

Modelagem probabilística do índice $IGPM_{12}$

Samantha G. Oliveira^{1 †}, Luciano G. Batista¹, Eduardo C. Barbosa¹, Paulo C. Emiliano¹,
Maurício S. Lacerda¹, Kamila A. de Oliveira¹

¹*Departamento de Estatística da Universidade Federal de Viçosa - UFV*

Resumo: *O objetivo desse trabalho foi ajustar uma distribuição aos dados de $IGPM_{12}$ e estimar a probabilidade deste índice assumir determinados valores que podem ser prejudiciais para pessoas físicas, jurídicas e investidores que possuem compromissos ou ativos financeiros cuja remuneração é atrelada ao $IGPM$. Para escolher entre as distribuições normal, cauchy, logística e LS t-Student foram analisados o teste de Kolmogorov-Smirnov e o Critério de Informação de Akaike (AIC). O modelo LS t Student com parâmetros $\hat{\mu} \approx 7,29$, $\hat{\sigma} \approx 3,99$ e $\hat{\nu} \approx 2,20$ foi selecionado e com ele estimou-se as probabilidades de interesse. Concluiu-se que o cenário mais provável ($p \approx 62\%$) envolvia um reajuste desse índice entre 0% e 10% para os próximos 12 meses.*

Palavras-chave: *Assimetria; curtose; máxima verossimilhança.*

Probabilistic modeling of the $IGPM_{12}$ index

Abstract: *The objective of this work was to adjust a distribution to the $IGPM_{12}$ data and estimate the probability of this index assuming certain values that may be specific to individuals, legal entities and investors who have commitments or financial investments whose investors are linked to the $IGPM$. To choose between the normal, Cauchy, logistic and LS t-Student distributions, the Kolmogorov-Smirnov test and the Akaike Information Criterion (AIC) were analyzed. The LS t Student model with parameters $\hat{\mu} \approx 7.29$, $\hat{\sigma} \approx 3.99$ and $\hat{\nu} \approx 2.20$ was selected and with it the probabilities of interest were estimated. It is concluded that the most likely scenario ($p \approx 62\%$) involves an adjustment of this index between 0% and 10% for the next 12 months.*

Keywords: *Asymmetry; kurtosis; maximum likelihood.*

Introdução

O Índice Geral de Preços de Mercado, conhecido como $IGPM$, é um indicador econômico calculado e divulgado mensalmente pela Fundação Getúlio Vargas (FGV-IBRE, 2001). Sua composição é uma média ponderada de três índices inflacionários: 60% do IPA (Índice de Preços ao Produtor Amplo), 30% do IPC (Índice de Preços ao Consumidor) e 10% do $INCC$ (Índice Nacional de Custo da Construção). Essa ponderação, realizada através de 3 diferentes índices, reflete variações importantes da economia brasileira, o que torna o $IGPM$ uma medida de inflação mais abrangente e até mesmo mais real do que o $IPCA$. Por esse motivo tal índice costuma ser utilizado para reajustar contratos de aluguel e receitas de empresas dos setores de energia elétrica, saneamento, petróleo e gás, entre outros (De Souza, 2004).

Adicionalmente, o $IGPM$ possui alta volatilidade (Mattos, 1986). Isso implica que os retornos desse índice possuem variações significativas em torno da média, o que torna a probabilidade de valores extremos ocorrerem razoavelmente alta. Nesse sentido, o presente trabalho visa ajustar um modelo probabilístico aos dados de $IGPM$ e estimar algumas probabilidades de interesse, como por exemplo: $P(\text{DEFLAÇÃO}) = P(IGPM < 0\%)$, $P(0\% \leq IGPM < 10\%)$, $P(10\% \leq IGPM < 20\%)$ e $P(IGPM \geq 20\%)$.

Os resultados obtidos poderão ser utilizados, principalmente, para que pessoas físicas (PF), empresários (PJ) e até mesmo investidores especializados (*stakeholders*) possam ter uma ideia

[†]Autor correspondente: samantha.gouvea@ufv.br

geral do quão provável seria a ocorrência de um reajuste extremo em seus contratos de aluguel/receitas, investimentos, dívidas ou qualquer outro tipo de compromisso financeiro cuja remuneração esteja atrelada ou indexada ao *IGPM*.

Referencial teórico

Coefficiente de assimetria

O terceiro momento em relação à média, denotado por μ_3 , expressa-se como a expectativa da diferença cúbica entre uma variável aleatória X e sua média $E(X)$, representado por $E[X - E(X)]^3$. Esse conceito é frequentemente referido como medida de assimetria ou simplesmente assimetria, conforme destacado por Mood et al. (1974). O valor do terceiro momento oferece *insights* cruciais sobre a natureza e a configuração da distribuição de probabilidade dos dados.

