

## Precificando Criptomoedas: uma Aplicação de Modelos GAMLSS

**Bruno César Martins Oliveira<sup>1, a</sup>**

**José Lamartine Távora Junior<sup>2, b</sup>**

<sup>a</sup> Doutorando em Economia, PIMES/UFPE

<sup>b</sup> Professor Titular, PIMES/UFPE

### Área Temática: Apreçamento de Ativos

**Resumo:** Criptomoedas são ativos digitais com variado potencial financeiro e têm apelo especulativo por exibirem alta volatilidade. Tentativas de precificar este tipo de ativo geralmente usam o estimador de mínimos quadrados ordinários (MQO) que assume que o excesso de retorno das criptomoedas segue uma distribuição de probabilidade do tipo normal, mas existem evidências que essa hipótese não é apropriada. O presente trabalho admite distribuições assimétricas e de caudas pesadas e estima modelos de precificação por meio de modelos de classe GAMLSS (*Generalized Additive Models for Location, Scale and Shape*) que permitem relaxar a hipótese de normalidade. Comparando os resultados obtidos por MQO e por GAMLSS, o desempenho dos modelos é analisado por seis indicadores e cinco deles evidenciam um melhor desempenho dos modelos GAMLSS. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

**Palavras-chave:** Criptomoedas, MQO, GAMLSS, distribuições de probabilidade.

**Códigos JEL:** C58; E4; E5; G1; G2.

**Abstract:** Cryptocurrencies are digital assets with wide financial potential and their high volatility form an appeal for speculation. In trying to pricing this class of assets, usually, the ordinary least square estimator (OLS), that assumes the excess returns follow a normal probability distribution, is deployed; however, there is evidence that such assumption is not adequate. The present work assumes asymmetric and heavy-tailed distributions and estimates asset pricing models through GAMLSS models (Generalized additive models for location, scale and shape) which allow for relaxation of normality hypothesis. Comparing the results obtained

---

<sup>1</sup> [bruno.cmoliveira@ufpe.br](mailto:bruno.cmoliveira@ufpe.br)

<sup>2</sup> [jose.tavorajr@ufpe.br](mailto:jose.tavorajr@ufpe.br)

with OLS and GAMLSS, the performance of the models is measured with six indicators and five of them point to a better performance of GAMLSS models. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

**Keywords:** Cryptocurrencies, OLS, GAMLSS, normality, probability distributions.

**JEL Codes:** C58; E4; E5; G1; G2.

## 1. Introdução:

Nos últimos anos, o mercado de criptomoedas tem se desenvolvido de forma rápida e abrangente, atraindo visões contrárias e favoráveis. Esse mercado multibilionário é parte do desenvolvimento de novas tecnologias criadas como resposta à crise financeira de 2008 (Banco Mundial, 2018).

Embora a literatura sobre criptomoedas se expanda rapidamente, lacunas relevantes persistem (Bariviera; Merediz-Sòla, 2021). Os investidores em criptomoedas estariam então interessados em entender quais fatores têm potencial de explicar as variações de retornos desses ativos e devem ser levados em contas em seu processo de tomada de decisão.

Com base nos modelos tradicionais de precificação de ativos, alguns trabalhos propõem a adaptação deles ao contexto das criptomoedas (Jia; Goodell; Shen, 2022; Shen; Urquhart; Wang, 2020). Fatores diferentes são sugeridos pelos autores, mas ao estimar esses modelos por mínimos quadrados ordinários (MQO), é assumido que os excessos de retornos da criptomoedas seguem uma distribuição normal de probabilidade.

Entretanto, há evidência de que as criptomoedas mostram excessos de retornos que seguem distribuições de probabilidades assimétricas e com caudas pesadas, isto é, a massa de probabilidade se concentra nos extremos da distribuição (Szczygielski; Karathanasopoulos; Zaremba, 2020).

A partir da hipótese de que a normalidade é inadequada para a modelagem dos retornos de criptomoedas, a principal contribuição proposta neste trabalho é a aplicação dos modelos GAMLSS (*Generalized additive models for location, scale and shape*) propostos por Rigby e Stasinopoulos (2005) que permitem considerar outras distribuições e parâmetros adicionais, como a assimetria, para o estudo dos retornos de criptomoedas.

O trabalho busca responder a seguinte questão: modelos de classe GAMLSS produzem resultados mais adequados do que o MQO para a precificação de criptomoedas? Para isso, os dois métodos são aplicados na estimação do modelo de três fatores para criptomoedas (Liu;

Tsyvinski; Wu, 2022), os resultados são diagnosticados visualmente e os modelos comparados por indicadores de desempenho.

Os resultados obtidos por GAMLSS apresentam melhorias, em relação àqueles obtidos por MQO, em todos os diagnósticos, exceto a raiz do erro quadrado médio. Exercícios de robustez, em que a comparação é feita para duas subamostras e para GAMLSS com mais parâmetros modelados, corroboram o melhor desempenho do GAMLSS frente ao MQO. Os resultados corroboram o do trabalho mais próximo a este que é o de Regis (2021) que compara o MQO e GAMLSS para a precificação de ações do mercado brasileiro e também conclui pelo melhor desempenho do GAMLSS.

Mostrando que o GAMLSS é mais apropriado para a estimação desses modelos, a pesquisa também se relaciona e contribui para a literatura sobre precificação de criptomoedas (Liu; Tsyvinski; Wu, 2022; Shen; Urquhart; Wang, 2020) e sobre quais distribuições de probabilidade empregar na modelagem destes ativos (Chan *et al.*, 2017; Szczygielski; Karathanasopoulos; Zaremba, 2020; Osterrieder, 2016; Punzo; Bagnato, 2022) indicando opções com assimetria e caudas pesadas. Para além das criptomoedas, a pesquisa ainda contribui com evidências de que o GAMLSS também pode beneficiar estudos considerando variáveis dependentes que exibam distribuições de probabilidades diferentes da normal.

## 2. Metodologia:

Este capítulo trata do processo para a estimação dos modelos de precificação de criptomoedas e dos dados a serem utilizados. O processo pode ser visto em quatro etapas principais: i) a estimação por MQO; ii) a estimação por GAMLSS; iii) a avaliação e comparação dos resultados e; iv) exercícios de robustez.

### 2.1. Estimação por Mínimos Quadrados Ordinários:

O objetivo principal dessa etapa é obter estimativas para o modelo da Equação 2.1 minimizando o quadrado dos erros e comparar esses resultados com aqueles estimados por modelos GAMLSS. O modelo de precificação a ser estimado segue Liu, Tsyvinski e Wu (2022) e tem o excesso de retornos  $(R_i - R_f)_t$  como variável dependente e os fatores explicativos desse excesso são prêmio de mercado  $(R_M - R_f)$ , tamanho (CSMB) e impulso (CMOM). A equação abaixo ilustra o modelo:

$$(R_i - R_f)_t = \beta_0 + \beta_{CMKT_i}(R_M - R_f)_t + \beta_{CSMB_i}CSMB_t + \beta_{CMOM_i}CMOM_t + \varepsilon_{it}, \quad (2.1)$$

onde  $R_i$  é o retorno do portfólio  $i$ ,  $R_M$  é o portfólio de mercado representado pelo retorno

semanal médio de uma carteira composta por todas as criptomoedas na amostra e  $R_f$  é o retorno de uma carteira livre de risco representada pelo retorno mensal de títulos do Tesouro Americano (Liu; Tsyvinski; Wu, 2022). Os dados são indexados em  $t$  períodos que, com base na literatura, são definidos como semanas (Huijbregts, 2020; Liu; Tsyvinski; Wu, 2022; Shen, Urquhart e Wang, 2020). Os demais parâmetros,  $\beta_0$  e  $\varepsilon_i$ , são o intercepto e o erro do modelo respectivamente.

A variável dependente do modelo é o excesso de retorno, isto é, a diferença entre  $R_{it}$  e  $R_{ft}$  e os retornos são ponderados pelo valor de capitalização dos ativos (produto entre preço e quantidade em circulação) dando peso maior àquelas criptomoedas de grande porte em cada portfólio (Liu; Tsyvinski; Wu, 2022; Regis, 2021).

O fator de tamanho (CSMB) é a diferença entre os retornos médios de portfólios compostos por criptomoedas segundo o valor de capitalização, em outras palavras, o retorno médio do portfólio com as maiores criptomoedas, ou *big* (acima do percentil 70%) subtraídos do retorno médio do portfólio com as menores criptomoedas, ou *small* (abaixo do percentil 30%) (Liu; Tsyvinski; Wu, 2022).

Uma construção semelhante é empregada para construir o fator de impulso (CMOM). Ainda tendo Liu, Tsyvinski e Wu (2022) como referência, as criptomoedas são ordenadas conforme os retornos obtidos em um período de três semanas e classificadas em grupos de maior impulso, ou *high*, (acima do percentil 70%), menor impulso, ou *low*, (abaixo do percentil 30%) e médio impulso, ou *medium* (entre os percentis 30% e 70%).

As intersecções entre os dois grupos de tamanho e os três de impulso formam seis portfólios: i) *small* e *high* (SH); ii) *small* e *medium* (SM); iii) *small* e *low* (SL); iv) *big* e *high* (BH); v) *big* e *medium* (BM) e; vi) *big* e *low* (BL). O fator CMOM é então dado pela diferença entre a média dos retornos de SH e BH e a média dos retornos de SL e BL (Liu; Tsyvinski; Wu, 2022).

Seguindo Regis *et al.* (2023), o modelo MQO é estimado para cada uma das carteiras SH, SM, SL, BH, BM e BL com erros padrão HC2 robustos à heterocedasticidade (Horn; Horn; Duncan, 1975) e os portfólios, antes da estimação propriamente dita, são diagnosticadas com o gráfico quantil-quantil e o teste de Jarque-Bera (1980) em busca da validade da hipótese de normalidade.

## 2.2. Estimação por GAMLSS:

Embora seja possível, com a aplicação dos modelos GAMLSS, estimar parâmetros de

localização, escala e forma, como os coeficientes na Equação 2.1 informam o efeito médio dos fatores no excesso de retorno, o parâmetro média ( $\mu$ ) é o de maior interesse neste artigo.

Assim como em Regis *et al.* (2023), assume-se a função identidade como função de ligação e que o modelo não possui termos não paramétricos, chegando assim a uma versão totalmente paramétrica do modelo GAMLSS. Adaptando o modelo dos autores para o caso das criptomoedas e aplicando os três fatores já discutidos, temos o seguinte submodelo para o parâmetro  $\mu$  dado por:

$$g_1(\mu_i) = \beta_0 + \beta_{CMKT_t}(R_M - R_f)_t + \beta_{CSMB_t}CSMB_t + \beta_{CMOM_t}CMOM_t \quad (2.2)$$

onde  $g_1(\mu_i)$  captura o parâmetro de média da distribuição do excesso de retornos do portfólio  $i$ . O modelo também é estimado para cada uma das seis carteiras como discutido na subseção anterior.

Resta ainda escolher quais famílias de distribuição a assumir para a estimação do parâmetro de média com a aplicação do modelo GAMLSS. Seguindo Punzo e Bagnato (2022) que estudam particularmente o ajuste de famílias de distribuição aos excessos de retornos de criptomoedas, são testadas as seguintes famílias para variáveis contínuas na linha real disponíveis no pacote estatístico *gamlss* (Stasinopoulos *et al.*, 2023): Gaussiana exponencial (exGAUS), Gumbel e Gumble reversa (GU e RG), SU de Johnson original (JSUo) e SU de Johnson parametrizada (JSU), logística (LO), NET (t-Normal exponencial), normal (NO), exponencial de potência (PE) e sua reparametrização (PE2), secante-arco-secante hiperbólica (SHASH) e secante-arco-secante hiperbólica original (SHASHo), normal assimétrica de tipos 1 e 2 (SN1 e SN2), potência exponencial assimétrica tipos 1, 2, 3 e 4 (SEP1, SEP2, SEP3 e SEP4), t assimétrica (SST), t assimétrica tipos 1, 2, 3, 4 e 5 (ST1, ST2, ST3, ST4 e ST5) e família t (TF e TF2). Para apresentações formais das supracitadas famílias de distribuições, ver Rigby *et al.* (2017).

