$
Priori 3	0.038	0.9946	115.02	169.20	-	$0.1 \cdot \beta_i$
Priori 4	0.038	0.9946	115.02	169.20	-	$40000 \cdot \beta_i$
Experimento 5						
Dados gerados	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	**Intercepto**	
	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,10)$	Não	
	β_1	β_2	β_3	β_4	β_5	**Variância** β_i
Valores reais	**0.1**	**0.1**	**100**	**100**	**-**	-
Priori 1	0.1	0.1	100	100	-	$0.1 \cdot \beta_i$
Priori 2	0.1	0.1	100	100	-	$40000 \cdot \beta_i$
Priori 3	0.038	0.9946	115.02	169.20	-	$0.1 \cdot \beta_i$
Priori 4	0.038	0.9946	115.02	169.20	-	$40000 \cdot \beta_i$

Notas:
⋆ $\mathcal{U}(a,b)$ denota uma variável aleatória com distribuição uniforme no intervalo real $[a,b]$.
⋆ Cada experimento foi repetido 1000 vezes para os seguintes tamanhos amostrais: 30, 100, 500, 1000 e 5000.
⋆ Para experimentos onde o intercepto está presente, β_1 representa o coeficiente do intercepto, β_2 é o coeficiente de X_1. β_3 é o coeficiente de X_2 e assim por diante. Caso o experimento seja sem intercepto, β_i é simplesmente o coeficiente de X_i.
⋆ Para cada experimento, foram amostrados n valores para cada X_i (das distribuições uniforme correspondentes) que foram mantidos fixos em cada uma das 1000 repetições.
⋆ Para gerar os valores de Y_i, a cada experimento gerou-se um vetor de erros com distribuição Normal padrão que foi multiplicado por um peso proporcional à soma das médias de cada X_i multiplicada pelo respectivo coeficiente β_i dividido pelo desvio padrão de X_i. Adicionalmente, cada experimento foi repetido considerando os pesos sem a divisão pelo desvio padrão, totalizando 2 versões de cada um dos 5 experimentos.

Fonte – Elaboração própria com base em dados simulados.

As estimativas para β_4 e β_5 foram as que menos tiveram variação devida ao aumento de n, o que também é esperado pois estão associados com os vetores da matriz X que tem maiores valores: X_4 foi gerado com valores entre 3500 e 5000, enquanto que X_4 multiplica valores que vão de 50mil a 650mil. Este último parâmetro, β_5, praticamente teve estimativas invariantes ao método de estimação, à priori utilizada, ao tamanho amostral e magnitude dos erros do PGD. Ainda em relação aos pesos dados aos erros do processo, observa-se que as estimativas na presença de maior volatilidade são também mais incertas: os histogramas produzidos (figuras 11 até 15) apresentam maior amplitude, sendo que alguns incluem faixas negativas (nenhum parâmetro é nulo ou negativo nos processos gerados).

Nos valores onde o tamanho amostral é menor ($n = 30$), as estimativas utilizando a Priori 2 foram ligeiramente superiores às estimativas de MQO, comparando o EQM. Porém, quando são observados os histogramas, quase não há diferença nas estimativas obtidas por ambos métodos.

No experimento 2 foram utilizados quatro vetores $X_1, \ldots, X_4$ com mesma especificação, todos foram amostrados de distribuições uniformes no intervalo de 0 a 10, com a finalidade de controlar a fonte de variabilidade oriunda da magnitude dos valores das variáveis exógenas. As estimativas geradas para o coeficiente do intercepto continuam sendo bastante voláteis quando comparadas com as demais, para ambos pesos multiplicando o vetor de erros (tabelas 11 e 12). Essa variabilidade, além de afetar as estimativas das médias e EQM, aparece também nos histogramas das figuras 16 a 20.

Nos histogramas construídos para estimativas onde o tamanho da amostra é menor (figuras 16a, 16c, 17a, 17c, 18a, 18c, 19a, 19c, 20a e 20c) observa-se que os extremos do eixo horizontal do gráfico tem variações grandes, que em alguns casos é da ordem de -200 a +200. As melhores estimativas para o intercepto ainda são aquelas produzidas pela Priori 1 (figuras 17a e 17c), conforme esperado. Porém, para os demais parâmetros, as diferenças nas estimativas já não são tão grandes já a partir de $n = 500$, sugerindo que a magnitude das variáveis exógenas têm influência nos resultados da estimação. Outro fator que parece exercer influência é o valor real de β_j: valores maiores já apresentam maior convergência de resultados, em média, nos valores amostrais pequenos e conforme n aumenta observa-se uma diminuição maior do EQM.

Para o experimento 3, foram reproduzidas as mesmas especificações do experimento 2, porém com ausência de intercepto, ou seja, removeu-se o coeficiente β_1. As estimativas obtidas (tabelas 13 e 14) foram muito similares, em média, àquelas do experimento 2, porém com um EQM ligeiramente inferior, em particular no maior tamanho de amostra (5000). Nas estimativas com $n = 30$, a estimativa utilizando a Priori 2 apresentou valores de EQM ligeiramente inferiores às estimativas de MQO para ambas especificações do peso dos erros. Os intervalos de variação dos histogramas (figuras 21 a 25) são igualmente similares aos coeficientes correspondentes do modelo com intercepto.

Por fim, nos experimentos 4 e 5 procurou-se investigar qual o efeito em coeficientes que tenham mesmo valor no PGD na presença ou não de intercepto no modelo (e na próxima seção isso é avaliado quanto à presença de erro de medida, usando a mesma especificação). Mais uma vez as estimativas do intercepto (exceto utilizando a Priori 2) apresentaram maiores diferências, em média, para o valor real, além de alto EQM. No experimento 4, ambos coeficientes β_1 e β_2 tem o mesmo valor de 0.1 no PDG, porém as estimativas para β_1 ficaram em torno de -0.039 (exceto na Priori 2 - Tabela 15). Há uma aparente melhora nas estimativas, em média, para o modelo com maior variabilidade de erros, considerando o modelo com intercepto, tanto nos valores da Tabela (16) como nos histogramas, que apresentaram menor variação no eixo horizontal (figuras 26 a 30). Já os demais coeficientes, quando comparados entre o experimento 4 e 5, são bastante similares.

Analisando isoladamente os resultados do modelo 5 percebe-se que a estimativa dos coeficientes β_1 e β_2 tem valores muito próximos quando o tamanho da amostra é grande, tanto para o modelo com pesos de erros menores (Tabela 17) quanto maiores (Tabela 18). Isso sugere que de fato o problema estaria em estimar o intercepto, independente do método utilizado, se bayesiano[4] ou clássico. Para os valores de amostra mais baixos, excetuando a priori 1, as demais estimativas são mais distantes para β_1 do que para β_2. É possível que os valores de X_1 gerados sejam distintos dos valores de X_2 o suficiente para causar essa diferença uma vez que o coeficiente tem valor pequeno. Já para β_3 e β_4, não parecem haver diferenças grandes entre eles e mesmo as estimativas para pequenos tamanhos de amostra são bastante próximas, em média, do valor verdadeiro, apenas apresentando um alto EQM que diminui para valores pequenos (em magnitude, quando comparados com o valor do parâmetro) à medida que a amostra aumenta.

3.2 SIMULAÇÃO COM ERRO DE MEDIDA

Nesta seção são apresentados os resultados para simulações incluindo erro de medida nos valores de X utilizados para a estimação do modelo. De acordo com Gujarati (2006), em aplicações práticas podem ocorrer diversos problemas na mensuração das variáveis, sejam valores faltantes, erros de transcrição, erros de cálculo, etc.

Assim como nos experimentos da seção anterior, foram utilizados 5 tamanhos amostrais (n) diferentes e, para cada um deles, foram realizadas 10000 repetições. Os passos utilizados para geração dos dados são os seguintes:

1. São criados 4 vetores de variáveis independentes, $X_1 \dots X_4$ (de tamanho n), amostrados de maneira independente de distribuições uniforme com parâmetros $a = 0$ e $b = 10$;

[4] Com exceção das estimativas com a Priori 1, que é aquela onde os valores da média à priori são iguais ao PGD e a variância é pequena.

2. O modelo gerado não tem intercepto e os valores verdadeiros dos parâmetros β_i são 0.1, 0.1, 100 e 100 (similar ao experimento 5);

3. Criam-se 10000 vetores de erros ε_t, que têm tamanho n e são amostrados a partir de uma distribuição normal padrão. Multiplicando os vetores, há um escalar de *peso*, que é igual a $\sum_{i=1}^{4} \bar{X}_i \cdot \beta_i \cdot {}^{1}/_{SD(X_i)}$ (onde $\bar{X}_i$ é a média amostral do vetor gerado X_i e $SD(X_i)$ seu desvio padrão, também amostral);

4. Utilizando a relação $Y = X\beta + \epsilon$ são calculados 10000 vetores de variáveis dependentes;

5. São simulados dois vetores de tamanho n que tem observações independentes de uma distribuição normal, chamados de u e v. A variância destes erros foi alterada entre as especificações para verificar o seu efeito nas estimativas;

6. O vetor X_2 do passo (1) é somado com u e e o vetor X_4, também do passo (1) é somado com v, dando origem aos vetores X_2^* e X_4^*. É criada, então, uma matriz X^* que tem como colunas X_1, X_2^*, X_3 e X_4^* e é utilizada como matriz na estimação dos parâmetros do modelo.

As etapas seguintes de estimação são as mesmas já abordadas nos experimentos da seção 3.1, apenas trocando a matriz X por X^*. Assim como nos experimentos anteriores, foram consideradas 4 prioris diferentes, alterando as médias e variâncias, conforme a Tabela (2). É importante notar que os valores utilizados a priori foram os mesmos do experimento 5, para comparação dos resultados.

Tabela 2 – Especificação das prioris, valores reais dos parâmetros β_i e processos de geração das variáveis independentes na simulação com erro de medida

Dados gerados	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	**Intercepto**
	U(0,10)	U(0,10)	U(0,10)	U(0,10)	Não
	β_1	β_2	β_3	β_4	Variância β_i
Valores reais	**0.1**	**0.1**	**100**	**100**	-
Priori 1	0.1	0.1	100	100	$0.1 \cdot \bar{\beta}_i$
Priori 2	0.1	0.1	100	100	$40000 \cdot \bar{\beta}_i$
Priori 3	0.038	0.9946	115.02	169.20	$0.1 \cdot \bar{\beta}_i$
Priori 4	0.038	0.9946	115.02	169.20	$40000 \cdot \bar{\beta}_i$

Fonte – Elaboração própria.

Como já existe uma componente de variabilidade devida ao erro de medida, nesta seção **não** foram consideradas duas especificações para a magnitude da variância dos erros do processo gerador dos dados (que é dada pelo escalar *peso*). Porém, foram testadas diferentes prioris para $\underline{s}^2$, que é a priori para a variância dos erros. Nos experimentos

anteriores isso não havia sido alterado (era sempre igual a 20% da média de Y) As estimativas por MQO e à posteriori considerando 4 prioris diferentes, bem como as especificações de experimento utilizadas, estão na Tabela (2).

Optou-se por fazer os histogramas dos valores estimados apenas da simulação 1, que estão na Tabela (19), para fins comparativos com o resultados do experimento 5 (Tabela 17). Comparando as estimativas obtidas utilizando a primeira especificação (topo da Tabela 17) com as estimativas do experimento 5 para o mesmo tamanho de amostra, observa-se que nenhuma estimativa com erro de medida para os dois primeiros coeficientes foi satisfatória, sendo até mais do que 10 vezes o valor real do parâmetro. Para o coeficiente β_1, a melhor estimativa obtida, tanto em termos de valor médio como de EQM foi utilizando a Priori 3. Isso é um pouco surpreendente uma vez que esta priori utiliza um valor distante (0.038) do verdadeiro (0.1) e, ao mesmo tempo, tem pouca variabilidade.

As estimativas dos coeficientes β_3 e β_4, quando comparadas com o valor verdadeiro dos parâmetros no PGD, são mais próximas dos valores reais, porém apresentam um alto EQM. Como o valor da amostra é muito grande nesta primeira especificação, as estimativas bayesianas são "puxadas" para o valor do estimador de MQO, que em uma especificação com erro de medida é um estimador viesado. Com relação a isso, é interessante notar que na terceira e quarta especificações do experimento (Tabela 17), onde o tamanho da amostra é reduzido ($n = 30$), as estimativas de β_1 usando a Priori 1 tem um desempenho bastante superior do que o estimador de MQO, porém as estimativas usando as prioris 2 e 4 ficam próximas ao estimador clássico.

A variância dos erros de medida também desempenha um papel interessante nas estimativas. Na segunda especificação (tabela 17) onde a variância do erro de medida é igual a 0.1, as estimativas em geral ficaram melhores em média e com menor EQM se comparadas com a última espeficicação. Porém é necessário observar que houve alteração na variância à priori dos erros do PGD, que também estão mais voláteis na última espeficicação.

Ainda com relação à variância do erro de medida, comparando a terceira e a quarta especificações é possível ver o que ocorre quando apenas um dos erros de medida tem sua volatilidade alterada. Mesmo que apenas a componente X_4^* esteja relacionada com o erro v, que teve variância multiplicada por 100 vezes, as estimativas dos coeficientes das outras variáveis também sofreram mudanças, sendo que β_1 e β_2 foram os que apresentaram maior diferença (possivelmente devido ao valor verdadeiro do parâmetro ser muito pequeno). Quando o tamanho da amostra é aumentado (para a especificação 5) junto com a volatilidade dos erros do PGD, as estimativas ficam ainda piores.

A influência do tamanho da amostra pode ser percebida comparando as especificações 2 ($n = 5000$) e 3 ($n = 30$). Ambas especificações apresentam estimativas, em média, próximas dos valores reais para β_4 e β_3, porém com o tamanho de amostra maior o EQM é igualmente menor, o que é esperado pela convergência do estimador clássico da variância (lembrando que para n grande as estimativas bayesianas convergem para as de MQO). A análise é complementada pelos histogramas da Figura (36).

4 APLICAÇÃO A DADOS REAIS

Para exemplificar o uso do modelo de regressão linear normal com priori conjugada natural apresentado no capítulo 1, será desenvolvida uma aplicação baseada no artigo de Bonini e Pozzobon (2016), que fala de prêmio salarial e discriminação de gênero nos salários de alguns setores produtivos da região Sul do Brasil. Diferentemente do artigo original, nesta monografia serão analisados apenas os dados do setor de tecnologia de informação (TI) para o estado de Santa Catarina. Optou-se por essa simplificação de redução do modelo original pois o foco da monografia é apresentar o método de estimação.

Este capítulo está dividido da seguinte forma: uma breve introdução ao referencial teórico sobre mercado de trabalho e prêmio salarial, além do modelo para o problema são apresentados, seguida pela descrição dos dados da amostra. As estimativas utilizando o mesmo modelo de regressão normal com priori conjugada natural normal gama são comparadas com as estimativas de MQO e depois é feita a previsão, pelos dois métodos, para 10 elementos de fora da amostra. Ressalta-se, também, que a análise tem como objetivo verificar se o modelo estimado utilizando as técnicas bayesianas irá produzir resultados consistentes com os obtidos no trabalho original de Bonini e Pozzobon (2016) e por isso não foi realizada a interpretação econômica das estimativas de maneira mais aprofundada.

4.1 PRÊMIO SALARIAL E DISCRIMINAÇÃO DE GÊNERO NO MERCADO DE TRABALHO DE TI

A falta de representatividade feminina em carreiras nas áreas de ciência, tecnologia, engenharias e matemática (STEM[1]) tem sido discutida na literatura por suas implicações para as mulheres e para a sociedade como um todo. Por exemplo, a falta de mulheres participantes em estudos clínicos levou à posterior descoberta de que os efeitos de alguns medicamentos eram diferentes nas mulheres em comparação com os homens. Além disso, carreiras nas áreas de STEM usualmente tem um salário mais alto em comparação com as demais áreas, acentuando o gap salarial de gênero (KAHN; GINTHER, 2017). Bonini e Pozzobon (2016) apontam para a crescente demanda de profissionais nessa área que não é acompanhada pela oferta, o que gera a pressão para aumento dos salários.

Proposta por Becker na década de 50, a teoria do capital humano relaciona as variáveis intrínsecas aos indivíduos com a sua produtividade. Especificamente, a teoria assume que educação formal e treinamento dentro da empresa determinam atividades

[1] Do inglês *science, technology, engineering and mathematics.*

cognitivas que por sua vez impactam em habilidades e produtividade. A produtividade, de acordo com a teoria neoclássica, deve ser igual a remuneração do fator mão-de-obra, que é o salário. Sendo assim, existe uma ligação entre as habilidades da pessoa (mensuradas pela educação e treinamento) e a sua remuneração.

A teoria do capital humano entra em um arcabouço econométrico através das funções mincerianas, que permitem uma avaliação quantitativa do efeito das características dos trabalhadores sobre o salário. Essa abordagem pode ser espandida, usando a decomposição de Oaxaca, para avaliação do efeito que a discriminação que determinados grupos sofrem tem nos salários.

O modelo econométrico utilizado tem o mesmo formato do modelo de Bonini e Pozzobon (2016), que possui a seguinte forma funcional:

$$\begin{aligned}\ln(\text{salário}) = \beta_0 + \beta_1 \cdot \text{Sup. Inc.} + \beta_2 \cdot \text{Sup. Comp.} + \beta_3 \cdot \text{Pós-Grad.} \\ + \beta_4 \cdot \text{Idade} + \beta_5 \cdot \text{Idade}^2 + \beta_6 \cdot \text{T. Emp.} + \beta_7 \cdot \text{Gênero} + \varepsilon,\end{aligned} \tag{4.1}$$

onde:

- *ln(salário)* é o vetor com os logarítmos naturais dos salários, em reais, dos trabalhadores;
- β_0 é o coeficiente de intercepto, que representa a heterogeneidade devida a variáveis não observadas no modelo e comuns a todos indivíduos;
- *Sup. Inc.*, *Sup. Comp.* e *Pós-Grad.* são vetores de variáveis binárias (com valores 0 e 1) que identificam se a escolaridade máxima do indivíduo é superior incompleto, superior completo ou pós graduação, respectivamente. Quando as três variáveis são iguais a 0 significa que o grau de escolaridade máximo da pessoa é ensino médio;
- *Idade* e *Idade*^{2} representam os vetores com as idades, em anos, e seus valores ao quadrado. O quadrado da idade é utilizado por assumir que os retornos marginais no salário são decrescentes para anos de vida;
- *T. Emp.* é o vetor com os tempos de emprego (do emprego atual) dos indivíduos, mensurados em meses;
- *Gênero* é um vetor com a variável dicotômica que é 0 para homens e 1 para mulheres. Uma vez que o gênero de um indivíduo não é uma "habilidade", o coeficiente associado a esta variável mede o grau de discriminação salarial entre homens e mulheres;
- É assumido que os termos do vetor ε são independentes, identicamente distribuídos e não correlacionados com as variáveis explicativas.

4.2 DESCRIÇÃO DA BASE DE DADOS

Bonini e Pozzobon (2016) utilizaram dados da Relação Anual de Informações Sociais (RAIS) do Ministério do Trabalho para o ano de 2011. Como mencionado no início do capítulo, no trabalho original foram analisados os três estados da Região Sul do Brasil, porém para esta monografia optou-se por utilizar uma versão reduzida da base de dados contendo apenas as informações dos trabalhadores de TI que trabalhavam no Estado de Santa Catarina. Ao todo, a amostra utilizada neste trabalho compreende 4328 trabalhadores. Destes, 10 foram selecionados para servirem de validação para previsão de dados fora da amostra, fazendo com que o modelo tenha sido estimado com base em 4318 observações.

A variável resposta é o logaritmo neperiano do salário hora. Sendo assim, os coeficientes estimados representam a variação salarial, em termos percentuais, entre os diferentes atributos (mantendo os demais constantes). Além do salário, outras três variáveis são quantitativas: tempo no emprego atual (em meses), idade (em anos) e idade ao quadrado (em anos ao quadrado). As estatísticas descritivas para as variáveis quantitativas (removendo as 10 observações que foram utilizadas para previsão fora da amostra) encontram-se na tabela (3):

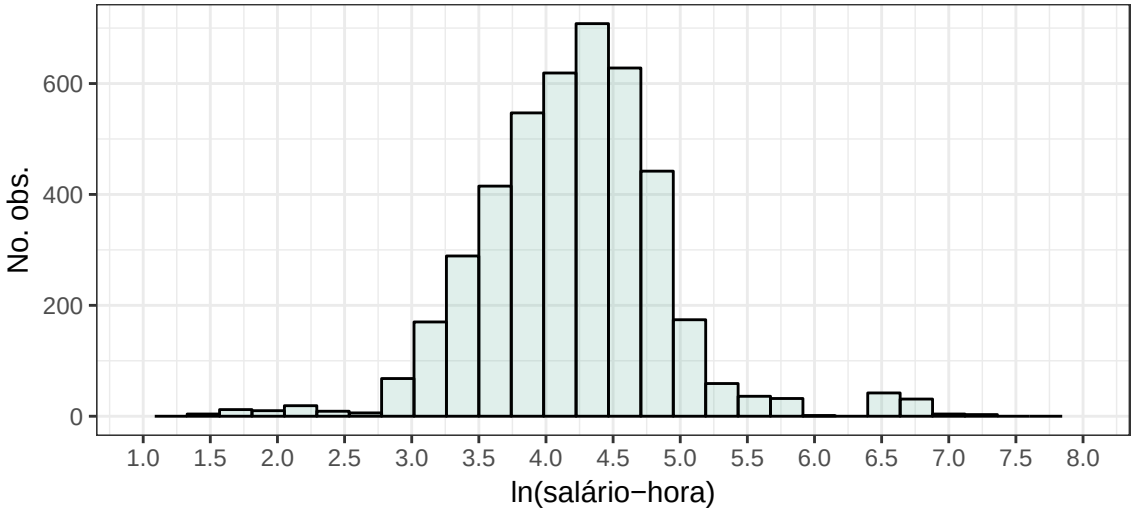
Tabela 3 – Estatísticas descritivas das variáveis quantitativas do modelo

	ln(Salário-Hora)*	Tempo de Emprego†	Idade**	(Idade)^2
N^o de Observações	4318	4318	4318	4318
Mínimo	1.38	0.10	17.00	289.00
1° quartil	3.78	9.70	26.00	676.00
Mediana	4.22	16.90	29.00	841.00
Média	4.21	34.59	30.58	989.10
3° quartil	4.60	37.10	34.00	1156
Máximo	7.32	433.40	72.00	5184
Desvio Padrão	0.69	54.26	7.35	527.32

* O salário-hora é medido em reais (R\$) de 2011 e ln(Salário-Hora) denota seu logaritmo na base neperiana.
† Tempo de emprego é medido em meses.
** Idade é medida em anos e, consequentemente, idade ao quadrado, denotada por (Idade)^2, tem como unidade anos ao quadrado.
Fonte – Elaboração própria utilizando dados da RAIS de trabalhadores de setores de TI no estado de Santa Catarina para o ano de 2011.

Tomando como base os valores dos logarítmos dos salários, os salários hora em reais variam de R$ 3.98 a R$ 1513.00, o que evidencia a heterogeneidade da variável resposta entre os indivíduos. O salário hora mediano é de R$ 68.23 enquanto que o salário médio é de R$ 89.40, indicando relativa simetria na distribuição da variável. O histograma do logaritmo dos salários está na Figura (4) e é possível perceber que de fato a distribuição está centralizada em torno da média de 4.2, com uma ligeira assimetria positiva (cauda para a direita), que é característica de dados de renda.

Figura 4 – Histograma do logaritmo do salário hora dos trabalhadores de TI do Estado de Santa Catarina para o ano de 2011



Fonte – Elaboração própria com base nos dados da RAIS.

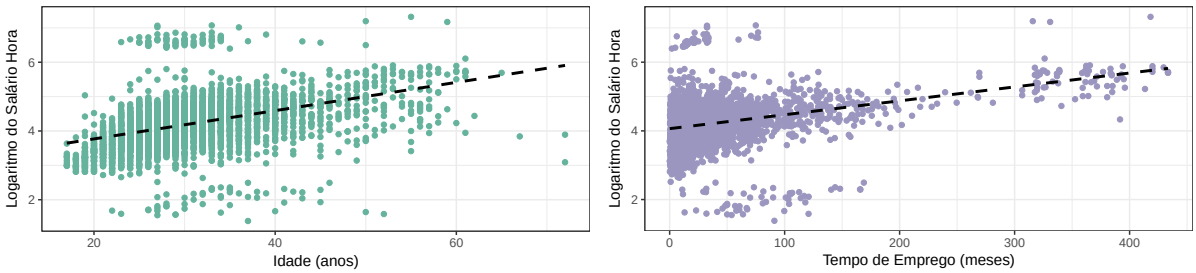
Com relação ao tempo de emprego, a maior parte dos trabalhadores (75%) tem no máximo 37 meses no serviço atual, o que corresponde a pouco mais de 3 anos, sendo que a mediana no tempo de emprego é de 16 meses e 27 dias. A idade dos trabalhadores também apresenta uma variação grande, indo de 17 a 72 anos. Embora o intervalo seja amplo, a mediana indica que até 50% dos trabalhadores tem no máximo 29 anos, evidenciando uma concentração de pessoas jovens nesta amostra.

Na figura (5) são apresentados os gráficos de dispersão tanto da idade como do tempo de emprego versus o logaritmo do salário-hora. Em ambos há uma tendência linear positiva, porém com grande concentração de valores menores no eixo x. Em particular, para a idade (Figura 5a), alguns outliers que apresentam pouca idade e elevados salários. Para o tempo de emprego (Figura 5b), há uma concentração ainda maior de baixos valores, porém há três pequenos grupos que se destacam: a) pessoas com pouco tempo de emprego e alto salário; b) pessoas com pouco tempo de emprego e c) salário bem abaixo da média e pessoas com muito tempo de emprego e altos salários. Em ambos gráficos da Figura (5) estão plotadas linhas tracejadas cujas inclinações e interceptos foram obtidos calculando a reta de regressão linear simples das variáveis.

As variáveis categóricas que compõe o modelo são as de escolaridade e gênero. A amostra é formada de pessoas com 4 níveis de escolaridade máximo: ensino médio, ensino superior incompleto, ensino superior e pós graduação. Para cada uma delas foi criada uma variável indicadora que é igual a 1 caso a pessoa tenha no máximo o grau de instrução correspondente e zero caso contrário, sendo que o ensino médio é a categoria base e portanto não é incluída no modelo para evitar multicolinearidade.

Ainda com relação às variáveis categóricas, para a variável gênero adotou-se como

Figura 5 – Gráficos de dispersão das variáveis explicativas quantitativas contra a variável explicada no modelo



(a) Distribuição da idade e logaritmo do salário hora

(b) Distribuição do tempo de emprego e logaritmo do salário hora

Fonte – Elaboração própria com base nos dados da RAIS.

categoria basal o sexo masculino, de forma que um homem que cursou no máximo o segundo ano de graduação teria o preenchimento de suas informações no banco de dados como descrito na tabela (4).

Tabela 4 – Exemplo de preenchimento do banco de dados

Variável	Gênero	Sup. Incompleto	Superior	Pós Graduação
Valor	0	1	0	0

Fonte – Elaboração própria.

Tabela 5 – Proporções nas categorias das variáveis categóricas incluídas no modelo

	Sim		**Não**	
	n	%	**n**	%
Homem*	3285	76,08	1033	23,92
Ensino superior incompleto†	697	16,14	3621	83,86
Ensino superior completo†	3201	74,13	1117	25,87
Pós graduação†	29	0,0067	4289	99,33

* A categoria "Homem" é considerada a categoria base para a variável do modelo "Gênero".

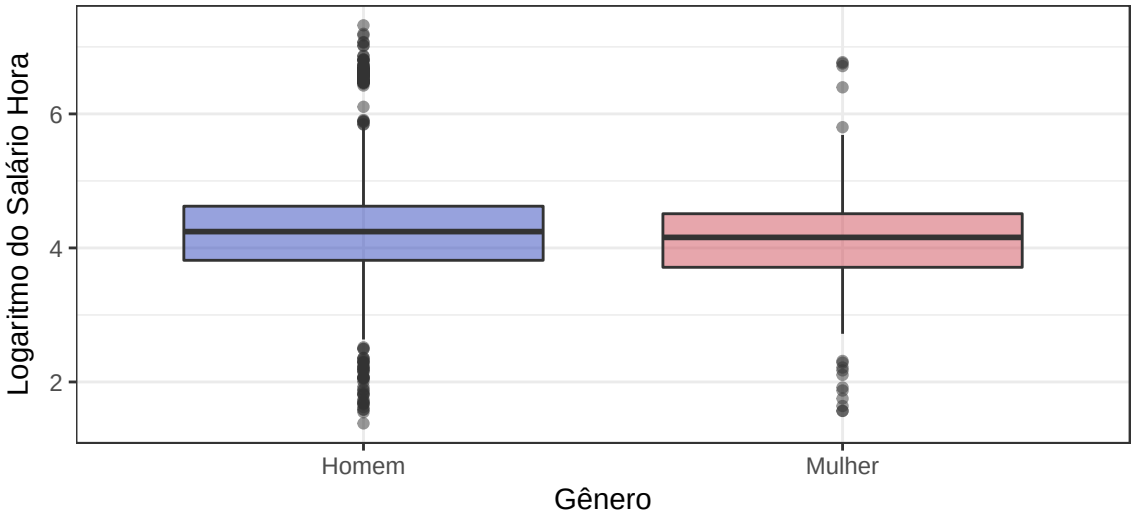
† As variáveis de escolaridade representam o grau de escolaridade máximo do indivíduo, isto é, uma pessoa que tenha cursado até o primeiro ano da faculdade terá "Sim" para ensino superior incompleto e "Não" para ensino superior completo e pós graduação. A categoria base da variável escolaridade é "Ensino Médio".

Fonte – Elaboração própria utilizando dados da RAIS de trabalhadores de setores de TI no estado de Santa Catarina para o ano de 2011.

As mulheres representam pouco menos de um quarto do total da amostra ($23,92\%$), o que vai ao encontro das evidências da literatura atual, que aponta diferenças significativas na proporção de mulheres que participam de carreiras STEM. Com relação à escolaridade, grande parte da amostra (74.12%) é formada de pessoas cujo maior grau de instrução é ensino superior completo, versus 0.01% com pós-graduação e 16.13% com ensino superior

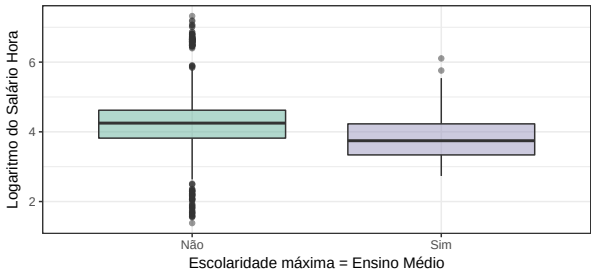
incompleto - os cerca de 10% restantes tem apenas ensino médio completo. A tabela (6) contém as proporções das variáveis categóricas na amostra.

Figura 6 – Boxplot do logaritmo do salário hora estratificado por gênero para trabalhadores de TI no Estado de Santa Catarina no ano de 2011

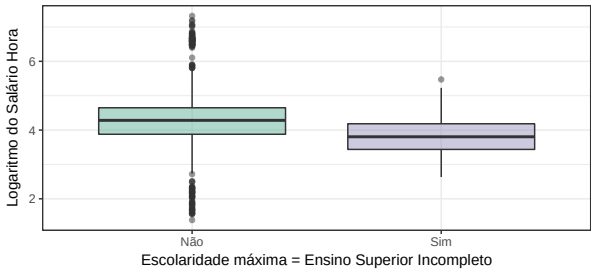


Fonte – Elaboração própria com base nos dados da RAIS.

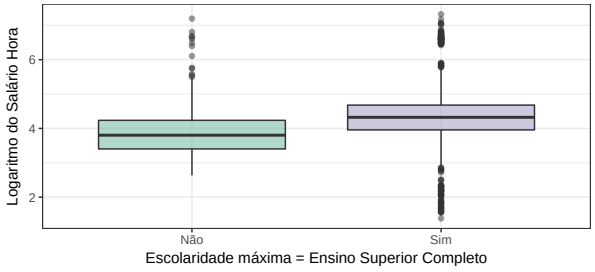
Figura 7 – Boxplots do logaritmo do salário hora por escolaridade máxima para trabalhadores de TI no Estado de Santa Catarina no ano de 2011



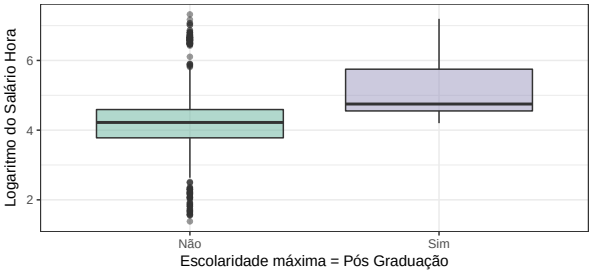
(a) Distribuição do logaritmo do salário hora para trabalhadores que cursaram até o ensino médio versus os demais trabalhadores



(b) Dist. do log. do salário hora para trabalhadores que cursaram até o ensino superior incompleto versus os demais trabalhadores



(c) Dist. do log. do salário hora para trabalhadores que cursaram até o final do ensino superior versus os demais trabalhadores



(d) Distribuição do logaritmo do salário hora para trabalhadores que finalizaram pós graduação versus os demais trabalhadores

Fonte – Elaboração própria com base nos dados da RAIS.

A análise bivariada simples entre gênero e salário (Figura 6) não evidencia diferenças no salário de homens e mulheres. Embora os homens apresentem alguns valores de salários mais elevados (representados pelos outliers na parte superior do gráfico), a maior parte da distribuição dos salários de ambos sexos está concentrada na mesma região. No entanto, ao contrastar as categorias de escolaridade máxima pelo logaritmo do salário (Figura 7) é possível perceber que os níveis de escolaridade máxima mais baixos (ensino médio e superior incompleto) estão relacionados com salários mais baixos, sugerindo a existência de um prêmio salarial para os concluintes do curso superior e pós graduação.

4.3 DETERMINAÇÃO DA PRIORI

Foram utilizadas como médias a priori dos coeficientes os valores estimados por Bonini e Pozzobon (2016): intercepto $\underline{\beta}_0 = 2.07$; Superior Incompleto $\underline{\beta}_1 = 0.12$; Superior Completo $\underline{\beta}_2 = 0.45$; Pós-graduação $\underline{\beta}_3 = 1.07$; Idade $\underline{\beta}_4 = 0.13$; Idade ao quadrado $\underline{\beta}_5 = -0.0014$; Tempo de emprego $\underline{\beta}_6 = 0.0027$; Gênero $\underline{\beta}_7 = -0.15$. Como a base de dados desta monografia pode apresentar diferenças da utilizada em Bonini e Pozzobon (2016) e podem haver diferenças no método de estimação, as variâncias a priori de todos os coeficientes foram escolhidas como iguais a 100.

Para a priori da variância dos erros, $\underline{s}_0^2$, adotou-se o valor de 1, uma vez que a variável resposta está no intervalo de 1 a 8 (Figura 4). Ou seja, assumindo que os erros são normalmente distribuídos com média zero e variância unitária, 95% dos erros estariam no intervalo $(-1.96, 1.96)$, o que seria condizente com os valores de salário-hora desta amostra. Para não dar muito peso aos valores da priori, os graus de liberdade $\underline{\nu}$ foram fixados em 44. De acordo com Koop (2003), isto significa que o peso da priori em relação à amostra é de aproximadamente 1% (pois $\underline{\nu}/N = 44/4318 \approx 0.01$). Assumiu-se que não há covariância a priori entre $\underline{\beta}_i$ e $\underline{\beta}_j$ para $i \neq j$.

4.4 RESULTADOS DO MODELO

A tabela (6) resume os resultados estimados, utilizando o método de mínimos quadrados ordinários e a priori normal gama conjugada. Para cada um dos parâmetros é apresentada a média a priori, a estimativa e erro padrão calculada pelo método dos mínimos quadrados ordinários, a média e o desvio padrão a posteriori, o intervalo de credibilidade [2] de 95% a posteriori e a probabilidade a posteriori de que $\bar{\beta}_j$ seja positivo.

[2] Intervalos de credibilidade, na econometria bayesiana, se diferenciam substancialmente dos intervalos de confiança da abordagem clássica. Informalmente, um conjunto com α% de credibilidade é um subconjunto C da posteriori para o qual a probabilidade do parâmetro θ pertencer a C é igual a α. Isso se distingue da visão clássica onde o intervalo é uma função da amostra e portanto é uma variável aleatória (os parâmetros, sob o enfoque tradicional, são constantes e portanto não estão associados a probabilidades). Como para uma mesma densidade a posteriori existem infinitos intervalos que

Tabela 6 – Valores estimados usando MQO e o método bayesiano com priori normal gama conjugada

Parâmetro	Priori	MQO	D.P. MQO	Posteriori	D.P. Posteriori	IC95%	$P(\bar{\beta}_j \geq 0)$
Constante	2.0712	1.49258	0.1238	1.49286	0.1249	[1.2481 ; 1.7376]	>0.9999
Sup. Inc	0.1227	0.09413	0.0368	0.09411	0.0371	[0.0213 ; 0.1669]	0.99438
Sup. Comp.	0.4487	0.3738	0.0319	0.37381	0.0321	[0.3108 ; 0.4368]	>0.9999
Pós	1.0708	1.07101	0.1124	1.07102	0.1134	[0.8487 ; 1.2934]	>0.9999
Idade	0.1285	0.11962	0.0072	0.11961	0.0073	[0.1053 ; 0.1339]	>0.9999
Idade^2	-0.0014	-0.00132	0.0001	-0.00132	0.0001	[-0.0015 ; -0.0011]	<0.0001
Temp. Emp.	0.0027	0.00252	0.0002	0.00252	0.0002	[0.0021 ; 0.0029]	>0.9999
Gênero	-0.1537	-0.13239	0.0208	-0.13239	0.0210	[-0.1736 ; -0.0912]	<0.0001

Fonte – Elaboração própria.

As estimativas bayesianas estão bastante próximas das estimativas de MQO, o que era esperado considerando que o tamanho da amostra é bastante elevado ($n = 4318$) e a precisão dos βs à priori foi baixa. Conforme discutido nos capítulos anteriores, a média à posteriori do modelo de regressão linear normal com priori normal gama conjugada é uma média ponderada da média à priori e do estimador de MQO, sendo que o peso do primeiro termo é inversamente proporcional à variância à priori.

Os intervalos de credibilidade obtidos são relativamente estreitos em relação à média dos parâmetros, de forma que possibilitam fazer afirmações mais assertivas sobre os valores das médias obtidos. Em particular, apenas o intervalo de credibilidade para o intercepto e para o coeficiente associado com escolaridade máxima igual a ensino superior completo não tem interseção com os valores da priori, o que evidencia a concordância entre os métodos de estimação.

Os parâmetros que no trabalho original de Bonini e Pozzobon (2016) tiveram a hipótese de nulidade rejeitadas apresentaram, com a abordagem bayesiana, uma alta probabilidade[3] de serem positivos (maior do que 99.99%). No caso do parâmetro da idade ao quadrado e do coeficiente associado a gênero, a probabilidade de serem positivos é menor do que 0.01%, o que é consistente com a literatura da área: pessoas mais jovens irão dedicar um tempo maior para capacitação e educação formal (pois estão no início de suas carreiras), o que faz com que a variação (gradiente) do salário seja maior. Isto é, pessoas com menor idade vão apresentar maior retorno marginal no salário enquanto que pessoas com mais idade tem retorno marginal decrescente a cada ano a mais de vida.

Quanto à escolaridade, os resultados estimados são consistentes em sinal com os estimados obtidos por Bonini e Pozzobon (2016) sendo bastante próximos no caso

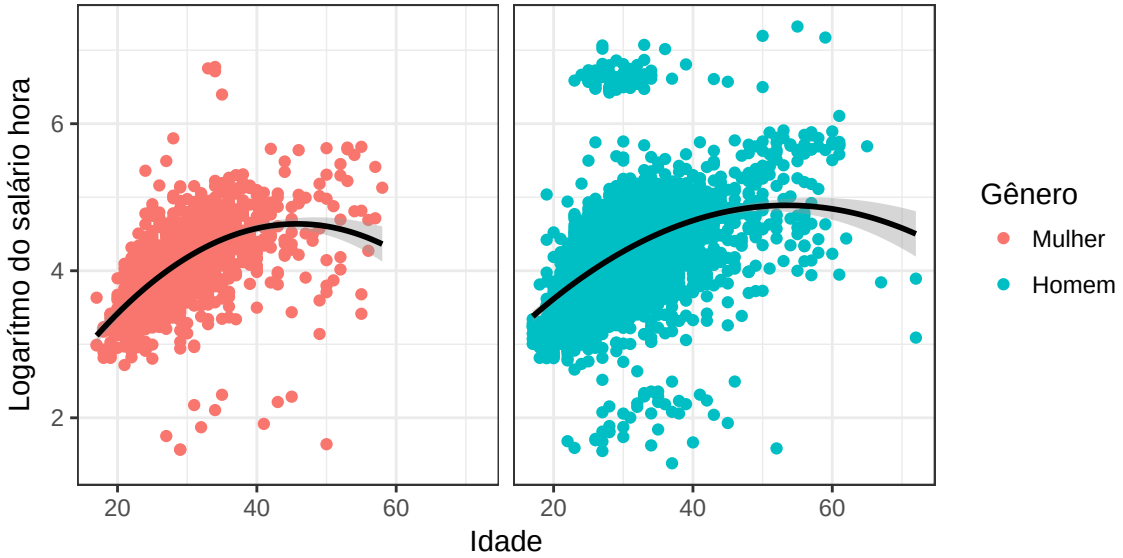
satisfazem $\mathbb{P}(\theta \subset C) = \alpha\%$, é comum utilizar o intervalo de credibilidade HPD (de *highest probability density* ou de maior densidade a posteriori). O IC HDR com probabilidade $\alpha\%$ é a região da posteriori que satisfaz 3 propriedades: i) a densidade da posteriori em qualquer ponto da região é maior ou igual que a densidade da posteriori em qualquer outro ponto fora; ii) a amplitude da região é a menor possível; iii) a probabilidade de θ pertencer à região é igual a α (ROBERT, 2007). Para uma definição formal, sugere-se a leitura do capítulo 2.5 de Geweke (2005).

[3] A probabilidade de um determinado coeficiente ser maior do que um valor é calculada com base na distribuição t teórica apresentada no capítulo 1.

de superior incompleto e pós graduação. Mesmo com as diferenças entre os valores obtidos e os valores da literatura, as estimativas estão consistentes com a teoria pois os coeficientes associados com escolaridades mais elevadas são maiores. Além disso, os coeficientes corroboram com as evidências encontradas na etapa descritiva do trabalho: as diferenças salariais de pessoas com ensino médio e ensino superior incompleto eram similares entre si e diferentes do padrão para indivíduos com ensino superior completo e pós graduação (Figura 7).

A idade apresenta coeficiente positivo enquanto que a idade ao quadrado, conforme esperado, tem sinal negativo. Isto está em conformidade com a literatura: no início da jornada profissional os indivíduos buscam maior capacitação profissional de forma que os rendimentos aumentam com maior velocidade. Esse efeito "quadrático" da idade nos rendimentos pode ser visto na Figura (8), representado pela linha preta em ambos gráficos. Uma possível justificativa para que o efeito não seja tão acentuado quanto as curvas encontradas em outros trabalhos é devida ao fato da amostra ser composta principalmente de pessoas jovens.

Figura 8 – Ajuste quadrático do logaritmo do salário hora por idade, estratificado por gênero



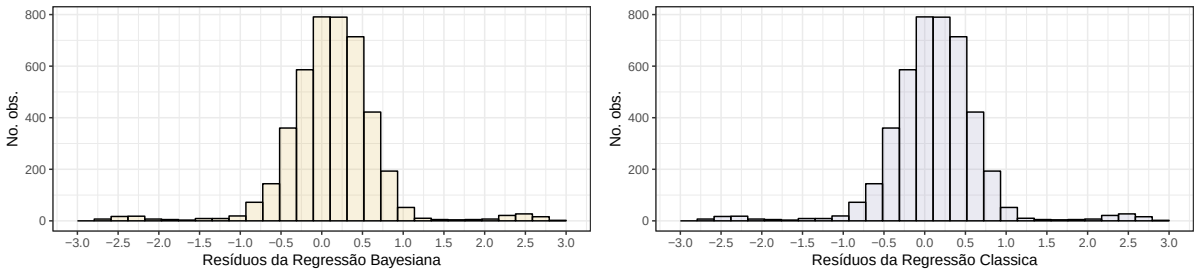
Fonte – Elaboração própria com base nos dados da RAIS.

O coeficiente associado ao gênero foi negativo, porém de magnitude ligeiramente menor (em módulo) que o original do trabalho de Bonini e Pozzobon (2016). Como a categoria basal é "homem", isso implica que mulheres tem um prêmio salarial menor, se comparado com trabalhadores do sexo masculino, mesmo controlando todas as outras variáveis. É interessante notar que esse efeito não era muito evidente na análise bivariada simples (Figura 6), reforçando a necessidade de inclusão de covariadas de controle no modelo de regressão.

Um método de verificação da qualidade de ajuste do modelo é a análise dos resíduos

da regressão: se o modelo consegue captar por completo todas as fontes de variação do logaritmo do salário hora, é esperado que as diferenças entre valores estimados e valores reais sejam um ruído branco gaussiano. A Figura (9) contém os histogramas dos resíduos tanto para o modelo bayesiano como para o modelo estimado por MQO. Ambas curvas aparentam ser simétricas e com formato de "sino", o que corresponde ao formato de um histograma de variáveis normais.

Figura 9 – Histograma dos resíduos dos modelos estimados com a metodologia bayesiana e por MQO (regressão clássica)

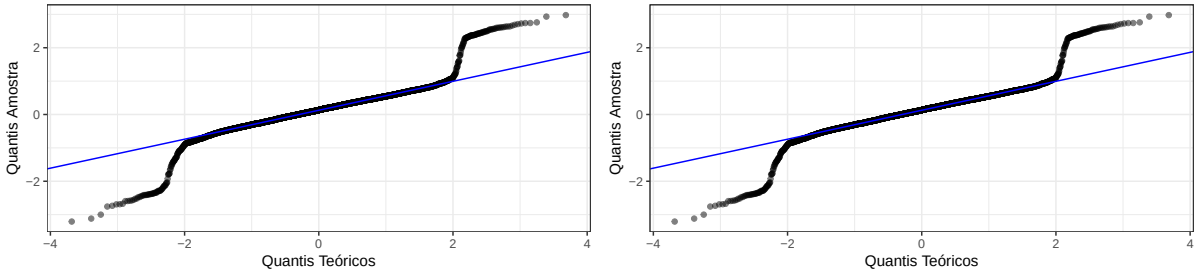


(a) Modelo Bayesiano (b) Modelo estimado com MQO

Fonte – Elaboração própria.

Embora os histogramas dos gráficos (9a) e (9b) pareçam simétricos em torno de zero, chama a atenção a persistência de valores nas extremidades, que é uma característica de distribuições com "caudas pesadas". O qqplot é um gráfico que compara os quantis teóricos de uma distribuição (neste caso, da distribuição normal) com os quantis da distribuição amostral obtida. No caso da distribuição normal, como seu histograma é similar ao de outras distribuições (em particular, ao da distribuição t e da distribuição de Cauchy), o qqplot é uma ferramenta útil para diagnóstico do modelo.

Figura 10 – qqplot dos resíduos dos modelos estimados com a metodologia bayesiana e por MQO (regressão clássica)



(a) Modelo Bayesiano (b) Modelo estimado com MQO

Fonte – Elaboração própria.

No qqplot (Figura 10), caso a distribuição amostral seja consistente com a distribuição teórica, os pontos estariam totalmente espalhados sobre a linha azul dos gráficos (10a) e (10b). Nota-se que de fato existem anomalias nas caudas, pois os pontos saem da linha azul em ambos métodos de estimação. Essa característica de caudas nos resíduos

é geralmente atribuída à presença de heterocedasticidade nos dados, isto é, os erros não seguem a mesma distribuição para todos os indivíduos. Como foi um dos pressupostos do modelo, isso gera má especificação do processo gerador dos dados. A presença de heterocedasticidade não causa viés de estimativas, mas compromete os testes de hipóteses dos coeficientes, no caso da estimação por MQO. Com relação à estimação Bayesiana, isso afeta o formato da verossimilhança. Além disso, em um modelo com matriz de covariância geral (que permite erros heterocedásticos), não existe densidade analítica para a posteriori, que terá outro formato. O capítulo 5 de Koop (2003) apresenta essa generalização do modelo de regressão, que não é objeto de estudo desta monografia.

