

UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS – ICSA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA APLICADA - PPEA

DISSERTAÇÃO EM ECONOMIA APLICADA

BACKTESTING DO VALUE-AT-RISK (VaR): UMA PERSPECTIVA SOB A ANÁLISE
ESTATÍSTICA SEQUENCIAL

BRUNA MARQUES AGUIAR MACHADO

Mariana, MG
2020

BRUNA MARQUES AGUIAR MACHADO

***BACKTESTING DO VALUE-AT-RISK (VaR): UMA PERSPECTIVA SOB A ANÁLISE
ESTATÍSTICA SEQUENCIAL***

Dissertação apresentada ao Programa de Pós-Graduação em Economia Aplicada do Instituto de Ciências Sociais Aplicadas da Universidade Federal de Ouro Preto, como requisito parcial à obtenção do Título de Mestre em Economia Aplicada.

Área de concentração: Economia Aplicada

Orientador: Prof. Dr. Ivair Ramos Silva
Coorientador: Prof. Dr. Júlio César Araújo
Silva Júnior

Mariana, MG
2020

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

M149b Machado, Bruna Marques Aguiar .
Backtesting do value-at-risk (VaR) [manuscrito]: uma perspectiva sob
a análise estatística sequencial. / Bruna Marques Aguiar Machado. - 2020.
86 f.

Orientador: Prof. Dr. Ivair Ramos Silva.
Coorientador: Prof. Dr. Júlio César Araújo Silva Júnior.
Dissertação (Mestrado Acadêmico). Universidade Federal de Ouro
Preto. Instituto de Ciências Sociais Aplicadas. Programa de Pós-
Graduação em Economia Aplicada.
Área de Concentração: Economia Aplicada.

1. Avaliação (Finanças). 2. Economia. 3. Probabilidades. 4. Risco
(Economia). 5. Risco financeiro. I. Júnior, Júlio César Araújo Silva . II. Silva,
Ivair Ramos. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 330.101.8



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA



FOLHA DE APROVAÇÃO

Bruna Marques Aguiar Machado

Backtesting do Value-at-Risk(VaR): Uma perspectiva sob a Análise Estatística Sequencial

Membros da banca

Carlos Eduardo da Gama Torres - Doutor - Universidade Federal de Ouro Preto
Hudson Torrent - Doutor - Universidade Federal do Rio Grande do Sul
Júlio César Araújo da Silva Júnior - Doutor - Universidade Federal de Viçosa

Versão final

Aprovado em 30 de setembro de 2020

De acordo

Ivair Ramos Silva



Documento assinado eletronicamente por **Ivair Ramos Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 05/10/2020, às 01:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0089551** e o código CRC **9C9FAEE2**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.007310/2020-02

SEI nº 0089551

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

AGRADECIMENTOS

É chegado o momento em que posso traduzir em palavras todo o sentimento de gratidão que me acompanhou durante todo esse processo de amadurecimento pessoal e profissional. A conclusão deste trabalho é mais do que um trabalho acadêmico, é um sonho alcançado. Os frutos que colho hoje são originários das sementes que plantei há muito tempo, bem antes do início do mestrado, e que só foi possível com o apoio e amor dos que se fizeram presentes nesse longo processo.

Agradeço a Deus, às vibrações do universo e toda e qualquer energia divina que me acompanha. Nos momentos de desespero e sufoco, foram a eles que voltei minhas preces para que me ajudassem a reconectar com a minha verdadeira essência e pudesse enxergar além das dificuldades encontradas pelo caminho.

À minha família que sempre depositou confiança em todos os caminhos que desejei trilhar. Em especial à minha mãe e minha avó Vera, nossas raízes estão conectadas de uma forma tão profunda que toda a bondade que há no coração de vocês reverbera em mim, só cheguei até aqui porque tenho vocês.

Ao meu companheiro de vida Caio, quem me acompanhou no início, meio e fim desse processo, segurando na minha mão e me apoiando incondicionalmente em todas as minhas decisões. O amor que existe entre nós é sublime.

Aos meus amigos de Mariana e Pedro Leopoldo que estiveram comigo mesmo na distância, obrigada por me incentivarem a continuar e a não perder as esperanças. A nossa irmandade me faz acreditar que o significado de família vai além dos laços sanguíneos.

Ao meu orientador Ivair, me faltam palavras para expressar tamanha admiração pela forma a qual você lida com a pesquisa e com as pessoas. Agradeço todo o conhecimento compartilhado até aqui e espero que possamos colher mais frutos desse trabalho.

Ao meu coorientador Júlio, sua ajuda foi fundamental na etapa final da pesquisa, obrigada pelo carinho e atenção de sempre.

Aos meus professores do PPEA/UFOP, em especial, aos professores Héder, Fernanda e Mirian por me confiarem atividades acadêmicas importantes nesse processo de desenvolvimento profissional, aprendo muito com vocês.

Ao meu amigo de turma Caio, sua amizade foi muito importante na caminhada até aqui. Todas as alegrias e tristezas compartilhadas foram fundamentais para o nosso amadurecimento. Espero que nossa parceria perdure no tempo.

À Universidade Federal de Ouro Preto, por oferecer ensino gratuito e de excelente qualidade. Foram sete bons anos desfrutando de todo o conhecimento que a universidade pública pôde me proporcionar. Tenho orgulho em dizer de onde vim.

À PROPP/UFOP por ter subsidiado esta pesquisa, sem esse apoio eu não estaria aqui defendendo este trabalho.

*Ocupo muito de mim com o meu desconhecer.
Sou um sujeito letrado em dicionários.
Não tenho mais que 100 palavras.
Pelo menos uma vez por dia me vou no
Morais ou no Viterbo –
A fim de consertar a minha ignorãça,
mas só acrescenta.
Despesas para minha erudição tiro nos
almanaques:
- Ser ou não ser, eis a questão.
Ou na porta dos cemitérios:
- Lembra que és pó e que ao pó tu
volarás.
Ou no verso das folhinhas:
- Conhece-te a ti mesmo.
Ou na boca do povinho:
- Coisa que não acaba no mundo é gente
besta
e pau seco.
Etc
Etc
Etc
Maior que o infinito é a encomenda.
(Manoel de Barros - Livro das Ignorãças)*

RESUMO

O objetivo dessa dissertação foi propor uma mudança de paradigma acerca da construção do *backtesting* do *Value-at-Risk* (VaR). Uma vez que os *backtests* discutidos na literatura podem ser interpretados como um processo pertencente ao controle estatístico de qualidade, essa dissertação apresenta argumentos de que os *backtests* podem ser elucidados pela análise estatística sequencial. Dessa forma, a partir dos estudos de simulações em que se investiga o comportamento dos *backtests* a partir da visão dos gráficos de controle, foi constatado que esses *backtests* retornam probabilidades do erro do Tipo I maiores que o nível de significância estipulado no início do experimento, o que pode trazer sérias consequências para a tomada de decisão quanto a adequabilidade do VaR. Como nova proposta de *backtesting* a partir da análise sequencial, foi construído uma função “*gasto de alpha*” capaz de controlar a real probabilidade do erro do Tipo I do experimento, associados a valores críticos exatos. Além disso, o novo *backtesting* tem alto poder em rejeitar um modelo VaR em um contexto que o risco relativo observado é muito maior que o risco especificado pela hipótese nula.

Palavras-chave: *Value-at-Risk*, *backtesting*, probabilidade do erro do Tipo I, análise sequencial, gráficos de controle, “*gasto de alpha*”.

ABSTRACT

The aim this thesis was to propose a paradigm shift on the construction of Value-at-Risk (VaR) backtesting. Since the backtests discussed in the literature can be interpreted as a process belonging to statistical quality control, this thesis presents arguments that the backtests can be elucidated by the sequential statistical analysis. Thus, the simulation studies was formulated to verify the behavior of the backtests from the view of the control charts, it was found that these backtests return Type I error probabilities greater than the significance level stipulated at the beginning of the experiment, which can have serious consequences for decision making regarding the suitability of VaR. As a new proposal for backtesting based on sequential analysis, an alpha spending function was built, capable of controlling a real type I error probability of the experiment, associated with exact critical values. In addition, the new backtesting has a high power to reject a VaR model in a context where the relative risk is much greater than the risk specified by the null hypothesis.

Key words: Value-at-Risk, backtesting, type I error probability, sequential analysis, control charts, alpha spending.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – Gráfico de controle convencional.....	44
FIGURA 2 – Gráfico de controle <i>bootstrap</i>	49
FIGURA 3 – “ <i>Gasto de alpha</i> ” para cada <i>backtesting</i> realizado	73
FIGURA 4 – Novos valores críticos	74

LISTA DE TABELAS

TABELA 1 – Zona dos <i>backtests</i> baseados no número de exceções	25
TABELA 2 – Erro do Tipo I e Tipo II no QCRM	48
TABELA 3 – Tamanho dos testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$	67
TABELA 4 – Valores críticos estimados, na escala do valor-p, para os testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$	68
TABELA 5 – Poderes estimados para os testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$	69
TABELA 6 – Medida de desempenho do teste sequencial com função “ <i>gasto de alpha</i> ” do tipo potência e $\rho = 0,5$	75

LISTA DE QUADROS

QUADRO 1 – Revisão de literatura dos principais métodos do VaR e <i>backtesting</i> empregados nos estudos empíricos	27
--	----

SUMÁRIO

INTRODUÇÃO.....	14
CAPÍTULO 1 - VaR E <i>BACKTESTING</i> : REVISÃO DE LITERATURA E PRINCIPAIS DEFINIÇÕES.....	17
1.1 VaR como medida de risco: Revisão de literatura.....	18
1.2 <i>Backtesting</i> : Revisão de literatura e principais testes.....	25
1.2.1 Teste de Kupiec.....	29
1.2.2 Teste de Christoffersen.....	31
1.2.3 Duration.....	33
1.2.4 Quantil Dinâmico (DQ).....	35
1.2.5 <i>Backtesting</i> via regressão quantílica.....	37
CAPÍTULO 2 - CONTROLE ESTATÍSTICO DE QUALIDADE VERSUS ANÁLISE ESTATÍSTICA SEQUENCIAL.....	41
2.1 Controle Estatístico de Qualidade.....	42
2.1.1 Gráficos de controle de Shewhart.....	42
2.1.2 Gráficos de controle CUSUM.....	45
2.1.3 Gráficos de controle EWMA.....	46
2.1.4 Aplicação dos gráficos de controle para o VaR.....	47
2.2 Análise Sequencial.....	51
2.2.1 Teste da Razão de Probabilidade Maximizado (MaxSPRT) para dados Binomiais.....	52
2.2.2 Função “ <i>gasto de alpha</i> ”.....	54
2.2.3 Medidas de desempenho: Função Poder, tempo médio de sinalização e tempo esperado de vigilância.....	55
2.2.4 Aplicações da análise sequencial na economia.....	57
CAPÍTULO 3 - O PROBLEMA DOS TESTES MÚLTIPLOS: DEMONSTRAÇÃO TEÓRICA E ESTUDO DE SIMULAÇÕES.....	60
3.1 O problema dos testes múltiplos para o <i>backtesting</i>	61

3.1.1 Exemplo prático de múltiplos testes para o <i>backtesting</i>	61
3.2 Estudo de simulação.....	63
3.2.1 Procedimentos metodológicos.....	63
3.2.2 Resultados e discussões.....	67
CAPÍTULO 4 - “ <i>GASTO DE ALPHA</i> ”: UMA NOVA ABORDAGEM PARA O <i>BACKTESTING</i>	71
4.1 Consolidando a função no <i>software R</i>	71
4.1.1 Resultados e discussões.....	73
CONSIDERAÇÕES FINAIS	77
REFERÊNCIAS BIBLIOGRÁFICAS	80

INTRODUÇÃO

O risco de mercado é definido como aquele que está inerente às mudanças nas condições de mercado e que podem impactar negativamente o lucro de uma instituição ou até mesmo o valor de seus ativos e passivos. Os fatores de riscos associados às operações no mercado financeiro podem ser a taxa de juros, o risco cambial, risco de inadimplência, risco de *spread*, risco de ações e risco envolvendo preços de *commodities*.