### 2.3. Avaliação de Desempenho:

Para verificar se o GAMLSS produz resultados melhores do que o MQO e qual dos modelos de fatores é o mais apropriado, as métricas empregadas por Regis *et al.* (2023) são seguidas, a saber:

- a)  $A|a_i|$  (a média dos interceptos, onde  $A$  é um operador de média aritmética, e  $a_i$  é o intercepto do  $i$ -ésimo modelo);
- b)  $A|a_i|/A|R_i^{diff}|$  e  $A|a_i|^2 / A|(R_i^{diff})^2|$  ( $R_i^{diff}$  é a diferença entre o excesso de retorno do portfólio  $i$  e o excesso de retorno do portfólio de mercado);

- c) erro absoluto médio (a média aritmética do módulo da diferença entre o valor observado da variável independente e o valor estimado para os T períodos analisados);
- d) raiz do erro médio quadrado (a raiz quadrada da média aritmética do quadrado da diferença entre o valor observado da variável independente e o valor estimado para os T períodos analisados) e;
- e) pseudo- $R^2$ .

Também serão realizados, de acordo com os autores, outros diagnósticos para a robustez de seus resultados como verificação da adequação dos dados aos modelos por meio dos *worm plots* (Buuren e Frediks, 2001), de gráficos com valores observados versus valores previstos que indicam adequação do modelos aos dados (Regis, 2021).

Os modelos são comparados por meio de critérios como o de informação de Akaike (1974) e o critério Bayesiano de Schwarz (1978) escolhendo como mais apropriado aquele que apresentar o menor valor; quando os critérios indicarem famílias distribucionais distintas, será escolhida aquela sugerida pelo critério de Akaike (Regis, 2021).

Em suma, o processo de estimação pode ser resumido nos seguintes passos:

- a) construção das variáveis;
- b) estimação do modelo de três fatores por MQO;
- c) estimação do modelo de três fatores por GAMLSS com a distribuição de melhor ajuste para cada carteira;
- d) estimação do modelo de três fatores por GAMLSS com a distribuição de melhor ajuste comum para as seis carteiras;
- e) avaliação de desempenho dos resultados de cada método de estimação e;
- f) exercícios de robustez.

Adiantando os resultados, o GAMLSS de fato mostra melhor desempenho do que o MQO para estimar o modelo de precificação e alguns exercícios de robustez envolvem avaliar os modelos MQO e GAMLSS em duas subamostras calculadas a partir da semana mediana da amostra completa e; com base nos padrões dos *worm plots*, ajustar o modelo GAMLSS para modelar a concentração de massa de probabilidade nos extremos das distribuições dos portfólios adicionando parâmetros distribucionais.

#### 2.4. Dados:

Dados sobre criptomoedas como valores passados para o par Bitcoin-dólar (e outras criptomoedas) e informações relevantes como capitalização de mercado (produto entre preço

da criptomoeda e sua quantidade emitida) estão disponíveis em *websites* que agregam dados de várias *exchanges*, como CoinGecko. CoinGecko (2024) disponibiliza essas informações até mesmo para criptomoedas que foram encerradas, o que evitaria viés de sobrevivência (Liu; Tsyvinski; Wu, 2022). Esse tipo de viés contaminaria resultados quando criptomoedas que não estão listadas são excluídas da análise (Ammann *et al.*, 2022).

Os históricos de criptomoedas foram obtidos entre 2 e 5 de outubro de 2024 para todas as criptomoedas disponibilizadas (ativas e inativas) em CoinGecko. Essas informações são filtradas mantendo criptomoedas onde o valor de capitalização e o preço são não positivos e o valor de capitalização não é menor que um milhão de dólares (Liu; Tsyvinski; Wu, 2022).

Manteve-se ainda as informações de criptomoedas em que as observações estão registradas nos sete dias da semana. Uma semana é considerada como o período começando às 00:00:00 da segunda-feira e encerrando às 23:59:59 do domingo.

Uma limitação dos dados é a falta de classificação para as criptomoedas inativas, já que para as ativas, é possível acessar informações se a criptomoeda é associada a algum setor como o de jogos, esportes ou se tem preço atrelado a alguma moeda restringindo extensões para a análise.

Para a construção do fator de mercado, assim como em Dunbar e Owusu-Amoako (2022), a taxa livre de risco a ser utilizada será a taxa mensal de títulos do Tesouro americano obtida do *website* de French (2024) que, quando acessados, iam até a última semana de agosto de 2024,

Dessa forma, a amostra contém informações para 592 observações semanais de um total de 7.532 criptomoedas que compreendem o período entre a semana de 28 de abril de 2013 a de 25 de agosto de 2024,

Os processos como manipulação dos dados e estimação dos modelos estatísticos foram feitos por meio da linguagem de programação R (R Core Team, 2023). Especificamente para a estimação do modelo GAMLSS, Rigby e Stasinopoulos (2007) apresentam o que chamam de *software* GAMLSS, um conjunto de pacotes estatísticos como o *gamlss* (Stasinopoulos *et al.*, 2023) que facilita o processo de estimação dos modelos GAMLSS e permite o teste de várias distribuições e diagnóstico de resultados.

### **3. Resultados e Discussão:**

Neste capítulo são apresentados e discutidos os resultados para os modelos de precificação de criptomoedas apresentados previamente. Primeiramente, os dados utilizados

nas estimações do modelo são analisados. Em seguida, são expostos os resultados relacionados à regressão por MQO e alguns diagnósticos como gráficos, teste de normalidade e matriz de correlação. Depois, aqueles que foram obtidos por GAMLSS são expostos e seus desempenhos são comparadas com os resultados por MQO. Finalmente, alguns exercícios de robustez são apresentados.

A Tabela 1 abaixo apresenta estatísticas descritivas para os excessos de retorno semanal de cada uma das seis carteiras (SL, SM, SH, BL, BM e BH).

Tabela 1 – Estatísticas Descritivas dos Excessos de Retornos Carteiras

	SL	SM	SH	BL	BM	BH
Máximo	0,04	0,28	0,04	0,07	5,41	0,22
Mínimo	-0,12	-0,12	-0,12	-0,15	-0,16	-0,40
Média	-0,03	0,03	0,03	-0,03	-0,02	-0,03
Mediana	-0,01	-0,01	0,01	-0,02	-0,02	0,01
Desvio-padrão	0,04	0,04	0,04	0,04	0,23	0,05
Assimetria	-1,25	-0,26	-1,25	-0,68	22,95	-0,76
Curtose	3,30	9,42	3,30	2,81	547,94	9,24

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

A tabela mostra que o portfólio BM apresenta o maior excesso de retorno máximo (aproximadamente 500%), seguido de SM (aproximadamente 28%) e BH (aproximadamente 23%), os menores valores aproximados de excesso de retorno mínimo são dos portfólios BH (-40%) e BM (-16%). Esses valores ilustram a volatilidade característica das criptomoedas.

Comparando SL a BL, SM a BM e SH a BH, as carteiras formadas por criptomoedas de maior capitalização de mercado atingem excessos de retorno maior que aquelas formadas por criptomoedas de baixa capitalização, mas também apresentam menores valores.

Com exceção da carteira BM, com aproximadamente -2% de excesso de retorno, as demais carteiras têm aproximadamente o mesmo valor de -3%, próximo do retorno de BM. Considerando apenas o excesso médio de retorno, é possível dizer que os portfólios são semelhantes.

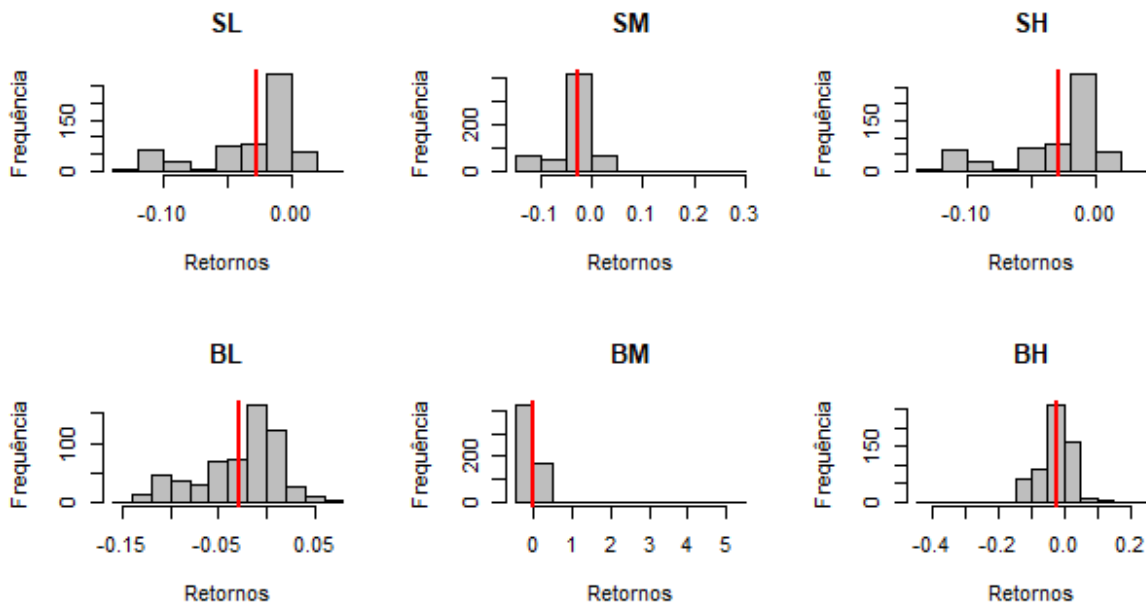
Os valores da mediana para os portfólios não desviam significativamente da média,

sendo essa diferença maior para o portfólio BH (excesso de retorno médio de  $-2,57\%$  e mediana de  $1,24\%$ ). Outra estatística interessante que a Tabela 1 apresenta é o desvio-padrão, uma medida da variabilidade dos excessos dos retornos, que é maior para a carteira BM ( $0,2276$ ) e menor para SL ( $0,03710418$ ) e SH ( $0,03715234$ ).

Com valores de assimetria negativo para todas as carteiras (exceto BM) e de curtose (ou achatamento) maiores que 3 (exceto BL) tem-se que a cauda esquerda das curvas de distribuição de probabilidade dessas carteiras é maior que a direita (para BM, é a cauda direita que é notavelmente mais pesada) e que as curvas são mais pesadas que a curva da distribuição normal (caso em que a curtose é 3) (para BM, a curva é menos pesada) (Regis, 2021). Vale notar que ao contrário das demais carteiras, a curtose de BM (aproximadamente  $547,94$ ) é a que mais se distancia da distribuição normal, seguida da carteira BH (aproximadamente  $9,24$ ).

Também cabe inspecionar visualmente o ajuste dos dados à distribuição normal, hipótese essencial para a estimação de modelos de precificação por MQO. Histogramas são gráficos úteis nesse sentido, pois permitem uma visualização de como os dados se concentram ao longo da distribuição.