4.5 PREVISÃO FORA DA AMOSTRA

Em modelos econométricos é comum, além do interesse nos parâmetros estimados (a fim de verificar as interações entre variáveis), realizar previsões para fora da amostra. A previsão é útil em situações que envolvem uma variável explicada de difícil obtenção/mensuração, mas que se relaciona com variáveis explicativas através de um processo que pode ser estimado por um modelo. Essas variáveis explicativas usualmente serão mais acessíveis em termos de coleta e podem então serem utilizadas para prever comportamentos da variável dependente quando novas observações surgem.

Para este trabalho realizou-se um exercício de previsão fora da amostra utilizando 10 indivíduos. Como existe uma grande proporção de indivíduos com características similares (homens com ensino superior completo e com menos de 30 anos), a amostra de 10 indivíduos foi selecionada por conveniência, uma vez que a seleção aleatória resultava sempre em um grupo pouco heterogêneo. A amostra foi separada do banco de dados antes do processo de estimação, com a finalidade de evitar excesso de ajuste dos dados. A tabela (7) contém as informações dos indivíduos selecionados.

Tabela 7 – Informações dos 10 indivíduos selecionados para previsão fora da amostra

Indivíduo	**ln(salário-hora)**	**Sup. Inc.**	**Superior**	**Pós-Grad.**	**Idade**	**Idade^2**	**T. Emprego**	**Gênero**
4010	4.65	Não	Sim	Não	29	841	31.1	Homem
1268	4.72	Não	Sim	Não	27	729	6.1	Homem
3634	4.77	Não	Sim	Não	25	625	17.9	Homem
810	3.90	Não	Sim	Não	20	400	16.9	Homem
3204	3.32	Sim	Não	Não	22	484	19.9	Mulher
1	3.47	Não	Não	Não	31	961	10.5	Homem
662	4.17	Não	Sim	Não	27	729	19.9	Homem
2015	3.77	Não	Sim	Não	23	529	19.2	Mulher
671	3.30	Não	Sim	Não	19	361	13.9	Homem
2413	6.70	Não	Não	Sim	37	1369	32.5	Mulher

Fonte – Elaboração própria com base na RAIS.

Foram realizadas previsões utilizando a densidade preditiva a posteriori descrita no capítulo 1 deste trabalho e também as previsões utilizando as estimativas de MQO

clássicas. A tabela (8) tem o valor real, valor predito e intervalo de credibilidade bayesianos e valores preditos e intervalos de confiança pelo método clássico.

Tabela 8 – Valores reais e previstos usando os dois métodos de previsão, para os 10 indivíduos fora da amostra

Indivíduo	Valor Real	Est. Bayesiana	Int. Cred. 95%	MQO	IC95%
4010	4.64957	4.30780	[3.6324 ; 4.9832]	4.30780	[3.1675 ; 5.4481]
1268	4.72352	4.15277	[3.4773 ; 4.8282]	4.15276	[3.0124 ; 5.2931]
3634	4.77206	4.08010	[3.4045 ; 4.7557]	4.08009	[2.9397 ; 5.2205]
810	3.89796	3.77543	[3.0990 ; 4.4519]	3.77539	[2.6343 ; 4.9165]
3204	3.32079	3.49967	[2.8229 ; 4.1765]	3.49965	[2.3582 ; 4.6411]
1	3.46981	3.96341	[3.2863 ; 4.6405]	3.96341	[2.8217 ; 5.1051]
662	4.17107	4.18760	[3.5121 ; 4.8631]	4.18759	[3.0473 ; 5.3279]
2015	3.76843	3.83802	[3.1618 ; 4.5142]	3.83800	[2.6971 ; 4.9789]
671	3.29927	3.69954	[3.0228 ; 4.3763]	3.69949	[2.5581 ; 4.8409]
2413	6.69687	5.13866	[4.4397 ; 5.8376]	5.13868	[3.9787 ; 6.2987]

Fonte – Elaboração própria.

As previsões usando ambos métodos são muito similares, em média, porém existem diferenças nos intervalores gerados pois os intervalos de credibilidade são mais estreitos que os intervalos de confiança, para esta amostra. É importante ressaltar que um intervalo de credibilidade não tem a mesma interpretação de um intervalo de confiança e portanto as comparações diretas devem ser feitas com parcimônia. Ambos modelos erraram na previsão subestimando o valor dos rendimentos do indivíduo 2413, possivelmente pelo fato da pessoa em questão se diferenciar da grande maioria da amostra por ter pós graduação, ser mulher, idade média maior (37 anos) e um tempo de emprego mais elevado (32.5 meses). Além disso, para o indivíduo 3634 o intervalo de credibilidade não tem interseção com o valor verdadeiro, porém a diferença do extremo superior do intervalo para o valor real é pequena.

5 CONSIDERAÇÕES FINAIS

Métodos de econometria bayesiana surgem como uma alternativa aos chamados modelos clássicos e são especialmente úteis pela flexibilidade na incorporação de informações a priori, atualização do modelo com novos dados e também no tratamento dos parâmetros como variáveis aleatórias. Em particular, métodos bayesianos permitem a realização de inferências inclusive para eventos que nunca foram observados, algo que a metodologia clássica não comporta.

A pesquisa realizada teve como objetivo investigar a metodologia bayesiana em modelos econométricos. Mais especificamente, foram apresentados os resultados para o modelo normal de regressão linear múltipla (MNRLC), que é um dos primeiros modelos a serem aprendidos em disciplinas de econometria básica. Todos os resultados obtidos foram comparados com as estimativas calculadas utilizando o estimador de mínimos quadrados ordinários, tanto para os resultados teóricos quanto os resultados de simulação e resultados da aplicação a dados reais.

Através da revisão bibliográfica do capítulo 2, foi feita toda a derivação do modelo utilizando a priori conjugada normal-gama, que leva à solução fechada (analítica) do modelo. Os resultados foram expressos utilizando as quantidades de MQO, o que permitiu tirar intuições importantes do modelo quanto ao comportamento para pequenas a grandes amostras, além da influência dos hiperparâmetros na determinação da posteriori. Embora as distribuições a posteriori sejam conhecidas na literatura da área, não foi encontrada nenhuma referência onde todos os passos dos cálculos estivessem descritos, de maneira que este trabalho constitui como um guia detalhado do MNRLC com a priori conjugada natural.

Os resultados das simulações no capítulo 3 foram consistentes com os resultados teóricos. Para grandes tamanhos de amostra de fato as estimativas bayesianas convergiram para as estimativas de MQO. Além disso, foram analisados os efeitos das alterações nas médias e variâncias a priori dos coeficientes: prioris com maior variância (mais difusas) acarretaram em maior erro quadrático médio, em geral. Na aplicação com erro de medida, foi observado que mesmo utilizando o valor verdadeiro dos coeficientes como priori, associado com uma baixa variância, não foi suficiente para evitar o viés das estimativas.

A aplicação feita no capítulo 4 envolve os dados de trabalhadores e trabalhadoras do setor de TI do Estado de Santa Catarina para o ano de 2011 e foi baseada no trabalho de Bonini e Pozzobon (2016). Os resultados estimados, utilizando a metodologia bayesiana, foram consistentes com o trabalho original, sendo que as estimativas bayesianas produziram intervalos de credibilidade mais estreitos que os intervalos de confiança das estimativas de

MQO. Foi explorada também a possibilidade de cálculos a posteriori, como a probabilidade de que os parâmetros sejam positivos. Todos os resultados obtidos estavam de acordo com aqueles obitidos por Bonini e Pozzobon (2016). Uma possível limitação do uso do modelo para estes dados é a matriz de covariância dos erros, que neste trabalho é assumida como homocedástica. Como visto na análise dos resíduos no capítulo 4, tanto para as estimativas de MQO como no modelo bayesiano, há um comportamento não gaussiano possível de ser observado nas caudas da distribuição dos erros, sugerindo a presença de heterocedasticidade.

Como limitações do trabalho, não foram exploradas outras possibilidades de priori além da conjugada natural. Apesar desta priori possuir como vantagem os resultados a posteriori em fórmula fechada, ela limita a especificação dos parâmetros β e h, que não pode ser feita de forma independente. Como mostrado no capítulo 2, a priori do modelo pode ser fatorada como o produto $p(\beta, h) = p(\beta|h)p(h)$, de forma que se a precisão h for muito pequena, essa incerteza estará contida em $p(\beta|h)$ e a priori dos β necessariamente será menos informativa. Especificações possíveis para a priori no MNRL e que não foram apresentadas é a priori normal gama independente (que resolve o problema citado anteriormente, mas não tem fórmula fechada da posteriori) ou ainda uma priori totalmente difusa, onde não é dado nenhum peso para a priori e a posteriori colapsa exatamente nas quantidades de MQO.

Além das prioris acima citadas, também não foi explorada a questão da comparação de modelos utilizando a razão de chances, que compara a probabilidade de dois modelos diferentes serem os verdadeiros dado o que foi observado nos dados. Essas especificações podem ser exploradas em trabalhos futuros, além do modelo com matriz de covariância geral dos erros, que demanda o uso de métodos como o Amostrador de Gibbs ou Metropolis Hastings para estimativa dos parâmetros.

REFERÊNCIAS

BAUWENS, L.; LUBRANO, M.; RICHARD, J.-F. *Bayesian inference in dynamic econometric models*. [S.l.]: Univ. Press, 2003. Citado 6 vezes nas páginas 19, 30, 32, 34, 35 e 72.

BONINI, P.; POZZOBON, F. Discriminação salarial feminina e o prêmio salarial de ti na indústria. *Análise Econômica (UFRGS)*, v. 34, p. 193–223, 2016. Citado 11 vezes nas páginas 7, 16, 17, 45, 46, 47, 51, 52, 53, 57 e 58.

CALDEIRA, J. F.; MOURA, G. V.; SANTOS, A. A. Previsões macroeconômicas baseadas em modelos tvp-var: evidências para o brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 69, n. 4, p. 407–428, 2015. Citado na página 19.

CASELLA, G.; BERGER, R. L. *Statistical inference*. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2. Citado 3 vezes nas páginas 19, 20 e 27.

COGLEY, T.; SARGENT, T. J. Evolving post-world war ii us inflation dynamics. *NBER macroeconomics annual*, MIT Press, v. 16, p. 331–373, 2001. Citado na página 19.

COGLEY, T.; SARGENT, T. J. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics*, Elsevier, v. 8, n. 2, p. 262–302, 2005. Citado na página 19.

DAVIDSON, R.; MACKINNON, J. G. *Estimation and inference in econometrics*. 1. ed. [S.l.: s.n.], 1993. Citado 2 vezes nas páginas 19 e 21.

DAVIDSON, R.; MACKINNON, J. G. *Econometric Theory and Methods*. 1. ed. [S.l.]: Oxford University Press, 2004. Citado na página 29.

De Castro, M. R. et al. Samba: Stochastic analytical model with a bayesian approach. *Brazilian Review of Econometrics*, v. 35, n. 2, p. 103–170, 2011. Citado na página 19.

DEL NEGRO, M.; SCHORFHEIDE, F. Bayesian macroeconometrics. In: GEWEKE, J.; KOOP, G.; VAN DIJK, H. (Ed.). *The Oxford Handbook of Bayesian Econometrics*. Oxford: Oxford University Press, 2012. cap. 7. Citado na página 13.

DOAN, T.; LITTERMAN, R.; SIMS, C. Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, Taylor & Francis, v. 3, n. 1, p. 1–100, 1984. Citado na página 19.

GEWEKE, J. Using simulation methods for bayesian econometric models: inference, development,and communication. *Econometric Reviews*, v. 18, n. 1, p. 1–73, 1999. Citado 2 vezes nas páginas 13 e 14.

GEWEKE, J. Bayesian econometrics and forecasting. *Journal of Econometrics*, Elsevier, v. 100, n. 1, 2001. Citado na página 18.

GEWEKE, J. *Contemporary Bayesian econometrics and statistics*. [S.l.]: John Wiley & Sons, 2005. v. 537. Citado na página 52.

GEWEKE, J.; KOOP, G.; Van Dijk, H. (Ed.). *The Oxford handbook of Bayesian econometrics*. [S.l.]: Oxford University Press, 2011. Citado na página 19.

GREENBERG, E. *Introduction to Bayesian econometrics*. [S.l.]: Cambridge University Press, 2014. Citado 2 vezes nas páginas 13 e 14.

GREENE, W. H. *Econometric analysis*. 7. ed. [S.l.: s.n.], 2012. Citado na página 21.

GUJARATI, D. N. *Econometria Básica*. 4. ed. [S.l.: s.n.], 2006. Citado 2 vezes nas páginas 36 e 41.

KAHN, S.; GINTHER, D. *Women and STEM*. [S.l.], 2017. Citado na página 45.

KASTNER, G.; FRÜHWIRTH-SCHNATTER, S.; LOPES, H. F. Efficient bayesian inference for multivariate factor stochastic volatility models. *Journal of Computational and Graphical Statistics*, Taylor & Francis, n. just-accepted, 2017. Citado na página 19.

KOOP, G. *Bayesian econometrics*. [S.l.]: Wiley, 2003. Citado 12 vezes nas páginas 16, 19, 24, 29, 30, 31, 32, 34, 35, 36, 51 e 55.

KOOP, G.; POIRIER, D. J.; TOBIAS, J. L. *Bayesian econometric methods*. [S.l.]: Cambridge University Press, 2007. Citado 3 vezes nas páginas 15, 19 e 30.

MCGRAYNE, S. B. *A teoria que não morreria: como a lei de Bayes decifrou o código Enigma, perseguiu submarinos russos e emergiu triunfante de dois séculos de controvérsias*. 1. ed. São Paulo: [s.n.], 2015. 480 p. Citado na página 18.

MITTELHAMMER, R. C. *Mathematical statistics for economics and business*. 2. ed. [S.l.]: Springer, 2013. Citado na página 20.

MORGAN, M. S. *The history of econometric ideas*. [S.l.]: University Press, 2003. Citado na página 13.

PAULINO, C. D. M.; AMARAL., T. M. A.; MURTEIRA, B. *Estatística Bayesiana*. [S.l.]: Fundação Calouste Gulbenkian, 2003. Citado na página 13.

PHILIPOV, A.; GLICKMAN, M. E. Multivariate stochastic volatility via wishart processes. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 24, n. 3, p. 313–328, 2006. Citado na página 19.

POIRIER, D. J. A report from the battlefront. *Journal of Business & Economic Statistics*, v. 7, n. 1, p. 137, 1989. Citado na página 16.

POIRIER, D. J. The growth of bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, v. 1, n. 4, p. 969–979, 2006. Citado na página 16.

POZRIKIDIS, C. *An introduction to grids, graphs, and networks*. [S.l.]: Oxford University Press, 2014. Citado na página 73.

ROBERT, C. P. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. [S.l.]: Springer, 2007. Citado 3 vezes nas páginas 19, 21 e 52.

SEARLE, S. R. *Matrix algebra useful for statistics*. [S.l.]: John Wiley & Sons, 1982. Citado 3 vezes nas páginas 68, 74 e 75.

WOOLDRIDGE, J. M. *Introductory econometrics: A modern approach*. [S.l.]: Nelson Education, 2015. Citado 2 vezes nas páginas 28 e 29.

ZELLNER, A. *An introduction to Bayesian inference in econometrics*. [S.l.: s.n.], 1971. Citado na página 18.

ZELLNER, A. Applications of bayesian analysis in econometrics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, JSTOR, v. 32, p. 23–34, 1983. Citado na página 19.

ZELLNER, A. Bayesian econometrics. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 253–269, 1985. Citado na página 18.

Apêndices

APÊNDICE A – DEMONSTRAÇÕES DOS PRINCIPAIS RESULTADOS APRESENTADOS NO CAPÍTULO 2

A VEROSSIMILHANÇA

Os próximos passos mostram que de fato a função de verossimilhança dada na equação (2.24) é o produto de uma densidade Normal para $\beta|h$ vezes o núcleo de uma densidade gama para h.

O primeiro passo é reescrever o termo $(Y - X\beta)'(Y - X\beta)$ de (2.28) somando e subtraindo $X\hat{\beta}$:

$$\begin{aligned}(Y - X\beta)'(Y - X\beta) = \\ &= (Y - X\beta + X\hat{\beta} - X\hat{\beta})'(Y - X\beta + X\hat{\beta} - X\hat{\beta}) \\ &= ((Y - X\hat{\beta}) - (X\beta - X\hat{\beta}))'((Y - X\hat{\beta}) - (X\beta - X\hat{\beta})) \\ &= ((Y - X\hat{\beta})' - (\beta - \hat{\beta})'X')((Y - X\hat{\beta}) - X(\beta - \hat{\beta})) \qquad \text{(A.1)}\end{aligned}$$

Então, faz-se a distributiva para obter:

$$(Y-X\hat{\beta})'(Y-X\hat{\beta})-(Y-X\hat{\beta})'X(\beta-\hat{\beta})-(\beta-\hat{\beta})'X'(Y-X\hat{\beta})+(\beta-\hat{\beta})'X'X(\beta-\hat{\beta}) \quad \text{(A.2)}$$

Os termos centrais de (A.2) são iguais a zero pois a expressão $(Y - X\hat{\beta})'X$ é igual a zero (consequentemente, seu transposto também é). Isto decorre da hipótese de ortogonalidade entre o vetor de resíduos ε e a matriz X.

De forma que conclui-se que:

$$(Y - X\beta)'(Y - X\beta) = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \qquad \text{(A.3)}$$

O próximo passo é mostrar que o termo $(Y - X\hat{\beta})'(Y - X\hat{\beta})$, que é a soma de quadrado dos resíduos de MQO (denotada por SQR), pode ser escrito como $SQR = Y'Y - \hat{\beta}X'X\hat{\beta}$. Primeiro, aplica-se a distributiva para obter $(Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - Y'X\hat{\beta} - (X\hat{\beta})'Y + (X\hat{\beta})'(X\hat{\beta})$. Uma vez que $Y'X\hat{\beta}$ é um produto interno e pode ser denotado por $\langle Y, X\hat{\beta}\rangle$, valem as propriedades de comutatividade, distributividade, bilinearidade, multiplicação por escalar e associatividade no produto por escalar.

Utilizando a comutatividade, tem-se:

$$\begin{aligned}
-Y'X\hat{\beta} - (X\hat{\beta})'Y &= -\langle Y, X\hat{\beta}\rangle - \langle X\hat{\beta}, Y\rangle \\
&= -\langle X\hat{\beta}, Y\rangle - \langle X\hat{\beta}, Y\rangle \\
&= -Y'X\hat{\beta} - Y'X\hat{\beta} \\
&= -2Y'X\hat{\beta} \qquad \text{(A.4)}
\end{aligned}$$

De forma que:

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \qquad \text{(A.5)}$$

O último termo pode ser reescrito da seguinte forma:

$$\begin{aligned}
\hat{\beta}X'X\hat{\beta} &= \\
&= \hat{\beta}X'X\overbrace{(X'X)^{-1}X'Y}^{\hat{\beta}} \\
&= \hat{\beta}'X'Y \\
&= (X\hat{\beta})'Y \\
&= Y'X\hat{\beta} \qquad \text{(A.6)}
\end{aligned}$$

Fazendo a substituição em (A.5), tem-se:

$$\begin{aligned}
(Y - X\hat{\beta})'(Y - X\hat{\beta}) &= \\
&= Y'Y - 2Y'X\hat{\beta} + Y'X\hat{\beta} \\
&= Y'Y - Y'X\hat{\beta} \\
&= \underbrace{Y'Y - \hat{\beta}'X'X\hat{\beta}}_{SQR} \qquad \text{(A.7)}
\end{aligned}$$

Logo, pode-se escrever a equação (A.3) como:

$$(Y - X\beta)'(Y - X\beta) = Y'Y - \hat{\beta}'X'X\hat{\beta} + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \qquad \text{(A.8)}$$

E a verossimilhança descrita em (2.24) será:

$$p(Y|\beta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} exp\left\{-\frac{h}{2}\left[\nu s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right]\right\} \qquad \text{(A.9)}$$

Utilizando o fato de que $\nu = N - k$, pode-se escrever $h^{\frac{N}{2}} = h^{\frac{\nu+k}{2}} = h^{\frac{\nu}{2}} \cdot h^{\frac{k}{2}}$. Além disso, pode-se escrever s^2 como $\frac{1}{s^{-2}}$ para reorganizar (A.9) e chegar na equação (2.28).

A DENSIDADE A POSTERIORI

A densidade a posteriori, por definição, é proporcional ao produto da verossimilhança e da priori. Utilizando (2.5) junto com as expressões da verossimilhança (2.28) e da priori (2.29), tem-se:

$$
\begin{aligned}
p(\beta, h|y) &\propto p(Y|\beta, h)p(\beta, h) \\
&= \left[\frac{1}{(2\pi)^{\frac{N}{2}}} h^{\frac{k}{2}} \exp\left[-\frac{h}{2}\underbrace{(\beta-\hat{\beta})'X'X(\beta-\hat{\beta})}_{(*)}\right]\right]\left[h^{\frac{\nu}{2}} \exp\left[-\frac{h}{2}\frac{\nu}{s^{-2}}\right]\right] \\
&\quad \cdot \chi h^{\frac{k+\underline{\nu}-2}{2}} \exp\left[-\frac{h}{2}(\underline{\nu}\underline{s}^2 + \underbrace{(\beta-\underline{\beta})'\underline{V}^{-1}(\beta-\underline{\beta})}_{(**)})\right]
\end{aligned} \tag{A.10}
$$

O termo χ é a constante da distribuição Gama, dada por:

$$
\chi = (2\pi)^{-\frac{k}{2}} \left[\left(2\frac{\underline{s}^{-2}}{\underline{\nu}}\right)^{\frac{\underline{\nu}}{2}} \Gamma\left(\frac{\underline{\nu}}{2}\right)\right]^{-1} |\underline{V}|^{-\frac{1}{2}}
$$

onde $\Gamma(\cdot)$ é a função gama.

Uma vez que os termos $(*)$ e $(**)$ de (A.10) são expoentes de mesma base, podem ser somados:

$$
\begin{aligned}
&(\beta-\hat{\beta})'X'X(\beta-\hat{\beta}) + (\beta-\underline{\beta})'\underline{V}^{-1}(\beta-\underline{\beta}) = \\
&= \beta'X'X\beta - \beta'X'X\hat{\beta} - \hat{\beta}'X'X\beta + \hat{\beta}'X'X\hat{\beta} + \beta'\underline{V}^{-1}\beta - \beta'\underline{V}^{-1}\underline{\beta} - \underline{\beta}'\underline{V}^{-1}\beta + \underline{\beta}'\underline{V}^{-1}\underline{\beta} \\
&= \underbrace{\beta'\left[\underline{V}^{-1} + X'X\right]\beta - \beta'\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right] - \left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]'\beta}_{(***)} + \underbrace{\hat{\beta}'X'X\hat{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{\beta}}_{(\bigstar)}
\end{aligned} \tag{A.11}
$$

O termo $(***)$ é uma forma quadrática do tipo $\beta'A\beta - \beta'b - b'\beta$, onde $A = [\underline{V}^{-1} + X'X]$ e $b = \left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]$. Somando e diminuindo $b'A^{-1}b$ e ignorando temporariamente o termo negativo, $-b'A^{-1}b$, pode-se escrever (***) como:

$$
\begin{aligned}
&\beta'\left[\underline{V}^{-1} + X'X\right]\beta - \beta'\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right] - \left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]'\beta \\
&\quad + \left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]'\left[\underline{V}^{-1} + X'X\right]^{-1}\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]
\end{aligned} \tag{A.12}
$$

De forma que (A.12) é um termo do tipo $\beta'A\beta - \beta'b - b'\beta + b'A^{-1}b$. Como A é a soma de uma matriz de variância de posto completo e $X'X$ é simétrica (também de posto completo), vale que $A^{-1}A = AA^{-1} = I$ e também $A' = A$. Portanto:

$$\beta' A\beta - \beta' b - b'\beta + b' A^{-1} b =$$
$$\beta' A\beta - \beta' A A^{-1} b - b' A^{-1} A\beta + b' A^{-1} A A^{-1} b \tag{A.13}$$

Agora, considerando $\overline{V} = A^{-1}$ e $\overline{\beta} = A^{-1}b$, a expressão (A.13) será $\beta'\overline{V}^{-1}\beta - \beta'\overline{V}\overline{\beta} - \overline{\beta}'\overline{V}\beta + \overline{\beta}'\overline{V}\overline{\beta}$, que pode ser escrito como $(\beta - \overline{\beta})'\overline{V}^{-1}(\beta - \overline{\beta})$, onde $\overline{V} = A^{-1} = [\underline{V}^{-1} + X'X]^{-1}$ e $\overline{\beta} = A^{-1}b = \overline{V}\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]$.

Resta trabalhar com (★) de (A.11) junto com o termo $-b'A^{-1}b$ que havia sido desconsiderado até então. O objetivo dos próximos quatro passos é escrever $\hat{\beta}'X'X\hat{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - b'A^{-1}b$ como $(\hat{\beta} - \beta)'[\underline{V} + (X'X)^{-1}]^{-1}(\hat{\beta} - \beta)'$. Utilizando a definição de b, tem-se:

$$\hat{\beta}'X'X\hat{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - b'A^{-1}b =$$
$$\hat{\beta}'X'X\hat{\beta} + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - \underbrace{\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]'\left[\underline{V}^{-1} + X'X\right]^{-1}\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]}_{(a)} \tag{A.14}$$

Passo 1: Reescrever o termo (a) da equação (A.14)
Usando a definição de $\overline{V}$ em (a) e fazendo a distributiva, obtem-se:

$$\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right]'\overline{V}\left[\underline{V}^{-1}\underline{\beta} + X'X\hat{\beta}\right] =$$
$$= \beta'\left(\underline{V}^{-1}\right)'\overline{V}\underline{V}^{-1}\underline{\beta} + \hat{\beta}'X'X\overline{V}\underline{V}^{-1}\underline{\beta} + (\underline{V}^{-1}\underline{\beta})'\overline{V}X'X\hat{\beta} + \hat{\beta}'X'X\overline{V}X'X\hat{\beta} \tag{A.15}$$

Então, (A.14), após rearranjar os termos e substituir (a) por (A.15) será:

$$\hat{\beta}'\underbrace{\left[X'X - X'X\overline{V}X'X\right]}_{(b)}\hat{\beta} + \underline{\beta}'\underbrace{\left[\underline{V}^{-1} - \underline{V}^{-1}\overline{V}\underline{V}^{-1}\right]}_{(c)}\underline{\beta} - \underline{\beta}\underline{V}^{-1}\overline{V}X'X\hat{\beta} - \hat{\beta}'X'X\overline{V}\underline{V}^{-1}\underline{\beta} \tag{A.16}$$

Onde foi usado o fato de que $\underline{V}$ é uma matriz simétrica (pois é uma matriz de covariância), de maneira que $\underline{V}'^{-1} = \underline{V}^{-1}$.

Passo 2: Mostrar que o termo (b) de (A.16) é igual a $X'X\overline{V}\underline{V}^{-1}$.

$$\mathbb{I} = [\underline{V}^{-1} + X'X]^{-1}[\underline{V}^{-1} + X'X]$$
$$\mathbb{I} = [\underline{V}^{-1} + X'X]^{-1}\underline{V}^{-1} + [\underline{V}^{-1} + X'X]^{-1}X'X$$
$$\mathbb{I} - [\underline{V}^{-1} + X'X]^{-1}X'X = [\underline{V}^{-1} + X'X]^{-1}\underline{V}^{-1}$$
$$\mathbb{I} - \overline{V}X'X = \overline{V}\underline{V}^{-1}$$
$$X'X(\mathbb{I} - \overline{V}X'X) = X'X(\overline{V}\underline{V}^{-1}) \tag{A.17}$$

Com $\mathbb{I}$ correspondendo à matriz identidade.

Passo 3: Mostrar que $\underline{V}^{-1} - \underline{V}^{-1}\overline{V}\underline{V}^{-1} = \underline{V}^{-1}\overline{V}X'X$

$$\begin{aligned}
\mathbb{I} &= \overline{V}\overline{V}^{-1} = \\
&= \overline{V}\left[\underline{V}^{-1} + X'X\right] \\
&= \overline{V}\underline{V}^{-1} + \overline{V}X'X
\end{aligned}$$

Juntando as duas extremidades, tem-se que:

$$\begin{aligned}
\mathbb{I} &= \overline{V}\underline{V}^{-1} + \overline{V}X'X \\
\mathbb{I} - \overline{V}\underline{V}^{-1} &= \overline{V}X'X \\
\underline{V}\underline{V}^{-1}\left[\mathbb{I} - \overline{V}\underline{V}^{-1}\right] &= \underline{V}\underline{V}^{-1}\overline{V}X'X \\
\underline{V}^{-1}\left[\mathbb{I} - \overline{V}\underline{V}^{-1}\right] &= \underline{V}^{-1}\overline{V}X'X \\
\underline{V}^{-1} - \underline{V}^{-1}\overline{V}\underline{V}^{-1} &= \underline{V}^{-1}\overline{V}X'X
\end{aligned}$$

Passo 4: Mostrar que o termo $X'X\overline{V}\underline{V}^{-1}$ (do passo 3) é igual a $(\underline{V} + (X'X)^{-1})^{-1}$.

$$\begin{aligned}
X'X\overline{V}\underline{V}^{-1} &= X'X\left[\underline{V}^{-1} + X'X\right]^{-1}\underline{V}^{-1} = \\
&= X'X\left[\underline{V}^{-1} + ((X'X)^{-1})^{-1}\right]^{-1}\underline{V}^{-1} \\
&= X'X(X'X)^{-1}(\underline{V} + (X'X)^{-1})^{-1}\underline{V}\underline{V}^{-1} \\
&= (\underline{V} + (X'X)^{-1})^{-1}
\end{aligned}$$

Na penúltima etapa do passo 4 foi utilizado no termo entre colchetes o fato que se as inversas estão bem definidas, então $(A^{-1} + B^{-1})^{-1} = A(A+B)^{-1}B = B(A+B)^{-1}A$ (onde $A = \underline{V}$ e $B = (X'X)^{-1}$) (SEARLE, 1982). Isso pode ser verificado multiplicando um termo pelo inverso do outro, para obter a matriz identidade:

$$\begin{aligned}
(A^{-1} + B^{-1})A(A+B)^{-1}B &= \\
&= A^{-1}A(A+B)^{-1}B + B^{-1}A(A+B)^{-1}B \\
&= (A+B)^{-1}B + B^{-1}A(A+B)^{-1}B \\
&= (\mathbb{I} + B^{-1}A)(A+B)^{-1}B \\
&= (B^{-1}B + B^{-1}A)(A+B)^{-1}B \\
&= B^{-1}(B+A)(A+B)^{-1}B \\
&= B^{-1}\mathbb{I}B \\
&= B^{-1}B = \mathbb{I}
\end{aligned}$$

Usando o mesmo resultado, conclui-se que $\underline{V}^{-1}\left[\underline{V}^{-1} + X'X\right]^{-1}X'X = [\underline{V} + (X'X)^{-1}]^{-1}$.

Juntando os passos 1, 2, 3 e 4, pode-se reescrever a equação (A.16) como:

$$\begin{aligned}
&\hat{\beta}' \overbrace{X'X\bar{V}\underline{V}^{-1}}^{X'X-X'X\bar{V}X'X} \hat{\beta} + \underline{\beta}' \overbrace{\underline{V}^{-1}\bar{V}X'X}^{\underline{V}^{-1}-\underline{V}^{-1}\bar{V}\underline{V}^{-1}} \underline{\beta} - \underline{\beta}\underline{V}^{-1}\bar{V}X'X\hat{\beta} - \hat{\beta}'X'X\bar{V}\underline{V}^{-1}\underline{\beta} = \\
&= \hat{\beta}' \left[\underline{V} + (X'X)^{-1}\right]^{-1} \hat{\beta} + \underline{\beta}' \left[\underline{V} + (X'X)^{-1}\right]^{-1} \underline{\beta} + \underline{\beta}' \left[\underline{V} + (X'X)^{-1}\right]^{-1} \hat{\beta} + \hat{\beta}' \left[\underline{V} + (X'X)^{-1}\right]^{-1} \\
&= \left(\hat{\beta} - \underline{\beta}\right)' \left[\underline{V} + (X'X)^{-1}\right]^{-1} (\hat{\beta} - \underline{\beta})
\end{aligned}$$

Com os resultados acima, o núcleo da densidade a posteriori conjunta para β e h dado y pode ser escrita como:

$$p(\beta, h|y) \propto \bar{\chi} \cdot \exp\left\{-\frac{h}{2}\left[(\beta - \bar{\beta})'\bar{V}^{-1}(\beta - \bar{\beta}) + (\hat{\beta} - \underline{\beta})'\bar{V}^{-1}(\hat{\beta} - \underline{\beta}) + \nu s^2 + \underline{\nu}\underline{s}^2\right]\right\} \tag{A.18}$$

Onde $\bar{\chi} = \frac{h^{\frac{\nu+k}{2}}}{(2\pi)^{\frac{N}{2}}}\chi h^{\frac{k+\nu-2}{2}}$ e χ foi definido previamente.

A DENSITADE A POSTERIORI MARGINAL PARA β

A densidade posterior marginal de β, denotada por $p(\beta|y)$ é obtida integrando a densidade à posteriori conjunta $p(\beta, h|y)$ em relação a h.

$$\int p(\beta, h|y)dh = \bar{\chi}\int h^{\frac{\bar{\nu}+k-2}{2}} \exp\left[-\frac{h}{2}\left(\bar{\nu}\bar{s}^2 + (\beta - \bar{\beta})'V^{-1}(\beta - \bar{\beta})\right)\right] dh \tag{A.19}$$

Definindo $\nu^* \equiv \bar{\nu} + k$ e $\mu^* \equiv \left(\nu^*(\bar{\nu}\bar{s}^2 + (\beta - \bar{\beta})'V^{-1}(\beta - \bar{\beta}))\right)^{-1}$, (A.19) pode ser escrita como o núcleo de uma densidade gama com vetor de parâmetros (ν^*, μ^*) vezes a constante $\bar{\chi}$:

$$\bar{\chi}\int h^{\frac{\bar{\nu}+k-2}{2}} \exp\left(-\frac{h}{2}\left[\frac{\nu^*}{\mu^*}\right]\right) dh \tag{A.20}$$

Sabendo que $\int \mathcal{G}(\nu^*, \mu^*)dh = 1$ e multiplicando apenas o conteúdo da integral em (A.20) pelo inverso da constante de integração da densidade gama, obtem-se:

$$\int \underbrace{\left[\left(\frac{2\mu^*}{\nu^*}\right)^{\frac{\nu^*}{2}} \Gamma\left(\frac{\nu^*}{2}\right)\right]^{-1}}_{a} h^{\frac{\nu^*-2}{2}} \exp\left[-\frac{h}{2}\frac{\nu^*}{\mu^*}\right] dh = 1 \tag{A.21}$$

Logo,

$$\int \underbrace{h^{\frac{\nu^*-2}{2}} \exp\left[-\frac{h}{2}\frac{\nu^*}{\mu^*}\right]}_{\text{Núcleo da gama}} dh = \frac{1}{a^{-1}} = a \tag{A.22}$$

Isto é:

$$\int h^{\frac{\nu^*-2}{2}} \exp\left[-\frac{h}{2}\frac{\nu^*}{\mu^*}\right] = \left[\left(\frac{2\mu^*}{\nu^*}\right)^{\frac{\nu^*}{2}} \Gamma\left(\frac{\nu^*}{2}\right)\right] \tag{A.23}$$

Usando (A.23) em (A.20), obtem-se:

$$p(\beta|y) = \bar{\chi}\left[\left(\frac{2\mu^*}{\nu^*}\right)^{\frac{\nu^*}{2}} \Gamma\left(\frac{\nu^*}{2}\right)\right] = \frac{(2\pi)^{-\frac{k}{2}}|\bar{V}|^{-\frac{1}{2}}}{\Gamma\left(\frac{\bar{\nu}}{2}\right)/\Gamma\left(\frac{\nu^*}{2}\right)}\left[\frac{\left(\frac{2}{\bar{\nu}\bar{s}^2+\bar{Q}}\right)^{\frac{\bar{\nu}+k}{2}}}{\left(\frac{2}{\bar{\nu}\bar{s}^2}\right)^{\frac{\bar{\nu}}{2}}}\right] \tag{A.24}$$

Onde $\bar{Q} = (\beta - \bar{\beta})'\bar{V}^{-1}(\beta - \bar{\beta})$.

Trabalhando apenas a parcela dentro de colchetes na expressão (A.24), tem-se:

$$\left[\frac{\left(\frac{2}{\bar{\nu}\bar{s}^2+\bar{Q}}\right)^{\frac{\bar{\nu}+k}{2}}}{\left(\frac{2}{\bar{\nu}\bar{s}^2}\right)^{\frac{\bar{\nu}}{2}}}\right] = \left[\frac{\left(\frac{1}{\bar{\nu}\bar{s}^2+\bar{Q}}\right)^{\frac{\bar{\nu}+k}{2}}(2)^{\frac{\bar{\nu}+k}{2}}}{\left(\frac{1}{\bar{\nu}\bar{s}^2}\right)^{\frac{\bar{\nu}}{2}}(2)^{\frac{\bar{\nu}}{2}}}\right]$$

Juntando a última expressão com o termo $(2\pi)^{-\frac{k}{2}}$, (A.24) pode ser escrita como:

$$\frac{\pi^{-\frac{k}{2}}\Gamma\left(\frac{\bar{\nu}+k}{2}\right)}{|\bar{V}|^{\frac{1}{2}}\Gamma\left(\frac{\bar{v}}{2}\right)}\left[\frac{(\bar{\nu}\bar{s}^2)^{\frac{\bar{\nu}}{2}}}{(\bar{v}\bar{s}^2+\bar{Q})^{\frac{\bar{\nu}+k}{2}}}\right] = \frac{\pi^{-\frac{k}{2}}\Gamma\left(\frac{\bar{\nu}+k}{2}\right)}{|\bar{V}|^{\frac{1}{2}}\Gamma\left(\frac{\bar{v}}{2}\right)}\left[\frac{(\bar{s}^2)^{\frac{\bar{\nu}}{2}}(\nu)^{\frac{\bar{\nu}}{2}}}{(\bar{s}^2)^{\frac{\bar{\nu}+k}{2}}\left(\bar{\nu}+\frac{\bar{Q}}{\bar{s}^2}\right)}\right]$$

Que é igual a:

$$\frac{\bar{\nu}^{\frac{\bar{\nu}}{2}}\Gamma\left(\frac{\bar{\nu}+k}{2}\right)}{\pi^{\frac{k}{2}}\Gamma\left(\frac{\bar{\nu}}{2}\right)}|\bar{s}^{-2}\bar{V}|^{-\frac{1}{2}}\left[\bar{\nu}+(\beta-\bar{\beta})'(\bar{s}^{-2}\bar{V})^{-1}(\beta-\bar{\beta})\right]^{-\frac{\bar{\nu}+k}{2}} \tag{A.25}$$

E a expressão acima é a de uma variável aleatória que segue uma distribuição $t(\bar{\beta}, \bar{s}^2\bar{V}, \bar{\nu})$.

A DENSIDADE PREDITIVA A POSTERIORI

Usando o mesmo procedimento já apresentado no texto, a verossimilhança dos novos dados pode ser escrita como:

$$\begin{aligned} p(y^*|\beta, h) &= \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\exp\left\{-\frac{h}{2}(y^*-X^*\beta)'(y^*-X^*\beta)\right\} \\ &= \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}}\exp\left\{-\frac{h}{2}\left[\nu s^2-(\beta-\hat{\beta})'X^{*'}X^*(\beta-\hat{\beta})\right]\right\} \end{aligned} \tag{A.26}$$

Enquanto que a posteriori dada em (2.30) pode ser escrita como:

$$p(\beta,h|y) = \underbrace{\frac{|\bar{\nu}|^{-1/2}}{\left(\frac{2\bar{s}^{-2}}{\bar{\nu}}\right)^{\bar{\nu}^2/2}\Gamma\left(\frac{\bar{\nu}}{2}\right)(2\pi)^{k/2}}}_{\bar{\chi}} h^{\frac{\bar{\nu}+k-2}{2}} \exp\left\{-\frac{h}{2}\left(\bar{\nu}\bar{s}^2 + (\beta-\bar{\beta})'\bar{V}^{-1}(\beta-\bar{\beta})\right)\right\}$$

$$= \bar{\chi}\ \underbrace{h^{\frac{\bar{\nu}+k-2}{2}} \exp\left\{-\frac{h}{2}\frac{\bar{\nu}}{\bar{s}^{-2}}\right\}}_{\textbf{(a)}} \underbrace{\exp\left\{-\frac{h}{2}(\beta-\bar{\beta})'\bar{V}^{-1}(\beta-\bar{\beta})\right\}}_{\text{(b)}} \tag{A.27}$$

O termo (a) de (A.27) é quase[1] o núcleo de uma distribuição Gama($\bar{s}^2$, $\bar{\nu}$) para h e o termo em (b) é o núcleo de uma densidade Normal com parâmetros $\bar{\beta}$ e $h\bar{V}$ para $\beta|h$.

Os próximos passos irão mostrar que $y^*|y \sim t\left(X^*\bar{\beta}, \bar{s}^2\left\{\mathbb{I}_N + X^*\bar{V}X^{*'}\right\}, \bar{\nu}\right)$ (onde $\mathbb{I}_N$ é a matriz identidade de dimensão N). Por definição, isso significa que:

$$p(y^*|y) = \frac{1}{C_t}\left|\bar{s}^2\left(\mathbb{I}_N + X^*\bar{V}X^{*'}\right)\right|^{-\frac{1}{2}}$$
$$\left[\bar{\nu} + (y^* - X^*\bar{\beta})'\left(\bar{s}^2\left(\mathbb{I}_N + X^*\bar{V}X^{*'}\right)\right)^{-1}(y^* - X^*\bar{\beta})\right]^{-\left(\frac{\bar{\nu}+N}{2}\right)} \tag{A.28}$$

Onde $C_t = \frac{\pi^{K/2}\Gamma\frac{\bar{\nu}}{2}}{\bar{\nu}^{\bar{\nu}/2}\Gamma\left(\frac{\bar{\nu}+N}{2}\right)}$, que é a constante da distribuição t.

Utilizando (A.26) e (A.27) em (2.33), tem-se que:

$$p(y^*|y) = \int\int \frac{h^{N/2}}{(2\pi)^{N/2}}\bar{\chi}\exp\left\{-\frac{h}{2}(y^* - X^*\beta)'(y^* - X^*\beta)\right\} h^{\frac{\bar{\nu}+k-2}{2}}$$
$$\exp\left\{-\frac{h}{2}\left(\bar{\nu}\bar{s}^2 + (\beta-\bar{\beta})'\bar{V}^{-1}(\beta-\bar{\beta})\right)\right\} dh\ d\beta$$
$$= \int\int \frac{h^{\frac{N+\bar{\nu}+k-2}{2}}}{(2\pi)^{N/2}}\bar{\chi}\exp\left\{-\frac{h}{2}\bar{\nu}\bar{s}^2\right\}$$
$$\exp\left\{-\frac{h}{2}\underbrace{\left[(y^* - X^*\beta)'(y^* - X^*\beta) + (\beta-\bar{\beta})'\bar{V}^{-1}(\beta-\bar{\beta})\right]}_{(*)}\right\} dh\ d\beta \tag{A.29}$$

Trabalhando com o termo (*) de (A.29), obtem-se:

$$\beta'\underbrace{(\bar{V}^{-1} + X^{*'}X)}_{(A)}\beta - \beta'\underbrace{(X^{*'}y^* + \bar{V}^{-1}\bar{\beta})}_{(b)} - (y^{*'}X^* + \bar{\beta}'\bar{V}^{-1})\beta + \underbrace{y^{*'}y^* + \bar{\beta}'\bar{V}^{-1}\bar{\beta}}_{(r)} \tag{A.30}$$

Que pode ser visto como $\beta'A\beta - \beta'b - b'\beta + r$, onde o termo r não depende de β.

[1] Integrando a posteriori em relação a β resta um termo $h^{-\frac{k}{2}}$ que é o termo faltante para completar a densidade Gama para h.

Somando e diminuindo um termo $b'A^{-1}b$ chega-se em:

$$\begin{aligned}
\beta'A\beta - \beta'b - b'\beta + b'A^{-1}b \underbrace{-b'A^{-1}b + r}_{(c)} &= \\
&= \beta'A\beta - \beta'b - b'\beta + b'A^{-1}b + c \\
&= \beta'A\beta - \beta'AA^{-1}b - b'AA^{-1}\beta + b'A^{-1}AA^{-1}b \qquad \text{(A.31)}
\end{aligned}$$

Considerando $\overline{\Sigma} = A^{-1}$ e $\bar{\mu} = A^{-1}b$, pode-se reescrever (A.31) como:

$$\begin{aligned}
&\beta\overline{\Sigma}^{-1}\beta - \beta'\overline{\Sigma}^{-1}\bar{\mu} - \bar{\mu}'\overline{\Sigma}^{-1}\beta + \bar{\mu}'\overline{\Sigma}^{-1}\bar{\mu} + c \\
&= (\beta - \bar{\mu})'\,\overline{\Sigma}^{-1}(\beta - \bar{\mu}) + c \qquad \text{(A.32)}
\end{aligned}$$

Então, (A.29) é:

$$\begin{aligned}
p(y^*|y) = &\int\int \frac{h^{\frac{N+\bar{\nu}+k-2}{2}}}{(2\pi)^{N/2}}\bar{\chi}\exp\left\{-\frac{h}{2}\frac{\bar{\nu}}{\bar{s}^{-2}}\right\} \\
&\exp\left\{(\beta - \bar{\mu})'\,\overline{\Sigma}^{-1}(\beta - \bar{\mu})\right\}\exp\left\{-\frac{h}{2}c\right\} dh\ d\beta \\
= &\int \frac{h^{(N+\bar{\nu}+k-2)/2}}{(2\pi)^{N/2}}\bar{\chi}\exp\left\{-\frac{h}{2}\left(\frac{\bar{\nu}}{\bar{s}^{-2}} + c\right)\right\}\int \underbrace{\left[\exp\left\{-\frac{h}{2}(\beta - \bar{\beta})'\overline{\Sigma}^{-1}(\beta - \bar{\mu})\right\} d\beta\right]}_{\text{Núcleo de uma normal para } \beta|h} dh \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.33)}
\end{aligned}$$

Utilizando o resultado de que a integral do núcleo é o inverso da constante de uma densidade (BAUWENS; LUBRANO; RICHARD, 2003), a integral em relação a β que aparece em (A.33) será simplificada:

$$\begin{aligned}
\int \exp\left\{-\frac{h}{2}(\beta - \bar{\beta})'\overline{\Sigma}^{-1}(\beta - \bar{\mu})\right\} d\beta = \\
&= \frac{\frac{1}{h^{k/2}}}{(2\pi)^{k/2}|\overline{\Sigma}|^{1/2}} \\
&= \frac{(2\pi)^{k/2}|\overline{\Sigma}|^{1/2}}{h^{k/2}} \qquad \text{(A.34)}
\end{aligned}$$

De maneira que (A.33) é:

$$\begin{aligned}
p(y^*|y) &= \int \frac{h^{(N+\bar{\nu}-2)/2}}{(2\pi)^{N/2}}\bar{\chi}\frac{h^{k/2}}{h^{k/2}}(2\pi)^{k/2}|\overline{\Sigma}|^{1/2}\exp\left\{-\frac{h}{2}\left(\bar{\nu}\bar{s}^{-2} + c\right)\right\} dh \\
&= \frac{\bar{\chi}(2\pi)^{k/2}}{(2\pi)^{N/2}}|\overline{\Sigma}^{1/2}|\int h^{(N+\bar{\nu}-2)/2}\exp\left\{-\frac{h}{2}(\bar{\nu}\bar{s}^2 + c)\right\} dh
\end{aligned}$$

Definindo $\nu^* = N + \bar{\nu}$ e $\frac{\nu^*}{\mu^*} = \bar{\nu}\bar{s}^2 + c$ e substituindo acima, tem-se:

$$p(y^*|y) = \frac{\overline{\chi}(2\pi)^{k/2}}{(2\pi)^{N/2}}|\overline{\Sigma}^{1/2}| \int \underbrace{h^{(\nu^*-2)/2} \exp\left\{-\frac{h}{2}\frac{\nu^*}{\mu^*}\right\}}_{\text{Núcleo de uma Gama}(s^*,\nu^*)} dh \tag{A.35}$$

Usando novamente o resultado de que a integral do núcleo é o inverso da constante, obtem-se:

$$\begin{aligned}
\int h^{(\nu^*-2)/2} \exp\left\{-\frac{h}{2}\frac{\nu^*}{\mu^*}\right\} dh = & \\
&= \left(2\frac{\mu^*}{\nu^*}\right)^{\nu^*/2} \Gamma\left(\frac{\nu^*}{2}\right) \\
&= \left(\frac{2}{\bar{\nu}\bar{s}^2 + c}\right)^{\frac{N+\bar{\nu}}{2}} \Gamma\left(\frac{N+\bar{\nu}}{2}\right)
\end{aligned} \tag{A.36}$$

Substituindo em (A.35), as integrais são eliminadas e chega-se em:

$$p(y^*|y) = \frac{\overline{\chi}(2\pi)^{k/2}}{(2\pi)^{N/2}}|\overline{\Sigma}|^{1/2} \left(\frac{2}{\bar{\nu}\bar{s}^2 + c}\right)^{\frac{N+\bar{\nu}}{2}} \Gamma\left(\frac{N+\bar{\nu}}{2}\right) \tag{A.37}$$

com $\overline{\chi} = \frac{|\overline{V}|^{-1/2}}{\left(2\frac{\bar{s}^2}{\bar{\nu}}\right)^{\bar{\nu}/2}\Gamma\left(\frac{\bar{\nu}}{2}\right)(2\pi)^{k/2}}$.