A criação do conglomerado de normas internacionais, formado pelo Comitê de Basileia, foi um importante marco regulatório no que tange a regulação bancária. Seu escopo trata sobre os requisitos mínimos de capital para cobrir possíveis riscos associados ao risco de crédito, risco operacional, risco de liquidez e ao risco de mercado. A partir desse marco, todas as instituições bancárias inseridas no conglomerado prudencial devem seguir as recomendações do Comitê de Basileia para se precaverem de possíveis desequilíbrios no sistema financeiro.

No que tange ao risco de mercado, o Comitê de Basileia recomenda a utilização do cálculo do *Value-at-Risk* (VaR) como medida de risco. O VaR mede a pior perda possível de um portfólio em um determinado período, associado a um nível de significância. Há uma vasta literatura dedicada a estudar diferentes metodologias de cálculo para o VaR (ABAD; BENITO; LÓPEZ, 2014; BALI; MO; TANG, 2008; BERTOWITZ; CHRISTOFFERSEN; PELLETIER, 2011; BROWNLEES; GALLO, 2010; DIAS, 2013; ENGLE; MANGANELLI, 2004; NIETO; RUIZ, 2016).

A fim de verificar a adequabilidade do VaR, o Comitê de Basileia também exige que as instituições adotem um teste de aderência, denominado de *backtesting*, cuja premissa baseia-se em contar quantas vezes o retorno de um portfólio ultrapassou a estimativa do VaR e classificar essas exceções de acordo com cores de sinalização. Nieto e Ruiz (2016) argumentam que uma subestimação do modelo VaR pode levar a instituição a adotar níveis de capital insuficientes para cobrir possíveis perdas financeiras, aumentando assim o risco de falência. No entanto, uma superestimação do modelo de risco afeta a lucratividade da instituição, devido ao alto nível de capital empregado erroneamente.

De forma a contribuir para a construção de *backtests* mais precisos, a literatura se encarregou em investigar metodologias mais robustas de *backtesting*, a fim de analisar outras propriedades além das recomendadas pelo Comitê de Basileia, desenvolvendo *backtests* mais poderosos em detectar qualquer indício de um modelo de risco mal especificado.

Alguns *backtests* podem ser destacados: o teste de Kupiec (1998); teste de Christoffersen (1998); teste de *duration* proposto por Christoffersen e Pelletier (2004); quantil dinâmico elaborado por Engle e Manganelli (2004) e teste de regressão quantílica proposto por Gaglianone *et al.* (2011). Esses *backtests* analisam diferentes propriedades de cobertura, como: hipótese de independência das exceções, duração entre as exceções observadas e hipótese de correlação entre as exceções.

O *backtesting* pode ser encarado como um monitoramento de processos, uma vez que seu objetivo é monitorar um VaR já especificado e detectar rapidamente qualquer mudança que ocorra na estimação do modelo de risco. Pode-se destacar duas áreas importantes no monitoramento de processos: o controle estatístico de qualidade e análise estatística sequencial.

No controle estatístico de qualidade, o objetivo é monitorar um processo que esteja sob controle e detectar de maneira eficaz qualquer alteração nos parâmetros que geram o processo, como a média e a variância. Em contrapartida, a análise estatística sequencial deseja verificar se os parâmetros intrínsecos do processo, isto é, a média e a variância, são verdadeiros desde o início do monitoramento, sem passar pela terminologia de “controle”. No que tange as semelhanças entre as duas áreas, destaca-se a utilização de limites inferiores e superiores para verificar quando os parâmetros do processo ultrapassam esses limiares. Como medidas de desempenho, tanto os gráficos de controle, como a análise estatística sequencial empregam o tempo médio amostral até que o monitoramento é interrompido.

Pena, Rivera e Mata (2006) arquitetam um *backtesting* a partir da abordagem dos gráficos de controle, cuja formulação do teste de hipóteses tem o objetivo de minimizar a probabilidade do erro do Tipo II. Dessa forma, os autores acreditam que pode haver um equilíbrio entre os erros do Tipo I e Tipo II, bem como na definição da região de rejeição do teste.

Uma abordagem mais apropriada para o *backtesting* e que ainda não foi proposta na literatura é a partir da análise estatística sequencial (WALD, 1945; JENNISON; TURNBULL, 2000). Um teste construído sobre as premissas da análise sequencial leva em consideração a acumulação progressiva dos dados ao longo do tempo, de forma a se obter respostas rápidas, mas ainda que precisas, para o monitoramento contínuo de processos. Assim, é utilizado um nível de significância global, considerando conjuntamente os níveis de significância de cada teste, assegurando, assim, um real controle da probabilidade do erro do Tipo I global sob o nível α desejado.

Pelo fato de que os *backtests* propostos na literatura não levam em consideração a problemática dos testes múltiplos, ou seja, que não consideram o fato dos dados se acumularem

ao longo do tempo, estes não são testes confiáveis para verificação da aderência do VaR. Portanto, a realização de sucessivos testes, sem levar em consideração os testes realizados anteriormente, faz inflar a probabilidade do erro do Tipo I, isto é, o nível de significância global da análise é maior que o pretendido. Em um teste de hipóteses, quando o tamanho do teste é superior ao nível de significância desejado, o teste não é válido e, portanto, não é de nível α .

Em face do exposto, essa dissertação tem como objetivo principal propor uma mudança de paradigma na construção dos testes de hipóteses que fundamentam os *backtests*. Mais especificamente, este trabalho visa (I) apresentar a fundamentação dos testes de hipóteses dos principais *backtests* discutidos pela literatura; (II) elaborar um paralelo entre os gráficos de controle e a análise sequencial de modo a traçar sob qual perspectiva o *backtesting* poderia ser estruturado; e (III) propor uma abordagem para um novo *backtesting* a partir da análise estatística sequencial, apresentando uma função “*gasto de alpha*” capaz de controlar adequadamente a inflação da probabilidade do erro do Tipo I.

Essa dissertação está dividida em quatro capítulos além dessa introdução. O Capítulo 1 apresenta uma revisão de literatura sobre as metodologias de cálculo do VaR e os principais *backtests* disponíveis. O Capítulo 2 traz uma comparação entre as abordagens dos gráficos de controle e da análise estatística sequencial. Isto é necessário para que a mudança de paradigma proposta seja justificada. O Capítulo 3 elucida o problema dos testes múltiplos quando não se considera que os dados se acumulam no tempo. Adicionalmente, um estudo de simulação é apresentado a fim de investigar o comportamento de um *backtesting* realizado sucessivas vezes ao longo do tempo. Por fim, o Capítulo 4 se encarrega de propor uma nova abordagem para o *backtesting* a partir da função “*gasto de alpha*”, onde será possível controlar a real probabilidade do erro do Tipo I associada aos novos valores críticos, bem como frisar a importância de se fornecer cruciais medidas de desempenho tais como o tempo médio de sinalização e tempo esperado de vigilância.

CAPÍTULO 1 - VAR E BACKTESTING: REVISÃO DE LITERATURA E PRINCIPAIS DEFINIÇÕES

A instabilidade no mercado financeiro experimentada no final da década de 1980 incentivou pesquisadores, *policymakers*¹ e órgãos regulatórios a lançarem um novo olhar sobre os instrumentos de mensuração de risco. Crises financeiras podem ser desencadeadas pela ineficiência do gerenciamento de riscos, principalmente no que tange às metodologias utilizadas para o cálculo desses riscos. (KING; WADHWANI, 1990; ENGLE; MANGANELLI, 2004; CHEN *et al.*, 2011; BILIR, 2016).

Risco configura-se como um elemento de incerteza. Uma maneira de quantificar o risco é utilizando-se de uma distribuição de probabilidade a qual pode ser expressa em termos de valores de capital (EMMER, KRATZ; TASCHE, 2015). Engle e Manganelli (2004) ressaltam a importância das medidas de risco no que se refere à tomada de decisões de supervisores, decisões sobre investimentos, alocação de capital e adequação às normas regulatórias internacionais.

Bancos comerciais são os principais agentes nas negociações no mercado financeiro (BERKOWITZ; O'BRIEN, 2002). Isso se deve ao fato da ampla carteira de serviços oferecidos ao mercado, como operações de *swaps* cambiais, negociações incluindo ações, debêntures e outros tipos de derivativos. Consequentemente, essas instituições estão mais expostas ao risco e às variações no mercado (BIS, 2017).

O Comitê de Basileia de Supervisão Bancária, ou simplesmente Comitê de Basileia, foi um marco regulatório no que concerne a regulamentação do sistema bancário mundial. A partir desse marco foi possível discutir e consolidar regras quantitativas e qualitativas a respeito dos requisitos mínimos de capital para os diferentes tipos de riscos associados às operações das instituições bancárias, inaugurando assim o *Basel I*² em 1988.

Para além dos riscos associados às exigências de capital, o *Basel II* se empenha em fornecer orientações acerca dos modelos internos de risco do banco, incorporando o cálculo do VaR como base para medir suas necessidades de capital de risco de mercado (BIS, 2001;

¹ *Policymakers*: pessoas que atuam nas decisões e formações de políticas internacionais.

² O *Basel I* (<https://www.bis.org/publ/bcbs04a.pdf>) foi o primeiro conjunto de normas arquitetado pelo Comitê de Basileia em 1988, cuja principal finalidade era a implementação de uma estrutura padrão mínima de capital de 8% até o fim de 1992. Foi só no *Basel II* (<https://www.bis.org/publ/bcbs107.pdf>), em 1996, que foi ampliado as pautas referentes não só aos riscos de capital, mas também ao risco de crédito e ao risco de mercado, recomendando que os bancos deveriam incorporar o cálculo do VaR para medir suas necessidades de capital de risco de mercado. Dentre outras medidas de risco, o *Basel III* (<http://www.bis.org/publ/bcbs193.pdf>), estabelecido em 2011, também prevê a estimativa do VaR, bem como seu *backtesting* (SHARMA, 2012).