Figura 1 – Histogramas



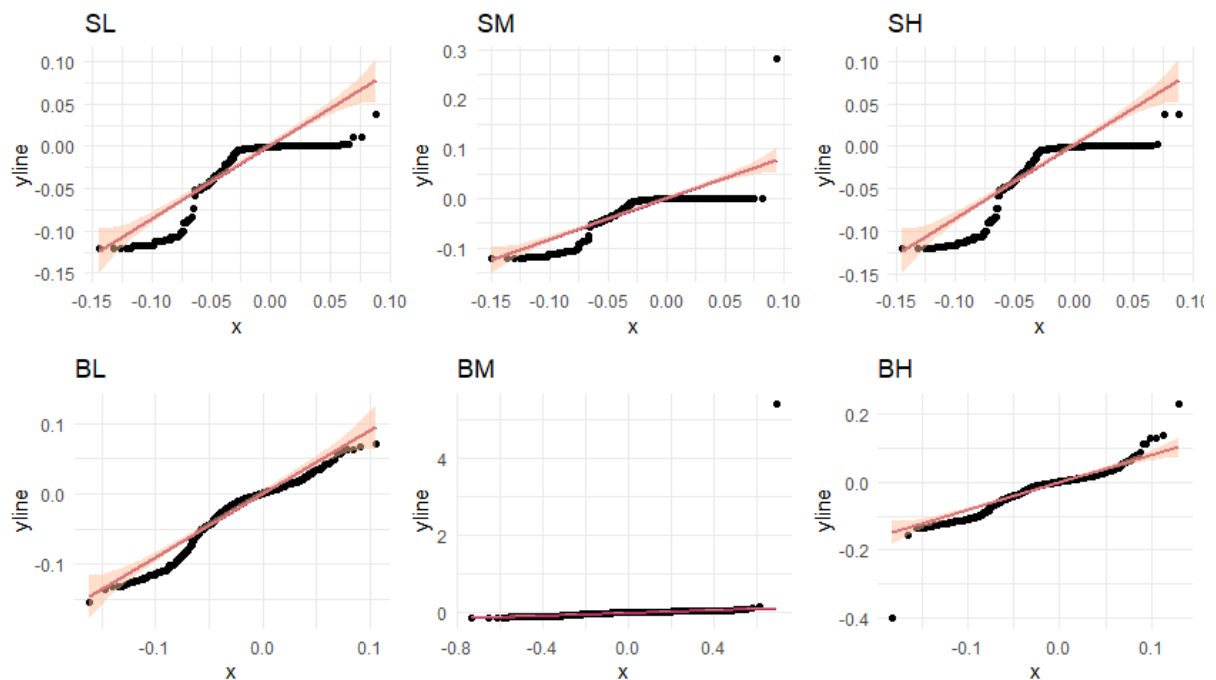
Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

A linha vermelha nos gráficos representa o excesso de retorno médio observado para cada portfólio. Assim é possível verificar que SL e SH são carteiras com distribuições semelhantes e com observações mais frequentes à direita da média. BL é o portfólio que tem

uma distribuição mais ampla e BM, a carteira com menor. SM, BM e BH têm maior concentração em volta da média; essa maior frequência junto à média é uma característica da distribuição normal (caso em que a massa de distribuição é simétrica em relação à média.)

Para verificar, especificamente, o ajuste dos excessos de retorno à distribuição normal, são empregados os gráficos quantil-quantil que mostram a inadequação da distribuição para a modelagem desses dados. Os gráficos indicam alguns pontos com considerável distância da diagonal principal em vermelho e outros que, além disso, estão fora dos envelopes em laranja, região onde é válida a hipótese de distribuição normal com nível de confiança de 95%, sugerindo que os quantis dos dados desviam dos quantis teóricos de uma distribuição normal, caso em que estariam sobre a diagonal vermelha (Almeida; Loy; Hoffman, 2018; Regis, 2021).

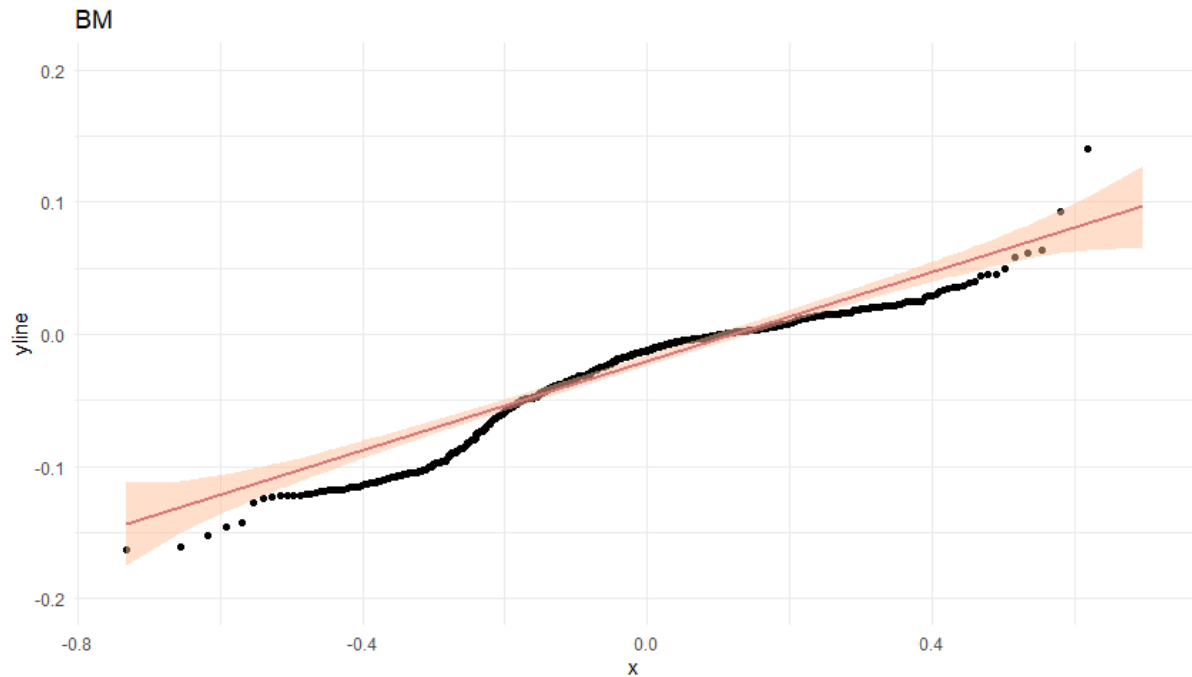
Figura 2 – Gráficos Quantil-Quantil



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

A carteira BM tem um ponto com desvio da diagonal principal grande o suficiente para distorcer a escala dos gráficos e fazer os envelopes e diagonal parecerem uma única linha. O gráfico abaixo limita os eixos para se concentrar onde há maior parte dos pontos, o que permite visualizar que esse portfólio tem comportamento similar ao de BL e BH.

Figura 3 – Gráfico Quantil-Quantil Ampliado para a Carteira BM



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

As carteiras SL, SM e SH têm os desvios mais frequentes e notáveis. BL, BM e BH têm menos pontos com desvio significativos, mas como as demais também concentram os maiores desvios em pontos mais distantes do centro da diagonal. Isso indica que a distribuição dos retornos possui caudas pesadas com massa de probabilidade não modelada quando assume-se que seguem a distribuição normal (Regis, 2021).

Para verificar a normalidade, o teste Jarque-Bera de normalidade também é aplicado. A hipótese nula é de que a distribuição é normal (Jarque; Bera, 1980) e é rejeitada para todas as carteira como indicado pelos p-valores apresentados na Tabela 2.

O teste de Jarque-Bera, os histogramas, os gráficos quantil-quantil discutidos e as estatísticas descritivas evidenciam que há divergência da distribuição das carteiras formadas com criptomoedas em relação à distribuição normal comumente assumida, motivando a adoção de métodos que considerem outras distribuições como o GAMLSS. Os resultados para essas estimações são apresentados e discutidos na próxima seção.

Tabela 2 – Teste Jarque-Bera de Normalidade

Carteira	Estatística Jarque-Bera	p-valor
SL	157,19	0,00***
SM	1022,10	0,00***
SH	155,58	0,00***
BL	46,77	0,00***
BM	7376924,00	0,00***
BH	1016,40	0,00***

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Antes de estimar os modelos de precificação com GAMLSS, é válido apresentar a matriz de correlação referente aos três fatores a serem utilizados na estimação a seguir.

Tabela 3 – Matriz de Correlação entre os Fatores do Modelo

	CMKT	CSMB	CMOM
CMKT	1,0000	-0,0558	0,0373
CSMB	-0,0558	1,0000	0,2027
CMOM	0,0373	0,2027	1,0000

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Em valor absoluto, os coeficientes de correlação entre os fatores CMKT e CSMB (único coeficiente negativo) e entre CMKT e CMOM são os mais próximos de zero. O maior coeficiente é o da correlação entre CSMB e CMOM (aproximadamente 0,2) e, como a correlação perfeita ocorre quando o valor absoluto do coeficiente é igual a um, pode-se considerar que não há grande correlação entre os fatores aliviando preocupações com problemas de multicolinearidade.

### 3.1. Estimação por MQO:

Para analisar a adequação da estimação do modelo de precificação de criptomoedas proposto, o método de mínimos quadrados ordinários é estimado a fim de comparação com os resultados a serem obtidos com o método GAMLSS. A tabela abaixo apresenta os resultados por MQO, a primeira coluna indica a carteira a qual o modelo na Equação 2.1 está aplicado e

as colunas seguintes apresentam suas respectivas estimativas e as estatísticas t são apresentadas entre parênteses.

Tabela 4 – Resultados de Regressão MQO

MQO	$\beta_0$	$\beta_{CMKT}$	$\beta_{CSMB}$	$\beta_{CMOM}$
SL	-0,0248 (-2,7509)**	0,1452 (0,4570)	0,0011 (0,0842)	0,1096 (1,4146)
SM	0,0240 (-2,6192)**	0,1483 (0,4612)	0,0024 (0,1673)	-0,1961 (-0,6085)
SH	-0,0249 (-2,7752)**	0,1469 (0,4653)	0,0038 (0,2651)	0,1248 (1,5852)
BL	-0,0234 (-2,2594)*	0,1776 (0,4859)	0,0021 (0,1513)	-0,5504 (-3,7477)***
BM	0,0408 (2,1471)*	2,4567 (3,6671)***	-0,0070 (-0,2632)	-0,4194 (-1,7422).
BH	-0,0234 (-2,2410)*	0,1758 (0,4786)	-0,0006 (-0,0429)	1,4343 (9,5786)***

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Notas: Estatísticas t em parênteses. Significância estatística: \*\*\*: p-valor < 0; \*\*: p-valor < 0,001; \*: p-valor < 0,01; . : p-valor < 0,1.

Inicialmente, nota-se que apenas o intercepto é a única estimativa estatisticamente significativa para todas as carteiras e isso para os níveis de significância de 1%, apenas as carteiras de baixa capitalização de mercado são significantes para níveis de 0,1%. A carteira BM apresenta maior intercepto (ainda que próximo de zero) sugerindo que é o portfólio com mais informações não captadas pelos outros fatores.

Essa carteira também apresenta o maior coeficiente (e único estatisticamente significativo) para o fator de mercado, 2,4567, indicando que o aumento de 1% desse fator implica um aumento de excesso de retorno de aproximadamente 2,46% para o portfólio. As demais carteiras teriam aumento entre 0,145% e 0,178%, mas nenhum desses coeficientes têm significância estatística.

O fator de tamanho apresenta coeficientes insignificantes estatisticamente para todas as carteiras e todos próximos de zero sugerindo a inocuidade desse fator. A insignificância estatística também se apresenta para o fator de impulso dos portfólios formados por criptomoedas de tamanho pequeno, mas significativo para carteiras com criptomoedas de tamanha grande. BL e BM têm coeficientes de -0,55 e -0,41 sugerindo que quando o fator de

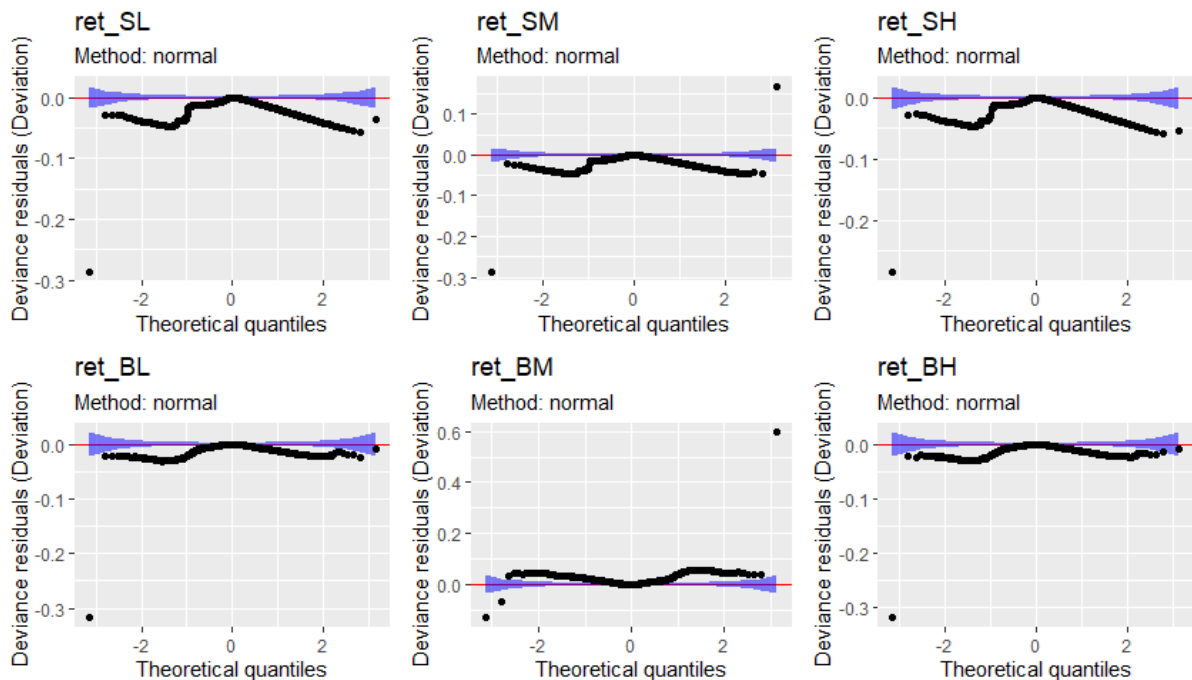
impulso de três semanas aumenta em 1%, seus excessos de retornos diminuem 0,55% e 0,41% respectivamente. Já para a carteira BH, os resultados sugerem que essa variação do fator leva a um aumento do excesso de retorno de 1,43%.

