

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

U LISBOA

UNIVERSIDADE
DE LISBOA

Backtesting Expected Shortfall

André Lourenço Tarelho

Mestrado em
Matemática Financeira

Orientador:

Professor Doutor, Joaquim Paulo Viegas Ferreira de Carvalho, Professor Auxiliar
Convidado,
ISCTE-IUL

Setembro, 2025

Departamento de Finanças
Departamento de Matemática

Backtesting Expected Shortfall

André Lourenço Tarelho

Mestrado em
Matemática Financeira

Orientador:

Professor Doutor, Joaquim Paulo Viegas Ferreira de Carvalho, Professor Auxiliar
Convidado,
ISCTE-IUL

Setembro, 2025

Agradecimento

O meu sincero agradecimento ao Professor Paulo Viegas de Carvalho, pela orientação, ajuda e disponibilidade que demonstrou durante o desenvolvimento desta dissertação.

Obrigado aos meus pais por sempre me darem força para alcançar os meus objetivos e acreditarem em mim. Obrigado à minha irmã, avós, tios e primos pelo apoio incondicional que sempre me deram.

Obrigado a todos os meus amigos e colegas que conheci na licenciatura, que sempre me apoiaram e ajudaram a chegar ao fim deste percurso, em particular, ao José, à Beatriz, ao João, à Rita e à Luana.

Obrigado a todos os meus amigos que me acompanham desde sempre e que me deram sempre o apoio para ultrapassar todos os desafios.

Por fim, obrigado aos meus colegas da equipa de Risco da Caixa Geral de Depósitos pelo apoio e pela motivação, em especial à Beatriz.

Resumo

O *Expected Shortfall* (ES) tem vindo a ganhar relevância como medida de risco de mercado, nomeadamente após a sua adoção pelo enquadramento regulatório de Basileia, em substituição do *Value at Risk* (VaR). Apesar de apresentar vantagens claras na captura de perdas extremas, o ES levanta dificuldades acrescidas no processo de backtesting, dada a ausência de um procedimento simples e amplamente aceite para a sua validação. Esta dissertação analisa diferentes metodologias propostas na literatura para o backtesting do ES, avaliando a sua aplicabilidade prática e comparando o desempenho de distintos modelos de estimação de risco. Através de uma aplicação empírica sobre dados de mercado, procura-se identificar as vantagens e limitações de cada abordagem e discutir as suas implicações para a gestão do risco e para o cumprimento dos requisitos regulamentares.

Palavras-chave: Expected Shortfall, Value at Risk, Backtesting, Regulação Financeira

Classificação JEL: G28, G32

Abstract

The *Expected Shortfall* (ES) has gained increasing relevance as a market risk measure, particularly following its adoption by the Basel regulatory framework as a replacement for the traditional *Value at Risk* (VaR). While ES offers clear advantages in capturing extreme losses, it poses additional challenges for backtesting, given the absence of a simple and widely accepted validation procedure. This dissertation analyzes different methodologies proposed in the literature for backtesting ES, assessing their practical applicability and comparing the performance of different risk estimation models. Through an empirical application on market data, the study seeks to identify the advantages and limitations of each approach and to discuss their implications for risk management and regulatory compliance.

Keywords: Expected Shortfall, Value at Risk, Backtesting, Financial Regulation

JEL Classification: G28, G32

Índice

Agradecimento	i
Resumo	iii
Abstract	v
Lista de Tabelas	ix
Lista de Figuras	xi
Glossário	xiii
Capítulo 1. Introdução	1
Capítulo 2. Revisão da Literatura	3
Capítulo 3. Definições	7
3.1. Propriedades das Medidas de Risco	7
3.1.1. Coerência e propriedades relacionadas	7
3.1.2. Elicitabilidade	8
3.1.3. <i>Backtestability</i>	8
3.2. Value at Risk (VaR)	9
3.3. Expected Shortfall	9
3.4. VaR vs Expected Shortfall	10
3.5. Backtesting	11
3.5.1. Backtesting VaR	12
3.5.2. Backtesting Expected Shortfall	13
Capítulo 4. Metodologia	21
4.1. Descrição da amostra	21
4.2. Metodologia	22
4.2.1. Modelos ARCH e GARCH	23
4.2.2. Normal e GARCH-Normal	23
4.2.3. H&W	25
Capítulo 5. Dados e Análise	27

vii

5.1. Resultados dos Testes de Backtesting	27
5.2. Requisitos de Capital	32
5.3. Impactos de uma variação na carteira	34
Capítulo 6. Conclusões	39
Referências Bibliográficas	43
Anexo A : Descrição do portfólio	45
Anexo B: Estimação dos Valores de VaR/ES	47

Lista de Tabelas

3.1 Limites das zonas de backtesting do FRTB	12
3.2 Classificação por zonas do teste M&P	20
4.1 Exposição da carteira a fatores de risco	22
5.1 Resultados dos testes em janela deslizante	28
5.2 Classificação dos testes em janelas discretas - Modelo <i>Norm</i>	30
5.3 Classificação dos testes em janelas discretas - Modelo <i>GARCH-Norm</i>	31
5.4 Classificação dos testes em janelas discretas - Modelo H&W	32
5.5 Multiplicadores de capital Basel e M&P	33
5.6 Nova exposição da carteira	35
5.7 Resultados dos testes em janela deslizante - Nova carteira	36
5.8 Classificação dos testes de backtesting em janelas discretas para a nova carteira - Modelo <i>GARCH-Norm</i>	37
6.1 Estatísticas descritivas do portfólio	45

Lista de Figuras

3.1 VaR 99% vs ES 97,5%	11
5.1 Comparação entre os multiplicadores de capital para o GARCH-NORM	34
5.2 Comparação dos multiplicadores de capital entre as carteiras	38
6.1 PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo Norm	47
6.2 PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo GARCH-Norm	47
6.3 PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo H&W	48

Glossário

A - Zona Amarela

ARCH - Autoregressive Conditionally Heteroskedastic

BCBS - Basel Committee on Banking Supervision

ES - Expected Shortfall

EUR - Euro

EWMA - Exponentially Weighted Moving Averages

FRTB - Fundamental Review of the Trading Book

GARCH - Generalized Autoregressive Conditional Heteroskedasticity

GBP - British Pound Sterling

HKD - Hong Kong Dollar

H&W - Hull e White

LRT - Likelihood Ratio Test

M&P - Teste de Moldenhauer and Pitera

MN - Distribuição Multinomial

P&L - Lucros e Perdas

R - Zona Vermelha

USD - United States Dollar

V - Zona Verde

VaR - Value at Risk

Z1 - Teste 1 de Acerbi e Szekely

Z2 - Teste 2 de Acerbi e Szekely

ZR - Ridge test de Acerbi e Szekely

CAPÍTULO 1

Introdução

O *Value at Risk* (VaR) consolidou-se, nas últimas décadas, como a medida de risco mais utilizada no setor financeiro. A sua popularidade deve-se sobretudo à simplicidade de interpretação, sendo que o VaR corresponde à maior perda esperada num horizonte temporal, com base num dado nível de confiança. Em termos práticos, fornece aos gestores de risco e reguladores uma estimativa clara e objetiva do potencial de perda, tornando-se uma referência central para o cálculo de requisitos de capital em instituições financeiras.

Apesar da sua ampla utilização, o VaR apresenta algumas limitações relevantes. Uma das críticas mais recorrentes prende-se com a sua insensibilidade a eventos extremos. Embora o VaR quantifique a probabilidade de ocorrência de perdas superiores a um certo valor, não fornece informação sobre a magnitude dessas perdas. Assim, em períodos de elevada volatilidade ou sob cenários de crise financeira, o VaR pode subestimar de forma significativa o risco. Além disso, a escolha do nível de confiança não elimina a incerteza associada às perdas que ultrapassam esse limiar, limitando a sua utilidade na gestão do risco extremo.

Como resposta a estas limitações, a comunidade académica e os reguladores têm vindo a promover a adoção do *Expected Shortfall* (ES). Ao contrário do VaR, o ES consegue captar a frequência das perdas acima do valor do VaR, bem como a sua severidade, correspondendo ao valor esperado das perdas que ocorrem além do quantil definido para o VaR. Desta forma, o ES é uma medida mais coerente e prudente, refletindo melhor o impacto de choques extremos nos mercados. Por este motivo, o novo enquadramento regulamentar de Basileia (FRTB – *Fundamental Review of the Trading Book*) determinou a substituição do VaR pelo ES no cálculo dos requisitos de capital regulamentar.

Contudo, a adoção do ES coloca desafios significativos do ponto de vista do backtesting. Enquanto o VaR pode ser facilmente validado através da contagem de exceções, isto é, a comparação entre o número observado de quebras e o esperado para um dado nível de confiança, o mesmo não acontece com o ES. Como o ES se baseia na média das perdas extremas, não existe um teste direto e universalmente aceite que permita verificar de forma simples a sua qualidade preditiva. Esta dificuldade em avaliar empiricamente o desempenho do ES motivou o desenvolvimento, na literatura recente, de diversos métodos de backtesting, cada um com diferentes pressupostos estatísticos e graus de complexidade.

A relevância prática deste tema é evidente. Uma vez que os requisitos de capital das instituições financeiras dependem diretamente dos resultados dos testes de backtesting, torna-se

essencial assegurar que as metodologias utilizadas para validar o ES são robustas, fiáveis e consistentes ao longo do tempo. Modelos que subestimem o risco podem levar a insuficiência de capital em cenários de crise, enquanto modelos excessivamente conservadores podem conduzir a custos de capital desnecessariamente elevados, afetando a competitividade das instituições.

Neste contexto, a presente dissertação tem como objetivo analisar e comparar alguns dos métodos mais relevantes de backtesting do Expected Shortfall, aplicando-os a diferentes modelos de estimação de risco. Pretende-se avaliar em que medida é possível realizar um backtesting eficaz do ES, identificando as vantagens, limitações e implicações práticas de cada abordagem. Com esta análise, procura-se contribuir para a compreensão do problema e oferecer evidência empírica sobre a adequação das metodologias propostas.

Assim sendo, a estrutura da tese organiza-se da seguinte forma. Após esta introdução, o segundo capítulo apresenta uma revisão da literatura, contextualizando a evolução das medidas de risco e dos testes de backtesting aplicáveis. O terceiro capítulo descreve os dados utilizados e as metodologias implementadas, tanto no que respeita à estimação do VaR e do ES, como relativamente aos testes de validação. O quarto capítulo expõe os resultados empíricos obtidos, discutindo-os em função das implicações para a prática de gestão de risco e para o enquadramento regulamentar. Finalmente, o quinto capítulo apresenta as conclusões, resumindo os principais contributos do trabalho e sugerindo possíveis linhas de investigação futura.

CAPÍTULO 2

Revisão da Literatura

Durante muitos anos, o *Value at Risk* (VaR) foi considerado a medida de risco predominante no setor financeiro, amplamente utilizada tanto por instituições financeiras como por reguladores para quantificar o risco de mercado. No entanto, é amplamente conhecido que esta medida não pertence à classe das medidas de risco coerentes, conforme definidas por Artzner, Delbaen, Eber, e Heath (1999), uma vez que não satisfaz a propriedade de subaditividade. Esta limitação implica que a diversificação de carteiras pode, em certos casos, conduzir a um VaR agregado superior à soma dos VaR individuais, contrariando um princípio fundamental da gestão de risco.

Para um determinado nível de confiança α , o VaR corresponde ao quantil α da distribuição de lucros e perdas (P&L), podendo ser estimado de diversas formas, como assumindo uma distribuição paramétrica dos retornos ou utilizando valores históricos.

O *Expected Shortfall* (ES) pode ser calculado a partir do VaR e representa o valor médio das perdas que excedem o nível do VaR. Esta medida tem sido progressivamente adotada como referência por parte de reguladores e instituições financeiras, sobretudo após a sua consagração no enquadramento do *Fundamental Review of the Trading Book* (FRTB) do *Basel Committee on Banking Supervision* (BCBS). Em contraste com o VaR, o ES proporciona uma visão mais completa do risco extremo, ao considerar a média das perdas além do quantil correspondente e é considerado uma medida de risco coerente, conforme demonstrado por Acerbi e Tache (2002). Contudo, a transição do VaR para o ES trouxe desafios significativos em termos de validação empírica, uma vez que a natureza do ES torna o seu *backtesting* mais complexo e menos direto.

A possibilidade de realizar o *backtesting* do ES chegou a ser posta em causa quando Gneiting (2011) demonstrou que o ES não pode ser considerado uma medida de risco eliciável. Para que uma medida de risco seja eliciável, é necessário que minimize o valor esperado de uma função *score*. Esta prespetiva foi posteriormente apoiada por outros autores, como Chen (2014) e Carver (2013). Contudo, esta ideia foi contrariada por outros autores, como Acerbi e Szekely, que mostraram ser possível o *backtesting* do ES, mesmo não sendo esta medida eliciável, defendendo ainda que a elicitabilidade nada tem a ver com o *backtesting*. Embora o ES não seja eliciável de forma isolada, Nolde e Ziegel (2017) demonstraram que o par (VaR,ES) é conjuntamente eliciável, o que fornece uma base teórica para testes combinados que exploram a informação do quantil e da média da cauda.

Enquanto que o VaR é avaliado através da contagem de exceções, o ES, por constituir uma medida da perda média na cauda da distribuição, não possui um critério binário natural de violação. Esta limitação motivou diversas investigações académicas e práticas para desenvolver metodologias de *backtesting* adequadas ao ES, que capturassem tanto a frequência, como a severidade das perdas extremas.

A literatura existente sobre o *backtesting* do VaR é extensa, com importantes contributos, como o teste de Kupiec (1995) e o teste de Christoffersen (1998), que se tornaram padrões na indústria para avaliar a precisão desta medida. No entanto, essas abordagens revelaram-se insuficientes para o ES, uma vez que se limitam a verificar a existência de exceções e ignoram completamente a magnitude das perdas que ocorrem além do VaR.

Já a literatura sobre o *backtesting* do ES é mais reduzida, embora tenha vindo a crescer de forma consistente ao longo dos últimos anos, sobretudo após a publicação do FRTB. Inicialmente, surgiram propostas para estender ou adaptar estes testes de VaR para o ES, muitas vezes com fraco poder estatístico ou sensibilidade inadequada à subestimação do risco extremo. Esta constatação motivou o desenvolvimento de diferentes classes de testes especificamente desenhados para o ES, procurando avaliar de forma mais robusta e informativa tanto a frequência como a severidade das perdas na cauda da distribuição.

Uma das abordagens inovadoras foi proposta por Kratz et al. (2018) através dos testes multinomiais implícitos. O princípio abaseia-se na representação do ES como uma média ponderada de múltiplos quantis, o que permite aplicar testes estatísticos ao número de exceções observadas em diferentes níveis de VaR. Desta forma, o problema contínuo de verificação do ES é convertido num conjunto de avaliações discretas, mais simples de implementar e analisar.