A expressão em (A.37) não depende mais dos parâmetros β e h. Os passos seguintes manipulam a expressão para que esteja com a mesma organização que a densidade da distribuição t. O primeiro passo é verificar quem é o produto $|\overline{V}|^{-1/2} \cdot |\overline{\Sigma}|^{1/2}$, em que o termo $\overline{\Sigma} \equiv A^{-1}$ é dado por $\left(\overline{V}^{-1} + X^{*'}X^*\right)^{-1}$.

$$\begin{aligned}
|\overline{V}|^{-1/2} \cdot |\overline{\Sigma}|^{1/2} &= |\overline{V}|^{-1/2} \cdot |\left(\overline{V}^{-1} + X^{*'}X^*\right)^{-1}|^{1/2} \\
&= |\overline{V}|^{-1/2} \cdot |\left(\overline{V}^{-1} + X^{*'}X^*\right)|^{-1/2} \\
&= |\overline{V}\left(\overline{V}^{-1} + X^{*'}X^*\right)|^{-1/2} \\
&= |\overline{V}\overline{V}^{-1} + \overline{V}X^{*'}X^*|^{-1/2} \\
&= |\mathbb{I}_k + \overline{V}X^{*'}X^*|^{-1/2} = |\mathbb{I}_N + X^*\overline{V}X^{*'}|^{-1/2}
\end{aligned}$$

Onde na última igualdade foi utilizado o teorema de Sylvester para determinantes (POZRIKIDIS, 2014).

Assim, $p(y^*|y)$ é:

$$p(y^*|y) = \frac{(2\pi)^{k/2}}{(2\pi)^{N/2}(2\pi)^{k/2}} \frac{\Gamma\left(\frac{N+\bar{\nu}}{2}\right)}{\Gamma\left(\frac{\bar{\nu}}{2}\right)} \left|\mathbb{I}_N + X^*\overline{V}X^{*'}\right|^{-1/2} \underbrace{\left(\frac{2}{\bar{\nu}\bar{s}^2 + c}\right)^{\frac{N+\bar{\nu}}{2}} \frac{1}{\left(2\frac{\bar{s}^{-2}}{\bar{\nu}}\right)^{\bar{\nu}/2}}}_{\mathcal{A}} \tag{A.38}$$

Trabalhando apenas com o termo $\mathcal{A}$ de (A.38):

$$\begin{aligned}
\mathcal{A} &= \frac{2^{N/2}2^{\bar{\nu}/2}}{(\bar{\nu}\bar{s}^2+c)^{\frac{N+\bar{\nu}}{2}}}\frac{1}{2^{\bar{\nu}/2}(\bar{\nu})^{-\bar{\nu}/2}\bar{s}^{-\frac{2\bar{\nu}}{2}}} \\
&= \frac{2^{N/2}\bar{\nu}^{\bar{\nu}/2}\bar{s}^{\bar{\nu}}}{(\bar{\nu}\bar{s}^2+c)^{\frac{N+\bar{\nu}}{2}}} \\
&= \frac{2^{N/2}\bar{\nu}^{\bar{\nu}/2}\bar{s}^{\bar{\nu}}}{(\bar{s}^2(\bar{\nu}+c\bar{s}^{-2}))^{\frac{N+\bar{\nu}}{2}}} \\
&= \frac{2^{N/2}\bar{\nu}^{\bar{\nu}/2}\bar{s}^{\bar{\nu}}}{\bar{s}^{2\left(\frac{N+\bar{\nu}}{2}\right)}(\bar{\nu}+c\bar{s}^{-2})^{\frac{N+\bar{\nu}}{2}}} \\
&= 2^{N/2}\ \bar{\nu}^{\bar{\nu}/2}\ \bar{s}^{-N}(\bar{\nu}+c\bar{s}^{-2})^{-\frac{N+\bar{\nu}}{2}}
\end{aligned} \tag{A.39}$$

Inserindo (A.39) em (A.38):

$$p(y^*|y) = \frac{2^{N/2}\ \bar{\nu}^{\bar{\nu}/2}}{2^{N/2}\ \pi^{N/2}}\ \frac{\Gamma\left(\frac{N+\bar{\nu}}{2}\right)}{\Gamma\left(\frac{\bar{\nu}}{2}\right)}\left|\bar{s}^2\ (\mathbb{I}_N + X^*\overline{V}X^{*'})\right|^{-1/2}\ (\bar{\nu}+c\bar{s}^{-2})^{-\frac{N+\bar{\nu}}{2}} \tag{A.40}$$

O termo c de (A.40) foi definido em (A.30) e é dado por:

$$\begin{aligned}
c &= r - b'A^{-1}b = y^{*'}y^* + \overline{\beta}'\overline{V}^{-1}\overline{\beta} - b'A^{-1}b \\
&= y^{*'}y^* + \overline{\beta}'\overline{V}^{-1}\overline{\beta} - (X^{*'}y^* + \overline{V}^{-1}\overline{\beta})'\underbrace{(\overline{V}^{-1} + X^{*'}X^*)^{-1}}_{Q^{-1}}(X^{*'}y^* + \overline{V}^{-1}\overline{\beta}) \\
&= y^{*'}y^* + \overline{\beta}'\overline{V}^{-1}\overline{\beta} - X^{*'}y^*Q^{-1}X^{*'}y^* - (\overline{V}^{-1}\overline{\beta})'Q^{-1}X^{*'}y^* \\
&\quad - y^{*'}X^{*'}Q^{-1}\overline{V}^{-1}\overline{\beta} - (\overline{V}^{-1}\overline{\beta})'Q^{-1}\overline{V}^{-1}\overline{\beta} \\
&= y^{*'}(y^* - X^*Q^{-1}X^{*'}y^*) + \overline{\beta}'(\overline{V}^{-1} - \overline{V}^{-1}Q^{-1}\overline{V}^{-1})\overline{\beta} - 2\overline{\beta}'\overline{V}^{-1}Q^{-1}X^{*'}y^* \\
&= y^{*'}\underbrace{(\mathbb{I}_N - X^*(\overline{V} + X^{*'}X)^{-1}X^{*'})}_{(\mathcal{B})}y^* + \overline{\beta}'\underbrace{(\overline{V}^{-1} - \overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}\overline{V}^{-1})}_{(\mathcal{C})}\overline{\beta} \\
&\quad - 2\overline{\beta}'\underbrace{\overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'}}_{(\mathcal{D})}y^*
\end{aligned} \tag{A.41}$$

Os termos destacados em (A.41) podem ser reescritos da seguinte maneira:

- $(\mathcal{B})$ será $\mathbb{I}_N - X^*(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'} = (\mathbb{I}_N + X^*\overline{V}X^{*'})$, usando o resultado de Searle (1982) que estabelece que $(\mathbb{I} + AB)^{-1} = \mathbb{I} - A(\mathbb{I} + BA)^{-1}B$. Para verificar, basta tomar $A = X^*\overline{V}$ e $B = X'$:

$$\begin{aligned}
\mathbb{I} - X^*\overline{V}(\mathbb{I} + X^{*'}X^*\overline{V})^{-1}X^{*'} &= \\
&= \mathbb{I} - X^*\overline{V}(\overline{V}^{-1}\overline{V} + X^{*'}X^*\overline{V})^{-1}X^{*'} \\
&= \mathbb{I} - X^*\overline{V}((\overline{V}^{-1} + X^{*'}X^*)\overline{V})^{-1}X^{*'}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{I} - X^*\overline{V}(\overline{V})^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^* \\
&= \mathbb{I} - X^*(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'} \qquad \text{(A.42)}
\end{aligned}
$$

- $(\mathcal{C})$ é $(\overline{V}^{-1} - \overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}\overline{V}^{-1}) = X^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}X^*$, pois:

$$
\begin{aligned}
(\overline{V}^{-1} - \overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}\overline{V}^{-1}) &= \\
&= X^{*'}X^* - X^{*'}X^*(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'}X^* \\
&= X^{*'}(X^* - X^*(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'}X) \\
&= X^{*'}(\mathbb{I} - X^*(\overline{V}^{-1} + X^{*'}X^*)^{-1})X^* \\
&= X^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}X^*
\end{aligned}
$$

  Onde na primeira igualdade foi utilizado o resultado de Searle (1982) para matrizes dado por $A - A(A+B)^{-1}A = B - B(A+B)^{-1}B$, considerando $A = \overline{V}^{-1}$ e $B = X^{*'}X^*$. Adicionalmente, na última igualdade, foi utilizado o resultado calculado em $(\mathcal{B})$.

- $(\mathcal{D})$ é $\overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'} = X^*(\mathbb{I} + X^*\overline{V}X^{*'})$ pois:

$$
\begin{aligned}
\overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*)^{-1}X^{*'} &= \\
&= \overline{V}^{-1}(\overline{V}^{-1} + X^{*'}X^*\overline{V}\overline{V}^{-1})^{-1}X^{*'} \\
&= \overline{V}^{-1}\left((\mathbb{I} + X^{*'}X^*\overline{V})\overline{V}^{-1}\right)^{-1} \\
&= \overline{V}^{-1}(\overline{V}^{-1})^{-1}(\mathbb{I} + X^{*'}X^*\overline{V})^{-1}X^{*'} \\
&= (\mathbb{I} + X^{*'}X^*\overline{V})^{-1}X^{*'} \\
&= X^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}
\end{aligned}
$$

  Na última igualdade foi utilizado o resultado $(\mathbb{I} + AB)^{-1}A = A(\mathbb{I} + BA)^{-1}$ (SEARLE, 1982).

Substituindo os termos $(\mathcal{B})$, $(\mathcal{C})$ e $(\mathcal{D})$ na equação (A.41), tem-se que:

$$
\begin{aligned}
c &= y^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}y^* + \overline{\beta}'X^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}X^*\overline{\beta} - 2\overline{\beta}'X^{*'}(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}y^* \\
&= (y^* - X^*\overline{\beta})'(\mathbb{I} + X^*\overline{V}X^{*'})^{-1}(y^* - X^*\overline{\beta}) \qquad \text{(A.43)}
\end{aligned}
$$

Por fim, substituindo (A.43) em (A.40), é obtida a forma final da densidade preditiva a posteriori:

$$
\begin{aligned}
p(y^*|y) = \frac{\overline{\nu}^{\overline{\nu}/2}\Gamma\left(\frac{N+\overline{\nu}}{2}\right)}{\pi^{N/2}\Gamma\left(\frac{\overline{\nu}}{2}\right)} &\left|\overline{s}^2\ (\mathbb{I}_N + X^*\overline{V}X^{*'})\right|^{-1/2} \\
&\cdot \left[\overline{\nu} + (y^* - X^*\overline{\beta})'\left(\overline{s}^{-2}(\mathbb{I} + X^*\overline{V}X^{*'})\right)^{-1}(y^* - X^*\overline{\beta})\right]^{-\frac{N+\overline{\nu}}{2}} \qquad \text{(A.44)}
\end{aligned}
$$

APÊNDICE B – TABELAS E GRÁFICOS DOS RESULTADOS DAS SIMULAÇÕES DO CAPÍTULO 3

EXPERIMENTO 1

Tabela 9 – Resultados do primeiro experimento - considerando os pesos dos erros menores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]	β_5	EQM[β_5]
Valor Real	**2**		**100**		**0.6**		**5**		**0.01**	
					$n = 30$					
MQO	6.11761	178086	99.73911	254.58	0.59872	0.11189	4.99955	0.01144	0.01	$< 10^{-5}$
Priori 1	2.00003	10^{-5}	99.94956	8.8695	0.59943	0.06303	5.00012	0.00219	0.01	$< 10^{-5}$
Priori 2	4.98782	93767	99.73663	254.25	0.59887	0.11056	4.99979	0.0078	0.01	$< 10^{-5}$
Priori 3	5.09592	9.5848	3.29537	9351.8	0.75021	0.129	5.09113	0.01158	0.0098	$< 10^{-5}$
Priori 4	5.93396	135063	99.69962	254.22	0.59875	0.11121	4.99964	0.00959	0.01	$< 10^{-5}$
					$n = 100$					
MQO	-0.31872	78845.20395	100.08226	70.37695	0.60121	0.02982	5.00019	0.00306	0.01	$< 10^{-5}$
Priori 1	1.99997	2e-05	100.03118	15.07553	0.60056	0.02057	4.99984	7e-04	0.01	$< 10^{-5}$
Priori 2	0.00356	58450.8855	100.07992	69.34039	0.60112	0.02816	5.00014	0.00248	0.01	$< 10^{-5}$
Priori 3	5.12958	9.79437	5.02644	9020.01438	0.72487	0.03865	5.08702	0.00833	0.00997	$< 10^{-5}$
Priori 4	0.08345	69734.5532	100.06942	69.89649	0.60111	0.02908	5.00013	0.0028	0.01	$< 10^{-5}$
					$n = 500$					
MQO	4.04642	13665.9299	99.94503	12.54504	0.599	0.00506	4.99972	0.00057	0.01	$< 10^{-5}$
Priori 1	2.00018	0.00011	99.965	8.26884	0.59951	0.00412	5.00007	0.00016	0.01	$< 10^{-5}$
Priori 2	3.99013	12924.47029	99.94538	12.51308	0.59901	0.00501	4.99973	0.00055	0.01	$< 10^{-5}$
Priori 3	5.23573	10.47064	14.23742	7355.38872	0.63546	0.00546	5.08797	0.00789	0.01007	$< 10^{-5}$
Priori 4	4.07072	13367.90608	99.94307	12.53195	0.59899	0.00504	4.99972	0.00056	0.01	$< 10^{-5}$
					$n = 1000$					
MQO	2.2483	6622	99.98923	6.05774	0.59977	0.00224	4.99997	0.00026	0.01	$< 10^{-5}$
Priori 1	2.00004	0.00019	99.99099	4.9073	0.59984	0.00169	5.00001	6e-05	0.01	$< 10^{-5}$
Priori 2	2.2447	6432	99.98925	6.05561	0.59977	0.00223	4.99997	0.00025	0.01	$< 10^{-5}$
Priori 3	5.22229	10.38441	22.71169	5973.73646	0.68746	0.00936	5.06156	0.00385	0.01013	$< 10^{-5}$
Priori 4	2.26815	6547	99.98823	6.0568	0.59977	0.00224	4.99997	0.00026	0.01	$< 10^{-5}$
					$n = 5000$					
MQO	1.57739	1447	100.01853	1.30716	0.59995	0.00051	5.0001	6e-05	0.01	$< 10^{-5}$
Priori 1	1.99964	0.00106	100.01626	1.21841	0.59984	4e-04	5.00003	2e-05	0.01	$< 10^{-5}$
Priori 2	1.57862	1439	100.01852	1.30697	0.59995	0.00051	5.0001	6e-05	0.01	$< 10^{-5}$
Priori 3	5.55976	12.67874	57.40728	1814.54086	0.64833	0.00274	5.03876	0.00152	0.01003	$< 10^{-5}$
Priori 4	1.58237	1444	100.01831	1.30707	0.59995	0.00051	5.0001	6e-05	0.01	$< 10^{-5}$

Notas técnicas:

⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;

⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.

⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;

⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;

⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

Tabela 10 – Resultados do primeiro experimento - considerando os pesos dos erros maiores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]	β_5	EQM[β_5]
Valor Real	**2**		**100**		**0.6**		**5**		**0.01**	
					$n = 30$					
MQO	50.6857	28175247	99.95006	29848	0.58744	12.81169	4.98532	1.19166	0.01006	$< 10^{-5}$
Priori 1	2.00026	0.00078	99.96117	1183	0.60004	6.9498	4.99364	0.21092	0.01007	$< 10^{-5}$
Priori 2	34.95587	12910119	99.90356	29735	0.59162	11.74926	4.98807	0.72597	0.01006	$< 10^{-5}$
Priori 3	5.09491	9.58352	3.34189	9344	0.90705	10.6348	5.09596	0.2921	0.00971	$< 10^{-5}$
Priori 4	43.41252	1999106	99.8925	29764	0.58945	12.24124	4.98664	0.94196	0.01006	$< 10^{-5}$
					$n = 100$					
MQO	-33.6635	7391476	100.31335	7520	0.57822	2.77109	5.0116	0.32087	0.01001	$< 10^{-5}$
Priori 1	1.99923	0.00329	100.08001	1924	0.57121	2.13279	5.00526	0.08016	0.01	$< 10^{-5}$
Priori 2	-29.87833	5905741	100.29602	7488	0.57733	2.68971	5.01094	0.27607	0.01001	$< 10^{-5}$
Priori 3	5.11908	9.75033	5.37814	8959	0.46521	2.3664	5.13886	0.10174	0.01002	$< 10^{-5}$
Priori 4	-31.90135	6748346	100.29665	7505	0.57779	2.73584	5.01131	0.30148	0.01001	$< 10^{-5}$
					$n = 500$					
MQO	23.67215	1805996	99.77427	1475	0.58906	0.64406	4.99725	0.06975	0.01	$< 10^{-5}$
Priori 1	2.00183	0.01311	99.88145	1034	0.59554	0.47548	5.00087	0.01742	0.01	$< 10^{-5}$
Priori 2	23.05421	1704476	99.77671	1474	0.58925	0.63511	4.99736	0.06686	0.01	$< 10^{-5}$
Priori 3	5.19394	10.28791	15.08315	7234	0.68156	0.49072	5.0799	0.02314	0.01006	$< 10^{-5}$
Priori 4	23.46902	1765131	99.77341	1475	0.58913	0.64046	4.99729	0.06859	0.01	$< 10^{-5}$
					$n = 1000$					
MQO	3.95409	875597	99.78599	783.66039	0.60647	0.28932	4.99903	0.03538	0.01	$< 10^{-5}$
Priori 1	2.00034	0.02908	99.81436	629.05012	0.60686	0.23765	4.99937	0.00855	0.01	$< 10^{-5}$
Priori 2	3.92765	852065	99.78613	782.96372	0.60647	0.28799	4.99904	0.03466	0.01	$< 10^{-5}$
Priori 3	5.30817	11.13585	23.97843	5816	0.7954	0.27472	5.05114	0.01092	0.00998	$< 10^{-5}$
Priori 4	3.96535	866253	99.78507	783.37005	0.60647	0.2888	4.99903	0.03509	0.01	$< 10^{-5}$
					$n = 5000$					
MQO	0.00568	166781	99.97236	157.88088	0.59736	0.0624	5.00088	0.00686	0.01	$< 10^{-5}$
Priori 1	1.99812	0.14724	99.96364	147.46914	0.59682	0.05015	5.00053	0.00189	0.01	$< 10^{-5}$
Priori 2	0.01096	165898	99.97234	157.85889	0.59736	0.06233	5.00088	0.00684	0.01	$< 10^{-5}$
Priori 3	5.58768	13.82801	59.00134	1732	0.65125	0.05266	5.03554	0.00308	0.01003	$< 10^{-5}$
Priori 4	0.01189	166434	99.97216	157.87167	0.59736	0.06237	5.00088	0.00685	0.01	$< 10^{-5}$

Notas técnicas:

⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;

⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.

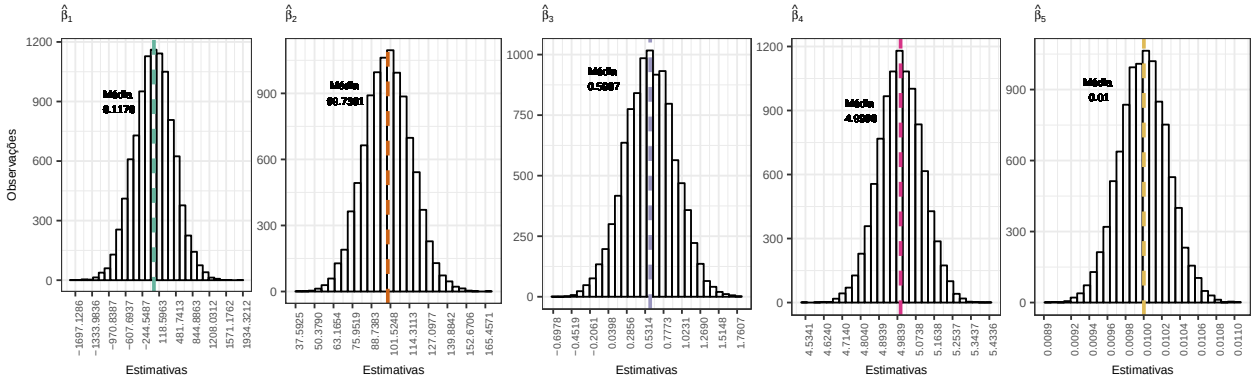
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pelo parâmetro β_i;

⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;

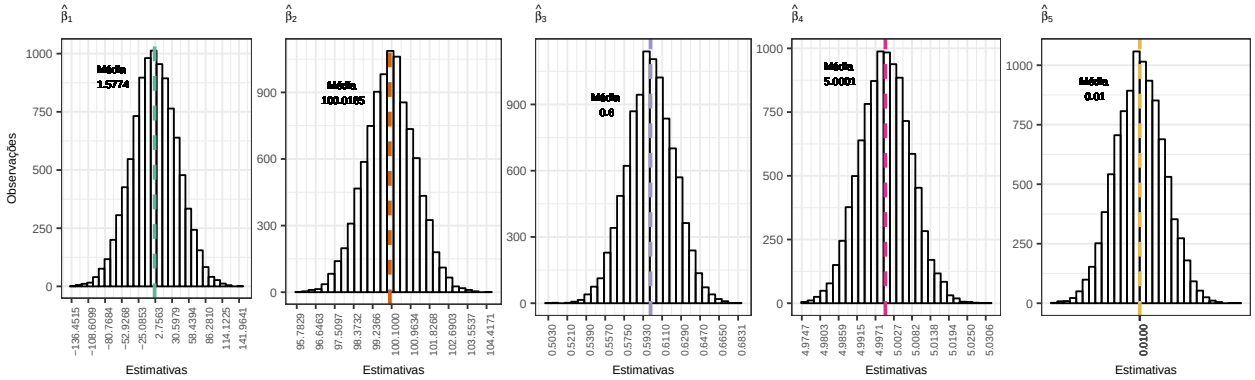
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

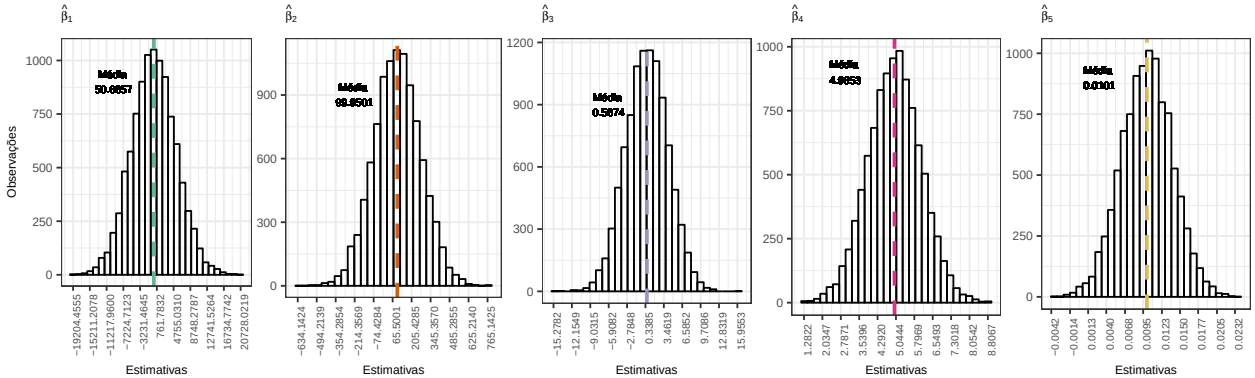
Figura 11 – Comparação das estimativas de MQO no experimento 1 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



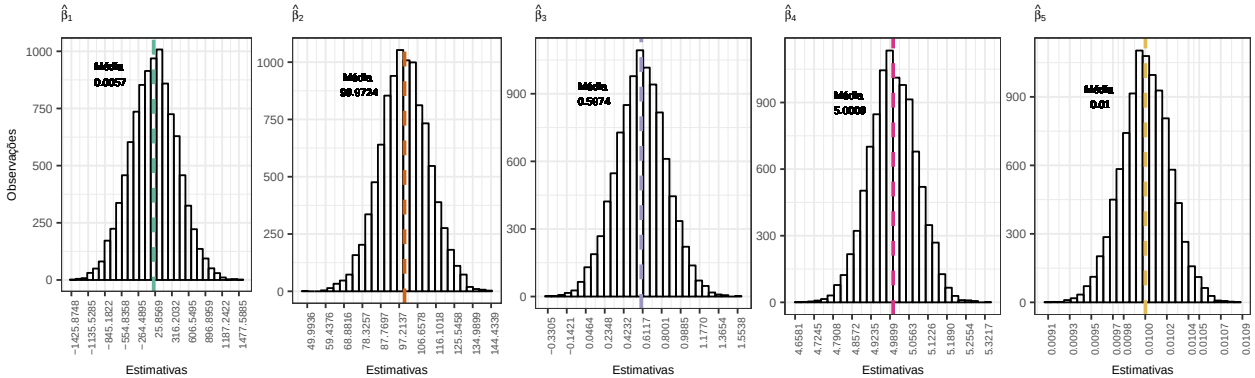
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



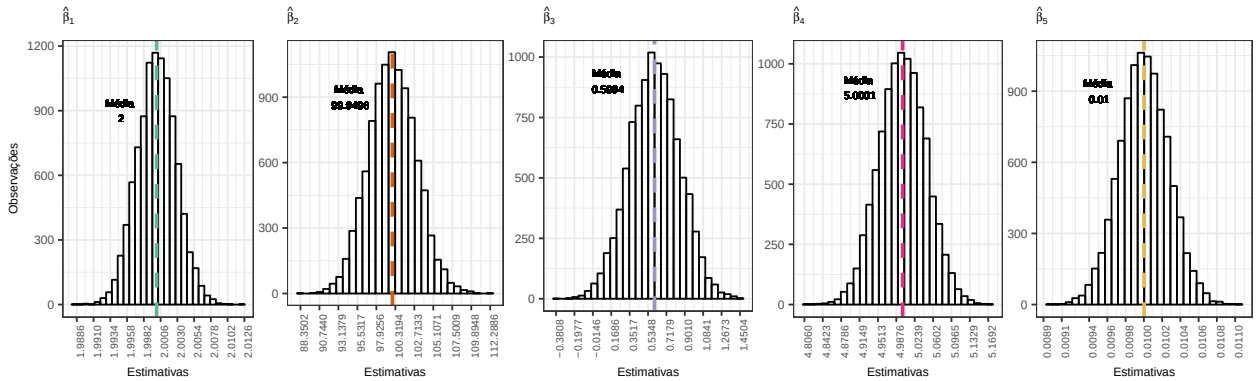
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



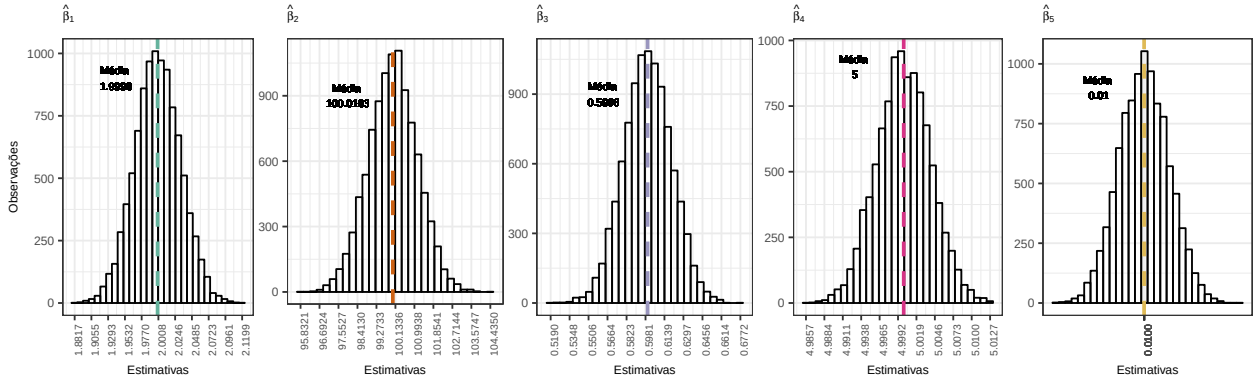
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

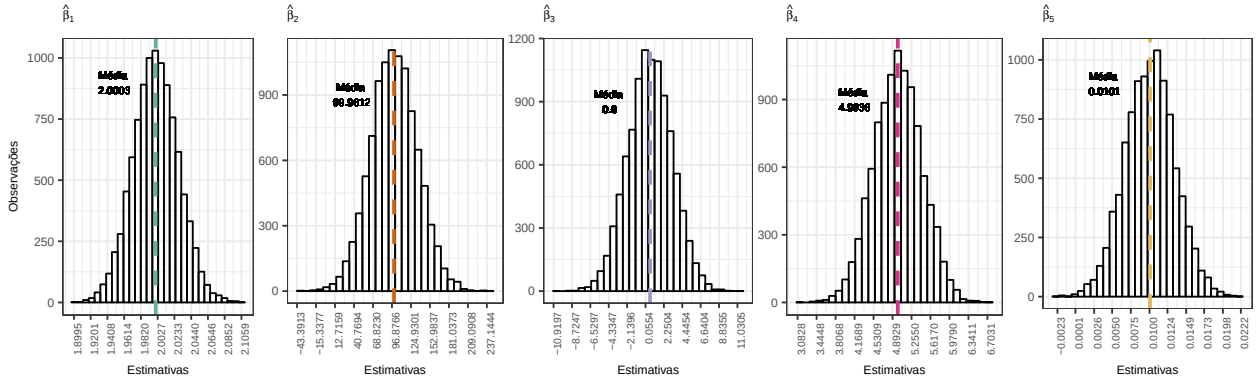
Figura 12 – Comparação das estimativas usando a Priori 1 no experimento 1 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



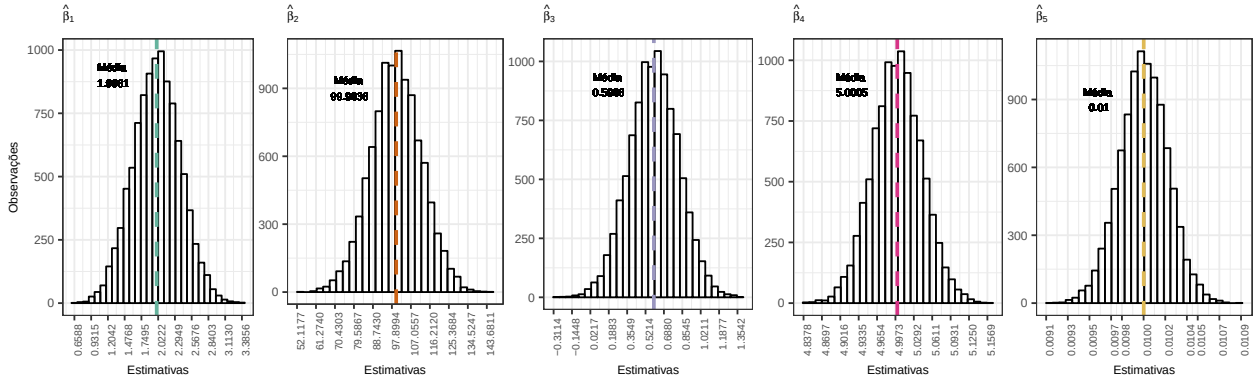
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



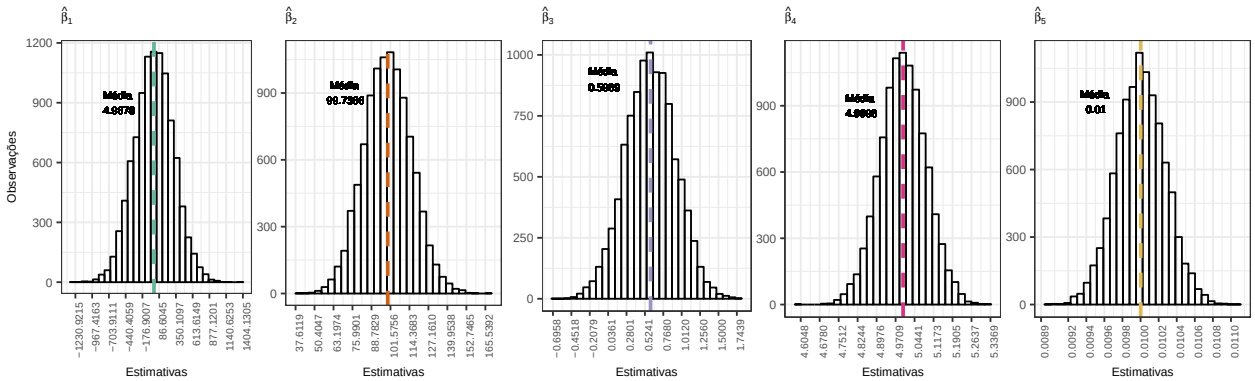
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



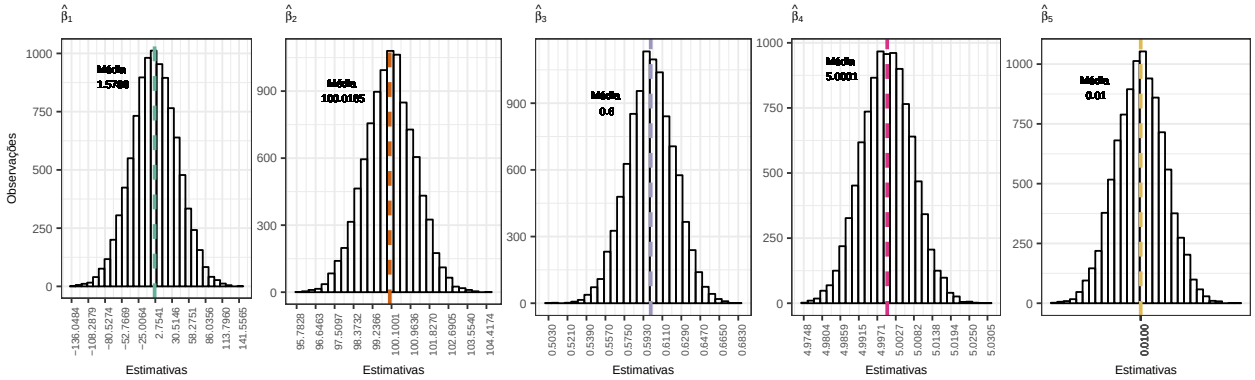
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

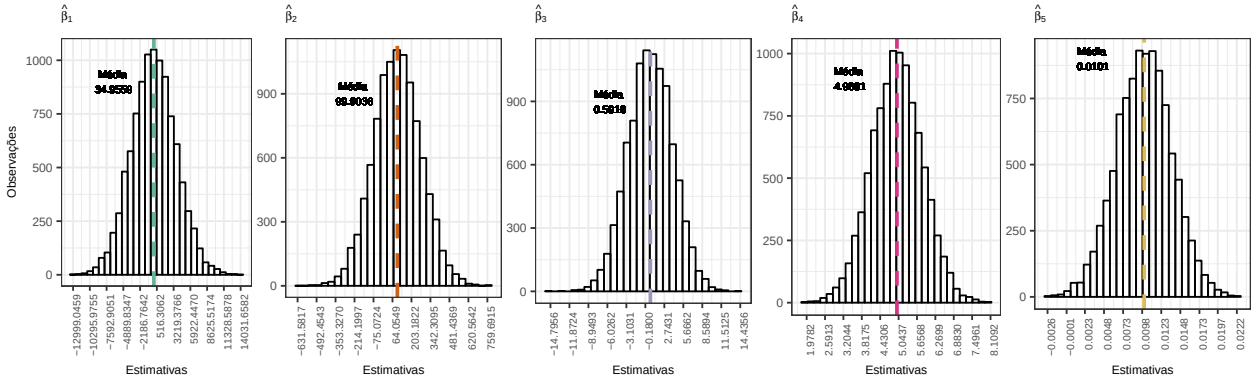
Figura 13 – Comparação das estimativas usando a Priori 2 no experimento 1 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



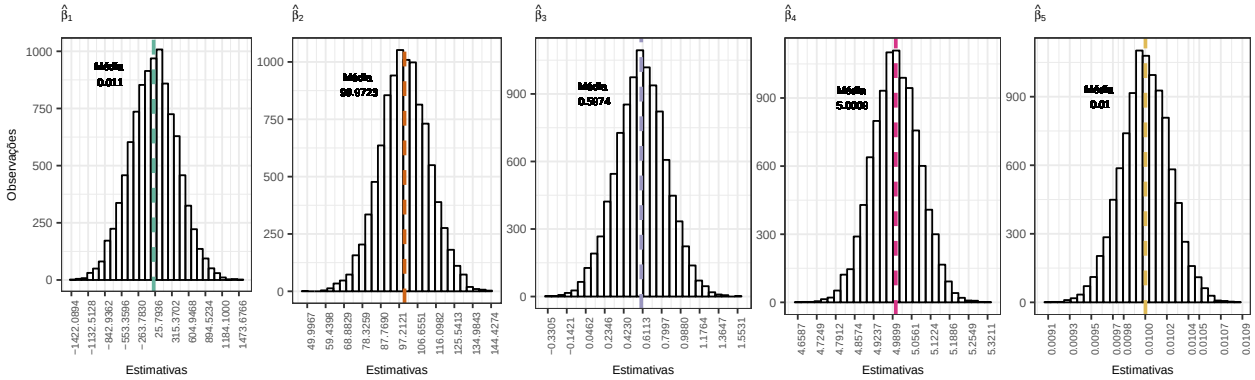
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



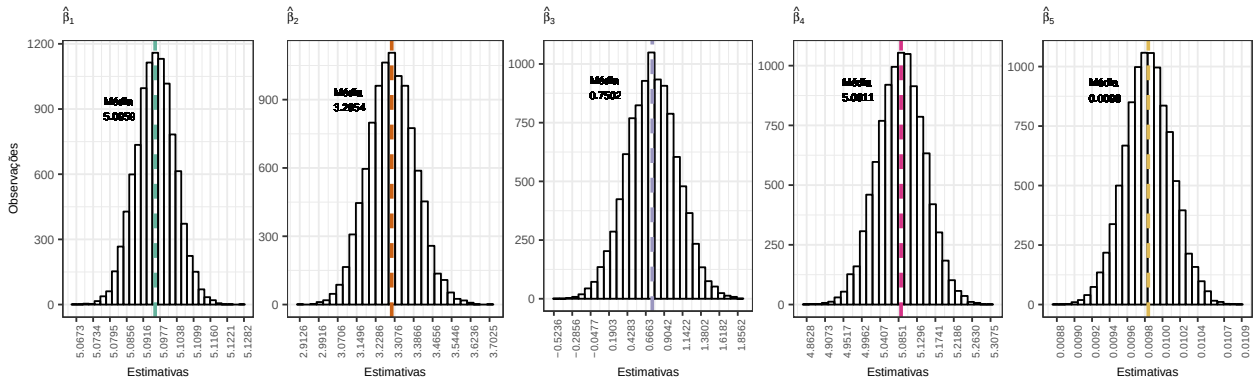
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



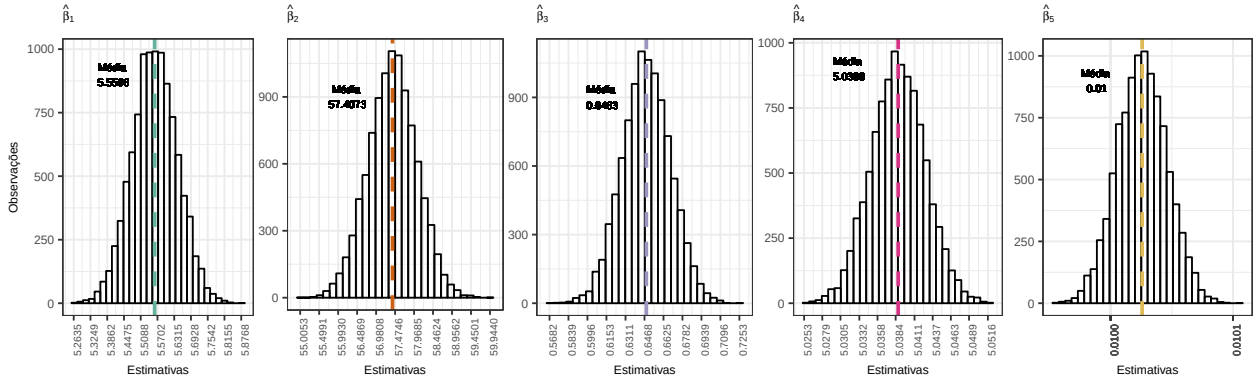
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

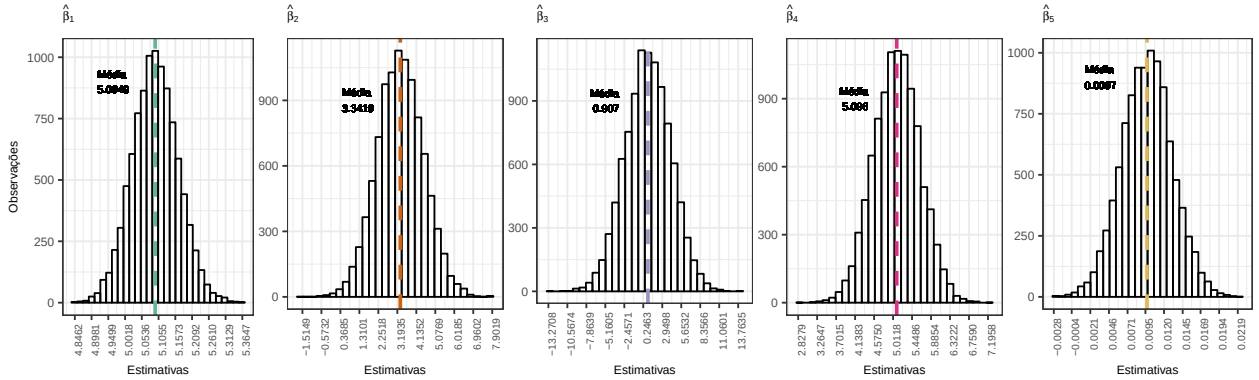
Figura 14 – Comparação das estimativas usando a Priori 3 no experimento 1 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



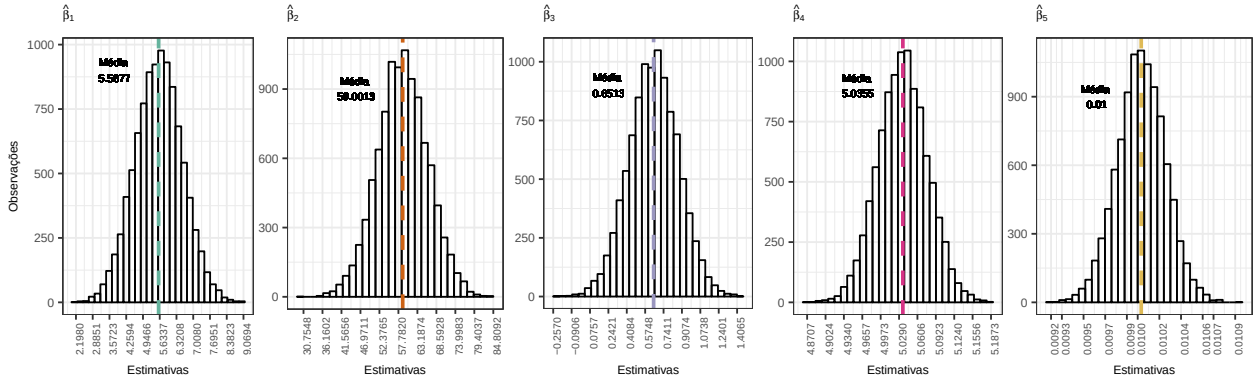
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



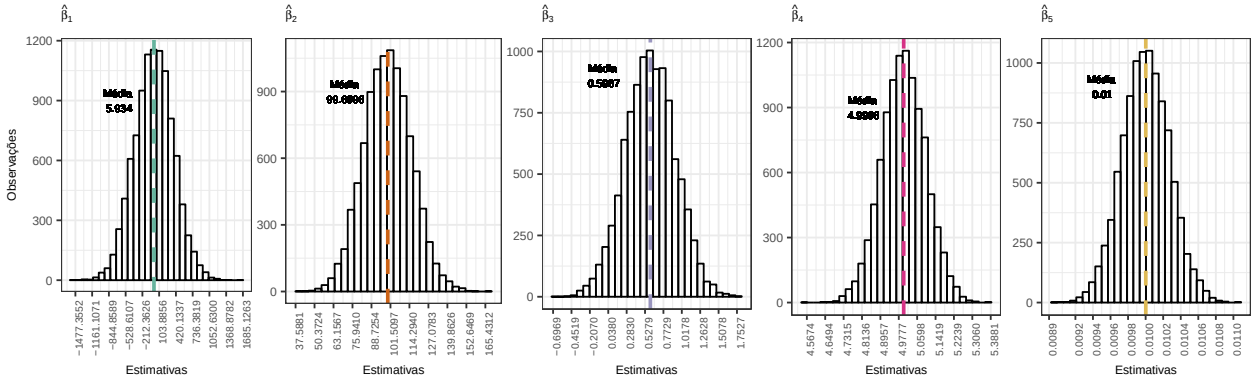
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



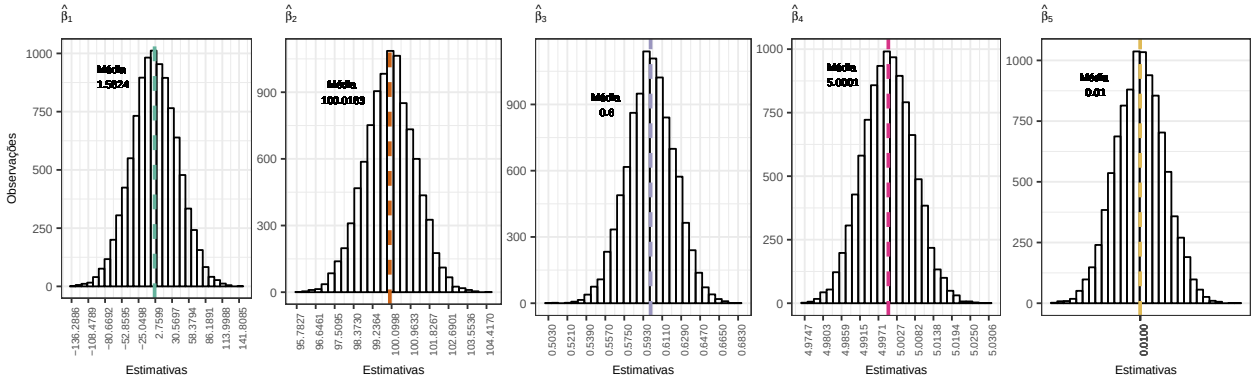
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

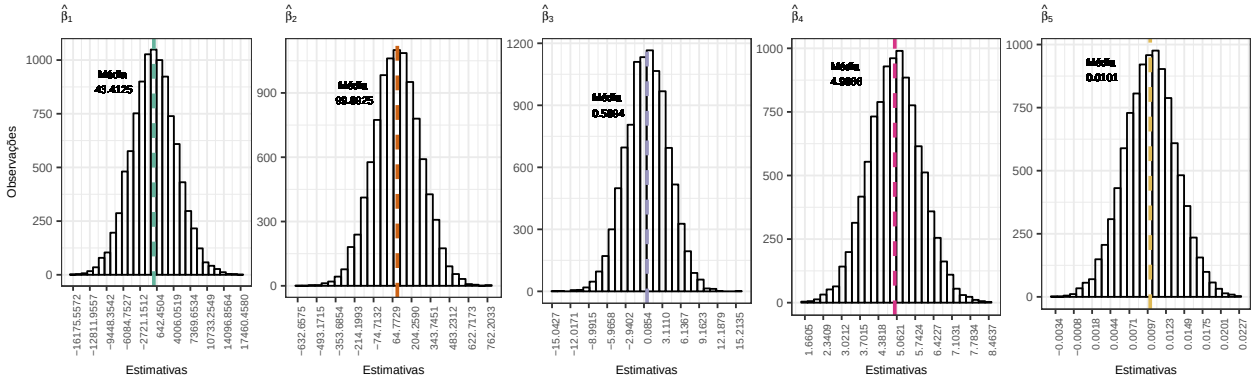
Figura 15 – Comparação das estimativas usando a Priori 4 no experimento 1 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



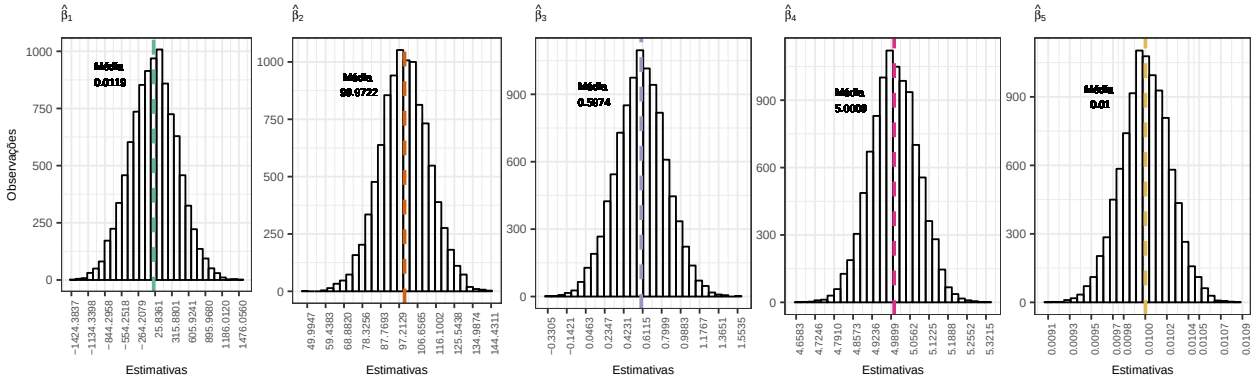
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$. **Observação**: *Para $\bar{\beta}_5$ (extrema direita), como os valores obtidos são muito próximos da média, não foi possível colocar valores no eixo X usando 4 casas decimais sem que a visualização ficarsse prejudicada pelo tamanho reduzido da área de plotagem.*



(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

EXPERIMENTO 2

Tabela 11 – Resultados dos segundo experimento - considerando os pesos dos erros menores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	1		10		50		100	
				$n = 30$				
MQO	-0.60743	19702.86449	10.19974	266.91286	49.91912	220.83042	100.11038	217.79118
Priori 1	0.99828	0.17874	10.04174	53.16744	49.86406	120.42264	100.04437	120.97574
Priori 2	-0.6041	19621.27691	10.19947	266.38281	49.91895	220.61495	100.11024	217.64479
Priori 3	0.25604	0.55603	6.3803	49.34654	49.07149	134.8549	105.64429	164.00048
Priori 4	-0.60074	19403.3528	10.19918	264.96951	49.91878	220.03969	100.11012	217.25395
				$n = 100$				
MQO	0.01446	6853	10.14591	85.45062	50.09327	90.52181	99.99114	85.24094
Priori 1	0.99666	0.32194	10.06523	36.87026	50.03419	52.78038	99.94756	58.17111
Priori 2	0.0151	6844	10.14586	85.41097	50.09323	90.48523	99.99111	85.21615
Priori 3	0.25842	0.55771	8.34392	33.03585	50.09094	53.81113	102.08331	64.35086
Priori 4	0.01506	6821	10.14586	85.30449	50.09323	90.38679	99.99112	85.14943
				$n = 500$				
MQO	0.90421	1700	10.02539	20.15581	50.07523	21.16772	99.95935	20.417
Priori 1	0.99852	0.70867	10.01903	12.74621	50.06919	14.2104	99.95431	13.63009
Priori 2	0.90422	1700	10.02538	20.15405	50.07523	21.16593	99.95934	20.41513
Priori 3	0.26214	0.58206	9.5771	12.28683	50.20151	14.21364	100.44386	13.84289
Priori 4	0.90387	1699	10.0254	20.14933	50.07526	21.16109	99.95937	20.41007
				$n = 1000$				
MQO	0.86317	822.35844	9.997	10.53705	49.99788	9.5513	100.04679	10.03891
Priori 1	0.99522	0.9897	9.98922	7.01438	49.98979	6.38606	100.03895	7.25169
Priori 2	0.86318	822.24342	9.997	10.53661	49.99788	9.55087	100.04679	10.03853
Priori 3	0.26661	0.61306	9.76401	6.87707	50.06057	6.38564	100.33435	7.34726
Priori 4	0.86301	821.93275	9.99701	10.53543	49.99789	9.54971	100.0468	10.03752
				$n = 5000$				
MQO	0.89451	147.50658	10.01785	1.84219	49.99409	1.77725	99.99406	1.81725
Priori 1	0.98278	1.98719	10.01259	1.30377	49.98889	1.27857	99.98865	1.27063
Priori 2	0.89452	147.50281	10.01785	1.84217	49.99409	1.77723	99.99406	1.81723
Priori 3	0.29213	0.87251	10.00343	1.28129	50.03934	1.26823	100.07541	1.2625
Priori 4	0.89448	147.49263	10.01785	1.84214	49.9941	1.7772	99.99406	1.8172

Notas técnicas:

⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;

⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.

⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;

⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;

⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

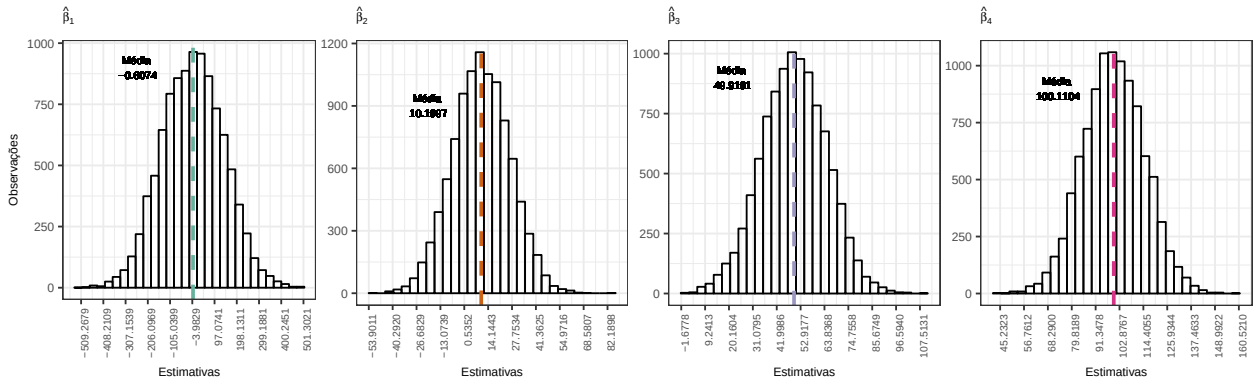
Tabela 12 – Resultados do segundo experimento - considerando os pesos dos erros maiores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	1		10		50		100	
				$n = 30$				
MQO	0.80417	1822.85807	10.03033	22.86407	50.01226	23.04962	99.98315	20.8943
Priori 1	0.99983	0.05138	10.00964	4.61022	50.00378	9.81439	99.97706	11.62243
Priori 2	0.80459	1815.02127	10.0303	22.82392	50.01223	23.02237	99.98313	20.87096
Priori 3	0.25327	0.55781	6.3389	16.51192	50.64874	10.69967	105.63396	44.58432
Priori 4	0.79989	1794.09708	10.03062	22.71696	50.01251	22.94962	99.9834	20.80862
				$n = 100$				
MQO	0.31494	675.27405	10.02123	8.67058	50.02562	7.82245	100.0467	8.31571
Priori 1	0.9973	0.09734	9.97934	3.59142	49.99068	5.19514	100.00425	5.22967
Priori 2	0.3154	674.37111	10.0212	8.66639	50.0256	7.82027	100.04668	8.31261
Priori 3	0.25794	0.55137	8.11019	6.51186	50.23505	5.31561	102.11081	9.77056
Priori 4	0.31481	671.94248	10.02123	8.65512	50.02563	7.81439	100.04672	8.30427
				$n = 500$				
MQO	0.96701	120.60475	10.00478	1.53995	50.00268	1.56229	100.00313	1.5621
Priori 1	0.99936	0.22062	10.00262	0.9678	50.00076	1.08983	100.00143	1.17127
Priori 2	0.96701	120.57533	10.00478	1.53983	50.00268	1.56219	100.00313	1.56202
Priori 3	0.26545	0.54323	9.5948	1.08731	50.07622	1.09777	100.48801	1.41054
Priori 4	0.96669	120.49589	10.0048	1.53951	50.0027	1.56189	100.00315	1.56178
				$n = 1000$				
MQO	1.04097	59.25808	9.99504	0.74737	49.99369	0.72363	100.00636	0.75614
Priori 1	1.00157	0.29819	9.99743	0.52295	49.99606	0.50469	100.00875	0.5181
Priori 2	1.04097	59.25073	9.99504	0.74734	49.99369	0.7236	100.00636	0.75611
Priori 3	0.26545	0.54323	9.5948	1.08731	50.07622	1.09777	100.48801	1.41054
Priori 4	1.0408	59.23084	9.99505	0.74728	49.9937	0.72353	100.00638	0.75603
				$n = 5000$				
MQO	0.99332	12.73354	10.00147	0.15448	49.99327	0.154	100.00569	0.15224
Priori 1	0.99884	0.61726	10.00113	0.10902	49.99295	0.11066	100.00535	0.10545
Priori 2	0.99332	12.73324	10.00147	0.15447	49.99327	0.15399	100.00569	0.15224
Priori 3	0.29971	0.5268	9.99909	0.10729	50.04256	0.11116	100.08687	0.11165
Priori 4	0.99328	12.73241	10.00147	0.15447	49.99327	0.15399	100.00569	0.15224

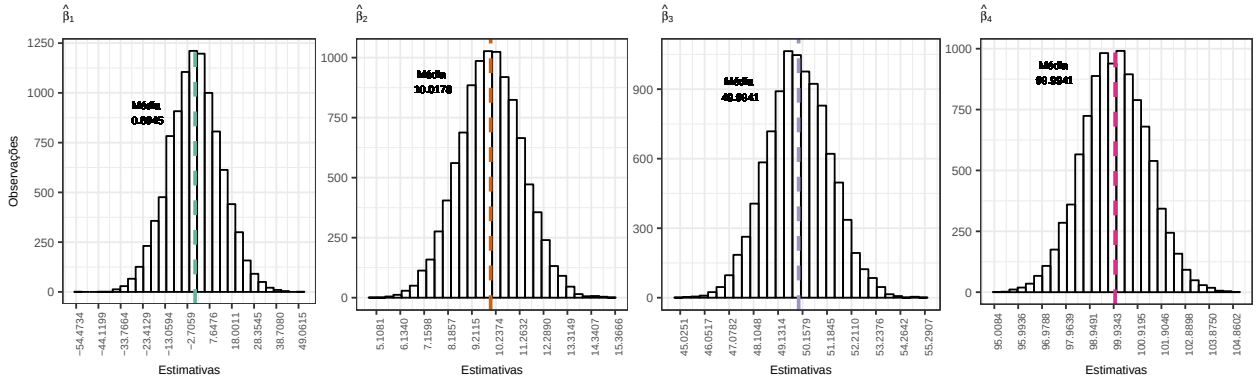
Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pelo parâmetro β_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

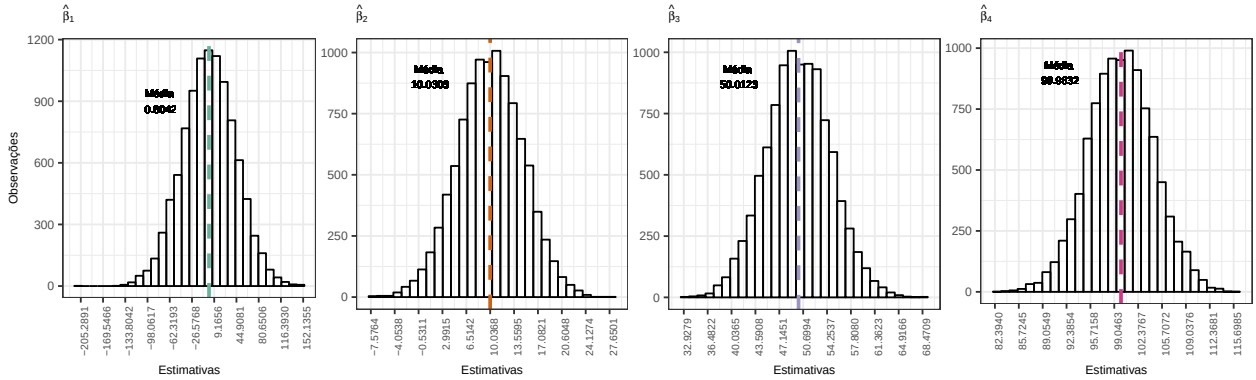
Figura 16 – Comparação das estimativas de MQO no experimento 2 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



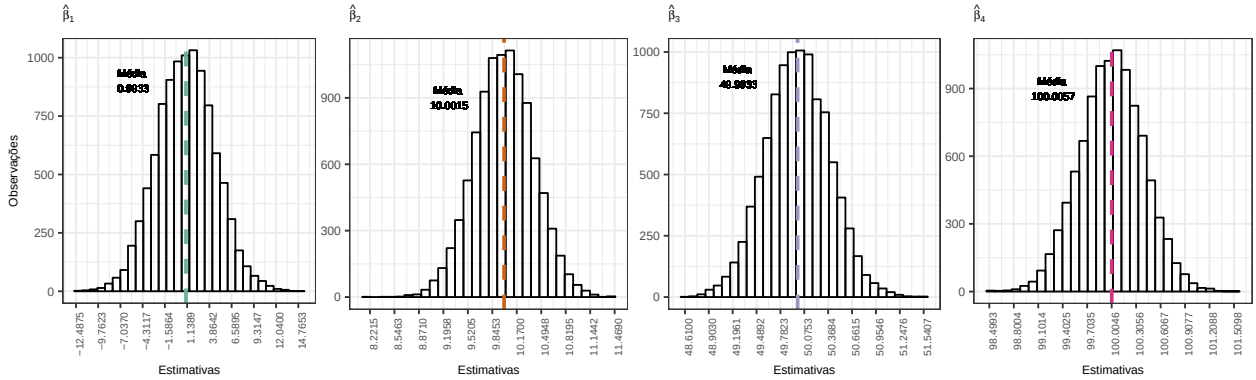
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



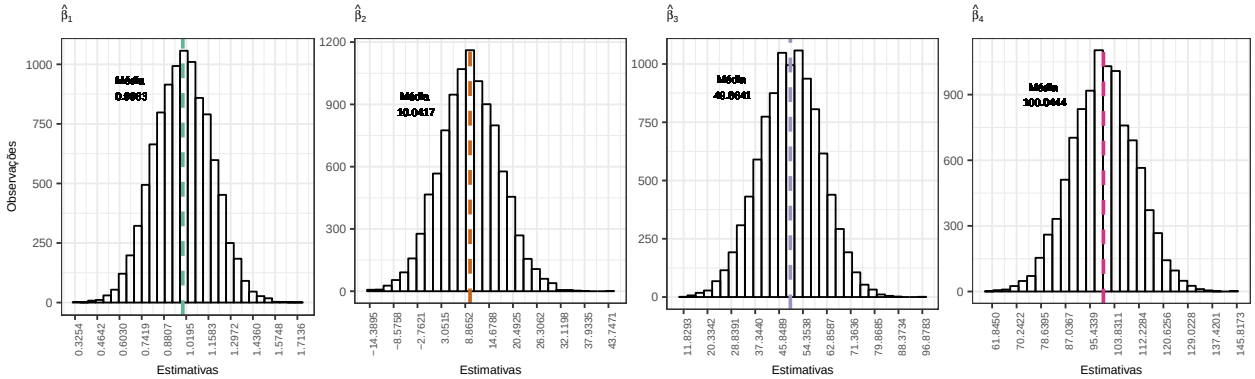
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



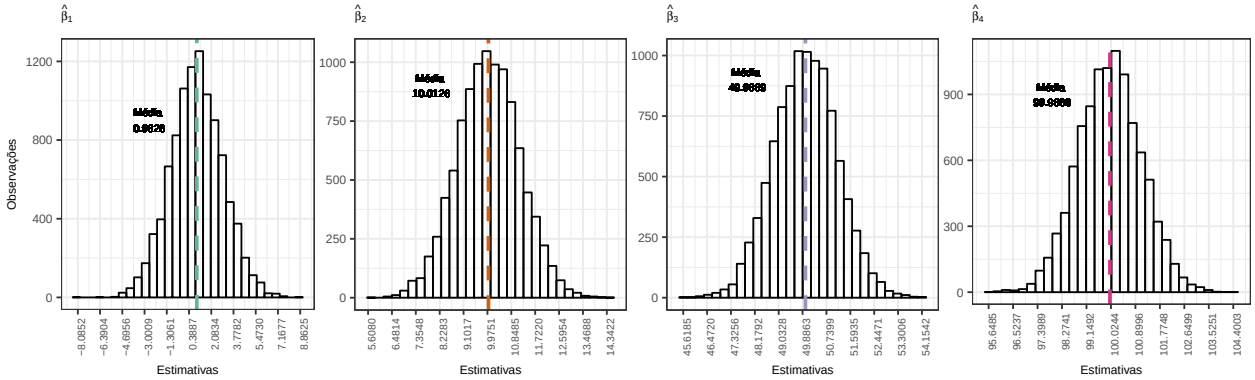
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

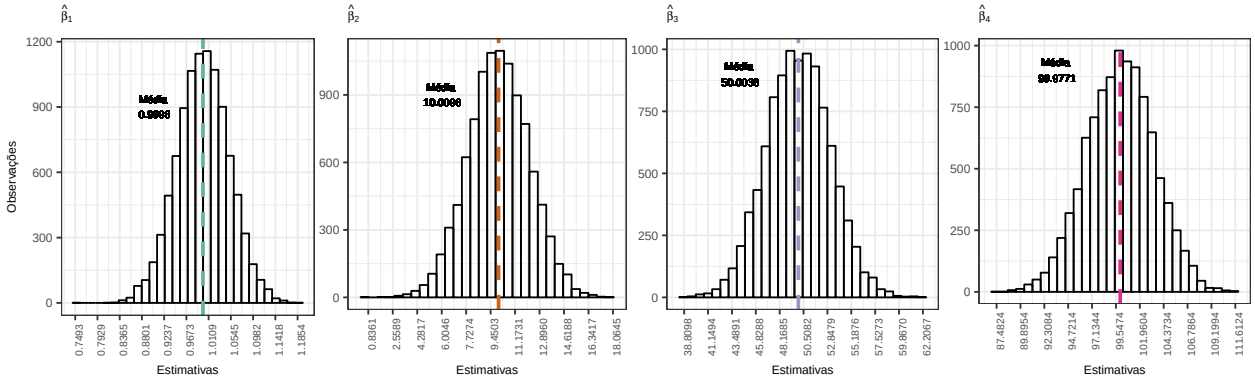
Figura 17 – Comparação das estimativas usando a Priori 1 no experimento 2 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



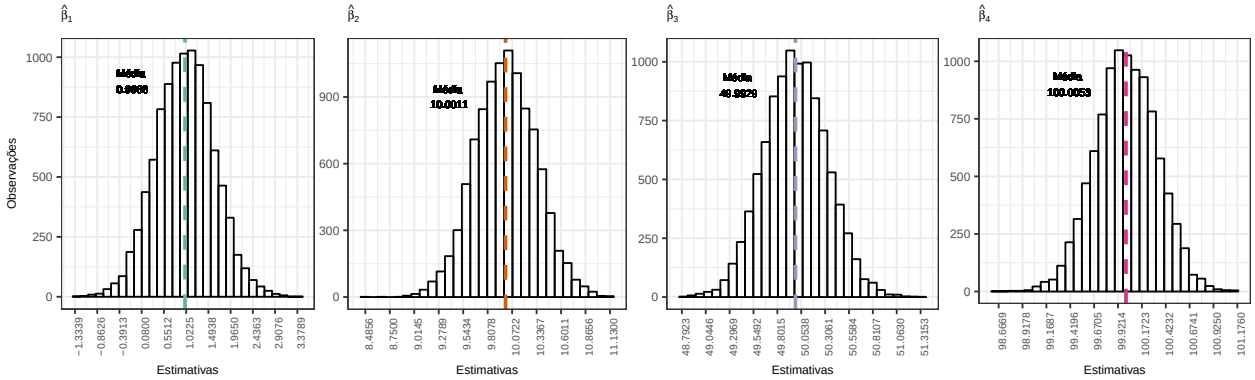
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



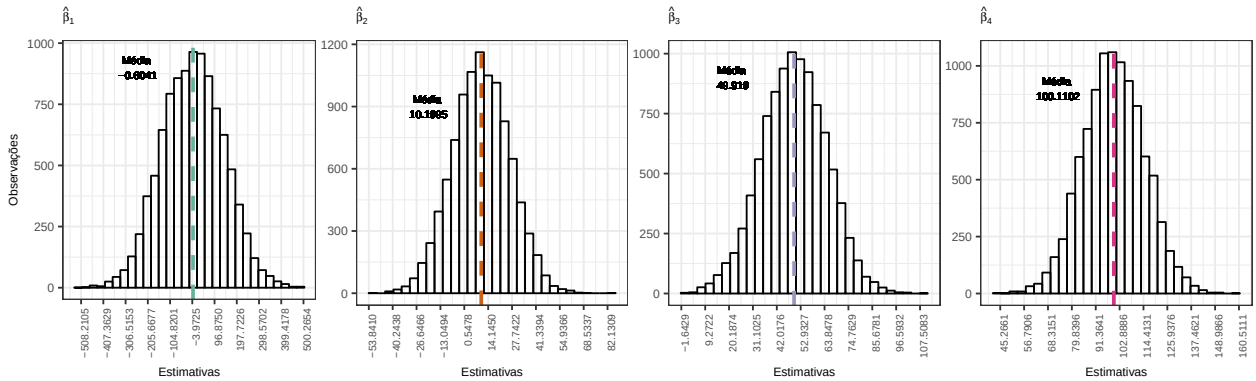
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



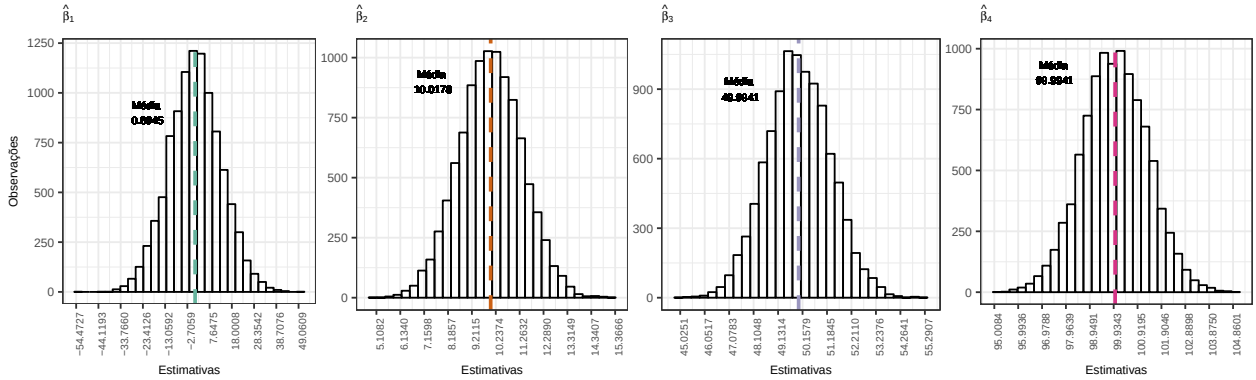
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

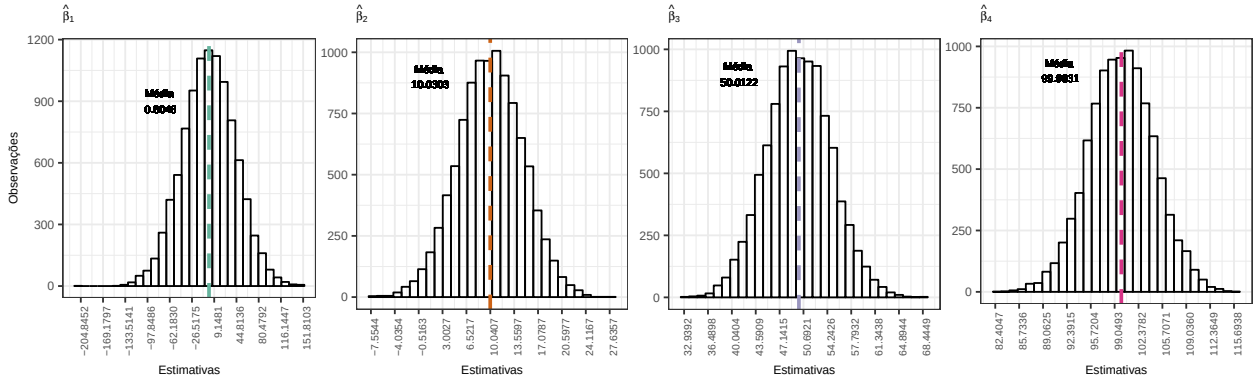
Figura 18 – Comparação das estimativas usando a Priori 2 no experimento 2 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



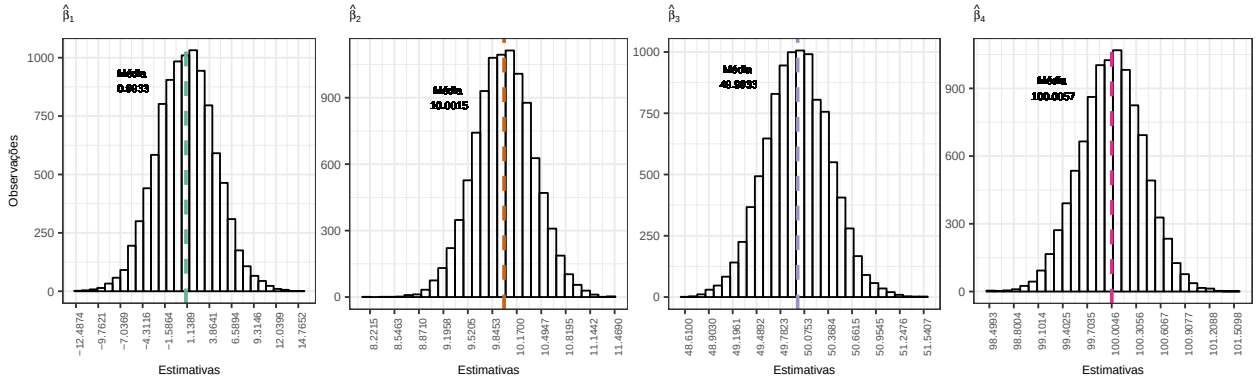
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



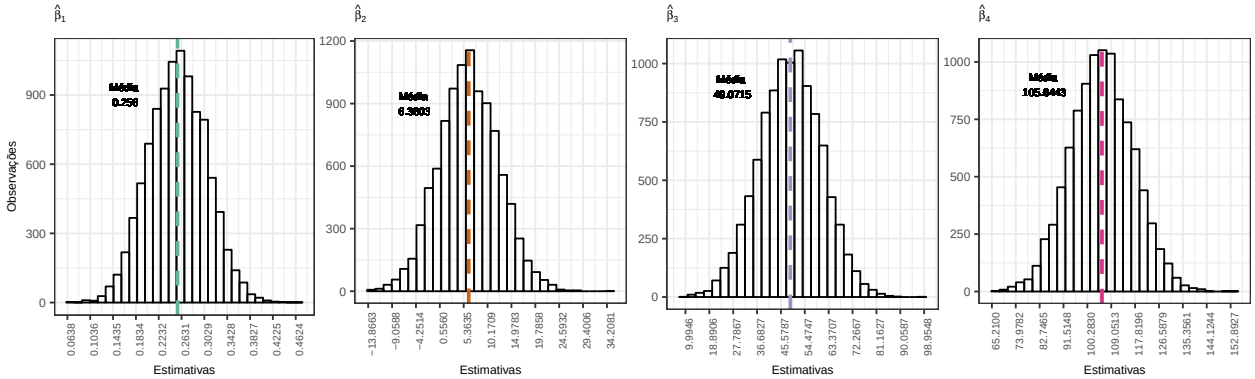
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



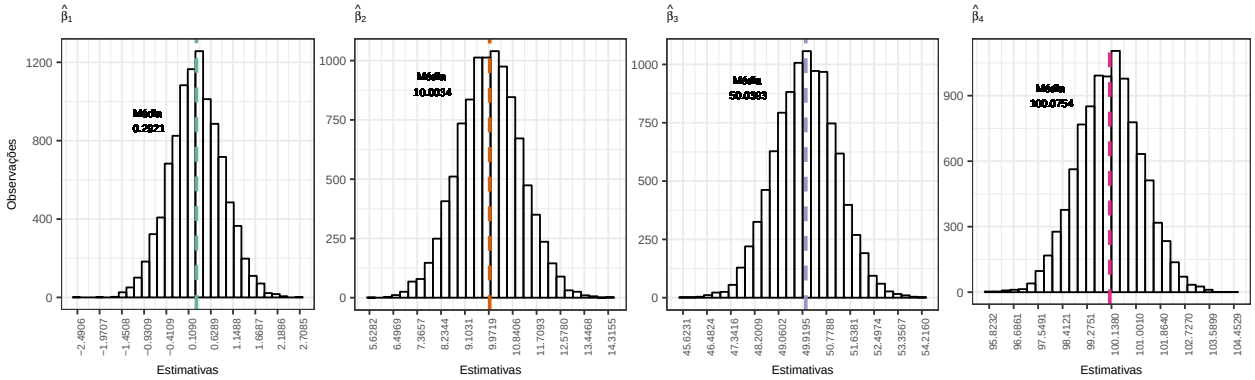
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

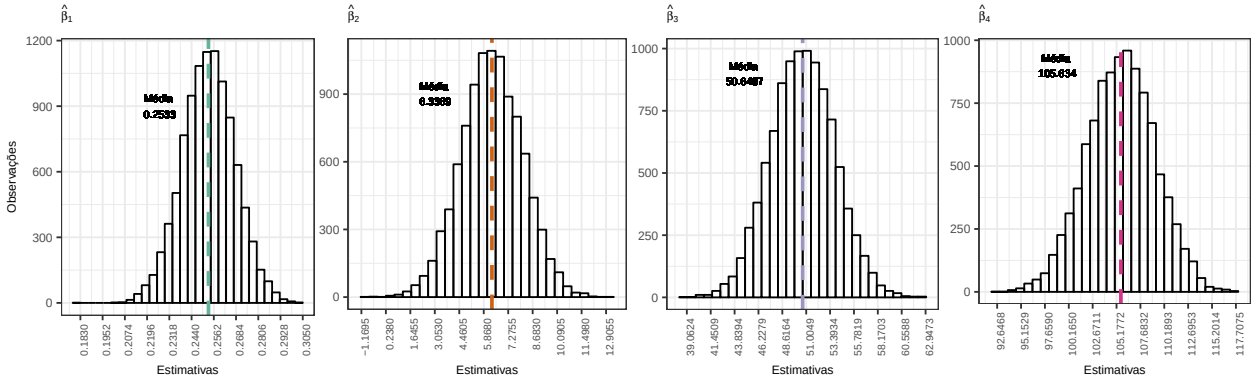
Figura 19 – Comparação das estimativas usando a Priori 3 no experimento 2 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



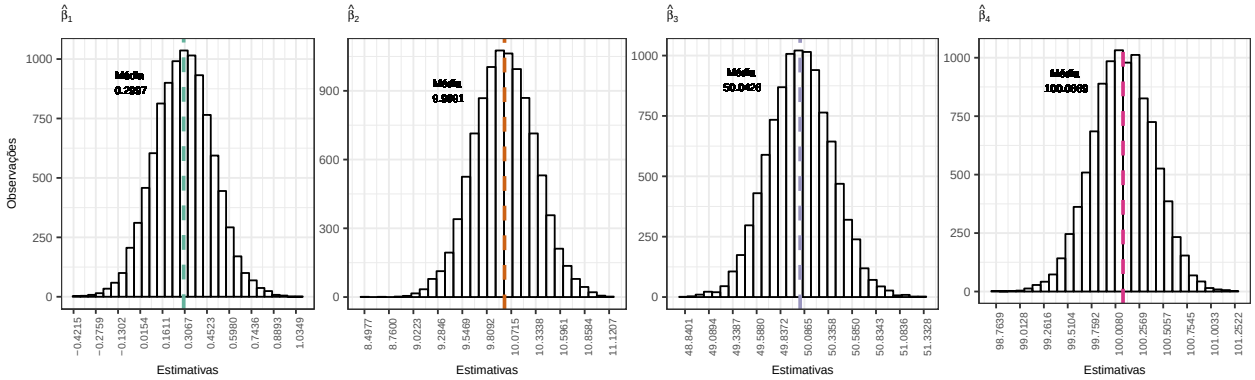
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



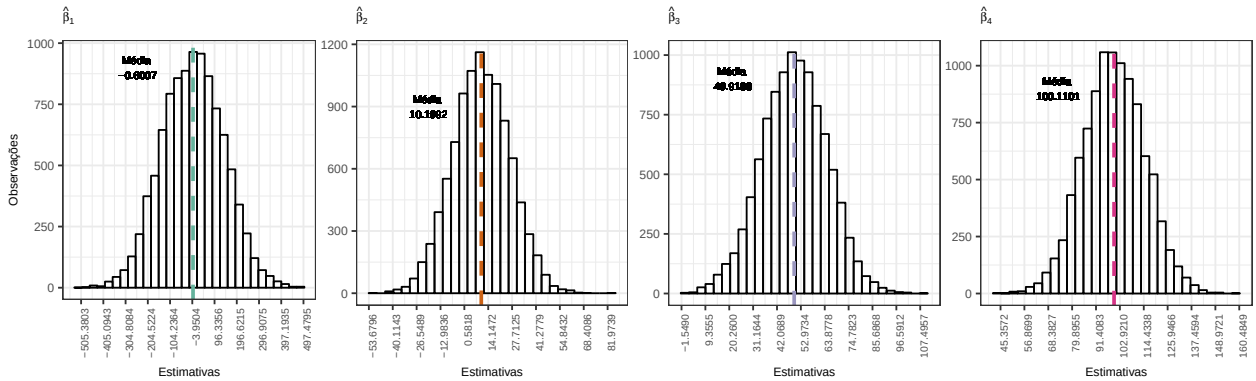
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



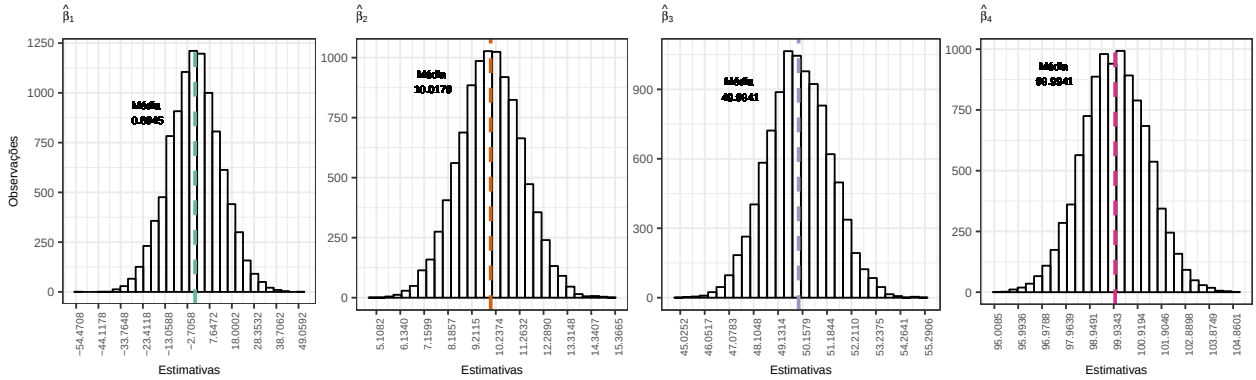
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

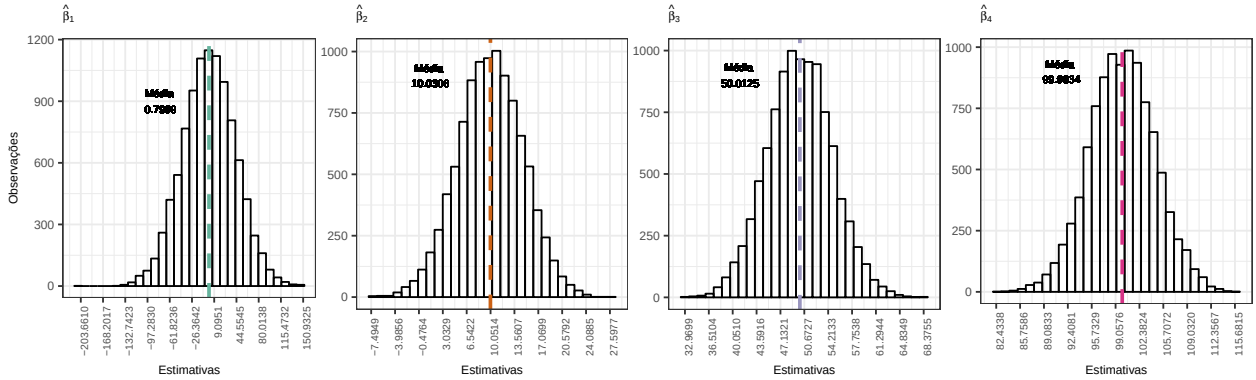
Figura 20 – Comparação das estimativas usando a Priori 4 no experimento 2 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



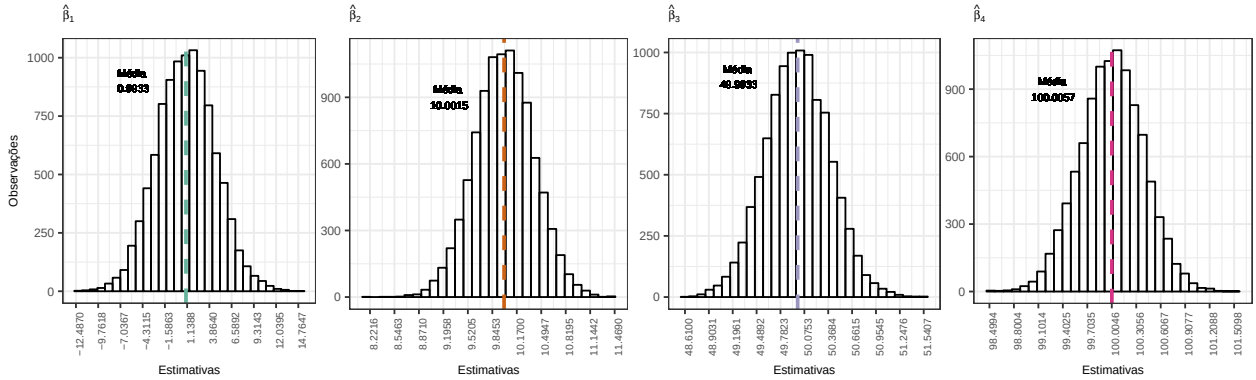
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

EXPERIMENTO 3

Tabela 13 – Resultados do terceiro experimento - considerando os pesos dos erros menores

	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	10		50		100	
			$n = 30$			
MQO	10.04814	224.96702	49.76318	172.97366	100.06944	269.4888
Priori 1	10.02657	64.89359	49.79405	129.54403	100.06462	146.09466
Priori 2	10.04814	224.96605	49.76318	172.97353	100.06944	269.48812
Priori 3	26.04153	399.20855	4.84525	2040.12085	115.09885	369.78681
Priori 4	10.04836	224.96657	49.76181	172.96465	100.07018	269.4856
			$n = 100$			
MQO	10.12464	85.46367	49.90171	59.13473	99.98805	68.2877
Priori 1	10.0941	49.19495	49.91644	49.2975	99.99923	58.07817
Priori 2	10.12464	85.46354	49.90171	59.1347	99.98805	68.28767
Priori 3	34.31701	650.14648	11.17949	1509.71593	111.63509	188.97032
Priori 4	10.12491	85.46336	49.90125	59.13373	99.98821	68.28753
			$n = 500$			
MQO	9.97857	12.3469	50.02496	12.99326	100.02765	11.38997
Priori 1	9.97961	11.27444	50.02434	12.49325	100.02728	11.11953
Priori 2	9.97857	12.3469	50.02496	12.99326	100.02765	11.38997
Priori 3	20.75218	125.7214	27.86317	493.98066	107.98328	73.69196
Priori 4	9.97862	12.34689	50.02486	12.9932	100.02768	11.38996
			$n = 1000$			
MQO	9.93827	6.70745	50.01543	6.83776	100.05116	6.69846
Priori 1	9.93974	6.39219	50.01481	6.71242	100.05046	6.59788
Priori 2	9.93827	6.70745	50.01543	6.83776	100.05116	6.69846
Priori 3	16.25636	45.1577	35.55791	211.98902	105.91828	41.06226
Priori 4	9.93829	6.70744	50.01537	6.83774	100.05119	6.69846
			$n = 5000$			
MQO	9.99413	1.28774	50.00489	1.26328	100.0015	1.29327
Priori 1	9.99416	1.2757	50.00488	1.25894	100.00149	1.28942
Priori 2	9.99413	1.28774	50.00489	1.26328	100.0015	1.29327
Priori 3	11.51987	3.56435	46.48543	13.44144	101.48298	3.45791
Priori 4	9.99413	1.28774	50.00488	1.26328	100.0015	1.29327

Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;
⋆ *EQM* significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

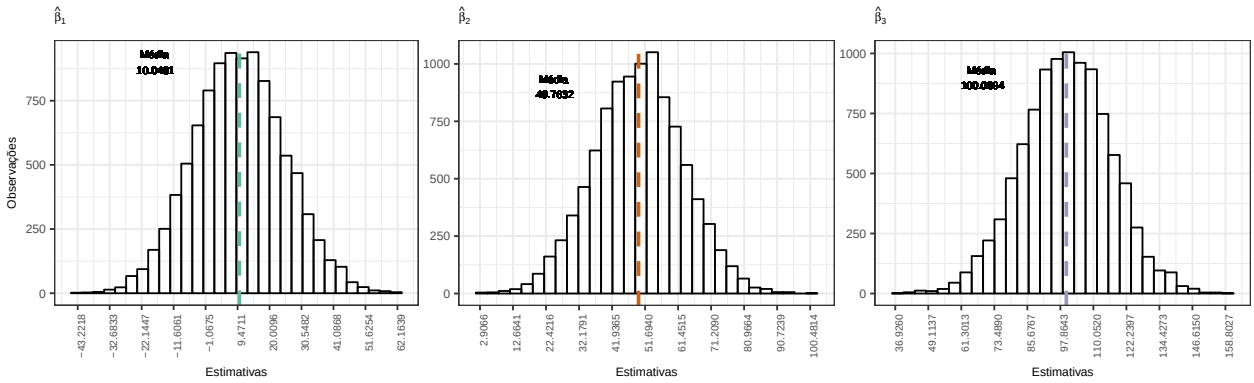
Tabela 14 – Resultados do terceiro experimento - considerando os pesos dos erros maiores

	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	10		50		100	
			$n = 30$			
MQO	9.95383	20.39645	50.05203	18.34817	99.99647	18.82007
Priori 1	9.97495	6.37054	50.0388	12.13435	99.99259	13.78209
Priori 2	9.95383	20.39637	50.05203	18.34814	99.99647	18.82005
Priori 3	36.0425	690.21454	4.57825	2063.24502	115.99915	267.01085
Priori 4	9.95452	20.39614	50.05056	18.34692	99.99716	18.81981
			$n = 100$			
MQO	9.98123	5.28365	49.99684	4.52268	100.01209	5.57126
Priori 1	9.98492	3.37058	49.99594	3.96936	100.00981	4.72433
Priori 2	9.98123	5.28364	49.99684	4.52268	100.01209	5.57126
Priori 3	26.87065	288.73484	11.55651	1478.12046	114.71915	220.86085
Priori 4	9.9814	5.28363	49.9964	4.5226	100.01228	5.57125
			$n = 500$			
MQO	10.01243	1.15588	50.00354	1.02322	99.97727	1.05596
Priori 1	10.01185	1.05032	50.00368	0.99104	99.97763	1.02435
Priori 2	10.01243	1.15588	50.00354	1.02322	99.97727	1.05596
Priori 3	19.12075	84.19767	29.18727	433.51097	108.34099	70.48895
Priori 4	10.01247	1.15588	50.00345	1.02322	99.97731	1.05596
			$n = 1000$			
MQO	9.98303	0.51657	49.99915	0.50319	100.01444	0.53726
Priori 1	9.98341	0.49353	49.99903	0.49549	100.01423	0.52917
Priori 2	9.98303	0.51657	49.99915	0.50319	100.01444	0.53726
Priori 3	15.38282	29.45245	36.41309	184.86807	105.88281	35.08961
Priori 4	9.98305	0.51657	49.9991	0.50319	100.01446	0.53726
			$n = 5000$			
MQO	9.9999	0.10979	50.00065	0.10601	100.0003	0.10416
Priori 1	9.9999	0.10873	50.00064	0.10561	100.0003	0.10386
Priori 2	9.9999	0.10979	50.00065	0.10601	100.0003	0.10416
Priori 3	11.69931	2.99408	46.41476	12.94507	101.36442	1.96335
Priori 4	9.9999	0.10979	50.00064	0.10601	100.00031	0.10416

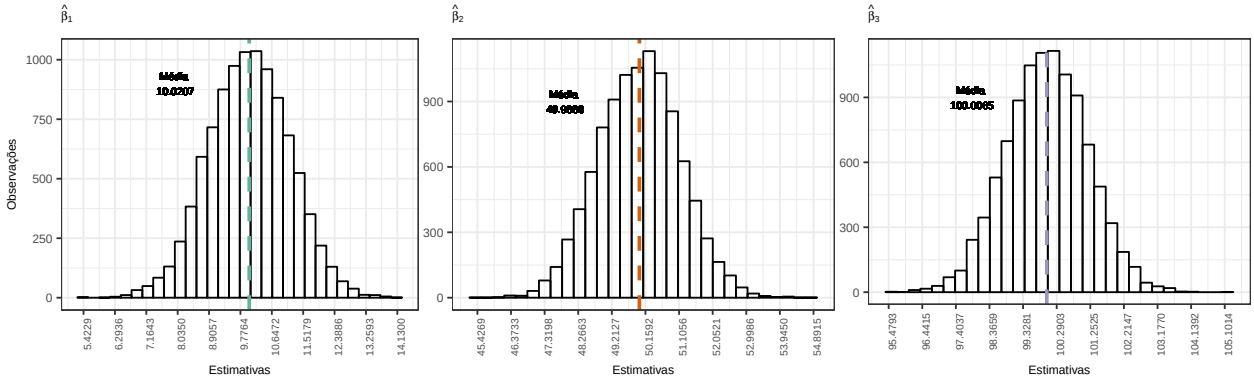
Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pelo parâmetro β_j;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

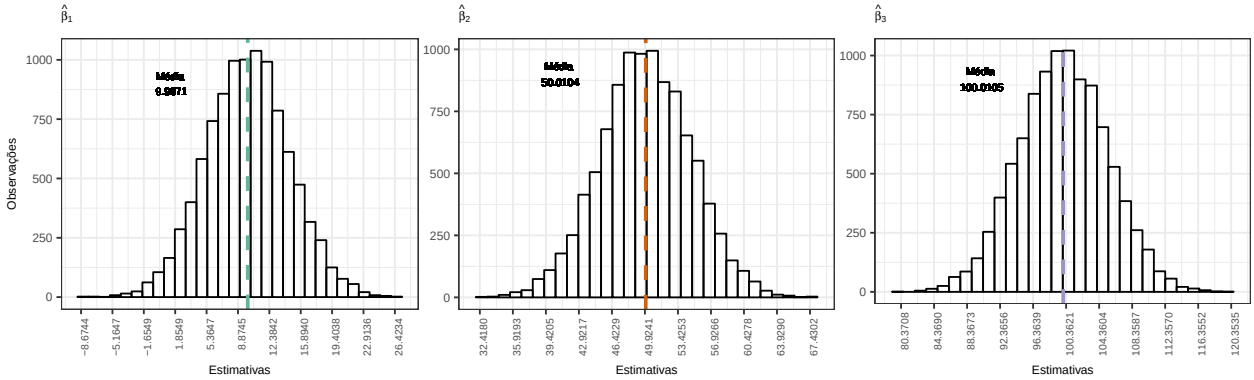
Figura 21 – Comparação das estimativas de MQO no experimento 3 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