De forma geral, os resultados por MQO indicariam que os fatores captam grande parte da variação dos excessos de retornos das carteiras (vide coeficiente próximos de zero), mas que eles não são estatisticamente significativos principalmente para carteiras com criptomoedas de baixa capitalização.

Notadamente, o excesso de retorno é afetado significativamente pelo fator de mercado apenas para o portfólio BM e o fator de impulso para BL, BM e BH. Já o fator de tamanho não teria nenhum impacto nos excessos de retornos de nenhuma das seis carteiras.

A confiabilidade destes resultados depende da qualidade de ajuste do modelo aos dados. Um diagnóstico útil para isso são os *worm plots* apresentados abaixo.

Figura 4 – *Worm Plots* para a Regressão MQO



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Os dados apresentam semelhança no formato em que os pontos se dispõem, com maior semelhança entre os gráficos para SL, SM e SH. Já o padrão em BM parece girar o de BL e BH. Outra semelhança compartilhada entre os *worm plots* dos seis portfólios é a ocorrência de pontos extremos com maior desvio do eixo de origem (linha vermelha horizontal) indicando

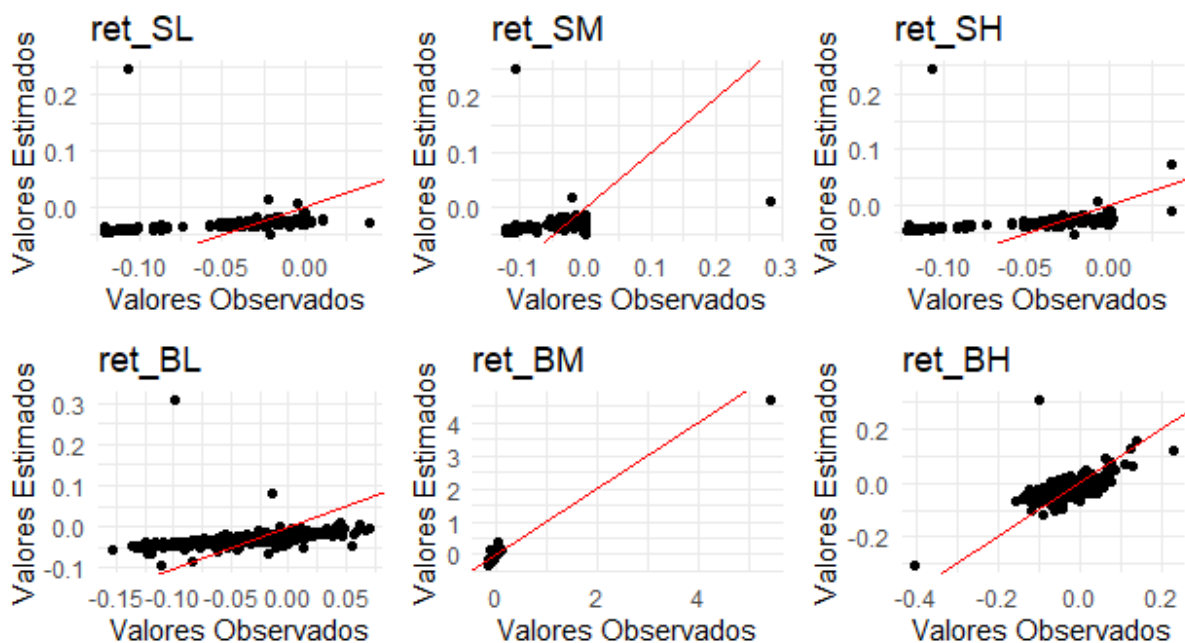
que esses têm mais distância aos valores que teriam se os resíduos seguissem a distribuição normal.

Também para as seis carteiras, há poucos pontos mais centrais aos gráficos que estão na região azul que representa um intervalo de confiança de 95% de que os resíduos estão de acordo com a distribuição assumida. Para os pontos nas extremidades esquerda e direita há certa aproximação dessa região, mas poucos pontos se encontram nela.

É, contudo, a observação de que a maioria dos pontos nos *worm plots* se encontra fora da região de confiança que serve como maior evidência que o MQO não ajusta bem os dados a esse modelo que assume uma distribuição de probabilidade normal.

O baixo ajuste desse método é corroborado pelos gráficos abaixo que representam os valores observados plotados contra os valores estimados. A linha vermelha indica igualdade entre os valores observados e os preditos pelo modelo, assim um bom modelo teria os pontos sobre a diagonal ou muito próximo a ela. Não é o caso dos resultados estimados com MQO como mostra a figura a seguir:

Figura 5 – Gráficos Valores Observados *versus* Previstos para a Regressão MQO



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

BM é o portfólio com mais pontos perto da diagonal vermelha, o baixo número de pontos indica que as observações têm valores semelhantes, diferente dos outros portfólios que apresentam mais pontos distintos entre si. Há de se notar também que os pontos, especialmente

para SL, SH e BL, parecem formar uma reta, mas que não se aproximam da linha que representa a igualdade entre valores previstos e observados.

Dessa forma, conclui-se que há inadequação do MQO aos dados de criptomoedas de acordo com os gráficos acima, o que inspira cuidados ao considerar os resultados apresentados na Tabela 4 como apropriados.

### 3.2. Estimação por GAMLSS:

Assumindo que os problemas de ajuste do MQO se dão pela suposição de normalidade, ela é relaxada e outras distribuições de probabilidades são consideradas. Para isso, é aplicado o método GAMLSS segundo a Equação 2.2.

Ajustando as distribuições disponíveis no pacote *gamlss* aos excessos de retornos, os critérios AIC e BIC de informação sugerem quatro famílias como as de melhor ajustes às carteiras. Ambos critérios concordam ao indicar a mesma família nos seis casos. As carteiras SL e SH se ajustam melhor à distribuição de Johnson-SU original (JSUo). Já SM e os três portfólios com criptomoedas de alta capitalização têm sugestões diferentes famílias entre si.

Os critérios indicam a distribuição assimétrica exponencial de potência de tipo 3 (SEP3) para SM e BL, assimétrica t de tipo 3 (ST3) para BM e assimétrica exponencial de potência de tipo 2 (SEP2) para BH. De fato, JSUo, ST3, SEP2 e SEP3 são capazes de modelar casos em que as distribuições têm caudas mais pesadas que a distribuição normal (Rigby *et al.*, 2017), característica desejada para a modelagem de retornos de criptomoedas (Osterrieder, 2016; Szczygielski, Karathanasopoulos; Zaremba, 2020).

Considerando essas famílias na estimação de modelos GAMLSS para o excesso de retorno de seus respectivos portfólios, chega-se aos resultados são apresentados na Tabela 5.

Os resultados obtidos por GAMLSS apresentam alguma semelhança com aqueles da Tabela 4 em relação aos interceptos. Todos são próximos de zero e, com exceção da carteira BM, são estatisticamente significativos. A proximidade de zero indica que há pouca variância na variável dependente que não é explicada pelos fatores do modelo.

A Tabela 5 mostra significância estatística em todos os seis casos para o fator de mercado, ao contrário do primeiro conjunto de resultados onde isso ocorre somente para o portfólio BM. Todos os coeficientes, próximos de um, indicam que um aumento de 1% no fator de mercado implicaria um aumento entre 0,9% e 1% nos excessos de retornos.

Também há contraste na significância do fator de tamanho que é estatisticamente insignificante para todas as carteiras nos resultados por MQO, mas que para os resultados

obtidos com GAMLSS, é estatisticamente significativo para as carteiras SM, BL e BH. Ambos conjuntos de resultados geram estimativas próximas de zero.

Tabela 5 – Resultados de Regressão GAMLSS

<b>GAMLSS</b>	$\beta_0$	$\beta_{CMKT}$	$\beta_{CSMB}$	$\beta_{CMOM}$
SL	-0,0054 (-9,6371)***	0,9067 (78,4415)***	-0,0027 (-1,0378)	0,0364 (1,4062)
SM	-0,0029 (-17,4543)***	0,9499 (230,8722)***	-0,0019 (-76,1533)***	0,0249 (7,5195)***
SH	-0,0054 (-9,5843)***	0,9067 (78,6404)***	0,0010 (0,2956)	0,0427 (1,6815).
BL	0,0007 (9,6875)***	0,9992 (527,0159)***	-0,0002 (-3,6983)***	-0,8923 (-147,1130)***
BM	0,0002 (0,3645)	1,0036 (98,2246)***	-0,0011 (-0,5117)	-0,0153 (-0,6084)
BH	0,0007 (20,0591)***	1,0061 (716,2514)***	-0,0028 (-162,3925)***	1,0983 (285,6090)***

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Notas: Estatísticas t em parênteses. Significância estatística: \*\*\*: p-valor < 0; \*\*: p-valor < 0,001; \*: p-valor < 0,01; . : p-valor < 0,1.

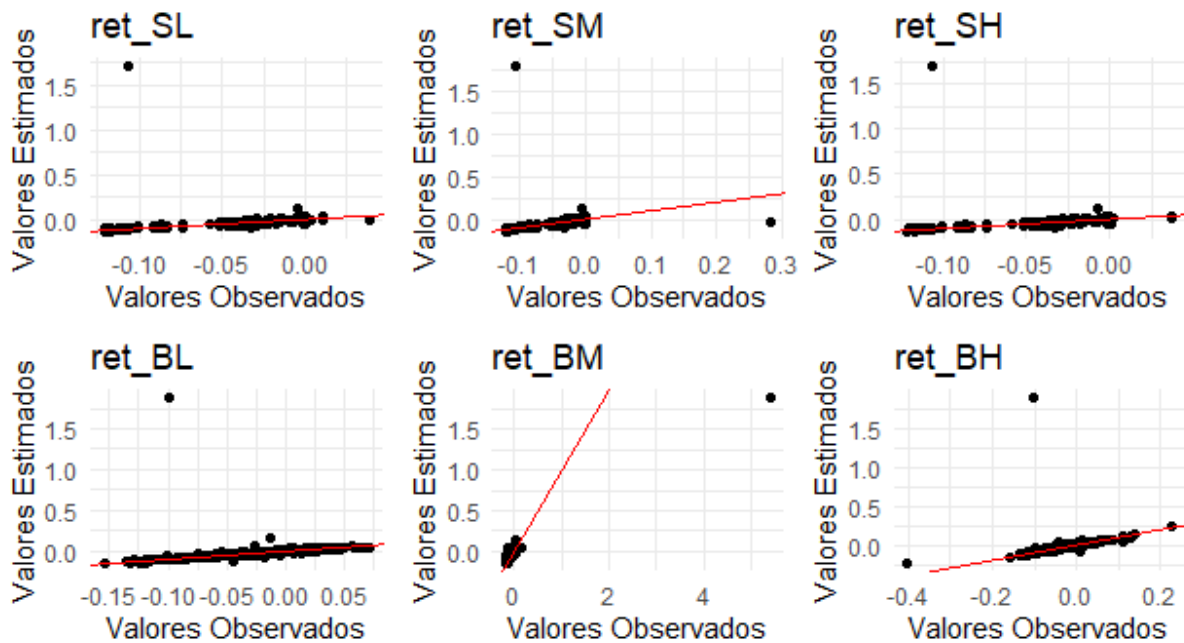
Por fim, o fator de impulso é significativo a um nível de significância de 5% para a carteira SH e a nível de 0% para SM, BL e BH. BL e BM são os únicos casos de coeficientes negativos para esse fator. Já o portfólio BH apresenta o maior coeficiente para esse fator, sugerindo que um aumento de aproximadamente 1,1% nos excessos de retornos segue a uma variação de 1% no fator de impulso.