Esta técnica permite a utilização de testes clássicos, como o teste de Pearson, para avaliar simultaneamente múltiplos níveis de risco, captando assim mais informação sobre a estrutura da cauda da distribuição. No entanto, este método pode apresentar distorções no tamanho do teste, sobretudo quando aplicada a quantis pequenos ou amostras de dimensão reduzida. Para mitigar este problema, Bayer e Dimitriadis (2019) propuseram alternativas, como a utilização do teste de Nass adaptado ao contexto multinomial, com o objetivo alinhar melhor a taxa de rejeição com o nível de significância definido e, assim, aumentar a fiabilidade das conclusões estatísticas.

Uma outra linha de investigação importante centra-se em testes que comparam o ES previsto com uma *proxy* empírica obtida através das perdas observadas. Entre os trabalhos desta área, destacam-se os de Acerbi e Szekely, que propõem o uso de estatísticas contínuas e diferenciáveis para medir a diferença entre o ES estimado e as perdas verificadas.

Esta abordagem leva à construção de testes, como os testes Z, que comparam diretamente as previsões com as perdas observadas, penalizando de forma contínua os desvios negativos.

Estes testes são facilmente computacionais, têm uma interpretação clara e permitem a estimação do valor-p por simulação Monte Carlo, uma vez que não possuem uma distribuição conhecida. No entanto, os próprios autores reconhecem que alguns destes testes, como o Z2, podem ser excessivamente sensíveis a pequenas imprecisões na estimativa do VaR, conduzindo a potenciais erros de classificação.

Já Bayer e Dimitriadis introduziram uma abordagem baseada em regressões condicionais para o *backtesting* do ES. O denominado *Intercept ESR Backtest* verifica se o termo constante de uma regressão entre as perdas observadas e o ES previsto é estatisticamente igual a zero, sob a hipótese de um modelo previamente especificado. Já o *Likelihood Ratio ESR Test* avalia simultaneamente se o termo constante e o declive da regressão são iguais a um, permitindo testar conjuntamente a ausência de enviesamento e a correta calibração do ES previsto.

Estes testes têm a vantagem de se basearem em modelos estatísticos clássicos e de serem facilmente interpretáveis. Além disso, apresentam bom desempenho em termos de poder estatístico, sendo capazes de detetar subestimações sistemáticas do risco. A inclusão da regressão permite também uma análise mais completa, captando desvios na calibração dos modelos de risco e fornecendo informação adicional sobre o relacionamento entre as previsões e as perdas observadas. Esta abordagem destaca-se das anteriores por avaliar o ES de forma direta, sem depender da informação do VaR, o que a torna particularmente útil em contextos onde a estimação do VaR possa ser menos fiável, ou não esteja disponível.

O *backtesting* do ES permanece um desafio técnico relevante, resultante da ausência de um critério binário de violação e das características desta medida de risco. As abordagens analisadas apresentam diferentes pontos fortes e limitações, variando no equilíbrio entre a robustez estatística, simplicidade de implementação e adequação a diferentes contextos de modelização. Esta diversidade de propostas demonstra que não existe uma solução universalmente superior, mas antes um conjunto de ferramentas cuja escolha deve considerar as características do modelo, o horizonte temporal e o objetivo da validação. A partir deste enquadramento, a presente dissertação irá implementar e comparar empiricamente diferentes classes de testes, de forma a avaliar o seu desempenho na validação de modelos de risco sob condições realistas de mercado.

CAPÍTULO 3

Definições

Com este capítulo, pretende-se entender algumas das propriedades matemáticas das medidas de risco, fornecer definições formais de VaR e de Expected Shortfall, bem como discutir as suas diferenças. Pretende-se também explorar o conceito de elicitabilidade e descrever como o backtesting destas duas medidas é feito hoje em dia.

3.1. Propriedades das Medidas de Risco

3.1.1. Coerência e propriedades relacionadas

Artner et al. (1999) introduziram o conceito de Medidas de Risco Coerentes, que necessitam de cumprir um conjunto de axiomas. A definição que se segue é baseada na de Emmer et al. (2015).

DEFINIÇÃO 3.1. *Uma medida de risco ρ diz-se **coerente** se são cumpridas as seguintes condições:*

- (1) **Homogeneidade:** ρ é homogénea se, para todas as variáveis aleatórias P e para $h \geq 0$:

$$\rho(hP) = h\rho(P)$$

- (2) **Subaditividade:** ρ é subaditiva se, para todas as variáveis aleatórias P_1 e P_2 :

$$\rho(P_1 + P_2) \leq \rho(P_1) + \rho(P_2)$$

- (3) **Monotonia:** ρ é monótona se, para todas as variáveis aleatórias P_1 e P_2 :

$$P_1 \leq P_2 \implies \rho(P_1) \leq \rho(P_2)$$

- (4) **Invariância por Translação:** ρ é invariante por translação se, para todas as variáveis aleatórias P e para $a \in \mathbb{R}$:

$$\rho(P - a) = \rho(P) - a$$

Do ponto de vista financeiro, cada um dos axiomas apresentados tem como objetivo capturar uma característica desejável numa medida de risco. Estes axiomas são essenciais para garantir que uma medida de risco se comporta de forma intuitiva e financeiramente sensata. No entanto, como será analisado mais adiante, nem todas as medidas de risco amplamente utilizadas satisfazem os quatro axiomas de coerência.

3.1.2. Elicitabilidade

O conceito de elicitação foi introduzido por Osband(1985) e aprofundado por Lambert et al. (2008), constitui uma propriedade matemática importante para a avaliação do desempenho de modelos de previsão. A definição que se apresenta de seguida é baseada na apresentada por Acerbi e Szekely (2017).

DEFINIÇÃO 3.2. Uma estatística $\mathbf{y}(X)$ diz-se **eliciável** se existir uma função score $S_{\mathbf{y}}(y, x)$ tal que $\mathbf{y}$ minimize o valor esperado dessa função, isto é,

$$\mathbf{y}(X) = \arg \min_y \mathbb{E}[S_{\mathbf{y}}(y, X)] \quad (3.1)$$

Assim sendo, a estatística $\mathbf{y}(X)$ pode ser interpretada como a melhor previsão possível, no sentido em que minimiza, em valor esperado, uma função de penalização, ou seja, quanto maior o valor desta função, pior a qualidade da previsão. Este conceito pode ser particularmente relevante no contexto das medidas de risco, no sentido em que permite identificar aquelas para as quais é possível construir critérios de avaliação objetivos e rigorosos.

3.1.3. Backtestability

Acerbi e Szekely (2017) destacaram o conceito de *Backtestability*, propondo uma abordagem para avaliar e ordenar previsões de medidas de risco. Nesta abordagem, previsões de menor qualidade são as que se afastam de zero. Esta propriedade permite-nos ainda a formulação de um teste de hipóteses capaz de avaliar as previsões geradas por um determinado modelo.

Apresenta-se de seguida a definição de *backtestable*, proposta por Acerbi e Szekely (2017):

DEFINIÇÃO 3.3. Uma estatística $\mathbf{y}$ de uma variável aleatória X é **backtestable**, se existe uma função $Z_{\mathbf{y}}(y, x)$, tal que

$$\mathbb{E}(Z_{\mathbf{y}}(y, x)) = 0$$

e,

$$\mathbb{E}(Z_{\mathbf{y}}(y_1, x)) < \mathbb{E}(Z_{\mathbf{y}}(y_2, x)), \text{ se } y_1 < y_2$$

Esta definição estabelece que a estatística é considerada *backtestable* se for possível construir uma função cujos valores esperados identificam previsões corretas (valor esperado nulo) e penalizam previsões incorretas de forma ordenada. Esta propriedade é importante para a validação de modelos de risco, sobretudo no caso de medidas em que não é tão direta a construção de procedimentos de avaliação.

Mais adiante, analisaremos de que forma estas propriedades se relacionam entre si e influenciam a escolha e avaliação das medidas de risco. De seguida, serão definidas as medidas de risco que servirão de base à análise desenvolvida ao longo deste trabalho.

3.2. Value at Risk (VaR)

O *Value at Risk* (VaR) é uma medida de risco amplamente utilizada na gestão de risco, destinada a quantificar a perda potencial máxima de um ativo ou portfólio, num determinado horizonte temporal, para um dado nível de confiança.

De acordo com Emmer et al. (2001), o VaR é formalmente definido da seguinte forma:

DEFINIÇÃO 3.4. *Considere-se uma variável aleatória L , que representa o lucro ou perda associado a um investimento num portfólio. O Value at Risk, ao nível de confiança $\alpha \in]0, 1[$, é definido como o quantil de ordem α da distribuição das perdas ao longo de um período pré-definido. Ou seja:*

$$VaR_\alpha(L) = q_\alpha(L) = \inf\{\ell : \mathbb{P}(L \leq \ell) \geq \alpha\} \quad (3.2)$$

Este resultado pode ser interpretado da seguinte maneira: “O VaR corresponde ao valor máximo que posso perder, com uma probabilidade de 99% durante um determinado período”.

É ainda útil salientar que, se os quantis forem calculados sob uma distribuição na qual os preços são normalmente distribuídos, então satisfazem a propriedade da subaditividade, desde que o nível de confiança seja inferior a 0,5. Neste caso, sendo L uma variável aleatória normal, temos que:

$$VaR_\alpha(L) = -\left(\mathbb{E}_{\mathbb{P}}[L] + \Phi^{-1}(\alpha) \cdot \sigma_{\mathbb{P}}(L)\right), \quad (3.3)$$

onde Φ representa a função de distribuição acumulada da normal padrão.

O VaR tornou-se uma ferramenta fundamental na regulamentação financeira, sendo amplamente utilizado pelas instituições bancárias para o cálculo dos requisitos mínimos de capital, conforme estabelecido nos acordos de Basileia II e Basileia III. Contudo, apesar da sua ampla aceitação, o VaR possui algumas limitações relevantes, como o facto de não ser, em geral, uma medida coerente de risco, devido à possível violação da propriedade da subaditividade, conforme demonstrado por Artzer et al. (1999).

Além disso, o VaR não fornece qualquer informação sobre a magnitude das perdas que excedem o valor limite, o que motivou o desenvolvimento de medidas alternativas. Uma dessas medidas é o Expected Shortfall, que será abordado na secção seguinte.

3.3. Expected Shortfall

Enquanto que o VaR indica a perda máxima apenas com um determinado nível de confiança, o Expected Shortfall consegue fornecer um maior detalhe ao considerar a perda média das observações mais penalizadoras, acima do valor do VaR. De uma forma mais formal, dado o conjunto de notações introduzidas anteriormente

$$ES_\alpha = -\mathbb{E}[L|L \leq -VaR_\alpha] \quad (3.4)$$

Tal como no caso do VaR, será útil observar que, se os quantis forem calculados assumindo que as perdas seguem uma distribuição normal, então existe também uma forma fechada para o *Expected Shortfall*. Conforme demonstrado por MCNeil et al. (2015), temos:

$$ES_\alpha = \mathbb{E}_\mathbb{P}[L] + \sigma \cdot \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha} \quad (3.5)$$

Onde ϕ é a função densidade de probabilidade da distribuição normal padrão.

A nível regulamentar, o VaR tem vindo a ser progressivamente substituído pelo *Expected Shortfall*, uma vez que este apresenta algumas propriedades vantajosas, nomeadamente a coerência, que o VaR, em geral, não satisfaz. Estas diferenças serão exploradas com maior detalhe na secção seguinte.

3.4. VaR vs Expected Shortfall

As duas principais medidas de risco utilizadas pelas instituições financeiras são o *Value at Risk* (VaR) e o *Expected Shortfall* (ES), conforme definidas nas secções anteriores. Apesar de o VaR não ser, em geral, uma medida de risco coerente, uma vez que não satisfaz a propriedade da subaditividade, tem sido a medida dominante na regulação bancária ao longo das últimas décadas. É também a medida utilizada no regime de regulação *Solvência II*, aplicável ao setor regulador na Europa, onde o Requisito de Capital de Solvência é definido como o VaR 99,5%.

Já o ES destaca-se do VaR por ser uma medida de risco coerente, conforme demonstrado por Acerbi e Tasche (2002). Para além disso, por ser uma medida de risco mais sensível a perdas acima do valor do VaR, o ES tem vindo a atrair uma maior atenção por parte dos reguladores. Como resultado da *Fundamental Review of the Trading Book* (FRTB), o ES a 10 dias, com um nível de confiança de 97,5%, foi adotado como a principal medida de risco para determinar os requisitos de capital regulamentar associados ao *trading book*, ao abrigo do novo quadro regulamentar estabelecido pelas normas de Basileia III.

De acordo com Emmer et al. (2015), na prática, a violação da propriedade da subaditividade pelo VaR não constitui um problema relevante, desde que os riscos subjacentes contenham variância ou média finita. No entanto, esta limitação torna-se mais crítica em contextos de cauda pesada, onde o comportamento extremo das perdas é mais provável.

A escolha do nível de confiança de 97,5% no ES, em substituição do VaR a 99%, prende-se com o facto de que, para distribuições gaussianas, os valores obtidos por ambas as medidas são praticamente idênticos. Contudo, para observações com caudas mais pesadas, o ES a

97,5% tende a ser superior ao VaR 99%, funcionando como um mecanismo de penalização para este tipo de risco, através da imposição de requisitos de capital mais exigentes.

No gráfico abaixo, podemos encontrar as diferenças entre o ES 97,5% e o VaR 99% para as distribuições normal e t-student com ν graus de liberdade.

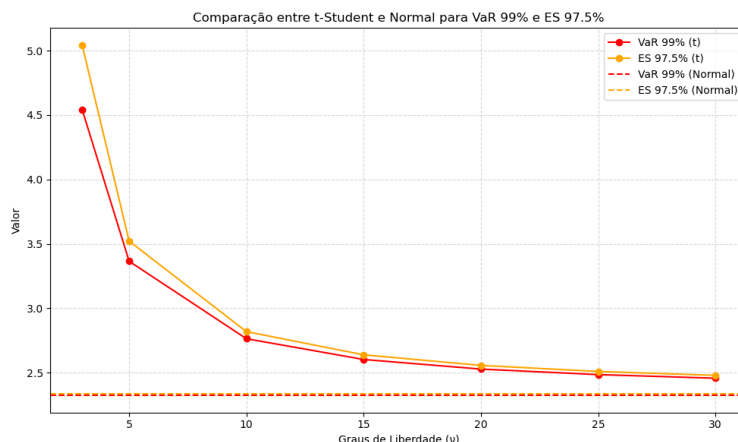


FIGURA 3.1. VaR 99% vs ES 97,5% — Comparação entre a distribuição t-Student e a Normal padrão para o VaR a 99% e o ES a 97,5%, em função dos graus de liberdade (ν).

A distribuição t-student apresenta um comportamento semelhante ao da normal, mas distingue-se por ter caudas mais pesadas, especialmente para valores baixos dos graus de liberdade. No gráfico acima, podemos observar que, tal como seria de esperar, quando menores forem os graus de liberdade, maior é a diferença entre os valores de VaR e ES da t-Student e os valores da normal padrão.