Em particular, uma distribuição simétrica exibe um terceiro momento (μ_3) igual a zero, enquanto distribuições assimétricas podem revelar um terceiro momento distinto de zero. Mais precisamente, uma distribuição com assimetria à esquerda é caracterizada por um terceiro momento negativo em relação à média, ao passo que uma distribuição assimétrica à direita é marcada por um terceiro momento positivo em relação à média. Uma métrica adimensional, o coeficiente de assimetria $\gamma_1 = \frac{\mu_3}{\sigma^3}$, é comumente utilizada para avaliar a assimetria de uma distribuição.

Um estimador para o coeficiente de assimetria (γ_1) é formulado como:

$$\hat{\gamma}_1 = \frac{\hat{\mu}^3}{\hat{\sigma}^3} \quad \text{em que} \quad \hat{\mu}^3 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3 \quad \text{e} \quad \hat{\sigma}^3 = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{3}{2}}.$$

Coefficiente de curtose

O quarto momento em relação à média, representado por μ_4 , desempenha um papel crucial na expressão do grau de achatamento dos dados, ou em outras palavras, na ponderação das caudas de uma distribuição. A prática padrão de normalização estabelece que a curtose de uma distribuição normal seja três, e o achatamento excessivo seja zero. O coeficiente de curtose pode ser expresso como $\gamma_2 = \frac{\mu_4}{\sigma^4}$ ou, de forma equivalente, $\gamma_3 = \frac{\mu_4}{\sigma^4} - 3$. Neste estudo, optamos pela primeira opção, que é a forma como o *software* livre R (R, 2023) estima esse coeficiente por meio do pacote *moments*. Para interpretar o coeficiente de curtose, podemos concluir que uma distribuição pode ser classificada como:

- i) Platicúrtica ($\gamma_2 < 3$): Distribuição com caudas mais suaves e um pico mais achatado em comparação com a distribuição normal;
- ii) Mesocúrtica ($\gamma_2 = 3$): Distribuição com caudas e pico semelhantes aos da distribuição normal;
- iii) Leptocúrtica ($\gamma_2 > 3$): Distribuição com caudas mais pesadas e um pico mais alto e estreito em comparação com a distribuição normal.

Uma maneira para estimar o coeficiente de curtose é apresentada a seguir:

$$\hat{\gamma}_2 = \frac{\hat{\mu}^4}{\hat{\sigma}^4} \quad \text{no qual} \quad \hat{\mu}^4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4 \quad \text{e} \quad \hat{\sigma}^4 = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2.$$

Distribuição Normal

A distribuição normal (ou gaussiana) é a distribuição mais importante e também a mais utilizada na estatística. Muitos fenômenos naturais se associam a sua forma e diversos conceitos teóricos são elaborados a partir do seu conhecimento (por exemplo, o Teorema Central do Limite). Sua função densidade de probabilidade, cuja forma é de sino, é definida a seguir:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty,$$

em que μ e σ^2 representam, respectivamente, a média (parâmetro de posição) e a variância (parâmetro de escala) da distribuição. Destaca-se que quando $\mu = 0$ e $\sigma^2 = 1$ a distribuição normal é conhecida como distribuição normal padrão $[N(0, 1)]$. Logo, se X é uma variável aleatória contínua com distribuição Normal(μ, σ^2), então:

$$E(X) = \mu \quad \text{e} \quad V(X) = \sigma^2.$$

A função suporte $[\ell(\mu; \sigma^2; \mathbf{x})]$ do modelo normal é:

$$\ell(\mu; \sigma^2; \mathbf{x}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Os estimadores de máxima verossimilhança para os parâmetros μ e σ^2 do modelo normal podem ser definidos analiticamente, sendo, respectivamente:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Distribuição Cauchy

A função densidade de probabilidade da distribuição Cauchy (ou de Lorentz) é definida como:

$$f(x) = \frac{1}{\pi\sigma \left[1 + \left(\frac{x-\theta}{\sigma} \right)^2 \right]} = \frac{1}{\pi} \left[\frac{\sigma}{(x-\theta)^2 + \sigma^2} \right], \quad -\infty < x < \infty,$$

em que θ e σ representam, respectivamente, os parâmetros de locação e escala da distribuição. Uma particularidade desse modelo probabilístico é que a sua média e variância não são definidas (ou divergem).

$$E(X) \rightarrow \infty \quad \text{e} \quad V(X) \rightarrow \infty.$$

Destaca-se que a distribuição Cauchy padrão pode ser obtida a partir da razão entre duas variáveis aleatórias independentes e com distribuição normal padrão. Ou seja:

$$U \sim N(0, 1) \quad \text{e} \quad V \sim N(0, 1) \quad \implies \quad X = \frac{U}{V} \sim \text{Cauchy}(0, 1).$$