Para avaliar a qualidade do ajuste quando o modelo de precificação é estimado por GAMLSS, analisa-se os gráficos de valores observados *versus* previstos na Figura 6.

Os pontos nesses gráficos estão bem distribuídos sobre a linha vermelha, indicando um ajuste satisfatório dos dados ao modelo, ainda que haja alguns pontos desviantes não capturados adequadamente pelo modelo.

Dessa forma, esses gráficos mostram um aprimoramento do ajuste dos dados quando o modelo GAMLSS é empregado na estimação de modelos de precificação de criptomoedas quando em comparação com as estimativas obtidas por mínimos quadrados ordinários.

Figura 6 – Gráficos Valores Observados *versus* Previstos para a Regressão GAMLSS



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Visualmente, também é possível analisar se há bom ajuste do modelo por meio dos *worm plots* na Figura 7. Nestes gráficos, a região de confiança é de cor cinza e a quantidade de pontos nessa região é notavelmente maior do que aquela nos *worm plots* para MQO.

Isso sugere que, nesse caso, os quantis dos resíduos são mais próximos dos quantis teóricos da distribuição aplicada a cada portfólio; portanto, sugerindo que o modelo GAMLSS se ajusta melhor aos dados. Há de se ressaltar que há alguns pontos no limite da região de confiança e outros de notável distância da linha horizontal na origem, sugerindo observações extremas que não se adequam bem ao modelo.

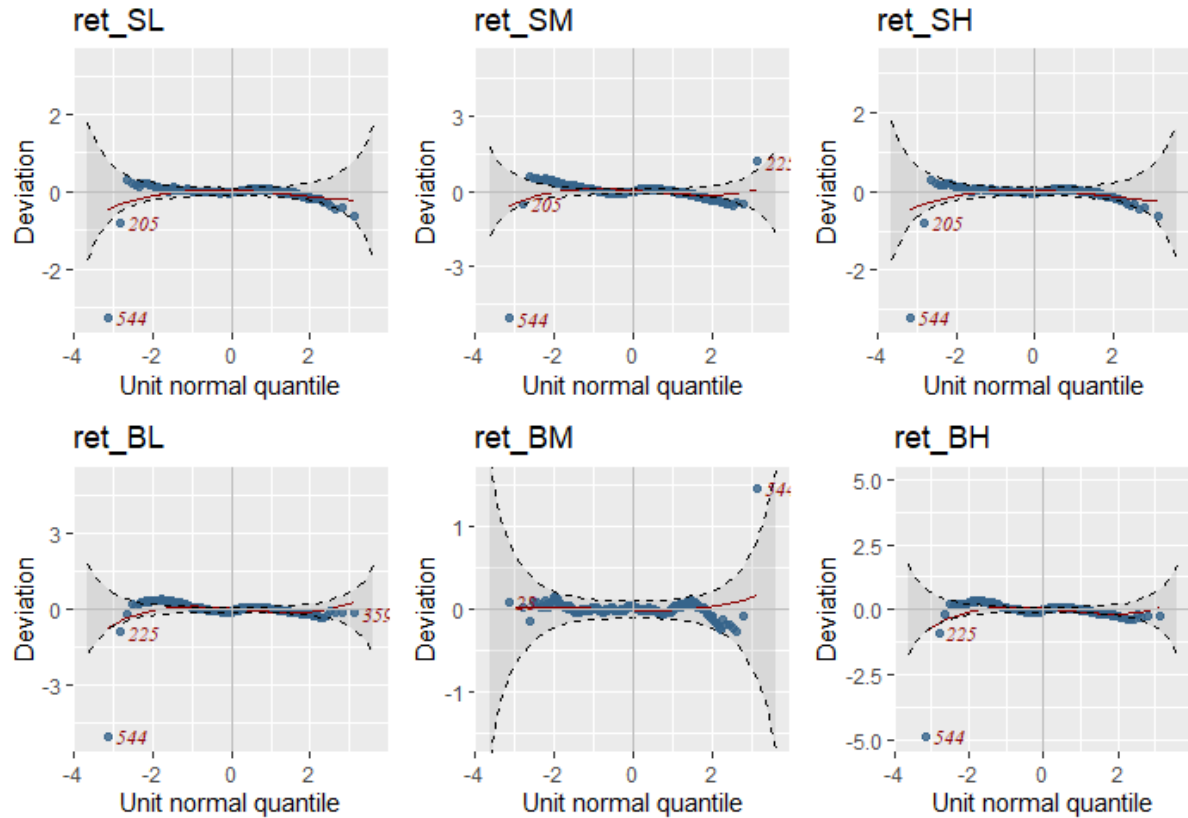
De todos os portfólios, é BM o que apresenta melhor ajuste com apenas um ponto fora da região de significância. Para os demais, os resíduos parecem formar o mesmo padrão onde, da esquerda para a direita, os padrões parecem ter alguma inclinação negativa e uma inclinação positiva mais ao centro antes de retomar a inclinação negativa.

Outra característica comum entre os *worm plots* são os desvios mais notáveis para os resíduos mais extremos. BH e SM parecem ser os dois casos em que há mais pontos fora da região cinza. Na Figura 8 (na subseção 3.4.1), são apresentados gráficos ampliados na região central que evidenciam esse comportamento.

Com mais resíduos na região de confiança nos seis gráficos quando comparados com aquela na Figura 4, há evidência de que o modelo GAMLSS, ao considerar distribuições de

probabilidade diferentes da normal assumida no MQO, aprimora o ajuste aos dados usados na precificação de criptomoedas. Para verificar a consistência dessas evidências, seis indicadores de desempenho foram calculados para cada um dos modelos e são exibidos na seção seguinte.

Figura 7 – *Worm Plots* para a Regressão GAMLSS



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

### 3.3. Avaliação de Desempenho:

Os indicadores empregados a seguir servem para comparar os dois modelos estimados e, com base na comparação, definir como modelo mais apropriado aquele que apresentar o menor valor das estatísticas; exceto para o pseudo- $R^2$ , quando o melhor modelo tem maior valor. A Tabela 6 mostra os valores dos indicadores para cada modelo e mostra em negrito o valor para o melhor modelo segundo aquele indicador.

De forma geral, o modelo GAMLSS se apresenta como a melhor alternativa segundo cinco dos seis indicadores e o MQO é o modelo indicado pelos dois restantes. Mesmo quando o MQO apresenta métricas mais favoráveis, elas não são tão discrepantes frente às de GAMLSS.

Tabela 6 – Avaliação de Desempenho

<b>Avaliação</b>	<i>MQO</i>	<i>GAMLSS</i>	<i>Melhor Modelo</i>
$A a_i $	0,0269	<b>0,0026</b>	GAMLSS
$A a_i /A R_i^{\text{diff}} $	6,5513	<b>0,6262</b>	GAMLSS
$A a_i^2 /A (R_i^{\text{diff}})^2 $	42,5662	<b>0,6359</b>	GAMLSS
$A MAE $	0,0311	<b>0,0129</b>	GAMLSS
$A RMSE $	<b>0,0436</b>	0,0938	MQO
$A \text{pseudo-R}^2 $	0,3094	<b>0,5144</b>	GAMLSS

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

A primeira estatística a ser analisada é a média dos interceptos das seis carteiras,  $A|a_i|$ , que deve ser zero, ou próximo de zero, para um modelo em que toda a variação é capturada pelas variáveis independentes (Fama; French, 2015). O modelo MQO apresenta um valor próximo de zero, mas é o GAMLSS que tem a métrica mais aproximada de zero e, dessa forma, o indicador  $A|a_i|$  sugere que o modelo GAMLSS é melhor que o MQO.

Os segundo ( $A|a_i|/A|R_i^{\text{diff}}|$ ) e terceiro indicadores ( $A|a_i^2|/A|(R_i^{\text{diff}})^2|$ ) são métricas de dispersão medindo a proporção de variação não explicada pelos modelos ( $A|a_i|$ ) por quanto os portfólios desviam do portfólio de mercado ( $A|R_i^{\text{diff}}|$ ), onde  $A|a_i^2|/A|(R_i^{\text{diff}})^2|$  é uma variação ajustada para possíveis erros na medição das variáveis (Regis, 2021). O modelo MQO apresenta os maiores valores para ambos indicadores (6,5513 e 42,5662, respectivamente) e diferem bastante daqueles do GAMLSS (0,6262 e 0,6359, respectivamente). Assim, esses indicadores sugerem que o modelo GAMLSS com média tem menor proporção da sua variação não explicada no modelo.

Os valores médio para o erro absoluto médio são bem próximos, porém o MQO ainda apresenta o maior erro (0,0311) frente ao GAMLSS (0,0129). A raiz do quadrado médio que, como o erro absoluto médio, também é uma medida de precisão dos modelos (Regis, 2021) mostra o MQO (0,0436) como mais preciso do que o GAMLSS (0,0938). A discrepância entre as conclusões implicadas pelas duas medidas de precisão pode-se explicar pelo maior peso que a raiz do quadrado médio dá a grandes erros (Regis, 2021).

Por fim, o pseudo  $R^2$  mostra que, em média, 30,94% da variação dos excessos de retorno é captada pelo modelo MQO. Para o modelo GAMLSS, essa métrica sugere um melhor desempenho, com média de 51,44% de variação explicada.

Esses indicadores são empregados para verificar qual modelo (ainda que imperfeito) que melhor explica o excesso de retorno de carteiras construídas diferentemente (Fama; French, 2015). Assim, com cinco indicadores dos seis apresentados na Tabela 6 e com os gráficos nas Figuras 6 e 7 que mostram resultados mais promissores para os ajustes dos modelos GAMLSS, considera-se que essa classe é mais adequada à estimação do modelo de precificação para criptomoedas.

Para verificar a robustez dessa consideração, alguns exercícios adicionais são realizados a seguir: o primeiro repete as estimações com MQO e GAMLSS para subdivisões da amostra principal e o segundo procura uma especificação para o GAMLSS com base nos ajuste refletidos por *worm plots*.

### 3.4. Exercícios de Robustez:

O seguinte conjunto de resultados mostra os indicadores aplicados na avaliação de desempenho anterior para duas subdivisões da amostra: amostra A e amostra B. Essas subdivisões são feitas com base na mediana para as semanas da amostra total, isto é, a amostra A apresenta 296 observações do período entre a semana de 28 de abril de 2013 a de 27 de dezembro de 2018; a amostra B contém as demais observações da amostra principal.