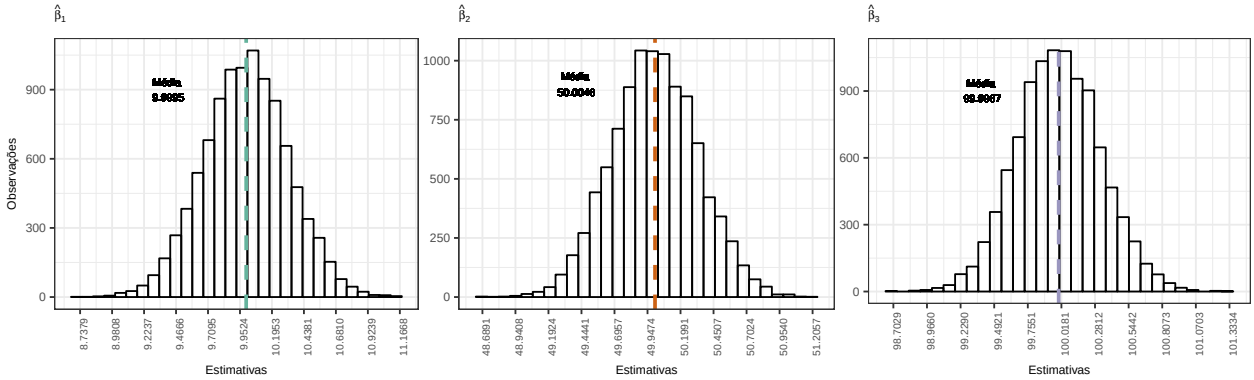
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



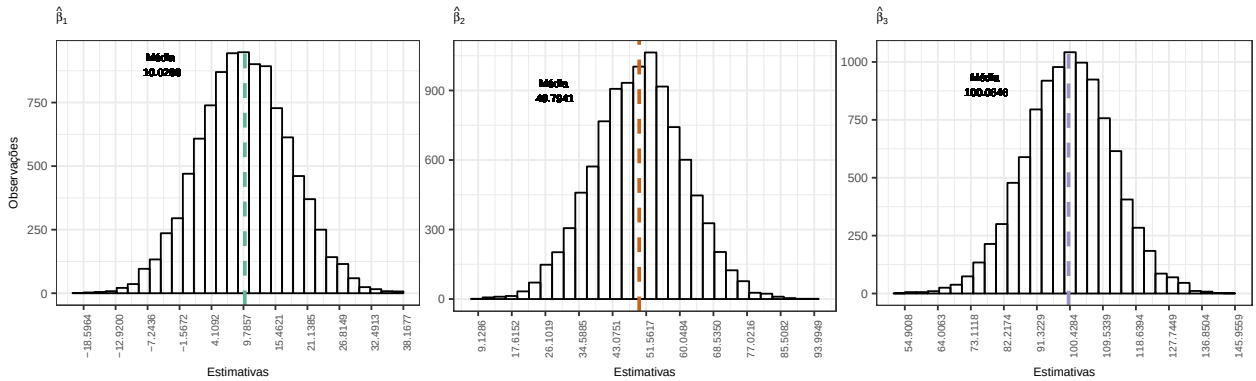
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



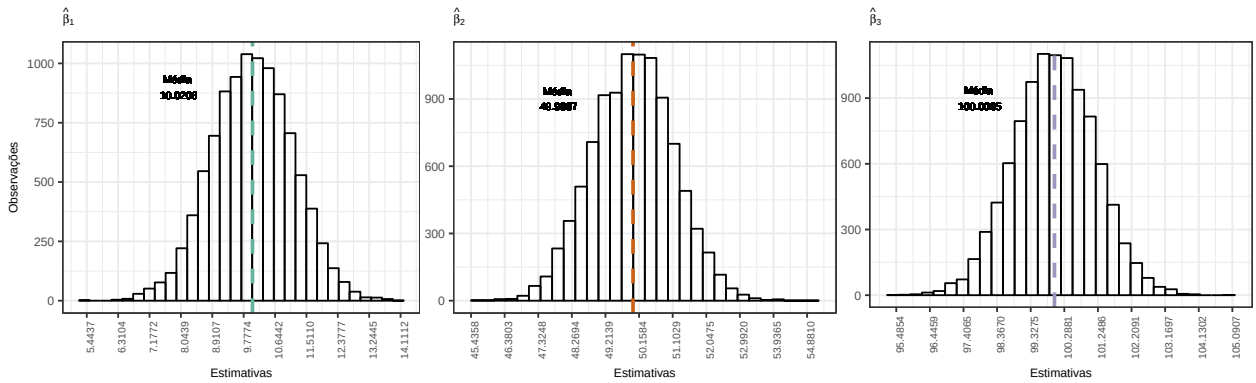
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

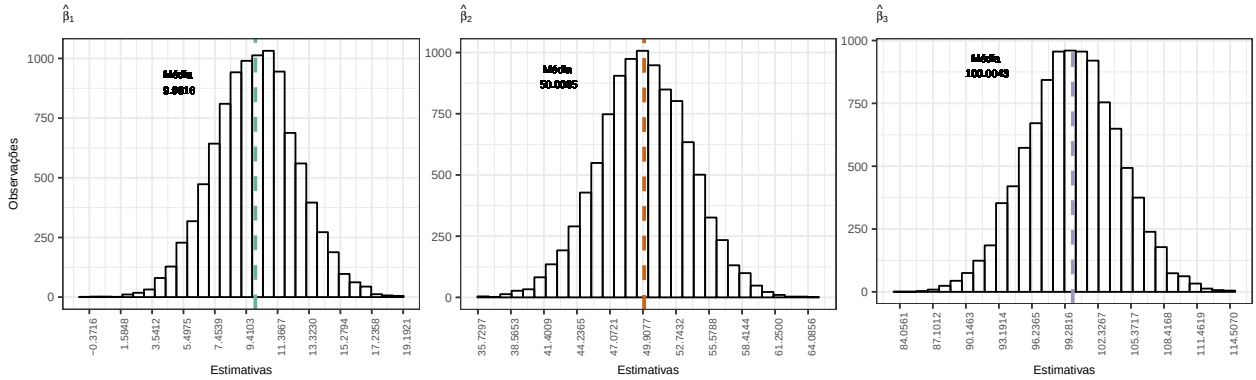
Figura 22 – Comparação das estimativas usando a Priori 1 no experimento 3 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



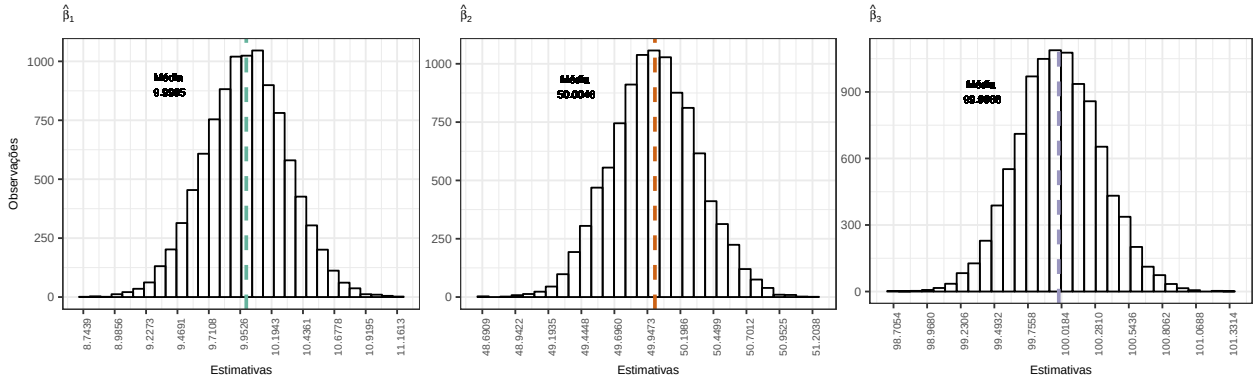
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



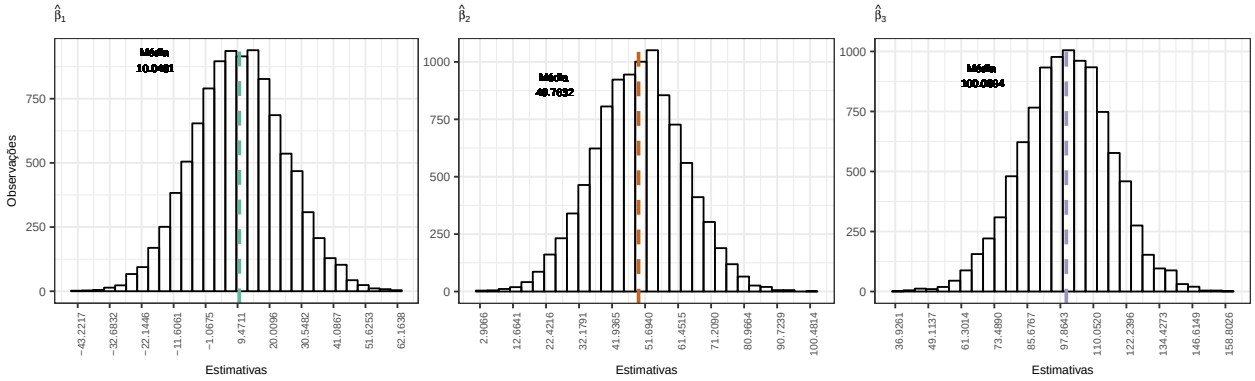
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



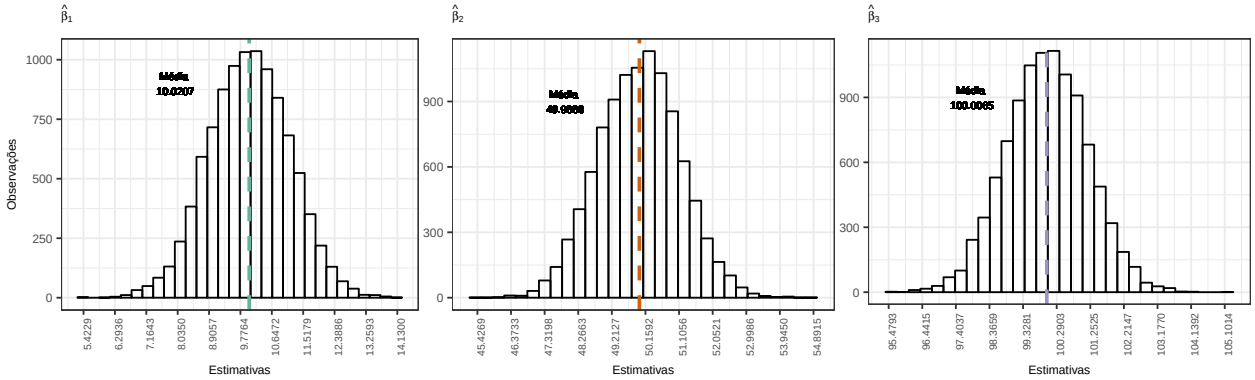
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em dados simulados.

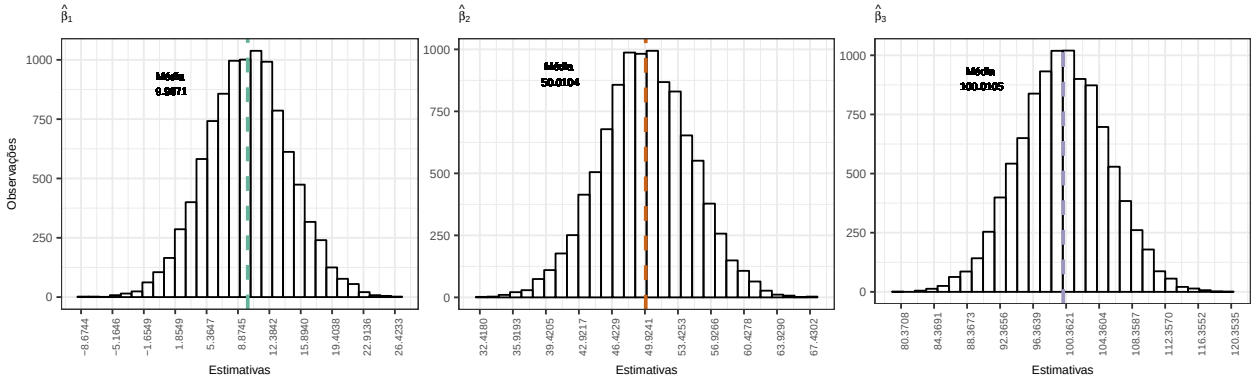
Figura 23 – Comparação das estimativas usando a Priori 2 no experimento 3 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



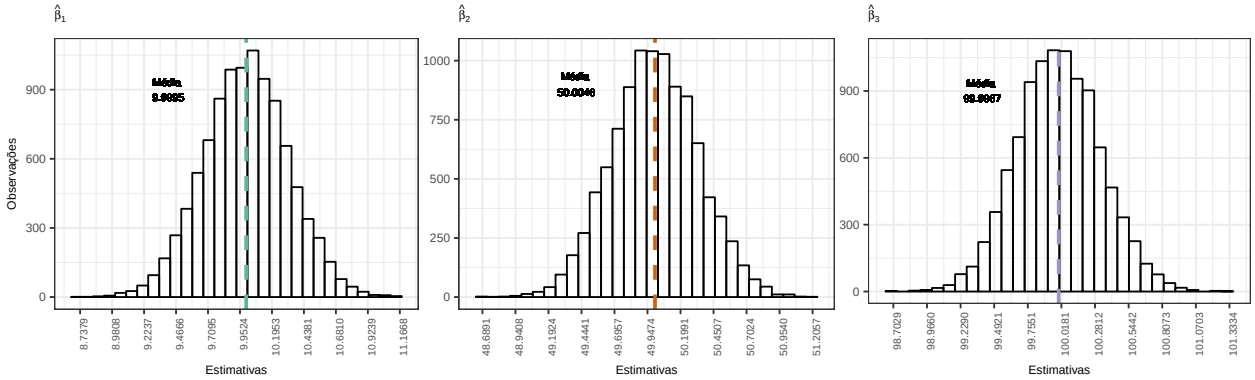
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



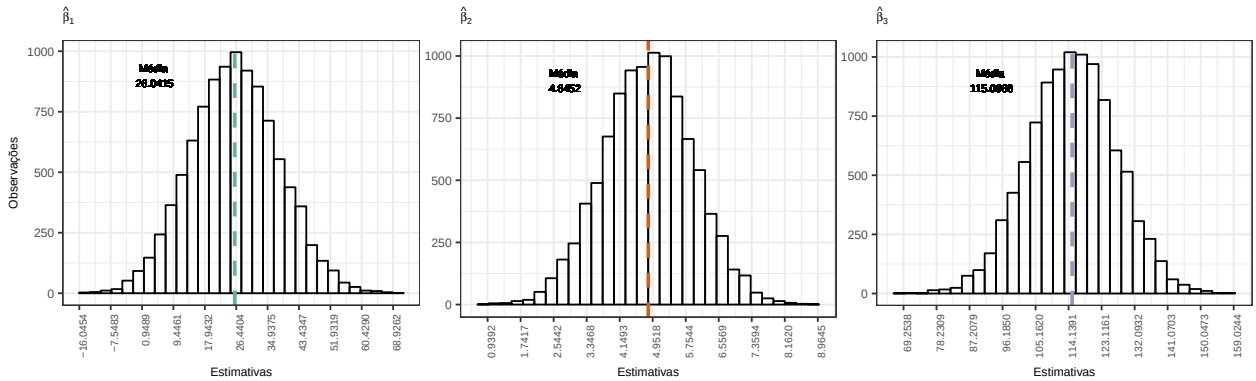
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



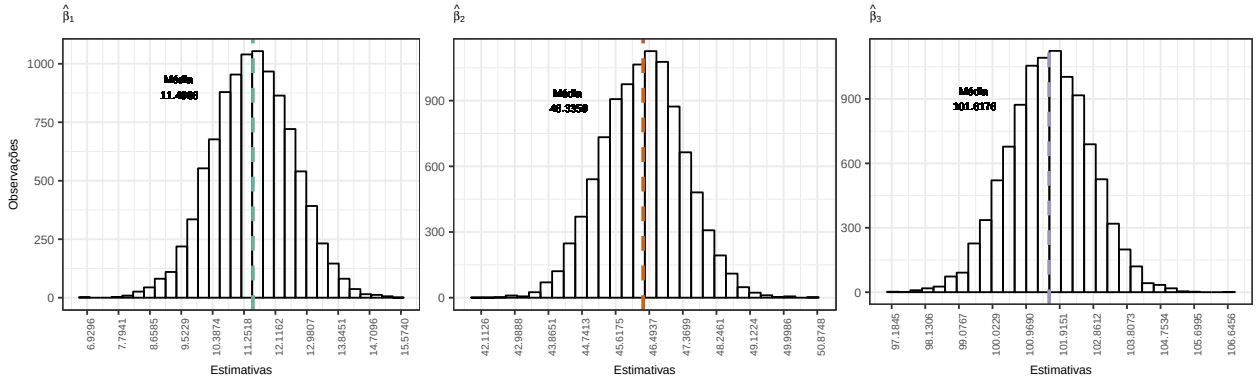
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

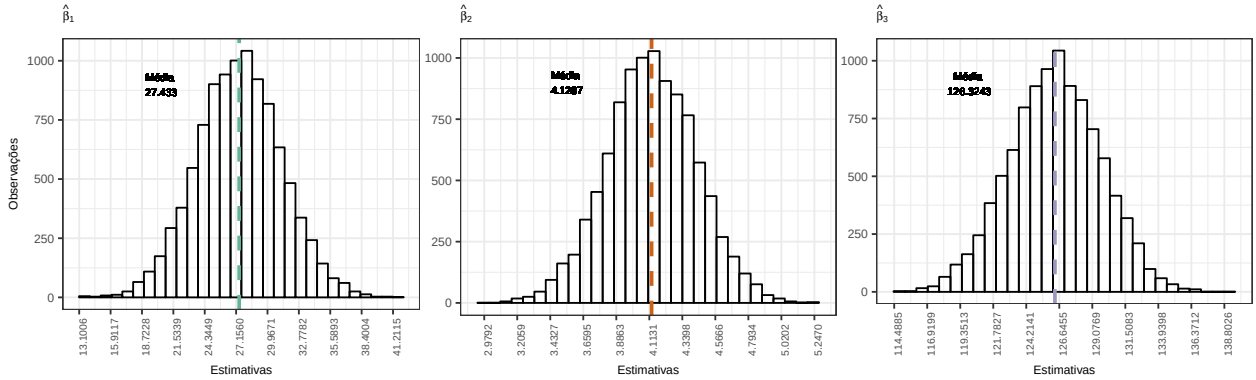
Figura 24 – Comparação das estimativas usando a Priori 3 no experimento 3 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



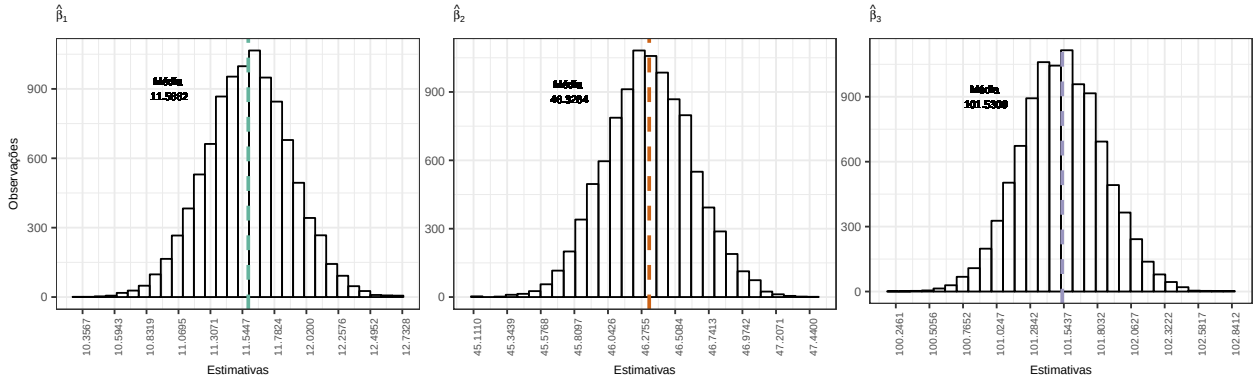
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



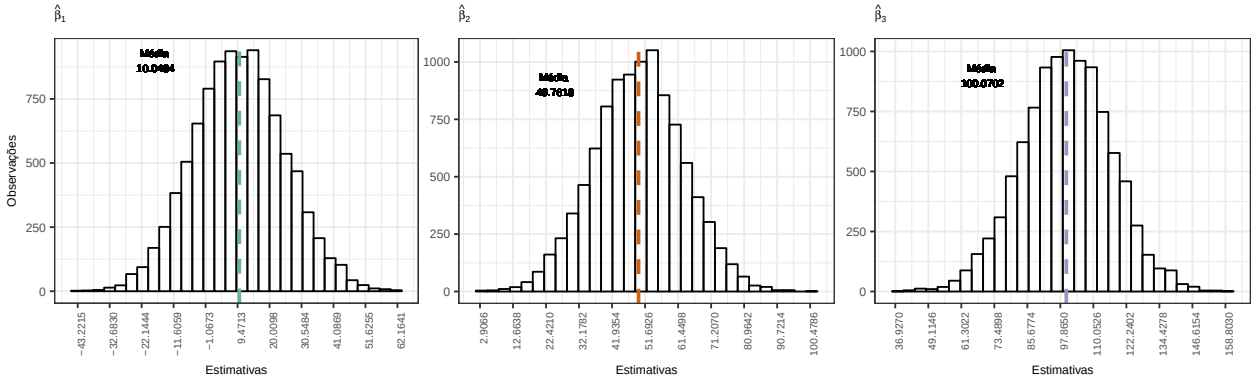
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



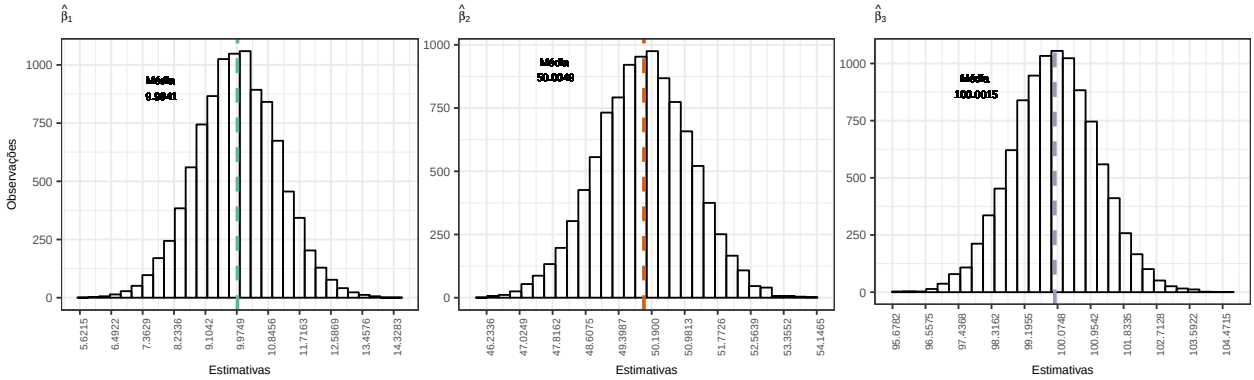
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

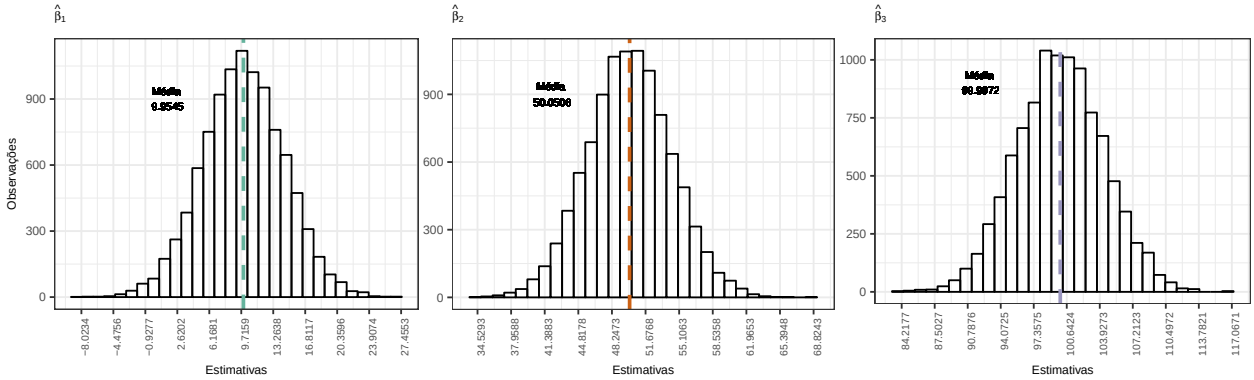
Figura 25 – Comparação das estimativas usando a Priori 4 no experimento 3 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



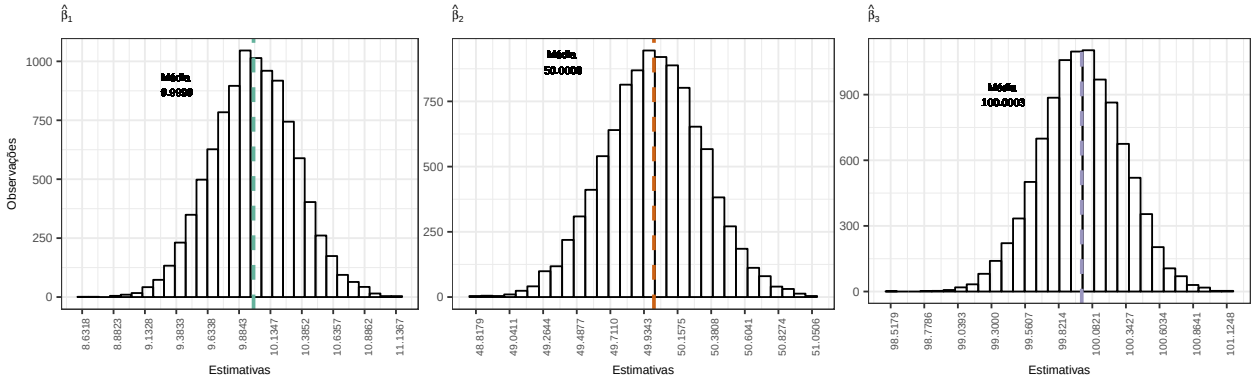
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

EXPERIMENTO 4

Tabela 15 – Resultados do quarto experimento - considerando os pesos dos erros menores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	0.1		0.1		100		100	
				$n = 30$				
QO	-3.00623	33385.75277	0.55689	450.05936	100.06345	427.73994	100.24769	414.83177
Priori 1	0.09989	0.0211	0.10269	0.03233	100.09859	163.12846	100.08134	235.49728
Priori 2	-2.92673	31707.49458	0.55157	442.46865	100.0599	424.22012	100.24295	408.99547
Priori 3	0.03628	0.00412	0.81123	3.16682	98.22694	175.15561	102.80542	262.20627
Priori 4	-2.81135	29244.61213	0.54411	431.98472	100.05453	419.0742	100.23606	400.4179
				$n = 100$				
MQO	-1.92748	10761.60624	0.27474	135.00981	100.17747	133.94397	100.03599	94.32509
Priori 1	0.09951	0.03776	0.1005	0.10856	100.05892	71.9548	99.96641	61.25633
Priori 2	-1.9082	10557.98081	0.27318	133.66121	100.17632	133.19426	100.03526	94.04652
Priori 3	0.03554	0.00431	0.55008	5.92512	99.46083	74.79234	100.44206	63.92212
Priori 4	-1.87944	10241.08947	0.27089	131.61987	100.17459	132.01725	100.03417	93.60888
				$n = 500$				
MQO	0.06952	2640.61064	0.1341	28.90009	99.96871	31.14333	99.98743	32.45653
Priori 1	0.10022	0.08084	0.1047	0.44319	99.97854	16.65864	99.99646	18.48732
Priori 2	0.06958	2631.0287	0.1341	28.86903	99.9687	31.10594	99.98742	32.41872
Priori 3	0.03538	0.00494	0.37384	8.33488	99.78424	18.19354	100.05101	19.75974
Priori 4	0.06935	2615.58149	0.13411	28.8217	99.96872	31.04537	99.98744	32.35746
				$n = 1000$				
MQO	-0.06331	1182.15908	0.11813	14.39823	100.03012	14.94464	99.97349	14.7419
Priori 1	0.09964	0.11182	0.1022	0.64859	100.0225	8.12801	99.96705	8.70714
Priori 2	-0.06318	1180.17529	0.11812	14.3915	100.03011	14.93656	99.97348	14.73516
Priori 3	0.03507	0.00576	0.26281	6.20639	99.91305	9.25644	99.9806	9.80589
Priori 4	-0.0631	1176.96557	0.11812	14.38129	100.0301	14.92341	99.97348	14.72419
				$n = 5000$				
MQO	-0.03817	237.14444	0.11362	2.80696	100.01285	2.88899	100.0008	2.90791
Priori 1	0.09824	0.22472	0.10337	0.7886	100.0058	1.87126	99.99342	1.85258
Priori 2	-0.03815	237.06184	0.11362	2.80663	100.01285	2.88872	100.0008	2.90762
Priori 3	0.03439	0.01134	0.1437	1.71945	99.98519	2.03795	99.99637	2.0271
Priori 4	-0.03814	236.92795	0.11362	2.80614	100.01284	2.88826	100.0008	2.90714

Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

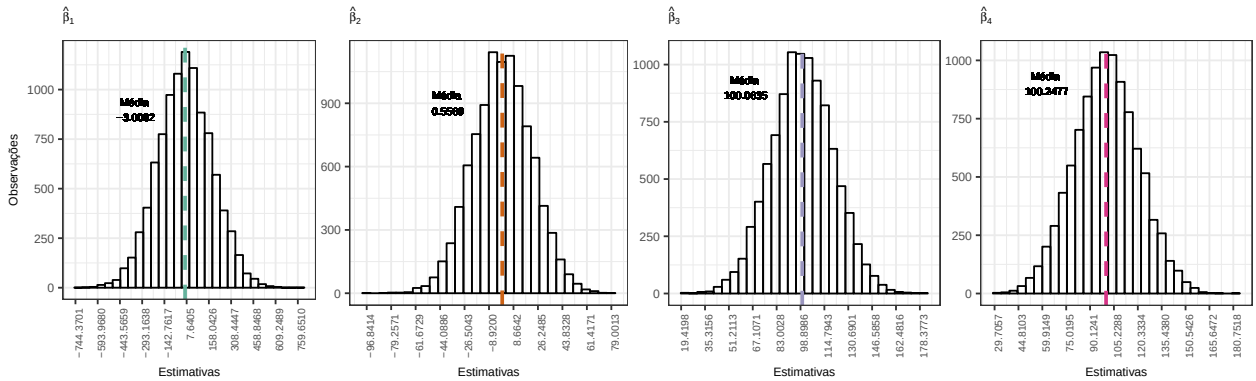
Tabela 16 – Resultados do quarto experimento - considerando os pesos dos erros maiores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	0.1		0.1		100		100	
				$n = 30$				
MQO	0.03328	3236.45112	0.24646	56.50916	99.94122	34.81807	99.94257	34.90852
Priori 1	0.10006	0.00601	0.10102	0.00212	99.98048	13.87752	100.00926	14.55873
Priori 2	0.03532	3064.83763	0.24627	55.66927	99.9411	34.03489	99.9425	34.48073
Priori 3	0.03616	0.00408	0.71798	0.56644	99.10049	15.12853	101.4697	17.44647
Priori 4	0.03339	2814.45459	0.24647	54.55792	99.9412	32.8841	99.94256	33.85663
				$n = 100$				
MQO	-0.15299	740.42528	0.14166	10.03116	99.99017	10.90651	100.01877	9.73643
Priori 1	0.09996	0.01142	0.10092	0.00884	99.98559	5.5928	100.01365	5.56739
Priori 2	-0.15126	730.38891	0.14155	9.99618	99.99006	10.86127	100.01867	9.70128
Priori 3	0.03585	0.00413	0.60556	0.76792	99.41285	6.11143	100.49636	6.00337
Priori 4	-0.14967	714.5274	0.14147	9.9453	99.98995	10.78942	100.01857	9.64545
				$n = 500$				
MQO	-0.0514	200.65758	0.10062	2.40977	100.02406	2.31603	100.01115	2.44139
Priori 1	0.0997	0.02418	0.09882	0.03848	100.01172	1.25556	99.99902	1.3132
Priori 2	-0.05115	199.99721	0.1006	2.40752	100.02404	2.31348	100.01114	2.43896
Priori 3	0.03527	0.00426	0.33747	0.75818	99.84383	1.39222	100.03708	1.45704
Priori 4	-0.05103	198.93224	0.1006	2.40411	100.02403	2.30935	100.01113	2.43501
				$n = 1000$				
MQO	0.0296	101.73796	0.11394	1.30045	99.99383	1.20872	100.004	1.21061
Priori 1	0.09994	0.03351	0.10245	0.0586	99.9927	0.7001	100.00286	0.71499
Priori 2	0.02966	101.56704	0.11394	1.29973	99.99382	1.20812	100.004	1.21004
Priori 3	0.03502	0.00436	0.26018	0.55496	99.89064	0.79519	100.0136	0.80124
Priori 4	0.02961	101.29087	0.11394	1.29863	99.99383	1.20716	100.004	1.2091
				$n = 5000$				
MQO	0.1444	20.23781	0.09968	0.2492	99.99947	0.23978	99.99697	0.23648
Priori 1	0.10075	0.06835	0.10148	0.06999	100.00245	0.15023	99.99994	0.1456
Priori 2	0.14439	20.23107	0.09968	0.24917	99.99947	0.23975	99.99697	0.23645
Priori 3	0.03543	0.00483	0.14209	0.15729	99.98259	0.16658	100.00189	0.16124
Priori 4	0.14435	20.22013	0.09969	0.24913	99.99947	0.23971	99.99697	0.23641

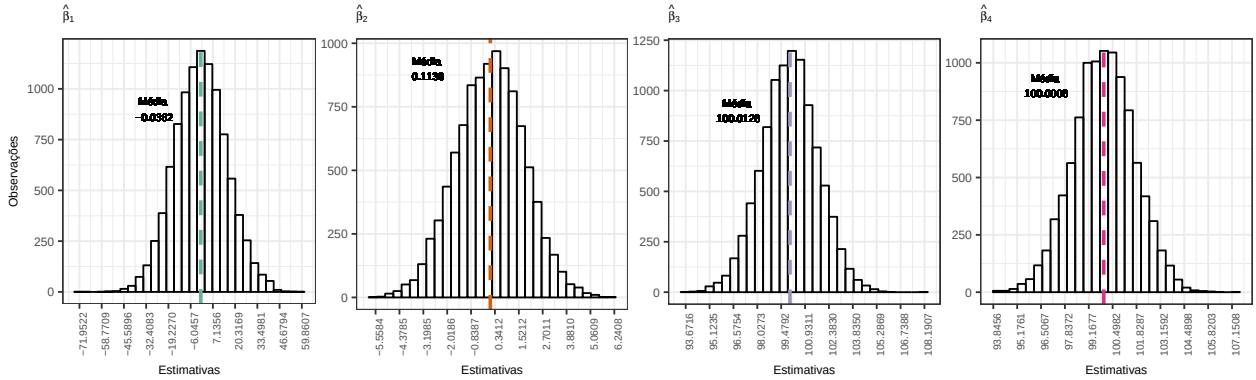
Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pelo parâmetro β_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

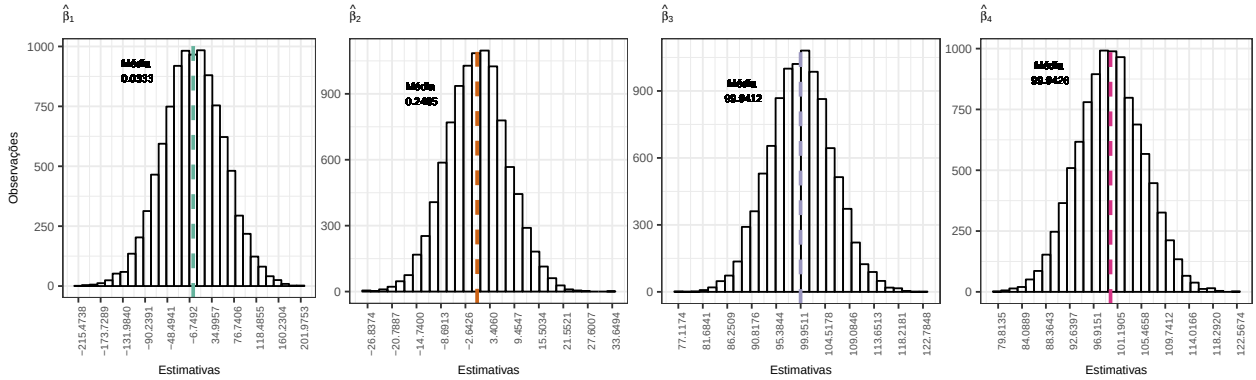
Figura 26 – Comparação das estimativas de MQO no experimento 4 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



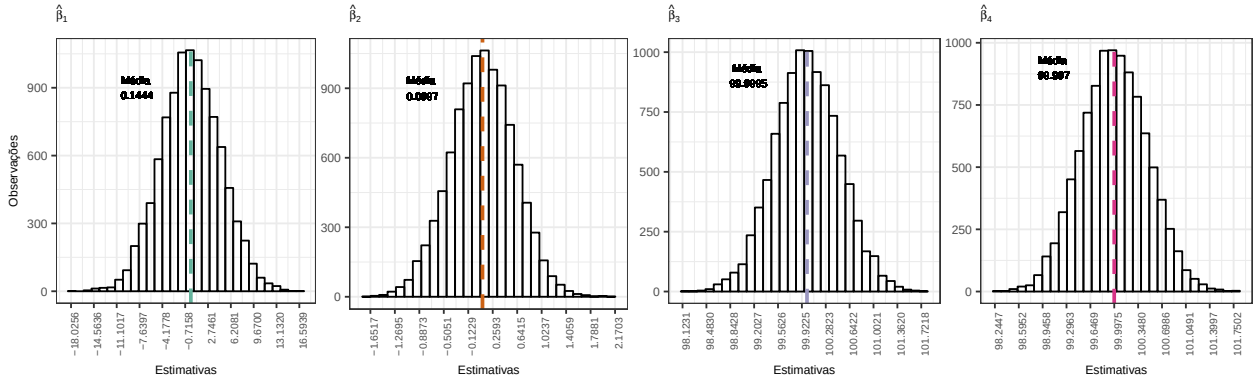
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



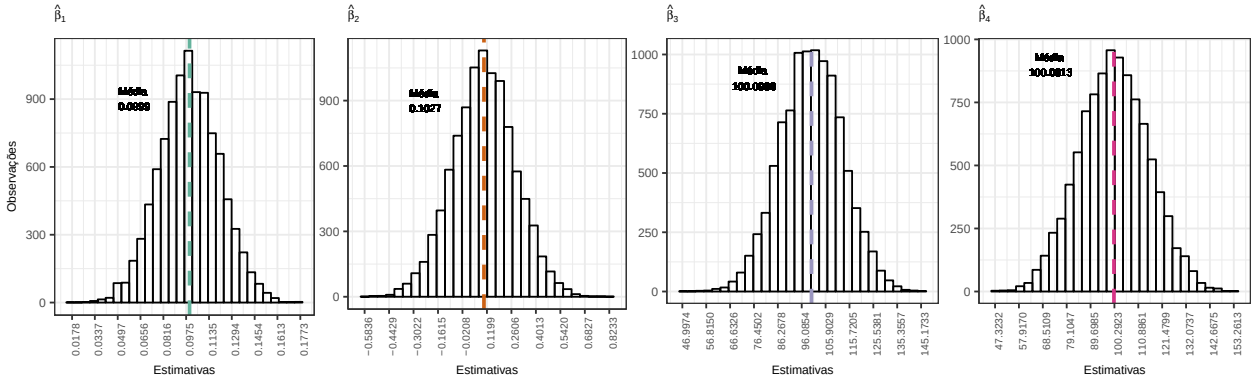
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



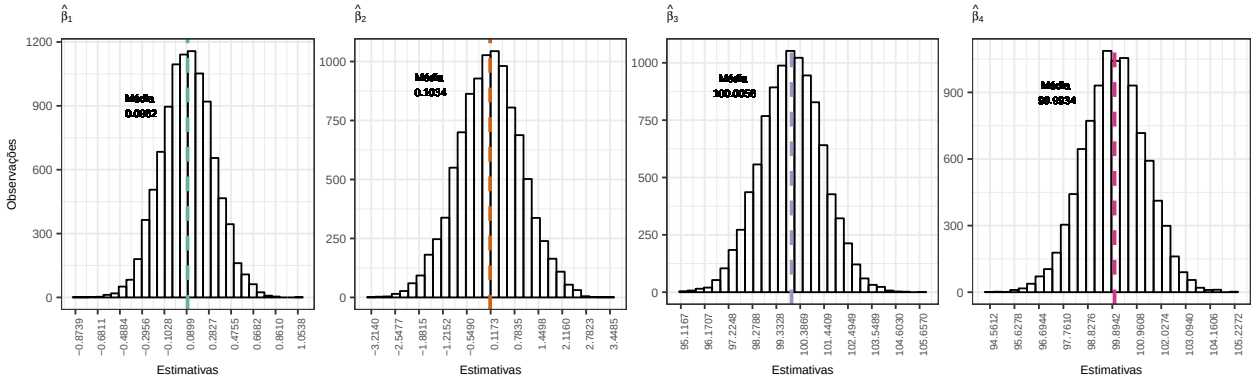
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

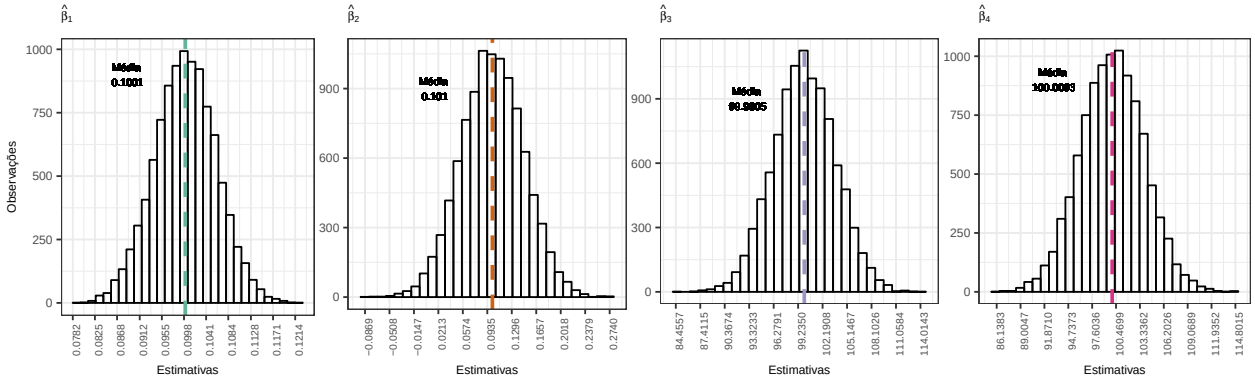
Figura 27 – Comparação das estimativas usando a Priori 1 no experimento 4 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



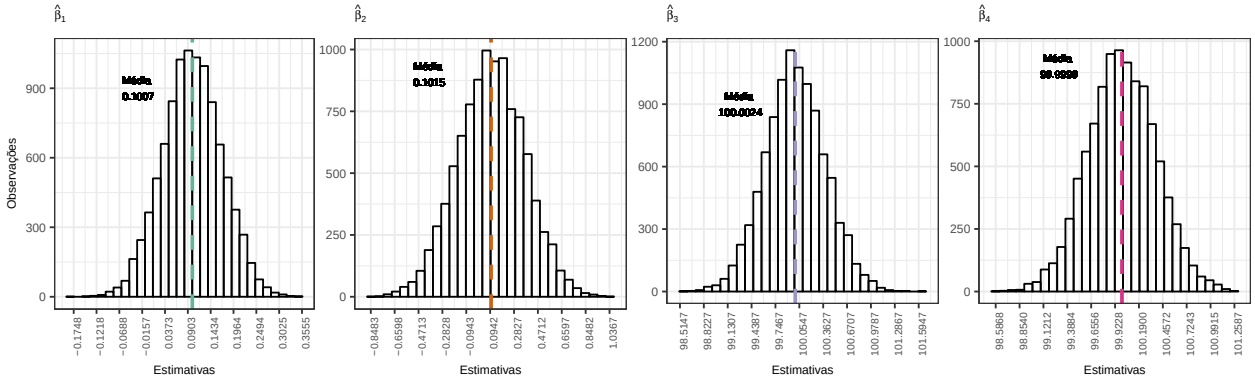
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



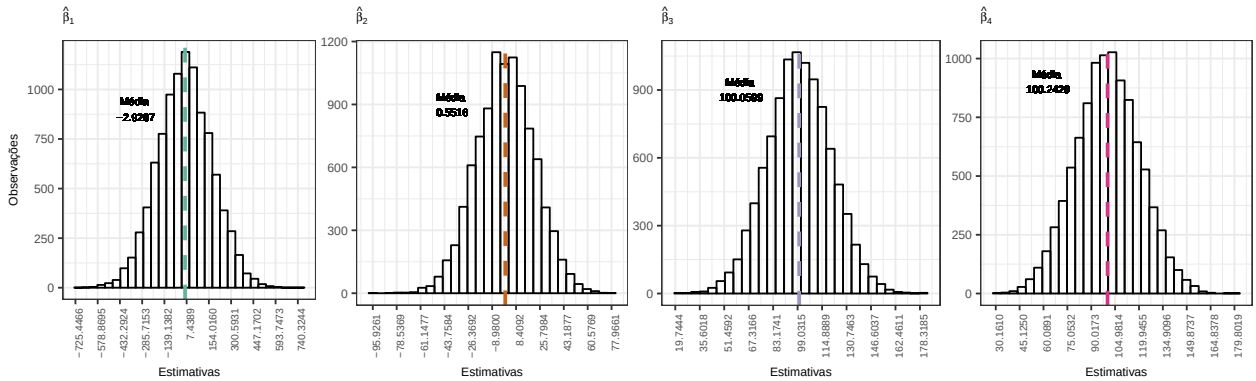
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



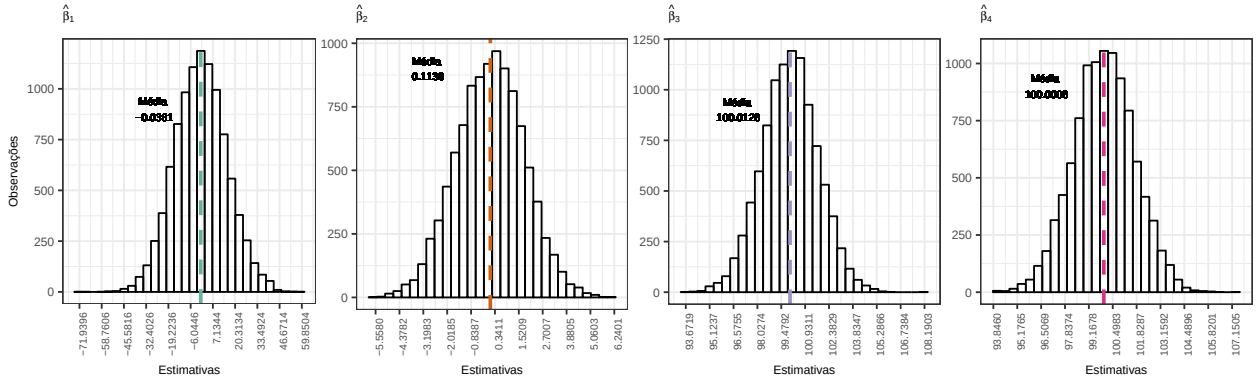
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

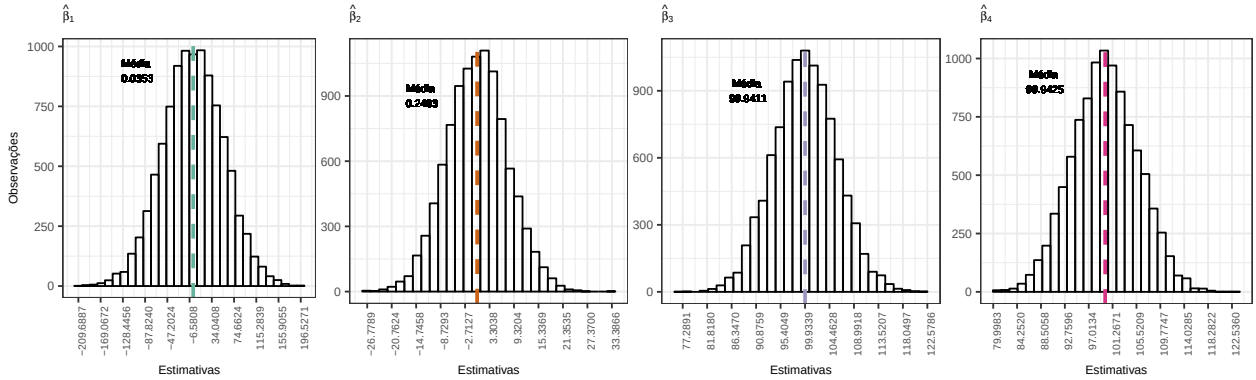
Figura 28 – Comparação das estimativas usando a Priori 2 no experimento 4 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