Apesar das vantagens referidas em relação ao ES, autores como Roccioletti (2016) e Osmundsen (2016) referem que o ES não é uma medida eliciável, ao contrário do VaR. No entanto, embora alguns autores, como Acerbi e Szekely (2016) argumentem que esta propriedade não compromete o backtesting do ES, outros autores, como Roccioletti (2016) afirmam que a ausência de elicitabilidade pode introduzir dificuldades adicionais na construção de procedimentos estatísticos robustos para a sua validação empírica.

3.5. Backtesting

Após a estimação das medidas de risco, é fundamental avaliar a sua capacidade preditiva e a adequação dos modelos utilizados. O *backtesting* consiste precisamente na comparação entre as previsões de risco e as perdas efetivamente observadas, permitindo validar a performance dos modelos de previsão e identificar eventuais deficiências.

Nesta secção, são apresentados os procedimentos de backtesting aplicados às previsões de VaR e ES obtidas com os diferentes modelos estimados.

3.5.1. Backtesting VaR

Segundo Christoffersen (1998), podemos considerar o *backtesting* baseado no *violation process*. Este processo diz respeito às falhas associadas ao intervalo de confiança, isto é, aos momentos em que o valor observado das perdas excede o limite previsto pelo VaR. Assim, podemos definir a função de violação da seguinte forma:

$$I_{\alpha,t} = \mathbb{1}(r_t < -VaR_{\alpha,t}), \quad (3.6)$$

onde r_t representa o retorno no instante t .

Ainda segundo Christoffersen (1998), se o modelo de VaR estiver bem especificado, então $I_{\alpha,t}$ deve seguir uma distribuição Bernoulli i.i.d. com probabilidade de sucesso α . Isto significa que o número total de violações de uma amostra com T dias segue uma distribuição $Binomial(T, \alpha)$. Com base neste pressuposto, é possível validar a exatidão de um modelo de VaR, comparando o número de exceções ao longo de um determinado período de tempo com a distribuição binomial esperada, avaliando assim a probabilidade de ocorrência do número de violações registadas ao longo do período de análise.

A prática descrita acima está alinhada com o *Fundamental Review of the Trading Book* (FRTB), desenvolvido pelo Comité de Basileia. O FRTB exige que as instituições realizem backtesting diário do VaR a 99% de confiança, com um horizonte de 10 dias, utilizando pelo menos 250 observações diárias consecutivas (aproximadamente 1 ano de dados). Dependendo do número de exceções observadas nesse período, o modelo recebe uma classificação que ajusta o multiplicador de capital regulatório aplicado ao requisito mínimo de capital. A Tabela 3.1 apresenta a classificação do FRTB por zonas, de acordo com o número de exceções observadas.

TABELA 3.1. Limites das zonas de backtesting do FRTB

Zona	# Exceções	Multiplicador	Probabilidade acumulada
Verde	0	1,50	8,11%
	1	1,50	28,58%
	2	1,50	54,32%
	3	1,50	75,81%
	4	1,50	89,22%
Amarela	5	1,70	95,88%
	6	1,76	98,63%
	7	1,83	99,60%
	8	1,88	99,89%
	9	1,92	99,97%
Vermelha	10 ou mais	2,00	99,99%

Existem também testes estatísticos para testar o *backtesting* do VaR. Estes testes baseiam-se no processo de violações definido anteriormente, analisando se a frequência das exceções observadas é consistente com as propriedades esperadas sob um modelo corretamente especificado.

Um dos testes mais utilizados é o teste de cobertura incondicional de Kupiec (1995). Este teste verifica se a proporção de exceções observadas ao longo do período de análise é estatisticamente compatível com a taxa de exceções esperada, dada pelo nível de confiança do VaR. Por outro lado, o teste de independência proposto por Christoffersen (1998) avalia se as exceções ocorreram de forma independente ao longo do tempo. Tal é relevante, uma vez que a presença de agrupamentos de exceções pode indicar que o modelo subestima o risco em certos períodos de maior volatilidade.

O teste da cobertura condicional combina os dois testes anteriores, permitindo avaliar simultaneamente a frequência de exceções e a sua independência. Para este estudo, iremos apenas utilizar o teste de cobertura incondicional de Kupiec (1995).

A estatística de teste para Kupiec (1995) segue da seguinte forma, para uma amostra de T observações:

$$POF = 2\ln \left(\left(\frac{1 - \hat{\alpha}}{1 - \alpha} \right)^{T - I(\alpha)} \left(\frac{\hat{\alpha}}{\alpha} \right)^{I(\alpha)} \right) \sim \chi_1^2 \quad (3.7)$$

Com $\hat{\alpha} = \frac{1}{T}I(\alpha)$, onde $I(\alpha) = \sum_{t=1}^T I_{\alpha,t}(\alpha)$ representa o número total de violações observadas, pode-se afirmar que, se a proporção de exceções de VaR, $\hat{\alpha} \times 100\%$, for exatamente igual a $\alpha \times 100\%$, então a estatística do teste POF assume o valor zero, indicando ausência de evidência contra a adequação do modelo de VaR. À medida que a proporção de violações se afasta de $\alpha \times 100\%$, a estatística do teste POF aumenta, refletindo evidência crescente de que a medida de VaR proposta subestima ou sobrestima sistematicamente o verdadeiro nível de risco da carteira.

Hipótese 1. Teste de cobertura incondicional de Kupiec:

$$H_0: \pi_{\text{obs}} = \alpha$$

$$H_1: \pi_{\text{obs}} \neq \alpha$$

Este teste será incluído nas análises mais adiante, em comparação com as metodologias existentes para o Expected shortfall.

3.5.2. Backtesting Expected Shortfall

A adoção do *Expected Shortfall* (ES) como medida regulamentar de risco pelo Comité de Basileia trouxe consigo vários desafios no que diz respeito à sua validação empírica, ou *backtesting*. Como vimos anteriormente, ao contrário do *Value-at-Risk* (VaR), o ES não é uma medida *eliciável*, o que significa que não existe uma função de *score* única que permita

comparar previsões de forma objetiva. Esta limitação suscitou um debate acadêmico em torno da questão de saber se a elicitabilidade é, ou não, uma condição necessária para o *backtesting* de medidas de risco.

Embora inicialmente a falta de elicitabilidade tenha sido vista como um obstáculo, vários autores como Acerbi e Szekely (2014); Emmer et al. (2015); Kratz et al. (2018) argumentam que o ES pode ser validado de forma indireta, por exemplo, através da aplicação de testes a múltiplos quantis do VaR, ou ainda mediante uma análise adequada da modelação da cauda da distribuição dos retornos.

Nas secções seguintes, serão descritas algumas das principais metodologias atualmente propostas na literatura para a realização do *backtesting* do ES, com particular foque naquelas que serão aplicadas no contexto deste trabalho.

Testes Multinomiais

Estes testes baseiam-se na aproximação proposta por Emmer et al. (2015), segundo a qual o Expected Shortfall (ES) pode ser aproximado por uma soma de Riemann da forma:

$$ES_{1-\alpha}(P_t) \approx \frac{1}{4} (\text{VaR}_{t,\alpha} + \text{VaR}_{t,0.75\alpha+0.25} + \text{VaR}_{t,0.5\alpha+0.5} + \text{VaR}_{t,0.25\alpha+0.75}) \quad (3.8)$$

Kratz et al. (2018) sugerem a utilização de um teste multinomial baseado em múltiplos níveis de VaR. Para formalizar esta abordagem, consideram-se N níveis de VaR, cujos respetivos níveis de confiança serão:

$$\alpha_j = \alpha + \frac{j-1}{N}(1-\alpha), \quad j = 1, \dots, N \quad (3.9)$$

Assume-se ainda que $\alpha_0 = 0$ e $\alpha_{N+1} = 1$, sendo α um determinado nível de referência, como por exemplo $\alpha = 0,975$.

Para um horizonte temporal de backtesting $T \in \mathbb{N}$, utiliza-se a função de violação $I_{\alpha,t}$, definida na equação 3.6, o que conduz à realização de N testes binomiais. Kratz et al. (2018) definem então X_t como

$$X_t = \sum_{j=1}^N I_{\alpha_j,t}, \quad t = 1, \dots, T \quad (3.10)$$

Assim, X_t conta o número de exceções nos diferentes níveis de VaR no instante t . É ainda utilizado um ajustamento que permite contabilizar simultaneamente todas as observações t no horizonte de *backtesting*. Para tal, é definida a seguinte função de contagem:

$$O_j = \sum_{t=1}^T \mathbf{1}_{(X_t = j)}, \quad j = 0, \dots, N \quad (3.11)$$

Note-se que qualquer violação do VaR_{t,α_j} implica necessariamente uma violação do $VaR_{t,\alpha_{j-1}}$, uma vez que, por definição, se tem $VaR_{t,\alpha_j} \geq VaR_{t,\alpha_{j-1}}$. Assim, O_j contabiliza o número de observações em que a perda ultrapassa o valor do VaR ao nível de confiança α_j , mas não excede o valor do VaR para o nível de confiança α_{j+1} .

Por definição, o vetor $\vec{O}$ deverá seguir a mesma distribuição que X_t , mas agregada ao longo dos T instantes do horizonte de *backtesting*. Assim, tem-se:

$$\vec{O} = \{O_0, \dots, O_N\} \sim MN(T, (\alpha_1 - \alpha_0, \dots, \alpha_{N+1} - \alpha_N)) \quad (3.12)$$

Partindo da suposição mais geral de que $\{O_0, \dots, O_N\} \sim MN(T, (\theta_1 - \theta_0, \dots, \theta_{N+1} - \theta_N))$ para parâmetros arbitrários $0 = \theta_0 < \theta_1, \dots, < \theta_N < \theta_{N+1} = 1$, as hipóteses do teste são dadas por:

Hipótese 2. Teste Multinomial:

$$H_0: \theta_j = \alpha_j$$

$$H_1: \theta_j \neq \alpha_j$$

A hipótese nula H_0 assume que as probabilidades teóricas de ocorrência das exceções, isto é, os incrementos entre os níveis de confiança sucessivos $(\alpha_{j+1} - \alpha_j)$ estão corretas. Esta hipótese implica que o modelo de risco utilizado gera previsões de VaR que estão em conformidade com a distribuição empírica das perdas observadas.

São propostos três testes estatísticos por Kratz et al. (2018): o Teste de Pearson, o teste de Nass e o LRT (*likelihood ratio test*). No entanto, dado que este último teste envolve maior complexidade técnica e é menos utilizado na prática, a sua implementação não será incluída no âmbito deste trabalho. Passaremos a definir as estatísticas de teste de Pearson e Nass.

a) Teste de Pearson (Pearson, 1900):

$$S_N = \sum_{j=0}^N \frac{(O_j - T(\alpha_{j+1} - \alpha_j))^2}{T(\alpha_{j+1} - \alpha_j)} \sim_{H_0} \chi_N^2 \quad (3.13)$$

Esta estatística mede o grau de divergência entre as exceções observadas e as frequências esperadas sob o modelo, assumindo que este está corretamente especificado. Se o valor da estatística S_N for suficientemente elevado, rejeita-se a hipótese nula H_0 , indicando que o modelo pode não estar a representar adequadamente a distribuição dos retornos.

b) Teste de Nass (Nass, 1959):

$$cS_N \sim_{H_0} \chi_2^\nu \quad (3.14)$$

com

$$\begin{aligned}
E[S_N] &= N \\
Var[S_N] &= 2N - \frac{N^2 + 4N + 1}{T} + \frac{1}{T} \sum_{j=0}^N \frac{1}{\alpha_{j+1} - \alpha_j} \\
c &= \frac{2E[S_N]}{Var[S_N]} \\
\nu &= c \cdot E[S_N]
\end{aligned}$$

O teste de Pearson pode revelar desvios quando a amostra é pequena. O teste de Nass ajusta a média e a variância de S_N , garantindo uma aproximação mais fiável à qui-quadrado. Esta correção melhora o desempenho do teste quando o número de observações é mais limitado. Kratz et al. (2018) mencionam que o teste de Nass com $N = 8$ deve ser preferível ao teste de Pearson com $N = 4$

Testes Z de Acerbi e Szekely

No âmbito do estudo sobre a elicitabilidade do *Expected Shortfall* (ES), Acerbi e Szekely (2014) propõem três métodos distintos para a avaliação da sua precisão e validade como medida de risco.

Para estes testes, assume-se que, para cada dia $t = 1, \dots, T$, X_t representa o retorno de uma carteira, que segue uma distribuição verdadeira, mas desconhecida, denotada por F_t . Esta distribuição é prevista com base num modelo de risco que assume uma distribuição P_t .

O primeiro teste baseia-se no valor esperado condicional

$$\mathbb{E} \left[\frac{X_t}{ES_{\alpha,t}} | X_t + VaR_{\alpha,t} < 0 \right] = 0 \quad (3.15)$$

a partir do qual, os autores determinam a estatística de teste

$$Z_1 = \frac{\sum_{t=1}^T \frac{X_t I_{\alpha,t}}{ES_{\alpha,t}}}{N_t}, \quad \text{com } N_t = \sum_{t=1}^T I_{\alpha,t} > 0 \quad (3.16)$$

Definem-se, então, as seguintes hipóteses para este teste

Hipótese 3. Teste 1 de Acerbi e Szekely (Z1):

$$H_0: P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t$$

$$H_1: ES_{\alpha,t}^F \geq ES_{\alpha,t} \quad \forall t \text{ e } > \text{ para algum } t$$

$$VaR_{\alpha,t}^F = VaR_{\alpha,t}, \quad \forall t$$

onde $P_t^{[\alpha]}(x) = \min(1, P_t(x)/\alpha)$ é a cauda da distribuição para $x < VaR_{\alpha,t}$.

Verifica-se que o VaR_α previsto continua correto sob H_1 , o que está em linha com a ideia de que este teste está condicionado à validação prévia do VaR. O segundo teste segue do seguinte valor esperado

$$ES_{\alpha,t} = -\mathbb{E} \left[\frac{X_t I_t}{\alpha} \right] \quad (3.17)$$

o que serve para definir a estatística de teste Z_2

$$Z_2 = \sum_{t=1}^T \frac{X_t I_{\alpha,t}}{T \alpha ES_{\alpha,t}} + 1 \quad (3.18)$$

As hipóteses para este teste são, então,

Hipótese 4. Teste 2 de Acerbi e Szekely (Z2):

$$H_0: P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t$$

$$H_1: ES_{\alpha,t}^F \geq ES_{\alpha,t} \quad \forall t \text{ e } > \text{ para algum } t$$

$$VaR_{\alpha,t}^F = VaR_{\alpha,t}, \quad \forall t$$

Apesar destes dois testes partilharem a mesma estrutura de hipóteses, diferem significativamente na forma como operacionalizam essa avaliação. O primeiro teste baseia-se num valor esperado condicional, aplicado exclusivamente às observações que se encontram na cauda esquerda da distribuição, ou sejam aquelas cuja perda excede o VaR. Esta característica torna o teste especialmente sensível à forma da cauda e à dispersão das perdas extremas, sendo eficaz na deteção de eventuais falhas, nomeadamente se este não captar corretamente a severidade dos eventos mais raros.