Adicionalmente, quando $\theta = 0$ e $\sigma = 1$ a distribuição Cauchy converge para uma distribuição t Student padrão, com apenas 1 grau de liberdade ou $\nu = 1$:

$$\text{Cauchy}(0, 1) \sim t_1.$$

A função suporte $[\ell(\theta; \sigma^2; \mathbf{x})]$ para a distribuição Cauchy é:

$$\ell(\theta; \sigma; \mathbf{x}) = -n \ln(\pi\sigma) - \sum_{i=1}^n \ln \left[1 + \left(\frac{x_i - \theta}{\sigma} \right)^2 \right].$$

No entanto, os estimadores de máxima verossimilhança para os parâmetros θ e σ do modelo Cauchy são tradicionalmente obtidos via métodos numéricos, utilizados para resolver o seguinte sistema de equações:

$$\begin{cases} \frac{\partial \ell(\theta; \sigma; \mathbf{x})}{\partial \theta} = 0 \\ \frac{\partial \ell(\theta; \sigma; \mathbf{x})}{\partial \sigma} = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^n \frac{2(x_i - \hat{\theta})}{\hat{\sigma}^2 + (x_i - \hat{\theta})^2} = 0 \\ \sum_{i=1}^n \frac{2(x_i - \hat{\theta})^2}{\hat{\sigma}[\hat{\sigma}^2 + (x_i - \hat{\theta})^2]} - \frac{n}{\hat{\sigma}} = 0 \end{cases}.$$

Distribuição Logística

Desenvolvida no início dos anos de 1800 pelo professor Pierre François Verhulst, a distribuição logística é contínua e possui caudas mais pesadas do que a de uma distribuição normal (assim como a distribuição t-Student). Sua função de distribuição acumulada $[F(\cdot)]$ é utilizada para o desenvolvimento do clássico modelo de regressão logística. A função densidade de probabilidade desse modelo probabilístico é descrita a seguir:

$$f(x) = \frac{e^{-\left(\frac{x-\mu}{\sigma}\right)}}{\sigma \left[1 + e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]}, \quad -\infty < x < \infty,$$

Com μ e σ representando, respectivamente, os parâmetros de localização e escala dessa distribuição. Adicionalmente:

$$E(Y) = \mu \quad \text{e} \quad V(Y) = \frac{\pi^2 \sigma^2}{3}.$$

A função suporte $[\ell(\mu; \sigma; \mathbf{x})]$ para a distribuição logística é:

$$\ell(\mu; \sigma; \mathbf{x}) = \frac{n\mu}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n x_i - n \ln(\sigma) - 2 \sum_{i=1}^n \ln \left[1 + e^{-\left(\frac{x_i - \mu}{\sigma}\right)} \right].$$

Já os estimadores de máxima verossimilhança para os parâmetros μ e σ do modelo logístico são obtidos via métodos numéricos, utilizados para resolver o seguinte sistema de equações:

$$\begin{cases} \frac{\partial \ell(\mu; \sigma; \mathbf{x})}{\partial \mu} = 0 \\ \frac{\partial \ell(\mu; \sigma; \mathbf{x})}{\partial \sigma} = 0 \end{cases} \Rightarrow \begin{cases} \frac{1}{\hat{\sigma}} \left[n - 2 \sum_{i=1}^n \frac{L_i}{1+L_i} \right] = 0 \\ \frac{1}{\hat{\sigma}^2} \left[\frac{2 \sum_{i=1}^n (x_i - \hat{\mu}) L_i}{1+L_i} \right] - n \hat{\mu} + \sum_{i=1}^n x_i = 0 \end{cases}, \quad \text{com} \quad L_i = e^{-\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)}.$$

Distribuição *Location-Scale t-Student*

A distribuição t-Student padrão, ou apenas t-Student, é obtida a partir da seguinte relação entre duas variáveis aleatórias contínuas (Z e V) e independentes:

$$X = \frac{Z}{\sqrt{\frac{V}{\nu}}} \sim t_\nu, \quad \text{em que} \quad Z \sim N(0, 1) \quad \text{e} \quad V \sim \chi_\nu^2.$$

A função densidade de probabilidade da distribuição t-Student é dada por:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad -\infty < x < \infty,$$

com $\Gamma(u) = \int_0^\infty t^{u-1} e^{-t} dt$, para $u > 0$. Tem-se ainda que:

$$E(X) = 0 \quad \text{e} \quad V(X) = \frac{\nu}{\nu - 2}, \quad \text{para } \nu > 2.$$