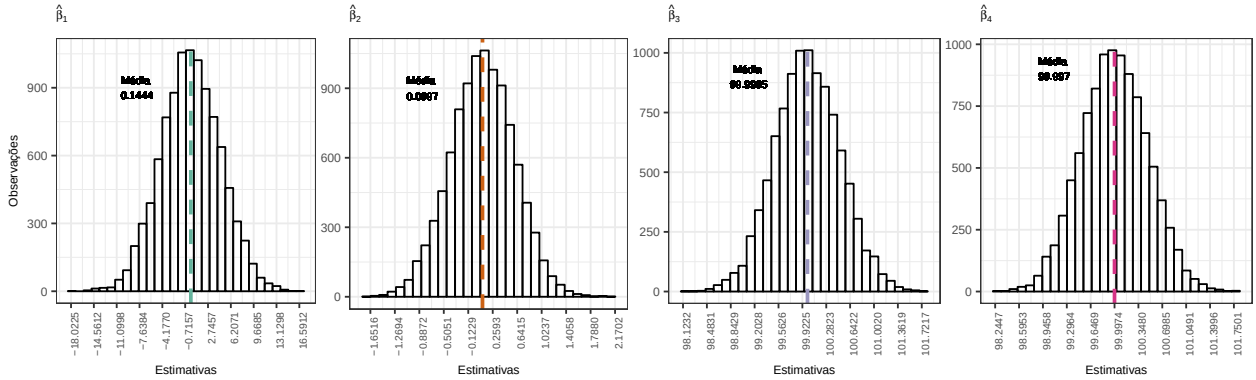
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



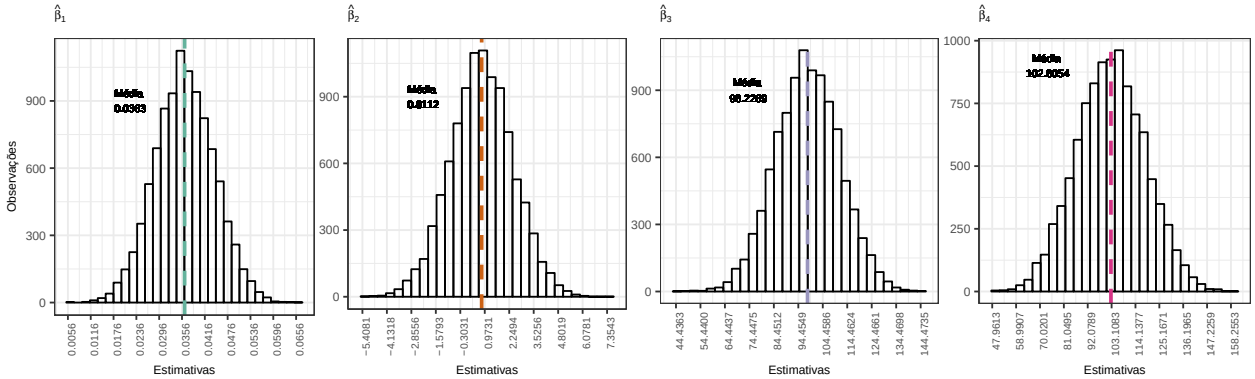
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



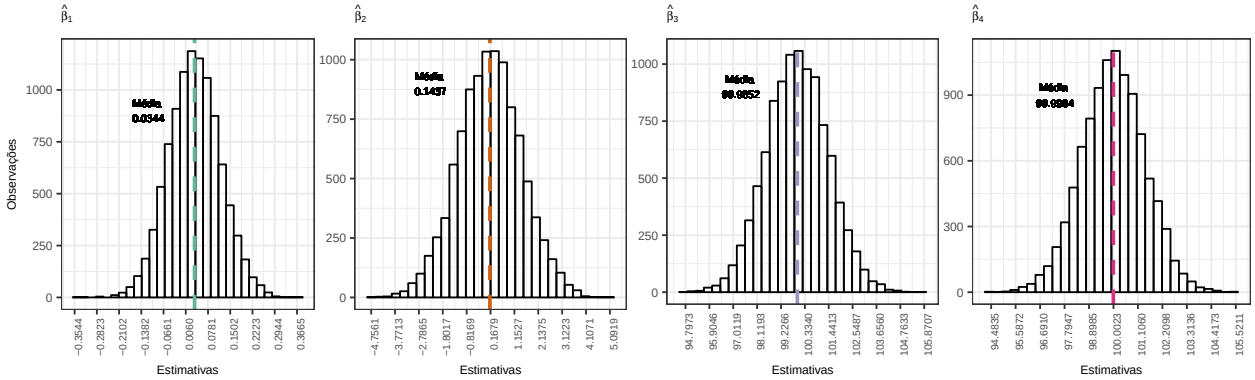
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

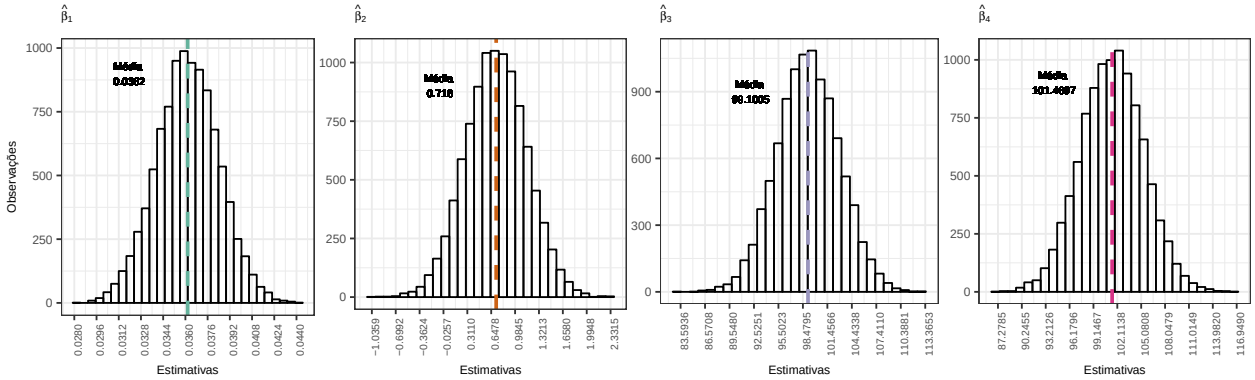
Figura 29 – Comparação das estimativas usando a Priori 3 no experimento 4 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



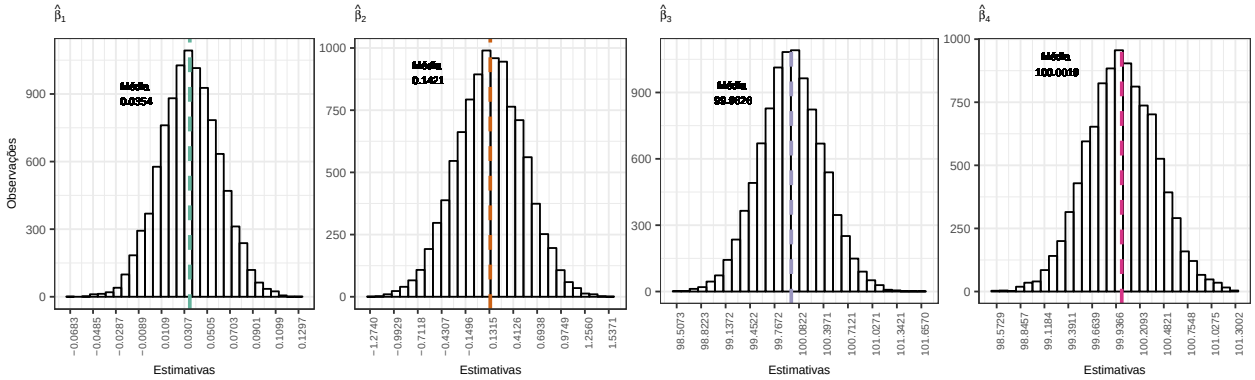
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



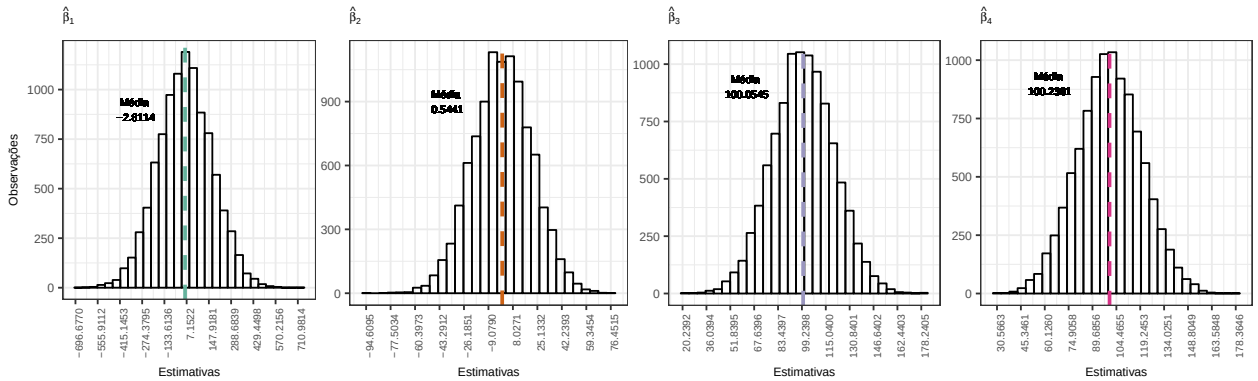
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



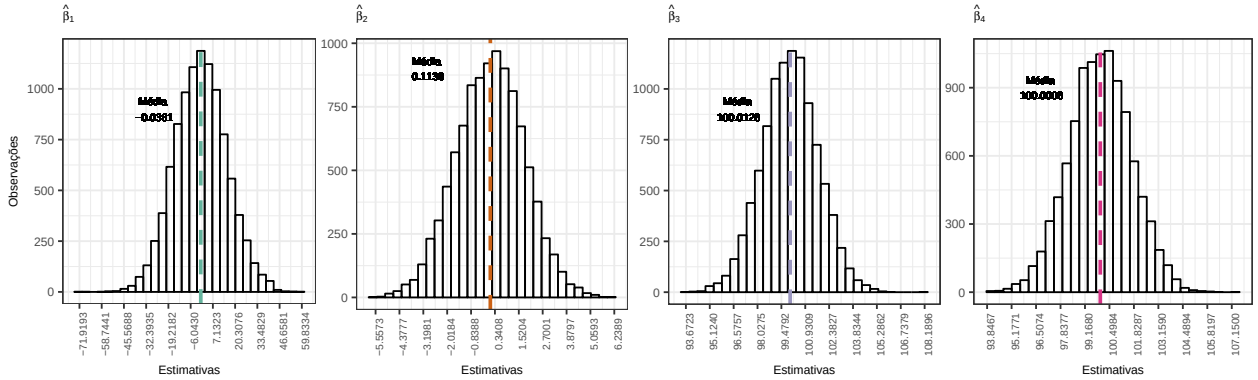
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

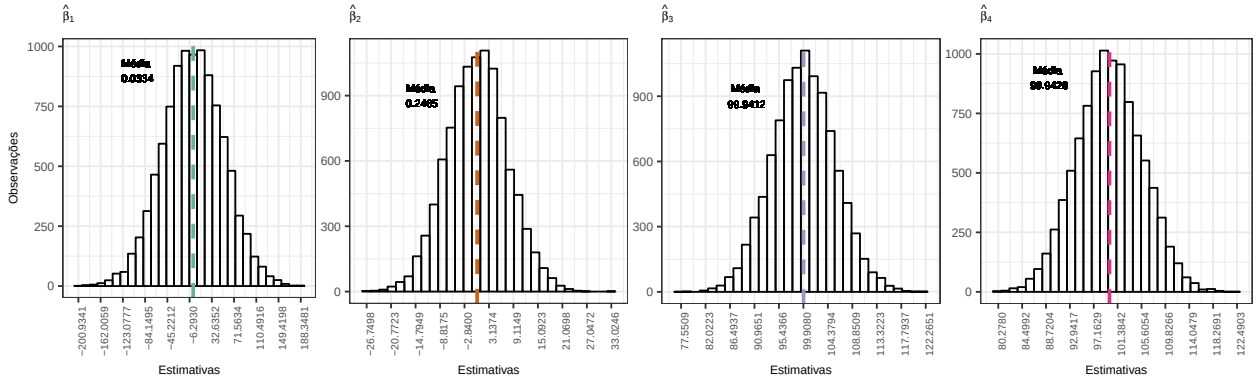
Figura 30 – Comparação das estimativas usando a Priori 4 no experimento 4 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



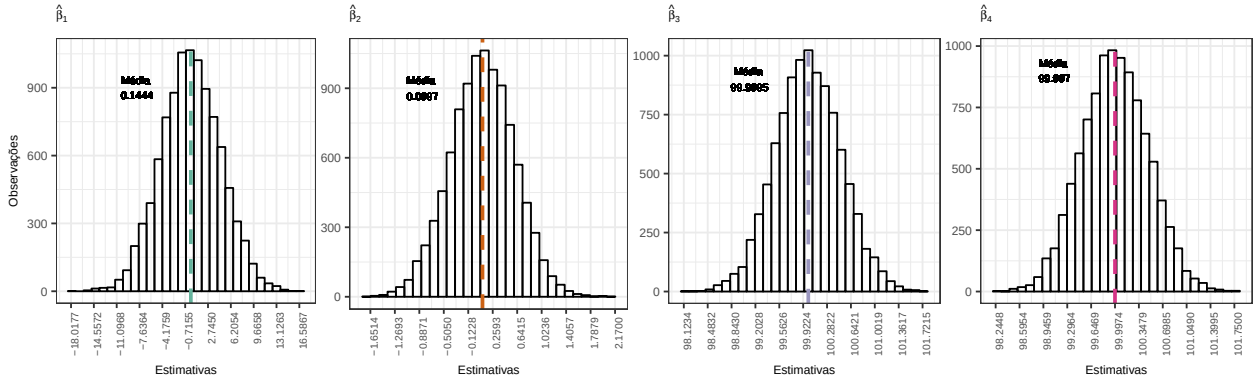
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

EXPERIMENTO 5

Tabela 17 – Resultados do quinto experimento - considerando os pesos dos erros menores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	0.1		0.1		100		100	
				$n = 30$				
MQO	-0.011	448.28511	0.25688	1065.87145	100.43091	634.72833	99.65835	497.1654
Priori 1	0.09924	0.19912	0.10057	0.01732	100.3932	302.2934	99.73673	214.77596
Priori 2	-0.01095	448.02523	0.25678	1064.51322	100.43093	634.5109	99.65839	496.91821
Priori 3	0.03049	0.01061	0.73239	1.98739	99.1148	316.85123	101.20182	229.83743
Priori 4	-0.01097	447.65193	0.25691	1065.68429	100.43088	634.61968	99.65834	497.12489
				$n = 100$				
MQO	-0.01291	107.85217	0.16066	87.50891	99.99357	116.41105	99.95904	122.74759
Priori 1	0.0974	0.26947	0.10097	0.0895	99.98031	68.23338	99.92289	86.43815
Priori 2	-0.0129	107.82936	0.16065	87.49335	99.99357	116.40622	99.95904	122.74337
Priori 3	0.02525	0.01607	0.60516	5.4434	99.13874	72.58047	100.87721	90.38819
Priori 4	-0.0129	107.79675	0.16066	87.50276	99.99357	116.40653	99.95904	122.73728
				$n = 500$				
MQO	0.08424	22.58156	0.08503	23.05454	100.02876	22.33508	99.97887	21.96064
Priori 1	0.09721	0.62763	0.09735	0.37666	100.01782	16.69496	99.96892	16.67246
Priori 2	0.08424	22.58071	0.08503	23.05364	100.02876	22.33495	99.97887	21.96053
Priori 3	0.02739	0.0696	0.36586	7.72269	99.84814	17.93801	100.04927	17.92606
Priori 4	0.08424	22.57954	0.08503	23.05423	100.02876	22.3349	99.97887	21.96048
				$n = 1000$				
MQO	0.11	10.42087	0.15993	9.95198	99.94467	10.25533	99.97156	9.74172
Priori 1	0.10609	0.7208	0.11506	0.51548	99.96511	7.74548	99.9942	7.39486
Priori 2	0.11	10.42067	0.15993	9.95178	99.94467	10.25529	99.97156	9.74169
Priori 3	0.03218	0.10264	0.31842	5.17239	99.87283	8.47436	100.0077	8.28796
Priori 4	0.11	10.42038	0.15993	9.95191	99.94467	10.25527	99.97156	9.74168
				$n = 5000$				
MQO	0.09632	2.14468	0.09857	2.25605	99.99373	2.25267	100.01039	2.19957
Priori 1	0.09768	0.85084	0.09887	0.74416	99.99304	1.95576	100.00968	1.90089
Priori 2	0.09632	2.14467	0.09857	2.25604	99.99373	2.25267	100.01039	2.19957
Priori 3	0.04843	0.27185	0.15311	1.80858	99.98157	2.06425	100.02304	2.01137
Priori 4	0.09632	2.14466	0.09857	2.25605	99.99373	2.25267	100.01039	2.19957

Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

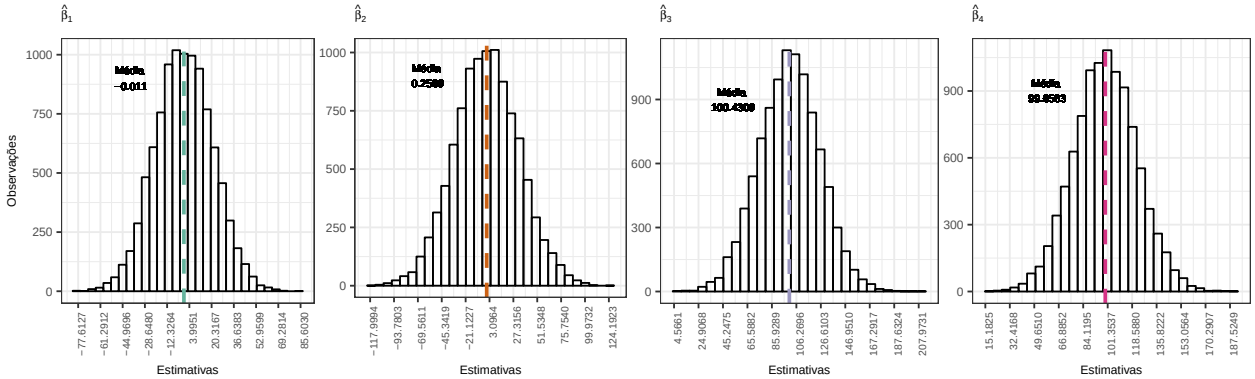
Tabela 18 – Resultados do quinto experimento - considerando os pesos dos erros maiores

	β_1	EQM[β_1]	β_2	EQM[β_2]	β_3	EQM[β_3]	β_4	EQM[β_4]
Valor Real	0.1		0.1		100		100	
				$n = 30$				
MQO	0.2304	40.95481	-0.03353	32.44289	100.06566	29.02134	99.98186	35.04743
Priori 1	0.10069	0.05326	0.09895	0.00362	100.08065	18.44944	99.94913	18.16317
Priori 2	0.23034	40.92575	-0.03348	32.42372	100.06568	29.01823	99.98185	35.04269
Priori 3	0.03016	0.00528	0.63499	0.58008	98.88869	20.53341	101.70282	22.28248
Priori 4	0.23022	40.88542	-0.03345	32.43391	100.06572	29.01331	99.98188	35.04563
				$n = 100$				
MQO	0.08483	9.78207	0.14676	9.94116	99.98283	9.17602	99.98331	11.01704
Priori 1	0.09997	0.0963	0.10137	0.00957	100.00026	6.56471	99.99025	6.9867
Priori 2	0.08483	9.7803	0.14675	9.93937	99.98283	9.17574	99.98331	11.01661
Priori 3	0.02631	0.00677	0.63325	0.85099	99.35222	7.20185	100.5899	7.57785
Priori 4	0.08481	9.77781	0.14677	9.94057	99.98283	9.17594	99.98331	11.01602
				$n = 500$				
MQO	0.09575	1.83253	0.1041	1.81241	100.00346	1.84548	99.98527	1.94254
Priori 1	0.09955	0.18407	0.10042	0.03372	100.0034	1.43381	99.98525	1.40598
Priori 2	0.09575	1.83247	0.1041	1.81234	100.00346	1.84547	99.98527	1.94253
Priori 3	0.02794	0.01076	0.35998	0.71014	99.83457	1.56503	100.0566	1.52686
Priori 4	0.09575	1.83237	0.1041	1.81239	100.00346	1.84546	99.98527	1.94252
				$n = 1000$				
MQO	0.11593	0.9449	0.10772	0.90068	99.99162	0.90362	99.98634	0.87373
Priori 1	0.10437	0.22852	0.10287	0.05236	99.99828	0.67456	99.99376	0.65844
Priori 2	0.11593	0.94489	0.10772	0.90066	99.99162	0.90362	99.98634	0.87373
Priori 3	0.03285	0.01465	0.27508	0.51758	99.90637	0.77484	100.02829	0.7208
Priori 4	0.11593	0.94486	0.10772	0.90067	99.99162	0.90362	99.98634	0.87373
				$n = 5000$				
MQO	0.09647	0.18884	0.10249	0.18682	100.00187	0.18339	100.00348	0.18149
Priori 1	0.09802	0.25788	0.10128	0.06679	100.00173	0.15857	100.00334	0.15693
Priori 2	0.09647	0.18884	0.10249	0.18682	100.00187	0.18339	100.00348	0.18149
Priori 3	0.04946	0.02788	0.15257	0.15439	99.99196	0.16686	100.01695	0.16494
Priori 4	0.09647	0.18884	0.10249	0.18682	100.00187	0.18339	100.00348	0.18149

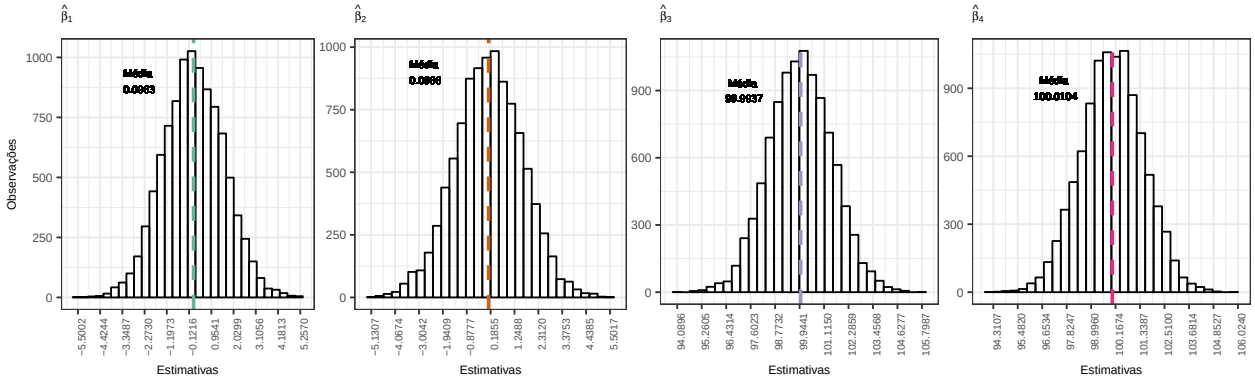
Notas técnicas:
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pelo parâmetro β_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (1) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

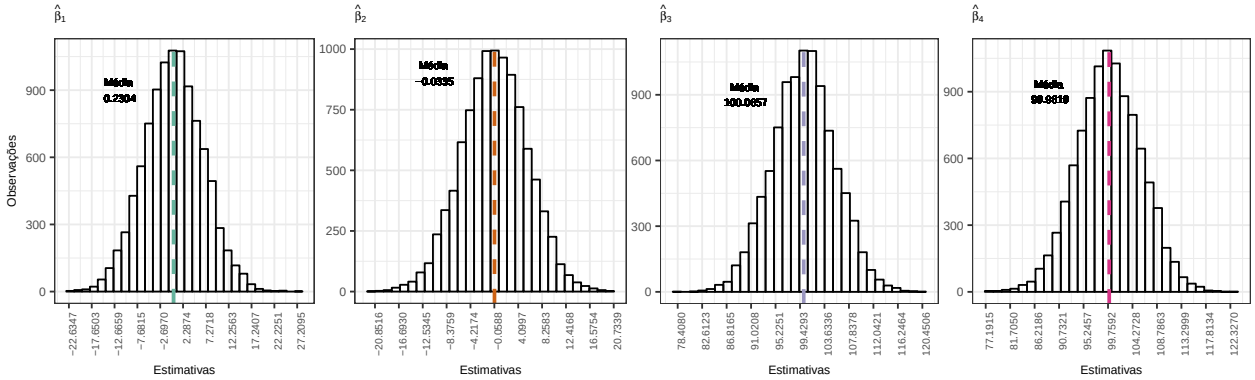
Figura 31 – Comparação das estimativas de MQO no experimento 5 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



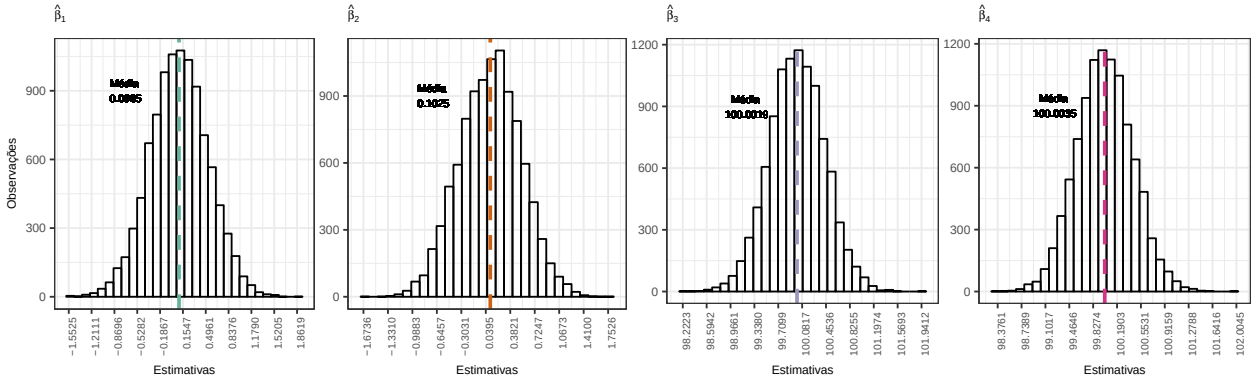
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



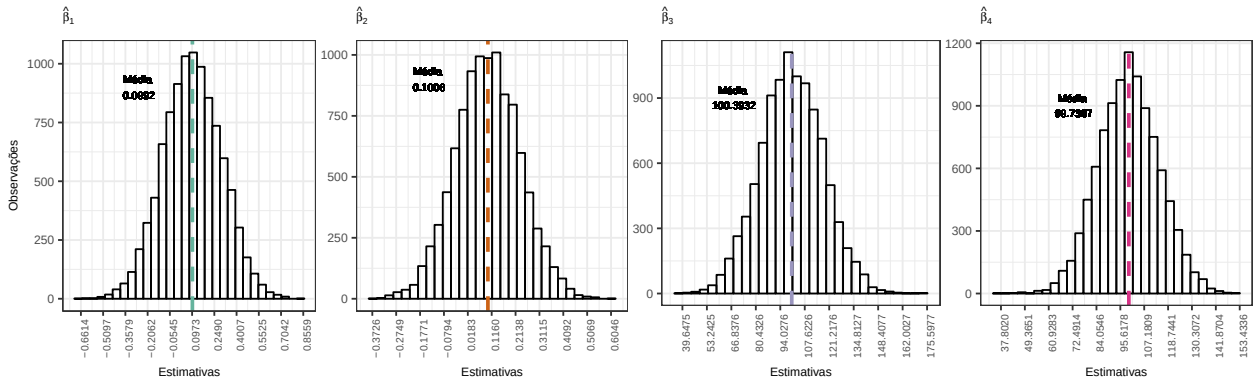
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



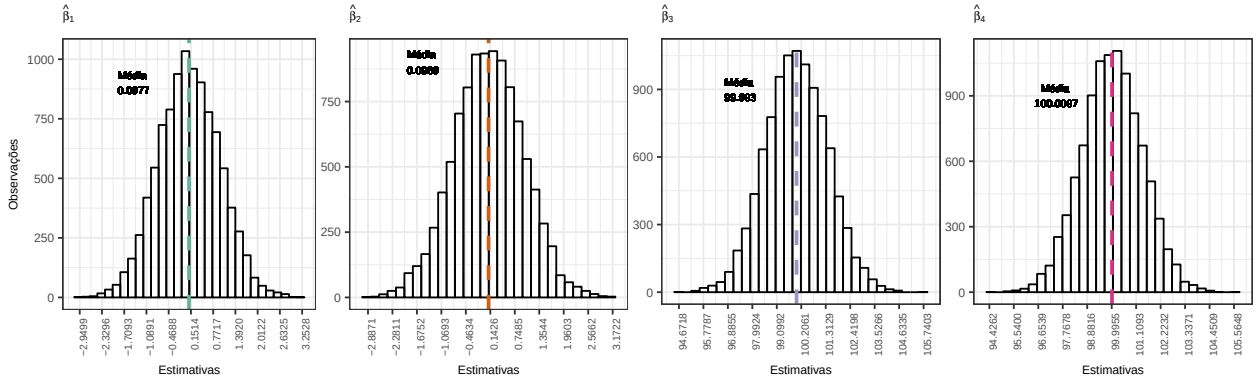
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

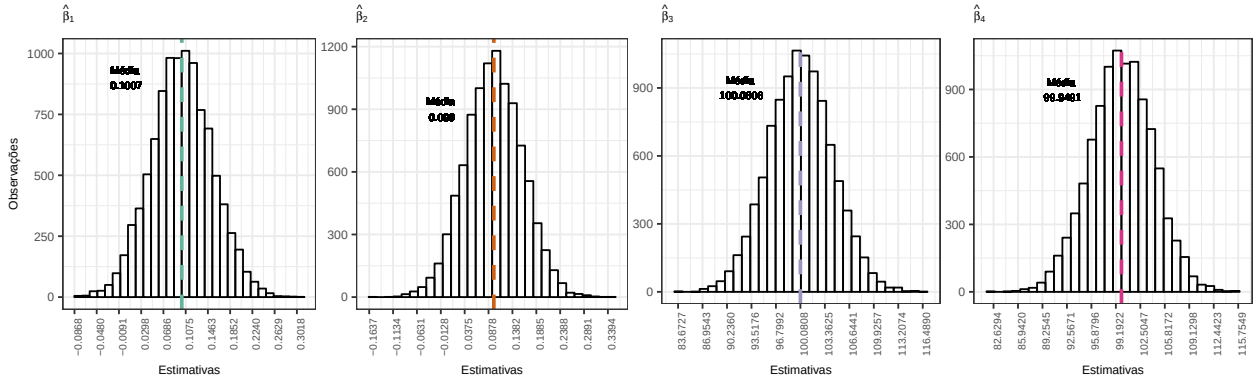
Figura 32 – Comparação das estimativas usando a Priori 1 no experimento 5 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



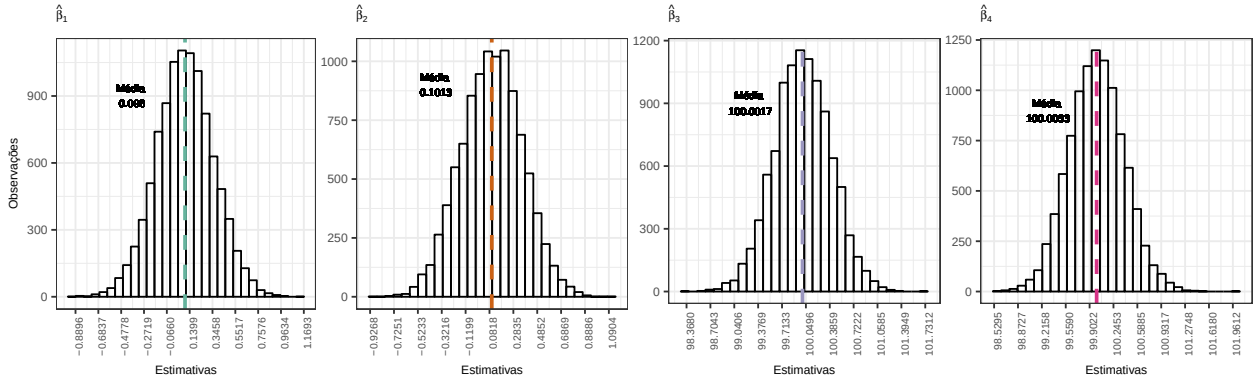
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



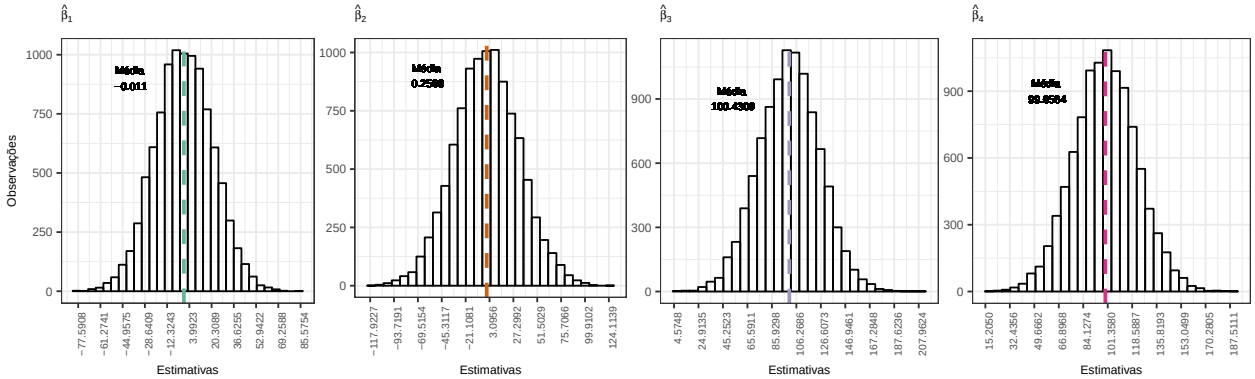
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



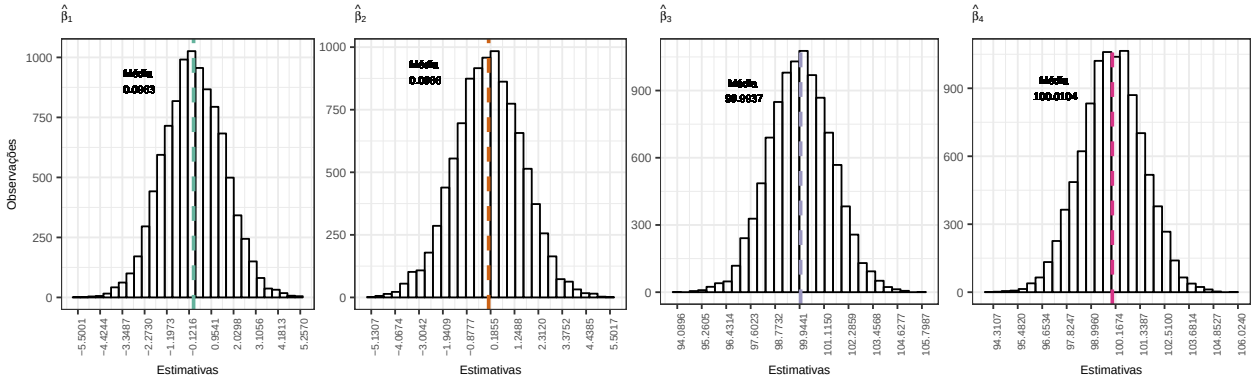
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

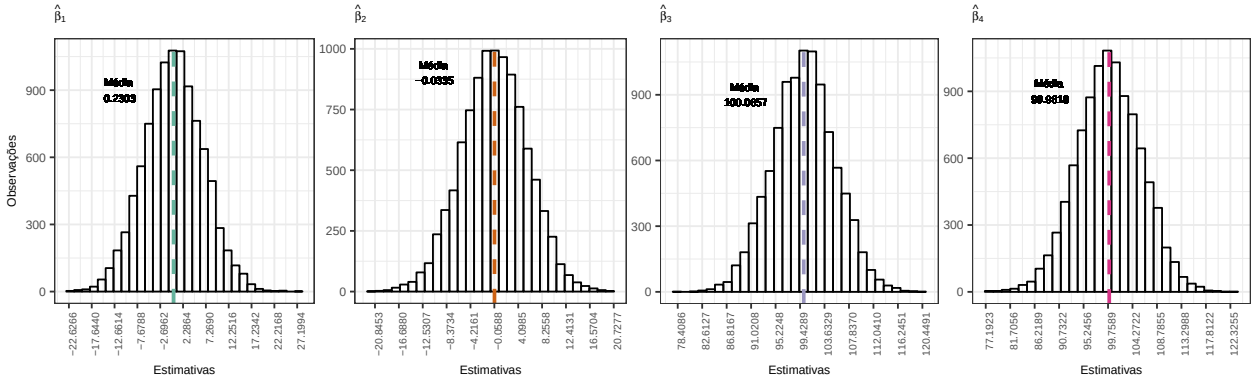
Figura 33 – Comparação das estimativas usando a Priori 2 no experimento 5 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



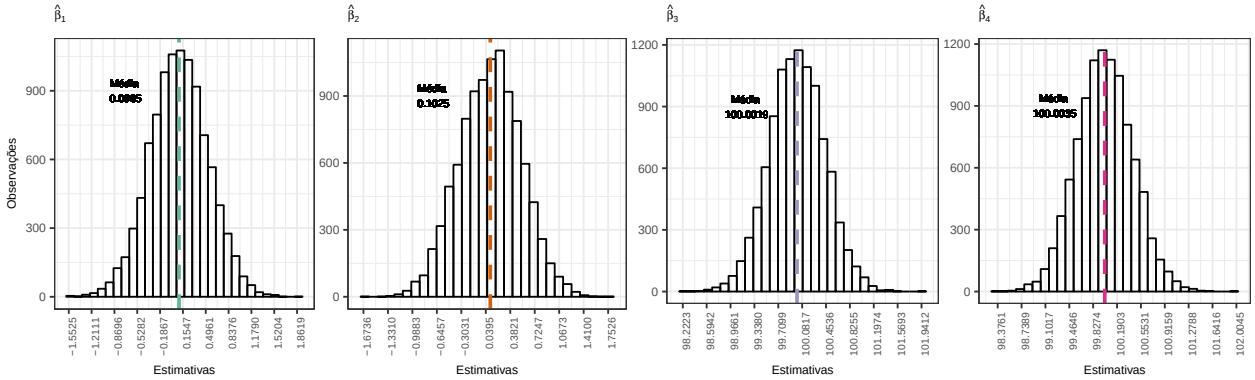
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



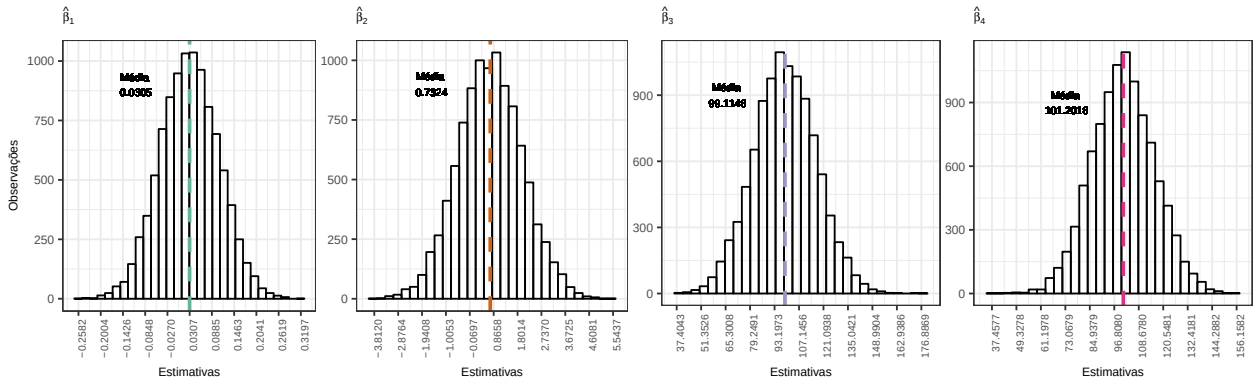
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



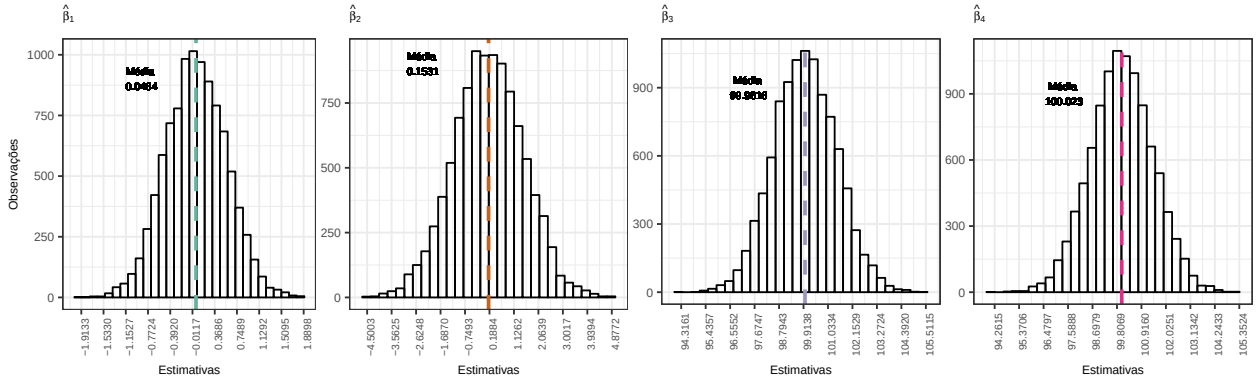
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

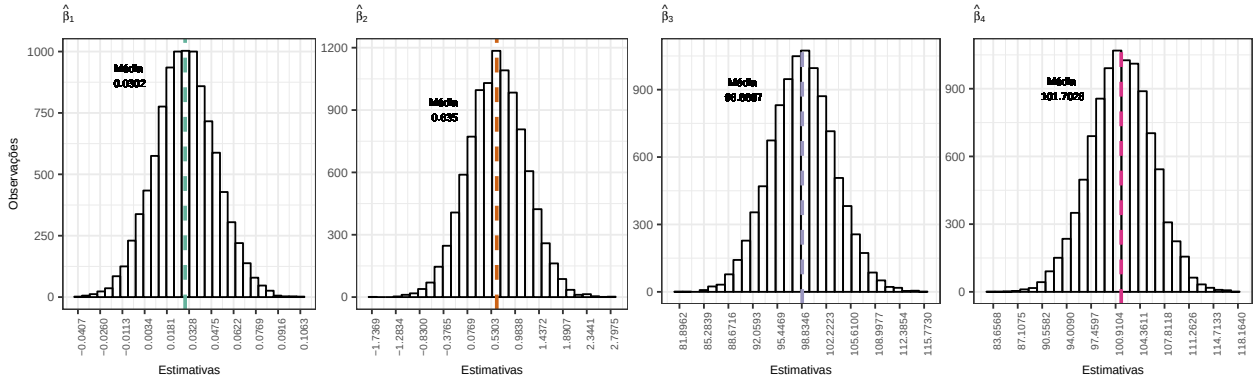
Figura 34 – Comparação das estimativas usando a Priori 3 no experimento 5 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



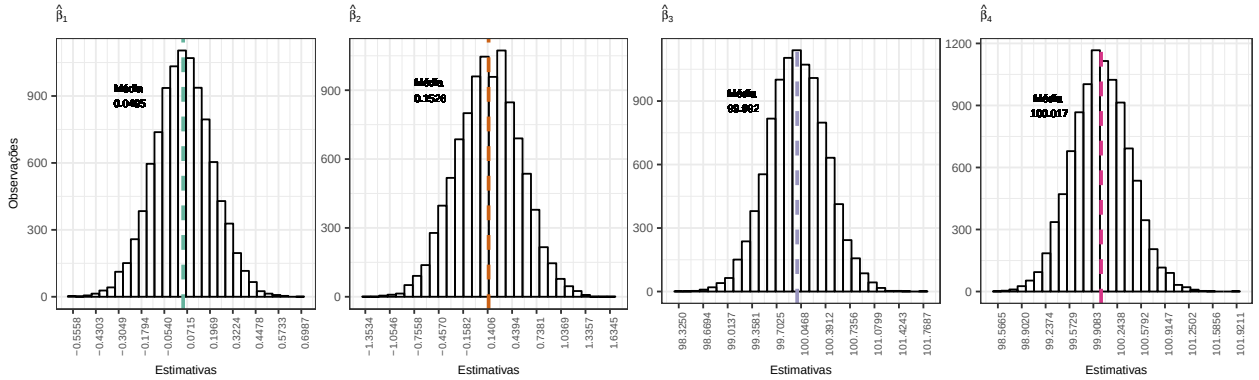
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1/_{SD(X_i)}$



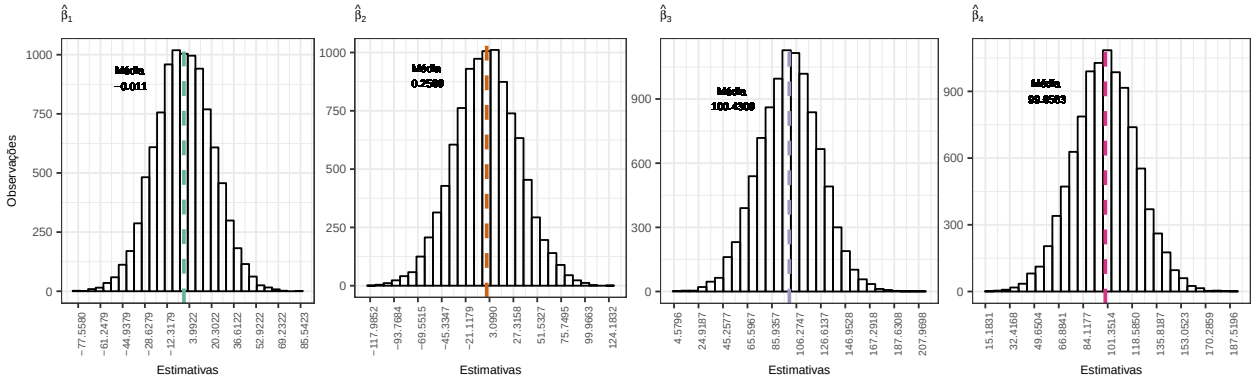
(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



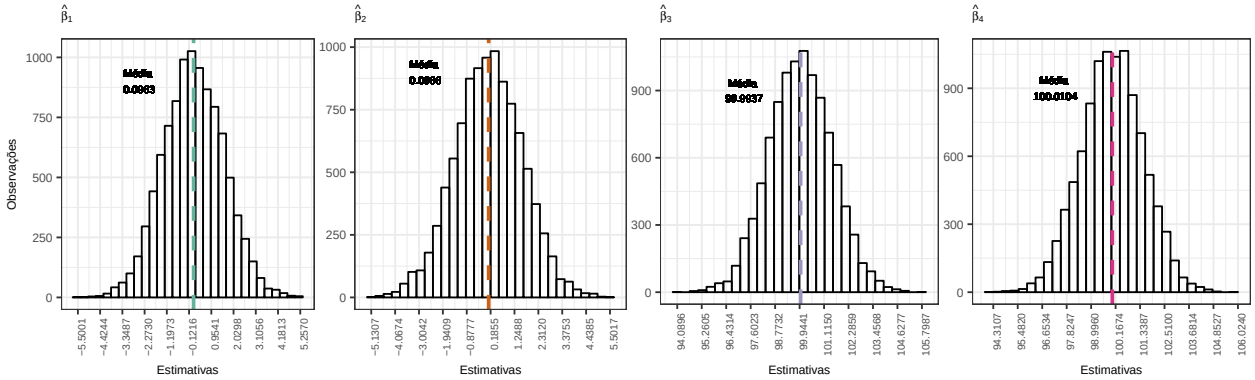
(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

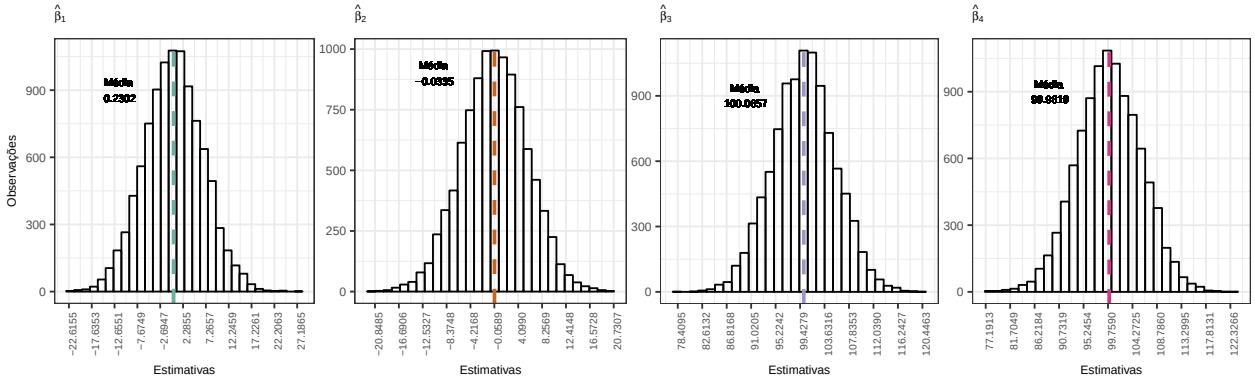
Figura 35 – Comparação das estimativas usando a Priori 4 no experimento 5 para $n = 30$ e $n = 5000$ e as duas especificações de pesos dos erros



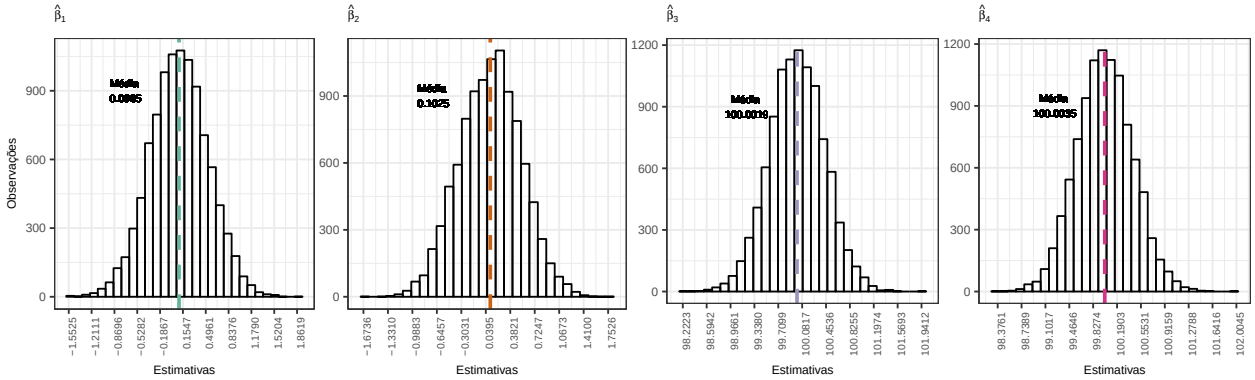
(a) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^{1}/_{SD(X_i)}$



(c) $n = 30$, $peso = \sum \bar{X}_i \cdot \beta_j$



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j$

Fonte – Elaboração própria com base em resultados simulados.