Já o segundo teste adota uma abordagem mais agregada, ao considerar todas as observações ponderadas pela ocorrência ou não de exceções ao VaR. A estatística de teste compara a média ponderada das perdas observadas com o valor previsto do ES, possibilitando a identificação de desvios sistemáticos ao longo do tempo. Por este motivo, o teste Z2 é particularmente útil em contextos onde o objetivo é avaliar a calibração global do modelo de risco, sendo mais robusto face à variabilidade pontual das observações mais extremas.

Em termos práticos, a aplicação destes testes enfrenta algumas limitações importantes. Uma das principais dificuldades prende-se com a escassez de observações na cauda da distribuição, sobretudo para níveis de confiança elevados (ex: $\alpha = 0,975$), o que compromete a fiabilidade estatística do teste Z1. Adicionalmente, ambos os testes assumem a validade do modelo de VaR, o que pode levantar questões metodológicas: se o VaR estiver mal calibrado, as conclusões sobre o ES podem ser enviesadas, já que os testes partem do princípio de que

a fronteira da cauda está corretamente posicionada. Esta dependência pode ser interpretada como uma limitação, dado que, na prática, os modelos de VaR e ES são muitas vezes calibrados de forma conjunta.

Relativamente ao terceiro teste proposto por Acerbi e Szekely (2014), Z_3 , optou-se por não o implementar nesta dissertação. Este teste consiste numa combinação linear das estatísticas Z_1 e Z_2 , com pesos adaptativos que visam otimizar a sua potência em diferentes regimes de mercado. No entanto, essa abordagem exige a estimação de parâmetros adicionais para calibrar a combinação ótima, o que introduz complexidade computacional e riscos de sobreajuste.

Teste ZR de Acerbi e Szekely

Após a introdução dos testes Z por Acerbi e Szekely (2014), os autores apresentam, em Acerbi e Szekely (2017), uma variação que será designada por ZR, a qual assume que o ES permite a aplicação de um *ridge backtest*. Nesta abordagem, a estatística pode ser formulada como o valor mínimo do valor esperado de uma função *score*, aplicada a uma estatística auxiliar. A principal vantagem deste método reside penalização contínua e suave dos desvios, em contraste com métodos que se baseiam apenas na contagem binária de exceções.

Para a formulação do teste, os autores recorrem aos resultados de Rockafellar e Uryasev (2002), segundo os quais o VaR e o ES podem ser obtidos a partir da minimização do valor esperado de uma função convexa. Em particular:

$$VaR_{\alpha,t} = \arg \min_y E \left[y + \frac{1}{\alpha} (X_t + y)^- \right] \quad (3.19)$$

$$ES_{\alpha,t} = \min_y E \left[y + \frac{1}{\alpha} (X_t + y)^- \right] \quad (3.20)$$

Partindo das expressões anteriores, e dado que a notação utilizada por Acerbi e Szekely (2017) é bastante geral, opta-se por adotar a notação utilizada por Catarino (2023) para expressar a estatística de teste $\bar{Z}_{ES}$

$$\bar{Z}_{ES} = \frac{1}{T} \sum_{t=1}^T [ES_{\alpha,t} - \hat{ES}_{\alpha}] \quad (3.21)$$

com

$$\hat{ES}_{\alpha} = VaR_{\alpha,t} + \frac{1}{\alpha} (r_t + VaR_{\alpha,t})^- \quad (3.22)$$

assim, $\bar{Z}_{ES}$ representa a média das estatísticas individuais $Z_{ES,t} = ES_{\alpha,t} - VaR_{\alpha,t} - \frac{1}{\alpha} (r_t + VaR_{\alpha,t})^-$

Em termos práticos, o teste ZR compara o ES previsto pelo modelo com um estimador empírico, obtido a partir das perdas observadas e do VaR correspondente. A estatística

é calculada para cada observação e agregada ao longo do período de backtesting, sendo o seu valor comparado com uma distribuição empírica obtida via simulação Monte Carlo para determinação do valor-p.

Uma vez que os autores não definem as hipóteses para este teste, será novamente seguida a abordagem proposta por Catarino (2023), onde se escrevem as hipóteses para este teste de forma semelhante à Hipótese 4:

Hipótese 5. Teste ZR de Acerbi e Szekely (ZR):

$$H_0: P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t$$

$$H_1: ES_{\alpha,t}^F \geq ES_{\alpha,t} \quad \forall t \text{ e } > \text{ para algum } t$$

$$VaR_{\alpha,t}^F = VaR_{\alpha,t}, \quad \forall t$$

Seguirei igualmente a abordagem proposta por Catarino (2023) para a implementação deste teste, no cálculo do valor-p. Como a estatística de teste não converge para nenhuma distribuição conhecida, é necessário simular sob a hipótese nula que a previsão da distribuição dos retornos está correta.

Assim, após o cálculo da estatística de teste observada, Z_{obs} , procede-se do seguinte modo:

- (1) Simular de um vetor $\vec{X}_i$ de T valores independentes, gerados segundo $X_t \sim P_t$, para $t = 1, \dots, T$;
- (2) Calcular o valor da estatística de teste simulada Z_{sim}^i , a partir de $\vec{X}_i$;
- (3) Repetir este processo M vezes, de forma a obter a distribuição empírica de Z ;
- (4) Estimar o valor-p da estatística de teste observada, Z_{obs} :

$$p = \frac{1}{M} \sum_{i=1}^M \mathbb{1}_{Z_{sim}^i \geq Z_{obs}} \quad (3.23)$$

Assim, o modelo é rejeitado com 95% de confiança (Zona amarela) quando $p \geq 0,95$ e com 99,99% de confiança (zona vermelha) quando $p \geq 0,9999$.

Caso o teste seja aplicado a um modelo histórico, em vez de um modelo paramétrico, a simulação de $\vec{X}_i$ a partir da distribuição assumida, é substituída por uma amostra aleatória retirada diretamente dos retornos passados, a partir da qual calculamos Z_{sim}^i .

Nos resultados apresentados neste trabalho, são considerados $T = 250$ e $M = 20,000$

Teste de Moldenhauer e Pitera (M&P)

O teste de Moldenhauer & Pitera, também conhecido como (*secured position test*) foi proposto como uma alternativa ao tradicional backtesting de modelos de risco. O seu objetivo é verificar diretamente se a estimativa do ES é adequada, sem recorrer a testes de hipóteses assentes em distribuições, como acontece noutros procedimentos vistos anteriormente.

Para tal, dado um retorno observado r_t e o ES estimado ES_t ao nível de confiança α , os autores definem a chamada *posição segura* como S_t

$$S_t = r_t + ES_t. \quad (3.24)$$

Intuitivamente, S_t representa a posição do portfólio após acrescentar o capital determinado pelo ES. Se o modelo for adequado, a posição segura deverá manter-se maioritariamente positiva.

A estatística de teste proposta pelos autores baseia-se no somatório acumulado das piores realizações da posição segura. Para tal, ordenam-se as observações $\{S_t\}_{t=1}^T$ de forma crescente, denotadas por $S_{(1)} \leq S_{(2)} \leq \dots \leq S_{(T)}$, e considera-se

$$G = \sum_{t=1}^T \mathbb{1} \left\{ \sum_{k=1}^t S_{(k)} < 0 \right\} \quad (3.25)$$

Assim, G mede quantas das somas cumulativas dos piores resultados do portfólio já coberto pelo capital ES_t permanecem negativas. Quanto maior for G , maior a evidência de que o ES não fornece uma cobertura adequada para os cenários adversos.

À semelhança de outros testes de backtesting, pode-se introduzir uma classificação em zonas, em linha com o enquadramento regulatório do FRTB. Típicamente adota-se:

TABELA 3.2. Classificação por zonas do teste M&P

Backtesting Zone	G (M&P test)
Green	≤ 11
Amber	$12 - 24$

Comparativamente ao procedimento tradicional de Basileia, baseado no número de excessos do VaR, este teste apresenta duas diferenças fundamentais. Em primeiro lugar, avalia diretamente o ES. Em segundo lugar, ao contrário do teste de Basileia que apenas contabiliza a frequência de violações ao VaR, o teste de Moldenhauer & Pitera incorpora também a severidade dos piores resultados através da estatística G . Assim, fornece uma avaliação mais robusta e prudencial da adequação do capital baseado em ES.

CAPÍTULO 4

Metodologia

Este capítulo descreve os procedimentos metodológicos adotados no âmbito da presente dissertação. Serão apresentados os critérios seguidos para a construção da carteira, o processo de preparação dos dados, bem como os modelos utilizados para a estimação das medidas de risco *Value at Risk* (VaR) e *Expected Shortfall* (ES).

4.1. Descrição da amostra

De forma a construir uma carteira representativa e diversificada do ponto de vista geográfico e setorial, selecionaram-se oito ações de referência de empresas cotadas nos principais mercados internacionais: a Apple e a Microsoft (Estados Unidos), a Siemens e a Bayer (Zona Euro), a Tencent e a Ping An (Asia) e a Unilever e a HSBC (Reino Unido). A escolha destas ações foi motivada pela sua elevada capitalização bolsista, liquidez e papel relevante nos respetivos mercados, permitindo captar as dinâmicas de economias desenvolvidas de diferentes regiões geográficas.

Ao contrário dos índices acionistas, estas ações incorporam risco idiossincrático, ou seja, risco específico de cada empresa, o que introduz uma componente adicional de variabilidade nos retornos da carteira. Esta opção permite uma análise mais realista do comportamento de carteiras compostas por ativos individuais, como ocorre frequentemente na prática por parte de investidores institucionais e particulares.

As séries de preços utilizadas correspondem a preços de fecho ajustados, obtidos através da plataforma *Bloomberg*. Esta fonte garante a fiabilidade e a consistência dos dados, os quais foram previamente ajustados para eventos corporativos como dividendos e desdobramentos de ações.

A amostra abrange o período de 2 de janeiro de 2008 a 30 de maio de 2025, totalizando 4 497 observações diárias. A escolha deste horizonte temporal teve como objetivo capturar diferentes regimes de volatilidade nos mercados financeiros, incluindo períodos de estabilidade e episódios de crise, como, por exemplo, a crise de 2008 e os impactos da pandemia de COVID-19. Esta diversidade de condições de mercado é essencial para uma avaliação robusta das medidas de risco.

A carteira foi construída de acordo com a Tabela 4.1, com alocações definidas de forma a refletir a relevância económica e financeira dos mercados representados. Atribuiu-se um

maior peso às ações de empresas com maior capitalização e liquidez, assegurando uma composição que respeita os princípios de diversificação e equilíbrio entre diferentes geografias. Dado que os ativos estão denominados em diferentes moedas, nomeadamente USD, EUR, HKD e GBP, a exposição ao risco cambial foi igualmente considerada no processo de construção da carteira.

TABELA 4.1. Exposição da carteira a fatores de risco

Fator de Risco	Exposição(%)
Apple	0,18
Microsoft	0,18
Siemens	0,10
Bayer	0,10
Tencent	0,08
Ping An	0,08
Unilever	0,14
HSBC	0,14
USD/EUR	0,36
HKD/EUR	0,24
GBP/EUR	0,20

4.2. Metodologia

Após a obtenção do PnL da carteira, procedeu-se ao cálculo do VaR 99% e do ES 97,5%. Nesta secção, apresentam-se os modelos utilizados para a estimação destas medidas. O principal objetivo da utilização de abordagens distintas consiste em comparar a sua capacidade de capturar o risco de cauda associado à carteira em análise.

Foram implementados três modelos distintos. O primeiro é um modelo simples, paramétrico, designado por *Normal*, que assume a normalidade dos retornos e volatilidade constante, servindo como referência base. O segundo, designado por *GARCH-Normal*, é também um modelo paramétrico, mas que incorpora a alteração da volatilidade através de um modelo GARCH. Por fim, foi considerado o modelo semi-paramétrico de Hull & White, que combina a abordagem de ajustamento dos retornos históricos com a estimação da volatilidade através do método EWMA.

Os modelos mencionados abaixo, bem como os já referidos testes de *backtesting* foram implementados em *Python*.

4.2.1. Modelos ARCH e GARCH

O modelo ARCH (*Autoregressive Conditionally Heteroskedastic*) foi introduzido por Engle em 1982 com o objetivo de modelar a heterocedasticidade condicional em séries temporais financeiras e económicas. Este modelo acabou por se tornar amplamente utilizado para diversas variáveis económicas e financeiras.

Um processo ARCH é composto por duas equações: a equação da média condicional e a equação da variância condicional. No caso geral de um modelo ARCH(q), temos:

$$y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{tk} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_t^2) \quad (4.1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 \quad (4.2)$$

onde $\alpha_0 > 0$ e $\alpha_i \geq 0$ para $i = 1, \dots, q$, garantindo a positividade da variância condicional. Além disso, para assegurar a estacionaridade do processo, impõe-se que a soma dos coeficientes α_i seja inferior a 1.

Uma extensão natural do modelo ARCH é o modelo GARCH (Generalized Autoregressive Conditional Heteroskedasticity), proposto por Bollerslev (1986), que permite que a variância condicional dependa não apenas de choques passados, mas também dos seus próprios valores defasados. O caso mais comum é o modelo GARCH(1,1), no qual a variância condicional é influenciada tanto pelo erro do período anterior como pela variância condicional do período anterior:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \cdots + \beta_k y_{t-k} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_t^2) \quad (4.3)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (4.4)$$

Apesar de ser possível generalizar este modelo para ordens superiores (GARCH(p,q)), o modelo GARCH(1,1) é amplamente utilizado na prática e revela-se, em muitos casos, suficiente para capturar fenómenos como o *volatility clustering*, frequentemente observados nos retornos financeiros. Por essa razão, será utilizado no âmbito deste estudo.

4.2.2. Normal e GARCH-Normal

O primeiro modelo, denominado por *Normal*, estima o *Expected Shortfall* (ES) assumindo uma distribuição normal dos retornos. Para tal, foi implementada uma metodologia baseada em janelas móveis (*rolling window*) de 250 observações diárias, equivalentes a aproximadamente um ano de dados para séries temporais financeiras.

Em cada janela, calcularam-se a média (μ_t) e o desvio padrão (σ_t) dos retornos diários percentuais da carteira. Estas estatísticas serviram de base para determinar o VaR 99% e o ES 97,5%.

O VaR é então estimado pela seguinte expressão:

$$VaR_t^{\text{Norm}} = \mu_t + \sigma_t \cdot z_\alpha \quad (4.5)$$

onde z_α representa o quantil da distribuição normal padrão correspondente ao nível de confiança α .