Em algumas situações, no entanto, há a necessidade de modelar determinado fenômeno utilizando uma distribuição t -Student que não seja a padrão. Por isso o modelo *Location-Scale t-Student* ou LS t -Student é empregado. Para sua definição, admita a relação estocástica:

$$Y = \mu + \sigma X \sim t_\nu(\mu, \sigma^2), \quad \text{sendo } X \sim t_\nu.$$

Logo, a função densidade de probabilidade da distribuição LS t -Student é:

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\sigma^2} \Gamma(\frac{\nu}{2})} \left(1 + \frac{1}{\nu} \left[\frac{y - \mu}{\sigma} \right]^2 \right)^{-\frac{(\nu+1)}{2}}, \quad -\infty < y < \infty,$$

sendo que:

$$E(Y) = \mu, \quad \text{para } \nu > 1 \quad \text{e} \quad V(Y) = \left(\frac{\nu}{\nu - 2} \right) \sigma^2, \quad \text{para } \nu > 2.$$

A função suporte $[\ell(\mu, \sigma^2, \nu; \mathbf{y})]$ para a distribuição LS t -Student é:

$$\ell(\mu; \sigma; \nu; \mathbf{y}) = n \ln \left[\Gamma \left(\frac{\nu + 1}{2} \right) \right] - \frac{n}{2} \ln(\nu\pi\sigma^2) + n \ln \left[\Gamma \left(\frac{\nu}{2} \right) \right] + \sum_{i=1}^n \ln \left[1 + \frac{1}{\nu} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right]^{-\frac{(\nu+1)}{2}}.$$

Naturalmente, os estimadores de máxima verossimilhança para os parâmetros μ , σ e ν do modelo LS t -Student precisam ser obtidos via métodos numéricos, resolvendo um complexo sistema de equações que envolve as derivadas parciais de $[\ell(\mu, \sigma^2, \nu; \mathbf{y})]$ com relação a cada um dos parâmetros. Maiores detalhes podem ser consultados em Lenth (1989). A Figura abaixo ilustra curvas da distribuição LS t -Student para diferentes valores de μ , σ^2 e ν .

Teste de Kolmogorov Smirnov

Considere uma amostra aleatória X_1, X_2, \dots, X_n com uma função de distribuição empírica denotada por $F_n(x)$, de tal forma que:

$$F_n(x) = \widehat{P}[X \leq x] = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i),$$

no qual $I_{(-\infty, x]}(X_i)$ é dado por:

$$I_{(-\infty, x]}(X_i) = \begin{cases} 1, & \text{se } x \leq X_i \\ 0, & \text{se } x > X_i \end{cases}$$

Em termos simples, o teste não paramétrico de Kolmogorov-Smirnov (KS) (Conover, 1971) é amplamente utilizado em análises estatísticas para verificar a adequação de um modelo teórico aos dados observados ou para verificar se duas amostras são estatisticamente semelhantes. Ele compara a função de distribuição acumulada empírica ($F_n(x)$) da amostra com uma distribuição de referência teórica ($F_X(x)$). As hipóteses do teste de KS são definidas como:

$$\begin{cases} H_0 : F_X(x) = F_n(x) \\ H_1 : F_X(x) \neq F_n(x) \end{cases}$$

A estatística do teste, D_n , é definida como:

$$D_n = \sup_x |F_n(x) - F_X(x)|,$$

em que D_n é o supremo ou o menor limite superior de todas as diferenças específicas resultantes de $|F_n(x) - F_X(x)|$. Para obter mais detalhes sobre a distribuição exata de D_n e, conseqüentemente, a fórmula para calcular o valor p , consulte Marsaglia et al. (2003).

Critério de Informação de Akaike

O Critério de Informação de Akaike (AIC), proposto por Hirotugu Akaike (1974) em 1974, é uma medida utilizada para selecionar um modelo estatístico entre um conjunto de modelos. Ele oferece uma estimativa da qualidade relativa dos modelos estatísticos para um determinado conjunto de dados.

A equação do AIC é dada por:

$$\text{AIC} = -2 \ln L(\hat{\theta}) + 2k,$$

no qual $L(\hat{\theta})$ denota o valor da função de verossimilhança ajustada do modelo, e k representa o número de parâmetros a serem estimados. Conforme discutido por Emiliano et al. (2014), modelos com valores de AIC mais baixos são considerados preferíveis.

Metodologia

Os dados utilizados neste estudo são valores históricos e atuais do índice $IGPM_{12}$, produzido e divulgado mensalmente pela Fundação Getúlio Vargas (FGV). Mais especificamente, foram obtidas 271 observações que compreendem o período de dezembro do ano 2000 até junho do ano de 2023. O acesso a essas informações foi realizado através da plataforma gratuita denominada como “Dados de mercado”, cujo *link* para consulta é: <https://www.dadosdemercado.com.br>.

