

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-Graduação em Estatística

Carla Paula Moreira Soares

**Uma Abordagem via Mistura Finita para
Modelos de Regressão Linear com Erro nas
Variáveis**

Belo Horizonte

2020

Carla Paula Moreira Soares

Uma Abordagem via Mistura Finita para Modelos de Regressão Linear com Erro nas Variáveis

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Estatística

Orientadora: Profa. Dra. Rosângela Helena Loschi

Coorientador: Prof. Dr. Marcos Oliveira Prates

Belo Horizonte

2020

Soares,.Carla Paula Moreira.

S676u Uma abordagem via mistura finita para modelos de regressão linear com erros nas variáveis [manuscrito] / Carla Paula Moreira Soares. - 2020.
56 f. il.

Orientadora: Rosângela Helena Loschi
Coorientador. Marcos Oliveira Prates

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento Estatística.
Referências: f.44-46

1. Estatística – Teses. 2. Erro de medida – Teses. 3. Mistura fina – Teses. 4. Teoria bayesiana de decisão estatística – Teses. 5. Análise de regressão- Teses. I. Loschi,Rosângela Helena. II. Prates, Marcos Oliveira. III.Universidade Federal de Minas Gerais; Instituto de Ciências Exatas, Departamento de Estatística IV.Título.

CDU 519.2(043)



ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO DA ALUNA CARLA PAULA MOREIRA SOARES, MATRICULADA NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 06 DE NOVEMBRO DE 2020.

Aos 06 dias do mês de novembro de 2020, às 14h30, em reunião pública virtual 255 (conforme orientações para a atividade de defesa de dissertação durante a vigência da Portaria PRPG nº 1819) OU na sala 2076 do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de dissertação da aluna Carla Paula Moreira Soares, intitulada: "Uma Abordagem via Mistura Finita para Modelos de Regressão Linear com Erros nas Variáveis", requisito final para obtenção do Grau de Mestre em Estatística. Abrindo a sessão, a Senhora Presidenta da Comissão, Prof(a). Rosângela Helena Loschi, passou a palavra à aluna para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da aluna. Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença da aluna e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

- (x) Aprovada.
- () Reprovada com resubmissão do texto em ____ dias.
- () Reprovada com resubmissão do texto e nova defesa em ____ dias.
- () Reprovada.

Rosângela Helena Loschi

Marcos Oliveira Prates

Cristiano de Carvalho Santos

Manuel Jesus Galea Rojas

O resultado final foi comunicado publicamente à aluna pela Senhora Presidenta da Comissão. Nada mais havendo a tratar, a Presidenta encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 06 de novembro de 2020.

Observações:

1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

FOLHA DE APROVAÇÃO

Uma Abordagem via Mistura Finita para Modelos de Regressão Linear com Erros nas Variáveis

CARLA PAULA MOREIRA SOARES

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 06 de novembro de 2020, pela banca constituída pelos membros:

Prof(a). Rosângela Helena Loschi - Orientador
DEST/UFMG

Prof(a). Marcos Oliveira Prates - Coorientador
DEST/UFMG

Prof(a). Cristiano de Carvalho Santos
DEST/UFMG

Prof(a). Manuel Jesus Galea Rojas
PUC - CHILE

Belo Horizonte, 6 de novembro de 2020.

Agradecimentos

Aos meus orientadores, Rosangela e Marcos, pelas horas de dedicação, aprendizado, paciência, ajuda e compreensão durante todo o período de trabalho juntos.

Agradeço a toda minha família pelo incentivo e por nunca me deixarem desistir.

Ao meu noivo Raphael que sempre me ajudou e acreditou que tudo ia dar certo mesmo nos momentos mais difíceis.

Aos meus colegas do mestrado pelo apoio e amizade.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) agradeço pela concessão da bolsa que permitiu minha dedicação aos estudos e à pesquisa.

O pensamento estatístico um dia será tão necessário para a cidadania eficiente quanto a capacidade de ler e escrever.

- H. G. Wells

Resumo

Em modelos de regressão linear com erros de medida, ocorre com grande frequência de a suposição usual de normalidade para o erro de medida não ser a mais adequada para os dados em questão. Isso pode ser evidenciado em casos que o erro de medida apresenta um comportamento que não coincide com os de diferentes subgrupos da população. O presente trabalho propõe uma distribuição de mistura finita de *skew-normal* com um ponto de massa em zero. Essa distribuição permite uma flexibilidade nos erros acomodando tanto simetria quanto assimetria no mesmo. Para realizar inferência Bayesiana desenvolve-se um algoritmo do tipo *Gibbs* com passo *Metropolis-Hasting*. Para avaliar o desempenho das estimativas apresenta-se um estudo de simulação com diferentes simetrias e assimetrias no erro de medida e também com conjuntos de dados reais.

Palavras-chaves: Erro de medida, Mistura finita, Regressão Linear, Inferência Bayesiana.

Abstract

In linear regression models with measurement errors it is usually common that the assumption of symmetric normal distribution for measurement error is not the most adequate for the data at hand. This can be evidenced in cases where the measurement error presents behavior that does not coincide with those of different population subgroups. This work proposes a finite mixture distribution of skew-normal with a mass point at zero. This distribution allows flexibility in errors, accommodating both symmetry and asymmetry in the same. To carry out Bayesian inference, an algorithm of the type Gibbs with Metropolis-Hasting step is developed. To evaluate the performance of the estimates, a simulation study is presented with different symmetries and asymmetries in the measurement error and applied to a real data set.

Key-words: Measurement Error, Finite Mixture, Linear Regression, Bayesian Inference.

Sumário

1	Introdução	10
2	Modelos com Erro de Medida	13
2.1	Modelos com Erro de Medida Aditivos	13
2.2	Conceitos Importantes	14
2.3	Modelo de Regressão Linear Simples com Erro de Medida	15
3	Uma Abordagem Flexível para o Erro de Medida	18
3.1	Modelo de Regressão Linear com Mistura Finita para o Erro de Medida	18
3.1.1	Suposições sobre o Erro de Medida	20
3.2	Modelo de Regressão Linear Simples com Mistura Finita de <i>Skew-Normal</i> para o Erro de Medida	21
3.2.1	Condições para o Erro de Medida	23
3.2.2	Função de Verossimilhança	24
3.2.3	Distribuições a <i>Priori</i>	25
3.2.4	Distribuição a <i>Posteriori</i>	26
3.2.5	Distribuições Condicionais Completas	27
3.3	CrITÉrios de Seleção de Modelos	31
4	Estudo de Simulação	32
4.1	Efeito das especificações a <i>priori</i> para \mathbf{p} nas estimativas a <i>posteriori</i>	33
4.2	Dados com diferentes assimetrias para o erro de medida	34
5	Aplicação	42
5.1	Banco de dados: rendimento do milho	42
5.2	Banco de dados: gastos ambulatoriais	44
6	Conclusões	46
	Referências	47
A	Função de Verossimilhança	50
A.1	Distribuição Conjunta (y_i, w_i) no Modelo com Erro Normal	50
A.2	Contas da Função de Verossimilhança para o Modelo Proposto	52
B	Distribuição <i>Half-Normal</i>	56
C	Resultados de Simulação	58

1 Introdução

Nos modelos de regressão, em algumas situações ocorre que uma ou mais covariáveis são medidas com erro, nestes casos, o valor exato da covariável não é observado perfeitamente e o que se observa é o valor real acrescido de algum erro de medição. Esses erros decorrem de falhas nos instrumentos de medição; falhas nos métodos e técnica de coletas de dados; por processamento inadequado de dados; falhas dos equipamentos em decorrência de desgaste; pela falta de calibração ou até mesmo pelas condições ambientais. Situações desse tipo são comuns no cotidiano e, para esses casos, devem ser utilizados modelos que considerem em sua estrutura erros de medição. Se uma análise estatística é feita ignorando que a covariável foi mensurada com erro, os resultados podem levar à inferências viesadas (Stefanski, 1985). Isso pode afetar a precisão das estimativas dos parâmetros e levar à perda de poder para detectar as relações entre as variáveis (Carroll et al., 2006).

Suponha que exista apenas uma variável medida com erro e seja w_i seja o valor observável desta variável para o i -ésimo indivíduo. A variável w_i se decompõe em duas partes: a primeira não observável representa o verdadeiro valor da variável (variável latente) x_i e a segunda representa o erro de medida u_i atribuído a variável observada. Na literatura, considera-se duas estruturas para modelar o erro de medida conhecidas por estrutura aditiva e multiplicativa. Neste trabalho, considera-se apenas a estrutura aditiva. O modelo aditivo assume que o valor observável w_i é decomposto na soma da variável latente x_i e do erro de medida u_i , ou seja, $w_i = x_i + u_i$. Uma outra abordagem para o modelo aditivo é o modelo de Berkson (Berkson, 1950), o qual assume que a variável latente x_i é decomposta na soma de w_i e do erro de medida u_i , $x_i = w_i + u_i$.

Há três classes de modelos com erro nas covariáveis: modelos funcional, estrutural e ultraestrutural. A distinção entre os modelos funcional e estrutural foi discutida pela primeira vez em Kendall (1951, 1952). No modelo funcional, a variável não observada x_i é fixada e considerada como uma constante desconhecida. No modelo estrutural, a variável latente x_i é considerada como aleatória e atribuí-se uma distribuição a ela. O modelo ultraestrutural é uma generalização dos dois primeiros, no qual se assume que as covariáveis não observadas são variáveis aleatórias não identicamente distribuídas (Dolby, 1976). Do ponto de vista Bayesiano, modelos funcional e estrutural são equivalentes, uma vez que se atribui uma distribuição *a priori* para todos os elementos desconhecidos.

Conforme apresentado por Sprent (1990), os primeiros estudos sobre os modelos de regressão quando a variável explanatória está sujeita a erros de medição foram apresentados no século XIX por Adcock em (1877, 1878). Desde então, na literatura, existem

diversos trabalhos que tratam de modelos de regressão com erro de medida. Para um estudo mais abrangente sobre os modelos com erros de medida recomendam-se os trabalhos de Kendall e Stuart (1979), Fuller (1987), Cheng e Ness (1999), Gustafson (2003), Carroll et al. (2006), Buonaccorsi (2010); Buzas et al. (2014), Grace (2016), entre outros.

Muitos trabalhos abordando modelos de regressão linear estrutural supõem normalidade tanto para o verdadeiro valor da covariável x_i , quanto para o erro de medida u_i . A suposição de normalidade tende a simplificar a construção da função de verossimilhança e a facilitar o processo de inferência. Porém em algumas situações, a suposição de normalidade não é a mais adequada para descrever o comportamento dos dados. Existem trabalhos envolvendo os modelos de regressão com erros aditivos que fazem uso de outras distribuições para x_i e u_i que se adequam melhores aos dados. Para evitar transformações nos dados, Arellano-Valle et al. (2005) estendem o modelo simétrico normal ao atribuir a distribuição *skew*-normal para a variável latente x_i , para o erro do modelo e para o erro de medição u_i , contendo como caso especial a distribuição normal. Em modelos com erros de medida multivariados, o uso de distribuições de mistura de escala *skew*-normal sob diferentes contextos para modelar a incerteza dos erros e da variável latente podem ser vistos em Lachos et al. (2009), Lachos et al. (2010) e Labra et al. (2011).

Outra possibilidade é considerar misturas finitas de distribuições que possuem grande flexibilidade em descrever dados com diferentes assimetrias, curtoses e pesos nas caudas. Carroll et al. (1999), aborda o problema de heterogeneidade dos dados por meio de misturas de densidade normais. No trabalho de Cabral et al. (2014), no contexto em que tanto a variável resposta quando a variável observada são medidos com erros, eles modelam conjuntamente a covariável não observada e os erro aleatórios por uma mistura de distribuição *t*-student permitindo modelar dados com grandes flexibilidades acomodando assimetria, caudas e multimodalidade.

A suposição de normalidade para os erros de medida também pode se tornar inadequada se a variável mensurada tem um comportamento que difere entre sub-grupos da população, podendo não apresentar erro em algum destes sub-grupos e esse erro ser positivo ou negativo para outros sub-grupos. Como exemplo, considere uma pesquisa de opinião na qual uma das perguntas feitas seja em relação a sua renda mensal. Conforme mencionado por Fuller (1987), a variável renda é habitualmente medida com erro. Para responder o entrevistado muitas vezes pode omitir o fato de ganhar muito respondendo que tem uma renda menor. Em contra partida quem recebe pouco pode responder que ganha mais. Esse tipo de situação pode variar o erro de medida u_i de três formas: a primeira sem erro (quando o entrevistado fala a verdade); de forma positiva (quando o entrevistado responde que ganha um salário maior) e de forma negativa (quando o entrevistado responde que ganha um salário menor).

Motivado por esse tipo de situação, este trabalho propõe um modelo com erro

aditivo o qual modela o comportamento do erro de medida u_i através uma mistura finita de distribuições. A meta é construir um modelo que seja flexível e capaz de acomodar tanto simetria quanto assimetria no erro de medida. Principalmente, objetiva-se construir um modelo que permita identificar o tipo de erro que acomete a covariável mensurada em cada indivíduo: se foi medida sem erro, se seu valor foi medido acima do seu valor real (erro positivo) ou se foi medida abaixo do valor real (erro negativo). Isso é obtido assumindo que a incerteza sobre o erro de medida é descrita por uma mistura finita de distribuições composta por três componentes, a saber, uma com massa pontual em zero responsável por acomodar a mensuração sem erro e duas distribuições contínuas uma assimétrica negativa colocando massa substancial em valores negativos e outra assimétrica positiva colocando massa probabilística substancial em valores positivos. A abordagem Bayesiana é utilizada para fazer inferência sobre os parâmetros. O desempenho do modelo proposto avaliado através de estudo de simulação e da análise de um banco de dados proveniente do livro do Fuller (1987) e de um banco de dados de gastos ambulatoriais.

A organização deste trabalho se encontra da seguinte maneira. No Capítulo 2, apresenta-se uma revisão sobre os principais conceitos e motivação em modelos com erro de medida e do modelo de regressão linear simples com erro de medida. No Capítulo 3 são apresentadas as principais contribuições deste trabalho. Constrói-se o modelo proposto assumindo-se uma mistura finita para o erro de medida e discute-se algumas de suas características. No Capítulo 4 tem-se os resultados dos estudos de simulações para avaliar a eficiência do modelo proposto. No Capítulo 5 tem-se uma aplicação do modelo proposto em duas bases de dados reais. Finalmente, no Capítulo 6 apresenta-se as considerações finais com as principais conclusões obtidas.

2 Modelos com Erro de Medida

Neste capítulo, realiza-se uma breve revisão sobre os modelos com erros de medida, apresentando os tipos de modelos, os principais conceitos assim como as abordagens sobre os erros analisados. Em seguida, apresenta-se o modelo de regressão linear simples com erro de medida aditivo, estratégia que também será utilizada no Capítulo 3.

2.1 Modelos com Erro de Medida Aditivos

Suponha que a covariável de interesse seja medida com erro. Denote por w_i o valor observável e por x_i o valor real não observável desta covariável para a unidade $i = 1, \dots, n$. Desta forma, o verdadeiro valor da variável x_i é uma variável latente. Denote por u_i o erro de medida associado a variável observada para a unidade i .

Os **modelos clássicos** com erros de medida aditivos assumem que a variável observável w_i , $i = 1, 2, \dots, n$, se relaciona linearmente com o valor real x_i da covariável de interesse da seguinte forma:

$$w_i = x_i + u_i, \quad (2.1)$$

em que u_i é a variável aleatória independente e igualmente distribuída (i.i.d) representando o erro de medida com média igual a zero, garantindo que w_i é uma estimativa não viciada de x_i . Nos modelos clássicos assume-se que u_i é independente de x_i e tem-se que a variabilidade na variável latente x_i é menor do que na covariável w_i , uma vez que $Var(x_i) = Var(w_i) - Var(u_i)$. O **modelo de Berkson** aditivo (Berkson, 1950) supõe que a variável latente é decomposta da seguinte maneira:

$$x_i = w_i + u_i, \quad (2.2)$$

para $i = 1, 2, \dots, n$, com u_i independente de w_i . Sob este modelo assume-se que a variável x_i possui mais variabilidade do a variável observável w_i , uma vez que $Var(x_i) = Var(w_i) + Var(u_i)$.

De acordo com Carroll et al. (2006), na prática, a escolha entre os modelos (2.1) e (2.2) é simples. Os erros são do tipo considerado no modelo clássico, se a variável com erro for única para cada indivíduo e especialmente se a medição puder ser replicada, por exemplo, a medição da pressão sanguínea. Agora se para os indivíduos de um grupo ou estratos são dados os mesmos valores da variável propensa a erros, por exemplo, os mineiros em uma mina sujeitos a mesma exposição de poeira mas a verdadeira exposição é particular para cada indivíduo então o mais apropriado é o de Berkson. Uma suposição incorreta dos erros pode causar grandes problemas como ignorar o erro, conforme ressalta Stefanski (2000).

Em outras situações, a relação aditiva pode não ser apropriada. Nestes casos os modelos com erros multiplicativos surgem como uma alternativa. No modelo clássico multiplicativo a decomposição da variável observada é o produto do verdadeiro valor com erro de medida, $w_i = x_i u_i$. No modelo de Berkson multiplicativo tem-se que a decomposição da variável latente é o produto do verdadeiro valor da variável com o erro de medida, $x_i = w_i u_i$. Os modelos multiplicativos não serão considerados neste trabalho.

2.2 Conceitos Importantes

Definida a estrutura dos erros, um ponto relevante a ser discutido em modelos com erros nas covariáveis é o tratamento dado as variáveis latentes. Existem três tipos de abordagens discutidas na literatura:

- **Modelo Funcional:** a variável x_i é fixa e desconhecida. Assim, o número de parâmetros cresce de acordo com o tamanho da amostra. Neste caso, diz-se que x_i são parâmetros incidentais no modelo.
- **Modelo Estrutural:** a quantidade não observada x_i é considerada como uma variável aleatória independente e identicamente distribuída e, ainda, independente do erro de medição.
- **Modelo Ultraestrutural:** é uma generalização dos modelos funcional e estrutural, proposto por Dolby (1976). Assume-se que as variáveis não são indenticamente distribuídas. Considerando-se que as médias e variâncias do verdadeiro valor da covariável x_i são iguais, o modelo ultraestrutural se reduz ao modelo estrutural. Se é assumido que o verdadeiro valor da variância de x_i é nulo retorna-se ao modelo funcional.

No contexto Bayesiano, todos os parâmetros desconhecidos do modelo são tratados como variáveis aleatórias. Assim, dentro desta ótica não há diferenças entre os modelos estrutural e funcional.

Outra diferença importante está entre o **erro de medida diferencial** e o **erro de medida não diferencial**. Considerando um modelo geral com erro de medida em que y_i é a variável resposta diz-se que o erro não-diferencial ocorre quando w_i não contém informação sobre y_i , ou seja, quando a distribuição de y_i dado (w_i, x_i) depende somente de x_i isto é $f(y_i|w_i, x_i) = f(y_i|x_i)$. Caso contrário, se $f(y_i|w_i, x_i) \neq f(y_i|x_i)$ o erro de medida é diferencial. A característica principal da medição não-diferencial é que não precisa de informação adicional para prever y_i além da contida em x_i .

Em várias situações, o erro de medida pode ser considerado como não-diferencial. Segundo Buzas et al. (2014), se o erro de medição se deve exclusivamente ao erro de

análise do instrumento ou do laboratório, pode-se argumentar que o erro é não diferencial. Entretanto, existem situações em que isso não acontece. Carroll et al. (2006) citam uma excessão em relação aos estudos de caso-controle em que o erro geralmente é diferencial, pois nesse caso a covariável observada pode variar no decorrer do tempo e a variável resposta pode ser medida antes. Por exemplo, em um estudo no qual o interesse é o indicador de presença de câncer de mama e um preditor de interesse é a dieta da mulher antes do diagnóstico. Pelas características do estudo, o relato da mulher é feito após o diagnóstico (x_i). Uma mulher que desenvolve câncer de mama deve seguir uma dieta durante o tratamento e, como consequência, a dieta reportada depois do diagnóstico (w_i), assim w_i é claramente correlacionada com a variável resposta (y_i).