O ES é calculado como:

$$ES_t^{\text{Norm}} = \mu_t - \sigma_t \cdot \frac{\phi(z_{\alpha_{ES}})}{1 - \alpha_{ES}} \quad (4.6)$$

onde $\phi(\cdot)$ representa a função densidade da distribuição normal padrão.

Já o segundo modelo utilizado, designado por *GARCH-Normal*, recorre à modelação da volatilidade condicional através de um processo GARCH(1,1). Para este estudo, será assumido que a média dos retornos é zero, para este modelo.

O modelo parte do pressuposto de que os retornos r_t podem ser descritos como:

$$r_t = \sigma_t \epsilon_t, \quad (4.7)$$

onde σ_t representa a volatilidade condicional e $\epsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

Para capturar a heterocedasticidade dos retornos, ajustou-se um modelo GARCH(1,1) com média condicional nula em janelas móveis de 250 observações. Assim, para cada janela, estima-se a volatilidade condicional σ_t no último ponto da janela, considerando a equação de volatilidade condicional previamente apresentada (Equação 4.4).

Dado que se assume $\mu = 0$, as expressões para o Value-at-Risk (VaR) a 99% e o Expected Shortfall (ES) a 97,5% tornam-se:

$$VaR_t = \sigma_t \cdot z_\alpha, \quad (4.8)$$

$$ES_t = -\sigma_t \cdot \frac{\phi(z_\alpha)}{\alpha}, \quad (4.9)$$

onde z_α corresponde ao quantil da normal padrão associado à probabilidade de excedência do VaR, e $\phi(\cdot)$ é a função densidade da normal padrão. A equação de volatilidade condicional permite modelar a evolução dinâmica do risco ao longo do tempo, ajustando as estimativas de VaR e ES às condições observadas em cada janela.

4.2.3. H&W

O terceiro modelo implementado, designado por H&W (Hull & White), segue a metodologia proposta por Hull e White (1998), na qual os retornos históricos são ajustados pela razão entre a volatilidade prevista e a volatilidade observada em cada instante.

A estimação da volatilidade será feita através do esquema EWMA (*Exponentially Weighted Moving Averages*). Assim, para cada observação t , a respetiva volatilidade é dada por,

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2 \quad (4.10)$$

onde λ é o *smoothing factor*, responsável por definir o peso relativo atribuído aos retornos passados. Seguindo a prática comum na literatura, adotou-se o valor $\lambda = 0,94$.

Após a estimação da volatilidade condicional σ_t para cada instante temporal, construiu-se para cada t uma amostra de retornos históricos ajustados. Para isso, consideraram-se os 250 retornos precedentes, ajustados pela razão entre a volatilidade prevista para t e a volatilidade condicional estimada em cada ponto da janela, ou seja,

$$\tilde{r}_s^{(t)} = r_s \cdot \frac{\hat{\sigma}_t}{\sigma_s}, \quad \text{para } s = t - 250, \dots, t - 1, \quad (4.11)$$

Este procedimento, proposto por Hull & White (1998), tem como objetivo capturar o impacto da evolução da volatilidade sobre a distribuição condicional dos retornos, ajustando os cenários históricos de forma a refletirem o nível de risco atual.

O *Value-at-Risk* (VaR) para cada instante t foi então obtido como o quantil empírico α da amostra ajustada, i.e.,

$$\text{VaR}_{t,\alpha} = q_\alpha \left(\{\tilde{r}_s^{(t)}\}_{s=t-250}^{t-1} \right) \quad (4.12)$$

e o *Expected Shortfall* (ES) para cada instante t foi calculado como a média das observações da amostra ajustada que representam perdas iguais ou mais severas do que o VaR, isto é, todos os $\tilde{r}_{H\&W}^{(t)}$ inferiores ou iguais ao valor do VaR empírico. Formalmente,

$$\text{ES}_{t,\alpha} = \frac{1}{N_{\text{tail}}} \sum_{s=t-250}^{t-1} \tilde{r}_s^{(t)} \cdot \mathbb{1} \{ \tilde{r}_s^{(t)} \leq \text{VaR}_{t,\alpha} \} \quad (4.13)$$

onde N_{tail} representa o número de observações que se encontram na cauda esquerda da distribuição (i.e., aquelas com $\tilde{r}_s^{(t)} \leq \text{VaR}_{t,\alpha}$). Este procedimento assegura que o ES reflete a média das piores perdas estimadas condicionalmente à volatilidade prevista para cada momento.

Esta escolha metodológica tende a resultar numa abordagem conservadora, dado que o ajustamento pela razão $\frac{\sigma_t}{\sigma_s}$ pode levar a uma ligeira sobrestimação do risco, sobretudo em períodos em que a volatilidade recente σ_t é superior à média da janela histórica.

CAPÍTULO 5

Dados e Análise

Neste capítulo apresentam-se as características dos dados utilizados e algumas estatísticas descritivas relevantes para o estudo. As análises são acompanhadas de tabelas e representações gráficas, de forma a facilitar a interpretação dos resultados.

5.1. Resultados dos Testes de Backtesting

O VaR 99% e o ES 97,5% foram estimados de acordo com os modelos descritos no capítulo anterior. Após a sua implementação, aplicaram-se diariamente os testes estatísticos apresentados, utilizando uma janela móvel de 250 dias, em conformidade com o procedimento adotado pelo FRTB.

Para o backtesting diário do VaR, implementou-se o teste binomial de contagem de exceções, denominado mais adiante por "Basel", bem como o teste de Kupiec. Já para o ES, foram aplicados os testes de Pearson e de Nass, com $N = 2, 4$ e 8 e ainda o M&P.

Os testes Z de Acerbi e Szekely foram excluídos desta análise inicial, devido ao elevado esforço computacional associado à utilização de simulações de Monte Carlo para o cálculo das respetivas estatísticas de teste, o que tornaria a execução diária menos eficiente.

A partir deste ponto, sempre que se justifique, serão apresentados exclusivamente os valores-p resultantes dos testes estatísticos, com base nos quais, para cada dia, cada um dos modelos foi classificado em zonas segundo o critério estabelecido pelo FRTB. A classificação dos modelos obedeceu à seguinte regra:

- (1) **Zona Verde (V)** — quando o valor-p é superior a 0,05;
- (2) **Zona Amarela (A)** — quando o valor-p se encontra entre 0,01 e 0,05;
- (3) **Zona Vermelha (R)** — quando o valor-p é inferior a 0,01.

A tabela 5.1 apresenta os resultados dos testes efetuados aos resultados do *backtesting* realizados com janelas móveis de 250 observações, evidenciando o desempenho dos diferentes modelos na estimação do VaR e do ES. Os resultados são apresentados em termos percentuais, refletindo a proporção de janelas classificadas em cada uma das zonas acima definidas.

Desde logo, é possível observar que os três modelos apresentam desempenhos bastante distintos. O modelo Norm evidencia fragilidades significativas, sendo que a proporção de janelas na zona verde é reduzida, enquanto que as zonas amarela e vermelha surgem com uma frequência superior ao desejável. Este resultado seria expectável dado o carácter simplista do

TABELA 5.1. Resultados dos testes em janela deslizando

Teste		Norm			GARCH-Norm			H&W		
		V	A	R	V	A	R	V	A	R
VaR 99%	Basel	50,14	36,79	13,07	41,57	58,10	0,33	74,59	25,41	0,00
	Kupiec	46,49	9,99	43,52	62,81	9,70	27,48	81,75	9,07	9,19
ES 97,5%	Pearson N=2	92,25	6,15	1,60	93,38	5,79	0,82	100,00	0,00	0,00
	Pearson N=4	88,81	9,23	1,95	86,20	7,75	6,05	100,00	0,00	0,00
	Pearson N=8	65,47	28,05	6,48	76,40	14,88	8,71	100,00	0,00	0,00
	Nass N=2	72,82	8,36	18,82	85,07	4,90	10,03	100,00	0,00	0,00
	Nass N=4	69,34	11,28	19,38	72,16	11,71	16,13	100,00	0,00	0,00
	Nass N=8	63,09	18,68	18,23	74,47	14,30	11,23	100,00	0,00	0,00
	M&P	58,24	31,94	9,82	53,37	45,93	0,71	100,00	0,00	0,00

modelo, que assume a normalidade dos retornos e uma volatilidade constante, não captando a variabilidade real observada nos mercados financeiros.

O modelo GARCH-Norm apresenta uma melhoria substancial, ao incorporar a heterocedasticidade da volatilidade. Globalmente, verifica-se um aumento das classificações na zona verde e uma redução de resultados nas restantes zonas. Ainda assim, importa notar que em determinados testes este modelo não supera de forma consistente o desempenho do Norm, revelando algumas limitações relacionadas com a assunção da normalidade dos resíduos.

Por fim, o modelo H&W alcança resultados quase perfeitos em todos os testes, colocando a totalidade das janelas na zona verde. Embora tal desempenho sugira uma robustez elevada na estimação do risco, também pode indicar uma sobreavaliação do risco. Em termos práticos, isto traduz-se numa postura excessivamente conservadora, que garante cobertura total do risco extremo, mas à custa de potenciais requisitos de capital superiores ao necessário.

No caso do VaR, a diferença entre modelos torna-se evidente. O modelo Norm obtém apenas cerca de metade das janelas na zona verde no teste Basel, com uma elevada proporção dos resultados na zona amarela. Já quando aplicado o teste de Kupiec, que avalia de forma mais rigorosa a consistência estatística da frequência de exceções, o desempenho deste modelo degrada-se ainda mais, com cerca de 40% das janelas em vermelho. Isto confirma que a suposição de normalidade com volatilidade constante é incapaz de captar a verdadeira dinâmica das perdas extremas.

Já o modelo GARCH-Norm, reduz ligeiramente as observações nas zonas vermelha e amarela, sendo que em Basel quase não se verificam janelas na zona vermelha, embora se acumule mais de metade dos resultados na zona amarela, sugerindo que este modelo ainda tende a subestimar as perdas mais acentuadas em determinados períodos. No teste de Kupiec, o GARCH-Norm mostra uma melhoria clara face ao Norm, com a maioria dos resultados na zona verde, embora continue a apresentar quase 30% na zona vermelha. O H&W, por sua vez, destaca-se com o desempenho mais estável, sendo que tanto no Basel como no Kupiec a proporção de janelas em verde é largamente maioritária, e a percentagem em

vermelho é quase nula. Ou seja, entre os três, é aquele que mais se aproxima da consistência exigida pela regulamentação.

Ao analisar o ES através dos testes de Pearson e de Nass, torna-se ainda mais visível a diferença entre os modelos. O Norm, que já se mostrava fraco no VaR, degrada-se rapidamente quando aumenta a granularidade dos testes. No teste de Pearson, o desempenho do modelo é aceitável com $N = 2$, mas cai de forma pronunciada quando se aumenta para $N = 4$ e sobretudo $N = 8$, refletindo uma clara incapacidade de reproduzir a estrutura das caudas da distribuição dos retornos. O GARCH-Norm acompanha a mesma tendência apresentando percentagens mais elevadas de janelas em verde para $N = 2$, e perdendo estabilidade à medida que a exigência aumenta, com resultados no vermelho que rondam até aos 8%. O H&W, em contraste, mantém 100% das janelas em verde independentemente do valor de N , demonstrando que a sua capacidade de capturar a forma das caudas não é afetada pela granularidade do teste. Esta estabilidade é um ponto forte, mas também levanta a questão de se o modelo não será demasiado conservador, já que resultados perfeitos em todos os cenários são raros.

A mesma conclusão pode ser retirada dos testes de Nass, que são mais severos por introduzirem um ajustamento adicional à estatística. Aqui, tanto o Norm como o GARCH apresentam dificuldades acrescidas. O primeiro acumula entre 18% e 19% de janelas em vermelho, independentemente do valor de N , confirmando que o modelo é inadequado para capturar as perdas extremas. O GARCH, embora mais robusto, ainda apresenta entre 10% e 16% de resultados no vermelho, evidenciando que o simples ajuste da volatilidade não elimina o problema estrutural da suposição de normalidade. Mais uma vez, o H&W mantém resultados perfeitos em todas as configurações, reforçando a sua posição como modelo mais adequado. É de notar que, à medida que o número de classes N aumenta, os testes se tornam progressivamente mais exigentes, e essa pressão expõe as debilidades dos modelos paramétricos, mas não afeta o H&W.

Por fim, o teste de M&P oferece uma perspetiva alternativa, baseada no conceito de posição segura, como vimos anteriormente. Os resultados voltam a distinguir claramente os três modelos. O Norm falha de forma recorrente, com uma percentagem relevante na zona amarela e quase 10% no vermelho, revelando insuficiência na cobertura do risco extremo. O GARCH apresenta uma melhoria, quase eliminando o vermelho, mas com quase metade das janelas em amarelo, sinal de que a cobertura é muitas vezes apenas marginal. O H&W, em contraste, volta a alcançar 100% em zona verde, demonstrando cobertura total em todos os períodos.

De forma a analisar os resultados dos testes Z de Acerbi e Szekely, introduzidos anteriormente, aplicaram-se estes testes em intervalos discretos. Foram também incluídos novamente nesta análise o teste de Kupiec para o VaR e o teste M&P. Para o modelo GARCH-NORM foi

ainda incluído o teste Basel, que será útil mais adiante. Em concreto, consideraram-se trinta e uma amostras de 250 dias úteis, espaçadas de 130 dias (aproximadamente meio ano). Para cada um destes intervalos foram calculadas as estatísticas de teste definidas previamente e, de igual modo, o resultado foi classificado segundo as zonas do FRTB: Verde (V), Amarelo (A) ou Vermelho (R).

A Tabela 5.2 apresenta os resultados obtidos para o modelo Norm. Volta-se a verificar que o modelo revela fragilidades, sobretudo nos períodos de maior volatilidade. O teste de Kupiec alterna com frequência entre verde e vermelho, evidenciando que a taxa de exceções do VaR nem sempre se encontra calibrada. Entre os testes para o ES, o Z1 concentra muitos resultados no amarelo, indicando desvios persistentes, mas moderados, do ES face ao valor observado. Este teste sinaliza sobretudo desalinhamentos moderados, pois a sua formulação avalia o ES sob a hipótese de que o VaR está corretamente calibrado. Já o teste Z2, mais sensível à forma da cauda, acumula uma percentagem mais elevada de intervalos na zona vermelha. Quanto aos testes ZR e M&P, estes tendem a ser mais benignos, sugerindo que o nível médio do ES não está sistematicamente desajustado, ainda que a distribuição das caudas o esteja.