CANDELON *et al.* 2011; ESCANCIANO; PEI, 2012). É importante destacar que as medidas para mensuração do risco de mercado poderiam ser aplicadas às instituições bancárias e não bancárias, desde que todas estejam expostas a esse tipo de risco (BIS, 2001).

Aliado ao cálculo do VaR, o *Basel II* preconiza a utilização de *backtests*, o qual basicamente compara os valores obtidos pela estimação do VaR com as alterações reais ocorridas em um portfólio. Nieto e Ruiz (2016) reforçam essa ideia argumentando que a validação da estimativa do VaR é um passo importante para o gerenciamento de riscos, de modo que as instituições possam adequar seu capital frente a possíveis perdas financeiras.

Visto isso, este capítulo aborda a literatura acerca do VaR e do *backtesting*, apresentando principais definições e estudos empíricos sobre o tema, bem como introduz conceitos matemáticos elementares para o cálculo do *backtesting*.

1.1 VaR como medida de risco: Revisão de literatura

O *RiskMetrics*³ (1996) conceitua o VaR como uma ferramenta a qual se pode mensurar a máxima alteração potencial no valor de uma carteira de ativos financeiros, dado um horizonte de risco e coeficiente de confiança pré-definidos. Huang (2010) corrobora a definição argumentando que o VaR é a quantia que uma carteira de investimentos pode perder de acordo com um limiar de probabilidade escolhido em um determinado período de tempo. O autor ainda acrescenta que, dentro de uma definição estatística, o VaR corresponde a um quantil da distribuição do retorno do ativo e/ou portfólio cuja probabilidade cumulativa subjacente é igual a α .

O VaR como medida de risco é atrativa dentre vários fatores podendo ser aplicado a qualquer tipo de atividade e a todos os tipos de risco; mede o fator de risco, bem como suas sensibilidades e pode ser comparado entre diferentes mercados (ALEXANDER, 2008). Outros fatores favoráveis que atribuem uma popularidade ao VaR são sua simplicidade e o fato de que o VaR diminui o risco de mercado atrelado a qualquer carteira, traduzido em apenas um valor monetário (ENGLE; MANGANELLI, 2004).

³ O *RiskMetrics* é um documento técnico elaborado pelo grupo americano J.P Morgan com o propósito de consolidar uma metodologia para a mensuração de risco de mercado, com sua primeira versão lançada em 1994. O documento disponibiliza informações gerais sobre os riscos de mercado, como esses devem ser estimados, bem como a interpretação dos resultados obtidos. Apresenta também técnicas estatísticas para analisar o comportamento dos retornos dos ativos, para assim, encontrar o método mais adequado para a mensuração do risco.

Os parâmetros básicos para a estimativa do VaR são o nível de significância do teste α ou de forma equivalente, o coeficiente de confiança correspondente a $1-\alpha$ e a definição do horizonte de risco, o qual representa o período de tempo medido em dias de negociações sobre os quais o VaR é estimado (JORION, 2006).

A fórmula geral do VaR pode ser definida da seguinte forma, seguindo a notação de Huang (2010):

$$VaR(P, \alpha, \delta t) = P \cdot k \cdot \sigma \sqrt{\delta t}, \text{ em que } \Pr(\delta P \geq k \cdot \sigma \cdot \sqrt{\delta t}) = \alpha \quad (1)$$

em que P representa o valor atual (medidas monetárias) do portfólio, k é o número de desvios-padrões corresponde a α , α é o nível de significância desejado, δt é o horizonte de tempo, σ representa o desvio-padrão da distribuição dos retornos do portfólio e δP é o valor absoluto da mudança do portfólio.

Existem quatro principais propriedades que fazem do VaR uma medida atrativa de risco, em um contexto que se pode considerar diversos tipos de distribuições contínuas, assim como descrevem Gaglianone *et al.* (2011) e Emmer, Kratz e Tasche (2015):

- (I) Monotonicidade: se o portfólio Z prevalecer sobre o portfólio Y , de modo que o *payoff* de Z seja tão bom quanto ao *payoff* de Y , então Y deve conter um risco menor ou igual a Z ;
- (II) Homogeneidade: a medida de risco é proporcional a escala do portfólio;
- (III) Subaditividade: a medida de risco do portfólio não pode ultrapassar a soma das medidas de risco das carteiras menores que o compõem;
- (IV) Ativo livre de risco: o risco total do investimento será diminuído se um ativo livre de risco for adicionado ao portfólio.

Nesta perspectiva, qualquer metodologia utilizada para o cálculo do VaR envolve três pressupostos básicos: (I) apresentar uma fórmula para o cálculo do VaR levando em consideração as variáveis no tempo passado, bem como os parâmetros; (II) demonstrar um procedimento que seja capaz de estimar parâmetros desconhecidos e (III) oferecer um teste que verifique a precisão do VaR estimado (ENGLE; MANGANELLI, 2004).

A literatura acerca das diferentes metodologias para se estimar o VaR é vasta, incluindo métodos paramétricos, semi-paramétricos e não paramétricos.

Os métodos paramétricos utilizam o cálculo de volatilidade dos ativos para a construção da estimativa a qual pode ser realizada por meio do modelo de média móvel exponencialmente

ponderada (EWMA) proposto pela *RiskMetrics*, como também através do modelo de heterocedasticidade condicional autorregressiva generalizada (GARCH⁴) de Bollerslev (1986).

As metodologias semi-paramétricas como a teoria do valor extremo (EVT⁵), o modelo CAViaR proposto por Engle e Manganelli (2004) e o método de simulação histórica filtrada (FSH⁶) mesclam métodos em que é possível realizar estimativas sobre um parâmetro conhecido, bem como utilizar de uma abordagem empírica para estimar a distribuição do retorno dos ativos.

Os métodos não paramétricos mais conhecidos e abrangentes na literatura para a mensuração do VaR são os de simulação histórica, simulação de Monte Carlo e o de regressão quantílica, em que a distribuição dos retornos dos ativos é deduzida empiricamente sem a necessidade de se conhecer algum parâmetro.

Embora exista uma gama de métodos para se estimar um VaR é importante destacar que o método de simulação histórica e o FSH são as abordagens mais utilizadas, cerca de 75% dos bancos comerciais nos EUA, Canadá e Europa utilizam esse método (PÉRIGNON; SMITH, 2010 *apud* GAGLIANONE *et al.*, 2011; ESCANCIANO; PEI, 2012; NIETO; RUIZ, 2016).

O modelo de simulação histórica para cálculo do VaR tem como hipótese principal a de que todas as possíveis variações futuras de um portfólio foram experimentadas no passado e que a distribuição histórica simulada é idêntica à distribuição de retornos sobre o horizonte de risco do futuro. Ainda mais, essa abordagem não exige o conhecimento da distribuição dos retornos dos ativos, configurando uma vantagem ao método. Além disso, o modelo de simulação histórica pode ser aplicado a carteiras lineares e não-lineares, mesmo que esse último caso se utilize de cálculos mais complexos, como é o caso de uma carteira de opções (ALEXANDER, 2008).

Em vista as diferentes formas de cálculo de VaR, surgem na literatura diferentes estudos realizando avaliações empíricas, tais como o de Berkowitz e O'Brien (2002) e Berkowitz, Christoffersen e Pelletier (2011), os quais estimam um VaR diário com coeficiente de confiança de 99%, através da análise de uma distribuição de uma carteira histórica de perdas e ganhos

⁴ GARCH: *Generalized autoregressive conditional heteroskedasticity*. A sigla em português é traduzida como modelo de heterocedasticidade condicional autorregressiva generalizada.

⁵ *Extreme Value Theory* (EVT): esse método utiliza a família paramétrica para descrever a cauda da distribuição, e de maneira similar, os valores centrais da distribuição dos retornos são usados para estimar empiricamente a função de distribuição. Ver mais em Longin (1996) e Embrechts (2009).

⁶ FHS (Filtered Historical Simulation): Esse método utiliza a simulação histórica para simular possíveis cenários e diferentes preços dos ativos, incorporando também modelos de volatilidade como um GARCH (1,1), minimizando o problema de heterocedasticidade dos retornos dos ativos. Melhor dizendo, é um método que combina estimação não-paramétrica (simulação histórica) e estimação paramétrica (GARCH(1,1)). Ver mais em Pritsker (2006).

(P&L⁷) de grandes bancos americanos, e, de modo geral, as previsões do VaR ultrapassaram o nível de nominal de cobertura.

Assim como fizeram Berkowitz e O'Brien (2002) e Berkowitz, Christoffersen e Pelletier (2011), Pérignon, Deng e Wang (2008) investigam se os seis maiores bancos canadenses superestimam o cálculo do VaR. Para isso, os autores utilizam do método de simulação histórica para estimar um VaR para um nível de significância de 1%. Confirmam, assim, sua hipótese alegando que os bancos comerciais são extremamente conservadores ao definir o VaR. Para exemplificar, no período analisado de 7354 dias, existem apenas 2 valores que ultrapassam o nível de significância, enquanto era de se esperar 74 exceções para um VaR com essa janela temporal.

Por sua vez, Escanciano e Pei (2012) utilizam do método de simulação histórica e FHS para estimar um VaR (níveis de significância de 1% e 5%) para as ações da *Walt Disney*, *General Electric* e *Merck & Company*, todas negociadas na *New York Stock Exchange*. Nesse estudo, os autores esperavam que a estimativa do VaR via FHS seria melhor do que a de simulação histórica, na premissa que o FHS leva em consideração a volatilidade dos ativos, porém foi constatado que ambas metodologias apresentam falhas na previsão.

Apesar do método de simulação histórica e FHS serem as abordagens mais utilizadas para se calcular o VaR, a literatura também se encarrega em explorar os métodos semi-paramétricos. Autores como Chan e Gray (2006) utilizam o EVT comparado ao método de simulação histórica para estimação de um VaR a níveis de significância de 0,5%, 1% e 5% para o mercado de eletricidade dos EUA, Nova Zelândia, Canadá, Escandinávia e Austrália. O mercado de eletricidade é muito peculiar e isso resulta em distribuições empíricas dos retornos com uma forma não padronizada, sendo assim, não é recomendado a utilização de métodos como o de simulação histórica para estimar a distribuição de retorno futura. Dessa forma, os autores combinam modelos de volatilidade ao EVT e chegam no AR-EGARCH-EVT⁸, apresentando a melhor previsão do VaR.

Utilizando de ativos menos peculiares do que o mercado de eletricidade, mas ainda tratando-se do mercado americano, Dias (2013) investiga a performance do VaR para portfólios com diferentes capitalizações de mercado, envolvendo ações da *New York Stock Exchange* (NYSE), *American Stock Exchange* (AMEX) e *National Association of Securities Dealers Automated Quotations* (NASDAQ). A autora utiliza do método de simulação histórica e FSH,

⁷ P&L: sigla para o termo *profit e loss*.