2.3 Modelo de Regressão Linear Simples com Erro de Medida

Considere um modelo de regressão linear simples, no qual a variável resposta observada y_i , $i = 1, 2, \dots, n$, esteja associada a uma única covariável x_i e essa seja medida com erro. Assumindo a abordagem de modelo clássico aditivo, tem-se o seguinte modelo:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i, \\ w_i &= x_i + u_i, \end{aligned} \tag{2.3}$$

em que β_0 denota o intercepto, β_1 é o efeito da covariável, x_i e w_i denotam as variáveis explicativa verdadeira (variável latente) e a observada, respectivamente; e_i é o erro do modelo e u_i o erro de mensuração. Assuma que os erros aleatórios e_i e u_i são independentes entre si e também independentes de x_i com $e_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ para $\sigma_e^2 > 0$. Sob o enfoque estrutural assume-se que $x_i \stackrel{iid}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ com $\mu_x \in \mathbb{R}$, $\sigma_x^2 > 0$ e $u_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_u^2)$ com $\sigma_u^2 > 0$. Sob tais supostos,

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{ind}{\sim} \mathcal{N}_3 \left[\begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}; \text{diag}(\sigma_e^2, \sigma_u^2, \sigma_x^2) \right]. \tag{2.4}$$

Um modelo mais geral, classificado como heteroscedástico pode ser construído, assumindo-se que as variâncias dos erros σ_u^2 e σ_e^2 variam de observação para observação. O modelo em (2.4), em que as variâncias são constantes é denominado homoscedástico.

Denote por $\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_e^2)$ os parâmetros de interesse no modelo e assuma que n unidades foram independentemente observadas. Seja $\mathbf{y} = (y_1, \dots, y_n)$, $\mathbf{w} = (w_1, \dots, w_n)$ e $\mathbf{x} = (x_1, \dots, x_n)$ e assuma erro não-diferencial. Das expressões (2.3)

e (2.4) a verossimilhança é obtida através da integração conjunta dos dados, na qual

$$\begin{aligned}
 f(\mathbf{y}, \mathbf{w}) &= \prod_{i=1}^n \int_{\Omega_{x_i}} f(y_i, w_i, x_i) dx_i \\
 &= \prod_{i=1}^n \int_{\Omega_{x_i}} f(y_i, w_i | x_i) f(x_i) dx_i \\
 &= \prod_{i=1}^n \int_{\Omega_{x_i}} f(y_i | x_i, w_i) f(w_i | x_i) f(x_i) dx_i.
 \end{aligned} \tag{2.5}$$

A integral na expressão (2.5) dada pela distribuição conjunta de (w_i, y_i) resulta na seguinte distribuição normal bivariada com vetor de médias e matriz de covariâncias dados por

$$\begin{pmatrix} y_i \\ w_i \end{pmatrix} \underset{ind}{\sim} \mathcal{N}_2 \left[\begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right]. \tag{2.6}$$

Assim, a função de verossimilhança é dada por

$$\begin{aligned}
 f(\mathbf{y}, \mathbf{w}) &= (2\pi)^{-\frac{n}{2}} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{n}{2}} \exp \left\{ \frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[\frac{(y_i - (\beta_0 + \beta_1 \mu_x))^2}{\beta_1^2 \sigma_x^2 + \sigma_e^2} + \right. \right. \\
 &\quad \left. \left. \frac{(w_i - \mu_x)^2}{\sigma_x^2 + \sigma_u^2} - 2\rho \left(\frac{(y_i - (\beta_0 + \beta_1 \mu_x))}{\sqrt{\beta_1^2 \sigma_x^2 + \sigma_e^2}} \right) \left(\frac{(w_i - \mu_x)}{\sqrt{\sigma_x^2 + \sigma_u^2}} \right) \right] \right\},
 \end{aligned} \tag{2.7}$$

em que

$$\rho = \frac{\beta_1 \sigma_x^2}{[(\beta_1^2 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)]^{\frac{1}{2}}}.$$

A distribuição conjunta (2.6) é não identificável e, portanto, não existem estimadores consistentes para β_0 e β_1 (Kendall e Stuart, 1979), uma vez que não há estatísticas suficientes independentes garantindo a estimação de todos os parâmetros desconhecidos. Utilizando o teorema da família exponencial é possível notar que há seis parâmetros $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_e^2, \sigma_u^2)$ a serem estimados e somente cinco estatísticas suficientes dadas por

$$\begin{aligned}
 \bar{w} &= \frac{1}{n} \sum_{i=1}^n w_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad S_{ww} = \frac{1}{n} \sum_{i=1}^n (w_i - \bar{w})^2, \quad S_{yy} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2, \\
 S_{wy} &= \frac{1}{n} \sum_{i=1}^n (w_i - \bar{w})(y_i - \bar{y}).
 \end{aligned}$$

Essas estatísticas $(\bar{w}, \bar{y}, S_{ww}, S_{yy}, S_{wy})$, são estatísticas suficientes para $(\mu_x, \beta_0 + \beta_1 \mu_x, \sigma_x^2 + \sigma_u^2, \beta_1^2 \sigma_x^2 + \sigma_e^2, \beta_1 \sigma_x^2)$. O parâmetro μ_x é o único identificável sem nenhuma suposição adicional, uma vez que \bar{w} é um estimador consistente para μ_x . Assim, para os demais parâmetros surge a necessidade de restrições para reduzir os números de parâmetros para identificar o modelo. Algumas das suposições mais conhecidas conforme listado por Cheng e Ness (1997) e Bolfarine et al. (1992) consideram:

- (i) a razão das variâncias dos erros $\lambda = \sigma_e^2/\sigma_u^2$ conhecida;
- (ii) σ_e^2 conhecida;
- (iii) σ_u^2 conhecida;
- (iv) β_0 conhecido com $\mu_x \neq 0$ e $\beta_1 \neq 0$;
- (v) o fator de atenuação conhecido $k_x = \sigma_x^2/(\sigma_u^2 + \sigma_x^2)$, $k_x \in [0, 1]$.

Todas essas suposições reduzem o número de parâmetros a serem estimados tornando o modelo identificável. A suposição (i) sobre a razão das variâncias dos erros tem sido amplamente utilizada conforme apresentado por Cheng e Ness (1999). Considerar o intercepto conhecido (iv) pode ser adequado em algumas situações conforme são vistos em Chan e Mak (1979), Aoki et al. (2001,2003) e Labra et al. (2011). A suposição (v) também chamada de razão de confiabilidade é usada com frequência nas áreas de Psicologia, Ciências Sociais e Biologia nas quais é possível conhecer previamente o k_x (Fuller, 1987). Essa suposição através da abordagem Bayesiana pode ser vista em Rodrigues e Baba (1994) e Bolfarine e Cordani (1993).

Segundo Bolfarine et al. (1992), outra forma de contornar o problema de identificabilidade do modelo estrutural seria obter réplicas dos valores de pelo menos uma das variáveis o que possibilita outra fonte para estimar as variâncias dos erros. Na ausência de informação sobre a variância do erro, ainda se pode obter estimativas dos parâmetros, se os dados disponíveis contiverem uma variável instrumental. Detalhes podem ser vistos em (Carroll et al. (2006)).

No contexto de inferência clássica, os métodos de estimação mais usados são baseados nas abordagens de máxima verossimilhança e pseudo-verossimilhança. Entre esses se tem o método *naive*, método da máxima verossimilhança e método calibração da regressão. Esses métodos não serão tratados neste trabalho e podem ser encontrados em Fuller (1987).

No próximo capítulo, será introduzido um novo modelo de regressão linear com erro aditivo em que assumimos uma mistura finita de distribuições para modelar a incerteza sobre o erro de medida.

3 Uma Abordagem Flexível para o Erro de Medida

Como mencionado na introdução deste trabalho, há algumas covariáveis usualmente medidas com erro em que o comportamento medido não é o mesmo para diferentes sub-grupos da população. Neste contexto, a suposição de que o erro tem a mesma distribuição para todos os sub-grupos da população não é razoável.

Objetivando apresentar uma resposta para esses tipos de situações, este capítulo propõe uma abordagem mais flexível para os modelos com erro de medição em que modela-se o comportamento do erro de medida através de uma mistura finita de distribuições. Misturas de distribuições tem sido amplamente utilizadas para construir modelos mais robustos capazes de acomodar assimetria, diferentes curtoses e populações heterogêneas. Serão apresentados e discutidos todas as suposições adotadas assim como o processo de inferência utilizado.

3.1 Modelo de Regressão Linear com Mistura Finita para o Erro de Medida

Suponha que uma amostra de tamanho n seja independente e selecionada na população. Seja y_i a variável resposta de interesse avaliada no indivíduo i . Assuma que y_i é linearmente correlacionada com as covariáveis (x_i, \mathbf{b}_i) de tal forma que

$$y_i = \beta_0 + \beta_1 x_i + \boldsymbol{\alpha}^t \mathbf{b}_i + e_i, \quad (3.1)$$

em que $y_i \in \mathbb{R}$ é a variável resposta para a unidade i , $(x_i, \mathbf{b}_i) \in \mathbb{R} \times \mathbb{R}^k$ denotam as variáveis explicativas para a i -ésima unidade, $(\beta_0, \beta_1, \boldsymbol{\alpha})$ são os parâmetros da regressão, $\boldsymbol{\alpha}$ é um vetor $k \times 1$ e e_i denota o erro aleatório para o indivíduo i , $e_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ com $\sigma_e^2 \in \mathbb{R}_+$.

Suponha que as covariáveis \mathbf{b}_i são exatamente mensuradas e que a covariável x_i não é perfeitamente medida, isto é, x_i é o valor real e não observável da covariável. Denote por w_i o valor observado de x_i . Assuma, ainda, que exista uma relação linear com erro aditivo entre x_i e w_i tal que

$$w_i = x_i + u_i, \quad (3.2)$$

em que u_i denota o erro aleatório de medida associado à unidade i . Os erros e_i e u_i são variáveis aleatórias independentes e assumamos que a variável latente x_i é independente dos erros e_i e u_i .

Neste trabalho, adota-se a abordagem vista como estrutural do ponto de vista da estatística clássica de modelos com erros nas covariáveis. Seguindo algumas propostas na literatura, Carroll et al. (2006), modela-se a incerteza sobre a variável latente considerando $x_i \stackrel{iid}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ com $\mu_x \in \mathbb{R}$ e $\sigma_x^2 \in \mathbb{R}_+$. A novidade no modelo proposto está na forma como se modela a incerteza sobre o erro de medida. Admite-se que exista heterogeneidade na população no que diz respeito ao erro de medição das unidades desse. Para acomodar a heterogeneidade modela-se o comportamento do erro u_i através de uma mistura de distribuições, isto é,

$$u_i \stackrel{iid}{\sim} \sum_k p_k \mathcal{F}_{ik}(\cdot), \quad (3.3)$$

em que $\mathcal{F}_{ik}(\cdot)$ é uma medida de probabilidade com função de densidade de probabilidade (f.d.p) representada por $f_{ik}(\cdot)$ e $p_k \in (0, 1)$ denota os pesos dos componentes $k \in \mathbb{R}_+$ da mistura tal que $\sum_k p_k = 1$.

No contexto deste estudo particular, assume-se que a variável w_i é mensurada sem erros para parte das unidades e que sua distribuição possui assimetria positiva ou negativa para outros sub-grupos. A fim de construir um modelo mais robusto incorporando essas características que permitam acomodar possível heterogeneidade entre os estados do erro de medida, podendo ser simétrico e assimétrico, a mistura proposta na expressão (3.3), é representada pela seguinte estrutura hierárquica:

$$\begin{aligned} u_i | \mathbf{Z}_i, M_{ik} &\stackrel{D}{=} Z_{i0}M_{i0} + Z_{i1}M_{i1} + Z_{i2}M_{i2}, \\ \mathbf{Z}_i | \mathbf{p} &\stackrel{iid}{\sim} \mathcal{M}(1, p_0, p_1, p_2), \\ M_{i0} &\stackrel{iid}{\sim} \mathcal{F}_0(\cdot), \\ M_{i1} &\stackrel{iid}{\sim} \mathcal{F}_1(\cdot), \\ M_{i2} &\stackrel{iid}{\sim} \mathcal{F}_2(\cdot), \end{aligned} \quad (3.4)$$

em que M_{i0} , M_{i1} e M_{i2} são variáveis aleatórias independentes; \mathcal{F}_0 denota uma massa pontual em zero, \mathcal{F}_1 uma medida de probabilidade com assimetria positiva; \mathcal{F}_2 uma medida de probabilidade com assimetria negativa; $\mathcal{M}(\cdot)$ denota a distribuição multinomial; $\mathbf{Z}_i = (Z_{i0}, Z_{i1}, Z_{i2})$ é um vetor tridimensional indicador do grupo da população ao qual unidade i pertence em que Z_{ik} se a i -ésima unidade pertence ao grupo k e $Z_{ik} = 0$ caso contrário; $\mathbf{p} = (p_0, p_1, p_2)$ denota o vetor dos pesos componentes da mistura tal que $p_k \geq 0$ e $\sum_{k=0}^2 p_k = 1$. Se \mathcal{F}_1 tem suporte positivo e \mathcal{F}_2 tem suporte negativo tem-se, como caso particular, a estrutura em que a população é subdividida em um grupo em que x_i é perfeitamente medida, grupo em que o valor observado de x_i é abaixo do valor real e em que o valor observado de x_i é acima do valor real.

Seja $f_k(u_i | \gamma_k)$ funções de densidade de probabilidade de $\mathcal{F}_i(\cdot)$ em que γ_k é o vetor de parâmetros para tais distribuição, para $k = 0, 1, 2$. Sob tais supostos, a distribuição de

u_i é uma mistura de distribuições, uma vez que,

$$\begin{aligned} f(u_i) &= \sum_{\mathbf{Z}} f(u_i|\mathbf{Z}_i, \gamma_k) f(\mathbf{Z}_i|\mathbf{p}) \\ &= \sum_{k=0}^2 p_k f_k(u_i|\gamma_k). \end{aligned}$$

3.1.1 Suposições sobre o Erro de Medida

Ao assumir que w_i deve ser uma estimativa não viciada de x_i , espera-se que o erro de medição u_i tenha média igual a zero. Esta condição impõe certas restrições no modelo proposto como será visto a seguir. Da representação hierárquica em (3.4), a esperança de u_i é dada por

$$\begin{aligned} E(u_i) &= E[E(u_i|\mathbf{Z}_i, M_k)] \\ &= E[Z_{i0}E(M_0|\gamma_0) + Z_{i1}E(M_1|\gamma_1) + Z_{i2}E(M_2|\gamma_2)] \\ &= E(Z_{i0})E(M_0|\gamma_0) + E(Z_{i1})E(M_1|\gamma_1) + E(Z_{i2})E(M_2|\gamma_2) \\ &= p_0E(M_0|\gamma_0) + p_1E(M_1|\gamma_1) + p_2E(M_2|\gamma_2). \end{aligned}$$

Para que o erro de medida tenha média igual a zero, ou seja, $E(u_i) = 0$, deve-se garantir a seguinte relação

$$-p_2E(M_2|\gamma_2) = p_0E(M_0|\gamma_0) + p_1E(M_1|\gamma_1). \quad (3.5)$$

Para o caso particular em que \mathcal{F}_0 é uma massa pontual em zero, deve-se ter $-p_2E(M_2|\gamma_2) = p_1E(M_1|\gamma_1)$. Usando propriedade da variância condicional, a variância do erro de medida é obtida de como se segue:

$$\begin{aligned} Var(u_i) &= Var[E(u_i|\mathbf{Z}_i, \gamma_k)] + E[Var(u_i|\mathbf{Z}_i|\gamma_k)] \\ &= Var(Z_{i0})[E(M_0|\gamma_0)]^2 + Var(Z_{i1})[E(M_1|\gamma_1)]^2 + Var(Z_{i2})[E(M_2|\gamma_2)]^2 + \\ &\quad E(Z_{i0}^2)Var(M_0|\gamma_0) + E(Z_{i1}^2)Var(M_1|\gamma_1) + E(Z_{i2}^2)Var(M_2|\gamma_2) - \\ &\quad 2 \sum_{l=0}^1 \sum_{k=l+1}^2 Cov(Z_{il}M_{il}, Z_{ik}M_{ik}|\mathbf{Z}_i, \gamma_k) \end{aligned}$$

Para o caso em que \mathcal{F}_0 é uma massa pontual em zero, tem-se que

$$\begin{aligned} Var(u_i) &= p_1(1 - p_1)E(M_1|\gamma_1)^2 + p_2(1 - p_2)E(M_2|\gamma_2)^2 + p_1Var(M_1|\gamma_1) + \\ &\quad p_2Var(M_2|\gamma_2) - 2p_1p_2E(M_1|\gamma_1)E(M_2|\gamma_2). \end{aligned} \quad (3.6)$$

Até o momento, não foram especificados quaisquer funções de densidades de probabilidade para $\mathcal{F}_k(\cdot)$, $k = 0, 1, 2$. A estrutura apresentada em (3.4), permite a escolha de diferentes distribuições como componentes da mistura. Para avaliar a performance do

modelo proposto, na seção subsequente descreve-se a inferência com essas distribuições pertencentes à família de distribuições *skew-normal* (Azzalini (1985)). Essa escolha pode também ser feita como exemplo pela distribuição log-normal ou pela distribuição gama. O motivo da escolha da distribuição *skew-normal* é para garantir conjugabilidade nas distribuições condicionais completas que facilitam na implementação do modelo. O processo de inferência é construído considerando o modelo de regressão linear simples através da abordagem Bayesiana.

3.2 Modelo de Regressão Linear Simples com Mistura Finita de *Skew-Normal* para o Erro de Medida

Para simplificar, admita que o modelo apresentado em (3.1), possua apenas uma única covariável e que essa seja mensurada com erro. De modo a garantir que o erro de medida apresente simetria, assimetria positiva e negativa, assume-se a estrutura em (3.4), a distribuição de \mathcal{F}_0 representada por uma massa pontual em zero denotada por $\delta(0)$, as distribuições de \mathcal{F}_1 e \mathcal{F}_2 dadas pela distribuição *skew-normal* com parâmetro de locação $\mu \in \mathbb{R}$, parâmetro de escala $\sigma > 0$ e parâmetro de assimetria $\lambda \in \mathbb{R}$ denotadas por $M_1 \sim \mathcal{SN}(\mu_1, \sigma_1, \lambda_1)$ e $M_2 \sim \mathcal{SN}(\mu_2, \sigma_2, \lambda_2)$, respectivamente.

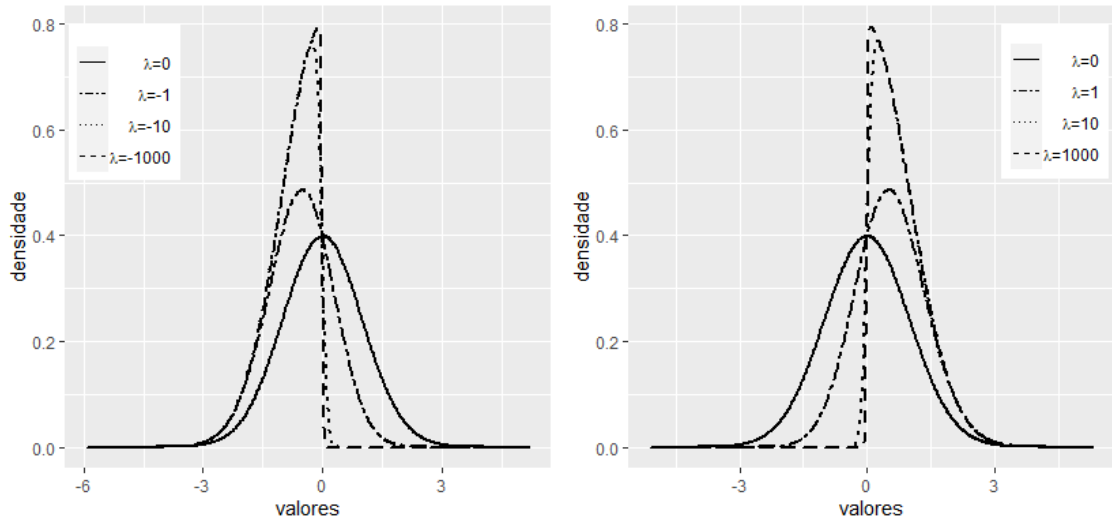
Dessa forma, assume-se a distribuição de u_i é hierarquicamente dada por

$$\begin{aligned}
 u_i | \mathbf{Z}_i, M_k &\stackrel{D}{=} Z_{i0}M_0 + Z_{i1}M_1 + Z_{i2}M_2, \\
 \mathbf{Z}_i | \mathbf{p} &\stackrel{iid}{\sim} \mathcal{M}(1, p_0, p_1, p_2), \\
 M_{i0} &\stackrel{iid}{\sim} \delta(0), \\
 M_{i1} &\stackrel{iid}{\sim} \mathcal{SN}(\mu_1, \sigma_1^2, \lambda_1), \\
 M_{i2} &\stackrel{iid}{\sim} \mathcal{SN}(\mu_2, \sigma_2^2, \lambda_2),
 \end{aligned} \tag{3.7}$$

em que $\lambda_1 \in \mathbb{R}_+$ e $\lambda_2 \in \mathbb{R}_-$. O modelo em (3.7) permite acomodar cenários que envolvem grupos de unidades que não experimentar erro de medida. A Figura 1, mostra que dependendo do valor atribuído a λ a distribuição *skew-normal* possui diferentes graus de assimetria. Para $\lambda_k = 0$ a distribuição torna-se a distribuição normal e quanto maior o valor de λ , maior é o grau de assimetria da distribuição a qual, para valores negativos de λ , é assimetria negativa e, para valores positivos de λ é assimetria positiva.