A análise inicial consistiu em entender o comportamento dos dados através de uma análise descritiva (cálculo de medidas de posição, de dispersão, de assimetria e de curtose). Além disso, foi construído o histograma de probabilidade dos valores de $IGPM_{12}$. De acordo com essas estatísticas e gráficos iniciais foram identificados na literatura alguns modelos ou distribuições de probabilidade que poderiam se adequar aos dados.

Os parâmetros dos modelos escolhidos foram estimados por máxima verossimilhança (Miller, 2011) e para validar o ajuste dos dados a essas distribuições foi utilizado o teste de Kolmogorov-Smirnov (KS), admitindo um nível de significância de 5%.

Dentre as distribuições selecionadas, aquelas que se adequaram aos dados ou que foram aprovadas pelo teste KS, a escolhida foi aquela que minimizou o Critério de Informação de Akaike (AIC).

O *software* livre R (R, 2023) foi utilizado para realizar as análises estatísticas e a estimação dos parâmetros dos modelos. Os pacotes (funções) utilizados foram: *moments* (*skewness*, *kurtosis*), *MASS* (*fitdistr*) e *fitdistrplus* (*fitdist*).

Resultados e discussão

A Tabela 1 apresenta a estatística descritiva dos valores de $IGPM_{12}$.

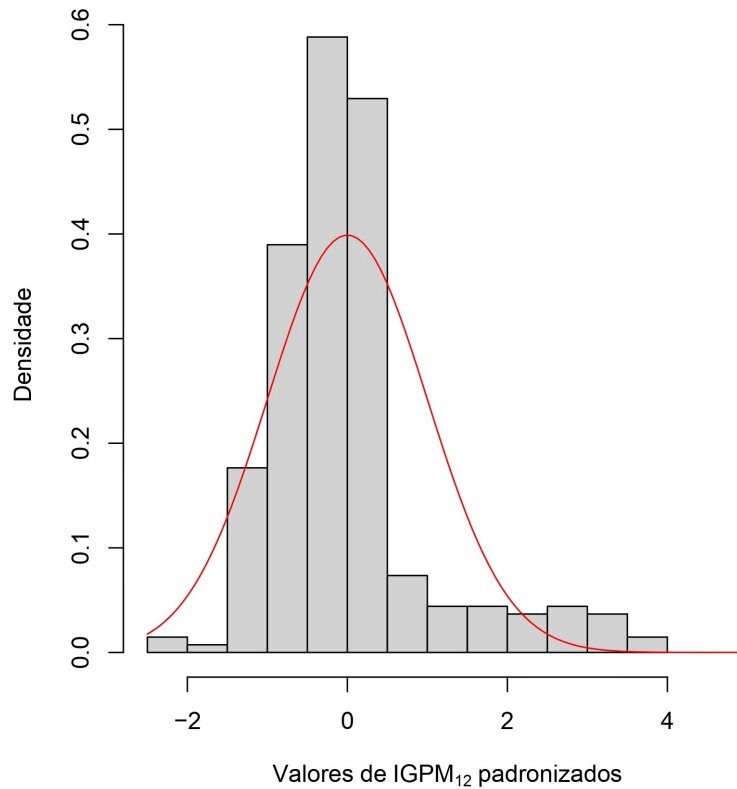
Note que a distribuição empírica ou dos dados parece ser positivamente assimétrica, uma vez que $Mo_X = 3,85 < Md_X = 7,55 < \bar{X} = 8,63$. Essa hipótese é corroborada pela estimativa do coeficiente de assimetria, que foi superior a 0, ou seja, $\hat{\gamma}_1 = 1,49 > 0$. Nesse caso a moda amostral seria um estimador mais recomendado para resumir os dados, sendo que o seu valor indica que no Brasil, historicamente, o $IGPM_{12}$ mais recorrente ou representativo estaria próximo de 3,85%.

Tabela 1: Estatística descritiva dos valores de $IGPM_{12}$.

| Estatísticas | Estimador | Estimativas |
|---------------------------|------------------|-------------|
| Mínimo | $X_{(1)}$ | -7,71 |
| Média | \bar{X} | 8,63 |
| Mediana | Md_X | 7,55 |
| Moda | Mo_X | 3,85 |
| Máximo | $X_{(n)}$ | 37,06 |
| Variância | S_X^2 | 56,29 |
| Desvio-padrão | S_X | 7,50 |
| Coeficiente de variação | CV_X | 86,92 |
| Coeficiente de assimetria | $\hat{\gamma}_1$ | 1,49 |
| Coeficiente de curtose | $\hat{\gamma}_2$ | 5,91 |

Fonte: Dos autores.