⁸AR: Modelo autorregressivo. EGARCH: Modelo de heterocedasticidade condicional autorregressiva generalizada exponencial. Ver mais em Nelson (1991).

bem como o EVT e suas variantes para estimar um VaR com grau de confiança de 99%. Surpreendentemente, os resultados apontam que o método de simulação histórica e FHS foram os que apresentaram melhor desempenho. Nesse caso, o método EVT não foi tão preciso devido ao grande volume de ações para a estimação do VaR, e a metodologia do EVT pode ser mais eficaz quando empregada em portfólios individuais, quando a distribuição dos retornos dos ativos não é tão discrepante.

Dentro da classe paramétrica para estimação do VaR, os modelos de volatilidade como o GARCH e suas variações são os métodos paramétricos mais populares, como descrevem Escanciano e Olmo (2010). Em um estudo utilizando o índice *Dow–Jones Industrial Average*⁹, Escanciano e Olmo (2011) mensuram um VaR com níveis de significância de 1%, 5% e 10%, utilizando-se de quatro diferentes metodologias: um GARCH (1,1) com inovações gaussianas; modelo de simulação histórica; o modelo FHS e a metodologia *RiskMetrics* empregando uma suavização exponencial do processo de volatilidade. Os resultados mostram que o método de simulação histórica obteve o pior desempenho para os três níveis de significância. A metodologia *RiskMetrics* apresentou melhor desempenho, considerando níveis de significância de 5% e 10%. Os modelos GARCH (1,1) e o FHS tiveram desempenho inferior para níveis a 1% e 5% se comparado ao nível de 10%.

Quando o VaR é estimado através de modelos de volatilidade como o GARCH, é relevante atentar para alguns “fatos estilizados” envolvendo séries temporais financeiras. Os fatos estilizados estão associados a alguns aspectos estatísticos empíricos como a heterocedasticidade condicional, distribuições do tipo “cauda pesada”, correlação positiva entre os retornos quadrados e absolutos e efeitos de agrupamento (GRIGOLETTO; LISSI, 2009; ERGEN 2015).

Levando em consideração esse contexto, Glasserman, Heidelberger e Shahabuddin (2002), Chan *et al.* (2007), Danielsson *et al.* (2013) e Ergen (2015) concentram seus estudos empíricos na investigação do fato estilizado envolvendo a distribuição de “cauda pesada” do retorno dos ativos para a mensuração do VaR. Ergen (2015) enfatiza que se faz necessário uma certa cautela às distribuições do tipo “cauda pesada” cuja probabilidade de perdas financeiras é muito maior do que as distribuições normais as quais não conseguem captar as assimetrias dos retornos dos ativos.

Há de se considerar ainda modelos de volatilidade que vão além dos comumente usados como o GARCH. Brownless e Gallo (2010), por exemplo, utilizam medidas de volatilidade

⁹ Esse índice é baseado na cotação das 30 maiores empresas dos Estados Unidos da América (EUA).

baseadas em dados de ultra alta frequência (UHFD¹⁰) para estimar um VaR para quatro ações da NYSE, sendo elas: *Boeing* (BA), *General Electric* (GE), *Johnson & Johnson* (JNJ) e *J. P. Morgan* (JPM). Os autores ressaltam que o período analisado é caracterizado por alta concentração de volatilidades em torno de uma volatilidade média. Os resultados empíricos demonstram que o UHFD possui melhor precisão se comparado aos modelos de simulação histórica, GARCH e *RiskMetrics*.

Por outro lado, Grigoletto e Lissi (2011) mensuram um VaR com nível de significância de 1% baseado no modelo GARCHDSK¹¹ para os oito maiores índices de ações internacionais: *Dow Jones*, S&P 500, *NYSE Composite*, CAC40, DAX, FTSE100, Nikkei225 e *Hangsen*. Os resultados indicam que o modelo GARCHDSK foi o que apresentou melhor desempenho em comparação ao GARCH-N (distribuição normal) e ao *RiskMetrics*; os modelos GARCH-t (distribuição *t-student*) e GARCHSK apresentaram desempenho semelhante ao GARCHDSK.

Autores como Chen *et al.* (2011) utilizam de uma abordagem bayesiana para mensurar o risco de ações de quatro importantes índices da Cooperação Econômica Ásia-Pacífico (APEC) sendo: Nikkei225, HANG SENG, *Korea Composite* (KOSPI) e o S&P500, considerando uma janela temporal de 1997 a 2009, divididos em períodos pré-crise de 2008 e pós-crise.

O método de Monte Carlo via cadeia de Markov (MCMC¹²) é empregado por Chen *et al.* (2011) para estimar o VaR e compará-lo aos modelos GARCH e suas seis variações (IGARCH, GARCH, TGARCH, GJR-GARCH, EGARCH e MS-GARCH) à metodologia *RiskMetrics* e aos modelos de volatilidade estocástica. Os resultados mostram que os modelos GARCH apresentaram melhor desempenho com relação aos modelos de volatilidade estocástica na maioria dos cenários como também demonstraram desempenho satisfatório ao nível de 5% de significância pós crise, enquanto os modelos de volatilidade estocástica assimétrica foram mais precisos no período pré-crise e, por fim, foi constatado que a previsão do VaR no pós-crise foi menos precisa e anti-conservadora.

Uma outra abordagem para previsão do VaR é o uso de modelos de quantil condicional dinâmico (KOENKER; XIAO, 2006), seguindo a premissa de que o VaR é caracterizado por

¹⁰ UHFD: Os dados de alta ultra frequência são os preços intra-diários de uma ação, considerando o número de segundos em um dia de negociação, iniciando a contagem no momento da abertura do mercado até o seu fechamento.

¹¹ GARCH with Dynamic Skewness and Kurtosis (GARCH com assimetria dinâmica e curtose). Esse modelo é capaz de descrever comportamentos assimétricos em uma série de tempo, permitindo a avaliação de possíveis dinâmicas em momentos de ordem superior. Ver mais em Grigoletto e Lissi (2009).

¹² Segundo Chen *et al.* (2011), o método de Monte Carlo via cadeia de Markov numa perspectiva bayesiana apresenta algumas vantagens como a eficiência na previsão do VaR de modo a realizar inferências exatas para amostras finitas; capta a incerteza dos parâmetros na previsão de probabilidade; mostra-se eficaz no tratamento de parâmetros não padronizados e de modelos complexos.

um quantil condicional da distribuição dos retornos de um portfólio (HUANG, 2010) e que a série financeira varia no tempo seguindo a dinâmica do retorno dos ativos (GOUREIOUX; KASIAK, 2008).

Seguindo essa premissa, o modelo de quantil condicional dinâmico amplamente empregado e discutido na literatura é o proposto por Engle e Manganelli (2004): o modelo autorregressivo condicional de valores em risco (CAViaR¹³), o qual especifica a evolução do quantil ao longo do tempo usando um processo autorregressivo para estimar o VaR. As volatilidades do retorno dos ativos se aglomeram ao longo do tempo, isto é, estatisticamente falando, a distribuição dos retornos é autocorrelacionada. Assim sendo, o VaR que está intrinsecamente relacionado ao desvio padrão da distribuição, possui comportamento semelhante, e, portanto, necessita-se uma especificação autorregressiva.

O estudo empírico de Engle e Manganelli (2004) estima o CAViaR para séries históricas da *General Motors* (GM), IBM e S&P500 considerando níveis de significância de 1% e 5%. Os resultados obtidos corroboram com a premissa de que há uma aglomeração das volatilidades nas caudas da distribuição e que o modelo estimado apresenta uma precisão significativa ao nível de 1% para todas as três séries.

Outra classe de modelos envolvendo quantis dinâmicos é o proposto por Goureloux e Kasiak (2008), o qual mensura o VaR por meio de um quantil aditivo dinâmico (DAQ¹⁴) capaz de garantir a monotonicidade das funções quantis condicionais. O estudo empírico mensura o VaR por meio do DAQ utilizando o índice do mercado da Bolsa de Toronto, apontando assim que o modelo não é rejeitado pelos testes de especificação.

Assim como Escanciano e Olmo (2011) que comparam várias metodologias para a estimação do VaR, Boucher *et al.* (2014) também analisam empiricamente o Índice *Dow Jones Industrial Average* e concluem que a estimação do VaR por meio de modelos dinâmicos como o CAViaR, GARCH (1,1) e EWMA se sobressaíram ao modelo Gaussiano e aos métodos usando densidades extremas. Os autores ressaltam ainda, que, surpreendentemente, o método de simulação histórica fica muito próximo de ser a melhor abordagem para previsão do VaR.

A partir dos estudos empíricos apresentados, é possível constatar que existem várias metodologias para a mensuração do VaR, levando-se em consideração abordagens paramétricas, semi-paramétricas e não paramétricas. De modo similar, a próxima seção demonstra os principais métodos de cálculo para a *backtesting*.

¹³ CAViaR: Conditional Autoregressive Value-at-Risk. Para estimar os parâmetros desse modelo é utilizada uma regressão quantílica arquitetada por Koenker e Bassett (1978).

¹⁴ DAQ: Dynamic Additive Quantile.

1.2 *Backtesting*: Revisão de literatura e principais testes

O teste de aderência para adequação do VaR é chamado de *backtesting*. O BIS (1996) conceitua o *backtesting* como estratégia adotada pelos bancos para medir o quão eficientes são os métodos de mensuração de risco. Com o acompanhamento diário das perdas e ganhos com medidas de risco estimadas por meio do VaR, o *backtesting* avalia o poder e a precisão da estratégia de mensuração de risco de um banco. Isto é, o *backtesting* consiste em comparar os resultados reais de negociação, aferidos em termos de ganhos ou perdas financeiras, com os valores estimados do VaR. Assim, um *backtesting* confiável deverá detectar, rapidamente, a ineficiência do modelo usado para mensuração de perdas/ganhos atípicos.

O BIS¹⁵ (2019) sugere que o *backtesting* do modelo de risco de todo banco deve ser calculado com base no número de exceções ao longo de 12 meses (ou 250 dias úteis), baseado em um VaR com um coeficiente de confiança de 99%. O teste se baseia simplesmente em calcular o número de vezes que os valores negociados não são cobertos pela medida de risco, ou seja, basta contabilizar o número de vezes que o valor observado “excede” o valor do VaR e verificar se as exceções estão dentro do intervalo de confiança de 99%.

A metodologia proposta pelo BIS (2019) para a aplicação do *backtesting* é categorizar o número de exceções em três zonas distintas: verde, âmbar e vermelha, considerando seus erros estatísticos.

Tabela 1- Zonas do *backtesting* baseado no número de exceções

Zona do <i>backtesting</i>	Número de exceções	Multiplicador dependente do <i>backtesting</i>	Probabilidade cumulativa
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%
Âmbar	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%

Fonte: BIS (2019)

¹⁵ O documento mais recente do Comitê de Basileia com as recomendações da aplicação do *backtesting* é o *Minimum capital requirements for market risk* (2019).

No que diz respeito à Tabela 1, a primeira coluna relaciona-se às zonas do *backtesting*: a zona verde corresponde a uma área em que os resultados do *backtesting* não apresentam problemas relacionados à precisão do modelo de risco; a zona âmbar relaciona os resultados do teste que são inconclusivos e a zona vermelha indica que as informações obtidas pelo *backtesting* possivelmente apontam problemas na mensuração do VaR.