TABELA 5.2. Classificação dos testes em janelas discretas - Modelo *Norm*

Período	Kupiec	Z ₁	Z ₂	Z _R	M&P	Período	Kupiec	Z ₁	Z ₂	Z _R	M&P
Dec/08 - Dec/09	A	A	V	V	V	Dec/17 - Dec/17	V	V	A	V	V
Jun/09 - Jun/10	V	A	V	V	V	Jun/17 - Jun/18	R	A	R	R	A
Dec/09 - Dec/10	V	A	V	V	V	Dec/18 - Dec/18	R	A	R	R	R
Jun/10 - Jun/11	V	V	V	V	V	Jun/18 - Jun/19	A	A	A	A	A
Dec/10 - Dec/11	R	R	R	A	A	Dec/19 - Dec/20	V	A	V	A	V
Jun/11 - Jun/12	A	R	R	A	A	Jun/19 - Jun/20	R	R	R	R	R
Dec/11 - Dec/12	V	V	V	V	V	Dec/20 - Dec/21	R	R	R	R	R
Jun/12 - Jun/13	V	V	V	V	V	Jun/20 - Jun/21	A	V	V	V	V
Dec/12 - Dec/13	V	V	V	V	V	Dec/21 - Dec/22	V	V	V	V	V
Jun/13 - Jun/14	V	V	V	V	V	Jun/21 - Jun/22	R	A	R	A	A
Dec/13 - Dec/14	A	V	A	V	V	Dec/22 - Dec/23	R	A	R	A	A
Jun/14 - Jun/15	R	A	R	A	A	Jun/22 - Jun/23	V	V	V	V	V
Dec/15 - Dec/15	R	A	R	R	R	Dec/23 - Dec/24	V	V	V	V	V
Jun/15 - Jun/16	A	A	R	A	A	Jun/23 - Jun/24	V	V	V	V	V
Dec/16 - Dec/16	V	V	V	V	V	Dec/24 - Dec/25	V	A	A	V	A
Jun/16 - Jun/17	V	V	V	V	V						

Total (%)	Kupiec	Z ₁	Z ₂	Z _R	M&P
V	51,61%	45,16%	51,61%	58,06%	58,06%
A	19,35%	41,94%	12,90%	25,81%	29,03%
R	29,03%	12,90%	35,48%	16,13%	12,90%

Na tabela 5.3 apresentam-se os resultados obtidos para o modelo GARCH-Norm. Observa-se um ganho nítido na frequência das exceções, sendo que, podemos observar que o teste de

Kupiec apresenta mais verdes e menos vermelhos, como no caso diário, refletindo um melhor o controlo da variância condicional. O ZR também melhora de forma consistente, indicando que o ES previsto se aproxima, em média, do valor esperado. Contudo, persistem muitos amarelos em Z1 e Z2. Tal confirma o que foi referido anteriormente: Embora a heterocedasticidade esteja razoalmente capturada, a hipótese normal para os retornos continua a não ser suficiente para as caudas. Pode-se observar que em períodos de crise surgem ainda alguns vermelhos pontuais, mas com magnitude e persistência inferiores Às do Norm.

TABELA 5.3. Classificação dos testes em janelas discretas - Modelo *GARCH-Norm*

Período	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
Dec/08 - Dec/09	V	V	V	V	A	V
Jun/09 - Jun/10	R	A	R	A	A	A
Dec/09 - Dec/10	R	A	R	A	V	A
Jun/10 - Jun/11	V	V	V	V	V	V
Dec/10 - Dec/11	R	A	R	A	A	A
Jun/11 - Jun/12	V	A	A	V	A	A
Dec/11 - Dec/12	V	V	V	V	V	V
Jun/12 - Jun/13	V	V	V	V	V	V
Dec/12 - Dec/13	V	A	V	A	V	V
Jul/13 - Jun/14	V	V	A	V	A	V
Jan/14 - Dec/14	A	A	V	A	V	A
Jul/14 - Jun/15	V	A	V	V	A	V
Jan/15 - Dec/15	V	V	A	V	V	V
Jul/15 - Jun/16	A	A	V	A	V	A
Jan/16 - Dec/16	V	V	V	V	V	V
Jul/16 - Jun/17	V	V	V	V	A	V

Período	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
Dec/17 - Dec/17	A	A	V	A	A	V
Jun/17 - Jun/18	A	A	V	A	V	A
Dec/18 - Dec/18	R	A	R	A	V	A
Jun/18 - Jun/19	A	A	V	A	V	A
Dec/19 - Dec/20	V	A	A	A	V	V
Jun/19 - Jun/20	R	A	R	R	R	A
Dec/20 - Dec/21	R	A	R	A	A	A
Jun/20 - Jun/21	V	V	V	V	V	V
Dec/21 - Dec/22	V	A	A	V	A	A
Jun/21 - Jun/22	V	A	A	V	A	A
Dec/22 - Dec/23	A	A	A	V	V	A
Jun/22 - Jun/23	V	V	V	V	V	V
Dec/23 - Dec/24	V	V	V	V	V	V
Jun/23 - Jun/24	V	A	V	V	V	V
Dec/24 - Dec/25	V	A	V	A	V	V

Total (%)	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
V	58,06%	38,71%	58,06%	38,71%	54,84%	58,06%
A	29,03%	63,33%	38,71%	41,94%	41,94%	41,94%
R	12,90%	0,00%	3,23%	19,35%	3,23%	0,00%

Já a tabela 5.4 apresenta os resultados para o modelo H&W. Mais uma vez, este modelo apresenta um desempenho substancialmente mais estável. Os testes de Kupiec, Z1 e Z2 validam em verde quase todo o horizonte temporal e o M&P valida integralmente o modelo. O único teste que introduz alguma nuance é o ZR, com alguns amarelos dispersos. Estes resultados para o ZR devem-se sobretudo ao "excesso de cautela" e não por subestimação do risco, isto é, os valores afastam-se do previsto por serem conservadores e não por falta de cobertura da cauda da distribuição.

Em resumo, os testes aplicados às janelas discretas confirmam a leitura da análise diária. O modelo Norm mantém as fragilidades já referidas acima. O modelo Norm-GARCH apresenta algumas melhorias com a introdução da volatilidade condicional. O H&W valida de forma estável todos os modelos, embora esta validação esteja assente numa sobreavaliação do risco.

TABELA 5.4. Classificação dos testes em janelas discretas - Modelo H&W

Período	Kupiec	Z_1	Z_2	Z_R	M&P	Período	Kupiec	Z_1	Z_2	Z_R	M&P
Dec/08 - Dec/09	V	V	V	A	V	Dec/17 - Dec/17	V	V	V	V	V
Jun/09 - Jun/10	V	V	V	V	V	Jun/17 - Jun/18	V	V	V	V	V
Dec/09 - Dec/10	V	V	V	V	V	Dec/18 - Dec/18	V	V	V	V	V
Jun/10 - Jun/11	V	V	V	V	V	Jun/18 - Jun/19	V	V	V	V	V
Dec/10 - Dec/11	A	V	A	V	V	Dec/19 - Dec/20	V	V	V	A	V
Jun/11 - Jun/12	V	V	V	V	V	Jun/19 - Jun/20	V	V	V	V	V
Dec/11 - Dec/12	V	V	V	V	V	Dec/20 - Dec/21	V	V	V	V	V
Jun/12 - Jun/13	V	V	V	V	V	Jun/20 - Jun/21	V	V	V	V	V
Dec/12 - Dec/13	V	V	V	V	V	Dec/21 - Dec/22	V	V	V	V	V
Jun/13 - Jun/14	V	V	V	V	V	Jun/21 - Jun/22	V	A	V	V	V
Dec/13 - Dec/14	V	V	V	V	V	Dec/22 - Dec/23	V	V	V	V	V
Jun/14 - Jun/15	V	V	V	V	V	Jun/22 - Jun/23	V	V	V	V	V
Dec/15 - Dec/15	V	V	V	V	V	Dec/23 - Dec/24	V	V	V	V	V
Jun/15 - Jun/16	V	V	V	V	V	Jun/23 - Jun/24	V	V	V	V	V
Dec/16 - Dec/16	V	V	V	V	V	Dec/24 - Dec/25	V	V	V	V	V
Jun/16 - Jun/17	V	V	V	A	V						

Total (%)	Kupiec	Z_1	Z_2	Z_R	M&P
V	96,77%	96,77%	96,77%	90,32%	100,00%
A	3,23%	3,23%	3,23%	9,68%	0,00%

5.2. Requisitos de Capital

Um dos principais objetivos das medidas de risco é a determinação dos requisitos mínimos de capital que as instituições financeiras devem deter para cobrir potenciais perdas de mercado. O enquadramento regulamentar do *Fundamental Review of the Trading Book* (FRTB), liga de forma direta os resultados dos testes de backtesting à determinação destes requisitos através da aplicação de multiplicadores.

Como já vimos, os resultados do backtesting classificam os modelos de risco em diferentes zonas (verde, amarela ou vermelha), consoante o número de violações registadas. Cada zona corresponde a um nível de prudência distinto e traduz-se em acréscimos de capital sobre o valor inicialmente calculado pelo modelo. Esta lógica de penalização procura garantir que, sempre que o modelo subestima o risco, o capital exigido seja aumentado, assegurando assim uma margem de segurança adicional.

Embora o FRTB tenha adotado o *Expected Shortfall* (ES) como medida principal de risco, o backtesting regulamentar permanece baseado no *Value at Risk* (VaR). Esta divergência levanta uma questão central: de que forma seria possível estender o mecanismo de multiplicadores, já consagrado para o VaR, a procedimentos de backtesting do ES. No caso do VaR, o procedimento regulamentar assenta num horizonte de 250 dias úteis e consiste na contagem do número de violações em que a perda observada supera o valor estimado. Dependendo

desse número de violações, o modelo é classificado numa das três zonas do *traffic-light approach*: até quatro violações, zona verde; entre cinco e nove, zona amarela; e dez ou mais, zona vermelha. A cada zona está associado um multiplicador aplicado ao requisito de capital, cujo objetivo é elevar o nível de cobertura até ao patamar prudencial de 99% de confiança. O fator mínimo é sempre 1,5, podendo atingir o valor de 2,0 nos cenários mais adversos.

Para o caso do ES, uma forma de estender esta lógica é através do teste de Moldenhauer e Pitera (2019). Neste procedimento, em vez do número de exceções de VaR, utiliza-se a estatística de teste G , que corresponde ao número de observações negativas da posição segura, conforme definido na equação 3.25. Sob um modelo corretamente especificado, o valor esperado de G é de cerca de 6,25, o que corresponde ao número médio de exceções esperado para o ES 97,5%. Os autores definem como zona verde até 11 observações negativas, como zona amarela entre 12 e 24, e como zona vermelha a partir de 25. Para alinhar este procedimento com o mecanismo regulamentar do VaR, é possível mapear o número de exceções de Basel para intervalos equivalentes de G , atribuindo-lhes os mesmos multiplicadores de capital. A correspondência entre exceções do VaR e da estatística G do teste M&P permite obter os resultados apresentados na Tabela 5.5.

TABELA 5.5. Multiplicadores de capital Basel e M&P

Número de Exceções	Taxa observada (π_{obs})	G (M&P test)	Multiplicador
5	2%	12, 13, 14	1,70
6	2,4%	15, 16	1,76
7	2,8%	17, 18, 19	1,83
8	3,2%	20, 21	1,88
9	3,6%	22, 23, 24	1,92
≥ 10	$\geq 4,0\%$	≥ 25	2,00

Deste modo, o teste de M&P permite construir um esquema de multiplicadores análogo ao do VaR, mas aplicado ao ES. Apesar da simplicidade e da coerência regulatória desta abordagem, importa sublinhar que a definição dos limiares de G carece de fundamentação estatística robusta, o que pode conduzir a alguma instabilidade nos requisitos de capital em períodos de maior volatilidade.

Para esta dissertação, a análise do impacto nos requisitos de capital será realizada apenas para o modelo GARCH-Norm. Esta opção prende-se com o facto de este modelo já incorporar a heterocedasticidade dos retornos através da estimação da volatilidade condicional, apresentando assim um maior realismo na mensuração do risco quando comparado com a abordagem Normal. Por outro lado, ao contrário do H&W, o GARCH-Norm não se revela excessivamente conservador, permitindo uma avaliação mais equilibrada da aplicação dos multiplicadores ao Expected Shortfall.

Seguindo a mesma lógica aplicada ao VaR no FRTB, os resultados do teste M&P podem então ser traduzidos em multiplicadores de capital equivalentes, tal como mostrado na Tabela 5.5. A comparação entre os multiplicadores implícitos do M&P e os definidos pelo Basel para o VaR permite avaliar de que forma a utilização de um procedimento de backtesting baseado no ES pode alterar os requisitos de capital ao longo do tempo. A Figura 5.2 apresenta a comparação entre os multiplicadores obtidos pelo Basel e pelo M&P, aplicados ao modelo GARCH-Norm.

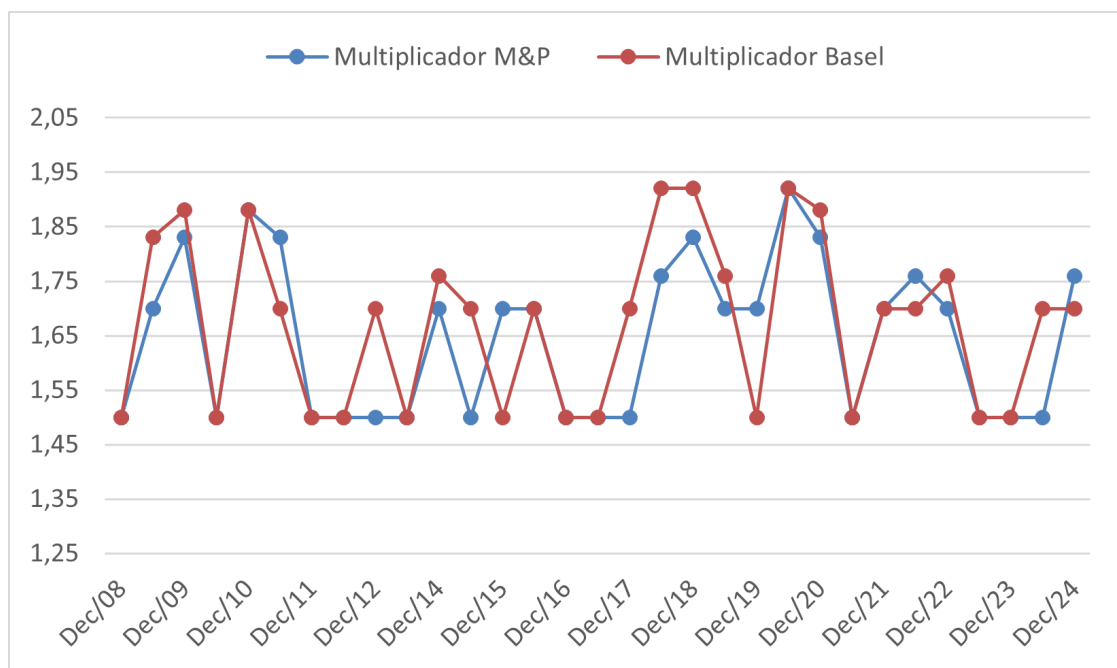


FIGURA 5.1. Comparação entre os multiplicadores de capital para o GARCH-NORM

A figura mostra que, apesar de em muitos períodos ambos os testes produzirem o mesmo multiplicador, existem momentos em que o M&P gera valores ligeiramente inferiores, sobretudo em fases de maior instabilidade dos mercados. Este resultado sugere que, embora o teste de M&P assegure coerência com a lógica do Basel, a sua utilização pode implicar requisitos de capital menos prudentes em determinados contextos de crise, salientando a importância de avaliar cuidadosamente o seu enquadramento regulatório. Conclui-se assim que a transição para um mecanismo de backtesting baseado no ES implicaria alterações não negligenciáveis nos requisitos de capital, podendo modificar de forma significativa a carga regulatória aplicada às instituições financeiras.