Tabela 7 – Avaliação de Desempenho (Duas Subamostras)

Avaliação	MQO			GAMLSS		
	Total	A	B	Total	A	B
$A a_i $	0,0269	0,0035	0,0301	<b>0,0026</b>	<b>0,0008</b>	<b>0,0058</b>
$A a_i /A R_i^{\text{diff}} $	6,5513	2,0021	3,6659	<b>0,6262</b>	<b>0,4446</b>	<b>0,6999</b>
$A a_i^2 /A (R_i^{\text{diff}})^2 $	42,5662	3,6569	13,4418	<b>0,6359</b>	<b>0,1809</b>	<b>0,8798</b>
$A MAE $	0,0311	<b>0,0096</b>	0,0219	<b>0,0129</b>	0,0099	<b>0,0142</b>
$A RMSE $	<b>0,0419</b>	<b>0,0147</b>	<b>0,0335</b>	0,0904	0,0178	0,1078
$A \text{pseudo-R}^2 $	0,3094	0,5648	0,6359	<b>0,5338</b>	<b>0,7374</b>	<b>0,7672</b>

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Assim, é possível verificar se os resultados se mantêm para ambos períodos,

especialmente o período da amostra A quando o mercado de criptomoedas era menos desenvolvido (Liu; Tsyvinski; Wu, 2022). Se os modelos GAMLSS são mesmo superiores, espera-se que a maioria dos indicadores para as duas subamostras sejam mais favoráveis a essa classe. De fato, é o que a Tabela 7 mostra.

A preferência pelo modelo GAMLSS é reforçada, pois as conclusões da comparação dos indicadores exibidos pelos modelos MQO e GAMLSS para a amostra total se repetem para as duas subamostras. A exceção é o erro absoluto médio que é menor (uma diferença de 0,0003) para o MQO quando a subamostra A é considerada.

De todo modo, segundo a maioria dos indicadores, conclui-se que os modelos GAMLSS tendem a apresentar desempenho superior ao MQO mesmo quando a comparação é repetida com outros períodos.

### 3.4.1. Parâmetros Distribucionais Adicionais:

A possibilidade de modelar outros parâmetros por meio de modelos GAMLSS é explorada no exercício desta subseção. Seguindo os padrões dos *worm plots* é possível ter uma indicação de qual parâmetro incluir na modelagem. Com pontos extremos nos *worm plots* da Figura 7, os gráficos são distorcidos e, por isso, eles são apresentados novamente na figura abaixo com seus limites verticais redefinidos com o objetivo de se concentrar onde há mais pontos e ter noção dos padrões que eles formam.

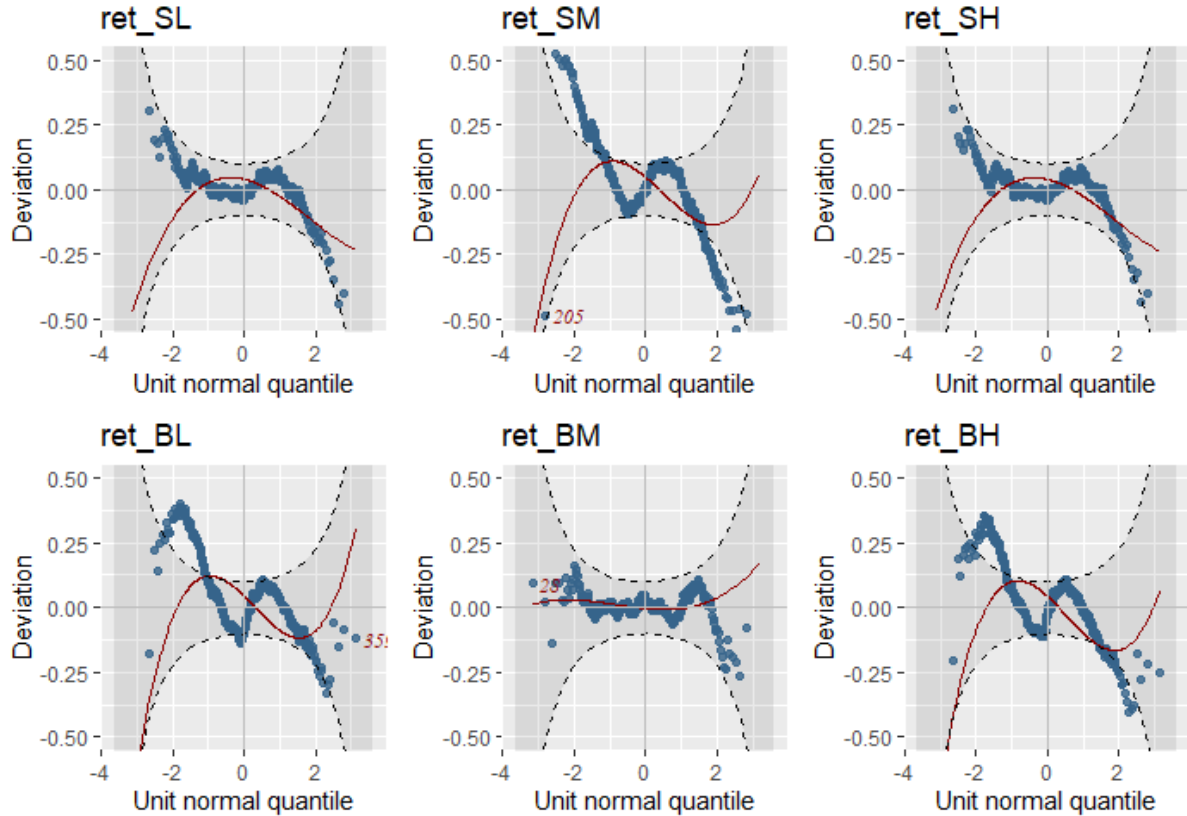
A curva vermelha resulta de ajustes cúbicos dos dados e é o padrão formado por ela que baliza a escolha do parâmetro (Mendonça, 2016). Para SL e SH, o padrão parece formar um “U” invertido que sugere assimetria negativa elevada; as demais carteiras formam um padrão que, a grosso modo, assemelha-se a um “S” curvado para baixo à esquerda e para cima à direita que indica que há necessidade de ajuste do parâmetro de curtose (Stasinopoulos *et al.*, 2015 *apud* Mendonça, 2016; Alcântara, 2018).

Esse último padrão é mais evidente em SM, BL e BH. Já para BM, a existência de um padrão é discutível, ainda que pareça seguir o padrão de “S” de forma mais pronunciada à direita. Por isso, para BM, BL, BH e SM, o modelo é reconsiderado para incluir o parâmetro de curtose ( $\tau$ ). As demais carteiras terão o parâmetro de assimetria ( $v$ ) como adição a sua estimação.

O objetivo em voltar a estimar os modelos GAMLSS, mas acrescentando a curtose à modelagem é verificar se o parâmetro é capaz de refinar o ajuste do modelo aos dados. Para isso, após a estimação, espera-se que os resíduos se encontrem mais próximos à linha de origem

e que todos, ou poucos, pontos estejam dentro da região cinza de significância.

Figura 8 – *Worm Plots* para o GAMLSS com Média (Ampliados)



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Para modelar esses parâmetros das carteiras, é necessário escolher as variáveis explicativas para este parâmetro que não necessariamente são as mesmas que o modelo inicial com a média como único parâmetro. Usando o critério de Akaike para fundamentar a escolha, tem-se que os fatores de mercado e de impulso são os melhores candidatos para explicar a curtose e assimetria de todas as carteiras. A exceção é SM, que tem sua curtose modelada apenas pelo fator de mercado.

Assim, os resultados para esses modelos estão apresentados a seguir na Tabela 8 com os coeficientes estimados para o excesso de retorno médio no Painel A e para o segundo parâmetro (assimetria ou curtose) do excesso de retorno no Painel B.

No Painel A, há algumas diferenças entre os valores estimados para o primeiro modelo GAMLSS. Os seis interceptos são todos próximos de zero como anteriormente e com significância estatística iguais e valores próximos quando em comparação com os interceptos na Tabela 5.

Tabela 8 – Resultados de Regressão GAMLSS com Dois Parâmetros

	SL	SM	SH	BL	BM	BH
Painel A ( $\mu$ )						
$\beta_0$	-0,0059 (-9,2080) ***	-0,0054 (-11,8741) ***	-0,0056 (-9,0366) ***	-0,0001 (-4,8883) ***	-0,0000 (-0,1415)	0,0007 (3,2392) **
$\beta_{CMKT}$	0,9454 (76,2158) ***	0,9238 (108,2296) ***	0,9455 (76,2870) ***	0,9875 (210,3965) ***	1,0030 (91,4326) ***	0,9999 (194,5317) ***
$\beta_{CSMB}$	-0,0034 (-0,3084)	-0,0021 (-8,5278) ***	0,0003 (0,2086)	-0,0000 (-0,2919)	-0,0019 (-0,5865)	-0,0026 (-19,8104) ***
$\beta_{CMOM}$	0,3955 (11,9087)	0,0422 (2,0074) *	0,3882 (11,9492) ***	-0,8691 (-69,8199) ***	-0,0119 (-0,4584)	1,0866 (77,0598) ***
Painel B						
	Assimetria ( $\nu$ )	Curtose ( $\tau$ )	Assimetria ( $\nu$ )	Curtose ( $\tau$ )	Curtose ( $\tau$ )	Curtose ( $\tau$ )
$\beta_0$	-0,1445 (-3,2535) **	-0,1341 (-12,3801) ***	-0,1096 (-2,4646) *	-0,2813 (-20,5704) ***	0,9440 (8,1483) ***	-0,3315 (-6,3013) ***
$\beta_{CMKT}$	5,1933 (9,2175) ***	-0,6126 (-4,4410) ***	5,1532 (9,5731) ***	-0,5702 (-3,4973) ***	-1,8481 (-1,6114)	-1,6901 (-2,1766) *
$\beta_{CMOM}$	23,0241 (11,7692) ***	-	21,4116 (10,7517) ***	1,0629 (5,2964) ***	-13,9661 (-2,8872) **	1,9506 (3,5642) ***

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Notas: Estatísticas t em parênteses. Significância estatística: \*\*\*: p-valor < 0; \*\*: p-valor < 0,001; \*: p-valor < 0,01; . : p-valor < 0,1.

Os coeficientes para o fator de mercado continuam todos estatisticamente significativos, positivos e próximos de um, mas seus valores são menores que os anteriores, ainda que essa diferença não seja expressiva.

Os coeficientes para o fator de tamanho também são semelhantes às anteriores (próximas de zero), mas há mudanças na significância estatística para BL cuja estimativa se torna estatisticamente insignificante.

Para o fator de impulso, os coeficientes são semelhantes em valor aos encontrados com o primeiro modelo GAMLSS, contudo vale notar que a estimativa da carteiras SM agora é significativa a partir do nível de 0%. As interpretações dos coeficientes no Painel A da Tabela 8 são semelhantes àquela da Tabela 5.

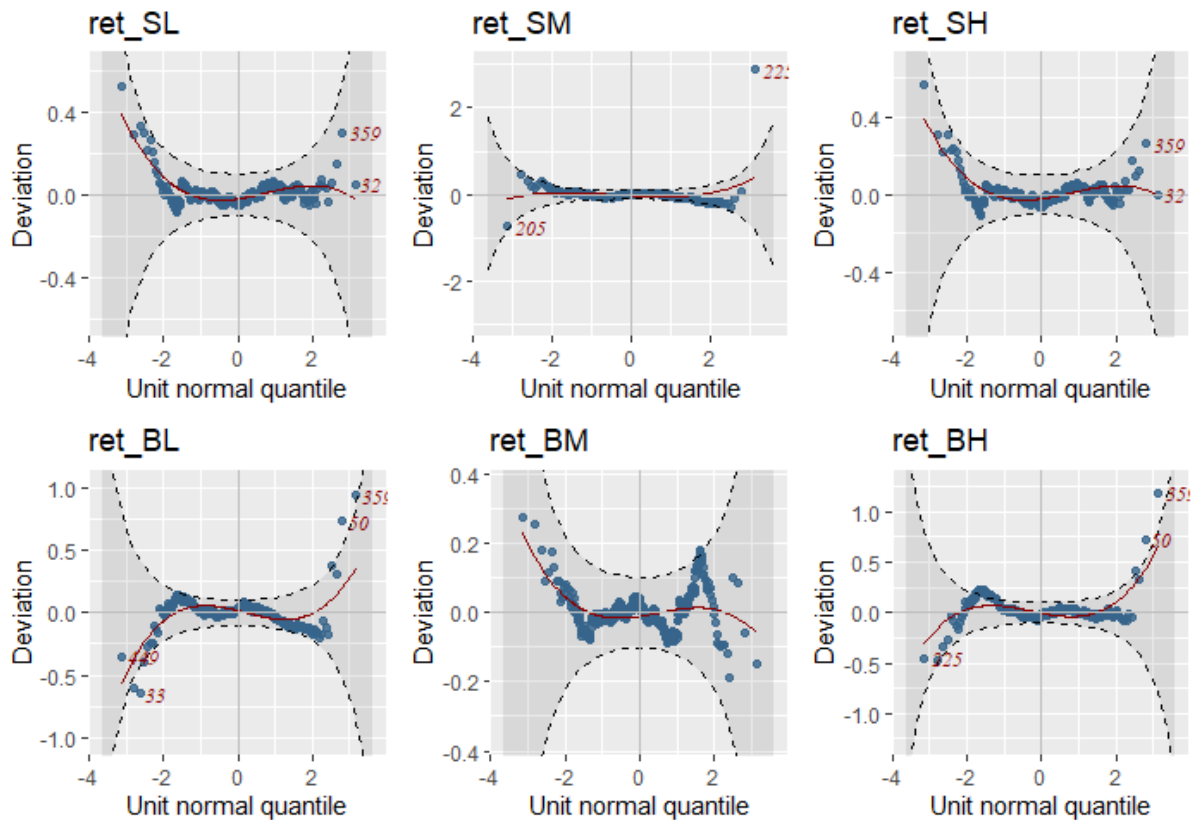
Para o Painel B, tem-se os coeficientes e o impacto desses na assimetria e curtose do excesso de retorno dos portfólios. Apenas as estimativas para o intercepto da carteira BM é positiva e todos são estatisticamente significantes variando os níveis de significância.