A segunda coluna trata-se do número de vezes em que as variações reais de um portfólio ultrapassam a estimativa do VaR, levando-se em consideração o nível de significância e horizonte de tempo propostos.

A terceira coluna refere-se ao fator de multiplicação para exigência de capital a qual o banco está sujeito, de acordo com o número de exceções obtidos no *backtesting*. Para cada fator de risco, o Comitê de Basileia estipula um adicional de capital que as instituições financeiras estão sujeitas para manterem provisionados a fim de se resguardarem de possíveis crises financeiras.

Como o Comitê de Basileia acredita que o *backtesting* não alcança um nível de 99% de cobertura, os valores do fator de multiplicação devem ser suficientes para retornar o modelo a um padrão de 99%. Além de uma exigência de capital mais alta para resultados dentro da zona amarela ou vermelha do *backtesting*, a autoridade supervisora pode suspender a utilização do modelo de capital de risco de mercado do banco (BIS, 2019).

Segundo o BIS (2019), as três zonas foram assim definidas a fim de balancear dois tipos de erros estatísticos, sendo eles: o erro do tipo I em que o modelo de risco preciso é classificado como impreciso com base em seu *backtesting* e o erro do tipo II, no qual o modelo impreciso não seja classificado como tal no resultado do seu *backtesting*.

O Comitê de Basileia ainda identifica algumas limitações estatísticas do *backtesting*, como, por exemplo, o fato de não haver um número limite de exceções que estime uma probabilidade baixa de se cometer o erro do Tipo I e o erro do Tipo II. Em contrapartida, os resultados a serem apresentados no Capítulo 4 deste trabalho, demonstrarão que na nova proposta de *backtesting* existem valores críticos exatos capazes de controlar a real probabilidade do erro do Tipo I.

A abordagem do *backtesting* proposto pelo Comitê de Basileia é baseada simplesmente em calcular o número de vezes em que as perdas financeiras de um portfólio ultrapassam as estimativas previstas pelo VaR. Diante disso, a literatura se encarrega em estudar métodos mais sofisticados e mais eficazes para a realização do *backtesting*.

As metodologias mais utilizadas para o cálculo do *backtesting* são: o teste de cobertura incondicional (KUPIEC, 1995), o teste de cobertura condicional e independência

(CHRISTOFFERSEN, 1998), o método de *duration* (CHRISTOFFERSEN; PELLETIER, 2004), quantil dinâmico (ENGLE; MANGANELLI, 2004) e método via regressão quantílica (GAGLIANONE *et al.*, 2011).

O Quadro 1 apresenta os principais estudos empíricos acerca das abordagens de cálculo para o VaR e o *backtesting*:

Quadro 1 - Revisão de literatura dos principais métodos do VaR e *backtesting* empregados em estudos empíricos.

	Autores	Value-at-Risk	Backtesting
1.	Berkowitz e O'Brien (2002)	- Simulação Histórica - GARCH	- Kupiec (1995) - Christoffersen (1998)
2.	Berkowitz, Christoffersen e Pelletier (2011)	- Simulação Histórica - GARCH - CAViaR	- Kupiec (1995) - Christoffersen (1998) - Duration de Christoffersen e Pelletier (2004)
3.	Dias (2013)	- Simulação Histórica - EVT	- Kupiec (1995) - Christoffersen (1998) - Metodologia proposta pelo <i>Basel II</i>
4.	Pérignon, Deng e Wang (2008)	- Simulação Histórica - GARCH - FHS - Metodologia <i>RiskMetrics</i>	- Kupiec (1995) - Abordagem de avaliação para previsão de densidade de Berkowitz (2001) - Combinação dos testes de Kupiec (1995) e Berkowitz (2001)
5.	Escanciano e Pei (2012)	- Simulação Histórica - FHS	- Kupiec (1995) - Christoffersen (1998) - Data Driven (D- test): combina um <i>backtesting</i> ponderado não paramétrico com o teste de Christoffersen (1998)
6.	Campbell (2007)	- Simulação Histórica - GARCH - EWMA	- Kupiec (1995) - Christoffersen (1998) - <i>Duration</i> de Christoffersen e Pelletier (2004) - Combinação dos testes de Christoffersen (1998) e Duration de Christoffersen e Pelletier (2004)
7.	Escanciano e Olmo (2010)	- GARCH - ARMA - Simulações de Monte Carlo	- Kupiec (1995) - Christoffersen (1998) - Teste LB de Bertowitz, Christoffersen e Pelletier (2011)

8.	Escanciano e Olmo (2011)	- Simulação Histórica - GARCH com inovações gaussianas - FHS - Metodologia <i>RiskMetrics</i>	- Kupiec (1995) - Christoffersen (1998) - Teste LB de Bertowitz, Christoffersen e Pelletier (2011)
9.	Brownlees e Gallo (2010)	- Simulação Histórica - GARCH - Metodologia <i>RiskMetrics</i>	- Christoffersen (1998) - Quantil Dinâmico, Engle e Manganelli (2004)
10.	Chan e Gray (2006)	- Simulação Histórica - GARCH e suas variações - EVT e suas variações	- Christoffersen (1998)
11.	Chen <i>et al.</i> (2011)	- GARCH - Metodologia <i>RiskMetrics</i> - Modelos de volatilidade estocástica - Simulações de Monte Carlo via cadeia de Markov (MCMC)	- Kupiec (1995) - Christoffersen (1998)
12.	Pérignon e Smith (2008)	- Simulação Histórica - GARCH - Metodologia <i>RiskMetrics</i> - FHS - Simulação Histórica Híbrida*	- Kupiec (1995) - Generalização multivariada do teste de Kupiec (1995)
13.	Grigoletto e Lisi (2011)	- GARCH e suas variações - Metodologia <i>RiskMetrics</i>	- Kupiec (1995) - Christoffersen (1998) - Metodologia proposta pelo <i>Basel II</i>
14.	Boucher <i>et al.</i> (2014)	- Simulação Histórica - GARCH - Metodologia <i>RiskMetrics</i> - CAViaR - Cornish Fischer** - GEV e GPD***	- Kupiec (1995) - Christoffersen (1998) - Abordagem de avaliação para previsão de densidade de Berkowitz (2001) ou magnitude
15.	Halbleib e Pohlmeier (2012)	- Simulação Histórica - GARCH e suas variações - CAViaR - Metodologia <i>RiskMetrics</i> - FHS - ARMA	- Christoffersen (1998) - Metodologia proposta pelo <i>Basel II</i>
16.	Nieto e Ruiz (2016)	- Simulação Histórica - GARCH - CAViaR - EVT - FHS - <i>Bootstrap</i>	- Quantil Dinâmico, Engle e Manganelli (2004) - Christoffersen (1998) - Teste LB de Bertowitz, Christoffersen e Pelletier (2011)

Fonte: Elaborado pelo autor.

* O método HS híbrido atribui pesos de probabilidade em declínio exponencial aos retornos de negociações passadas.

** Cornish-Fisher: esse método baseia-se na estimação de quantis dos retornos dos ativos, considerando seus primeiros cumulantes. Pode ser empregado para estimar quantis de distribuições não normais. Ver mais em Favre e Galeano (2002).

*** GEV e GDP: esses métodos estimam o VaR a partir das densidades extremas para o retorno dos ativos.

1.2.1 Teste de Kupiec

O teste baseado na frequência de falhas proposto por Kupiec (1995) tem como objetivo elucidar um teste que verifica a validação dos modelos de risco pelos bancos, os quais já foram recomendados nos primeiros documentos técnicos do Comitê de Basileia (1996).

Kupiec (1995) ressalta que apesar do Comitê de Basileia trazer as novas recomendações para os modelos de risco interno, a instituição não havia se preocupado em formalizar um teste estatístico para validar o VaR. Dessa forma, Kupiec (1995) traz a proposta de elaborar um *backtesting* que usa de perdas e ganhos históricos de um portfólio para verificar a precisão da estimativa de exposição ao risco de uma instituição.

Campbell (2007) observa que o método proposto por Kupiec (1995) tem um poder estatístico baixo para amostras pequenas, e quando a distribuição das potenciais perdas do portfólio é do tipo “cauda pesada”, as estimativas do valor crítico contêm viés significativo e erros padrão substanciais, mesmo em amostras grandes.

É interessante observar que, mesmo apresentando possíveis problemas de previsão, o teste de Kupiec é o método mais utilizado para o cálculo do *backtesting*, conforme mostra o Quadro 1, devido ao fato da simplicidade do teste baseado em frequência das falhas e do baixo custo de operação.

O teste de Kupiec pode ser representado da seguinte forma: defina VaR_{t+1}^p como um número construído no dia t de modo que as perdas da carteira no dia $t + 1$ sejam maiores do que a previsão do VaR_{t+1}^p com probabilidade p . Considere uma série temporal de previsões *ex-ante* do VaR e uma série de perdas *ex-post*, representado por PL , então pode-se definir a sequência de violações do VaR como:

$$I_{t+1}(p) = \begin{cases} 1 & \text{se } PL_{t+1} > VaR_{t+1}^p \\ 0 & \text{se } PL_{t+1} < VaR_{t+1}^p \end{cases} \quad (2)$$

ou seja, a sequência de exceções retorna 1 no dia $t+1$ se a perda naquele dia for maior do que o VaR previsto para aquele dia. Caso o VaR não seja excedido, a sequência de ocorrências retorna 0. Ao realizar o *backtesting* do modelo de risco, é construído a sequência $\{I_{t+1}\}_{t=1}^T$ ao longo de T dias, indicando quando ocorreram as exceções passadas.

Sendo assim, Kupiec (1995) analisa a propriedade de cobertura incondicional da proporção de falhas. Isto é, a hipótese nula do teste é que a proporção de falhas do teste

corresponde ao nível de tolerância do VaR, e tem como hipótese alternativa que a proporção de falhas é diferente que o nível de significância do VaR, ou seja:

$$\begin{aligned} H_0: p &= p^* \\ H_1: p &\neq p^* \end{aligned} \quad (3)$$

A proporção de falhas do teste segue uma distribuição Binomial com parâmetro p e que pode ser definido como 0,01 ou 0,05, a depender do nível de significância desejado pelo analista.

Desse modo, o teste de Kupiec baseado na proporção de falhas é dado pela razão de verossimilhanças da distribuição de uma binomial:

$$LR_{un}^K = -2 \ln \left(\frac{(p^*)^x (1 - p^*)^{n-x}}{\left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x}} \right) \sim \chi_1^2 \quad (4)$$

em que p^* é a probabilidade de falha sob a hipótese nula, x é o número de falhas da amostra e n é o tamanho total da amostra. Assintoticamente, como o número de observações n tende ao infinito, a estatística de teste segue uma distribuição *qui-quadrada* com um grau de liberdade.