A distribuição *skew-normal* possui algumas propriedades importantes, estudadas por Azzalini (2008), entre elas tem-se que se $M_{ik} \sim \mathcal{SN}(\mu_k, \sigma_k, \lambda_k)$, $\lambda_k \rightarrow \infty$ então M_{ik} tende em distribuição para uma distribuição *half-normal* com parâmetros de locação $\mu_k \in \mathbb{R}$ e escala $\sigma_k > 0$, denotada por $\mathcal{HN}(\mu_k, \sigma_k^2)$. Assim, para garantir que a distribuição de M_{i1} e M_{i2} tenha suporte positivo (negativo) basta assumir que $\lambda_1 \rightarrow \infty$ ($\lambda_2 \rightarrow -\infty$). Essa escolha de distribuições facilita para obtenção de contas fechadas. Detalhes sobre a distribuição *skew-normal* e *half-normal* podem ser vistas no Apêndice B.

Figura 1 – Densidades da distribuição *skew-normal* com diferentes graus de assimetria.



No caso em que para se modelar o comportamento do o erro de medida considera-se que M_{i1} e M_{i2} são distribuições *Half-normal*, os parâmetros de locação das distribuições de M_{i1} e M_{i2} permitem que somente os valores acima de μ_1 sejam considerados positivos e que somente os valores abaixo de μ_2 sejam considerados negativos, assim os parâmetros de locações podem ser interpretados como um limite de tolerância para que a variável seja considerada como sendo medida com e sem erro, uma vez que os valores dentro do intervalo entre eles são considerados como sem erro.

Nas Figuras 2 e 3 são mostrados alguns exemplos para a distribuição do erro. Na Figura 2 assume-se que $M_{i1} \sim \mathcal{HN}(0, 1; 1, 5)$ e $M_{i2} \sim \mathcal{HN}(-0, 1; 1, 5)$ e diferentes configurações para os pesos da mistura são assumidos. Na Figura 3, assume-se que (a) $M_{i1} \sim \mathcal{HN}(0, 1; 0, 2)$ e $M_{i2} \sim \mathcal{HN}(-0, 1; 8)$ e (b) $M_{i1} \sim \mathcal{HN}(0, 1; 8)$ e $M_{i2} \sim \mathcal{HN}(-0, 1; 0, 2)$, com diferentes configurações dos pesos da mistura.

Figura 2 – Exemplos de funções de densidades da distribuição do erro de medida com diferentes tipos simetrias.

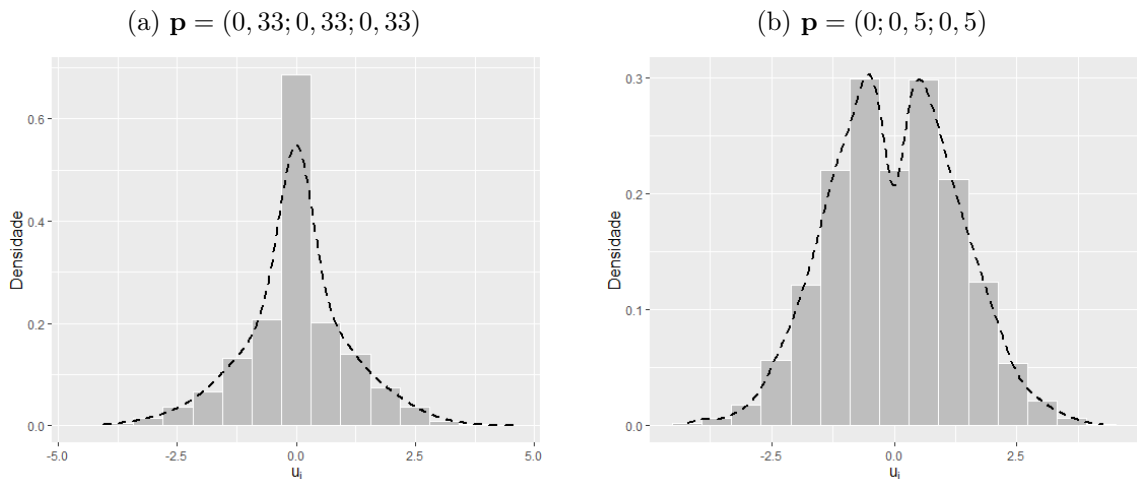
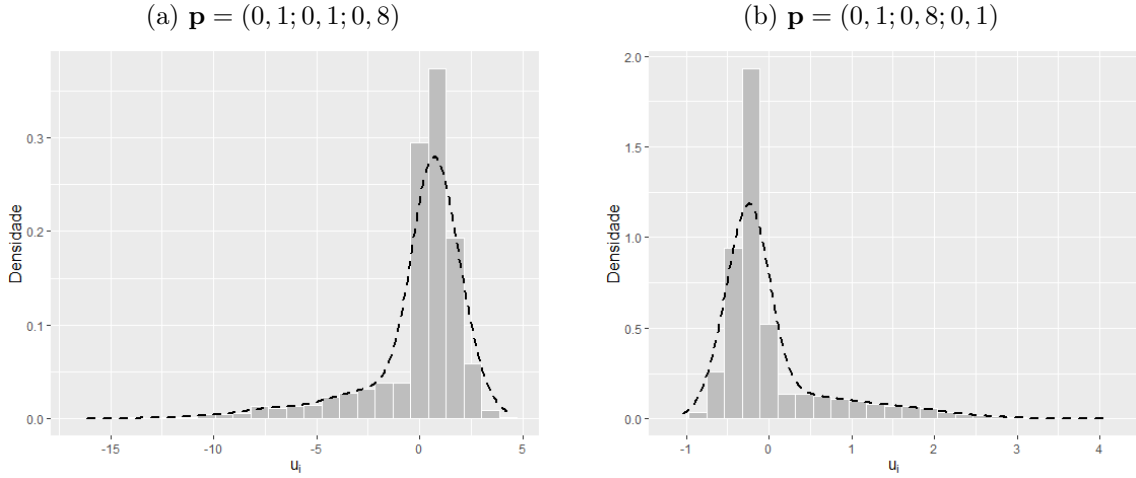


Figura 3 – Exemplos de funções de densidades da distribuição do erro de medida com diferentes tipos assimetrias.



Na próxima subseção apresenta-se algumas restrições a serem consideradas sobre o erro do modelo proposto quando se assume a distribuição *half*-normal para M_{i1} e M_{i2} .

3.2.1 Condições para o Erro de Medida

Para que w_i seja uma estimativa não viciada de u_i tem-se que garantir que o valor esperado de u_i seja igual a zero. Assumindo que $M_{ik} \sim \mathcal{HN}(\mu_k, \sigma_k^2)$, $k = 1, 2$, a partir da equação (3.5), e uma vez que $E(M_0|\sigma_0^2) = 0$, tem-se que

$$-p_2 E(M_2|\mu_2, \sigma_2^2) = p_1 E(M_1|\mu_1, \sigma_1^2). \quad (3.8)$$

Diferentes restrições são impostas no modelo dependendo dos valores assumidos para os pesos p_1 e p_2 em (3.8). Verifica-se as seguintes condições:

1. Se $p_1 = 0$ e $p_2 \neq 0$, tem-se que $-p_2 E(M_2|\mu_2, \sigma_2^2) = 0 \Rightarrow$ a média de M_2 tem que ser zero.
2. Se $p_1 \neq 0$ e $p_2 = 0$, tem-se que $p_1 E(M_1|\mu_1, \sigma_1^2) = 0 \Rightarrow$ a média de M_1 tem que ser zero.
3. Se $p_1 = 0$ e $p_2 = 0$, $E(u_i) = 0$ e $V(u_i) = 0$.
4. Se $p_1 \neq 0$ e $p_2 \neq 0$, tem-se que

$$\begin{aligned} -E(M_2|\mu_2, \sigma_2^2) &= \frac{p_1}{p_2} E(M_1|\mu_1, \sigma_1^2) \\ -\left(\mu_2 + \sigma_2 \sqrt{\frac{2}{\pi}}\right) &= \frac{p_1}{p_2} \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}}\right). \end{aligned} \quad (3.9)$$

Assim, o parâmetro de escala σ_2 é dado por

$$\sigma_2 = \frac{\frac{p_1}{p_2} \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}} \right) + \mu_2}{\sqrt{\frac{2}{\pi}}}. \quad (3.10)$$

Como o parâmetro de escala σ_2 precisa ser maior que zero, segue de (3.10) que a seguinte restrição tem que ser válida,

$$\mu_2 < -\frac{p_1}{p_2} \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}} \right).$$

Nesse caso, assumindo que \mathcal{F}_0 é uma massa pontual em zero e $M_{ik} \sim \mathcal{HN}(\mu_k, \sigma_k^2)$, $k = 1, 2$, a variância do erro de medida (3.6) torna-se

$$\begin{aligned} V(u_i) = & p_1(1-p_1) \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}} \right)^2 + p_2(1-p_2) \left(\mu_2 + \sigma_2 \sqrt{\frac{2}{\pi}} \right)^2 + p_1 \sigma_1^2 \left(1 - \frac{2}{\pi} \right) + \\ & p_2 \sigma_2^2 \left(1 - \frac{2}{\pi} \right) - 2p_1 p_2 \left(\mu_1 + \sigma_1 \sqrt{\frac{2}{\pi}} \right) \left(\mu_2 + \sigma_2 \sqrt{\frac{2}{\pi}} \right). \end{aligned} \quad (3.11)$$

A inferência realizada considera apenas o caso 4 no qual o erro de medida possui diferentes valores de simetria e assimetria. Para esse caso, quando os valores de p_1 e p_2 forem próximos dos limites de zero alguns cuidados são necessários ao analisar os resultados, uma vez que:

- Se $\lim p_1 \rightarrow 0 \Rightarrow \sigma_2^2 \rightarrow (\mu_2 / \sqrt{2/\pi})^2$;
- Se $\lim p_2 \rightarrow 0 \Rightarrow \sigma_2^2$ é indefinido;
- Se $\lim p_1 \rightarrow 0$ e $\lim p_2 \rightarrow 0 \Rightarrow \sigma_2^2$ é indefinido e $V(u_i) = 0$. Vale ressaltar um cuidado necessário ao observar esse tipo de comportamento. Esse modelo é um indicativo de um modelo sem erro de medida.

Nas próximas subseções tem-se a descrição do processo de inferência Bayesiana adotado.

3.2.2 Função de Verossimilhança

Assuma que n unidades foram independentemente observadas. Seja $\mathbf{y} = (y_1, \dots, y_n)$, $\mathbf{w} = (w_1, \dots, w_n)$, $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{u} = (u_1, \dots, u_n)$. Conforme apresentado em (2.5), para encontrar a função de verossimilhança é necessário determinar a fdp dada por $f(w_i|x_i)$. Como a distribuição do erro de medida é uma mistura finita a função de distribuição acumulada $F(w_i|x_i)$ é dada por

$$F(w_i|x_i) = P(w_i \leq w_i|x_i) = \int_{-\infty}^{w_i-x_i} [p_0 f_{i1}(t|\gamma_0) + p_1 f_{i1}(t|\gamma_1) + p_2 f_{i2}(t|\gamma_2)] dt.$$

O fato de $f(y_i, w_i)$ também ser uma mistura finita faz com que a distribuição de (y_i, w_i) dado x_i também seja uma mistura finita com uma estrutura mais complexa que leva a uma maior dificuldade na implementação computacional do modelo. Para contornar este problema a função de verossimilhança é construída considerando a distribuição de w_i condicionada em u_i de forma que $w_i|u_i \stackrel{ind}{\sim} \mathcal{N}(\mu_x + u_i, \sigma_x^2)$. Considerando esta abordagem o modelo é reescrito considerando que a variável latente x_i é dada pela diferença entre a variável observada menos o erro de medição, isto é, $x_i = w_i - u_i$. Assumindo esta estrutura o modelo é hierarquicamente representado por

$$\begin{aligned} y_i|w_i, u_i, \theta^* &\stackrel{ind}{\sim} \mathcal{N}(\beta_0 + \beta_1(w_i - u_i), \sigma_e^2), \\ w_i|u_i, \mu_x, \sigma_x^2 &\stackrel{ind}{\sim} \mathcal{N}(\mu_x + u_i, \sigma_x^2), \\ u_i|p_k, \gamma_k &\stackrel{iid}{\sim} \sum_{k=0}^2 p_k f_k(u_i|\gamma_k), \end{aligned} \quad (3.12)$$

em que $i = 1, \dots, n$ e $\theta^* = (\beta, \sigma_e^2, \mu_x, \sigma_x^2, p_0, p_1, p_2, \gamma_k)$. Assim, a função de verossimilhança pode ser obtida através da integração da função conjunta dos dados completos em relação a u_i , de modo que,

$$\begin{aligned} f(\mathbf{y}, \mathbf{w}) &= \prod_{i=1}^n \int f(y_i|w_i, u_i) f(w_i|u_i) f(u_i) du_i \\ &= \prod_{i=1}^n \int f(y_i|w_i, u_i) f(w_i|u_i) \sum_{k=0}^2 p_k f_k(u_i|\gamma_k) du_i \\ &= \prod_{i=1}^n \sum_{k=0}^2 p_k \int f(y_i|w_i, u_i) f(w_i|u_i) f_k(u_i|\gamma_k) du_i. \end{aligned} \quad (3.13)$$

Encontrar a função de verossimilhança integrando em u_i ao invés de integrar em x_i não altera nos resultados e proporciona a vantagem de facilitar na interpretação do erro de medida. Detalhes das contas usando esta abordagem para o modelo de regressão linear simples com o erro de medida distribuído normalmente podem ser vistos na Seção A.1 do Apêndice A. As contas detalhadas para $f(\mathbf{y}, \mathbf{w})$ usando a abordagem proposta em (3.7) são encontradas na Seção A.2 do Apêndice A.

Conforme discutido no Capítulo 2, o modelo de regressão linear com erro nas co-variáveis apresentam problemas de falta de identificabilidade, uma vez que a distribuição conjunta $f(y_i, w_i)$ não possui estatísticas suficientes independentes em número suficiente para garantir a estimação de todos os parâmetros do modelo. Para garantir a identificabilidade do modelo os parâmetros de locação μ_1 e μ_2 serão considerados conhecidos e nas simulações mais de uma restrição será adotada.

3.2.3 Distribuições a *Priori*

Em estatística Bayesiana a inferência sobre o parâmetro de interesse é realizada através da distribuição a *posteriori*. A distribuição a *posteriori* é proporcional ao produto

da distribuição a *priori* para as quantidades desconhecidas com a função de verossimilhança. Para encontrar a distribuição a *posteriori* primeiro é preciso definir as distribuições a *priori* para cada parâmetro desconhecido do modelo. Seguindo escolhas comumente feitas na literatura assume-se as seguintes distribuições a *priori*: $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, $\mu_x \sim \mathcal{N}(\mu_{xx}, \sigma_{xx}^2)$, $\sigma_e^2 \sim \mathcal{GI}(a_e, b_e)$, $\sigma_x^2 \sim \mathcal{GI}(a_x, b_x)$, $\sigma_1^2 \sim \mathcal{GI}(a_1, b_1)$, onde $\mathcal{GI}(a, b)$ representa a distribuição Gama Inversa, com parâmetros $a > 0$ e $b > 0$ e cuja função de densidade é

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp\left\{-\frac{b}{\sigma^2}\right\}; \quad \sigma^2 > 0, \quad a > 0 \text{ e } b > 0. \quad (3.14)$$

Neste caso, a média e variância dadas, respectivamente, por

$$E[\sigma^2] = \frac{b}{a-1}, \quad a > 1 \quad \text{e} \quad \text{Var}[\sigma^2] = \frac{b^2}{(a-1)^2(a-2)}, \quad a > 2.$$

Distribuições vagas ou pouco informativas para σ^2 são obtidas assumindo $a \rightarrow 0$ e $b \rightarrow 0$. Para descrever a incerteza a *priori* atribuída aos pesos do modelo $\mathbf{p} = (p_0, p_1, p_2)$ como habitualmente feito em modelos de mistura, considera-se a distribuição Dirichlet cuja função de densidade de probabilidade

$$\pi(\mathbf{p}) = \frac{1}{\text{Beta}(\boldsymbol{\alpha})} \prod_{k=0}^2 p_k^{\alpha_k-1}; \quad 0 < p_k < 1, \quad (3.15)$$

em que, $\text{Beta}(\boldsymbol{\alpha}) = \prod_{k=0}^2 \Gamma(\alpha_k) / \Gamma(\alpha_T)$, $\sum_{k=0}^2 p_k = 1$ e $\alpha_k > 0$, $k = 0, 1, 2$ e $\alpha_T = \sum_{k=0}^2 \alpha_k$. Sob a distribuição (3.15) a média e a variância marginais de p_k são dadas, respectivamente, por

$$E[p_k] = \frac{\alpha_k}{\alpha_T} \quad \text{e} \quad \text{Var}[p_k] = \frac{\alpha_k(\alpha_T - \alpha_k)}{\alpha_T^2(\alpha_T + 1)}.$$

3.2.4 Distribuição a *Posteriori*

Considerando as distribuições a *priori* para os parâmetros de interesse dadas na Seção (3.2.3) e a função de verossimilhança dada em (3.13), obtém-se que a distribuição a *posteriori* para $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{p}, \sigma_1^2)$ é

$$\pi(\boldsymbol{\theta} | \mathbf{y}, \mathbf{w}) \propto f(\mathbf{y}, \mathbf{w}) \pi(\boldsymbol{\beta}) \pi(\sigma_e^2) \pi(\mu_x) \pi(\sigma_x^2) \pi(\mathbf{p}) \pi(\sigma_1^2). \quad (3.16)$$

em que $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\mathbf{Z}_i = (Z_{i0}, Z_{i1}, Z_{i2})$ e $\mathbf{p} = (p_0, p_1, p_2)$.

A distribuição a *posteriori* (3.16), não é uma distribuição de probabilidade conhecida. Assim, a inferência será realizada via MCMC (*Markov Chain Monte Carlo*), mais especificamente, via Gibbs Sampling com passo Metropolis-Hasting. Para a implementação do amostrador de Gibbs, é necessário encontrar as distribuições condicionais completas a *posteriori* de todos os parâmetros desconhecidos. Na próxima seção deriva-se

tais distribuições assumindo-se em (3.7) que $M_{ik} \sim \mathcal{HN}(\mu_k, \sigma_k^2)$, $k = 1, 2$. Como M_{i0} é uma massa pontual em zero e devido a problemas computacionais em que a mistura contém distribuições contínuas com uma distribuição discreta, a distribuição de M_{i0} será representada pela distribuição normal com média zero e variância $\sigma_0^2 \in \mathbb{R}_+$. Conforme mencionado os valores de μ_1 e μ_2 determinam uma tolerância para que uma variável seja considerada como medida com erro, uma vez que os valores u_i acima de μ_1 são positivos, abaixo de μ_2 negativos e dentro de seu intervalo são considerados como nulos. Assim, a distribuição normal é escolhida de tal forma que o valor de σ_0^2 fique próximo dos valores de μ_1 e μ_2 .