De maneira geral, o coeficiente para o fator de mercado apresenta uma discrepância entre as seis carteiras com valores entre -0,6126 e 5,1933 que mostram que uma variação de 1% nesse fator pode implicar desde uma redução de 0,6126% a um aumento de 5,1933% no segundo parâmetro dos excessos de retornos dos portfólios.

O fator de impulso apresenta os maiores coeficientes do Painel B, implicando variações de 21,4116% a 23,0241% na assimetria das carteiras SH e SL. Em valor absoluto, os coeficientes desse fator também são os maiores para as demais carteiras onde modela a curtose.

Os *worms plots* na Figura 9 abaixo mostram um melhor ajuste quando comparados com os da Figura 7 quando apenas a média é modelada. Ainda há alguns desvios significativos dos resíduos em relação aos quantis teóricos, porém grande parte deles está na área de confiança entre os semicírculos indicando ajuste satisfatório para as seis carteiras.

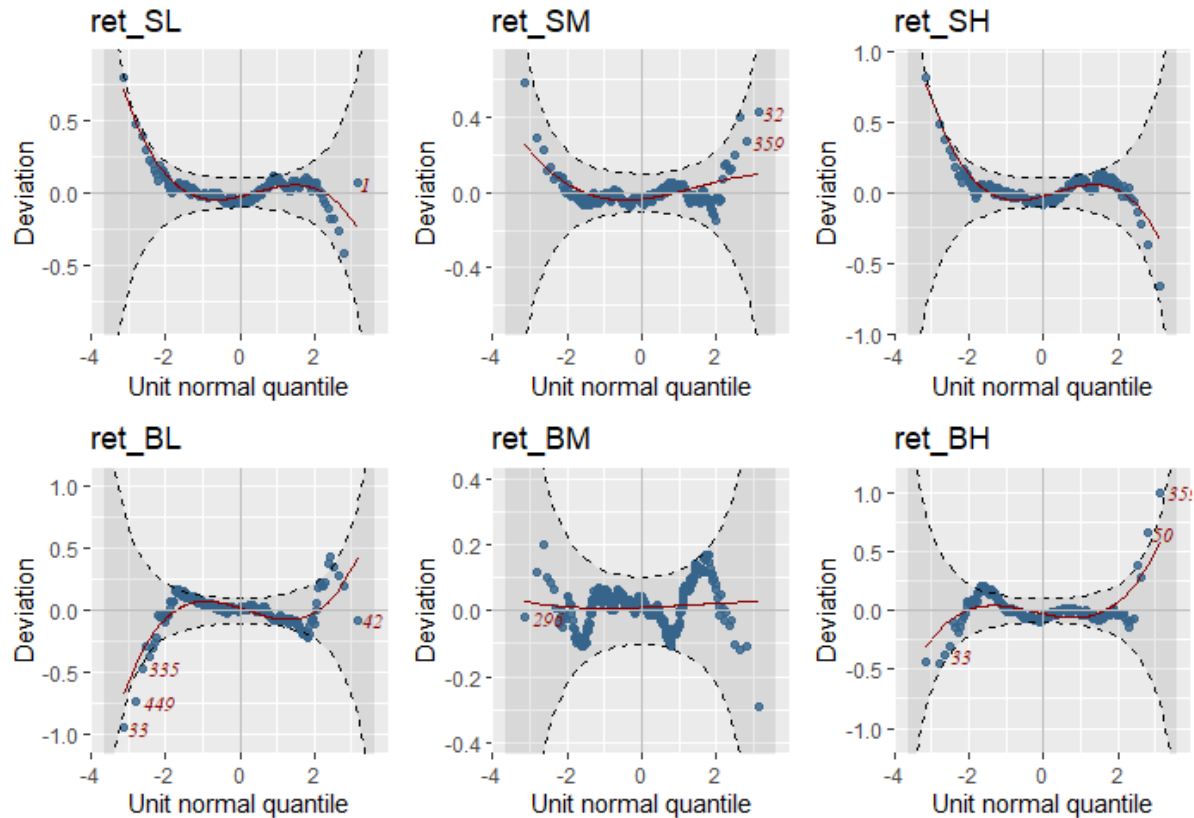
Figura 9 – Worm Plots para a Regressão GAMLSS com Média e um Parâmetro Adicional



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Esses resultados preliminares motivam a estimação de um modelo GAMLSS adicional onde a média, assimetria e curtose são modelados simultaneamente. Os resultados são apresentados nos *worm plots* na Figura 10.

Figura 10 – *Worm Plots* para GAMLSS com Três Parâmetros



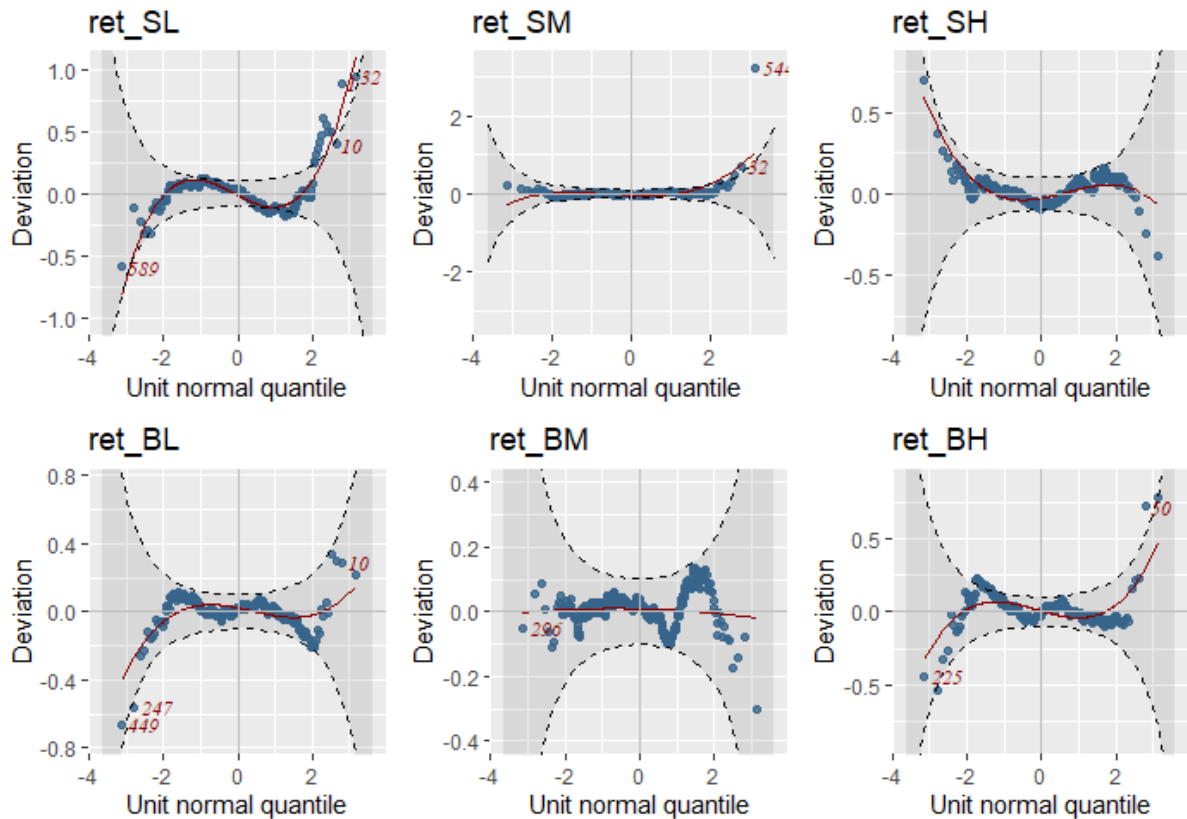
Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

De modo geral, não há mudanças significativas nos ajustes representados pelo *worm plots* da Figura 10 em relação aos da Figura 9; exceto para SM que neste caso tem todos os resíduos dentro da região de significância. Para os demais portfólios, Para SL, SH, BL e BM os resíduos mais extremos mostram desvio maior à reta de origem. No caso de BH, os *worm plots* não há diferenças notáveis. Ainda assim, a maioria dos resíduos se encontram na área cinza que indica um bom ajuste do modelo.

Por fim, também estima-se um modelo com os quatro parâmetros possíveis: média, variância, assimetria e curtose. Os *worm plots* deste modelo estão na Figura 11.

As carteiras SL e SH têm mais pontos fora da área cinza do que nos casos anteriores, indicando que adicionar o parâmetro de variância ao modelo com três parâmetros não melhora a qualidade de ajuste para essas carteiras. As quatro outras carteiras possuem mais de 95% dos resíduos entre as elipses e um comportamento próximo ao no caso anterior.

Figura 11 – Worm Plots para GAMLSS com Quatro Parâmetros



Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

Para finalizar a análise do modelo GAMLSS, as seis estatísticas de avaliação usadas anteriormente são calculadas para os modelos com dois e três parâmetros. A Tabela 9 apresenta os indicadores e destaca em **negrito** o valor mais favorável.

Segundo os três indicadores que utilizam o intercepto, o modelo com três parâmetros é o de melhor ajuste entre os modelos GAMLSS e frente ao MQO. Nota-se também os valores discrepantes entre essas três medidas para o GAMLSS com quatro parâmetros e os demais modelos da mesma classe sugerindo que há maior dispersão nesse modelo do que nos outros.

O erro absoluto médio continua indicando o modelo GAMLSS apenas com média, embora os valores dessa métrica sejam próximos entre as alternativas da mesma classe. Por sua vez, o MQO ainda é o modelo com menor raiz do erro médio quadrado sendo o único caso em que é recomendado em vez do GAMLSS. Finalmente, o pseudo- $R^2$  mostra que o modelo GAMLSS com quatro parâmetros explica, em média, 62,34% da variação dos excessos de retornos; maior valor entre os modelos, seguido por GAMLSS com três parâmetros (61,05%).

Tabela 9 – Avaliação de Desempenho (Comparação de Quatro Modelos)

Avaliação	MQO	GAMLSS			
		Média	Dois Parâmetros	Três Parâmetros	Quatro Parâmetros
$A a_i $	0,0269	0,0026	0,0032	<b>0,0013</b>	0,0176
$A a_i /A R_i^{\text{diff}} $	6,5513	0,6262	0,7706	<b>0,3215</b>	4,2800
$A a_i^2 /A (R_i^{\text{diff}})^2 $	42,5662	0,6359	0,9126	<b>0,2040</b>	92,5682
$A MAE $	0,0311	<b>0,0129</b>	0,0133	0,0140	0,0308
$A RMSE $	<b>0,0436</b>	0,0938	0,0910	0,0946	0,1330
$A \text{pseudo-R}^2 $	0,3094	0,5144	0,5813	0,6105	<b>0,6284</b>

Fonte: Elaboração própria a partir de dados de CoinGecko (2024) e French (2024).