A região crítica para rejeição da hipótese nula será: $RC(\alpha) = \{x \in R: x \leq c\}$, onde $c = \max \{x \in R: P(LR \leq x | H_0 \text{ verdadeira}) \leq \alpha\}$ corresponde ao valor crítico do *backtesting*, x é o número de exceções observados e α é o nível de significância desejado.

Kupiec (1995) destaca ainda que seria interessante realizar um *backtesting* seguindo um esquema de monitoramento contínuo, porém é mais conveniente coletar uma amostra de resultados com distribuição do tipo Bernoulli gerados pela performance do teste e realizar um *backtesting* para um tamanho de amostra fixo.

A limitação principal do teste de Kupiec, segundo Campbell (2007), está relacionada à dificuldade do teste em detectar medidas de VaR que estão sistematicamente sob risco, uma vez que esses testes possuem baixo poder estatístico para o tamanho de amostras sugerido pelo Comitê de Basileia. Alternativamente, o novo *backtesting* proposto a partir da análise sequencial evidenciado no Capítulo 4, apresenta alto poder estatístico no contexto em que o VaR apresenta um risco relativo muito alto. Significa dizer que se os retornos do portfólio forem muito superiores ao VaR estimado, o novo *backtesting* será capaz de detectar essa subestimação na maior parte do tempo.

Estudos empíricos que utilizam como *backtesting* o teste de Kupiec como o de Escanciano e Pei (2012), demonstram que tal *backtesting* para avaliar o VaR construído com base no método de simulação histórica e FSH pode apresentar falhas na detecção de um modelo

de risco subestimado. Isto é, a probabilidade de se rejeitar a hipótese nula não converge para um, conforme orienta a equação (2). Conseqüentemente, uma falha do *backtesting* resulta em subestimação dos multiplicadores dos fatores de capital e alocação de capital de risco insuficiente.

De maneira similar, Viridi (2011) estima um VaR utilizando a matriz de variância-covariância com nível de confiança de 95% para o Índice NIFTY 50 no período de 2006 a 2007, e dentre os procedimentos de *backtesting* também adota o teste de Kupiec para verificar a precisão do modelo de risco. Como resultado, o autor observa que o teste de Kupiec rejeita todos os VaR estimados, contrapondo com o teste de Christoffersen, por exemplo, o qual é capaz de diferenciar um modelo de risco preciso e não preciso. Isso corrobora com o fato do teste de Kupiec apresentar baixo poder estatístico.

Visto as limitações do teste de Kupiec, Christoffersen (1998) propõe então complementar o teste de cobertura incondicional, propondo um teste que, além de levar em consideração as exceções observadas, também examina a propriedade de independência da série de exceções do VaR. O método será descrito mais detalhadamente na próxima seção.

1.2.2 Teste de Christoffersen

O teste de proposto por Christoffersen (1998), conhecido também como teste de cobertura condicional, busca investigar se a probabilidade de uma violação do VaR depende ou não de uma violação ocorrida no dia anterior. Ou seja, se todas as exceções observadas devem ser consideradas como independentes. A premissa básica do teste é examinar a propriedade de independência da série de exceções do VaR, utilizando-se de uma cadeia de Markov¹⁶.

Christoffersen (1998) descreve que na hipótese de que as exceções são dependentes, a sequência de violações observadas pode ser descrita por meio de uma cadeia de Markov binária de primeira ordem, com a matriz de probabilidade de transição como:

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix} \quad (5)$$

em que $\pi_{ij} = \Pr(I_t = j | I_{t-1} = i)$. π_{11} corresponde a probabilidade de que o retorno no período $t + 1$ seja uma exceção, dado que no período t ocorreu uma exceção. O π_{01} é a

¹⁶ Cadeia de Markov: são processos em que a probabilidade condicional de qualquer evento no futuro depende apenas da distribuição de probabilidade do estado atual, independente dos eventos que ocorreram no passado (GEYER, 1992).

probabilidade que o retorno no período $t + 1$ seja uma violação dado que no período t não ocorreu uma violação.

Dessa forma, a função de verossimilhança para a matriz de primeira ordem será:

$$L(\Pi_1; I_1, I_2, \dots, I_t) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}} \quad (6)$$

em que n_{ij} é o número de observações com o valor i seguido por j . O estimador do log da máxima verossimilhança será:

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00} + n_{01}} & \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{10}}{n_{10} + n_{11}} & \frac{n_{11}}{n_{10} + n_{11}} \end{bmatrix} \quad (7)$$

Depois disso, Christoffersen (1998) estima outra cadeia de Markov de primeira ordem e testa a hipótese de que as exceções são independentes, de modo que:

$$\Pi_2 = \begin{bmatrix} 1 - \pi_2 & \pi_2 \\ 1 - \pi_2 & \pi_2 \end{bmatrix} \quad (8)$$

A função de máxima verossimilhança sob a hipótese nula será:

$$L(\Pi_2; I_1, I_2, \dots, I_t) = (1 - \pi_2)^{(n_{00} + n_{10})} \pi_2^{(n_{01} + n_{11})} \quad (9)$$

e o estimador do log da máxima verossimilhança é:

$$\hat{\Pi}_2 = \frac{n_{10} + n_{11}}{n_{00} + n_{10} + n_{01} + n_{11}} \quad (10)$$

Assim como no teste de Kupiec, Christoffersen (1998) constrói o teste de razão de verossimilhanças como:

$$LR_{ind}^{Ch} = -2 \ln \left(\frac{L(\hat{\Pi}_2; I_1, I_2, \dots, I_t)}{L(\hat{\Pi}_1; I_1, I_2, \dots, I_t)} \right) \sim \chi_1^2 \quad (11)$$

De modo similar a razão de verossimilhanças do teste de Kupiec, o teste de Christoffersen converge para uma distribuição *qui-quadrada* quando o tamanho da amostra tende ao infinito.

O teste de Christoffersen foi um dos testes utilizados como *backtesting* por Bali, Mo e Tang (2008) para verificar um VaR condicional mensurado por meio de uma distribuição

generalizada assimétrica t (SGT¹⁷) para ações da NYSE, AMEX e NASDAQ em um período entre 1963 a 2004. Dessa forma, foi verificado que tal modelo de risco apresenta ótimas previsões sobre os riscos de mercado, bem como consegue capturar a taxa de ocorrências e a extensão de eventos externos no mercado acionário americano.

Apesar do teste de Christoffersen apresentar propriedades de cobertura mais precisas se comparado ao teste de Kupiec e por testar a independência das violações, Wong (2010) alerta que esse teste não deve ser utilizado para *backtesting* regulatório¹⁸, por se tratar de um teste bicaudal. Isso significa que, em um teste bilateral, a hipótese alternativa implica a probabilidade em se rejeitar um VaR mais conservador.

Os *backtests*, como fora proposto por Kupiec (1995) e Christoffersen (1998), são baseados em variáveis binárias, ou seja, são testes em que revelam apenas se houve ou não alguma violação do VaR. Assim, Nieto e Ruiz (2016) e Gaglianone *et al.* (2011) chamam a atenção ao que diz respeito ao sacrifício de informações e na premissa de que as violações observadas no *backtesting* são eventos raros, indicando que em amostras finitas o número de exceções pode ser muito pequeno, atrapalhando a detecção de uma possível rejeição de um modelo mal especificado.

1.2.3 Duration

Os testes de *duration*, ou testes de independência com duração são uma classe de *backtesting* que verificam o número de observações entre duas exceções consecutivas do VaR. A partir da duração entre uma exceção e outra é possível detectar períodos em que o mercado está mais calmo (duração mais curta) ou mais turbulento (duração mais longa). O método de *duration* mais abordado na literatura segue a proposta de Christoffersen e Pelletier (2004) e Haas (2005).

Considere então:

$$D_i = t_i - t_{i-1} \quad (12)$$

em que t_i representa o dia da violação i do VaR.

¹⁷ SGT: Skewed generalized t . Esse modelo é capaz de realizar a estimação empírica da distribuição do retorno dos ativos, obtendo informação sobre assimetrias, “caudas pesadas” e leptocurtoses. A estimação das médias condicionais e volatilidade do retorno dos ativos é obtida utilizando modelos como o GARCH e suas variações (BALI; MO; TANG, 2008).

¹⁸ *Backtesting* regulatório seguindo as recomendações do Comitê de Basileia, por exemplo.

Assume-se como hipótese nula do teste de que o VaR está especificado de maneira correta, de forma que a duração entre duas exceções não deve ter memória e uma duração média de $\frac{1}{p}$ dias.

Christoffersen e Pelletier (2004) ainda demonstram a distribuição de probabilidade discreta referente a propriedade sem memória:

$$\begin{aligned}
 \Pr(D = 1) &= p \\
 \Pr(D = 2) &= (1 - p) p \\
 \Pr(D = 2) &= (1 - p)^2 p \\
 &\vdots \\
 \Pr(D = d) &= (1 - p)^{d-1} p
 \end{aligned} \tag{13}$$

Já a distribuição de probabilidade contínua da propriedade sem memória deve ser exponencial, portanto, de forma geral, sob a hipótese nula a distribuição da *duration* (D) deve seguir:

$$f_{\text{exp}}(D; p) = p e^{-pD} \tag{14}$$

De modo a determinar um teste estatístico de independência, Christoffersen e Pelletier (2004) utilizam da distribuição de Weibull como hipótese alternativa para a dependência da *duration*:

$$f_w(D; a, b) = a^b b D^{b-1} e^{-(aD)^b} \tag{15}$$

O método de *duration* pode identificar de maneira mais eficiente um VaR mal especificado se comparado a outros testes de independência utilizando uma cadeia de Markov, por exemplo, como observa Campbell (2007). No entanto, o autor releva uma preocupação diante da formulação da hipótese alternativa de ambos testes, visto que existe uma gama de adversidades em que a instituição estará exposta, dessa forma é necessário delinear cada possível cenário para que o teste possa detectar sistematicamente qualquer falha.

O estudo empírico de Candelon *et al.* (2011) discorre acerca do fato de que nem sempre é fácil delinear uma distribuição específica para a construção da hipótese alternativa, e por essa razão o método de *duration* formulado por Christoffersen e Pelletier (2004) não é muito difundido entre os analistas. Diante dessa limitação, os autores apresentam um backtesting

também usando *duration*, mas incorporando uma estrutura GMM¹⁹, utilizando momentos específicos associados a uma distribuição geométrica. A partir disso, o método pode melhorar o poder e tamanho do teste utilizando-se de aproximações contínuas, em detrimento das discretas.

Assim como Candelon *et al.* (2011), Ziggel *et al.* (2014) também busca melhorar a proposta de Christoffersen e Pelletier (2004) e demonstram um *backtesting* o qual além de considerar a independência das violações do VaR, detecta também a presença de *clusters* nesse processo. Essa nova abordagem permite calcular valores críticos mais precisos, deduzidos empiricamente por meio de simulações de Monte Carlo. A vantagem desse método é a identificação de um VaR mal especificado principalmente em cenários de *boom* econômico, em que as perdas financeiras se agrupam devido à alta volatilidade do mercado financeiro.