5.3. Impactos de uma variação na carteira

A motivação para introduzir diferentes composições de carteira prende-se com a necessidade de avaliar a robustez dos resultados obtidos nos testes de backtesting. Em teoria, a variação da carteira não deveria ter impacto relevante nos resultados destes testes, uma vez que o

objetivo do backtesting é aferir a qualidade do modelo de risco utilizado para estimar medidas como o VaR e o ES. Assim, um modelo corretamente especificado deverá manter a mesma capacidade preditiva independentemente da carteira considerada, dado que a avaliação incide sobre a adequação estatística da distribuição condicional dos retornos e não sobre o perfil específico dos ativos. No entanto, na prática, diferentes carteiras podem apresentar dinâmicas distintas de volatilidade, correlação e exposição a eventos extremos, o que pode influenciar tanto a frequência das exceções como a severidade das perdas. Deste modo, a análise da variação da carteira permite verificar se os resultados obtidos são consistentes e estáveis perante diferentes perfis de risco, assegurando que as conclusões não estão excessivamente dependentes da carteira inicial escolhida.

Para avaliar a robustez dos resultados, procedeu-se a uma alteração na composição da carteira inicial. Neste exercício, manteve-se o mesmo conjunto de ativos, mas ajustaram-se os pesos atribuídos a cada mercado, de modo a testar se diferentes distribuições de alocação teriam impacto nos resultados dos testes de backtesting. Assim, a carteira resultante apresenta a estrutura de exposições descrita na Tabela 5.6, permitindo uma análise comparativa direta com os resultados obtidos anteriormente.

TABELA 5.6. Nova exposição da carteira

Fator de Risco	Exposição
Apple	0,08
Microsoft	0,08
Siemens	0,14
Bayer	0,14
Tencent	0,18
Ping An	0,18
Unilever	0,10
HSBC	0,10
USD/EUR	0,16
HKD/EUR	0,36
GBP/EUR	0,20

Tal como no capítulo anterior, a validação dos resultados será conduzida apenas para o modelo GARCH-Norm. Esta escolha justifica-se novamente pelo facto de este modelo já ter demonstrado uma maior capacidade de captar a heterocedasticidade dos retornos e de refletir de forma realista a dinâmica de volatilidade, sem apresentar o excessivo conservadorismo observado no H&W. Desta forma, assegura-se a comparabilidade entre as duas carteiras e mantém-se o foco na avaliação do impacto das alterações de pesos nos requisitos de capital.

Seguidamente apresentam-se os resultados obtidos com a nova carteira, aplicando os testes de backtesting em janela deslizante ao modelo GARCH-Norm. A Tabela 5.7 resume as frequências observadas em cada zona.

TABELA 5.7. Resultados dos testes em janela deslizante - Nova carteira

Teste		GARCH-Norm		
		V	A	R
VaR 99%	Basel	44,32	50,19	5,49
	Kupiec	56,85	15,99	27,48
ES 97,5%	Pearson N=2	90,74	4,62	4,64
	Pearson N=4	86,48	6,81	6,71
	Pearson N=8	77,51	11,16	11,33
	Nass N=2	79,37	6,41	14,23
	Nass N=4	74,59	9,47	15,94
	Nass N=8	75,51	10,86	13,64
	M&P	56,62	43,25	1,13

Os resultados obtidos confirmam a estabilidade geral já observada anteriormente. Para o VaR, o teste de Basel apresenta uma distribuição equilibrada entre as zonas verde e amarela, enquanto o teste de Kupiec continua a evidenciar fragilidades, com uma percentagem significativa de classificações na zona vermelha. No que respeita ao ES, os testes de Pearson voltam a validar de forma robusta o modelo, com uma clara predominância de observações na zona verde, embora com uma ligeira dispersão acrescida quando se aumenta o número de níveis N . Os testes de Nass mostram uma maior sensibilidade, conduzindo a uma percentagem relevante de resultados na zona vermelha, enquanto o teste de M&P mantém o padrão de classificação observado anteriormente, com elevada proporção em zona amarela e apenas uma pequena fração em zona vermelha.

Em termos globais, conclui-se que a alteração na composição da carteira não alterou de forma significativa as conclusões do capítulo anterior. o GARCH-Norm mantém um desempenho consistente nos testes baseados no ES, ao passo que os testes de VaR revelam maior variabilidade entre métodos. Este resultado confirma a ideia inicial de que a variação da carteira, em princípio, não deveria afetar substancialmente a validação estatística dos modelos, reforçando assim a robustez das conclusões obtidas.

Para complementar a análise, apresentam-se de seguida os resultados obtidos em janelas discretas, de 250 observações com intervalos de 130 dias, aplicados à nova carteira com o modelo GARCH-Norm. A Tabela 5.8 mostra a classificação de cada período nos diferentes testes considerados, bem como o resumo percentual das frequências em cada zona.

A leitura da Tabela 5.8 confirma que a alteração da carteira não produziu mudanças substanciais face aos resultados anteriores. Os testes baseados no ES continuam a validar maioritariamente o modelo, com uma elevada percentagem de classificações na zona verde e

TABELA 5.8. Classificação dos testes de backtesting em janelas discretas para a nova carteira - Modelo *GARCH-Norm*

Período	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
Dec/08 - Dec/09	V	V	V	V	A	V
Jun/09 - Jun/10	V	A	V	A	A	V
Dec/09 - Dec/10	A	A	V	A	V	A
Jun/10 - Jun/11	V	V	A	V	V	V
Dec/10 - Dec/11	A	A	A	R	A	A
Jun/11 - Jun/12	A	A	A	R	A	A
Dec/11 - Dec/12	V	V	V	V	A	V
Jun/12 - Jun/13	V	A	V	A	V	V
Dec/12 - Dec/13	V	A	V	A	V	V
Jun/13 - Jun/14	V	V	A	V	A	V
Dec/13 - Dec/14	A	A	V	A	V	A
Jun/14 - Jun/15	V	V	V	V	A	V
Dec/14 - Dec/15	V	V	A	V	V	V
Jun/15 - Jun/16	A	A	V	A	V	A
Dec/15 - Dec/16	V	V	V	V	V	V
Jun/16 - Jun/17	V	V	V	V	A	V

Período	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
Dec/16 - Dec/17	A	A	V	A	A	V
Jun/17 - Jun/18	A	A	V	A	V	A
Dec/17 - Dec/18	A	R	V	A	V	A
Jun/18 - Jun/19	A	A	V	A	V	A
Dec/18 - Dec/19	V	A	A	A	A	V
Jun/19 - Jun/20	R	A	R	R	R	A
Dec/19 - Dec/20	R	A	A	R	A	A
Jun/20 - Jun/21	V	A	V	V	V	V
Dec/20 - Dec/21	R	R	A	R	A	A
Jun/21 - Jun/22	R	R	A	R	A	A
Dec/21 - Dec/22	V	A	A	A	V	A
Jun/22 - Jun/23	V	V	A	V	V	V
Dec/22 - Dec/23	V	V	A	V	V	V
Jun/23 - Jun/24	V	V	V	V	V	V
Dec/24 - Dec/25	V	A	V	A	V	V

Total (%)	Kupiec	Basel	Z ₁	Z ₂	Z _R	M&P
V	58,06%	35,48%	58,06%	38,71%	54,84%	58,06%
A	29,03%	54,84%	38,71%	41,94%	41,94%	41,94%
R	12,90%	9,68%	3,23%	19,35%	3,23%	0,00%

até ausência de resultados em zona vermelha no caso do M&P. Já os testes de VaR apresentam maior variabilidade, sendo que no teste de Basel os resultados concentram-se sobretudo na zona amarela, enquanto o Kupiec registra uma maior dispersão, incluindo cerca de 13% de classificações em zona vermelha. O teste Z_2 mantém o padrão de maior severidade, com quase 20% de observações em zona vermelha, confirmando a sua sensibilidade acrescida a desvios entre exceções observadas e esperadas. Em termos gerais, a consistência dos resultados reforça a conclusão de que a variação da carteira não altera de forma material a avaliação do modelo GARCH-Norm, consolidando a robustez das inferências obtidas.

A análise dos requisitos de capital constitui, no entanto, um passo essencial para compreender as implicações práticas dos resultados de backtesting. Como referido em capítulos anteriores, o FRTB estabelece uma ligação direta entre a classificação dos testes e a determinação dos multiplicadores de capital. Assim, ainda que em termos estatísticos a alteração da carteira não tenha modificado significativamente a avaliação dos modelos, importa verificar se esta variação influencia o capital regulatório exigido. Para esse efeito, analisam-se os multiplicadores implícitos do teste de Moldenhauer & Pitera (M&P) aplicados ao modelo GARCH-Norm.

A Figura 5.2 mostra a evolução dos multiplicadores de capital para as duas carteiras. Observa-se que, embora ambas sigam trajetórias semelhantes, a carteira alterada apresenta uma tendência para multiplicadores mais elevados, situando-se em vários períodos próximos dos valores máximos definidos pelo FRTB. Já a carteira original revela maior variabilidade,

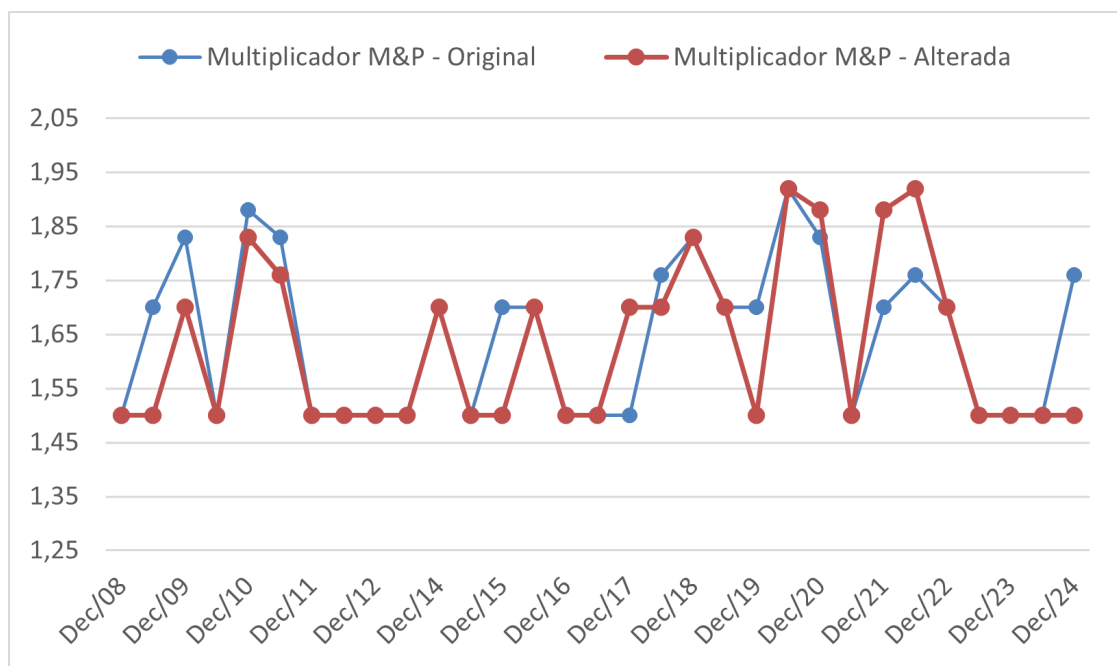


FIGURA 5.2. Comparação dos multiplicadores de capital entre as carteiras

com fases em que os multiplicadores se aproximam do mínimo regulamentar de 1.5 intercaladas com períodos de maior penalização.

Estes resultados indicam que a alteração da composição da carteira, apesar de não comprometer a robustez estatística do modelo, tem impacto direto no capital regulamentar exigido. Em particular, a carteira alterada conduz a uma exigência de capital sistematicamente superior, refletindo um perfil de risco mais conservador do ponto de vista prudencial. Em síntese, conclui-se que a variação da carteira não afeta a consistência dos testes estatísticos, mas traduz-se em diferenças relevantes nos requisitos de capital, mostrando que a composição do portfólio pode influenciar de forma determinante o nível de prudência imposto pelo enquadramento regulamentar.

CAPÍTULO 6

Conclusões

O presente trabalho teve como principal objetivo avaliar a validade estatística de diferentes modelos de mensuração de risco, em particular do *Expected Shortfall* (ES) e do *Value at Risk* (VaR), no contexto regulamentar do *Fundamental Review of the Trading Book* (FRTB). O estudo partiu do reconhecimento de que a introdução do ES como medida de risco regulamentar levanta desafios relevantes em termos de backtesting, uma vez que os mecanismos consagrados na regulação permanecem assentes no VaR. Este desfasamento cria um espaço de investigação que é simultaneamente académico e prático, dada a relevância direta para instituições financeiras sujeitas a requisitos prudenciais de capital.

No plano empírico, a análise recorreu a uma carteira de ações representativas de diferentes mercados geográficos e setores, de modo a assegurar um portefólio diversificado e realista. Foram implementados três modelos para estimação de risco: o modelo Normal, o GARCH-Norm e o H&W. Após um conjunto inicial de análises, optou-se por concentrar a avaliação mais aprofundada no GARCH-Norm, uma vez que este conseguiu aliar maior realismo na dinâmica de volatilidade a uma validação estatística robusta, sem incorrer no excesso de conservadorismo que caracterizou o H&W. A utilização de janelas móveis e discretas permitiu uma visão abrangente, distinguindo a avaliação em tempo contínuo da estabilidade das classificações em períodos definidos.

Os resultados obtidos permitem retirar várias conclusões relevantes. Em primeiro lugar, confirmou-se a expectativa de que os modelos baseados na normalidade são insuficientes para capturar a heterocedasticidade e a complexidade dos retornos financeiros, resultando em desempenho frágil nos testes de backtesting. Em segundo lugar, verificou-se que a introdução de volatilidade condicional via GARCH melhora significativamente a qualidade da mensuração de risco, traduzindo-se em maior percentagem de validações na zona verde e numa redução dos períodos críticos em zona vermelha. Em terceiro lugar, constatou-se que a escolha do teste estatístico exerce um impacto material sobre a avaliação: enquanto os testes de Pearson validaram de forma consistente o modelo GARCH-Norm, os testes de Nass e, sobretudo, o Z2 revelaram maior severidade, classificando com maior frequência em zona vermelha. Esta diversidade de resultados reforça a ideia de que não existe um teste universalmente superior, mas sim diferentes instrumentos complementares de diagnóstico.