3.2.5 Distribuições Condicionais Completas

Para obter-se as distribuições condicionais completas a *posteriori*, considera-se a densidade conjunta de todos os componentes do modelo, representada por

$$\begin{aligned} \pi(\mathbf{y}, \mathbf{w}, \mathbf{u}, \mathbf{Z}_i, \boldsymbol{\theta}) &\propto \pi(\mathbf{y}|\mathbf{w}, \mathbf{u}, \boldsymbol{\beta}, \sigma_e^2) \pi(\mathbf{w}|\mathbf{u}, \mu_x, \sigma_x^2) \pi(\mathbf{u}|\mathbf{Z}_i, \gamma_k) \pi(\mathbf{Z}_i|\mathbf{p}) \times \\ &\pi(\boldsymbol{\beta}) \pi(\sigma_e^2) \pi(\mu_x) \pi(\sigma_x^2) \pi(\mathbf{p}) \pi(\sigma_1^2), \end{aligned} \quad (3.17)$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{p}, \sigma_1^2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $\mathbf{Z}_i = (Z_{i0}, Z_{i1}, Z_{i2})$ e $\mathbf{p} = (p_0, p_1, p_2)$. Com as distribuições condicionais completas a *posteriori* determinadas, é possível obter-se amostras utilizando técnicas de simulação MCMC. Algumas distribuições condicionais completas a *posteriori* não apresentam forma fechada e serão estimadas via algoritmo de Metropolis-Hastings. Mais detalhes sobre algoritmo do tipo MCMC podem por exemplo, ser encontrados em Gamerman e Lopes (2006) e Bolstad e Curran (2016).

Distribuição Condicional Completa para $\boldsymbol{\beta}$: A distribuição condicional completa de $\boldsymbol{\beta}$ é proporcional a:

$$\begin{aligned} \pi(\boldsymbol{\beta}|\cdot) &\propto \pi(\mathbf{y}|\mathbf{w}, \mathbf{u}, \boldsymbol{\beta}, \sigma_e^2) \pi(\boldsymbol{\beta}) \\ &\propto \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1(w_i - u_i)))^2 \right\} \exp \left\{ -\frac{1}{2\sigma_\beta^2} (\boldsymbol{\beta}^T \boldsymbol{\beta} - 2\boldsymbol{\mu}_\beta^T \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}) \right\}. \end{aligned}$$

Defina

$$\Delta = \sigma_e^2 \boldsymbol{\Sigma}_\beta^{-1}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad e \quad \mathbf{d} = \begin{bmatrix} 1 & w_1 - u_1 \\ 1 & w_2 - u_2 \\ \vdots & \vdots \\ 1 & w_n - u_n \end{bmatrix}.$$

Então a distribuição condicional completa de $\boldsymbol{\beta}$ é proporcional a:

$$\pi(\boldsymbol{\beta}|\cdot) \propto \exp \left\{ -\frac{1}{2\sigma_e^2} [\boldsymbol{\beta}^T (\mathbf{d}^T \mathbf{d} + \Delta) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T (\mathbf{d}^T \mathbf{y} + \Delta \boldsymbol{\mu}_\beta)] \right\}.$$

ou seja, tem-se

$$\boldsymbol{\beta}|\cdot \sim \mathcal{N}_2((\mathbf{d}^T \mathbf{d} + \Delta)^{-1} (\mathbf{d}^T \mathbf{y} + \Delta \boldsymbol{\mu}_\beta), \sigma_e^2 (\mathbf{d}^T \mathbf{d} + \Delta)^{-1}).$$

Distribuição Condicional Completa para σ_e^2 : A distribuição condicional completa de σ_e^2 é proporcional a:

$$\begin{aligned} \pi(\sigma_e^2|\cdot) &\propto \pi(\mathbf{y}|\mathbf{w}, \mathbf{u}, \boldsymbol{\beta}, \sigma_e^2)\pi(\sigma_e^2) \\ &\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1(w_i - u_i)))^2\right\} \frac{b_e^{a_e}}{\Gamma(a_e)} (\sigma_e^2)^{-a_e-1} \exp\left\{-\frac{b_e}{\sigma_e^2}\right\} \\ &\propto (\sigma_e^2)^{-(a_e+\frac{n}{2}+1)} \exp\left\{-\frac{b_e + \frac{1}{2}(\mathbf{y} - \mathbf{d}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{d}\boldsymbol{\beta})}{\sigma_e^2}\right\}, \end{aligned}$$

o que implica,

$$\sigma_e^2|\cdot \sim \mathcal{GI}\left(a_e + \frac{n}{2}, b_e + \frac{1}{2}(\mathbf{y} - \mathbf{d}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{d}\boldsymbol{\beta})\right).$$

Distribuição Condicional Completa para μ_x : A distribuição condicional completa de μ_x é proporcional a:

$$\begin{aligned} \pi(\mu_x|\cdot) &\propto \pi(\mathbf{w}|\mathbf{u}, \mu_x, \sigma_x^2)\pi(\mu_x) \\ &\propto \exp\left\{-\frac{1}{2\sigma_x^2} \sum_{i=1}^n (w_i - u_i - \mu_x)^2\right\} \exp\left\{-\frac{1}{2\sigma_\mu^2} \mu_x^2\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (w_i - u_i - \mu_x)^2}{\sigma_x^2} + \frac{\mu_x^2}{\sigma_\mu^2} \right]\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left[\mu_x^2(n\sigma_x^{-2} + \sigma_\mu^{-2}) - 2\mu_x \left(\sigma_x^{-2} \sum_{i=1}^n (w_i - u_i) \right) \right]\right\}. \end{aligned}$$

Dessa forma, a distribuição condicional completa de μ_x tem forma fechada conhecida e tem-se que

$$\mu_x|\cdot \sim \mathcal{N}\left(\sigma_x^{-2} \sum_{i=1}^n (w_i - u_i)(n\sigma_x^{-2} + \sigma_\mu^{-2})^{-1}, (n\sigma_x^{-2} + \sigma_\mu^{-2})^{-1}\right).$$

Distribuição Condicional Completa para σ_x^2 : A distribuição condicional completa de σ_x^2 é proporcional a:

$$\begin{aligned} \pi(\sigma_x^2|\cdot) &\propto \pi(\mathbf{w}|\mathbf{u}, \mu_x, \sigma_x^2)\pi(\sigma_x^2) \\ &\propto (\sigma_x^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_x^2} \sum_{i=1}^n (w_i - u_i - \mu_x)^2\right\} \frac{b_x^{a_x}}{\Gamma(a_x)} (\sigma_x^2)^{-a_x-1} \exp\left\{-\frac{b_x}{\sigma_x^2}\right\} \\ &\propto (\sigma_x^2)^{-(a_x+\frac{n}{2}+1)} \exp\left\{-\frac{b_x + \frac{1}{2} \sum_{i=1}^n (w_i - u_i - \mu_x)^2}{\sigma_x^2}\right\}, \end{aligned}$$

consequentemente, tem-se que

$$\sigma_x^2|\cdot \sim \mathcal{GI}\left(a_x + \frac{n}{2}, b_x + \frac{1}{2} \sum_{i=1}^n (w_i - u_i - \mu_x)^2\right).$$

Distribuição Condicional Completa para \mathbf{p} : A distribuição condicional completa de $\mathbf{p} = (p_0, p_1, p_2)$ é proporcional a:

$$\begin{aligned} \pi(\mathbf{p}|\cdot) &\propto \pi(\mathbf{u}|\mathbf{Z}_i, \gamma_k)\pi(\mathbf{Z}_i|\mathbf{p})\pi(\mathbf{p}) \\ &\propto p_0^{\alpha_0-1} p_1^{\alpha_1-1} p_2^{\alpha_2-1} \prod_{i=1}^n p_0^{Z_{i0}} p_1^{Z_{i1}} p_2^{Z_{i2}} \left(\sigma_2^{-1} \exp \left\{ -\frac{1}{2\sigma_2^2} (u_i - \mu_2)^2 \right\} \right)^{Z_{i2}} \\ &\propto \left(\frac{1}{\sigma_2^2} \right)^{\frac{n_2}{2}} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=1}^{n_2} (u_i - \mu_2)^2 \right\} p_0^{\alpha_0-1+\sum_{i=1}^n Z_{i0}} p_1^{\alpha_1-1+\sum_{i=1}^n Z_{i1}} p_2^{\alpha_2-1+\sum_{i=1}^n Z_{i2}} \end{aligned}$$

em que $\sigma_2^2 = [(-p_1\mu_1 - p_2\mu_2 - p_1\sigma_1\sqrt{2}/\sqrt{\pi}) / (p_2\sqrt{2}/\sqrt{\pi})]^2$ e n_2 são os componentes pertencentes a Z_{i2} . Nesse caso, a distribuição condicional completa para os pesos não possui forma fechada e será necessário um passo de Metropolis para gerar-se da distribuição a *posteriori* de \mathbf{p} . Dada a observação $\mathbf{p}^{(j-1)}$ obtida na iteração $j - 1$, no passo Metropolis-Hastings, considera-se como distribuição proposta para gerar os candidatos \mathbf{p}^* a distribuição normal multivariada $\mathcal{N}_3(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$. O valor de $\boldsymbol{\mu}^*$ é o valor dos pesos no passo anterior e a matriz de covariância $\boldsymbol{\Sigma}^*$ foi escolhida tal que a taxa de aceitação do algoritmo estivesse em torno de 0,23. Para garantir que a soma dos pesos seja igual a um, tem-se a seguinte restrição

$$\frac{p_0^*}{p_0^* + p_1^* + p_2^*} + \frac{p_1^*}{p_0^* + p_1^* + p_2^*} + \frac{p_2^*}{p_0^* + p_1^* + p_2^*} = 1.$$

Assim, a probabilidade de aceitação para os pesos é dada por

$$Q = \min \left\{ 1, \frac{\pi(\mathbf{p}^*|\cdot)\pi(\mathbf{p}^*)}{\pi(\mathbf{p}^{(j-1)}|\cdot)\pi(\mathbf{p}^{(j-1)})} \right\}.$$

Distribuição Condicional Completa para \mathbf{Z} : Sejam $f_0(u_i|\gamma_0)$, $f_1(u_i|\gamma_1)$ e $f_2(u_i|\gamma_2)$ as respectivas funções de densidade de probabilidade de M_0 , M_1 e M_2 . Desta maneira a distribuição condicional completa de \mathbf{Z}_i é proporcional a

$$\begin{aligned} \pi(\mathbf{Z}_i|\cdot) &\propto \pi(\mathbf{u}|\mathbf{Z}_i, \gamma_k)\pi(\mathbf{Z}_i|\mathbf{p}) \\ &\propto f_0(u_i|\gamma_0)^{Z_{i0}} f_1(u_i|\gamma_1)^{Z_{i1}} f_2(u_i|\gamma_2)^{Z_{i2}} \frac{1!}{Z_{i0}!Z_{i1}!Z_{i2}!} p_0^{Z_{i0}} p_1^{Z_{i1}} p_2^{Z_{i2}} \\ &\propto \frac{1}{Z_{i0}!Z_{i1}!Z_{i2}!} (f_0(u_i|\gamma_0)p_0)^{Z_{i0}} (f_1(u_i|\gamma_1)p_1)^{Z_{i1}} (f_2(u_i|\gamma_2)p_2)^{Z_{i2}} \\ &\propto \frac{1}{Z_{i0}!Z_{i1}!Z_{i2}!} \left(\frac{f_0(u_i|\gamma_0)p_0}{f^*} \right)^{Z_{i0}} \left(\frac{f_1(u_i|\gamma_1)p_1}{f^*} \right)^{Z_{i1}} \left(\frac{f_2(u_i|\gamma_2)p_2}{f^*} \right)^{Z_{i2}}, \end{aligned}$$

em que para $k = 0, 1, 2$, $\sum_{k=0}^2 Z_{ik} = 1$ e $f^* = p_0f_0(u_i|\gamma_0) + p_1f_1(u_i|\gamma_1) + p_2f_2(u_i|\gamma_2) = 1$. Consequentemente, uma vez que

$$p_k^{**} = \frac{f_k(u_i|\gamma_k)p_k}{f^*} \in (0, 1),$$

tem-se que

$$\mathbf{Z}_i|\cdot \sim \text{Multinomial}(1, p_0^{**}, p_1^{**}, p_2^{**}).$$

Distribuição Condicional Completa para u_i : A condicional completa de u_i é

$$\begin{aligned}
\pi(u_i|\cdot) &\propto \pi(\mathbf{y}|\mathbf{w}, \mathbf{u}, \boldsymbol{\beta}, \sigma_e^2)\pi(\mathbf{w}|\mathbf{u}, \mu_x, \sigma_x^2)\pi(\mathbf{u}|\mathbf{Z}_i, \boldsymbol{\gamma}_k) \\
&\propto \exp\left\{-\frac{1}{2}\left[u_i^2\left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2}\right) - 2u_i\left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2}\right)\right]\right\} \times \\
&\quad \left[\mathbb{I}(Z_{i0} = 1) \exp\left\{-\frac{u_i^2}{2\sigma_0^2}\right\} + \mathbb{I}(Z_{i1} = 1) \exp\left\{-\frac{1}{2}\left(\frac{u_i^2}{\sigma_1^2} - 2u_i\frac{\mu_1}{\sigma_1^2}\right)\right\} \mathbb{I}(u_i > \mu_1) + \right. \\
&\quad \left. \mathbb{I}(Z_{i2} = 1) \exp\left\{-\frac{1}{2}\left(\frac{u_i^2}{\sigma_2^2} - 2u_i\frac{\mu_2}{\sigma_2^2}\right)\right\} \mathbb{I}(u_i < \mu_2)\right] \\
&\propto \mathbb{I}(Z_{i0} = 1) \exp\left\{-\frac{1}{2}\left[u_i^2\left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_0^2}\right) - 2u_i\left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2}\right)\right]\right\} \\
&\quad + \mathbb{I}(Z_{i1} = 1) \exp\left\{-\frac{1}{2}\left[u_i^2\left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_1^2}\right) - 2u_i\left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2} + \frac{\mu_1}{\sigma_1^2}\right)\right]\right\} \\
&\quad \mathbb{I}(u_i > \mu_1) + \mathbb{I}(Z_{i2} = 1) \exp\left\{-\frac{1}{2}\left[u_i^2\left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_2^2}\right) - 2u_i\left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2} + \frac{\mu_2}{\sigma_2^2}\right)\right]\right\} \\
&\quad \mathbb{I}(u_i < \mu_2),
\end{aligned}$$

ou seja,

$$\pi(u_i|\cdot) = \begin{cases} \phi(u_i|\xi_0^*, \nu_0^*), & \text{se } Z_{i0} = 1 \\ \phi(u_i|\xi_1^*, \nu_1^*)\mathbb{I}(u_i > \mu_1), & \text{se } Z_{i1} = 1 \\ \phi(u_i|\xi_2^*, \nu_2^*)\mathbb{I}(u_i < \mu_2), & \text{se } Z_{i2} = 1 \end{cases}$$

em que,

$$\begin{aligned}
\nu_0^* &= \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_0^2}\right)^{-1}, \quad \xi_0^* = \left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2}\right) \nu_0^*, \\
\nu_1^* &= \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_1^2}\right)^{-1}, \quad \xi_1^* = \left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2} + \frac{\mu_1}{\sigma_1^2}\right) \nu_1^*, \\
\nu_2^* &= \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_2^2}\right)^{-1}, \quad \xi_2^* = \left(-\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} + \frac{w_i - \mu_x}{\sigma_x^2} + \frac{\mu_2}{\sigma_2^2}\right) \nu_2^*.
\end{aligned}$$

Distribuição Condicional Completa para σ_1^2 : A distribuição condicional completa de σ_1^2 é

$$\begin{aligned}
\pi(\sigma_1^2|\cdot) &\propto \pi(\mathbf{u}|\mathbf{Z}_i, \sigma_1^2)\pi(\sigma_1^2) \\
&\propto (\sigma_1^2)^{-a_1 - \frac{n_1}{2} - 1} \exp\left\{-\frac{b_1 + \frac{1}{2}\sum_{i=1}^{n_1}(u_i - \mu_1)^2}{\sigma_1^2}\right\} + \\
&\quad (\sigma_2^2)^{-\frac{n_2}{2}} \exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=1}^{n_2}(u_i - \mu_2)^2\right\} (\sigma_1^2)^{-a_1 - 1} \exp\left\{-\frac{b_1}{\sigma_1^2}\right\},
\end{aligned}$$

em que n_1 é o número de unidades em que $Z_{i1} = 1$ e n_2 é o número de unidades em que $Z_{i2} = 1$. Neste caso, usando um passo de Metropolis-Hasting e considerando como

distribuição proposta para gerar os candidatos σ_1^{2*} a distribuição normal $\mathcal{N}(\mu^*, \sigma^*)$. O valor de μ^* é igual ao valor no passo anterior. Assim, a probabilidade de aceitação é dada por

$$Q = \min \left\{ 1, \frac{\pi(\sigma_1^{2*} | \cdot) \pi(\sigma_1^{2*})}{\pi(\sigma_1^{2(j-1)} | \cdot) \pi(\sigma_1^{2(j-1)})} \right\}.$$

A quantidade de σ^{2*} é regulada de forma a ter uma taxa de aceitação em torno de 0,44.

3.3 Critérios de Seleção de Modelos

Na inferência Bayesiana existem vários métodos que facilitam a escolha do melhor modelo ajustado para um determinado conjunto de dados. Em Spiegelhalter et al. (2002) os critérios de seleção no contexto Bayesiano são obtidos a partir da distribuição *a posteriori*. Neste trabalho, os critérios considerados são o *deviance information criterion* (DIC) e o logaritmo da verossimilhança pseudomarginal (LPML). O critério LPML é definido com base estatística na ordenada preditiva condicional (CPO), que é baseado no critério de validação cruzada. Seja $\theta = (\beta, \mu_x, \sigma_e^2, \sigma_x^2, \mathbf{p}, \sigma_1^2)$, e $D(\theta) = -2l(\theta | \mathbf{y}, \mathbf{w})$, em que $l(\theta | \mathbf{y}, \mathbf{w})$ é o logaritmo da função de verossimilhança do modelo dada por (3.13). Seja $\theta^{(l)}$, $m = 1, \dots, l$, a amostra MCMC válida (após descartar o *burn-in*). Usando amostras MCMC, a aproximação para o CPO_i , para a i -ésima observação pode ser obtida por

$$\widehat{CPO}_i = \left\{ \frac{1}{m} \sum_{l=1}^m \frac{1}{f(\mathbf{y}, \mathbf{w} | \theta^{(l)})} \right\}^{-1}.$$

Assim, o LPML é definido como

$$LPML = \sum_{i=1}^n \ln(\widehat{CPO}_i).$$

O DIC observado é definido como

$$DIC = D(\bar{\theta}) + 2\tau_D,$$

em que, $D(\bar{\theta}) = -2l(\bar{\theta} | \mathbf{y}, \mathbf{w})$, $\tau_D = \overline{D(\theta)} - D(\bar{\theta})$ e

$$\overline{D(\theta)} = -\frac{1}{m} \sum_{l=1}^m D(\theta^{(m)}).$$

Para comparação dos modelos, valores pequenos de DIC e valores maiores do LPML indicam os melhores modelos. No capítulo a seguir serão abordados alguns estudos de simulações referente a metodologia proposta.

4 Estudo de Simulação

Neste capítulo, tem-se por objetivo analisar a performance do modelo proposto por meio de estudos de simulações. Considera como estimativas *a posteriori* a média e o intervalo HPD com probabilidade 0,95. O programa foi implementado no *software* R versão (3.5.2) (R Core Team). Como anteriormente mencionado, os modelos com erros de medida apresentam problemas de identificabilidade, não sendo possível a estimação de todos os parâmetros sem inclusão de alguma restrição no modelo. Para contornar esse problema se assume como restrição que o intercepto é conhecido.