1.2.4 Quantil Dinâmico (DQ)

O *backtesting* a partir de um quantil dinâmico (*DQ test*), tal qual foi arquitetado por Engle e Manganelli (2004), verifica se as exceções observadas não estão correlacionadas com o conjunto de variáveis pertencentes às informações disponíveis quando o VaR foi calculado. O modelo proposto é baseado em uma regressão linear a qual inclui uma variável que representa a exceção (*hit*) do VaR, um conjunto de variáveis explicativas, valores defasados da variável *hit* e qualquer função do conjunto de informações passadas.

Assim como no teste de Christoffersen, o *DQ test* também avalia a condição de independência das variáveis, transformando uma série de retornos serialmente correlacionados e heterocedásticos em uma sequência de retornos serialmente independentes, de modo a testar se:

$$[I(y_t < f_t(\beta^0))]_{t=1}^T = \{I_t\}_{t=1}^T. \quad (16)$$

onde I é a função indicadora, y_t a série de retornos, $f_t(\beta^0)$ ²⁰ representa o modelo VaR calculado a partir da regressão linear associado ao quantil dos retornos do portfólio e I_t variáveis indicadoras no tempo t .

¹⁹ GMM (*Generalized method of moments*): O *backtesting* construído com base nesse método permite associar polinômios ortogonais a uma distribuição geométrica a fim de construir uma hipótese específica de acordo com a propriedade que se queira testar. Ver mais em Bontemps e Meddahi (2005).

²⁰ A regressão linear delineada por Engle e Manganelli (2004) nada mais é do que o Valor em Risco autorregressivo condicional (CAViaR), em que o quantil dos retornos do portfólio no tempo t é uma função do quantil dos retornos do período passado, levando-se em consideração também um operador de lag l o qual relaciona o conjunto de informações no tempo $t - 1$ com o quantil dos retornos no tempo t . Apesar dos autores

Entretanto, Engle e Manganeli (2004) afirmam que verificar a independência das variáveis indicadoras é apenas uma condição necessária, mas não suficiente para a construção do *backtesting* que avalia o quantil dos retornos do portfólio.

Considere então $\{z_t\}_{t=1}^T$ como uma sequência de variáveis aleatórias independentes, em que:

$$\begin{aligned} P(z_t = 1) &= \theta \\ P(z_t = -1) &= 1 - \theta \end{aligned} \tag{17}$$

Isso significa que a probabilidade de z_t apresentar autocorrelação positiva é igual a θ (nível de significância do VaR), no passo que a probabilidade de z_t apresentar autocorrelação negativa será $1 - \theta$. Todavia, analisar somente as probabilidades incondicionais das exceções é um erro extremo de medição do quantil, uma vez que essas probabilidades incondicionais são corretas e serialmente não correlacionadas, mas as probabilidades condicionais das exceções dadas pelo quantil ainda são serialmente correlacionadas. A inclusão de qualquer ruído na estimativa do quantil irá alterar a probabilidade condicional de uma exceção.

Desse modo, Engle e Manganeli (2004) propõem um novo teste a fim de sanar a limitação relacionada a probabilidade condicional de uma exceção:

$$Hit_t(\beta_0) = I(y_t < f_t(\beta^0)) - \theta \tag{18}$$

onde $Hit_t(\beta_0)$ é a exceção relacionada ao parâmetro regressivo. Se o retorno do portfólio (y_t) for menor que o quantil ($f_t(\beta^0)$), $Hit_t(\beta_0)$ assume o valor de $1 - \theta$, caso contrário $Hit_t(\beta_0)$ será $-\theta$. A expectativa condicional de $Hit_t(\beta_0)$ dada qualquer informação disponível no tempo $t - 1$ precisa ser 0, ou melhor, $Hit_t(\beta_0)$ precisa ser não correlacionado com seus valores de defasagem e com $f_t(\beta^0)$ e ter valor esperado de 0. Se $Hit_t(\beta_0)$ atender essas condições, então não haverá erro na medição do quantil, tal qual demonstrado em (18), nem autocorrelação nas exceções e trará a proporção correta das exceções.

A construção do DQ *test*, portanto, consiste em testar se $T^{\frac{1}{2}}X'(\hat{\beta})Hit(\hat{\beta})$ é diferente de 0, em que $X_t(\hat{\beta})$, $t = 1, \dots, T$ pode depender de $\hat{\beta}$, e $[Hit(\hat{\beta}) = [Hit_1(\hat{\beta}), \dots, Hit_T(\hat{\beta})]'$. Considere também a expressão $M_T = (X'(\beta^0) - E[T^{-1}X'(\beta^0)H\nabla f(\beta^0)])D_T^{-1} \times \nabla' f(\beta^0)$, em que H é uma matriz diagonal com informações binárias. Assim, Engle e Manganeli (2004)

elaborarem o DQ *test* com base no CAViaR, esse modelo de *backtesting* pode ser empregado para quaisquer métodos de mensuração do VaR.

delineiam então o DQ *test in-sample* (dentro da amostra) e o DQ *test out-of-sample* (fora da amostra), respectivamente:

$$DQ_{IS} = \frac{Hit'(\hat{\beta})X(\hat{\beta})(\widehat{M}_T\widehat{M}'_T)^{-1}X'(\hat{\beta})Hit'(\hat{\beta})}{\theta(1-\theta)} \sim \chi_q^2 \quad T \rightarrow \infty \quad (19)$$

em que \widehat{M}_T é a diferença entre $X'(\hat{\beta})$ e a função gradiente de $f'(\hat{\beta})$.

$$DQ_{OOS} = \frac{N_R^{-1}Hit'(\hat{\beta}_{TR})X(\hat{\beta}_{TR})[X'(\hat{\beta}_{TR})X(\hat{\beta}_{TR})]^{-1}X'(\hat{\beta}_{TR})Hit'(\hat{\beta}_{TR})}{\theta(1-\theta)} \sim \chi_q^2 \quad R \rightarrow \infty \quad (20)$$

onde TR representa o número de observações *in-sample* (dentro da amostra) e NR o número de observações *out-of-sample* (fora da amostra).

No estudo empírico de Rubia e Sanchis-Marco (2013), os autores salientam que o *backtesting* baseado no quantil dinâmico pode apresentar maior poder se comparado ao teste de Kupiec e o teste de Christoffersen, pois esses últimos tendem a aceitar mais a hipótese de que o VaR foi bem estimado, na medida que o DQ *test* tende a rejeitar mais modelos de VaR, identificando assim modelos mal especificados.

1.2.5 *Backtesting* via regressão quantílica

O *backtesting* construído a partir de uma regressão quantílica leva em consideração não só informações de ortogonalidade entre a variável binária e alguns instrumentos, mas também fornece uma condição suficiente para avaliar o desempenho do VaR a partir de um modelo quantil regressivo, diferentemente dos testes de Kupiec, teste de Christoffersen, o de *duration* e o quantil dinâmico, em que são baseados apenas nas informações em torno das variáveis binárias.

O *backtesting* utilizando regressão quantílica mais abordado na literatura é o proposto por Gaglianone *et al.* (2011) o qual utiliza de coeficiente aleatório para construir um teste de Wald na consolidação da hipótese nula de que o VaR esteja especificado de forma precisa.

O *backtesting* baseado em variáveis binárias é interpretado como teste do tipo multiplicadores de Lagrange (*LM tests*). Nieto e Ruiz (2016) ressaltam que os *LM tests* sacrificam muita informação, de tal forma que pode atrapalhar na decisão de aceitar ou rejeitar um modelo de risco. Em vista disso, usar o *LM test* pode reduzir o poder do *backtesting* em identificar um VaR mal especificado.

Gaglianone *et al.* (2011) observam ainda que ambos testes – LM e Wald – possuem equivalência assintótica na construção das hipóteses nula e alternativa, com a diferença que o teste de Wald pode fornecer um teste com maior poder para amostras finitas. Dessa forma, o *backtesting* via regressão quantílica usa mais informações para rejeitar um VaR mal especificado, ainda mais, é possível identificar quando e o porquê o VaR é especificado incorretamente.

O modelo proposto por Gaglianone *et al.* (2011) se constitui basicamente da medida de interesse do VaR, a qual é incorporada a uma classe geral de modelos para retorno dos ativos, diferentemente dos modelos que usam a volatilidade condicional como regressor. Dessa forma, é possível obter um VaR por meio de um teste de hipótese convencional. Considere a seguinte equação:

$$R_t = \beta_0(U_t) + \beta_1(U_t)VaR_t \quad (21)$$

em que R_t é um modelo de coeficiente aleatório, $U_t \sim iid U(0,1)$, $\beta_i(U_t)$, $i = 0,1$ apresentam comonotonicidade²¹ em U_t .

Levando em consideração a equação (17), o quantil condicional de R_t pode ser escrito como:

$$Q_{R_t}(\alpha|\mathcal{F}_{t-1}) = \beta_0(\alpha) + \beta_1(\alpha)VaR_t, \quad \text{para todo } \alpha \in (0,1) \quad (22)$$

Considere testar $Pr(R_t < VaR|\mathcal{F}_{t-1}) = \alpha$, onde \mathcal{F}_{t-1} é todo o conjunto de informação disponível em $t - 1$ e α é o nível de significância desejado. O modelo VaR pode ser testado a partir da seguinte hipótese nula:

$$H_0: \begin{cases} \beta_0(\alpha) = 0 \\ \beta_1(\alpha) = 1 \end{cases} \quad (23)$$

contra a hipótese alternativa geral. Em outras palavras, o modelo de previsão do quantil será preciso quando o intercepto da regressão for 0 (zero) e inclinação 1 (um).

De acordo com Gaglianone *et al.* (2011), o *backtesting* baseado em regressões quantílicas também é denominado por VQR *test*, cuja distribuição assintótica converge para uma qui-quadrado com dois graus de liberdade²², assim como observado nos testes de Kupiec e de Christoffersen.

²¹ Comonotonicidade: é uma propriedade estatística interessante a qual está relacionada a área de gerenciamento de riscos, visto que a comonotonicidade refere-se a uma dependência positiva perfeita entre os componentes de um vetor aleatório, ou seja, corresponde a uma estrutura de dependência mais arriscada observada em um determinado vetor. Para mais detalhes sobre demonstrações matemáticas ver Gaglianone *et al.* (2011).

²² Essa propriedade só é observada se as hipóteses de 1 a 4 do artigo de Gaglianone *et al.* (2011) forem garantidas.

A vantagem de se usar o *VQR test* é que esse *backtesting* pode ser aplicado tanto em situações nas quais o modelo utilizado para estimar o VaR é conhecido, por meio de um modelo GARCH, CAViaR ou um ARCH, por exemplo, quanto em casos cuja estimativa do VaR não é conhecida (GAGLIANONE *et al.*, 2011).