EXPERIMENTO COM ERRO DE MEDIDA

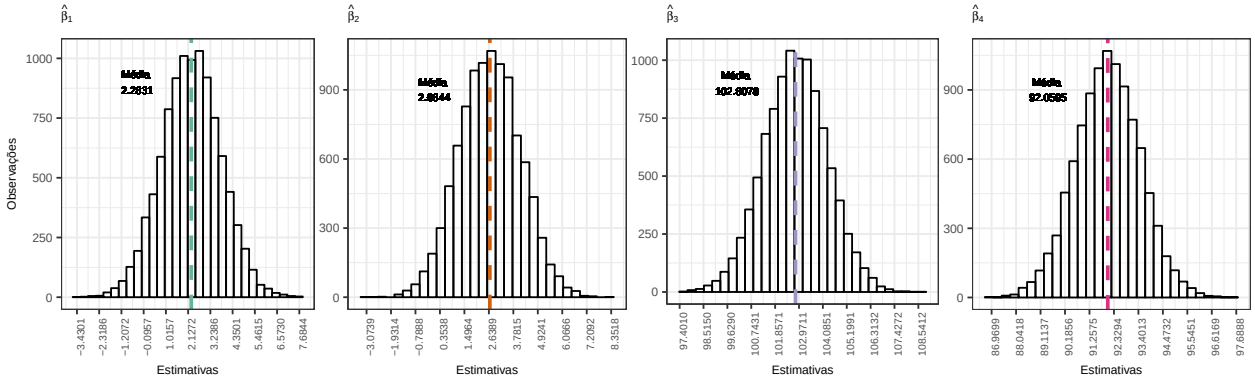
Tabela 19 – Resultados das estimativas na simulação de dados com erro de medida

	β_1	sd[β_1]	β_2	sd[β_2]	β_3	sd[β_3]	β_4	sd[β_4]
Valor Real	0.1		0.1		100		100	
$n = 5000,\ \underline{s}^2 = 0.2 \cdot (1/10000) \cdot \sum \bar{Y}_n,\ u, v \sim \mathcal{N}(0,1)$								
MQO	2.2831	6.9408	2.6844	8.72317	102.80781	10.07503	92.05951	65.03987
Priori 1	1.66761	0.90376	1.94058	4.20042	103.41435	13.55842	92.61066	56.36065
Priori 2	2.2831	6.94078	2.6844	8.72315	102.80781	10.07504	92.05951	65.03984
Priori 3	0.91557	0.9885	2.9802	9.9824	103.28098	12.74353	92.51593	57.82466
Priori 4	2.28309	6.94074	2.6844	8.72318	102.80781	10.07504	92.05951	65.03983
$n = 5000,\ \underline{s}^2 = 500,\ u, v \sim \mathcal{N}(0, 1/10)$								
MQO	0.06296	2.13131	0.15902	2.17133	100.03577	2.19431	99.94641	2.26309
Priori 1	0.089	0.58325	0.12175	0.34758	100.04043	1.79633	99.95133	1.81002
Priori 2	0.06296	2.13129	0.15902	2.17132	100.03577	2.1943	99.94641	2.26308
Priori 3	0.03133	0.09234	0.24004	1.55663	99.99471	1.95193	99.9607	1.97068
Priori 4	0.06296	2.13126	0.15902	2.17133	100.03577	2.1943	99.94641	2.26308
$n = 30,\ \underline{s}^2 = 500,\ u, v \sim \mathcal{N}(0, 1/10)$								
MQO	1.17173	698.67789	-0.39265	439.79634	100.28204	497.11204	99.72112	489.24301
Priori 1	0.10276	0.06873	0.09836	0.00772	100.37874	218.02318	99.94068	196.24139
Priori 2	1.17066	697.32489	-0.39228	439.26442	100.28227	496.89095	99.72134	488.97517
Priori 3	0.0315	0.00537	0.70182	1.06574	98.16963	240.18191	104.19433	240.48684
Priori 4	1.16883	695.13993	-0.39239	439.72862	100.28296	496.7294	99.7221	488.80626
$n = 30,\ \underline{s}^2 = 500,\ u \sim \mathcal{N}(0, 1),\ v \sim \mathcal{N}(0, 10)$								
MQO	44.2988	2148.52174	22.18726	718.31629	102.23586	322.49075	10.97564	7964.85031
Priori 1	0.48572	0.0893	0.37698	0.08412	150.08095	2602.91414	11.29156	7903.78111
Priori 2	44.2837	2147.00516	22.18344	717.91013	102.25126	322.43479	10.97488	7964.98366
Priori 3	0.17687	0.00705	3.40501	11.57525	149.55697	2552.37787	11.41339	7883.11174
Priori 4	44.2507	2143.84386	22.20358	718.96275	102.25989	322.48721	10.97602	7964.78376
$n = 5000,\ \underline{s}^2 = 5000,\ u \sim \mathcal{N}(0, 1),\ v \sim \mathcal{N}(0, 10)$								
MQO	29.01055	837.89959	25.71214	657.86663	127.17111	740.3149	9.67551	8158.74332
Priori 1	15.2113	0.5862	14.66171	212.38493	144.64219	1994.1277	11.43715	7843.59829
Priori 2	29.01048	837.89558	25.71209	657.86428	127.17119	740.31927	9.67552	8158.74185
Priori 3	6.46093	40.55266	30.80613	944.16192	139.14857	1534.13912	11.01654	7918.27889
Priori 4	29.01028	837.88399	25.71223	657.87142	127.17123	740.32139	9.67552	8158.74084
$n = 5000,\ \underline{s}^2 = 5000,\ u, v \sim \mathcal{N}(0, 1)$								
MQO	3.11849	11.14451	3.04393	10.67633	102.58744	8.80076	90.99392	83.06642
Priori 1	0.36562	0.0929	0.37285	0.08394	104.94794	25.9595	93.29455	46.42233
Priori 2	3.11841	11.14386	3.04386	10.67576	102.58751	8.80109	90.99398	83.0653
Priori 3	0.12178	0.00177	2.15415	4.56596	104.11894	18.51842	92.95696	51.1357
Priori 4	3.11819	11.1423	3.04401	10.67675	102.58754	8.80124	90.994	83.0648

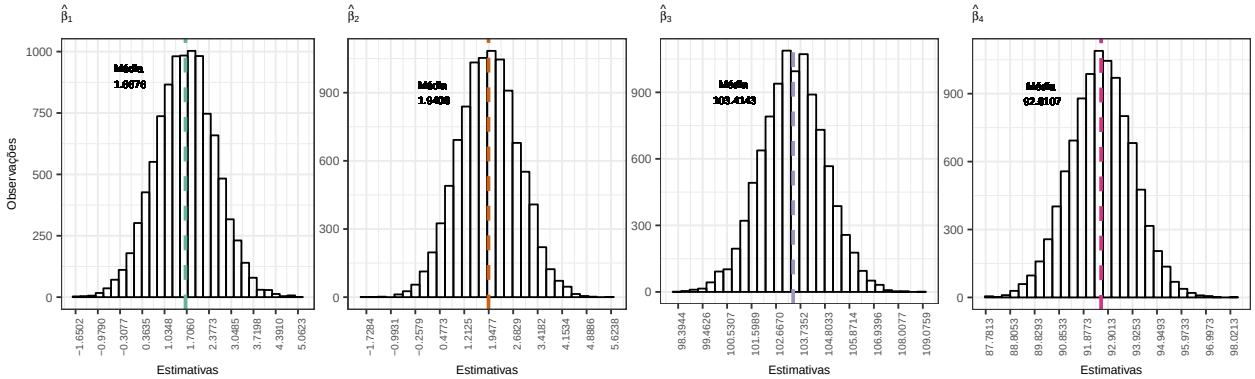
⋆ $x \sim \mathcal{N}(\mu, \sigma)$ significa que x segue uma distribuição normal de média μ e desvio padrão σ;
⋆ A linha **Valor Real** contém os valores de β_j utilizados no processo gerador dos dados;
⋆ Os valores são aproximados até a quinta casa decimal no caso de valores sem unidade significativas, até a quarta decimal no caso de haver uma unidade significativa, até a segunda no caso de dezena significativa, até a primeira no caso de centena significativa e arredondou-se na unidade, omitindo os decimais para valores maiores que o milhar.
⋆ O peso que multiplica os erros do processo gerador dos dados é calculado pela soma das médias dos vetores X_i (do verdadeiro processo) ponderadas pela razão do parâmetro β_i e o desvio padrão de X_i;
⋆ EQM significa erro quadrático médio e é calculado conforme a definição dada no capítulo 3 desta monografia;
⋆ As especificações de cada priori estão na tabela (2) (capítulo 3).

Fonte – Elaboração própria com base em resultados simulados.

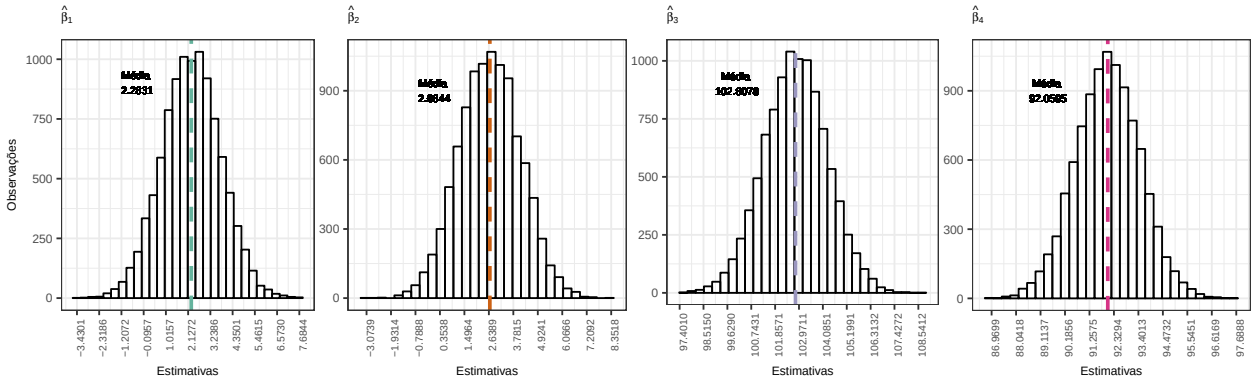
Figura 36 – Comparação das estimativas obtidas na primeira simulação com erro de medida



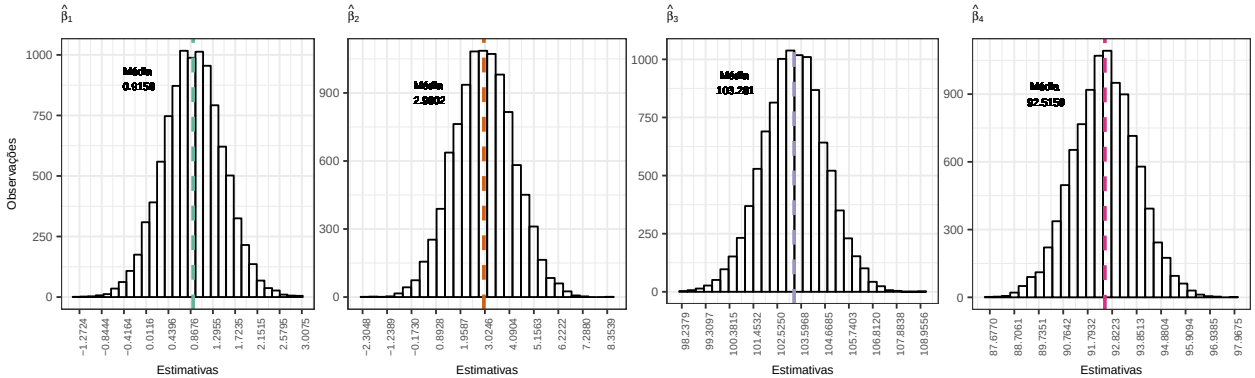
(a) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1\!/_{SD(X_i)}$, $\bar{s}^2 = 0.2 \cdot {}^1\!/_{10^3} \sum \bar{Y}_k$, $u, v \sim \mathcal{N}(0,1)$, estimativas de MQO.



(b) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1\!/_{SD(X_i)}$, $\bar{s}^2 = 0.2 \cdot {}^1\!/_{10^3} \sum \bar{Y}_k$, $u, v \sim \mathcal{N}(0,1)$, estimativas usando a Priori 1.



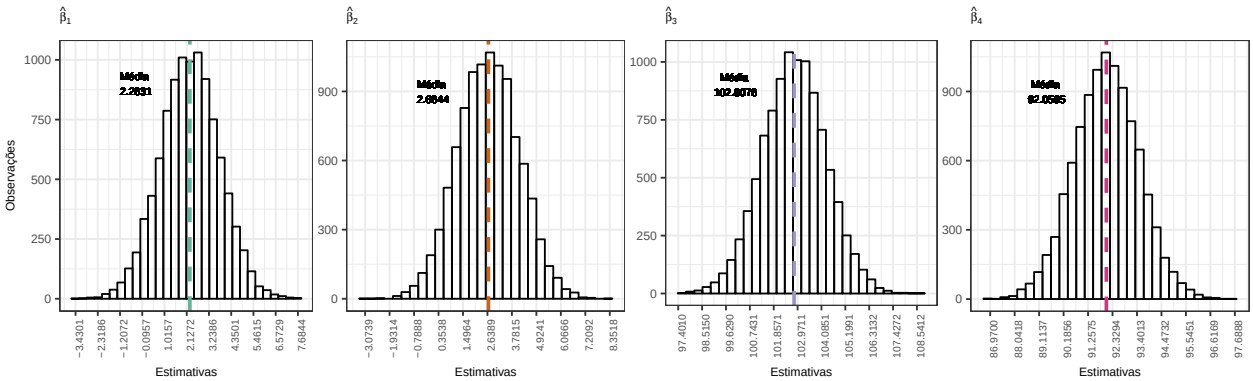
(c) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1\!/_{SD(X_i)}$, $\bar{s}^2 = 0.2 \cdot {}^1\!/_{10^3} \sum \bar{Y}_k$, $u, v \sim \mathcal{N}(0,1)$, estimativas usando a Priori 2.



(d) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1\!/_{SD(X_i)}$, $\bar{s}^2 = 0.2 \cdot {}^1\!/_{10^3} \sum \bar{Y}_k$, $u, v \sim \mathcal{N}(0,1)$, estimativas usando a priori 3.

Fonte – Elaboração própria com base em resultados simulados.

Figura 37 – Comparação das estimativas obtidas na primeira simulação com erro de medida (continuação)



(a) $n = 5000$, $peso = \sum \bar{X}_i \cdot \beta_j \cdot {}^1\!/_{SD(X_i)}$, $\bar{s}^2 = 0.2 \cdot {}^1\!/_{10^3} \sum \bar{Y}_k$, $u, v \sim \mathcal{N}(0,1)$, estimativas usando a Priori 4

Fonte – Elaboração própria com base em resultados simulados.

APÊNDICE C – CÓDIGOS EM R UTILIZADOS NA MONOGRAFIA

Neste apêndice estão todos os códigos desenvolvidos em linguagem R para a realização da monografia. Algumas partes dos códigos foram adaptados com quebras de linhas para melhor visualização no PDF final, então é sugerido que as quebras sejam retiradas para reprodução do código. Além disso, os códigos foram feitos em R Markdown, onde somente os códigos das áreas (*chunks*) em cinza são efetivamente códigos de R. Os pacotes utilizados sempre estão indicados.

Para reprodução de um *chunk*, é necessário habilitar a opção *eval*, escrevendo *eval* = *TRUE* (no lugar de *eval* = *FALSE*), se for copiado o código do R Markdown para um arquivo *.Rmd. Todos os códigos estão disponíveis no seguinte repositório: <https://github.com/aishameriane/TCC>, onde além dos arquivos Rmd também estão disponíveis os arquivos html correspondentes. As instruções estão no arquivo README do próprio repositório.

Figuras do Capítulo 2 - Distribuição Beta

Carregando os pacotes

```
list.of.packages <- c("ggplot2", "reshape2", "latex2exp")
new.packages <- list.of.packages[!(list.of.packages %in%
                                   installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
library(ggplot2, quietly = T)
library(reshape2, quietly = T)
library(latex2exp, quietly = T)
```

Gerando as figuras

O primeiro chunk gera 4 distribuições betas com parâmetros diferentes. Para rodar o código, utilize eval = TRUE.

```
p <- ggplot(data.frame(x = c(0,1)), aes(x=x)) +
     stat_function(fun = dbeta, args = list(.5, .5),
                   aes(colour = "Alpha = .5, Beta = .5"), size = 1) +
     stat_function(fun = dbeta, args = list(1, 1),
                   aes(colour = "Alpha = 1, Beta = 1"), size = 1) +
     stat_function(fun = dbeta, args = list(10, 10),
                   aes(colour = "Alpha = 10, Beta = 10"), size = 1) +
     stat_function(fun = dbeta, args = list(5, 2),
                   aes(colour = "Alpha = 5, Beta = 2"), size = 1) +

     scale_x_continuous(name = "Valores de X",
                        breaks = seq(0, 1, 0.2),
                        limits = c(0,1)) +
     scale_y_continuous(name = "Densidade") +

     ggtitle("") +
        scale_colour_brewer(palette="Dark2") +
        labs(colour = "Parâmetros")

p <- p + theme_bw()
p
```

Salva a figura em PDF (para ativar o chunk, apague o `eval = FALSE`). Lembre-se de editar os três pontos para editar o diretório para salvar a imagem no seu computador.

```
# Altura e largura são em inches, default = 7

pdf(file="...\\Fig-1.001.pdf", width = 6, height = 3)
p
dev.off()
```

Triplot

Para rodar o código, utilize eval = TRUE.

```
# Cria um vetor de thetas
theta  <- seq(0, 1, length = 100)

# Priori
alphapriori <- 10
betapriori <- 10
priori <- dbeta(theta, alphapriori, betapriori)

# Número de caras observadas
somay <- 75
# Tamanho da amostra
n <- 100
# Verossimilhança - o termo (n+1) é multiplicado para dar escala
vero <- (n+1)*(factorial(n)/(factorial(somay)*
                              factorial(n-somay)))*theta^somay*(1-theta)^(n-somay)

# Posteriori
alphapost <- alphapriori + somay
betapost <- betapriori + n - somay
post <- dbeta(theta, alphapost, betapost)

# Juntando os dados
df <- as.data.frame(cbind(theta, priori, vero, post))
df <- melt(df, theta)
levels(df$variable)[levels(df$variable) == "priori"] <- "Priori"
levels(df$variable)[levels(df$variable) == "vero"]   <- "Verossimilhança"
levels(df$variable)[levels(df$variable) == "post"]   <- "Posteriori"
head(df)

p <- ggplot(df, aes(theta,value, color = factor(variable))) +
  geom_line(size = 1)+
  scale_x_continuous(name = TeX("$\\theta$")) +
     scale_y_continuous(name = "") +
  ggtitle("") +
    scale_colour_brewer(palette="Dark2") +
    labs(colour = "Função") +
  theme_bw()

p
```

Salva a figura em PDF. Para ativar o chunk, apague o `eval = FALSE`. Lembre-se de editar os três pontos para editar o diretório para salvar a imagem no seu computador.

```
# Altura e largura são em inches, default = 7
pdf(file="...\\Fig-1.003.pdf", width = 6, height = 3)
p
dev.off()
```

“‘

Capítulo 3 - Simulações

Atenção: é necessário descomentar os trechos de códigos correspondentes às diferentes especificações de pesos e prioris, conforme desejado, além de mudar a opção `eval=FALSE` para `eval=TRUE`.

O código é único para todas as especificações, incluindo modelos com 3, 4 e 5 valores de β_i, de forma que é importante comentar/descomentar os trechos conforme a priori utilizada.

Instala e carrega os pacotes usados

```
list.of.packages <- c("ggplot2", "metRology", "reshape2", "gridExtra",
                      "knitr", "stargazer", "latex2exp")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
library(ggplot2, quietly = T)
library(metRology, quietly = T)
library(reshape2, quietly = T)
library(gridExtra, quietly = T)
library(knitr, quietly = T)
library(stargazer, quietly = T)
library(latex2exp, quietly = T)
options(scipen=999)
```

Gerando dados

```
gera_dados <- function(n, nsim, x1min, x1max, x2min = NULL, x2max = NULL, x3min = NULL, x3max = NULL,
                       x4min = NULL, x4max = NULL, x5min = NULL, x5max = NULL, beta_real, intercepto = 

  N         <- n
  minX1     <- x1min
  maxX1     <- x1max
  Beta_real <- beta_real
  nsim      <- nsim

  X1 <- round(runif(N, min = minX1, max = maxX1),2)
  X  <- X1

  if ((is.null(x2min) & is.null(x2max)) == FALSE ) {
    minX2 <- x2min
    maxX2 <- x2max
    X2 <- round(runif(N, min = minX2, max = maxX2),2)
    X  <- cbind(X, X2)
  }

  if ((is.null(x3min) & is.null(x3max)) == FALSE ) {
    minX3 <- x3min
    maxX3 <- x3max
    X3 <- round(runif(N, min = minX3, max = maxX3),2)
    X  <- cbind(X, X3)
  }
```

```
  if ((is.null(x4min) & is.null(x4max)) == FALSE ) {
    minX4 <- x4min
    maxX4 <- x4max
    X4 <- round(runif(N, min = minX4, max = maxX4),2)
    X  <- cbind(X, X4)
  }

  if (intercepto == TRUE) {
    X <- cbind(rep(1, N), X)
  }

  k <- ncol(X)

  if (length(beta_real) != ncol(X)) {
    stop("A dimensão de beta deve ser a mesma do vetor X acrescido do intercepto")
  }

  # Calcula os pesos
  # O peso é a média do X vezes o beta dividido pelo desvio padrão
  # No caso de haver intercepto, o peso é apenas o Beta_real[1]

  # pesos pequenos
   # if (intercepto == TRUE) {
   #   peso <- colMeans(X) * Beta_real * c(1,1/apply(X[,2:ncol(X)], 2, sd))
   # } else {
   #   peso <- colMeans(X) * Beta_real * 1/apply(X, 2, sd)
   # }

  # pesos grandes
   peso <- colMeans(X) * Beta_real * 0.1

  erro   <- list()
  Y_real <- list()

  for (i in 1:nsim) {
    erro[[i]] <- sum(peso) * rnorm(N, mean = 0, sd = 1)
    Y_real[[i]] <- as.matrix(X) %*% as.matrix(Beta_real, nrow = length(Beta_real))
                  + erro[[i]]
  }

  return(list(Y_real = Y_real, Beta_real = Beta_real, X = X, peso = peso, erro = erro))
}
```

Determinação da priori

Antes de estabelecer a priori, é gerarado um único conjunto de dados que vai servir para as simulações deste bloco.

```
##########################################################################################
# 1o experimento
# N (tamanho da amostra) = (30, 100, 500, 1000, 5000)
##########################################################################################
```

```
# nsim = 10000
# dados <- gera_dados(n = 30, nsim = nsim, x1min = 0, x1max = 10, x2min = 500,
#       x2max = 1000, x3min =3500, x3max = 5000, x4min = 50000,
#       x4max = 600000, beta_real = c(2, 100, 0.6, 5, 0.01), intercepto = TRUE)

##################################################################################
# 2o experimento
#N (tamanho da amostra) = (30, 100, 500, 1000, 5000)
##################################################################################

 nsim = 10000
 dados <- gera_dados(n = 5000, nsim = nsim, x1min = 0, x1max = 10, x2min = 0,
                     x2max = 10, x3min =0, x3max = 10, beta_real = c(1, 10, 50, 100),
                     intercepto = TRUE)

##################################################################################
# 3o experimento
#N (tamanho da amostra) = (30, 100, 500, 1000, 5000)
##################################################################################

 # nsim = 10000
 # dados <- gera_dados(n = 5000, nsim = nsim, x1min = 0, x1max = 10, x2min = 0,
 #        x2max = 10, x3min =0, x3max = 10,
 #        beta_real = c(10, 50, 100), intercepto = FALSE)

##################################################################################
# 4o experimento
#N (tamanho da amostra) = (30, 100, 500, 1000, 5000)
##################################################################################

 #nsim = 10000
 #dados <- gera_dados(n = 5000, nsim = nsim, x1min = 0, x1max = 10, x2min = 0,
 #          x2max = 10, x3min =0, x3max = 10, beta_real = c(0.1,  0.1, 100, 100),
 #          intercepto = TRUE)

##################################################################################
# 5 experimento
#N (tamanho da amostra) = (30, 100, 500, 1000, 5000)
##################################################################################

 # nsim = 10000
 # dados <- gera_dados(n = 5000, nsim = nsim, x1min = 0, x1max = 10, x2min = 0,
 #            x2max = 10, x3min =0, x3max = 10, x4min =0, x4max = 10,
 #            beta_real = c(0.1, 0.1, 100, 100), intercepto = FALSE)

##################################################################################
 ################################################################################
# 6 experimento - erro de medida
# Para o erro de medida, geramos os dados normalmente, porém precisamos de uma
# nova matriz de X para fazer as estimativas.
##################################################################################
##################################################################################
```

```
#  nsim = 10000
# #
# ## Passo 1 - obter os valores de Y corretos
#  dados <- gera_dados(n = 5000, nsim = nsim, x1min = 0, x1max = 10, x2min = 0,
#             x2max = 10, x3min =0, x3max = 10, x4min =0, x4max = 10,
#             beta_real = c(0.1, 0.1, 100, 100), intercepto = FALSE)
#
# ## Passo 2 - Criar dois vetores em X que tenham erro de medida
# ## Primeiro vou olhar como está a estrutura da matriz X
# # head(dados$X)
# #
# # ## Vou criar dois termos de erros independentes, com mesmo tamanho de X
# # ## Primeiro vou fazer uns erros pequeninhos
#  erro1 <- rnorm(nrow(dados$X), mean = 0, sd = 1)
#  erro2 <- rnorm(nrow(dados$X), mean = 0, sd = 1)
# # summary(erro1)
# # summary(erro2)
#  Xem <- cbind(dados$X[,1], dados$X[,2]+erro1, dados$X[,3], dados$X[,4] + erro2)
# # summary(dados$X)
# # summary(Xem)
# #
# # ## Agora crio uma nova matriz de dados onde o X vai ser o xerrado
#  dados_em <- dados
#  dados_em$X <- Xem
```

Priori com valores iguais aos verdadeiros e baixo desvio padrão

```
# Supondo que os betas tem distribuição Gama com médias próximas aos valores reais
#e desvio padrão igual a 10% das médias
# 1o EXPERIMENTO
# b0_in   <-  c(2, 100, 0.6, 5, 0.01)

# 2o experimento
 b0_in   <-  c(1, 10, 50, 100)

# 3o experimento
# b0_in   <-  c(10, 50, 100)

# # 4o, 5o experimentos e erro de medida
# b0_in   <-  c(0.1, 0.1, 100, 100)

s02_in   <- 0.2*mean(sapply(dados$Y_real, mean))
# Desvio padrão dos erros a priori -
# decidi usar 20% da média dos valores de y - antes era 5000
# s02_in   <- 5000 # para o erro de medida
nu0_in   <- 5          # Graus de liberdade a priori
h0_in    <- 1/s02_in   # Priori para h
stdh0_in <- 2/(s02_in*nu0_in)

# Para encontrar os elementos da matriz V, utilizamos a fórmula do texto
cons_in    <- (nu0_in * s02_in)/(nu0_in-2)
varbeta_in <- b0_in * 0.1
```

```
V0_in      <- diag(length(b0_in))*(varbeta_in/cons_in) # Matriz de escala da priori
```

Priori próxima dos valores verdadeiros e alto desvio padrão

```
# Supondo que os betas tem distribuição Gama com médias próximas
# aos valores reais e desvio padrão igual a 200x as médias
# 1o experimento
# b0_ind   <-  c(2, 100, 0.6, 5, 0.01)

# 2o experimento
 b0_ind   <-  c(1, 10, 50, 100)

# 3o experimento
# b0_ind   <-  c(10, 50, 100)

# # 4o, 5o experimentos e erro de medida
# b0_ind   <-  c(0.1, 0.1, 100, 100)

s02_ind   <- 0.2*mean(sapply(dados$Y_real, mean))
# Desvio padrão dos erros a priori
#s02_ind   <- 5000 #erro de medida
nu0_ind   <- 5       # Graus de liberdade a priori
h0_ind    <- 1/s02_ind   # Priori para h
stdh0_ind <- 2/(s02_ind*nu0_ind)

# Para encontrar os elementos da matriz V, utilizamos
#    a fórmula do texto do Koop
cons_ind    <- (nu0_ind * s02_ind)/(nu0_ind-2)
varbeta_ind <- b0_ind * 40000
V0_ind      <- diag(length(b0_ind))*(varbeta_ind/cons_ind)
# Matriz de escala da priori
```

Priori própria com valores bem distantes dos reais e baixo desvio padrão

```
# Supondo que os betas tem distribuição Gama com médias diferentes aos valores
#reais e desvio padrão igual a 10% das médias
# 1o experimento
# b0_ni   <-  c(5.1, 2.7, 15, 98.5, 37.7)

# 2o experimento -
#gerei dados de uniformes (0,1), (0,10), (0,100), (0,1000)
 b0_ni   <-  c(0.27, 6.22, 73.10, 720.29) # Priori Não-Informativa

# 3o experimento -
#gerei dados de uniformes (0,95), (0,25), (0,72)
# b0_ni   <-  c(48.17, 0.57, 61.09) # Priori Não-Informativa

# # 4o, 5o experimentos e erro de medida -
#gerei dados de uniformes (0,1), (0,1), (0,1000), (0,1000)
# b0_ni   <-  c(0.038, 0.9946, 115.02, 169.20)
```

```r
s02_ni   <- 0.2*mean(sapply(dados$Y_real, mean))
# Desvio padrão dos erros a priori
#s02_ni   <- 5000 #erro de medida
nu0_ni   <- 5       # Graus de liberdade a priori
h0_ni    <- 1/s02_ni   # Priori para h
stdh0_ni <- 2/(s02_ni*nu0_ni)

# Para encontrar os elementos da matriz V,
# utilizamos a fórmula do texto do Koop
cons_ni    <- (nu0_ni * s02_ni)/(nu0_ni-2)
varbeta_ni <- b0_ni * 0.1
V0_ni      <- diag(length(b0_ni))*(varbeta_ni/cons_ni)
# Matriz de escala da priori
```

Priori própria com valores bem distantes dos reais e alto desvio padrão

```r
# Supondo que os betas tem distribuição Gama com médias
#bem diferentes aos valores reais e desvio padrão igual
#a 200 vezes as médias
# 1o experimento
# b0_nid   <-  c(5.1, 2.7, 15, 98.5, 37.7)

# 2o experimento -
# gerei dados de uniformes (0,1), (0,10), (0,100), (0,1000)
 b0_nid   <-  c(0.27, 6.22, 73.10, 720.29) # Priori Não-Informativa

# 3o experimento -
#gerei dados de uniformes (0,95), (0,25), (0,72)
# b0_nid   <-  c(48.17, 0.57, 61.09) # Priori Não-Informativa

# 4o, 5o experimentos e erro de medida -
#gerei dados de uniformes (0,1), (0,1), (0,1000), (0,1000)
# b0_nid   <-  c(0.038, 0.9946, 115.02, 169.20)
# Priori não Informativa

s02_nid   <- 0.2*mean(sapply(dados$Y_real, mean))
# Desvio padrão dos erros a priori
#s02_nid   <- 5000 #erro de medida
nu0_nid   <- 5       # Graus de liberdade a priori
h0_nid    <- 1/s02_nid   # Priori para h
stdh0_nid <- 2/(s02_nid*nu0_nid)

# Para encontrar os elementos da matriz V,
#utilizamos a fórmula do texto do Koop
cons_nid    <- (nu0_nid * s02_nid)/(nu0_nid-2)
varbeta_nid <- b0_nid * 40000
V0_nid      <- diag(length(b0_nid))*(varbeta_nid/cons_nid)
# Matriz de escala da priori
```

Posteriori

```
conjugada_natural <- function(Y, X, nu0, s20, B0, V0, varargin = NULL) {

  # Cálculo das quantidades de MQO
  N      <- nrow(X)     # Tamanho da amostra a partir da matriz X
  k      <- ncol(X)     # Número de regressores
  nu     <- N-k         # Graus de liberdade
  XX     <- t(matrix(X, ncol = k)) %*% matrix(X, ncol=k)
  iXX    <- solve(XX)
  BMQO   <- iXX %*% (t(X) %*% Y)
  # Beta de MQO
  s2     <- (t(Y-X %*% BMQO) %*% (Y-X %*% BMQO))/nu
  # Estimativa não viesada da var dos erros
  CovMQO <- as.numeric(s2) * iXX
  # Matriz de covariância do beta de MQO
  stdMQO <- sqrt(diag(CovMQO))
  # Desvio padrão das estimativas de MQO

  # Parâmetros da posteriori

  nu1 <- nu0 + N # Graus de liberdade
  if (V0 == 0){
    iV0 <- diag(k)*0
  } else {
    iV0 <- solve(V0) # Precisão a posteriori
  }

  iV1   <- iV0 + XX    # Inverso da escala a posteriori
  V1    <- solve(iV1)  # Escala a posteriori
  B1    <- V1 %*% (iV0 %*% B0 + XX %*% BMQO)
  nus21 <- nu0 * s20 + nu * s2

  if (det(iV0) > 0) {
    nus21 <- nus21 + t(BMQO - B0) %*% solve(V0 + iXX) %*% (BMQO - B0)
    # SQR a posteriori
  } else {
    nus21 <- nus21
  }

  s21 <- nus21/nu1

  # Parâmetros da distribuição posterior marginal beta

  CovB <- V1 * as.numeric(nus21/(nu1-2))
  # Variância da distribuição t
  stdB <- sqrt(diag(CovB))
  # Desvio padrão da posteriori marginal de beta

  # Calcula os intervalos de maior densidade a posteriori de 95% e 99%

  lim95   <- qt(0.975, nu1) # Valor da t para o quantil 97.5
  lim99   <- qt(0.995, nu1) # Valor da t para o quantil 99.5
```

```
  HPDI95  <- round(cbind(B1 - lim95*stdB, B1 + lim95*stdB),4)
  HPDI99  <- round(cbind(B1 - lim99*stdB, B1 + lim99*stdB),4)

  probpos=vector()

  for (i in 1:k){
    tnorm = -B1[i,1]/sqrt(s21*V1[i,i])
    probpos[i] = 1 - pt(tnorm,nu1)
  }

  # Parâmetros da distribuição posterior marginal de h

  h1    <- 1/s21
  # Média da distribuição gama que determina h
  h1var <- 2/(nus21)
  # Variância da distribuição a posteriori marginal de h

  # Logaritmo da função de verossimilhança marginal

  lmlik <- vector()
  if (det(iV0 > 0)) {
    c <- lgamma(0.5*nu1) +
      0.5*nu0*log(nu0*s20)-lgamma(0.5*nu0)
             -0.5*N*log(pi) # constante de integração
    lmlik <- c+0.5*log(det(V1)/det(V0))
             -0.5*nu1*log(nus21)  # log da veross. marginal
  } else {
    lmlik <- 0
  }

  # Faz previsão de X* é passado como argumento para a função
  if (is.null(varargin) == FALSE) {
    Xestrela   <- varargin
    Yestrela   <- Xestrela %*% B1
    # Previsão Y*
    VYestrela  <-  (diag(nrow(Xestrela))+Xestrela %*% V1
                   %*% t(Xestrela))*as.numeric(s21)
    # Matriz de escala da previsão
    Varestrela <- VYestrela * nu1/(nu1 - 2)
    # Variância da previsão
  } else {
    Yestrela     <- NA
    VYestrela   <- NA
    Varestrela  <- NA
  }

  return(list(B1=B1, V1=V1, nu1=nu1, s21=s21, CovB=CovB,
             h1=h1, h1var=h1var, lmlik = lmlik, HPDI95 = HPDI95,
             HPDI99 = HPDI99, previsao = Yestrela,
             varprev = Varestrela, escalaprev = VYestrela,
             BMQO = BMQO, stdMQO = stdMQO, probpos = probpos))
}
```

A função abaixo faz retiradas de cada beta a posteriori.

Essa função seria para amostrar da posteriori, mas por enquanto não vai ser usada. Como esse é um procedimento relativamente tradicional, optei por deixar no código.

```
drawbetas <- function(n, betabarra, sVbarra, nubarra) {

  beta       <- vector()
  n          <- n
  betabarra  <- betabarra
  sVbarra    <- sVbarra
  nubarra    <- nubarra

  for (j in 1:length(betabarra)) {
    temp      <- vector()
    temp <- rt.scaled(n = n, df = nubarra,
                      mean = betabarra[j], sd = sVbarra[j])
    beta <- cbind(beta, temp)
  }
  return(beta)
}
```

Posteriori para a priori com valores iguais em média e baixo desvio padrão

```
beta_in   <- matrix(0, ncol = length(b0_in), nrow = nsim)
beta_MQO_in  <- matrix(0, ncol = length(b0_in), nrow = nsim)
ic95_in   <- matrix(0, ncol = 2*length(b0_in), nrow = nsim)
ic99_in   <- matrix(0, ncol = 2*length(b0_in), nrow = nsim)
stdMQO_in <- matrix(0, ncol = length(b0_in), nrow = nsim)

for (i in 1:nsim) {
  post_in <- conjugada_natural(Y = dados$Y_real[[i]],
                               X = dados$X, nu0 = nu0_in,
                               s20 = s02_in, B0 = b0_in,
                               V0 = V0_in, varargin = NULL)
  #post_in <- conjugada_natural(Y = dados_em$Y_real[[i]],
  #X = dados_em$X, nu0 = nu0_in, s20 = s02_in, B0 = b0_in,
  #V0 = V0_in, varargin = NULL) # Erro de medida
  beta_in[i,]      <- post_in$B1
  beta_MQO_in[i,]  <- post_in$BMQO
  ic95_in[i,]      <- as.vector(t(post_in$HPDI95))
  ic99_in[i,]      <- as.vector(t(post_in$HPDI99))
  stdMQO_in[i,]    <- post_in$stdMQO
}
```

Posteriori para a priori com valores iguais em média e alto desvio padrão

```
beta_ind      <- matrix(0, ncol = length(b0_ind), nrow = nsim)
beta_MQO_ind  <- matrix(0, ncol = length(b0_ind), nrow = nsim)
ic95_ind      <- matrix(0, ncol = 2*length(b0_ind), nrow = nsim)
ic99_ind      <- matrix(0, ncol = 2*length(b0_ind), nrow = nsim)
stdMQO_ind    <- matrix(0, ncol = length(b0_ind), nrow = nsim)
```

```
for (i in 1:nsim) {
  post_ind <- conjugada_natural(Y = dados$Y_real[[i]],
                                X = dados$X, nu0 = nu0_ind,
                                s20 = s02_ind, B0 = b0_ind,
                                V0 = V0_ind, varargin = NULL)
  #post_ind <- conjugada_natural(Y = dados_em$Y_real[[i]],
  #X = dados_em$X, nu0 = nu0_ind, s20 = s02_ind, B0 = b0_ind,
  #V0 = V0_ind, varargin = NULL) # Erro de medida
  beta_ind[i,]      <- post_ind$B1
  beta_MQO_ind[i,]  <- post_ind$BMQO
  ic95_ind[i,]      <- as.vector(t(post_ind$HPDI95))
  ic99_ind[i,]      <- as.vector(t(post_ind$HPDI99))
  stdMQO_ind[i,]    <- post_ind$stdMQO
}
```

Posteriori para a priori com valores distantes e pouco desvio padrão

```
beta_ni      <- matrix(0, ncol = length(b0_ni), nrow = nsim)
beta_MQO_ni  <- matrix(0, ncol = length(b0_ni), nrow = nsim)
ic95_ni      <- matrix(0, ncol = 2*length(b0_ni), nrow = nsim)
ic99_ni      <- matrix(0, ncol = 2*length(b0_ni), nrow = nsim)
stdMQO_ni    <- matrix(0, ncol = length(b0_ni), nrow = nsim)

for (i in 1:nsim) {
  post_ni <- conjugada_natural(Y = dados$Y_real[[i]],
                               X = dados$X, nu0 = nu0_ni,
                               s20 = s02_ni, B0 = b0_ni,
                               V0 = V0_ni, varargin = NULL)
  #post_ni <- conjugada_natural(Y = dados_em$Y_real[[i]],
  #X = dados_em$X, nu0 = nu0_ni, s20 = s02_ni, B0 = b0_ni,
  #V0 = V0_ni, varargin = NULL) #Erro de medida
  beta_ni[i,]      <- post_ni$B1
  beta_MQO_ni[i,]  <- post_ni$BMQO
  ic95_ni[i,]      <- as.vector(t(post_ni$HPDI95))
  ic99_ni[i,]      <- as.vector(t(post_ni$HPDI99))
  stdMQO_ni[i,]    <- post_ni$stdMQO
}
```

Posteriori para a priori com valores distantes e muito desvio padrão

```
beta_nid      <- matrix(0, ncol = length(b0_nid), nrow = nsim)
beta_MQO_nid  <- matrix(0, ncol = length(b0_nid), nrow = nsim)
ic95_nid      <- matrix(0, ncol = 2*length(b0_nid), nrow = nsim)
ic99_nid      <- matrix(0, ncol = 2*length(b0_nid), nrow = nsim)
stdMQO_nid    <- matrix(0, ncol = length(b0_nid), nrow = nsim)

for (i in 1:nsim) {
  post_nid <- conjugada_natural(Y = dados$Y_real[[i]],
                                X = dados$X, nu0 = nu0_nid,
                                s20 = s02_nid, B0 = b0_nid,
```

```
                                V0 = V0_nid, varargin = NULL)
  #post_nid <- conjugada_natural(Y = dados_em$Y_real[[i]],
  #X = dados_em$X, nu0 = nu0_nid, s20 = s02_nid, B0 = b0_nid,
  #V0 = V0_nid, varargin = NULL) # Erro de medida
  beta_nid[i,]      <- post_nid$B1
  beta_MQO_nid[i,]  <- post_nid$BMQO
  ic95_nid[i,]      <- as.vector(t(post_nid$HPDI95))
  ic99_nid[i,]      <- as.vector(t(post_nid$HPDI99))
  stdMQO_nid[i,]    <- post_nid$stdMQO
}
```

Junta os resultados em uma tabela

Função que calcula o EQM

```
EQM <- function (media_amostral, media_pop) {
  n <- length(media_amostral)
  EQM <- (1/n)*sum((media_amostral - media_pop)^2)
  EQM
}
```

Bloco para 5 betas

```
colnames(beta_in) <- c("Beta1", "Beta2", "Beta3", "Beta4", "Beta5")
#beta='\u03b2'

betas <- matrix(0, ncol = 2*length(b0_in)+1, nrow = 7)

betas[1,2:11] <- c(paste0("beta","_1"), paste0("EQM[","beta","_1","]"),
                paste0("beta","_2"), paste0("EQM[","beta","_2","]"),
                paste0("beta","_3"), paste0("EQM[","beta","_3","]"),
                paste0("beta","_4"), paste0("EQM[","beta","_4","]"),
                paste0("beta","_5"), paste0("EQM[","beta","_5","]"))

betas[2:7,1] <- c("Real", "MQO", "Priori 1",
                  "Priori 2", "Priori 3", "Priori 4")

betas[2,2:11] <- c(2, "", 100, "", 0.6, "", 5, "", 0.01, "")

beta_real <- dados$Beta_real

### COM EQM

betas[3,2:11] <- c(round(mean(beta_MQO_in[,1]),5),
                   round(EQM(beta_MQO_in[,1], beta_real[1]),5),
                   round(mean(beta_MQO_in[,2]),5),
                   round(EQM(beta_MQO_in[,2], beta_real[2]),5),
                   round(mean(beta_MQO_in[,3]),5),
                   round(EQM(beta_MQO_in[,3], beta_real[3]),5),
                   round(mean(beta_MQO_in[,4]),5),
                   round(EQM(beta_MQO_in[,4], beta_real[4]),5),
                   round(mean(beta_MQO_in[,5]),5),
                   round(EQM(beta_MQO_in[,5], beta_real[5]),5))
```

```
betas[4,2:11] <- c(round(mean(beta_in[,1]),5),
                   round(EQM(beta_in[,1], beta_real[1]),5),
                   round(mean(beta_in[,2]),5),
                   round(EQM(beta_in[,2], beta_real[2]),5),
                   round(mean(beta_in[,3]),5),
                   round(EQM(beta_in[,3], beta_real[3]),5),
                   round(mean(beta_in[,4]),5),
                   round(EQM(beta_in[,4], beta_real[4]),5),
                   round(mean(beta_in[,5]),5),
                   round(EQM(beta_in[,5], beta_real[5]),5))

betas[5,2:11] <- c(round(mean(beta_ind[,1]),5),
                   round(EQM(beta_ind[,1], beta_real[1]),5),
                   round(mean(beta_ind[,2]),5),
                   round(EQM(beta_ind[,2], beta_real[2]),5),
                   round(mean(beta_ind[,3]),5),
                   round(EQM(beta_ind[,3], beta_real[3]),5),
                   round(mean(beta_ind[,4]),5),
                   round(EQM(beta_ind[,4], beta_real[4]),5),
                   round(mean(beta_ind[,5]),5),
                   round(EQM(beta_ind[,5], beta_real[5]),5))

betas[6,2:11] <- c(round(mean(beta_ni[,1]),5),
                   round(EQM(beta_ni[,1], beta_real[1]),5),
                   round(mean(beta_ni[,2]),5),
                   round(EQM(beta_ni[,2], beta_real[2]),5),
                   round(mean(beta_ni[,3]),5),
                   round(EQM(beta_ni[,3], beta_real[3]),5),
                   round(mean(beta_ni[,4]),5),
                   round(EQM(beta_ni[,4], beta_real[4]),5),
                   round(mean(beta_ni[,5]),5),
                   round(EQM(beta_ni[,5], beta_real[5]),5))

betas[7,2:11] <- c(round(mean(beta_nid[,1]),5),
                   round(EQM(beta_nid[,1], beta_real[1]),5),
                   round(mean(beta_nid[,2]),5),
                   round(EQM(beta_nid[,2], beta_real[2]),5),
                   round(mean(beta_nid[,3]),5),
                   round(EQM(beta_nid[,3], beta_real[3]),5),
                   round(mean(beta_nid[,4]),5),
                   round(EQM(beta_nid[,4], beta_real[4]),5),
                   round(mean(beta_nid[,5]),5),
                   round(EQM(beta_nid[,5], beta_real[5]),5))


betas <- as.data.frame(betas)
stargazer(betas, summary=FALSE, rownames = FALSE,
          header = FALSE, type = 'latex')
```