Diferentes cenários são analisados. No primeiro estudo seção (4.1), as simulações foram realizadas com o propósito de avaliar o impacto nas inferências das escolhas das distribuições *priori* para os pesos da mistura, visto que na prática nem sempre é possível obter informações sobre esses elementos. No segundo estudo seção (4.2), a análise é dividida em quatro cenários em que diferentes valores para os pesos são assumidos para gerar os erros de medida.

Para cada um dos estudos gera-se a variável explicativa w_i por meio de uma distribuição normal com média $\mu_x + u_i$, $\mu_x = 2$ e variância σ_x^2 igual a 1,5. A variável u_i gerada por meio de uma mistura finita, na qual M_0 segue uma distribuição normal com média zero e parâmetro de escala σ_0^2 considerado igual a 10^{-5} , que representa uma normal concentrada em zero com desvio padrão igual a 0,003. As variáveis M_1 e M_2 com distribuição *half-normal* com os parâmetros de locação μ_1 e μ_2 considerados conhecidos e iguais a 0,01 e -0.01 , respectivamente. Entende-se que os valores dentro do intervalo $(-0,01, 0,01)$ estão a uma distância próxima de zero sendo praticamente nulos e portanto considerados como sem erro de medida. Os parâmetros de escala σ_1^2 e σ_2^2 escolhidos de forma que a variância do erro de medida considerada seja sempre menor que a variância do verdadeiro valor da variável, $\text{Var}(\mathbf{u}) < \text{Var}(\mathbf{x})$, uma vez que não faz sentido que a variância dos erros seja maior que a do verdadeiro valor dos dados. Isso indicaria que os dados possuem mais erro que valores medidos corretamente. A variável resposta y_i tem uma distribuição normal com média $\beta_1(w_i - u_i)$ em que $\beta_1 = 0,5$ e variância $\sigma_e^2 = 0,09$.

Nos ajustes, em todos estudos, adota-se as seguintes distribuições *a priori* não informativas para os parâmetros: $\beta_1 \sim \mathcal{N}(\mu_{\beta_1} = 0; \sigma_{\beta_1}^2 = 0,001)$; $\mu_x \sim \mathcal{N}(\mu_{xx} = 0, \sigma_{xx}^2 = 0,001)$; $\sigma_e^2 \sim \mathcal{GI}(a_e = 0,1, b_e = 0,1)$ e $\sigma_x^2 \sim \mathcal{GI}(a_x = 0,1, b_x = 0,1)$. Para os demais parâmetros *a priori* são especificadas em cada estudo. Em todos os estudos para garantir a identificabilidade do modelo, assumiu-se $\beta_0 = 0$.

Para o MCMC em cada réplica, considerou-se 150.000 iterações, sendo 50.000 para o período de aquecimento (*burn in*), para evitar as autocorrelação, saltos (*lag*) de

50 iterações foram consideradas. Como resultado, obteve-se uma amostra da distribuição *a posteriori* de tamanho 2.000.

4.1 Efeito das especificações a *priori* para \mathbf{p} nas estimativas a *posteriori*

Nesta seção, é conduzido um estudo de sensibilidade considerando diferentes escolhas para a distribuição a *priori* dos pesos. O objetivo tem-se por verificar o impacto das escolhas a *priori* dos parâmetros nas estimativas a *posteriori*. Para isso, os dados foram gerados assumindo o erro de medida simétrico e os pesos iguais a $p_0 = 0,4$, $p_1 = 0,3$ e $p_2 = 0,3$. Para a análise dos dados adota-se como distribuição a *priori* para σ_1^2 a distribuição $\mathcal{GI}(a_1 = 0,1, b_1 = 0,1)$. Considera-se três ajustes que se diferem pelas escolhas a *priori* para \mathbf{p} : (i) Uma distribuição muito informativa em que a distribuição *priori* é centrada nos seus valores reais e com desvio padrão $(0,002; 0,002; 0,002)$; $\mathbf{p} \sim \text{Dirichlet}(4000; 3000; 3000)$. (ii) $\mathbf{p} \sim \text{Dirichlet}(4; 3; 3)$ que é uma distribuição pouco informativa centrada nos valores reais com desvio padrão igual a $(0,14; 0,13; 0,13)$; (iii) $\mathbf{p} \sim \text{Dirichlet}(0,01; 0,01; 0,01)$ que é uma distribuição a *priori* não informativa. Neste estudo apenas uma réplica é considerada.

A Tabela 1 apresenta a média, mediana, desvio padrão e os intervalos HPD com probabilidade 0,95 a *posteriori* para todos os parâmetros. Com a *priori* não informativa tem-se um intervalo HPD maior para os pesos, mas os resultados da média e mediana a *posteriori* revelam-se próximos aos valores reais dos pesos independente da escolha a *priori*. Em médias as estimativas a *posteriori* dos demais parâmetros são próximas de seus valores reais sendo σ_1^2 e σ_2^2 os que parâmetros que apresentam os maiores desvio padrões e intervalos HPD com maiores amplitudes para os resultados usando a *priori* não informativas. Na prática, conhecer os valores dos pesos associados ao erro de medida é um ponto que pode ser inconveniente e restritivo. Neste estudo observa-se que mesmo que não se tenha informação a *priori* precisa sobre os pesos é possível obter-se estimativas razoáveis sobre os parâmetros as quais se mostram comparáveis à aquelas obtidas quando a distribuição a *priori* informativas são consideradas.

Tabela 1 – Resultados do estudo de simulação com diferentes distribuições *a priori* para os pesos.

$\pi(\mathbf{p})$	θ	Real	Média	Mediana	DP	HPD LI	HPD LS
Dir(4000,3000,3000)	β_1	0,50	0,4955	0,4953	0,0071	0,4818	0,5095
Dir(4,3,3)			0,4936	0,4935	0,0081	0,4783	0,5092
Dir(0,01;0,01;0,01)			0,4971	0,4973	0,0084	0,4805	0,5133
Dir(4000,3000,3000)	μ_x	2,00	2,0210	2,0210	0,0046	1,9350	2,1134
Dir(4,3,3)			2,0236	2,0240	0,0472	1,9325	2,1154
Dir(0,01;0,01;0,01)			2,0124	2,0119	0,0479	1,9167	2,1045
Dir(4000,3000,3000)	p_0	0,40	0,3999	0,3999	0,0015	0,3969	0,4028
Dir(4,3,3)			0,4183	0,4184	0,0596	0,3139	0,5273
Dir(0,01;0,01;0,01)			0,3854	0,3882	0,0667	0,2720	0,5043
Dir(4000,3000,3000)	p_1	0,30	0,3001	0,3001	0,0014	0,2975	0,3029
Dir(4,3,3)			0,2873	0,2865	0,0341	0,3139	0,3507
Dir(0,01;0,01;0,01)			0,3084	0,3076	0,0389	0,2373	0,3793
Dir(4000,3000,3000)	p_2	0,30	0,2999	0,2999	0,0015	0,2970	0,3029
Dir(4,3,3)			0,2944	0,2943	0,0305	0,2346	0,3514
Dir(0,01;0,01;0,01)			0,3062	0,3061	0,0320	0,2481	0,3630
Dir(4000,3000,3000)	σ_x^2	1,50	1,5633	1,5615	0,0872	1,3937	1,7319
Dir(4,3,3)			1,5794	1,5763	0,0924	1,3984	1,7602
Dir(0,01;0,01;0,01)			1,5591	1,5562	0,0848	1,3955	1,7278
Dir(4000,3000,3000)	σ_e^2	0,09	0,1054	0,1053	0,0095	0,0879	0,1245
Dir(4,3,3)			0,1098	0,1101	0,0104	0,0888	0,1292
Dir(0,01;0,01;0,01)			0,1062	0,1065	0,0125	0,0809	0,1288
Dir(4000,3000,3000)	σ_1^2	1,50	1,4246	1,4207	0,1139	1,2096	1,6514
Dir(4,3,3)			1,5092	1,4829	0,2319	1,0872	1,9621
Dir(0,01;0,01;0,01)			1,4090	1,3810	0,2156	1,0450	1,8496
Dir(4000,3000,3000)	σ_2^2	1,50	1,4262	1,4214	0,1152	1,2137	1,6631
Dir(4,3,3)			1,4278	1,4095	0,1939	1,0715	1,8234
Dir(0,01;0,01;0,01)			1,4177	1,4013	0,1779	1,0860	1,7720

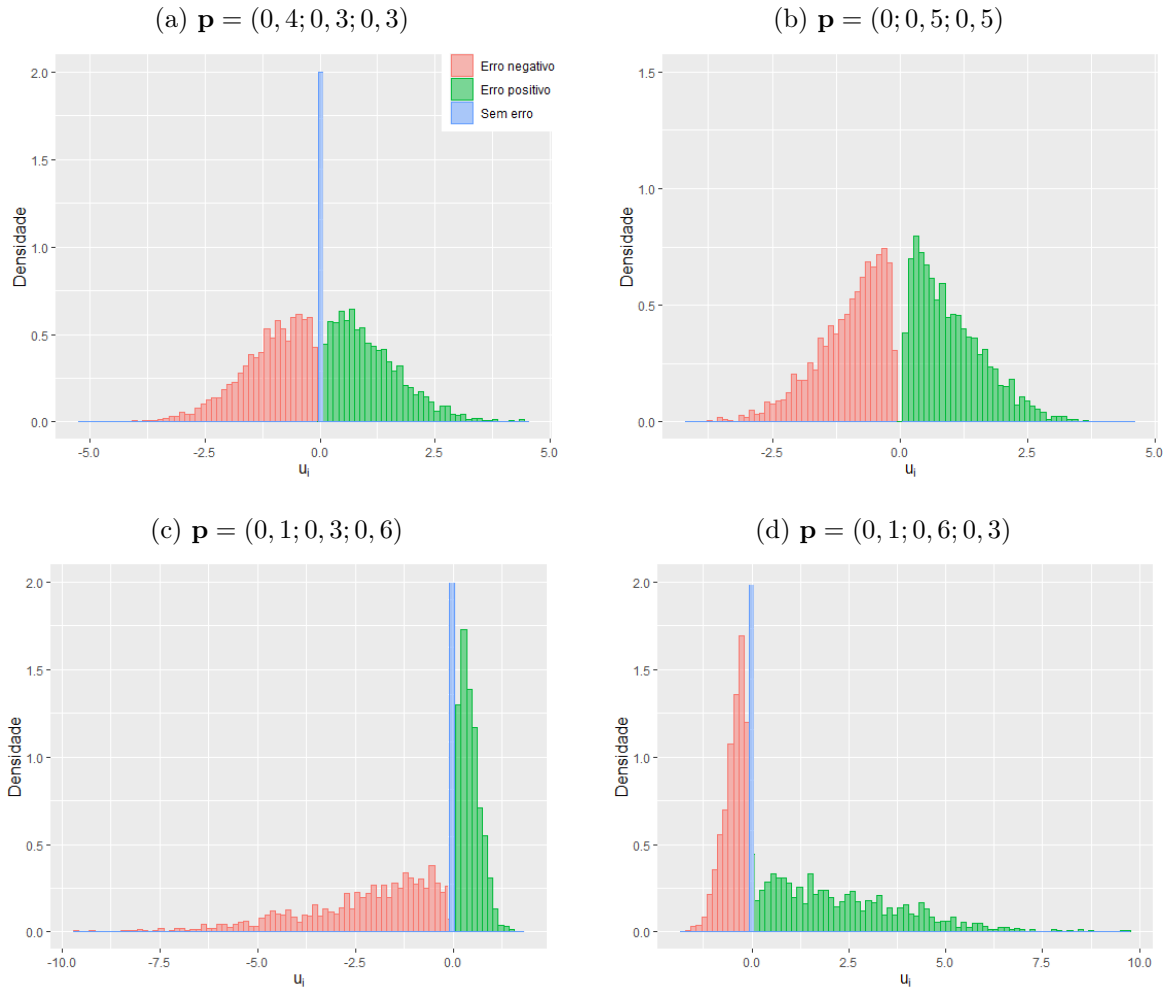
4.2 Dados com diferentes assimetrias para o erro de medida

Nesta seção, tem-se como propósito avaliar a performance do modelo para diferentes graus de assimetrias no erro de medida. Para isso, foram gerados 100 réplicas, em cada cenário considerando valores distintos para os pesos. Este número de réplicas foi escolhido em função do longo tempo computacional envolvido na obtenção das estimativas *a posteriori*. Os seguintes cenários foram considerados

- Cenário 1: erro de medida simétrico e pesos iguais a $(0,4;0,3;0,3)$;
- Cenário 2: erro de medida simétrico com todos os dados mensurados com erros e pesos iguais a $(0;0,5;0,5)$;
- Cenário 3: erro de medida assimétrico em que a maior na parte do erro possui massa positiva e pesos iguais a $(0,1;0,6;0,3)$;
- Cenário 4: erro de medida assimétrico em que a maior na parte do erro possui massa negativa e pesos iguais a $(0,1;0,3;0,6)$.

A Figura 4 apresenta exemplos das distribuições para u_i nos quatro cenários analisados. Os valores de σ_1^2 e σ_2^2 são diferentes em cada estudo, pois a escolha garante que a variância do erro de medida seja menor que a variância da variável latente. Baseado no estudo de sensibilidade dos pesos, a escolha da *priori* para os mesmos em todos os cenários é a distribuição Dirichlet(0,01;0,01;0,01) e uma distribuição *a priori* não informativa para σ_1^2 , $\mathcal{GI}(1, 1)$.

Figura 4 – Exemplos de funções de densidades da distribuição do erro de medida para os diferentes cenários analisados.



A fim de avaliar a performance do modelo algumas estatísticas são consideradas. Seja θ um elemento de $\boldsymbol{\theta} = (\beta_1, \mu_x, \sigma_e^2, \sigma_x^2, p_0, p_1, p_2, \sigma_1^2)$ e $\hat{\theta}_r$ a estimativa *a posteriori* de $\boldsymbol{\theta}$ para r -ésima réplica. Foram considerados a média e o desvio-padrão das estimativas dados respectivamente,

$$\bar{\theta} = \sum_{r=1}^R \frac{\hat{\theta}_r}{R},$$

$$SD_{\theta} = \sqrt{\sum_{r=1}^R \frac{(\hat{\theta}_r - \bar{\theta})^2}{(R-1)}}.$$

Também considerou-se o vício $\bar{\theta} - \theta$; variância $Var_{\theta} = SD_{\theta}^2$ e o erro quadrático médio $EQM = \sqrt{Vicio^2 + Var_{\theta}}$, em que R é o número de réplicas. Para obter o percentual de cobertura do intervalo HPD com 95% de confiança, o número de intervalos que contém o verdadeiro valor do parâmetro é calculado e dividido pelo número de réplicas.

Um ponto importante a ser verificado é em relação as classificações à *posteriori* de cada indivíduo segundo tipo de erro de medida que experimenta. A fim de verificar a quantidade de acertos das classificações, considera-se a moda a *posteriori* dos pesos da mistura para cada indivíduo, em cada uma das réplicas, e depois obtém a quantidade de acertos geral por meio da moda dessas réplicas.

No primeiro cenário, o erro de medida considerado é simétrico com pesos iguais a (0,4;0,3;0,3) e assume-se $\sigma_1^2 = \sigma_2^2 = 1,5$. Os resultados da média, vício, desvio padrão (SD), EQM e os intervalos de cobertura HPD são apresentados na Tabela 2, considerando os tamanhos amostrais n=250 e 500. Nota-se que as médias das estimativas obtidas são próximas dos seus valores reais. No entanto, como esperado para todos os parâmetros à medida que o tamanho da amostra aumenta, as estimativas tornam-se mais próxima dos valores reais. A Tabela 3 mostra a proporção ocorrida de cada tipo de erro de medida dentro de cada grupo de erros. Percebe-se que a proporção de indivíduos com erros positivos ou negativos indicados pelo modelo é inferior ao percentual real. A ocorrência da mudança de classificação do erro positivo para o erro negativo ou vice-versa não ocorre, indicando um bom comportamento na classificação uma vez que não ocorre a inversão de sinal no erro. A ocorrência dos erros positivos ou negativos classificados como sem erro, são os valores gerados mais próximos de zero que podem ser vistos na Figura 5. As taxas de acertos totais das classificações do tipo de erro são: 63% para n=250 e 65% para n=500.

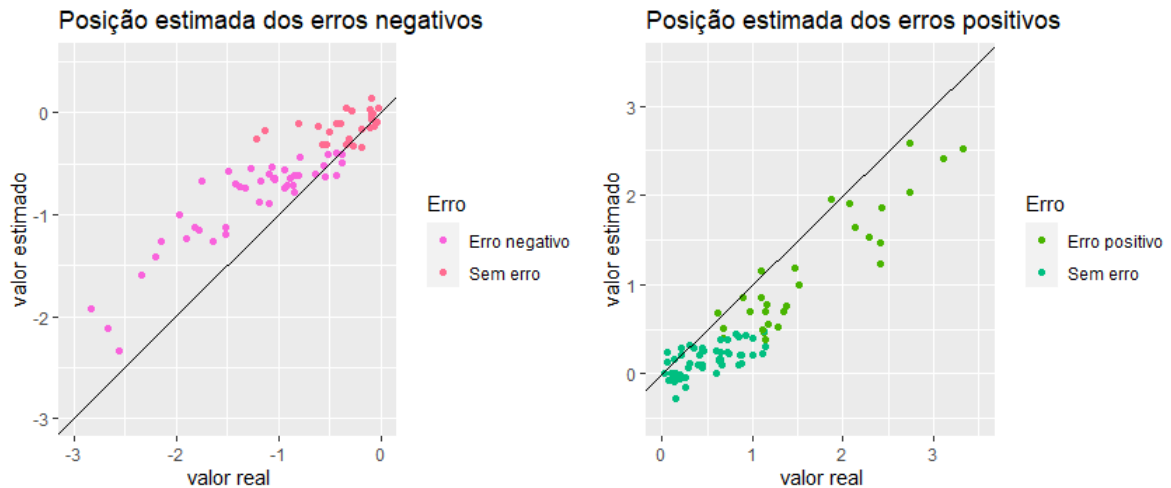
Tabela 2 – Resultados do estudo de simulação com $n=250$ e 500 , cenário 1.

n	Parâmetros	Valor Real	Média	Vício	SD	EQM	Cobertura HPD
250	β_1	0,50	0,4861	-0,0139	0,0107	0,0003	1,00
	μ_x	2,00	1,8788	-0,1212	0,0209	0,0151	0,99
	p_0	0,40	0,4359	0,0359	0,0900	0,0094	0,93
	p_1	0,30	0,2861	-0,0139	0,0467	0,0024	0,97
	p_2	0,30	0,2781	-0,0219	0,0496	0,0029	0,92
	σ_x^2	1,50	1,4200	-0,0800	0,0630	0,0104	1,00
	σ_e^2	0,09	0,1292	0,0392	0,0131	0,0017	0,84
	σ_1^2	1,50	1,4680	-0,0320	0,3036	0,0932	0,97
	σ_2^2	1,50	1,5491	0,0491	0,3163	0,1024	0,98
500	β_1	0,50	0,4932	-0,0068	0,0073	0,0001	1,00
	μ_x	2,00	2,0228	0,0228	0,0233	0,0156	0,98
	p_0	0,40	0,4412	0,0412	0,0402	0,0033	1,00
	p_1	0,30	0,2586	-0,0414	0,0311	0,0027	0,76
	p_2	0,30	0,3001	0,0001	0,0376	0,0014	0,95
	σ_x^2	1,50	1,6597	0,1597	0,0603	0,0291	0,90
	σ_e^2	0,09	0,1247	0,0347	0,0122	0,0014	0,66
	σ_1^2	1,50	1,5500	0,0500	0,2571	0,0686	0,97
	σ_2^2	1,50	1,1748	-0,3252	0,3288	0,2138	0,60

Tabela 3 – Classificação dos componentes da mistura para n iguais a 250 e 500, cenário 1.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
250	Sem erro	0,36	0,02	0,02
	Erro positivo	0,18	0,14	0,00
	Erro negativo	0,12	0,00	0,16
500	Sem erro	0,30	0,07	0,04
	Erro positivo	0,11	0,18	0,00
	Erro negativo	0,12	0,01	0,16

Figura 5 – Posição estimadas *a posteriori* dos erros gerados negativos e positivos para $n=250$, cenário 1.