Um aspeto particularmente relevante foi a análise do teste M&P, que, ao permitir uma correspondência direta com a lógica do Basel, oferece uma ponte natural para a extensão

do mecanismo de multiplicadores ao ES. Os resultados mostraram que, embora este teste seja simples e coerente com a prática regulamentar, pode gerar classificações excessivamente concentradas na zona amarela, levantando dúvidas sobre o seu caráter prudencial. Ainda assim, o M&P revelou-se útil para quantificar o impacto potencial de uma transição para backtesting do ES, permitindo construir multiplicadores de capital implícitos e compará-los com os definidos pelo Basel para o VaR.

A extensão da análise à variação da carteira trouxe uma dimensão adicional de robustez. Teoricamente, a composição do portfólio não deveria afetar os resultados estatísticos do backtesting, e os resultados confirmaram esta expectativa: os testes de ES mantiveram-se estáveis e consistentes mesmo após a alteração dos pesos. No entanto, a análise dos requisitos de capital revelou diferenças importantes: a carteira alterada originou multiplicadores mais elevados e estáveis, traduzindo-se numa exigência de capital sistematicamente superior. Esta evidência demonstra que, embora os testes estatísticos validem de forma semelhante os modelos, a dimensão prudencial depende da composição do portfólio, salientando o papel determinante que a carteira assume na regulação bancária.

Em termos práticos, as conclusões deste trabalho apontam para três implicações centrais. Primeiro, confirmam a superioridade de modelos com volatilidade condicional (como o GARCH-Norm) face a abordagens mais simplistas, sugerindo que as instituições financeiras devem privilegiar este tipo de modelização na sua mensuração de risco. Segundo, evidenciam a necessidade de um enquadramento regulamentar mais adaptado ao ES, capaz de refletir a diversidade de resultados entre testes e de incorporar métricas que assegurem prudência sem sobrecapitalização. Terceiro, demonstram que a gestão ativa da carteira pode ter efeitos indiretos nos requisitos de capital, mesmo quando a validade estatística dos modelos não é posta em causa, reforçando a importância de integrar a dimensão regulatória nas decisões de gestão de risco.

Naturalmente, este trabalho não está isento de limitações. A escolha de três modelos restringe a generalização das conclusões, deixando espaço para avaliar outras abordagens, como modelos com distribuições não normais, volatilidade estocástica ou métodos puramente históricos. Adicionalmente, a análise foi conduzida com dados e horizontes específicos, o que significa que resultados distintos poderão ser obtidos em outros contextos ou com carteiras de natureza diferente. Do mesmo modo, a aplicação prática dos testes de backtesting do ES permanece condicionada pela sua ausência na regulamentação formal, o que limita a comparação direta com os requisitos atualmente em vigor.

Como perspetivas de investigação futura, seria interessante aprofundar a comparação entre testes estatísticos em amostras mais amplas, explorar modelos alternativos de previsão de volatilidade e avaliar de forma mais detalhada o impacto dos testes ZR e ESR, propostos recentemente na literatura. Outro caminho relevante passa por estudar a interação entre

gestão de carteira e requisitos de capital, de modo a perceber até que ponto a regulação pode induzir comportamentos estratégicos por parte das instituições. Finalmente, a análise comparativa entre regimes de normalidade e regimes de cauda pesada poderia clarificar a robustez das conclusões em cenários de crise financeira.

Em suma, esta dissertação mostrou que, embora a transição para o ES represente um avanço conceptual e regulamentar, a sua operacionalização em termos de backtesting levanta desafios importantes. Os resultados empíricos confirmaram a robustez do modelo GARCH-Norm e a utilidade de vários testes estatísticos, mas também evidenciaram que a composição da carteira pode influenciar os requisitos de capital, mesmo quando a validade estatística dos modelos não é posta em causa. Assim, mais do que uma questão técnica, a evolução do backtesting do ES representa um debate de natureza prudencial, em que a estabilidade do sistema financeiro deve ser equilibrada com a proporcionalidade dos requisitos de capital aplicados às instituições.

Referências Bibliográficas

- Acerbi, C. (2019). Backtesting expected shortfall. *Risk Magazine*, September, 76–81.
- Acerbi, C., & Szekely, B. (2014). Backtesting expected shortfall. *Risk Magazine*, December.
- Acerbi, C., & Szekely, B. (2017). General properties of backtestable statistics. *SSRN Electronic Journal*.
- Acerbi, C., & Szekely, B. (2023). Backtestability and the ridge backtest. *Frontiers of Mathematical Finance*, 2(4), 497–521.
- Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7), 1487–1503.
- Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228.
- Artzner, P., Delbaen, F., Koch-Medina, P., & Munari, C. (2019). Robust acceptability and coherent risk measures. *Mathematical Finance*, 29(3), 699–736.
- Basel Committee on Banking Supervision (BCBS). (2016). Minimum capital requirements for market risk. Bank for International Settlements.
- Bayer, S., & Dimitriadis, T. (2019). Regression-based expected shortfall backtesting. *Journal of Business and Economic Statistics*, 39(1), 220–233.
- Berglund, E., & Markgren, A. (2022). Backtesting expected shortfall: A qualitative study for central counterparty clearing. Master’s thesis, Umeå University.
- Catarino, J. (2023). Backtesting Expected Shortfall: Comparative study and impact analysis on capital requirements. Dissertação de Mestrado, ISCTE Business School.
- Chen, J. M. (2014). Measuring market risk under the Basel Accords: VaR, stressed VaR, and expected shortfall. *Aestimatio, The IEB International Journal of Finance*, 8, 184–201.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Emmer, S., Klüppelberg, C., & Korn, R. (2001). Optimal portfolios with bounded capital at risk. *Mathematical Finance*, 11(4), 365–384.
- Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, 18(2), 31–60.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Gameiro, A. (2017). Avaliação empírica do risco de mercado: Expected Shortfall vs. Value-at-Risk. Dissertação de Mestrado, Faculdade de Ciências da Universidade de Lisboa / ISCTE.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Hull, J. & White, A. (1998). Incorporating volatility updating into the historical simulation method for Value at Risk. *Journal of Risk*, 1(1), 5–19.
- Kratz, M., Lok, Y., & McNeil, A. J. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking and Finance*, 88, 393–407.

- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84.
- Lambert, N. S., Pennock, D. M., & Shoham, Y. (2008). Eliciting properties of probability distributions. *Proceedings of the 9th ACM Conference on Electronic Commerce*, 129–138.
- McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools* (2nd ed.). Princeton University Press.
- Moldenhauer, F., & Pitera, M. (2018). Backtesting expected shortfall: A simple recipe? *SSRN Electronic Journal*.
- Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, 11(4), 1833–1874.
- Osband, K. H. (1985). Providing incentives for better cost forecasting. *International Journal of Forecasting*, 1(3), 289–305.
- Osmundsen, K. (2016). Backtesting Expected Shortfall: A comparative study of methods. Master’s thesis, Norwegian School of Economics.
- Pereira, A., & Ferreira, E. (2018). The use of GARCH models in VaR estimation. *Revista de Estudos em Finanças*, 6(2), 45–62.
- Roccioletti, F. (2016). Expected shortfall: Backtesting and applications. *International Journal of Financial Studies*, 4(2), 1–12.
- Spring, K. (2021). Backtesting the expected shortfall. *Junior Management Science*, 6(3), 590–636.
- Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking and Finance*, 32(7), 1404–1415.
- Yurttagüler, I. M. (2024). Analysis of Istanbul stock market returns volatility with ARCH and GARCH models. *Istanbul Journal of Economics*, 74(1), 37–58.

Anexo A : Descrição do portfólio

TABELA 6.1. Estatísticas descritivas do portfólio

Ação	Média	Máx.	Mín.	Mediana	Desvio- padrão	Assimetria
AAPL	0,0008	0,1467	-0,1833	0,0004	0,0204	-0,2562
MSFT	0,0007	0,1540	-0,1568	0,0000	0,0185	-0,0336
Siemens	0,0003	0,2157	-0,1873	0,0001	0,0188	-0,1436
Bayer	-0,0002	0,1257	-0,1980	0,0000	0,0189	-0,6661
Tencent	0,0009	0,2061	-0,1395	0,0000	0,0235	0,0862
PingAn	0,0002	0,1798	-0,1724	0,0000	0,0238	0,1534
Unilever	0,0002	0,1214	-0,1146	0,0000	0,0142	0,0461
HSBC	0,0000	0,1423	-0,2195	0,0000	0,0185	-0,3933
Portfólio	0,0004	0,1033	-0,0893	0,0008	0,0116	-0,3590

Anexo B: Estimação dos Valores de VaR/ES

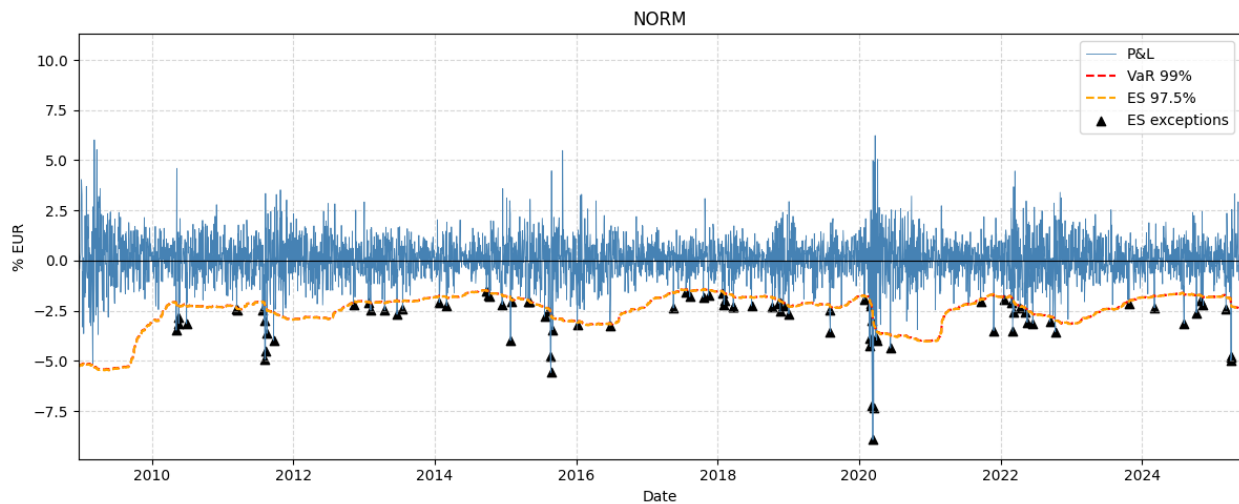


FIGURA 6.1. PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo Norm

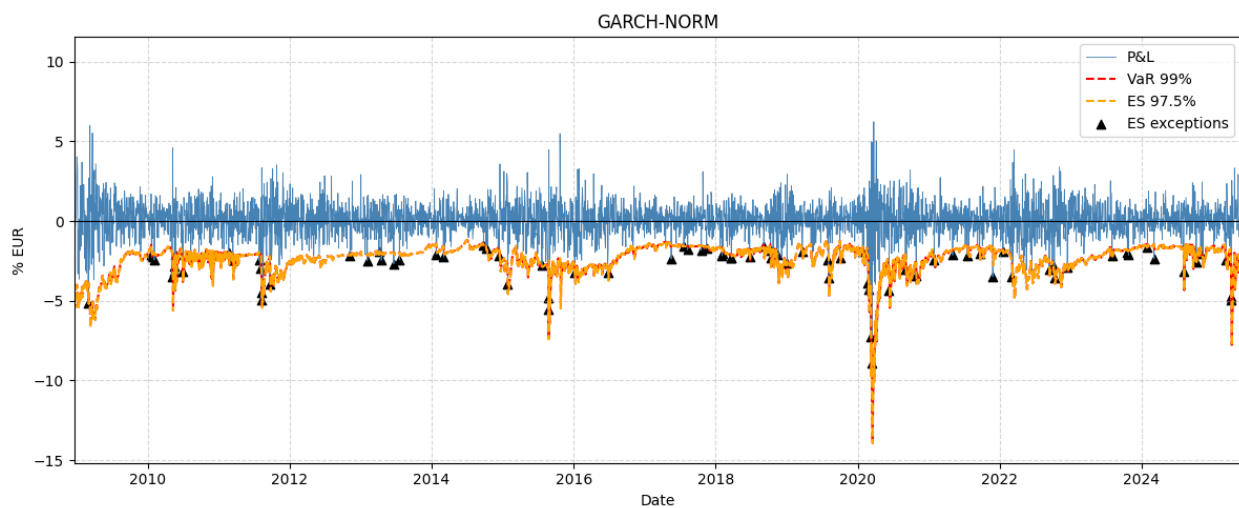


FIGURA 6.2. PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo GARCH-Norm

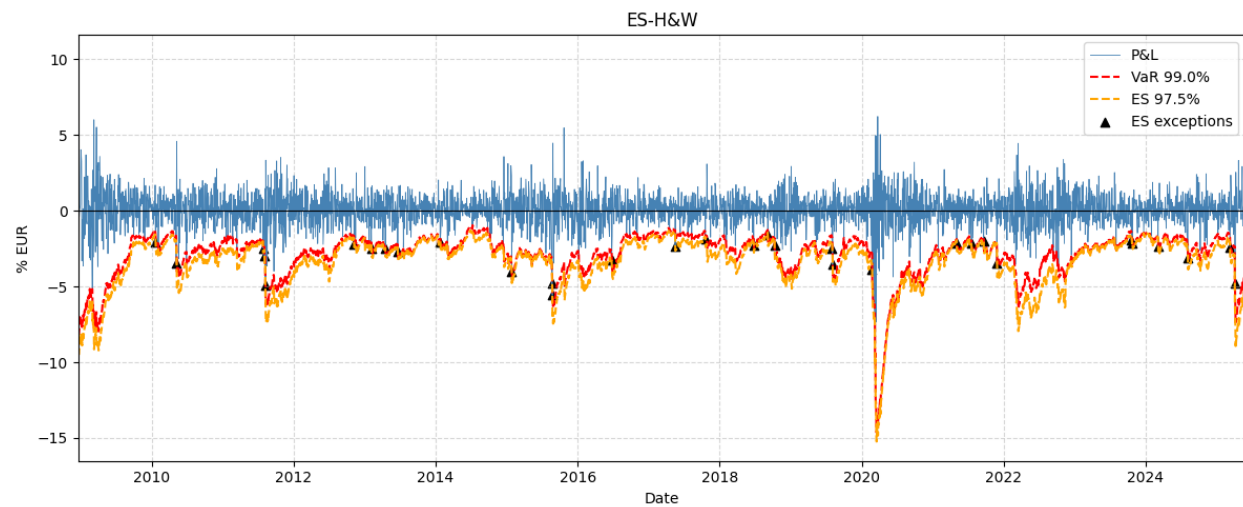


FIGURA 6.3. PnL do portfólio com o VaR 99% e o ES 97,5% - Modelo H&W