É possível observar que a distribuição dos dados é leptocúrtica, com caudas mais densas ou mais pesadas do que a de uma distribuição normal padrão, já que $\hat{\gamma}_2 = 5,91 > 3$. Logo, a probabilidade de valores tão ou mais extremos de $IGPM_{12}$ ocorrerem é maior do que a probabilidade que seria obtida se os dados originais tivessem sido gerados de uma distribuição normal padrão.

Figura 1: Histograma dos dados e curva de densidade de uma $N(0, 1)$.

Fonte: Dos autores.

Observe que embora a distribuição do $IGPM_{12}$ seja positivamente assimétrica [$\hat{\gamma}_1 = 1,49 > 0$], o fato dessa variável poder assumir valores negativos, por exemplo $X_{(1)} = -7,71\%$, inviabiliza a utilização de algumas distribuições que seriam mais apropriadas para essa modelagem, como a Gama, a log-normal e a Weibull. Nesse sentido, foram escolhidas outras quatro distribuições contínuas de probabilidade na tentativa de modelar os dados de forma satisfatória. Mais especificamente, as distribuições foram: Normal, Cauchy, Logística e a LS t Student. A tabela abaixo ilustra as estimativas de máxima verossimilhança dos parâmetros dessas distribuições, bem como os seus respectivos erros padrão.

Posteriormente, o teste de Kolmogorov-Smirnov foi utilizado para avaliar se os dados amostrais se ajustavam adequadamente as distribuições estimadas. A Tabela a seguir ilustra o valor p para o teste KS e também o resultado do Critério de Informação de Akaike (AIC) para cada uma dessas distribuições.

Como observado, a 5% de significância, apenas a distribuição logística e LS t -Student se ajustaram adequadamente aos dados segundo o teste KS. A distribuição LS t -Student obteve o menor valor de AIC ($AIC \approx 1797$) e por isso foi selecionada. Como complemento, a Figura 2 ilustra os gráficos da suporte do modelo LS t -Student ajustado (em cada gráfico é fixada a estimativa de um dos parâmetros), indicando que a combinação de estimativas que produz o seu valor máximo, $\ell(\hat{\mu}; \hat{\sigma}; \hat{\nu}; \mathbf{y}) \approx -896$, são os valores de $\hat{\mu} \approx 7,29$ e $\hat{\sigma} \approx 3,99$ e $\hat{\nu} \approx 2,20$.

Tabela 2: Estimativas de máxima verossimilhança dos parâmetros com seus respectivos erros padrão.

| Estimativas de máxima verossimilhança (Erro padrão da estimativa) | | | |
|--|--|--|--|
| Normal | Logística | Cauchy | LS t - Student |
| $\hat{\mu} \approx 8,63$ (0,454) | $\hat{\mu} \approx 7,73$ (0,376) | $\hat{\theta} \approx 7,28$ (0,292) | $\hat{\mu} \approx 7,29$ (0,309) |
| $\hat{\sigma} \approx 7,49$ (0,321) | $\hat{\sigma} \approx 3,69$ (0,194) | $\hat{\sigma} \approx 3,09$ (0,244) | $\hat{\sigma} \approx 3,99$ (0,328) |
| - | - | - | $\hat{\nu} \approx 2,20$ |
| - | - | - | (0,355) |

Fonte: Dos autores.

Tabela 3: Valor p para o teste KS e resultado do Critério de Informação de Akaike

| Valor p - Teste KS Valor de AIC | | | |
|---------------------------------------|---------------------------------------|---------------------------------------|--|
| Normal | Logística | Cauchy | LS t - Student |
| $p \approx 0\%$ $AIC \approx 1871$ | $p \approx 7\%$ $AIC \approx 1827$ | $p \approx 0\%$ $AIC \approx 1826$ | $p \approx 37\%$ $AIC \approx 1797$ |

Fonte: Dos autores.

Já a [Figura 3](#) apresenta o histograma de probabilidades dos dados e, sobreposto a este, a curva da distribuição LS t -Student ajustada, isto é, com os parâmetros estimados $\hat{\mu} \approx 7,29$ e $\hat{\sigma} \approx 3,99$ e $\hat{\nu} \approx 2,20$.