No segundo cenário o erro de medida é simétrico com pesos iguais a $(0;0,5;0,5)$, indicando que w_i nunca foi medido totalmente sem erro. Os valores de σ_1^2 e σ_2^2 são iguais a 1,3 e 1,3, respectivamente. Na Tabela 4, os resultados são apresentados para os tamanhos amostrais iguais a 250 e 500. Nota-se que mesmo sendo $p_0 = 0$, a média a *posteriori* indica que cerca de 1,5% dos indivíduos não experimentaram erro na covariável. Isso se deve ao fato do modelo ajustado considerar que existe os três componentes da mistura. Também nota-se uma sub-estimação do número de indivíduos cuja covariável experimenta um erro negativo, se $n=250$, e um erro positivo, se $n=500$.

Ao observar as classificações dos componentes da mistura na Tabela 5, nota-se que agora tem-se uma troca de classificação entre os erros positivos e negativos. Nota-se também não haver componentes classificados como sem erro, mas como a classificação é obtida pela moda geral das réplicas pode ser que em algum momento o indivíduo foi considerado como sem erro mas ao obter a moda ele tenha sido classificado como positivo ou negativo. Neste cenário, observa-se que as estimativas tendem a serem mais viciadas e os EQM tendem a serem maiores que no primeiro cenário. As taxas de acerto total das classificações do erro de mensuração são 91% para $n=250$ e 91% para $n=500$.

Tabela 4 – Resultados do estudo de simulação com $n=250$ e 500 , cenário 2.

n	Parâmetros	Valor Real	Média	viés	SD	EQM	Cobertura HPD
250	β_1	0,50	0,4890	-0,0110	0,0117	0,0003	1,00
	μ_x	2,00	1,9811	-0,0189	0,0336	0,0015	1,00
	p_0	0,00	0,0161	0,0161	0,0217	0,0007	0,00
	p_1	0,50	0,5487	0,0487	0,0280	0,0032	0,53
	p_2	0,50	0,4352	-0,0648	0,0258	0,0049	0,28
	σ_x^2	1,50	1,2808	-0,2192	0,1126	0,0607	0,69
	σ_e^2	0,09	0,1476	0,0576	0,0222	0,0038	0,36
	σ_1^2	1,30	1,0264	-0,2736	0,1249	0,0904	0,72
	σ_2^2	1,30	1,6385	0,3385	0,1935	0,1520	0,87
500	β_1	0,50	0,4738	-0,0262	0,0089	0,0008	1,00
	μ_x	2,00	2,1835	0,1835	0,0279	0,0345	0,05
	p_0	0,00	0,0150	0,0150	0,0300	0,0011	0,00
	p_1	0,50	0,4522	-0,0478	0,0282	0,0031	0,39
	p_2	0,50	0,5328	0,0328	0,0314	0,0021	0,54
	σ_x^2	1,50	1,7331	0,2331	0,1050	0,0653	0,58
	σ_e^2	0,09	0,1366	0,0466	0,0171	0,0025	0,36
	σ_1^2	1,30	1,3608	0,0608	0,1674	0,0317	0,93
	σ_2^2	1,30	0,9759	-0,3241	0,1267	0,1211	0,30

Tabela 5 – Classificação dos componentes da mistura para n iguais a 250 e 500, cenário 2.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
250	Sem erro	0,00	0,00	0,00
	Erro positivo	0,00	0,49	0,01
	Erro negativo	0,00	0,08	0,42
500	Sem erro	0,00	0,00	0,00
	Erro positivo	0,00	0,41	0,09
	Erro negativo	0,00	0,01	0,50

O terceiro cenário considera o erro de medida assimétrico com probabilidade maior para erros com assimetria positiva $\mathbf{p}=(0,3;0,6;0,1)$, $\sigma_1^2 = 0,2$ e $\sigma_2^2 = 7,54$. Os resultados do estudo de simulação são apresentados na Tabela 6, para os tamanhos amostrais $n=250$ e $n=500$. Na Tabela 7, a troca entre as posição real e a posição estimada dos erros negativos e dos erros positivos quase não ocorre, mas existe uma quantidade razoável de sem erro classificados como erros positivos. Isso pode ser devido ao fato de neste cenário a maior massa de probabilidade pertence ao erro positivo. Aqui a taxa de acerto total das classificações dos erros é em torno de 64%. Como nos outros cenários estimativas menos viciadas e com menor variabilidade são obtidas se $n=500$.

Tabela 6 – Resultados do estudo de simulação com $n=250$ e 500, cenário 3.

n	Parâmetros	Valor Real	Média	viés	SD	EQM	Cobertura HPD
250	β_1	0,50	0,5049	0,0049	0,0158	0,0003	0,92
	μ_x	2,00	1,9298	-0,0702	0,0543	0,0079	0,94
	p_0	0,30	0,3362	0,0362	0,1751	0,0320	0,58
	p_1	0,60	0,5672	-0,0328	0,1579	0,0260	0,59
	p_2	0,10	0,0966	-0,0034	0,0191	0,0004	0,89
	σ_x^2	1,50	1,2932	-0,2068	0,0452	0,0448	0,86
	σ_e^2	0,09	0,1016	0,0116	0,0113	0,0003	0,91
	σ_1^2	0,20	0,3507	0,1507	0,0771	0,0287	0,72
	σ_2^2	7,54	12,1432	4,6030	1,9634	25,0421	0,97
	500	β_1	0,50	0,5095	0,0095	0,0087	0,0002
μ_x		2,00	2,0117	0,0117	0,0296	0,0010	1,00
p_0		0,30	0,2852	-0,0148	0,1245	0,0157	0,69
p_1		0,60	0,5956	-0,0044	0,1113	0,0124	0,74
p_2		0,10	0,1192	0,0192	0,0146	0,0006	0,78
σ_x^2		1,50	1,4108	-0,0892	0,0290	0,0088	1,00
σ_e^2		0,09	0,0876	-0,0024	0,0072	0,0001	0,97
σ_1^2		0,20	0,2576	0,0576	0,0411	0,0050	0,95
σ_2^2		7,54	6,5379	-1,0024	0,5759	1,3364	0,99

Tabela 7 – Classificação dos componentes da mistura para n iguais a 250 e 500, cenário 3.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
250	Sem erro	0,14	0,15	0,01
	Erro positivo	0,18	0,43	0,00
	Erro negativo	0,02	0,01	0,07
500	Sem erro	0,09	0,21	0,01
	Erro positivo	0,11	0,48	0,00
	Erro negativo	0,02	0,01	0,08

No quarto cenário 4, como no cenário 3, o erro é assimétrico mas a probabilidade é maior na assimetria negativa, com pesos iguais a $(0,1;0,6;0,3)$ e $\sigma_1^2 = 8,5$ e $\sigma_2^2 = 0,23$. Os resultados apresentados na Tabela 8 para tamanhos amostrais iguais a 250 e 500 indica estimativas média a *posteriori* próximas dos seus valores reais. Para $n=500$ existe uma melhora nos resultados em relação a $n=250$. A Tabela 9 apresenta as classificações de cada componente da mistura, de forma similar ao cenário anterior, quase não ocorre troca entre erros positivos e erro negativos mas possui uma taxa maior de indivíduos que eram sem erro classificado no erro negativo. As taxas de acertos aumentam com o tamanho amostral. Com uma taxa de acerto total de 63% para $n=250$ e 65% para $n=500$.

Tabela 8 – Resultados do estudo de simulação com $n=250$ e 500, cenário 4.

n	Parâmetros	Valor Real	Média	viés	SD	EQM	Cobertura HPD
250	β_1	0,50	0,5129	0,0129	0,0106	0,0003	0,87
	μ_x	2,00	1,8843	-0,1157	0,0296	0,0143	0,89
	p_0	0,30	0,2519	-0,0481	0,1947	0,0402	0,42
	p_1	0,10	0,1162	0,0162	0,0197	0,0006	0,91
	p_2	0,60	0,6320	0,0320	0,1839	0,0348	0,40
	σ_x^2	1,50	1,2585	-0,2415	0,0468	0,0605	0,48
	σ_ϵ^2	0,09	0,1146	0,0246	0,0129	0,0008	0,53
	σ_1^2	8,50	5,1854	-3,3146	0,8957	11,7892	0,52
	σ_2^2	0,23	0,2009	-0,0252	0,1083	0,0124	0,66
500	β_1	0,50	0,4986	-0,0014	0,0083	0,0001	0,96
	μ_x	2,00	2,0479	0,0479	0,0247	0,0029	1,00
	p_0	0,30	0,2771	-0,0229	0,1070	0,0120	0,55
	p_1	0,10	0,1079	0,0079	0,0102	0,0002	0,95
	p_2	0,60	0,6150	0,0150	0,1000	0,0102	0,56
	σ_x^2	1,50	1,4493	-0,0507	0,0257	0,0032	1,00
	σ_ϵ^2	0,09	0,0991	0,0091	0,0089	0,0002	0,78
	σ_1^2	8,50	7,1469	-1,3531	0,6037	2,1954	0,96
	σ_2^2	0,23	0,2215	-0,0046	0,0639	0,0041	0,85

Tabela 9 – Classificação dos componentes da mistura para n iguais a 250 e 500, cenário 4.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
250	Sem erro	0,09	0,01	0,19
	Erro positivo	0,02	0,07	0,02
	Erro negativo	0,13	0,01	0,47
500	Sem erro	0,03	0,01	0,26
	Erro positivo	0,00	0,08	0,02
	Erro negativo	0,04	0,01	0,54

Comparando todos os resultados obtidos nas simulações, observa-se que as estimativas, em geral, são mais precisas à medida que aumenta o tamanho da amostral. Comparando os cenários 1 e 2, em que o erro de medida tem distribuição simétrica observa-se que se $n=500$ as estimativas menos viciadas são obtidas no cenário 1 que tem um número maior balanceado de unidades com erro positivo, com erro negativo e sem erro. Comparando os cenários 3 e 4, se $n=250$, verifica-se que o cenário com maior percentual de erro negativos fornece estimativas mais viciadas, exceto para p_1 e σ_2^2 . O Apêndice C, contém os resultados do estudo do cenário 1 e 2 no caso em que o tamanho amostral é igual a 1000.

5 Aplicação

Neste capítulo, para ilustrar o modelo proposto dois conjuntos de dados são analisados. As análises feitas consideram a comparação do modelo de regressão linear simples com distribuição de mistura finita para o erro de medida (MRLSM) com o modelo de regressão linear simples com distribuição normal para o erro de medida (MRLSN) e com o modelo sem erro de medida (MRLSSE). O ajuste do modelo com erro de medida normal e do modelo sem erro de medida foram feitos no *software* R versão (3.5.2) (R Core Team) e no JAGS, usando o pacote *rjags*, Plummer et al. (2019).

5.1 Banco de dados: rendimento do milho

Nesta seção o modelo proposto é aplicado a um conjunto de banco de dados reais retirados da página 19, Fuller (1987). O conjunto de dados possui amostras coletadas do rendimento do milho (\mathbf{y}) em função da quantidade de nitrogênio no solo (\mathbf{w}). Os dados são referentes a 11 amostras coletadas em Iowa nos Estados Unidos. Em Fuller (1987), o autor ajusta o modelo com a situação de identificabilidade em que a variância do erro de medida é conhecida e igual a 57, porém como no modelo proposto a variância de u_i é dada pela expressão (3.11), essa informação foi usada para obter alguma informação sobre a *priori* de σ_1^2 e para identificação do modelo o σ_x^2 é considerado como conhecido e igual ao valor obtido pelo autor de 247,8545.

A distribuição a *priori* de σ_u^2 em MRLSN foi escolhida de forma que sua média é igual a 57 e sua variância igual a 100. Para usar essas mesmas informações em MRLSM obteve-se *priori* de σ_1^2 de forma que $\text{Var}(\mathbf{u})=57$, os pesos p_1 e p_2 iguais e $p_0 = P(-0.01 < U < 0.01) = 0.01$. Assim, $p_1 = p_2 = 0.495$ e $\pi(\sigma_1^2)$ com $E(\sigma_1^2) = 155,717$ e $\text{Var}(\sigma_1^2) = 100$. A *priori* para os pesos foi a Dirichlet(0.01,0.01,0.01). Para as demais *priori* utilizadas nos dois ajustes assume-se: $\beta_1 \sim \mathcal{N}(0; 0,001)$; $\mu_x \sim \mathcal{N}(0; 0,001)$; $\sigma_e^2 \sim \mathcal{GI}(0, 1; 10)$.

Na Tabela 10, tem-se as estimativas médias, mediana e desvio padrão (DP) a *posteriori* para os três modelos. As estimativas do σ_e^2 do modelo com mistura são próximas as obtidas pelo modelo sem erro. Como os valores das estimativas média a *posteriori* dos pesos indicam que esse é um modelo em que 99% dos dados são medidos sem erro, as estimativas a *posteriori* de σ_1^2 e σ_2^2 foram omitidas da análise uma vez que p_1 e p_2 foram estimados próximos de zero. Em todos os parâmetros os desvios padrões no modelo de mistura são menores. Os três ajustes possuem estimativas dos parâmetros bem próximas o que reflete nas retas de regressão ajustadas, Figura 6.

Os intervalos HPD são apresentados na Tabela 11 em que é possível observar

intervalos menores no modelo com a mistura. Na Tabela 12, pelos critérios de comparação entre os modelos com erro de medida, o DIC e LPML indicam que o melhor modelo é o modelo sem erro de medida. Essa amostra possui 11 dados o que pode indicar não ter evidências suficiente para determinar que exista erro de medida nos dados.

Tabela 10 – Estimativas *posteriori* para todos os modelos, conjunto de dados de rendimento do milho.

θ	Modelo								
	MRLSM			MRLSN			MRLSSE		
	Média	Mediana	DP	Média	Mediana	DP	Média	Mediana	DP
β_0	64,3551	64,8354	8,3191	65,0619	65,9913	11,3878	64,9059	66,1618	11,2853
β_1	0,4629	0,4556	0,1152	0,4516	0,4408	0,1581	0,4535	0,43742	0,1584
μ_x	69,0112	69,0409	3,3353	68,8982	68,9069	4,7072	-	-	-
p_0	0,9855	0,9961	0,0210	-	-	-	-	-	-
p_1	0,0061	0,0008	0,0107	-	-	-	-	-	-
p_2	0,0083	0,0005	0,0173	-	-	-	-	-	-
σ_x^2	247,8545	247,8545	0,0000	247,8545	247,8545	0,0000	-	-	-
σ_e^2	79,2904	70,8699	36,9612	57,0511	62,6991	50,8886	76,5498	65,3595	47,5779
$Var(\mathbf{u})$	0,4179	0,4137	0,0586	2,1817	2,1647	0,1416	-	-	-

Figura 6 – Retas de regressão dos modelos em linha solida o MRLSM, em linha pontilhada MRLSN e em linha tracejada MRLSSE.

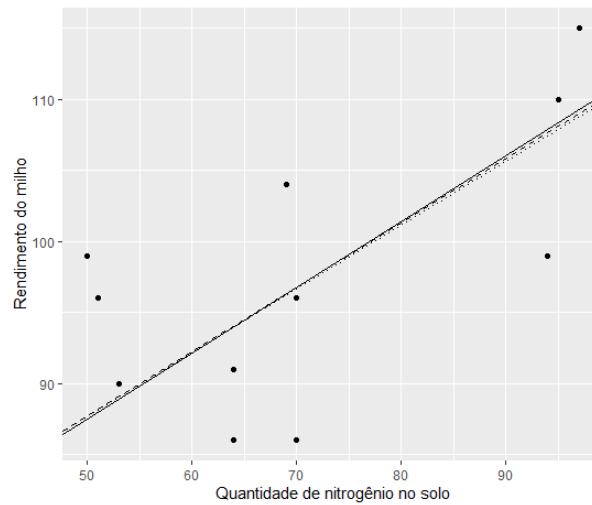


Tabela 11 – Intervalo HPD com 95% , conjunto de dados de rendimento do milho.

Parâmetros	Modelo					
	MRLSM		MRLSN		MRLSSE	
	LI	LS	LI	LS	LI	LS
β_0	47,2369	79,6580	42,3638	86,07954	39,4503	84,3543
β_1	0,2507	0,6992	0,1628	0,7702	0,1622	0,7780
μ_x	62,7110	75,7533	59,6568	77,7957	-	-
p_0	0,9409	1,0000	-	-	-	-
p_1	0,0000	0,0331	-	-	-	-
p_2	0,0000	0,0518	-	-	-	-
σ_e^2	29,0392	146,9394	21,5636	168,3496	21,7055	174,5031

Tabela 12 – Critério de comparação DIC e LPML, conjunto de dados de rendimento do milho.

	Modelo		
	MRLSM	MRLSN	MRLSSE
DIC	182,911	187,066	79,809
LPML	-99,671	-129,849	-40,097

5.2 Banco de dados: gastos ambulatoriais

Nesta seção, os dados analisados são provenientes de Cameron et al. (2009), disponíveis em "*2001 Medical Expenditure Panel Survey*". Os dados são constituídos por 3328 observações sobre os gastos ambulatoriais (\mathbf{y}), na escala logarítmica, com 526 observações iguais a zero. Possui variáveis como idade, gênero, status educacional, status de seguro, total de doenças crônicas, etnia e renda. Os mesmos foram analisados por Rocha et al. (2016) considerando o erro de medida na covariável renda. Aqui, a análise constitui-se em um modelo de regressão linear simples com a variável explicativa dada pela renda (\mathbf{w}) e sem as observações censuradas iguais a 526, resultando em um tamanho amostral igual a 2802.

Como na aplicação do banco de dados do rendimento do milho, o modelo a comparação do modelo proposto é feita comparando-se com o modelo com erro de mensuração normal e com o modelo sem erro de medida. O objetivo aqui é avaliar o comportamento da renda, deste modo para a identificação do modelo, considera-se o intercepto conhecido e igual a estimativa obtida do modelo sem erro de medida. As distribuições a *priori* para os pesos e de σ_1^2 foram Dirichlet(0,01;0,01;0,01) e $\mathcal{GI}(1, 1)$, respectivamente. Nos dois modelos considera-se as seguintes *priori* para os demais parâmetros: $\beta_1 \sim \mathcal{N}(0; 0, 001)$; $\mu_x \sim \mathcal{N}(0; 0, 001)$; $\sigma_e^2 \sim \mathcal{GI}(0, 1; 10)$. Antes dos ajustes, 20% dos dados foram retirados para avaliar a capacidade preditiva dos modelos. A capacidade preditiva é calculada através da média dos valores preditos menos os valores estimados ao quadrado.

A Tabela 13 mostra os resultados da média, mediana e desvio padrão dos parâmetros a *posteriori*. Os pesos indicam uma maior quantidade de renda informada corretamente sendo de 60% dos dados, 32% de erros positivos e 8% de erros negativos. Os indivíduos identificados com erro positivo são aqueles que possui renda média de \$9.622, 20, os classificados com erro negativo possui renda média de \$467, 30 e os que informaram a renda corretamente possui renda média de \$3.386, 80. Nota-se valores estimados diferentes para β_1 nos três modelos assim como os valores de σ_x^2 e $\text{Var}(\mathbf{u})$ nos modelos com erro. A estimativa de β_1 no modelo com mistura indica que a renda não é significativa para explicar os gastos ambulatoriais.

Os modelos com erro de mistura apresentam os menores desvios padrões e os

menores intervalos HPD, Tabela 14. Os critérios de comparação indicam que o melhor modelo é modelo sem erro, obtendo menor DIC e maior LPML, Tabela 15. Ao analisar a capacidade preditiva o modelo que possui a melhor capacidade preditiva é o modelo com mistura.

Tabela 13 – Estimativas *posteriori* para todos os modelos, conjunto de dados gastos ambulatoriais.