Bloco para 4 betas

```
colnames(beta_in) <- c("Beta1", "Beta2", "Beta3", "Beta4")
#beta='\u03b2'

betas <- matrix(0, ncol = 2*length(b0_in)+1, nrow = 7)

betas[1,2:(2*length(b0_in)+1)] <- c(paste0("beta","_1"),
                                    paste0("EQM["," "beta","_1","]"),
                paste0("beta","_2"), paste0("EQM["," "beta","_2","]"),
                paste0("beta","_3"), paste0("EQM["," "beta","_3","]"),
                paste0("beta","_4"), paste0("EQM["," "beta","_4","]"))

betas[2:7,1] <- c("Real", "MQO", "Priori 1",
                  "Priori 2", "Priori 3", "Priori 4")

betas[2,2:(2*length(b0_in)+1)] <- c(1, "", 10, "", 50, "", 100, "")

beta_real <- dados$Beta_real

# Com EQM

betas[3,2:(2*length(b0_in)+1)] <- c(round(mean(beta_MQO_in[,1]),5),
                                    round(EQM(beta_MQO_in[,1], beta_real[1]),5),
                    round(mean(beta_MQO_in[,2]),5),
                    round(EQM(beta_MQO_in[,2], beta_real[2]),5),
                    round(mean(beta_MQO_in[,3]),5),
                    round(EQM(beta_MQO_in[,3], beta_real[3]),5),
                    round(mean(beta_MQO_in[,4]),5),
                    round(EQM(beta_MQO_in[,4], beta_real[4]),5))

betas[4,2:(2*length(b0_in)+1)] <- c(round(mean(beta_in[,1]),5),
                                    round(sd(beta_in[,1]),5),
                    round(mean(beta_in[,2]),5),
                    round(EQM(beta_in[,2], beta_real[2]),5),
                    round(mean(beta_in[,3]),5),
                    round(EQM(beta_in[,3], beta_real[3]),5),
                    round(mean(beta_in[,4]),5),
                    round(EQM(beta_in[,4], beta_real[4]),5))

betas[5,2:(2*length(b0_in)+1)] <- c(round(mean(beta_ind[,1]),5),
                                    round(EQM(beta_ind[,1], beta_real[1]),5),
                    round(mean(beta_ind[,2]),5),
                    round(EQM(beta_ind[,2], beta_real[2]),5),
                    round(mean(beta_ind[,3]),5),
                    round(EQM(beta_ind[,3], beta_real[3]),5),
                    round(mean(beta_ind[,4]),5),
                    round(EQM(beta_ind[,4], beta_real[4]),5))

betas[6,2:(2*length(b0_in)+1)] <- c(round(mean(beta_ni[,1]),5),
                                    round(EQM(beta_ni[,1], beta_real[1]),5),
                    round(mean(beta_ni[,2]),5),
                    round(EQM(beta_ni[,2], beta_real[2]),5),
                    round(mean(beta_ni[,3]),5),
                    round(EQM(beta_ni[,3], beta_real[3]),5),
```

```
                    round(mean(beta_ni[,4]),5),
                    round(EQM(beta_ni[,4], beta_real[4]),5))

betas[7,2:(2*length(b0_in)+1)] <- c(round(mean(beta_nid[,1]),5),
                                    round(EQM(beta_nid[,1], beta_real[1]),5),
                    round(mean(beta_nid[,2]),5),
                    round(EQM(beta_nid[,2], beta_real[2]),5),
                    round(mean(beta_nid[,3]),5),
                    round(EQM(beta_nid[,3], beta_real[3]),5),
                    round(mean(beta_nid[,4]),5),
                    round(EQM(beta_nid[,4], beta_real[4]),5))


betas <- as.data.frame(betas)
stargazer(betas, summary=FALSE, rownames = FALSE,
          header = FALSE, type = 'latex')
```

Bloco para 3 betas

```
colnames(beta_in) <- c("Beta2", "Beta3", "Beta4")
#beta='\u03b2'

betas <- matrix(0, ncol = 2*length(b0_in)+1, nrow = 7)

betas[1,2:(2*length(b0_in)+1)] <- c(paste0("beta","_1"),
                                    paste0("EQM[","beta","_1","]"),
                paste0("beta","_2"), paste0("EQM[","beta","_2","]"),
                paste0("beta","_3"), paste0("EQM[","beta","_3","]"))

betas[2:7,1] <- c("Real", "MQO", "Priori 1", "Priori 2", "Priori 3", "Priori 4")

betas[2,2:(2*length(b0_in)+1)] <- c(10, "", 50, "", 100, "")

beta_real <- dados$Beta_real

# Com EQM

betas[3,2:(2*length(b0_in)+1)] <- c(round(mean(beta_MQO_in[,1]),5),
                                    round(EQM(beta_MQO_in[,1], beta_real[1]),5),
                    round(mean(beta_MQO_in[,2]),5),
                    round(EQM(beta_MQO_in[,2], beta_real[2]),5),
                    round(mean(beta_MQO_in[,3]),5),
                    round(EQM(beta_MQO_in[,3], beta_real[3]),5))

betas[4,2:(2*length(b0_in)+1)] <- c(round(mean(beta_in[,1]),5),
                                    round(EQM(beta_in[,1], beta_real[1]),5),
                    round(mean(beta_in[,2]),5),
                    round(EQM(beta_in[,2], beta_real[2]),5),
                    round(mean(beta_in[,3]),5),
                    round(EQM(beta_in[,3], beta_real[3]),5))

betas[5,2:(2*length(b0_in)+1)] <- c(round(mean(beta_ind[,1]),5),
                                    round(EQM(beta_ind[,1], beta_real[1]),5),
```

```
                    round(mean(beta_ind[,2]),5),
                    round(EQM(beta_ind[,2], beta_real[2]),5),
                    round(mean(beta_ind[,3]),5),
                    round(EQM(beta_ind[,3], beta_real[3]),5))

betas[6,2:(2*length(b0_in)+1)] <- c(round(mean(beta_ni[,1]),5),
                                    round(EQM(beta_ni[,1], beta_real[1]),5),
                    round(mean(beta_ni[,2]),5),
                    round(EQM(beta_ni[,2], beta_real[2]),5),
                    round(mean(beta_ni[,3]),5),
                    round(EQM(beta_ni[,3], beta_real[3]),5))

betas[7,2:(2*length(b0_in)+1)] <- c(round(mean(beta_nid[,1]),5),
                                    round(EQM(beta_nid[,1], beta_real[1]),5),
                    round(mean(beta_nid[,2]),5),
                    round(EQM(beta_nid[,2], beta_real[2]),5),
                    round(mean(beta_nid[,3]),5),
                    round(EQM(beta_nid[,3], beta_real[3]),5))

betas <- as.data.frame(betas)
stargazer(betas, summary=FALSE, rownames = FALSE, header = FALSE, type = 'latex')
```

Gráficos

Histogramas

Este código faz os histogramas sem a estimativa da função densidade e foi o que usei no trabalho final.

```
scaleFUN1 <- function(x) sprintf("%.4f", x)
scaleFUN2 <- function(x) sprintf("%.2f", x)
scaleFUN <- function(x) sprintf("%.4f", x)

cores <- c("#65b39d", "#cc6215", "#9a96bf", "#d63386", "#dfba56", "#d14f5a")

print("Estimativas de MQO")

p1 <- ggplot(data = data.frame(beta_MQO_in), aes(beta_MQO_in[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_1$"), y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN, breaks = round(seq(min(beta_MQO_in[,1]),
                                                         max(beta_MQO_in[,1]), by = (max(beta_MQO_in[,1
  geom_text(aes(quantile(beta_MQO_in[,1],0.03), 900,
                label=paste("Média\n",round(mean(beta_MQO_in[,1]),4))),
            colour="black", size=2.15)

p1 <- p1 + geom_vline(aes(xintercept=mean(beta_MQO_in[,1])),
            color= cores[1], linetype="dashed", size=1)
p1 <- p1 + theme_bw()  +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7), axis.text.y = element_text(size=7),
```

```
        axis.title.y = element_text(size = 7),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p2 <- ggplot(data = data.frame(beta_MQO_in), aes(beta_MQO_in[,2])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_2$"), y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_MQO_in[,2]),
                                        max(beta_MQO_in[,2]), by = (max(beta_MQO_in[,2])-min(beta_MQO_i
  geom_text(aes(quantile(beta_MQO_in[,2],0.03), 900,
                label=paste("Média\n",round(mean(beta_MQO_in[,2]),4))),
            colour="black", size=2.15)

p2 <- p2 + geom_vline(aes(xintercept=mean(beta_MQO_in[,2])),
            color= cores[2], linetype="dashed", size=1)
p2 <- p2 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p3 <- ggplot(data = data.frame(beta_MQO_in), aes(beta_MQO_in[,3])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_3$"), y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_MQO_in[,3]),
                                        max(beta_MQO_in[,3]), by = (max(beta_MQO_in[,3])-min(beta_MQO_i
  geom_text(aes(quantile(beta_MQO_in[,3],0.03), 900,
                label=paste("Média\n",round(mean(beta_MQO_in[,3]),4))),
            colour="black", size=2.15)

p3 <- p3 + geom_vline(aes(xintercept=mean(beta_MQO_in[,3])),
            color= cores[3], linetype="dashed", size=1)
p3 <- p3 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p4 <- ggplot(data = data.frame(beta_MQO_in),
             aes(beta_MQO_in[,4])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
```

```
  labs(title=TeX("$\\hat{\\beta}_4$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_MQO_in[,4]),
                                        max(beta_MQO_in[,4]), by = (max(beta_MQO_in[,4])-min(beta_MQO_i
  geom_text(aes(quantile(beta_MQO_in[,4],0.03), 900,
                label=paste("Média\n",round(mean(beta_MQO_in[,4]),4))),
            colour="black", size=2.15)

p4 <- p4 + geom_vline(aes(xintercept=mean(beta_MQO_in[,4])),
            color= cores[4], linetype="dashed", size=1)
p4 <- p4 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

# p5 <- ggplot(data = data.frame(beta_MQO_in), aes(beta_MQO_in[,5])) +
#   geom_histogram(aes(y = ..count..),
#                  alpha = .2, colour = "black", fill = "white") +
#   ggtitle("") +
#   labs(title=TeX("$\\hat{\\beta}_5$"),
# y = "Observações", x= "Estimativas") +
#   scale_x_continuous(labels=scaleFUN,
#     breaks = round(seq(min(beta_MQO_in[,5]),
#                        max(beta_MQO_in[,5]),
# by = (max(beta_MQO_in[,5])-min(beta_MQO_in[,5]))/10),4)) +
#   geom_text(aes(quantile(beta_MQO_in[,5],0.03), 900,
# label=paste("Média\n",round(mean(beta_MQO_in[,5]),4))), colour="black", size=2.15)
#
# p5 <- p5 + geom_vline(aes(xintercept=mean(beta_MQO_in[,5])),
#             color= cores[5], linetype="dashed", size=1)
# p5 <- p5 + theme_bw() +
#   theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
#plot.title = element_text(size=7), axis.text.y = element_text(size=7),
#axis.title.y = element_blank(), axis.title.x = element_text(size = 7),
#plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

#grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)

pdf(file="C:\\...\\mqoh.pdf", width = 10, height = 3)
#grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)    # 5 parâmetros
 grid.arrange(p1, p2, p3, p4, nrow = 1)              # 4 parâmetros
# grid.arrange(p1, p2, p3, ncol=3)                   # 3 parâmetros
dev.off()

##########################################################################################
##########################################################################################
print("Valores próximos dos reais, pouco desvio")

p1 <- ggplot(data = data.frame(beta_in), aes(beta_in[,1])) +
```

```
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_1$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_in[,1]),
                                        max(beta_in[,1]),
                                        by = (max(beta_in[,1])-min(beta_in[,1]))/10),4)) +
  geom_text(aes(quantile(beta_in[,1],0.03), 900,
                label=paste("Média\n",round(mean(beta_in[,1]),4))),
            colour="black", size=2.15)

p1 <- p1 + geom_vline(aes(xintercept=mean(beta_in[,1])),
            color= cores[1], linetype="dashed", size=1)
p1 <- p1 + theme_bw()  +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_text(size = 7),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))


p2 <- ggplot(data = data.frame(beta_in), aes(beta_in[,2])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_2$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_in[,2]),
                                        max(beta_in[,2]),
                               by = (max(beta_in[,2])-min(beta_in[,2]))/10),4)) +
  geom_text(aes(quantile(beta_in[,2],0.03), 900,
                label=paste("Média\n",round(mean(beta_in[,2]),4))),
            colour="black", size=2.15)

p2 <- p2 + geom_vline(aes(xintercept=mean(beta_in[,2])),
            color= cores[2], linetype="dashed", size=1)
p2 <- p2 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p3 <- ggplot(data = data.frame(beta_in), aes(beta_in[,3])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
```

```
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_3$"),  y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_in[,3]), max(beta_in[,3]), by = (max(beta_in[,3])-min(l
  geom_text(aes(quantile(beta_in[,3],0.03), 900,
                label=paste("Média\n",round(mean(beta_in[,3]),4))),
            colour="black", size=2.15)

p3 <- p3 + geom_vline(aes(xintercept=mean(beta_in[,3])),
            color= cores[3], linetype="dashed", size=1)
p3 <- p3 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p4 <- ggplot(data = data.frame(beta_in), aes(beta_in[,4])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_4$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_in[,4]), max(beta_in[,4]), by = (max(beta_in[,4])-min(l
  geom_text(aes(quantile(beta_in[,4],0.03), 900,
                label=paste("Média\n",round(mean(beta_in[,4]),4))),
            colour="black", size=2.15)

p4 <- p4 + geom_vline(aes(xintercept=mean(beta_in[,4])),
            color= cores[4], linetype="dashed", size=1)
p4 <- p4 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

# p5 <- ggplot(data = data.frame(beta_in), aes(beta_in[,5])) +
#   geom_histogram(aes(y = ..count..),
#                  alpha = .2, colour = "black", fill = "white") +
#   ggtitle("") +
#   labs(title=TeX("$\\hat{\\beta}_5$"),
#y = "Observações", x= "Estimativas") +
#   scale_x_continuous(labels=scaleFUN,
#  breaks = round(seq(min(beta_in[,5]),
#                     max(beta_in[,5]), by = (max(beta_in[,5])-min(beta_in[,5]))/10),4)) +
#   geom_text(aes(quantile(beta_in[,5],0.03), 900,
#  label=paste("Média\n",round(mean(beta_in[,5]),4))), colour="black", size=2.15)
#
```

```
# p5 <- p5 + geom_vline(aes(xintercept=mean(beta_in[,5])),
#             color= cores[5], linetype="dashed", size=1)
# p5 <- p5 + theme_bw() +
#   theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6), plot.title = element_text(size=7
#
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)

pdf(file="C:\\...\\p1h.pdf", width = 10, height = 3)
 grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)
# grid.arrange(p1, p2, p3, p4, nrow = 1)
# grid.arrange(p1, p2, p3, ncol=3)
dev.off()

#####################################################################
#####################################################################

print("Valores próximos dos reais, muito desvio")
p1 <- ggplot(data = data.frame(beta_ind), aes(beta_ind[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_1$"),  y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ind[,1]),
                                        max(beta_ind[,1]),
                                        by = (max(beta_ind[,1])-min(beta_ind[,1]))/10),4)) +
  geom_text(aes(quantile(beta_ind[,1],0.03), 900,
                label=paste("Média\n",round(mean(beta_ind[,1]),4))),
            colour="black", size=2.15)

p1 <- p1 + geom_vline(aes(xintercept=mean(beta_ind[,1])),
            color= cores[1], linetype="dashed", size=1)
p1 <- p1 + theme_bw()  +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_text(size = 7),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))


p2 <- ggplot(data = data.frame(beta_ind), aes(beta_ind[,2])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_2$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN, breaks =
                       round(seq(min(beta_ind[,2]),
                                 max(beta_ind[,2]),
                                 by = (max(beta_ind[,2])-min(beta_ind[,2]))/10),4)) +
  geom_text(aes(quantile(beta_ind[,2],0.03), 900,
                label=paste("Média\n",round(mean(beta_ind[,2]),4))),
```

```
            colour="black", size=2.15)

p2 <- p2 + geom_vline(aes(xintercept=mean(beta_ind[,2])),
            color= cores[2], linetype="dashed", size=1)
p2 <- p2 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7), axis.text.y = element_text(size=7),
        axis.title.y = element_blank(), axis.title.x = element_text(size = 7), plot.margin=unit(c(0.05,(

p3 <- ggplot(data = data.frame(beta_ind), aes(beta_ind[,3])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_3$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ind[,3]), max(beta_ind[,3]), by = (max(beta_ind[,3])-m
  geom_text(aes(quantile(beta_ind[,3],0.03), 900,
                label=paste("Média\n",round(mean(beta_ind[,3]),4))),
            colour="black", size=2.15)

p3 <- p3 + geom_vline(aes(xintercept=mean(beta_ind[,3])),
            color= cores[3], linetype="dashed", size=1)
p3 <- p3 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p4 <- ggplot(data = data.frame(beta_ind), aes(beta_ind[,4])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_4$"),  y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ind[,4]),
                                        max(beta_ind[,4]),
                                        by = (max(beta_ind[,4])-min(beta_ind[,4]))/10),4)) +
  geom_text(aes(quantile(beta_ind[,4],0.03), 900,
                label=paste("Média\n",round(mean(beta_ind[,4]),4))),
            colour="black", size=2.15)

p4 <- p4 + geom_vline(aes(xintercept=mean(beta_ind[,4])),
            color= cores[4], linetype="dashed", size=1)
p4 <- p4 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))
```

```
# p5 <- ggplot(data = data.frame(beta_ind), aes(beta_ind[,5])) +
#   geom_histogram(aes(y = ..count..),
#                  alpha = .2, colour = "black", fill = "white") +
#   ggtitle("") +
#   labs(title=TeX("$\\hat{\\beta}_5$"),
# y = "Observações", x= "Estimativas") +
#   scale_x_continuous(labels=scaleFUN,
#  breaks = round(seq(min(beta_ind[,5]),
#                     max(beta_ind[,5]),
#                     by = (max(beta_ind[,5])-min(beta_ind[,5]))/10),4)) +
#   geom_text(aes(quantile(beta_ind[,5],0.03), 900,
#  label=paste("Média\n",round(mean(beta_ind[,5]),4))),
#colour="black", size=2.15)
#
# p5 <- p5 + geom_vline(aes(xintercept=mean(beta_ind[,5])),
#             color= cores[5], linetype="dashed", size=1)
# p5 <- p5 + theme_bw() +
#   theme(axis.text.x = element_text(angle = 90, hjust = 1,
#size = 6), plot.title = element_text(size=7),
#axis.text.y = element_text(size=7),
#axis.title.y = element_blank(),
#axis.title.x = element_text(size = 7),
#plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))
#
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)

pdf(file="C:\\...\\p2h.pdf", width = 10, height = 3)
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)
grid.arrange(p1, p2, p3, p4, nrow = 1)
# grid.arrange(p1, p2, p3, ncol=3)
dev.off()

##############################################################
##############################################################

print("Valores distantes dos reais, pouco desvio")
p1 <- ggplot(data = data.frame(beta_ni), aes(beta_ni[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_1$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ni[,1]),
                                        max(beta_ni[,1]),
                                        by = (max(beta_ni[,1])-min(beta_ni[,1]))/10),4)) +
  geom_text(aes(quantile(beta_ni[,1],0.03), 900,
                label=paste("Média\n",round(mean(beta_ni[,1]),4))),
            colour="black", size=2.15)

p1 <- p1 + geom_vline(aes(xintercept=mean(beta_ni[,1])),
            color= cores[1], linetype="dashed", size=1)
p1 <- p1 + theme_bw()  +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6),
        plot.title = element_text(size=7), axis.text.y = element_text(size=7),
        axis.title.y = element_text(size = 7), axis.title.x = element_text(size = 7), plot.margin=unit(c


p2 <- ggplot(data = data.frame(beta_ni), aes(beta_ni[,2])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_2$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ni[,2]),
                                        max(beta_ni[,2]),
                                        by = (max(beta_ni[,2])-min(beta_ni[,2]))/10),4)) +
  geom_text(aes(quantile(beta_ni[,2],0.03), 900,
                label=paste("Média\n",round(mean(beta_ni[,2]),4))),
            colour="black", size=2.15)

p2 <- p2 + geom_vline(aes(xintercept=mean(beta_ni[,2])),
            color= cores[2], linetype="dashed", size=1)
p2 <- p2 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p3 <- ggplot(data = data.frame(beta_ni), aes(beta_ni[,3])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_3$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ni[,3]),
                                        max(beta_ni[,3]),
                                        by = (max(beta_ni[,3])-min(beta_ni[,3]))/10),4)) +
  geom_text(aes(quantile(beta_ni[,3],0.03), 900,
                label=paste("Média\n",round(mean(beta_ni[,3]),4))),
            colour="black", size=2.15)

p3 <- p3 + geom_vline(aes(xintercept=mean(beta_ni[,3])),
            color= cores[3], linetype="dashed", size=1)
p3 <- p3 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
```

```
        plot.margin=unit(c(0.05,0.05,0.05,0.05),
                         units = "cm"))

p4 <- ggplot(data = data.frame(beta_ni), aes(beta_ni[,4])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_4$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_ni[,4]),
                                        max(beta_ni[,4]),
                                        by = (max(beta_ni[,4])-min(beta_ni[,4]))/10),4)) +
  geom_text(aes(quantile(beta_ni[,4],0.03), 900,
                label=paste("Média\n",round(mean(beta_ni[,4]),4))),
            colour="black", size=2.15)

p4 <- p4 + geom_vline(aes(xintercept=mean(beta_ni[,4])),
            color= cores[4], linetype="dashed", size=1)
p4 <- p4 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05),
                         units = "cm"))

# p5 <- ggplot(data = data.frame(beta_ni), aes(beta_ni[,5])) +
#   geom_histogram(aes(y = ..count..),
#                  alpha = .2, colour = "black", fill = "white") +
#   ggtitle("") +
#   labs(title=TeX("$\\hat{\\beta}_5$"),
# y = "Observações", x= "Estimativas") +
#   scale_x_continuous(labels=scaleFUN,
#  breaks = round(seq(min(beta_ni[,5]), max(beta_ni[,5]),
# by = (max(beta_ni[,5])-min(beta_ni[,5]))/10),4)) +
#   geom_text(aes(quantile(beta_ni[,5],0.03),
#  900, label=paste("Média\n",round(mean(beta_ni[,5]),4))),
# colour="black", size=2.15)
#
# p5 <- p5 + geom_vline(aes(xintercept=mean(beta_ni[,5])),
#             color= cores[5], linetype="dashed", size=1)
# p5 <- p5 + theme_bw() +
#   theme(axis.text.x = element_text(angle = 90,
#hjust = 1, size = 6), plot.title = element_text(size=7),
#axis.text.y = element_text(size=7),
#axis.title.y = element_blank(),
#axis.title.x = element_text(size = 7),
#plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))
#
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)
```

```
pdf(file="C:\\...\\p3h.pdf", width = 10, height = 3)
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)
grid.arrange(p1, p2, p3, p4, nrow = 1)
# grid.arrange(p1, p2, p3, ncol=3)
dev.off()

##########################################################
#########################################################

print("Valores distantes dos reais, muito desvio")
p1 <- ggplot(data = data.frame(beta_nid), aes(beta_nid[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_1$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_nid[,1]),
                                        max(beta_nid[,1]),
                                        by = (max(beta_nid[,1])-min(beta_nid[,1]))/10),4))+
  geom_text(aes(quantile(beta_nid[,1],0.03), 900,
                label=paste("Média\n",round(mean(beta_nid[,1]),4))),
            colour="black", size=2.15)

p1 <- p1 + geom_vline(aes(xintercept=mean(beta_nid[,1])),
            color= cores[1], linetype="dashed", size=1)
p1 <- p1 + theme_bw()  +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_text(size = 7),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))


p2 <- ggplot(data = data.frame(beta_nid), aes(beta_nid[,2])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_2$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_nid[,2]),
                                        max(beta_nid[,2]),
                                        by = (max(beta_nid[,2])-min(beta_nid[,2]))/10),4)) +
  geom_text(aes(quantile(beta_nid[,2],0.03), 900,
                label=paste("Média\n",round(mean(beta_nid[,2]),4))),
            colour="black", size=2.15)

p2 <- p2 + geom_vline(aes(xintercept=mean(beta_nid[,2])),
            color= cores[2], linetype="dashed", size=1)
p2 <- p2 + theme_bw() +
```

```
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05),
                         units = "cm"))

p3 <- ggplot(data = data.frame(beta_nid), aes(beta_nid[,3])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_3$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_nid[,3]), max(beta_nid[,3]), by = (max(beta_nid[,3])-m
  geom_text(aes(quantile(beta_nid[,3],0.03), 900,
                label=paste("Média\n",round(mean(beta_nid[,3]),4))),
            colour="black", size=2.15)

p3 <- p3 + geom_vline(aes(xintercept=mean(beta_nid[,3])),
            color= cores[3], linetype="dashed", size=1)
p3 <- p3 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

p4 <- ggplot(data = data.frame(beta_nid), aes(beta_nid[,4])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black", fill = "white") +
  ggtitle("") +
  labs(title=TeX("$\\hat{\\beta}_4$"),
       y = "Observações", x= "Estimativas") +
  scale_x_continuous(labels=scaleFUN,
                     breaks = round(seq(min(beta_nid[,4]),
                                        max(beta_nid[,4]),
                                        by = (max(beta_nid[,4])-min(beta_nid[,4]))/10),4)) +
  geom_text(aes(quantile(beta_nid[,4],0.03), 900,
                label=paste("Média\n",round(mean(beta_nid[,4]),4))),
            colour="black", size=2.15)

p4 <- p4 + geom_vline(aes(xintercept=mean(beta_nid[,4])),
            color= cores[4], linetype="dashed", size=1)
p4 <- p4 + theme_bw() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 6),
        plot.title = element_text(size=7),
        axis.text.y = element_text(size=7),
```

```
        axis.title.y = element_blank(),
        axis.title.x = element_text(size = 7),
        plot.margin=unit(c(0.05,0.05,0.05,0.05), units = "cm"))

# p5 <- ggplot(data = data.frame(beta_nid), aes(beta_nid[,5])) +
#   geom_histogram(aes(y = ..count..),
#                  alpha = .2, colour = "black", fill = "white") +
#   ggtitle("") +
#   labs(title=TeX("$\\hat{\\beta}_5$"),
# y = "Observações", x= "Estimativas") +
#   scale_x_continuous(labels=scaleFUN,
#  breaks = round(seq(min(beta_nid[,5]), max(beta_nid[,5]),
#                     by = (max(beta_nid[,5])-min(beta_nid[,5]))/10),4)) +
#   geom_text(aes(quantile(beta_nid[,5],0.03), 900,
#  label=paste("Média\n",round(mean(beta_nid[,5]),4))),
#colour="black", size=2.15)
#
# p5 <- p5 + geom_vline(aes(xintercept=mean(beta_nid[,5])),
#             color= cores[5], linetype="dashed", size=1)
# p5 <- p5 + theme_bw() +
#   theme(axis.text.x = element_text(angle = 90,
#  hjust = 1, size = 6), plot.title = element_text(size=7),
#  axis.text.y = element_text(size=7),
#  axis.title.y = element_blank(),
#  axis.title.x = element_text(size = 7),
#  plot.margin=unit(c(0.05,0.05,0.05,0.05),
#                   units = "cm"))
#
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)

pdf(file="C:\\...\\p4h.pdf", width = 10, height = 3)
# grid.arrange(p1, p2, p3, p4, p5, ncol=5, nrow = 1)
grid.arrange(p1, p2, p3, p4, nrow = 1)
# grid.arrange(p1, p2, p3, ncol=3)
dev.off()
```

Capítulo 4 - Aplicação

Instala e carrega os pacotes usados

```
list.of.packages <- c("ggplot2", "metRology", "reshape2",
                      "gridExtra","knitr", "stargazer", "foreign",
                      "dplyr", "lmtest", "sandwich")
new.packages <- list.of.packages[!(list.of.packages
                                   %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
library(ggplot2, quietly = T)
library(metRology, quietly = T)
library(reshape2, quietly = T)
library(gridExtra, quietly = T)
library(knitr, quietly = T)
library(stargazer, quietly = T)
library(foreign, quietly = T)
library(dplyr, quietly = T)
library(lmtest, quietly = T)
library(sandwich, quietly = T)
options(scipen=999)
```

Leitura e descritivas dos dados

Arquivo dadosraistisc

```
dadosraistisc <- read.dta("C:\\...\\dadosraistisc.dta")
head(dadosraistisc)
glimpse(dadosraistisc)
tail(dadosraistisc)

# Utiliza a notação decimal
options(scipen=9999)

N <- nrow(dadosraistisc)     # Determina o tamanho da amostra
k <- 7                       # Número de regressores

# Junta as variáveis com o intercepto
dados <- cbind(dadosraistisc$lnsalh, rep(1,N),
               dadosraistisc[,c(17,18,21,6,14,11,10)])
dados <- as.matrix(dados)
dados <- dados[complete.cases(dados),]
colnames(dados) <- c("lnsalh", "cons", "supinc",
                     "sup", "pos", "idade", "idade2", "tempemp", "sexo")
summary(dados)
sd(dados[,1])

## Sorteio dos valores para serem previstos
#round(runif(6,min = 1,max = 4328),0)
```

```r
# 3997 1266 3624 1239 1156  809

# Acabei escolhendo no olho para ter bastante perfis diferentes
## Separa em dois bancos de dados
dados[c(3997, 1266, 3624, 809, 3196, 1,
        661, 2007, 670, 2405),]
novos_dados <- dados[c(3997, 1266, 3624, 809,
                       3196, 1, 661, 2007, 670, 2405),]
dados_antigos <- dados[-c(3997, 1266, 3624, 809,
                          3196, 1, 661, 2007, 670, 2405),]

dados_antigos2 <- data.frame(dados_antigos)

dados_antigos2$sexo[dados_antigos2$sexo == 1] <- 0
dados_antigos2$sexo[dados_antigos2$sexo == 2] <- 1
dados_antigos2$pos[dados_antigos2$pos == 2] <- 1

novos_dados2 <- data.frame(novos_dados)

novos_dados2$sexo[novos_dados2$sexo == 1] <- 0
novos_dados2$sexo[novos_dados2$sexo == 2] <- 1
novos_dados2$pos[novos_dados2$pos == 2] <- 1

summary(dados_antigos2[,"idade"])
sd(dados_antigos2[,"idade"])
nrow(dados_antigos2)

summary(dados_antigos2[,"idade2"])
sd(dados_antigos2[,"idade2"])

summary(dados_antigos2[,"tempemp"])
sd(dados_antigos2[,"tempemp"])

summary(dados_antigos2[,"lnsalh"])
sd(dados_antigos2[,"lnsalh"])

table(dados_antigos2[,"sexo"])
round(table(dados_antigos2[,"sexo"])/nrow(dados_antigos2), 4)

table(dados_antigos2[,"supinc"])
round(table(dados_antigos2[,"supinc"])/nrow(dados_antigos2), 4)

table(dados_antigos2[,"sup"])
round(table(dados_antigos2[,"sup"])/nrow(dados_antigos2), 4)

table(dados_antigos2[,"pos"])
round(table(dados_antigos2[,"pos"])/nrow(dados_antigos2), 4)

# Numeração das variáveis

# 4 - escolaridadeaps2005
# 5 - qtdhoracontr
# 6 - idade
```

```
# 10 - sexo
# 11 - tempemp
# 12 - salhr
# 13 - lnsalh
# 14 - idade2
# 16 - mdio
# 17 - supinc
# 18 - sup
# 19 - mes
# 20 - dout
# 21 - pos
```

```
# Gráficos

# Idade
p <- ggplot(as.data.frame(dados_antigos),
            aes(idade, lnsalh)) +
  geom_point(color = "#65b39d") +
  geom_smooth(method=lm, se=FALSE,
              linetype="dashed", color="black") +
  labs(title="", x = "Idade (anos)",
       y= "Logaritmo do Salário Hora") +
  theme_bw()
p
pdf(file="C:\\...\\Fig-4.002.pdf", width = 6, height = 3)
p
dev.off()

# Idade^2
p <- ggplot(as.data.frame(dados_antigos),
            aes(idade2, lnsalh)) +
  geom_point(color = "#9a96bf") +
  geom_smooth(method=lm, se=FALSE,
              linetype="dashed", color="black") +
  labs(title="", x = "Idade ao quadrado (anos ao quadrado)",
       y= "Logaritmo do Salário Hora") +
  theme_bw()
p
pdf(file="C:\\...\\Fig-4.003.pdf", width = 6, height = 3)
p
dev.off()

# Tempo de emprego
p <- ggplot(as.data.frame(dados_antigos),
            aes(tempemp, lnsalh)) +
  geom_point(color = "#9a96bf") +
  geom_smooth(method=lm, se=FALSE,
              linetype="dashed", color="black") +
  labs(title="", x = "Tempo de Emprego (meses)",
       y= "Logaritmo do Salário Hora") +
  theme_bw()
p
pdf(file="C:\\...\\Fig-4.004.pdf", width = 6, height = 3)
p
```

```
dev.off()

# Ensino médio
p <- ggplot(as.data.frame(dadosraistisc),
            aes(factor(mdio),lnsalh, fill=factor(mdio))) +
  geom_boxplot(alpha = 0.5) +
  ggtitle("") +
  scale_fill_manual(values = c("#65b39d", "#9a96bf"),
                    name = "", labels = c("Não", "Sim")) +
  scale_x_discrete(breaks=c(0,1), labels = c("Não", "Sim")) +
  labs(title="", y = "Logaritmo do Salário Hora",
       x= "Escolaridade máxima = Ensino Médio") +
  theme_bw()
p<-p + theme(legend.position="none")
p

pdf(file="C:\\...\\Fig-4.005.pdf", width = 6, height = 3)
p
dev.off()


# Superior Incompleto
p <- ggplot(as.data.frame(dados_antigos),
            aes(factor(supinc),lnsalh, fill=factor(supinc))) +
  geom_boxplot(alpha = 0.5) +
  ggtitle("") +
  scale_fill_manual(values = c("#65b39d", "#9a96bf"),
                    name = "", labels = c("Não", "Sim")) +
  scale_x_discrete(breaks=c(0,1), labels = c("Não", "Sim")) +
  labs(title="", y = "Logaritmo do Salário Hora",
       x= "Escolaridade máxima = Ensino Superior Incompleto") +
  theme_bw()
p <- p + theme(legend.position="none")
p

pdf(file="C:\\...\\Fig-4.006.pdf", width = 6, height = 3)
p
dev.off()

# Superior
p <- ggplot(as.data.frame(dados_antigos),
            aes(factor(sup),lnsalh, fill=factor(sup))) +
  geom_boxplot(alpha = 0.5) +
  ggtitle("") +
  scale_fill_manual(values = c("#65b39d", "#9a96bf"),
                    name = "", labels = c("Não", "Sim")) +
  scale_x_discrete(breaks=c(0,1), labels = c("Não", "Sim")) +
  labs(title="", y = "Logaritmo do Salário Hora",
       x= "Escolaridade máxima = Ensino Superior Completo") +
  theme_bw()
p <-p + theme(legend.position="none")
p
```

```
pdf(file="C:\\...\\Fig-4.007.pdf", width = 6, height = 3)
p
dev.off()

# Pós graduação
p <- ggplot(as.data.frame(dados_antigos),
            aes(factor(pos),lnsalh, fill=factor(pos))) +
  geom_boxplot(alpha = 0.5) +
  ggtitle("") +
  scale_fill_manual(values = c("#65b39d", "#9a96bf"),
                    name = "", labels = c("Não", "Sim")) +
  scale_x_discrete(breaks=c(0,2), labels = c("Não", "Sim")) +
  labs(title="", y = "Logaritmo do Salário Hora",
       x= "Escolaridade máxima = Pós Graduação") +
  theme_bw()
p <- p + theme(legend.position="none")
p

pdf(file="C:\\...\\Fig-4.008.pdf", width = 6, height = 3)
p
dev.off()

# Gênero
p <- ggplot(as.data.frame(dados_antigos),
            aes(factor(sexo),lnsalh, fill=factor(sexo))) +
  geom_boxplot(alpha = 0.5) +
  ggtitle("") +
  scale_fill_manual(values = c("#3146bc", "#d14f5a"),
                    name = "", labels = c("Homem", "Mulher")) +
  scale_x_discrete(breaks=c(1,2),
                   labels = c("Homem", "Mulher")) +
  labs(title="", y = "Logaritmo do Salário Hora",
       x= "Gênero") +
  theme_bw()
p <-p + theme(legend.position="none")
p

pdf(file="C:\\...\\Fig-4.009.pdf", width = 6, height = 3)
p
dev.off()
```

Modelo teórico

O modelo proposto é:

$$\ln y_i = \beta_0 + \beta_1 \text{supinc} + \beta_2 \text{sup} + \beta_3 \text{pos} + \beta_4 \text{idade} + \beta_5 \text{idade}^2 + \beta_6 \text{genero} + u_i$$

Onde:

- $\ln y_i$ é o logaritmo do salário hora do i-ésimo trabalhador;
- β_0 é uma variável latente que representa a heterogeneidade não observada no modelo;

- *supinc* é uma variável binária que indica se a escolaridade do indivíduo é igual a curso superior incompleto;
- *sup* é uma variável binária que indica se a escolaridade do indivíduo é igual a curso superior;
- *pos* é uma variável binária que indica se a escolaridade do indivíduo é igual a pós-graduação (mestrado, doutorado ou especialização);
- *idade* é a idade, em meses, do indivíduo e $idade^2$ é o seu valor ao quadrado;
- *genero* é uma variável binária que é zero se homem e igual a 1 se mulher.

Determinação da priori

Distribuição dos salários hora:

```
p1 <- ggplot(data = data.frame(dados),
             aes(dados[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black",
                 fill = "#65b39d") +
  ggtitle("") +
  labs(title="", y="No. obs.",
       x="Logaritmo do Salário Hora") +
  scale_x_continuous(breaks = seq(1, 8, 0.5),
                     limits = c(1,8)) +
  scale_colour_brewer(palette="Dark2") +
  theme_bw()

p1

# Altura e largura são em inches, default = 7
pdf(file="C:\\...\\Fig-4.001.pdf", width = 6, height = 3)
p1
dev.off()
```

```
# Seta uma priori informativa
# * Intercepto = 2.07$
# * Coef. Superior Incompleto  = 0.122687$
# * Coef. Superior Completo = 0.44866$
# * Coef. Pós-Graduação  = 1.070809$
# * Coef. Idade = 0.128502$
# * Coef. Idade$^2$  = -0.00143$
# * Coef. Tempo de emprego: 0.00271
# * Coef. Gênero: $\beta_6. = -0.15366$

b0   <- c(2.07, 0.122687, 0.44866, 1.070809,
          0.128502, -0.00143, 0.00271, -0.15366) # Priori Informativa

s02   <- 1  # Desvio padrão dos erros a priori
nu0   <- 44       # Graus de liberdade a priori
h0    <- 1/s02   # Priori para h
stdh0 <- 2/(s02*nu0)

# Para encontrar os elementos da matriz V, utilizamos a fórmula do texto
cons    <- (nu0 * s02)/(nu0-2)
varbeta <- c(100, 100, 100, 100, 100, 100, 100, 100)
```

```
V0_in   <- diag(8)*(varbeta/cons) # Matriz de escala da priori
```

Cálculos a posteriori

```
conjugada_natural <- function(Y, X, nu0, s20, B0, V0, varargin = NULL) {

  # Cálculo das quantidades de MQO
  N      <- nrow(X)    # Tamanho da amostra a partir da matriz X
  k      <- ncol(X)    # Número de regressores
  nu     <- N-k        # Graus de liberdade
  XX     <- t(matrix(X, ncol = k)) %*% matrix(X, ncol=k)
  iXX    <- solve(XX)
  BMQO   <- iXX %*% (t(X) %*% Y)
  # Beta de MQO
  s2     <- (t(Y-X %*% BMQO) %*% (Y-X %*% BMQO))/nu
  # Estimativa não viesada da var dos erros
  CovMQO <- as.numeric(s2) * iXX
  # Matriz de covariância do beta de MQO
  stdMQO <- sqrt(diag(CovMQO))
  # Desvio padrão das estimativas de MQO

  # Parâmetros da posteriori

  nu1 <- nu0 + N # Graus de liberdade
  if (V0 == 0){
    iV0 <- diag(k)*0
  } else {
    iV0 <- solve(V0) # Precisão a posteriori
  }

  iV1   <- iV0 + XX    # Inverso da escala a posteriori
  V1    <- solve(iV1)  # Escala a posteriori
  B1    <- V1 %*% (iV0 %*% B0 + XX %*% BMQO)
  nus21 <- nu0 * s20 + nu * s2

  if (det(iV0) > 0) {
    nus21 <- nus21 + t(BMQO - B0) %*%
      solve(V0 + iXX) %*% (BMQO - B0) # SQR a posteriori
  } else {
    nus21 <- nus21
  }

  s21 <- nus21/nu1

  # Parâmetros da distribuição posterior marginal beta

  CovB <- V1 * as.numeric(nus21/(nu1-2))
  # Variância da distribuição t
  stdB <- sqrt(diag(CovB))
  # Desvio padrão da posteriori marginal de beta

  # Calcula os intervalos de maior
```

```
#densidade a posteriori de 95% e 99%

lim95   <- qt(0.975, nu1) # Valor da t para o quantil 97.5
lim99   <- qt(0.995, nu1) # Valor da t para o quantil 99.5

HPDI95  <- round(cbind(B1 - lim95*stdB, B1 + lim95*stdB),4)
HPDI99  <- round(cbind(B1 - lim99*stdB, B1 + lim99*stdB),4)

probpos=vector()

for (i in 1:k){
  tnorm = -B1[i,1]/sqrt(s21*V1[i,i])
  probpos[i] = 1 - pt(tnorm,nu1)
}

# Parâmetros da distribuição posterior marginal de h

h1    <- 1/s21
# Média da distribuição gama que determina h
h1var <- 2/(nus21)
# Variância da distribuição a posteriori marginal de h

# Logaritmo da função de verossimilhança marginal

lmlik <- vector()
if (det(iV0 > 0)) {
  c <- lgamma(0.5*nu1) +
    0.5*nu0*log(nu0*s20)-lgamma(0.5*nu0)-
    0.5*N*log(pi) # constante de integração
  lmlik <- c+0.5*log(det(V1)/det(V0)) -
    0.5*nu1*log(nus21)  # log da veross. marginal
} else {
  lmlik <- 0
}

# Faz previsão de X* é passado como argumento para a função
if (is.null(varargin) == FALSE) {
  Xestrela   <- varargin
  Yestrela   <- Xestrela %*% B1
  # Previsão Y*
  VYestrela  <- (diag(nrow(Xestrela))+Xestrela %*%
                  V1 %*% t(Xestrela))*as.numeric(s21)
  # Matriz de escala da previsão
  Varestrela <- VYestrela * nu1/(nu1 - 2)
  # Variância da previsão
  HPDI95prev  <- round(cbind(Yestrela - lim95*diag(Varestrela),
                             Yestrela + lim95*diag(Varestrela)),4)
  HPDI99prev  <- round(cbind(Yestrela - lim99*diag(Varestrela),
                             Yestrela + lim99*diag(Varestrela)),4)
} else {
  Yestrela    <- NA
  VYestrela   <- NA
  Varestrela  <- NA
```

```
    HPDI95prev  <- NA
    HPDI99prev  <- NA
  }

  return(list(B1=B1, V1=V1, stdB=stdB, nu1=nu1,
              s21=s21, CovB=CovB, h1=h1, h1var=h1var,
              lmlik = lmlik, HPDI95 = HPDI95, HPDI99 = HPDI99,
              previsao = Yestrela, varprev = Varestrela,
              escalaprev = VYestrela, BMQO = BMQO, stdMQO = stdMQO,
              probpos = probpos, HPDI95prev = HPDI95prev,
              HPDI99prev= HPDI99prev))
}
```

Calculando

Verificando o ajuste

```
errosbayes <- dados_antigos[,1] - (estimativa$B1[1] +
                                     estimativa$B1[2] *
                                     dados_antigos[,3] +
                                     estimativa$B1[3] *
                                     dados_antigos[,4] +
                                     estimativa$B1[4] *
                                     dados_antigos[,5] +
                                     estimativa$B1[5] *
                                     dados_antigos[,6] +
                                     estimativa$B1[6] *
                                     dados_antigos[,7] +
                                     estimativa$B1[7] *
                                     dados_antigos[,8] +
                                     estimativa$B1[8] *
                                     dados_antigos[,9])

errosMQO <- dados_antigos[,1] - (estimativa$BMQO[1] +
                                   estimativa$BMQO[2] *
                                   dados_antigos[,3] +
                                   estimativa$BMQO[3] *
                                   dados_antigos[,4] +
                                   estimativa$BMQO[4] *
                                   dados_antigos[,5] +
                                   estimativa$BMQO[5] *
                                   dados_antigos[,6] +
                                   estimativa$BMQO[6] *
                                   dados_antigos[,7] +
                                   estimativa$BMQO[7] *
                                   dados_antigos[,8]  +
                                   estimativa$BMQO[8] *
                                   dados_antigos[,9])

ggQQ <- function(vetor) # argument: a linear model
{
```

```
    y <- quantile(vetor[!is.na(vetor)], c(0.25, 0.75))
    x <- qnorm(c(0.25, 0.75))
    slope <- diff(y)/diff(x)
    int <- y[1L] - slope * x[1L]
    d <- data.frame(resids = vetor)
    p <- ggplot(d, aes(sample=resids)) +
       stat_qq(alpha = 0.5) +
       geom_abline(slope = slope,
                   intercept = int, color="blue")+
       ggtitle("") +
       labs(title="", y = "Quantis Amostra",
            x= "Quantis Teóricos") +
       theme_bw()

    return(p)
}

ggQQ(errosbayes)

# Altura e largura são em inches, default = 7
pdf(file="C:\\...\\Fig-4.014.pdf", width = 6, height = 3)
ggQQ(errosbayes)
dev.off()

ggQQ(errosMQO)

# Altura e largura são em inches, default = 7
pdf(file="C:\\...\\Fig-4.015.pdf", width = 6, height = 3)
ggQQ(errosMQO)
dev.off()

# Histogramas

d <- data.frame(resids = errosbayes)
p1 <- ggplot(data = d, aes(d[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black",
                 fill = "#dfba56") +
  ggtitle("") +
  labs(title="", y = "No. obs.",
       x= "Resíduos da Regressão Bayesiana") +
  scale_x_continuous(breaks = seq(-3, 3, 0.5),
                     limits = c(-3,3)) +
  scale_colour_brewer(palette="Dark2") +
  theme_bw()

p1

# Altura e largura são em inches, default = 7
pdf(file="C:\\...\\Fig-4.012.pdf", width = 6, height = 3)
p1
dev.off()
```

```r
d <- data.frame(resids = errosMQO)
p1 <- ggplot(data = d, aes(d[,1])) +
  geom_histogram(aes(y = ..count..),
                 alpha = .2, colour = "black",
                 fill = "#9a96bf") +
  ggtitle("") +
  labs(title="", y = "No. obs.",
       x= "Resíduos da Regressão Classica") +
  scale_x_continuous(breaks = seq(-3, 3, 0.5),
                     limits = c(-3,3)) +
  scale_colour_brewer(palette="Dark2") +
  theme_bw()

p1

# Altura e largura são em inches, default = 7
pdf(file="C:\\...\\Fig-4.013.pdf", width = 6, height = 3)
p1
dev.off()
```