Por fim, com a distribuição LS t -Student ($\hat{\mu} \approx 7,29$; $\hat{\sigma} \approx 3,99$; $\hat{\nu} \approx 2,20$) ajustada aos dados históricos de $IGPM_{12}$, estimou-se as seguintes probabilidades:

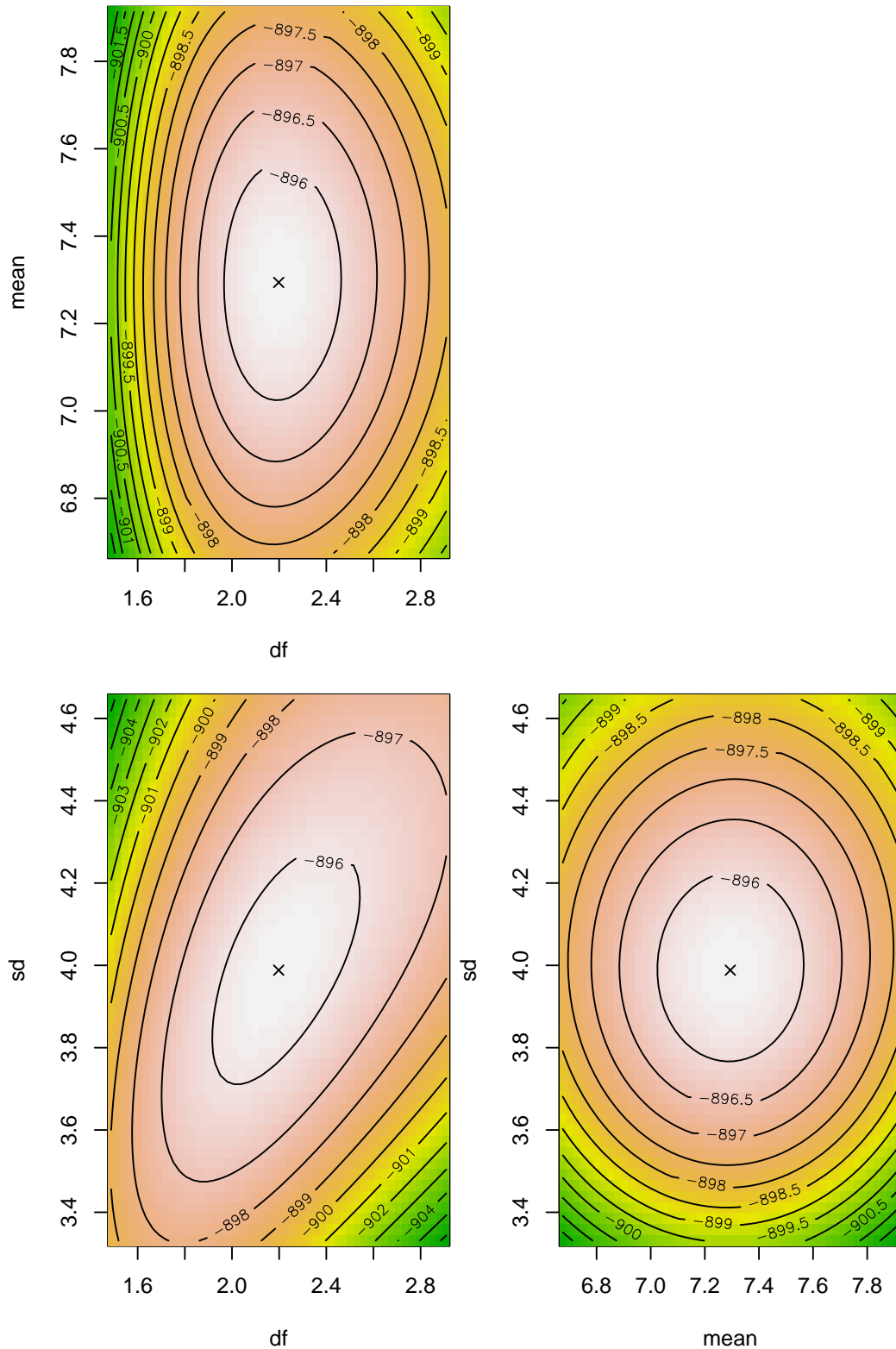
Tabela 4: Probabilidades estimadas

| Probabilidades estimadas |
|--|
| $P(IGPM_{12} < 0\%) \approx 10\%$ |
| $P(0\% \leq IGPM_{12} < 10\%) \approx 62\%$ |
| $P(10\% \leq IGPM_{12} < 20\%) \approx 24\%$ |
| $P(IGPM_{12} \geq 20\%) \approx 4\%$ |

Fonte: Dos autores.