θ	Modelo								
	MRLSM			MRLSN			MRLSSE		
	Média	Mediana	DP	Média	Mediana	DP	Média	Mediana	DP
β_0	6,1730	6,1730	0,0000	6,1730	6,1730	0,0000	6,1730	6,1730	6,1730
β_1	0,0001	0,0001	0,0001	0,1785	0,1785	0,0086	0,1169	0,1169	0,0085
μ_x	3,3902	3,3903	0,0107	3,3845	3,3840	0,0158	-	-	-
p_0	0,5939	0,5910	0,0613	-	-	-	-	-	-
p_1	0,3243	0,3258	0,0495	-	-	-	-	-	-
p_2	0,0818	0,0821	0,0144	-	-	-	-	-	-
σ_x^2	0,3587	0,3592	0,0196	0,1152	0,1131	0,0247	-	-	-
σ_e^2	2,1726	2,1711	0,0458	2,0129	2,0125	0,0623	1,9873	1,9853	0,0596
σ_1^2	0,1169	0,1109	0,0258	-	-	-	-	-	-
σ_2^2	1,9786	1,9343	0,2931	-	-	-	-	-	-
$Var(\mathbf{u})$	0,4179	0,4137	0,0586	2,1817	2,1647	0,1416	-	-	-

Tabela 14 – Intervalo HPD com 95% , conjunto de dados gastos ambulatoriais.

Parâmetros	Modelo					
	MRLSM		MRLSN		MRLSSE	
	LI	LS	LI	LS	LI	LS
β_1	-0,0001	0,0002	0,1631	0,1947	0,1001	0,1332
μ_x	3,3696	3,4104	3,3552	3,4147	-	-
p_0	0,4736	0,7027	-	-	-	-
p_1	0,2387	0,4269	-	-	-	-
p_2	0,0547	0,1058	-	-	-	-
σ_x^2	0,3229	0,3997	0,0727	0,1686	-	-
σ_e^2	2,0877	2,2640	1,8820	2,1316	1,8817	2,1135
σ_1^2	0,0772	0,1745	-	-	-	-
σ_2^2	1,4648	2,5589	-	-	-	-
$Var(\mathbf{u})$	0,3109	0,5317	1,9191	2,4721	-	-

Tabela 15 – Critério de comparação DIC e LPML, conjunto de dados gastos ambulatoriais.

	Modelo		
	MRLSM	MRLSN	MRLSSE
DIC	15089,85	14482,29	7890,67
LPML	-7609,43	-7243,75	-3945,37

6 Conclusões

Este trabalho introduz um modelo de regressão linear com erro nas variáveis no qual o erro foi modelado a partir de uma distribuição de mistura finita. Essa nova proposta considera que parte das unidades podem ser medidas sem erro, com erro positivo e com erro negativo. Isso permite que o modelo proposto acomode certas características do erro de mensuração que não são usualmente levadas em consideração.

Com o objetivo de avaliar a performance do modelo proposto o presente trabalho apresenta estudos de simulação. Primeiro estuda-se o efeito da escolha *a priori* dos pesos da mistura em que se percebe não haver sensibilidade nas suas escolhas. Em seguida, se estuda a eficácia da abordagem desenvolvida sob diferentes tipos de erro por meio de um estudo de simulação Monte Carlo. Nota-se que as estimativas são melhores quando o tamanho da amostra é maior, da mesma forma com os acertos em relação as classificações dos componentes da mistura *a posteriori*. Analisa-se também dois conjuntos de dados reais no qual também é feito o ajuste comparando o modelo sem erro de mensuração e o com erro mensuração normal. O primeiro, contém o banco de dados presente em Fuller (1987) em que estimativas apresentadas indicam um melhor comportamento do modelo com mistura comparado com o normal e com o modelo sem erro de medida. Os resultados indicam que esse seria um modelo sem erro, porém com 11 dados pode-se não ter evidências suficientes para determinar o erro de medida. O segundo banco de dados, aponta que a renda no modelo com mistura não foi significativa para explicar os gastos ambulatoriais. As estatísticas LPML e DIC mostram que o melhor modelo é aquele que não possui erro de medida, no trabalho de Rocha et al. (2016), eles obtiveram conclusões nas quais essas ferramentas são ineficientes na comparação dos modelos com erro de medida daqueles sem erro de medida. Conclui-se então que o modelo proposto se sobressai no que se refere a interpretabilidade dos dados, permitindo a classificação e quanto ao tipo de erro e também no poder preditivo.

Como trabalho futuro sugere-se o estudo teórico para as condições de identificabilidade do modelo, como o cálculo da matriz de informação de Fisher. Sugere-se também a inclusão de covariáveis medidas sem erro no estudo e o uso do coeficiente de atenuação para identificabilidade em vez do conhecimento prévio de alguma variável.

Referências

- AOKI, R.; BOLFARINE, H.; ACHCAR, J. A.; JR, L. P. D. Bayesian analysis of a multivariate null intercept errors-in-variables regression model. *Journal of biopharmaceutical statistics*, Taylor & Francis, v. 13, n. 4, p. 767–775, 2003.
- AOKI, R.; BOLFARINE, H.; SINGER, J. M. Null intercept measurement error regression models. *Test*, Springer, v. 10, n. 2, p. 441–457, 2001.
- ARELLANO-VALLE, R. B.; OZAN, S.; BOLFARINE, H.; LACHOS, V. Skew normal measurement error models. *Journal of Multivariate Analysis*, Elsevier, v. 96, n. 2, p. 265–281, 2005.
- AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, JSTOR, p. 171–178, 1985.
- AZZALINI, A. *A very brief introduction to the skew-normal distribution*. 2008.
- BERKSON, J. Are there two regressions? *Journal of the american statistical association*, Taylor & Francis, v. 45, n. 250, p. 164–180, 1950.
- BOLFARINE, H.; CORDANI, L. K. Estimation of a structural linear regression model with a known reliability ratio. *Annals of the Institute of Statistical Mathematics*, Springer, v. 45, n. 3, p. 531–540, 1993.
- BOLFARINE, H.; RODRIGUES, J.; CORDANI, L. O modelo de regressão com erros nas variáveis. *Texto de Minicurso do X SINAPE, Sao Paulo, SP*, 1992.
- BOLSTAD, W. M.; CURRAN, J. M. *Introduction to Bayesian statistics*. [S.l.]: John Wiley & Sons, 2016.
- BUONACCORSI, J. P. *Measurement error: models, methods, and applications*. [S.l.]: Chapman and Hall/CRC, 2010.
- BUZAS, J. S.; STEFANSKI, L. A.; TOSTESON, T. D. Measurement error. *Handbook of epidemiology*, Springer, p. 1241–1282, 2014.
- CABRAL, C. R. B.; LACHOS, V. H.; ZELLER, C. B. Multivariate measurement error models using finite mixtures of skew-student t distributions. *Journal of Multivariate Analysis*, Elsevier, v. 124, p. 179–198, 2014.
- CAMERON, A. C.; TRIVEDI, P. K. et al. *Microeconometrics using stata*. [S.l.]: Stata press College Station, TX, 2009. v. 5.
- CARROLL, R. J.; ROEDER, K.; WASSERMAN, L. Flexible parametric measurement error models. *Biometrics*, Wiley Online Library, v. 55, n. 1, p. 44–54, 1999.
- CARROLL, R. J.; RUPPERT, D.; CRAINICEANU, C. M.; STEFANSKI, L. A. *Measurement error in nonlinear models: a modern perspective*. [S.l.]: Chapman and Hall/CRC, 2006.
- CHAN, L. K.; MAK, T. K. On the maximum likelihood estimation of a linear structural relationship when the intercept is known. *Journal of Multivariate Analysis*, Academic Press, v. 9, n. 2, p. 304–313, 1979.

- CHENG, C.-L.; NESS, J. V. Structural and functional models revisited. In: *Proceedings of the 2nd international workshop on total least squares techniques and errors-in-variables modeling*. SIAM, Philadelphia. [S.l.: s.n.], 1997. p. 37–50.
- CHENG, C.-L.; NESS, J. W. V. Statistical regression with measurement error: Kendall's library of statistics 6. *Arnold, London*, 1999.
- DOLBY, G. R. The ultrastructural relation: A synthesis of the functional and structural relations. *Biometrika*, Oxford University Press, v. 63, n. 1, p. 39–50, 1976.
- FULLER, W. A. *Measurement error models*. [S.l.]: John Wiley & Sons, 1987.
- GAMERMAN, D.; LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. [S.l.]: Chapman and Hall/CRC, 2006.
- GRACE, Y. Y. *Statistical Analysis with Measurement Error Or Misclassification*. [S.l.]: Springer, 2016.
- GUSTAFSON, P. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. [S.l.]: CRC Press, 2003.
- KENDALL, M.; STUART, A. The advanced theory of statistics. vol. 2: Inference and relationship. *London: Griffin, 1979, 4th ed.*, 1979.
- KENDALL, M. G. Regression, structure and functional relationship part i. *Biometrika*, JSTOR, v. 38, n. 1/2, p. 11–25, 1951.
- KENDALL, M. G. Regression, structural and functional relationship part ii. *Biometrika*, JSTOR, v. 39, n. 1/2, p. 96–108, 1952.
- LABRA, F.; AOKI, R.; GARIBAY, V.; LACHOS, V. et al. Skew-normal distribution in the multivariate null intercept measurement error model. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 25, n. 2, p. 145–170, 2011.
- LACHOS, V.; GARIBAY, V.; LABRA, F.; AOKI, R. A robust multivariate measurement error model with skew-normal/independent distributions and bayesian mcmc implementation. *Statistical Methodology*, Elsevier, v. 6, n. 5, p. 527–541, 2009.
- LACHOS, V.; LABRA, F.; BOLFARINE, H.; GHOSH, P. Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics*, Taylor & Francis, v. 44, n. 6, p. 541–556, 2010.
- PLUMMER, M.; STUKALOV, A.; DENWOOD, M.; PLUMMER, M. M. Package 'rjags'. 2019.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<http://www.R-project.org/>>.
- ROCHA, G. H.; LOSCHI, R. H.; ARELLANO-VALLE, R. B. Bayesian mismeasurement t-models for censored responses. *Statistics*, Taylor & Francis, v. 50, n. 4, p. 841–869, 2016.
- RODRIGUES, J.; BABA, M. Bayesian estimation of a simple regression model with measurement errors. *Brazilian Journal of Probability and Statistics*, JSTOR, p. 107–118, 1994.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002.

SPRENT, P. Some history of functional and structural relationships. *Contemporary Mathematics*, v. 112, p. 3–15, 1990.

STEFANSKI, L. A. The effects of measurement error on parameter estimation. *Biometrika*, Oxford University Press, v. 72, n. 3, p. 583–592, 1985.

STEFANSKI, L. A. Measurement error models. *Journal of the American Statistical Association*, Taylor & Francis, v. 95, n. 452, p. 1353–1358, 2000.

A Função de Verossimilhança

Neste apêndice encontra-se as contas para obtenção da distribuição conjunta de (y_i, w_i) , no caso de suposição normal para o erro de medida e a obtenção da função de verossimilhança do modelo proposto no Capítulo 3.

A.1 Distribuição Conjunta (y_i, w_i) no Modelo com Erro Normal

Para o modelo de regressão linear simples apresentado em (2.4), em que $x_i = w_i - u_i$, considerando as seguintes distribuições para os parâmetros: $e_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ com $\sigma_e^2 > 0$, $x_i \stackrel{iid}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ com $\mu_x \in \mathbb{R}$, $\sigma_x^2 > 0$ e $u_i \stackrel{iid}{\sim} \mathcal{N}(\mu_u, \sigma_u^2)$ com $\mu_u \in \mathbb{R}$, $\sigma_u^2 > 0$. Se a distribuição de w_i for condicionada em relação ao erro de medida, de forma que $w_i|u_i \stackrel{iid}{\sim} \mathcal{N}(\mu_x + u_i, \sigma_x^2)$, tem-se que a distribuição conjunta de (y_i, w_i) é obtida como se segue

$$\begin{aligned}
 f(y_i, w_i) &= \int f(y_i|u_i, w_i) f(w_i|u_i) f(u_i) du_i \\
 &= \int \left(\frac{1}{2\pi\sigma_e^2} \right)^{1/2} \exp \left\{ -\frac{1}{2\sigma_e^2} (y_i - (\beta_0 + \beta_1 w_i - \beta_1 u_i))^2 \right\} \times \\
 &\quad \left(\frac{1}{2\pi\sigma_x^2} \right)^{1/2} \exp \left\{ -\frac{1}{2\sigma_x^2} (w_i - (\mu_x + u_i))^2 \right\} \times \\
 &\quad \left(\frac{1}{2\pi\sigma_u^2} \right)^{1/2} \exp \left\{ -\frac{1}{2\sigma_u^2} (u_i - \mu_u)^2 \right\} du_i \\
 &= (2\pi)^{-1} \left(\frac{1}{\sigma_e \sigma_x \sigma_u} \right)^{1/2} \exp \left\{ -\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{2\sigma_e^2} - \frac{(w_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_u^2}{2\sigma_u^2} \right\} \int \left(\frac{1}{2\pi} \right)^{\frac{1}{2}} \times \\
 &\quad \exp \left\{ -\frac{1}{2} \left[u_i^2 \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_u^2} \right) + 2u_i \left(\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{w_i - \mu_x}{\sigma_x^2} - \frac{\mu_u}{\sigma_u^2} \right) \right] \right\} du_i.
 \end{aligned}$$

Seja $b = \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_u^2} \right)$ e $c = \left(\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{(w_i - \mu_x)\sigma_u^2 + \mu_u \sigma_x^2}{\sigma_x^2 \sigma_u^2} \right)$, assim,

$$\begin{aligned}
 f(y_i, w_i) &= (2\pi)^{-1} \frac{1}{\sigma_e \sigma_x \sigma_u} \exp \left\{ -\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{2\sigma_e^2} - \frac{(w_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_u^2}{2\sigma_u^2} \right\} (b^{-1})^{\frac{1}{2}} \times \\
 &\quad \int \frac{1}{2\pi} \left(\frac{1}{b^{-1}} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2b^{-1}} (u_i^2 - 2u_i c b^{-1} + c^2 (b^{-1})^2 - c^2 (b^{-1})^2) \right\} du_i \\
 &= (2\pi)^{-1} \left(\frac{1}{\sigma_e^2 \sigma_x^2 \sigma_u^2 b} \right)^{\frac{1}{2}} \exp \left\{ -\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{2\sigma_e^2} - \frac{(w_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_u^2}{2\sigma_u^2} \right\} (b^{-1})^{\frac{1}{2}} \times \\
 &\quad \exp \left\{ \frac{(c b^{-1})^2}{2b^{-1}} \right\}
 \end{aligned}$$

$$\begin{aligned}
 f(y_i, w_i) &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2b} \left[\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{\sigma_e^2} \left(b - \frac{\beta_1^2}{\sigma_e^2} \right) + \right. \right. \\
 &\quad \left. \frac{(w_i - \mu_x)^2}{\sigma_x^2} \left(b - \frac{1}{\sigma_x^2} \right) + \frac{\mu_u^2}{\sigma_u^2} \left(b - \frac{1}{\sigma_u^2} \right) - \frac{2(w_i - \mu_x)\mu_u}{\sigma_x^2 \sigma_u^2} + \right. \\
 &\quad \left. \frac{2\beta_1(y_i - \beta_0 - \beta_1 w_i)((w_i - \mu_x)\sigma_u^2 + \mu_u \sigma_x^2)}{\sigma_e^2 \sigma_x^2 \sigma_u^2} \right] \left. \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -\frac{\sigma_e^2 \sigma_x^2 \sigma_u^2}{2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))} \times \right. \\
 &\quad \left[\frac{(y_i - \beta_0 - \beta_1 w_i)^2 (\sigma_x^2 + \sigma_u^2) - (w_i - \mu_x)^2 (\beta_1^2 \sigma_u^2 + \sigma_e^2) - \mu^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2)}{\sigma_e^2 \sigma_x^2 \sigma_u^2} \right. \\
 &\quad \left. \left. - \frac{2(w_i - \mu_x)\mu_u \sigma_e^2 + 2\beta_1(y_i - \beta_0 - \beta_1 w_i)((w_i - \mu_x)\sigma_u^2 + \mu_u \sigma_x^2)}{\sigma_e^2 \sigma_x^2 \sigma_u^2} \right] \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -[2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_u^2 - \sigma_x^2))]^{-1} \times \right. \\
 &\quad \left. [(y_i - \beta_0 - \beta_1 w_i)^2 (\sigma_x^2 + \sigma_u^2) + (w_i - \mu_x)^2 (\beta_1^2 \sigma_u^2 + \sigma_e^2) + \mu_u^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) \right. \\
 &\quad \left. - 2(w_i - \mu_x)\mu_u \sigma_e^2 + 2\beta_1(y_i - \beta_0 - \beta_1 w_i)((w_i - \mu_x)\sigma_u^2 + \mu_u \sigma_x^2)] \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -[2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_u^2 - \sigma_x^2))]^{-1} \times \right. \\
 &\quad \left. [(y_i - \beta_0)^2 (\sigma_x^2 + \sigma_u^2) - 2(y_i - \beta_0)\beta_1 w_i (\sigma_x^2 + \sigma_u^2) + \beta_1^2 w_i^2 (\sigma_x^2 + \sigma_u^2) + \right. \\
 &\quad \left. w_i^2 (\beta_1^2 \sigma_u^2 + \sigma_e^2) + 2w_i \mu_x (\beta_1^2 \sigma_u^2 + \sigma_e^2) + \mu_x^2 (\beta_1^2 \sigma_u^2 + \sigma_e^2) + \mu_u^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) - \right. \\
 &\quad \left. 2(w_i - \mu_x)\mu_u \sigma_e^2 + 2\beta_1(y_i - \beta_0)(w_i \sigma_u^2 - \mu_x \sigma_u^2 + \mu_u \sigma_x^2) - 2\beta_1^2 w_i^2 \sigma_u^2 - \right. \\
 &\quad \left. 2\beta_1^2 w_i (\mu_u \sigma_x^2 - \mu_x \sigma_u^2)] \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -[2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_u^2 - \sigma_x^2))]^{-1} \times \right. \\
 &\quad \left. [(y_i - \beta_0)^2 (\sigma_x^2 + \sigma_u^2) + w_i^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) + \mu_x^2 (\beta_1^2 \sigma_u^2 + \sigma_e^2) + \mu_u^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) + \right. \\
 &\quad \left. 2\mu_x \mu_u \sigma_e^2 - 2w_i (\beta_1 \mu \sigma_x^2 + \mu \sigma_e^2 + \mu_u \sigma_e^2) - 2\beta_1(y_i - \beta_0)(w_i \sigma_x^2 + \mu_x \sigma_u^2 - \mu_u \sigma_x^2)] \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -[2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_u^2 - \sigma_x^2))]^{-1} \times \right. \\
 &\quad \left. [(y_i - \beta_0)^2 (\sigma_x^2 + \sigma_u^2) + w_i^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) + \mu_x^2 (\beta_1^2 \sigma_x^2 + \beta_1^2 \sigma_x^2 - 2\beta_1^2 \sigma_x^2 + \beta_1^2 \sigma_u^2 + \sigma_e^2) + \right. \\
 &\quad \left. \mu_u^2 (\beta_1^2 \sigma_x^2 + \sigma_e^2) + 2\mu_x \mu_u \sigma_e^2 + 2w_i (\beta_1^2 \sigma_x^2 \mu_x - \beta_1^2 \sigma_x^2 \mu_x - \beta_1 \mu \sigma_x^2 - \mu \sigma_e^2 - \mu_u \sigma_e^2) - \right. \\
 &\quad \left. 2\beta_1(y_i - \beta_0)(w_i \sigma_x^2 + \mu_x \sigma_u^2 - \mu_u \sigma_x^2 + \mu_x \sigma_x^2 - \mu_x \sigma_x^2) - 2\beta_1^2 \sigma_x^2 \mu_x \mu_u + 2\beta_1^2 \sigma_x^2 \mu_x \mu_u] \right\} \\
 &= (2\pi)^{-1} [(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_x^2 + \sigma_u^2))]^{-\frac{1}{2}} \exp \left\{ -\frac{(\beta_1^2 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)}{2(\beta_1^2 \sigma_x^2 \sigma_u^2 + \sigma_e^2 (\sigma_u^2 - \sigma_x^2))} \times \right. \\
 &\quad \frac{1}{[(\beta_1^2 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)]^{-1}} \left[\frac{(y_i - \beta_0)^2 - 2(y_i - \beta_0)\beta_1 \mu_x + \beta_1^2 \mu_x^2}{\beta_1 \sigma_x^2 + \sigma_e^2} + \right. \\
 &\quad \frac{w_i^2 - 2w_i(\mu_x + \mu_u) + (\mu_x + \mu_u)^2}{(\sigma_x^2 + \sigma_u^2)} - \frac{2\beta_1 \sigma_x^2 (y_i - \beta_0)(w_i - \mu_x - \mu_u)}{(\beta_1 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)} + \\
 &\quad \left. \left. \frac{2\beta_1 \sigma_x^2 (\beta_1 \mu_x)(w_i - \mu_x - \mu_u)}{(\beta_1 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)} \right] \right\}.
 \end{aligned}$$

Para,

$$\rho = \frac{\beta_1 \sigma_x^2}{[(\beta_1^2 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)]^{\frac{1}{2}}},$$

tem-se que:

$$f(y_i, w_i) = \frac{1}{2\pi} \left[\frac{1}{(\beta_1^2 \sigma_x^2 + \sigma_e^2)(\sigma_x^2 + \sigma_u^2)(1 - \rho^2)} \right]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \times \left[\frac{(y_i - (\beta_0 + \beta_1 \mu_x))^2}{\beta_1^2 \sigma_x^2 + \sigma_e^2} + \frac{(w_i - (\mu_x + \mu_u))^2}{\sigma_x^2 + \sigma_u^2} - 2\rho \left(\frac{(y_i - (\beta_0 + \beta_1 \mu_x))}{\sqrt{\beta_1^2 \sigma_x^2 + \sigma_e^2}} \right) \times \left(\frac{(w_i - (\mu_x + \mu_u))}{\sqrt{\sigma_x^2 + \sigma_u^2}} \right) \right] \right\}. \quad (\text{A.1})$$

Assim, a distribuição conjunta de $(w_i, y_i)^t$ é dada pela seguinte distribuição normal bivariada com vetor de médias e matriz de covariâncias dados por:

$$\begin{pmatrix} y_i \\ w_i \end{pmatrix} \underset{\text{ind}}{\sim} \mathcal{N}_2 \left[\begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_u \end{pmatrix}; \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right]. \quad (\text{A.2})$$

Denote por $\phi_n(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi_n(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectivamente, a fdp e a fda de uma variável aleatória com distribuição normal n -variada com vetor de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$ e por $\phi(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ quando a distribuição for univariada. Assim, $(y_i, w_i)^t$ dada por (A.2) será representada por $\phi_2((y_i, w_i)^t | \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, em que,

$$\boldsymbol{\mu}^* = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_u \end{pmatrix} \text{ e } \boldsymbol{\Sigma}^* = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

Note-se que quando μ_u é igual a zero (A.2) a expressão retorna ao modelo apresentado em (2.6).