De modo geral, os indicadores reforçam a conclusão baseada nas comparações anteriores de que os modelos de classe GAMLSS, considerando distribuições de probabilidades que não a normal, têm desempenho melhor que o MQO na precificação de criptomoedas. Além disso, entre os modelos GAMLSS, há evidências de que um modelo com mais parâmetros não é necessariamente o melhor.

A possibilidade de analisar a distribuição dos retornos em parâmetros além da média com modelos GAMLSS, permitindo uma melhor caracterização dos retornos é a principal contribuição do estudo para a Economia. Com curtose e assimetria respondendo significativamente aos fatores de precificação, o trabalho encontra evidências de que investidores exigem um retorno maior para ativos com maior risco de eventos extremos. Além disso, indicam que eles construiriam seus portfólios considerando a distribuição inteira de retornos e não apenas o retorno médio.

## 6. Conclusões:

As criptomoedas são uma classe de ativos atrativos aos investidores, especialmente por sua alta volatilidade e capacidade de gerar lucros rapidamente quando negociados especulativamente. A literatura apresenta casos de modelos tradicionais de precificação de ações adaptados às criptomoedas, mas ao estimar esses modelos por métodos mínimos quadrados, esses modelos assumem a hipótese de que os excessos de retornos seguem uma distribuição normal de probabilidade.

Como as evidências apresentadas neste trabalho indicam, essa hipótese não é adequada às criptomoedas visto que seus excessos apresentam distribuições assimétricas e de caudas pesadas. Partindo da hipótese de que distribuições com esses comportamentos são mais indicadas para a modelagem de criptomoedas, o modelo de precificação de três fatores para criptomoedas é estimado por MQO e por GAMLSS admitindo distribuições de probabilidade que apresentam essas características.

Comparando os modelos visualmente e por seis indicadores de desempenho, tem-se que, à exceção da raiz do erro quadrado médio, todos indicadores favorecem modelos de classe GAMLSS. A magnitude dos resultados não é muito diferente entre os métodos, mas há casos de mudança de significância dos coeficientes sugerindo a importância de um método que considere a distribuição da variável dependente.

Analisando o ajuste do modelo por meio de *worm plots* conclui-se que os modelos GAMLSS consistem em uma melhoria em relação ao MQO mas, que há possibilidade de aperfeiçoamento em trabalhos futuros. É sugerido que combinações de parâmetros sejam testadas em busca de um modelo com maior ajuste; visto que a adição de parâmetros de distribuição ao modelo, como assimetria e curtose, ao modelo geraram menos desvios nas extremidades da distribuição. Também faz-se a sugestão de outros métodos para determinar quais distribuições assumir para o caso de criptomoedas e verificar se elas melhoram o ajuste.

Por fim, dada a natureza recente destes ativos, a principal recomendação para próximas pesquisas é considerar diferentes fatores que sejam capazes de aumentar a potência explicativa dos modelos de precificação de criptomoedas. Além disso, como fatores de maior poder explicativo, é possível buscar evidências de que investidores exigem maiores retornos considerando eventos extremos que estariam indicados em parâmetros de curtose e assimetria.

#### **Agradecimentos:**

Os autores gostariam de agradecer o apoio à realização do presente trabalho dado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

#### **Conflito de Interesses:**

Os autores declaram não haver conflito de interesses na elaboração ou na submissão do presente trabalho.

**Referências:**

AKAIKE, Hirotugu. A new look at the statistical model identification. **IEEE transactions on automatic control**, [S.l.], v. 19, n. 6, p. 716–723, dez. 1974. Disponível em: <https://ieeexplore.ieee.org/document/1100705>. Acesso em 11 set. 2023.

ALCÂNTARA, Cássio de. **Modelos aditivos generalizados para posição, escala e forma (Gamlss) na modelagem da área miocárdica sob risco de Necrose**. Orientador: Dr. Edmilson Rodrigues Pinto. 2018. 51 f. Trabalho de Conclusão de Curso (Bacharelado) (Graduação) – Curso de Bacharelado em Estatística, Universidade Federal de Uberlândia, Uberlândia, 2018.

ALMEIDA, Alexandre; LOY, Adam; HOFMANN, Heike. ggplot2 Compatible Quantile-Quantile Plots in R. **The R Journal**, [S.l.], v. 10, n. 2, p. 248-261, 2018. Disponível em: <https://journal.r-project.org/archive/2018/RJ-2018-051/index.html>. Acesso em 30 out 2024.

AMMANN, M.; BURDORF, T.; LIEBI, L.; STÖCKL, S. **Survivorship and Delisting Bias in Cryptocurrency Markets**. [S.l.], nov. 2022. (Working paper). Disponível em: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4287573](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4287573). Acesso em: 4 ago. 2023.

BUUREN, Stef van; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in medicine**, [S.l.], v. 20, n. 8, p. 1259–1277, abr. 2001. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/sim.746>. Acesso em: 5 ago. 2023.  
COINGECKO. **CoinGecko**, 2024. Página inicial. Disponível em: <https://coingecko.com>. Acesso em: 2 out. 2024.

DUNBAR, Kwamie; OWUSU-AMOAKO, Johnson. Cryptocurrency returns under empirical asset pricing. **International Review of Financial Analysis**, [S.l.], v. 82, [s.n.], [n.p.], jul 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1057521922001776>. Acesso em 7 abr 2023.

FAMA, Eugene F.; FRENCH, Kenneth. R. A five-factor asset pricing model. **Journal of Financial Economics**, [S.l.], v. 116, n. 1, p. 1–22, abr. 2015. Disponível em: [www.sciencedirect.com/science/article/abs/pii/S0304405X14002323](http://www.sciencedirect.com/science/article/abs/pii/S0304405X14002323). Acesso: 11 abr. 2023.

FRENCH, Kenneth R. Data Library. **Kenneth R. French**, 2024. Disponível em: [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Acesso: 04 out. 2024

HORN, S. D.; HORN, R. A.; DUNCAN, D. B. Estimating heteroscedastic variances in linear models. **Journal of the American Statistical Association**, [S.l.], v. 70, n. 350, p. 380-385, 1975. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1975.10479877>. Acesso 17 out 2024.

HUIJBREGTS, Menno. **An Asset Pricing Model for Cryptocurrencies**. Orientador: Lauren A. P. Swinkels. Co-orientador: Ronald Huisman. 2020. 45 f. Dissertação (mestrado) — Master Financial Economics, Erasmus University Rotterdam, Rotterdam, 2020.

JARQUE, Carlos M.; BERA, Anil K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. **Economics letters**, [S.l.], v. 6, n. 3, p. 255-259, 1980. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0165176580900245>. Acesso em 10 set 2024.

JIA, Boxiang; GOODELL, John W.; SHEN, Dehua. Momentum or reversal: Which is the appropriate third factor for cryptocurrencies? **Finance Research Letters**, [S.l.], v. 45, [s.n.], [n.p.], mar. 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1544612321002208>. Acesso: jun 2024.

JIA, Yuecheng; LIU, Yuzheng; YAN, Shu. Higher moments, extreme returns, and cross-section of cryptocurrency returns. **Finance Research Letters**, [S.l.], v. 39, [s.n.], [n.p.], mar. 2021. Disponível em: 101536. <https://www.sciencedirect.com/science/article/abs/pii/S1544612320303135>. Acesso: 18 jun 2024.

LIU, Yukun; TSYVINSKI, Aleh; WU, Xi. Common risk factors in cryptocurrency. **The Journal of Finance**, [S.l.], v. 77, n. 2, p. 1133–1177, fev. 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/jofi.13119>. Acesso em: 31 jul. 2023.

MENDONÇA, Inês Rodrigues. **Estimação da creatinina sérica basal através de modelos GAMLSS**. Orientadora: Dra. Maria Fernanda Nunes Diamantino. Coorientadora: Dra. Ana Luísa Trigoso Papoila da Silva. 2016. 87 f. Dissertação (mestrado) – Mestrado em Bioestatística, Universidade de Lisboa, Lisboa, 2016.

OSTERRIEDER, Joerg. **The statistics of bitcoin and cryptocurrencies**. [S.l.], nov. 2016. (*Working paper*). Disponível em: [https://papers.ssrn.com/sol3/Papers.cfm?abstract\\_id=2872158](https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2872158). Acesso em: 1 ago. 2023.

PUNZO, Antonio; BAGNATO, Luca. **Asymmetric Laplace scale mixtures for the distribution of cryptocurrency returns**. [S.l.], set 2022. Disponível em: <https://arxiv.org/abs/2209.12848>. Acesso em 22 jul 2024.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Versão 4.3.1. Vienna, Áustria: R Foundation for Statistical Computing, 2023. <https://www.R-project.org>.

REGIS, Renan Oliveira. **Precificação de Ativos: Uma Análise dos Cinco Fatores de Fama e French em Modelos GAMLSS**. Orientador: Dr. Raydonal Ospina Martinez. Coorientador: Wilton Bernardino da Silva. 2021. 93 f. Tese (Doutorado) – Doutorado em Economia, Universidade Federal de Pernambuco, Recife, 2021.

REGIS, R.O.; OSPINA, R.; BERNARDINO, W.; CRIBARI-NETO, F. **Asset pricing in the Brazilian financial market: five-factor GAMLSS modeling**. *Empirical Economics*, [S.l.], v. 64, n. 5, p. 2373–2409, maio 2023. Disponível em: <https://link.springer.com/article/10.1007/s00181-022-02316-3>. Acesso: 1 ago. 2023.

RIGBY, Robert A.; STASINOPOULOS, D. Mikis. Generalized additive models for location, scale and shape. **Journal of Statistical Software**, [S.l.], v. 23, n. 7, p. 1–46, dez. 2007. Disponível em: <https://www.jstatsoft.org/article/view/v023i07>. Acesso: 31 jul. 2023.

\_\_\_\_\_. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society Series C: Applied Statistics**, [S.l.], v. 54, n. 3, p. 507–554, abr. 2005.

Disponível em: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2005.00510.x>. Acesso: 31 jul. 2023.

RIGBY, R.; STASINOPOULOS, M.; HELLER, G.; DE BASTIANI, F. **Distributions for Modelling Location, Scale and Shape: Using GAMLSS in R**. [S.l.], 2017. Disponível em: <https://www.gamlss.com/wp-content/uploads/2023/06/DistributionsForModellingLocationScaleandShape-1.pdf>. Acesso: 12 out 2024.

SCHWARZ, Gideon. Estimating the dimension of a model. **The Annals of Statistics**, [S.l.], v. 16, n. 2, p. 461–464, mar. 1978. Disponível em: <https://www.jstor.org/stable/2958889>. Acesso: 11 set. 2023.

SHEN, Dehua; URQUHART, Andrew; WANG, Pengfei. A three-factor pricing model for cryptocurrencies. **Finance Research Letters**, [S.l.], v. 34, [s.n.], [n.p.], maio 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1544612319304519>. Acesso: 31 jul. 2023.

STASINOPOULOS, Mikis D.; RIGBY, Robert A.; BASTIANI, Fernanda De. GAMLSS: A distributional regression approach. **Statistical Modelling** [S.l.], v. 18, n. 3-4, p. 248-273, 2018. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1471082X18759144>. Acesso: 27 jul 2024.

STASINOPOULOS, M.; RIGBY, B.; VOUDOURIS, V.; AKANTZILIOTOU, C.; ENEA, M.; KIOSE, D. **gamlss: Generalised Additive Models for Location Scale and Shape**. Versão 5.4-12. [Londres], 2023. <https://www.gamlss.com/>

SZCZYGIELSKI, Jan Jakub; KARATHANASOPOULOS, Andreas; ZAREMBA, Adam. One shape fits all? A comprehensive examination of cryptocurrency return distributions. **Applied Economics Letters**, [S.l.], v. 27, n. 19, p. 1567–1573, dez. 2023. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/13504851.2019.1697420>. Acesso: 1 ago. 2023.