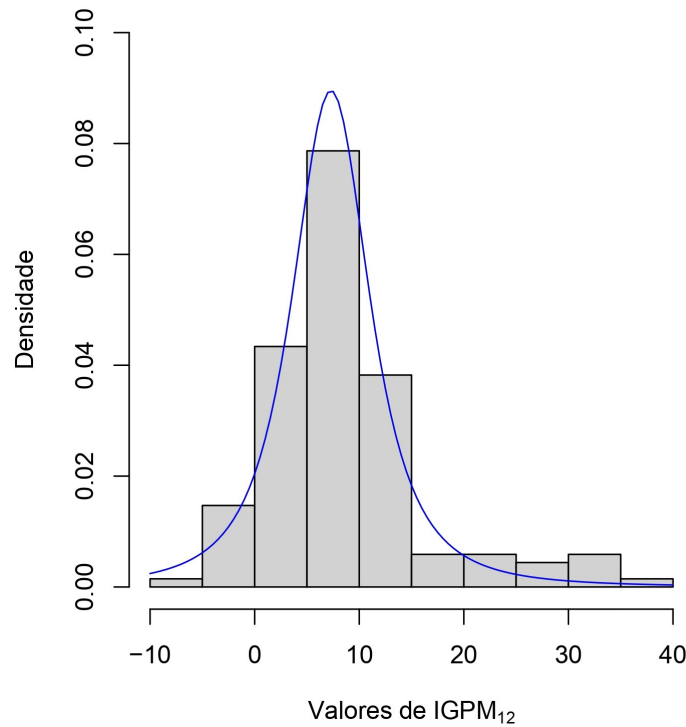
Os resultados demonstram que os contratos de locação, de receitas e de dívidas, cuja remuneração está indexada ao $IGPM$, possuem alta probabilidade ($p \approx 62\%$) de sofrerem reajustes entre 0% e 10%. Adicionalmente, é razoável ou não desprezível a probabilidade ($p \approx 24\%$) de que em 12 meses esses reajustes possam ser de algo entre 10% e 20%. Já reajustes superiores a 20% possuem baixa probabilidade ($p \approx 4\%$) e estão associados a eventos que normalmente ocorrem em situações extremas, como foi o caso da pandemia da Covid-19.

Figura 2: Função suporte ajustada do modelo t -Student ($\hat{\mu} \approx 7,29$, $\hat{\sigma} \approx 3,99$ e $\hat{\nu} \approx 2,20$).



Fonte: Dos autores.

Vale ainda destacar que existe uma probabilidade de 10% de que, no acumulado de 12 meses, o *IGPM* possa assumir valores negativos. Esse percentual reflete a probabilidade do país apresentar deflação, que seria a redução de preços em diversos produtos e serviços ofertados

Figura 3: Histograma da distribuição t - Student, sobre os dados de $IGPM_{12}$.

Fonte: Dos autores.

(Freitas e Cintra, 2008). Além disso, é importante ter em mente que os contratos financeiros reajustados pelo $IGPM$ deveriam sofrer reduções nos valores quando o acumulado desse índice fica negativo. Porém, o que se verifica na prática é que proprietários e credores, equivocadamente, apenas não alteram o valor da remuneração nessas situações.

Conclusão

A probabilidade de ocorrer deflação $P(IGPM_{12} < 0\%) = P(\text{DEFLAÇÃO}) = 9,87\%$. O mais provável é que, para os próximos 12 meses, os reajustes contratuais sejam de magnitude entre 0% e 10% do valor corrente. Existe ainda uma probabilidade razoável de que esses acréscimos fiquem entre 10% e 20%, bem como uma chance mínima que eles excedam 20%.

Agradecimentos

Os autores agradecem Capes, CNPq, Fapemig e a Universidade Federal de Viçosa que deram suporte financeiro para realização deste artigo.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Notre Dame, v. 19, n. 6, p. 717-723, 1974.
- CONOVER, W. J. *Practical Nonparametric Statistical*. John Wiley & Sons Inc., New York, 462 p.
- De SOUZA, J. Â. Para entender os índices de preços: uma visão moderna. *Revista de Economia Mackenzie*, v. 2, n. 2, 2004.

- EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. Information criteria: How do they behave in different models?. *Computational Statistics & Data Analysis*, v. 69, p. 141-153, 2014.
- LENTH, R. V. Algorithm AS 243 — Cumulative distribution function of the non-central t distribution. *Applied Statistics*, v. 38, n. 185–189, 1989.
- MARSAGLIA G., TSANG W. W., WANG J. Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, v. 8, n. 18, 2003.
- MATTOS, A. A inflação brasileira. *Uma abordagem pratica e moderna antes e depois do choque*. Petrópolis: Vozes, 1986.
- MILLER, R. B. *Maximum likelihood estimation and inference: with examples in R, SAS, and ADMB*. 1. ed. [S.l.]: Wiley, 2011.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. Singapore: McGraw-Hill, 1974. 564 p.
- R CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2023.