A.2 Contas da Função de Verossimilhança para o Modelo Proposto

Considere o modelo dado em (3.7), e assuma, neste caso, a função de verossimilhança $f(\mathbf{y}, \mathbf{w})$ do modelo proposto é obtida pela integração de $f(y_i, w_i, u_i)$ em relação a u_i , de forma que

$$f(\mathbf{y}, \mathbf{w}) = \prod_{i=1}^n \sum_{k=0}^2 p_k \int_{\Omega_{u_i}} f(y_i | w_i, u_i) f(w_i | u_i) f_k(u_i | \gamma_k) du_i,$$

substituindo pelas respectivas funções do modelo, tem-se que

$$\begin{aligned}
 f(\mathbf{y}, \mathbf{w}) &= \prod_{i=1}^n p_0 \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i), \sigma_e^2) \phi(w_i|\mu_x + u_i, \sigma_x^2) \phi(u_i|\sigma_0^2) du_i + \\
 &\quad \prod_{i=1}^n p_1 \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i), \sigma_e^2) \phi(w_i|\mu_x + u_i, \sigma_x^2) \phi(u_i|\mu_1, \sigma_1^2, \lambda_1) \times \\
 &\quad \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i + \prod_{i=1}^n p_2 \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i), \sigma_e^2) \times \\
 &\quad \phi(w_i|\mu_x + u_i, \sigma_x^2) \phi(u_i|\mu_1, \sigma_2^2, \lambda_2) \Phi\left(\lambda_2 \left(\frac{u_i - \mu_2}{\sigma_2}\right)\right) du_i \\
 &= \prod_{i=1}^n p_0 I_0 + \prod_{i=1}^n p_1 I_1 + \prod_{i=1}^n p_2 I_2. \tag{A.3}
 \end{aligned}$$

Resolvendo as integrais em (A.3) separadas tem-se

$$\begin{aligned}
 I_0 &= \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i = 0), \sigma_e^2) \phi(w_i|\mu_x + u_i = 0, \sigma_x^2) du_i \\
 &= \phi(y_i|\beta_0 + \beta_1 w_i, \sigma_e^2) \phi(w_i|\mu_x, \sigma_x^2) \\
 &= \left(\frac{1}{2\pi\sigma_e^2\sigma_x^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{(y_i - (\beta_0 + \beta_1 w_i))^2}{\sigma_e^2} + \frac{(w_i - \mu_x)^2}{\sigma_x^2}\right]\right\} \\
 &= \left(\frac{1}{2\pi\sigma_e^2\sigma_x^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{y_i^2 - 2y_i\beta_0 - 2y_i w_i \beta_1 + \beta_0^2 + w_i^2 \beta_1^2}{\sigma_e^2} + \frac{w_i^2 - 2w_i\mu_x + \mu_x^2}{\sigma_x^2}\right]\right\}.
 \end{aligned}$$

A integral I_1 pode ser reescrita de forma que

$$\begin{aligned}
 I_1 &= \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i), \sigma_e^2) \phi(w_i|\mu_x + u_i, \sigma_x^2) 2\phi(u_i|\mu_1, \sigma_1^2) \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i \\
 &= \int_{\Omega_{u_i}} \phi(y_i|\beta_0 + \beta_1(w_i - u_i), \sigma_e^2) \phi(w_i|\mu_x + u_i, \sigma_x^2) 2\phi(u_i|\mu_1, \sigma_1^2) \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i \\
 &= (2\pi)^{-1} \frac{1}{\sigma_e \sigma_x \sigma_1} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{2\sigma_e^2} - \frac{(w_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_1^2}{2\sigma_1^2}\right\} (b_1^{-1})^{\frac{1}{2}} \times \\
 &\quad \int_{\Omega_{u_i}} \frac{1}{2\pi} \left(\frac{1}{b_1^{-1}}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2b_1^{-1}}(u_i^2 - 2u_i c_1 b_1^{-1} + c_1^2 (b_1^{-1})^2 - c^2 (b_1^{-1})^2)\right\} \times \\
 &\quad \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i \\
 &= (2\pi)^{-1} \left(\frac{1}{\sigma_e^2 \sigma_x^2 \sigma_1^2 b_1}\right)^{\frac{1}{2}} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 w_i)^2}{2\sigma_e^2} - \frac{(w_i - \mu_x)^2}{2\sigma_x^2} - \frac{\mu_u^2}{2\sigma_u^2}\right\} \times \\
 &\quad \exp\left\{\frac{(c_1 b_1^{-1})^2}{2b_1^{-1}}\right\} \int_{\Omega_{u_i}} 2\phi(u_i|c_1 b_1^{-1}, b_1^{-1}) \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i \\
 &= \phi_2((y_i, w_i)^t | \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_1^*) \int_{\Omega_{u_i}} 2\phi(u_i|c_1 b_1^{-1}, b_1^{-1}) \Phi\left(\lambda_1 \left(\frac{u_i - \mu_1}{\sigma_1}\right)\right) du_i.
 \end{aligned}$$

em que,

$$\begin{aligned}
 \boldsymbol{\mu}_1^* &= \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1^* = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_1^2 \end{pmatrix}, \\
 b_1 &= \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_1^2}\right), \quad c_1 = \left(\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{(w_i - \mu_x)\sigma_1^2 + \mu_1 \sigma_x^2}{\sigma_x^2 \sigma_1^2}\right).
 \end{aligned}$$

Lema A.2.1. *Seja $U \sim \mathcal{N}_n(\mathbf{c}, \mathbf{C})$. Então para algum vetor fixo \mathbf{a} de dimensão n e alguma matriz fixa \mathbf{B} de dimensão $n \times k$, tem-se*

$$E[\Phi_n(\mathbf{B}U + \mathbf{a}|\mathbf{d}, \mathbf{D})] = \Phi_n(\mathbf{B}\mathbf{c} + \mathbf{a}|\mathbf{d}, \mathbf{D} + \mathbf{B}\mathbf{C}\mathbf{B}^T).$$

Assim,

$$I_1 = 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_1^*) \Phi \left(\lambda_1 \left(\frac{c_1 b_1^{-1} - \mu_1}{\sigma_1} \right) \middle| 0, 1 + \frac{\lambda_1^2}{b_1 \sigma_1^2} \right).$$

De forma similar a I_1 , o I_2 é obtido substituindo os parâmetros da *skew-normal* M_1 pelos da *skew-normal* M_2 , dessa forma

$$I_2 = 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_2^*, \boldsymbol{\Sigma}_2^*) \Phi \left(\lambda_2 \left(\frac{c_2 b_2^{-1} - \mu_2}{\sigma_2} \right) \middle| 0, 1 + \frac{\lambda_2^2}{b_2 \sigma_2^2} \right),$$

em que,

$$\boldsymbol{\mu}_2^* = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2^* = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_2^2 \end{pmatrix},$$

$$b_2 = \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_2^2} \right), \quad c_2 = \left(\frac{\beta_1 (y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{(w_i - \mu_x) \sigma_1^2 + \mu_2 \sigma_x^2}{\sigma_x^2 \sigma_2^2} \right).$$

Portanto, a função de verossimilhança do modelo proposto em (3.12), considerando o erro de medida modelado por uma mistura de *skew-normal* é dada por

$$\begin{aligned} f(\mathbf{y}, \mathbf{w}) &= \prod_{i=1}^n (p_0 \phi(y_i | \beta_0 + \beta_1 w_i, \sigma_e^2) \phi(w_i | \mu_x, \sigma_x^2))^{\Delta_0} \times \\ &\quad (p_1 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_1^*) \Phi(\lambda(c_1 b_1^{-1} - \mu_1) \sigma_1^{-1} | 0, 1 + \lambda^2 b_1^{-1} \sigma_1^2))^{\Delta_1} \times \\ &\quad (p_2 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_2^*, \boldsymbol{\Sigma}_2^*) \Phi(\lambda(c_2 b_2^{-1} - \mu_2) \sigma_2^{-1} | 0, 1 + \lambda^2 b_2^{-1} \sigma_2^2))^{\Delta_2}, \end{aligned}$$

em que, Δ_0 , Δ_1 e Δ_2 as partes associadas aos componentes sem erros, aos erros positivos e negativos, respectivamente.

Função de verossimilhança com a *half-normal*: Para o caso em que $\lambda \rightarrow +\infty$ e $\lambda \rightarrow -\infty$ as distribuições do erro positivo e do erro negativo são dadas pela distribuição *half-normal* com função de verossimilhança dada por

$$\begin{aligned} f(\mathbf{y}, \mathbf{w}) &= \prod_{i=1}^n [p_0 \phi(y_i | \beta_0 + \beta_1 w_i, \sigma_e^2) \phi(w_i | \mu_x, \sigma_x^2)]^{\Delta_0} \times \\ &\quad \left[p_1 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_1^*) \left(1 - \Phi \left((\mu_1 - c_1) b_1^{-\frac{1}{2}} \right) \right) \right]^{\Delta_1} \times \\ &\quad \left[p_2 2\phi_2((y_i, w_i)^t | \boldsymbol{\mu}_2^*, \boldsymbol{\Sigma}_2^*) \Phi \left((\mu_2 - c_2) b_2^{-\frac{1}{2}} \right) \right]^{\Delta_2}, \end{aligned}$$

em que,

$$\boldsymbol{\mu}_1^* = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1^* = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_1^2 \end{pmatrix},$$

$$\boldsymbol{\mu}_2^* = \begin{pmatrix} \beta_0 + \beta_1 \mu_x \\ \mu_x + \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2^* = \begin{pmatrix} \beta_1^2 \sigma_x^2 + \sigma_e^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_2^2 \end{pmatrix},$$

$$b_1 = \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_1^2} \right), \quad c_1 = \left(\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{(w_i - \mu_x)\sigma_1^2 + \mu_1 \sigma_x^2}{\sigma_x^2 \sigma_1^2} \right),$$

$$b_2 = \left(\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} + \frac{1}{\sigma_2^2} \right), \quad c_2 = \left(\frac{\beta_1(y_i - \beta_0 - \beta_1 w_i)}{\sigma_e^2} - \frac{(w_i - \mu_x)\sigma_1^2 + \mu_2 \sigma_x^2}{\sigma_x^2 \sigma_2^2} \right).$$

Δ_0 , Δ_1 e Δ_2 as partes associadas aos componentes sem erros, aos erros positivos e negativos, respectivamente.

B Distribuição *Half*-Normal

Conforme definido por Azzalini (1985), uma variável aleatória \mathcal{Z} tem distribuição *skew-normal* com parâmetro de locação μ , parâmetro de escala σ^2 e parâmetro de assimetria $\lambda \in \mathbb{R}$, representado por $\mathcal{Z} \sim \mathcal{SN}(\mu, \sigma^2, \lambda)$ e sua função de densidade de probabilidade é tal que

$$\phi_{SN}(z|\mu, \sigma^2, \lambda) = 2\phi(z|\mu, \sigma^2)\Phi(\lambda(z - \mu)/\sigma), \quad (\text{B.1})$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ são respectivamente, a fdp e a fda de uma variável aleatória com distribuição normal quando a distribuição é uni-variada. Note-se que, quando $\lambda = 0$ tem-se a distribuição normal. A média e a variância de \mathcal{Z} são respectivamente dadas por

$$E[\mathcal{Z}] = \mu + \sigma\sqrt{\frac{2}{\pi}}\delta, \quad \text{Var}[\mathcal{Z}] = \sigma^2\left(1 - \frac{2\delta^2}{\pi}\right), \quad (\text{B.2})$$

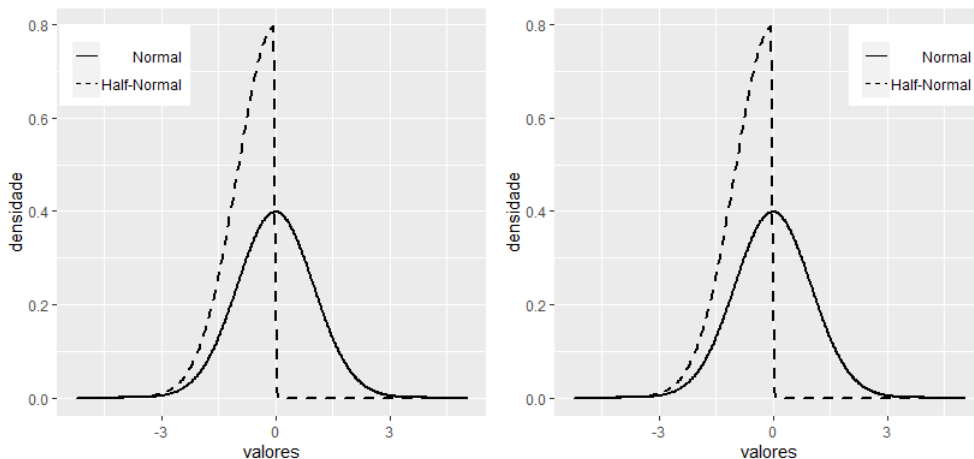
para $\delta = \lambda/\sqrt{1 + \lambda^2}$.

A distribuição *skew-normal* possui algumas propriedades importantes, estudadas por Azzalini (2008), que permitem que ela se relacione com outras distribuições, como é o caso das duas propriedades citadas a seguir.

1. Se $\mathcal{Z} \sim \mathcal{SN}(\mu, \sigma, 0)$ então $\mathcal{Z} \sim \mathcal{N}(\mu, \sigma)$. Nota-se que isso ocorre em (B.1) quando $\Phi(0) = 1/2$.
2. Se $\mathcal{Z} \sim \mathcal{SN}(\mu, \sigma, \lambda)$, $\lambda \rightarrow \infty$ então $\mathcal{Z} \xrightarrow{D} \mathcal{HN}(\mu, \sigma)$. Isso ocorre pois $\Phi(\lambda z) \rightarrow 1$ se $z > 0$ e $\Phi(\lambda z) \rightarrow 0$ se $z < 0$.

A segunda propriedade apresenta a distribuição *half-normal* que depende do valor escolhido para λ obtém-se a *half-normal*. A Figura 7 mostra o comportamento da fdp de uma variável com distribuição normal e distribuição *half-normal* definida na Expressão (B.3). Se $\mathcal{Z} \sim \mathcal{N}(0, 1)$

Figura 7 – Densidades das distribuições normal e *half-normal*.



então $M = \mu + \sigma|\mathcal{Z}|$ e M tem distribuição *half*-normal com parâmetro de locação e escala dados por μ e σ então sua função de densidade é expressa por

$$f(M) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (m - \mu)^2 \right\}, \quad m > \mu, \quad -\infty < \mu < \infty, \quad \sigma > 0. \quad (\text{B.3})$$

A média e a variância são respectivamente dadas por

$$E[M] = \mu + \sigma \sqrt{\frac{2}{\pi}}, \quad \text{Var}[M] = \sigma^2 \left(1 - \frac{2}{\pi} \right). \quad (\text{B.4})$$

Se M tem uma distribuição *half*-normal negativa $M = \mu - \sigma|\mathcal{Z}|$ a inferência da variável é feita de forma que $-M \sim \mathcal{HN}(-\mu, \sigma)$.

C Resultados de Simulação

Neste apêndice é apresentado as tabelas com os resultados do estudo de simulação da Seção 4.2 contendo os cenários não apresentados nessa seção.

Cenário 1: erro de medida simétrico e pesos iguais a $(0,4;0,3;0,3)$.

Tabela 16 – Resultados do estudo de simulação com $n=1000$, cenário 1.

n	Parâmetros	Valor Real	Média	Vício	SD	EQM	Cobertura HPD
1000	β_1	0,50	0,4994	-0,0006	0,0059	0,0000	0,99
	μ_x	2,00	2,0247	0,0247	0,0125	0,0008	1,00
	p_0	0,40	0,4131	0,0131	0,0352	0,0014	0,95
	p_1	0,30	0,2827	-0,0173	0,0180	0,0006	0,91
	p_2	0,30	0,3042	0,0042	0,0206	0,0004	0,99
	σ_x^2	1,50	1,5729	0,0729	0,0350	0,0065	1,00
	σ_e^2	0,09	0,1018	0,0118	0,0087	0,0002	0,97
	σ_1^2	1,50	1,7398	0,2398	0,1321	0,0750	0,95
	σ_2^2	1,50	1,5037	0,0037	0,1318	0,0174	1,00

Tabela 17 – Classificação dos componentes da mistura para n igual a 1000, cenário 1.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
1000	Sem erro	0,30	0,05	0,05
	Erro positivo	0,10	0,18	0,01
	Erro negativo	0,11	0,00	0,20

Cenário 2: erro de medida simétrico com todos os dados mensurados com erros e pesos iguais a $(0;0,5;0,5)$.

Tabela 18 – Resultados do estudo de simulação com $n=1000$, cenário 2.

n	Parâmetros	Valor Real	Média	viés	SD	EQM	Cobertura HPD
1000	β_1	0,50	0,4951	-0,0049	0,0822	0,0001	1,00
	μ_x	2,00	2,0536	0,0536	0,1423	0,0033	0,97
	p_0	0,00	0,0222	0,0222	0,1733	0,0014	0,00
	p_1	0,50	0,4966	-0,0034	0,1393	0,0004	0,86
	p_2	0,50	0,4812	-0,0188	0,1447	0,0008	0,74
	σ_x^2	1,50	1,5368	0,0368	0,2224	0,0038	0,99
	σ_e^2	0,09	0,1216	0,0316	0,1109	0,0011	0,47
	σ_1^2	1,30	1,1654	-0,1346	0,2786	0,0242	0,83
	σ_2^2	1,30	1,2431	-0,0569	0,3178	0,0134	0,86

Tabela 19 – Classificação dos componentes da mistura para n igual a 1000, cenário 2.

n	Posição real	Posição estimada		
		Sem erro	Erro positivo	Erro negativo
1000	Sem erro	0,00	0,04	0,00
	Erro positivo	0,01	0,40	0,10
	Erro negativo	0,00	0,10	0,40