Autores como Gerlach e Wang (2016) e Trucíos, Tiwari e Alqahtani (2019) utilizam o *VQR test* como um de seus métodos para *backtesting* de modo a testar a precisão da estimativa do VaR, sendo o primeiro estudo aplicado a diferentes índices de mercado (S&P500, NASDAQ, Hang Seng, Nikkei 225, DAX e FTSE 100) e o segundo aplicado ao mercado de criptomoedas (Bitcoin, Dash, DigiByte, Dogecoin, Litecoin, MaidSafeCoin e Vertcoin). Em ambos estudos o *VQR test* obteve destaque na investigação da precisão da maioria dos modelos de risco se comparado a outros testes como o de Kupiec, Christoffersen e Quantil Dinâmico.

Considerações sobre o capítulo

Este capítulo buscou demonstrar os principais estudos empíricos acerca do VaR e do *backtesting*, bem como apresentar os pressupostos matemáticos que circundam o cálculo do *backtesting* em suas diferentes abordagens.

O Comitê de Basileia foi um marco importante no que tange à regulamentação do sistema bancário internacional, o que inclui normas e recomendações para o cálculo do VaR e do *backtesting*. As principais instituições financeiras precisam estar em conformidade às normas internacionais para total transparência e para adequação eficiente de capital, assegurando assim o bom funcionamento do sistema financeiro.

As diferentes abordagens teóricas sobre o VaR evidenciam que o risco de mercado pode ser mensurado sob diversas formas, a depender das informações disponíveis e pelas hipóteses a qual o pesquisador/analista deseja assumir. É importante destacar que os autores que dissertam acerca do VaR buscam investigar qual o modelo de risco é mais preciso para mensurar possíveis perdas para determinada série de tempo. Isso significa que para determinado portfólio e/ou ativo, existe uma abordagem para o cálculo do VaR cuja estimação foi mais precisa para aquele conjunto de informações, não havendo, portanto, uma metodologia a qual se sobressai de outra.

Concomitantemente, o *backtesting* é método indispensável para verificar a adequação do modelo de risco, se este de fato foi estimado de maneira correta e apresenta resultados consistentes. Como no VaR, existe variadas metodologias para a aplicação do *backtesting*, tal como foi apresentado neste capítulo. A apresentação da construção dos testes estatísticos permite constatar semelhanças as quais se podem destacar a convergência dos testes para uma

distribuição qui-quadrado, bem como a observação de variáveis binárias; e diferenças entre os testes, salientando as hipóteses de independência, duração e formação de *clusters* entre as exceções, como também um *backtesting* calculado a partir de uma regressão quantílica.

Delineado o escopo do *backtesting*, o próximo capítulo discorrerá sobre dois campos da estatística em que o *backtesting* poderia ser ensaiado e que a literatura ainda não se encarregou de tal discussão.

CAPÍTULO 2 - CONTROLE ESTATÍSTICO DE QUALIDADE VERSUS ANÁLISE ESTATÍSTICA SEQUENCIAL

Uma vez definida a metodologia para o cálculo do VaR, faz-se necessário um monitoramento adequado para detectar rapidamente alguma mudança no modelo de risco, visto que quaisquer alterações na mensuração do VaR implicam em mudanças nas decisões de alocação de capital. Visto isso, é imprescindível que se defina um método para cálculo do *backtesting*.

Um dos ramos da estatística amplamente discutido na literatura (SHEWHART, 1932; MONTGOMERY; MASTRANGELO, 1991; BARNARD, 1959; KOUTRAS, BERSIMIS; MARAVELAKIS, 2007) que se encarrega no monitoramento de processos é o “controle estatístico de qualidade”. As ferramentas para controle estatístico de qualidade permitem a detecção rápida de mudanças na distribuição de probabilidades geradora de um processo estocástico de interesse. Isto é feito com o auxílio de gráficos de controle, os quais consistem basicamente no acompanhamento dos valores realizados do processo monitorado em comparação a limites de confiança inferiores e superiores.

Outra importante área da estatística no monitoramento de processos é a chamada “análise estatística sequencial” (WALD, 1945). Em análise sequencial, o objetivo é monitorar um processo estocástico de interesse visando estimar parâmetros de sua distribuição de probabilidades, ou testar hipóteses estatísticas sobre tais parâmetros tidos como características intrínsecas ao processo desde o início do monitoramento. Vale ressaltar que muitas características dos métodos sequenciais se parecem aos gráficos de controle, pois em tais casos é comum que se tenham limites inferiores e superiores para o acompanhamento do processo realizado ao longo do tempo.

Uma importante diferença entre os gráficos de controle e os gráficos de análise sequencial é a formulação das estatísticas de teste usados em cada uma das abordagens. Nos gráficos de controle, a estatística de teste é baseada apenas nas novas observações do processo que chegam para cada análise específica, ou seja, a estatística de teste não é baseada em dados cumulativos no tempo. Em sentido oposto, na análise sequencial a estatística de teste é calculada com informações cumulativas no tempo.

A literatura de *backtesting* tem sido fundamentada sob a abordagem dos gráficos de controle, como exposto no trabalho de Pena, Rivera e Mata (2006). Entretanto, seria natural questionar se o *backtesting* não deveria ser tratado como um problema de análise sequencial,

visto que o objetivo é testar a qualidade de ajuste de um modelo de risco em que os dados se acumulam no tempo.

Diante desse contexto, o objetivo desse capítulo é apresentar as principais definições acerca dos gráficos de controle e da análise sequencial, bem como suas propriedades estatísticas, com o intuito de elucidar as principais diferenças entre as duas abordagens e tentar delinear em qual escopo um *backtesting* pode estar inserido.

2.1 Controle Estatístico de Qualidade

O controle estatístico de qualidade é uma técnica muito utilizada principalmente na indústria para o monitoramento de produtos e processos, cujo principal objetivo é minimizar variabilidades que basicamente são determinadas pela média e desvio padrão populacionais de um processo (SAMOHYL, 2009).

Mingoti e Yassukawa (2008) descrevem que com o monitoramento adequado a partir do controle de qualidade é possível minimizar custos e perdas na produção, desde que o processo tenha seus valores centralizados nos parâmetros pré-estabelecidos e que sua variabilidade seja mínima.

Dentro do controle estatístico de qualidade destacam-se os gráficos de controle, método que visa monitorar uma ou mais variáveis aleatórias por meio de inferências estatísticas, utilizando limites inferiores e superiores para verificar se o processo em questão está estatisticamente sob controle. A seguir, serão apresentadas as abordagens dos gráficos de controle de Shewhart, CUSUM (*Cumulative Sum*) e EWMA (*Exponentially Weighted Moving Average*), bem como suas aplicações no contexto do VaR.

2.1.1 Gráficos de controle de Shewhart

Os gráficos de controle de Shewhart (1931) são os métodos mais abordados na literatura e de grande importância para o controle de qualidade dos processos. São elementos visuais de fácil utilização, cujo método baseia-se em delinear uma região em que as amostras coletadas periodicamente permaneçam dentro dos limites para que o processo esteja sob controle.

Existem três abordagens diferentes para os gráficos de controle, de acordo com Shewhart (1931):

- (i) Gráficos para controlar a média amostral \bar{X} ;

- (ii) Gráficos para controlar o desvio padrão (variabilidade) S;
- (iii) Gráficos para amplitude R.

Os gráficos para controle da média são as abordagens mais utilizadas. Mingoti e Yassukawa (2008) argumentam que o gráfico para controle médio de um processo constitui-se de um limite inferior de controle (LIC), limite superior de controle (LSC) e uma linha central (LC) representando o valor médio do processo sob controle. Os limites são calculados a partir das seguintes equações:

$$LIC = \mu - k\sigma \quad (24)$$

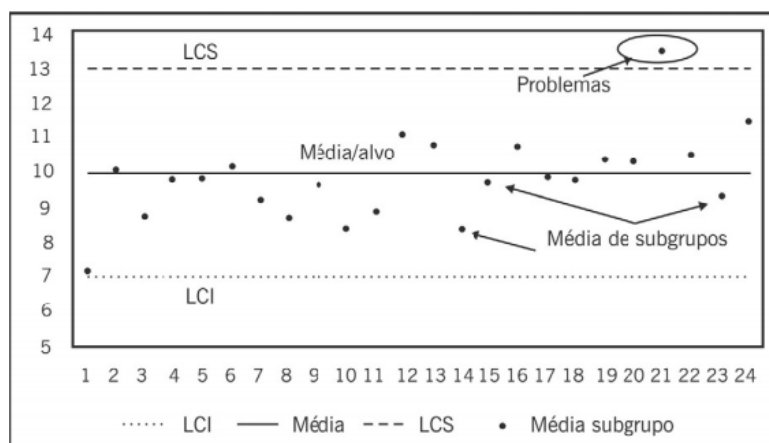
$$LSC = \mu + k\sigma \quad (25)$$

$$LC = \mu \quad (26)$$

sendo que μ e σ se refere a média e o desvio padrão do processo, respectivamente, e k é o percentil da distribuição normal padrão relacionado ao coeficiente de confiança, ou de confiabilidade estipulado pelo analista, o que geralmente é fixado em $k = 3$, representando um coeficiente de aproximadamente 99,7%, tal como argumentado por Koutras, Bersimis e Maravelakis (2007). Isso implicaria que a probabilidade de se rejeitar um processo que está sob controle (probabilidade do erro do Tipo I) é de 0,0027. Caso a distribuição das amostras não seja uma normal, recorre-se ao teorema central do limite, o qual demonstra que as médias das amostras seguem uma distribuição aproximadamente normal, quando o tamanho amostral no cálculo de \bar{X} é suficientemente grande.

A Figura 1 é uma representação de um gráfico de controle de Shewhart, onde o limite central é igual a 10 ($LC = 10$), o limite inferior é 7 ($LIC = 7$), limite superior é 13 ($LSC = 13$) e desvio padrão igual a 3 ($k = 3$). É possível notar que a média de um subgrupo ultrapassou o limite superior, o que indica problemas e que o processo está fora de controle.

Figura 1 - Gráfico de controle convencional



Fonte: Samohyl (2009)

A performance dos gráficos de controle pode ser avaliada pelo *Average Run Length* (ARL) - que em português seria o tempo médio de monitoramento - medida de desempenho que representa o tempo esperado até que ocorra a primeira ocorrência de um evento fora de controle. Shewhart (1931) divide o ARL em duas medidas: o *ARL in-control*, definida como o número médio de amostras observadas no gráfico até que um evento fora de controle seja sinalizado, mesmo que o processo ainda esteja sob controle, e o *ARL out-of-control* que mensura o número médio de amostras observadas no gráfico antes que um evento fora de controle ocorra quando o processo de fato já está fora de controle.

Jensen *et al.* (2006) demonstra que o ARL pode ser mensurado como:

$$ARL = \frac{1}{Pr(sinal)} \quad (27)$$

sendo o $Pr(sinal)$ a probabilidade de uma única amostra ficar fora dos limites de controle.

A partir da equação (26) é possível deduzir as equações para o *ARL in-control* e *ARL out-of-control*:

$$ARL_{in\ control} = \frac{1}{\alpha} \quad (28)$$

em que o α é a probabilidade de cometer o erro do Tipo I.

$$ARL_{out\ of\ control} = \frac{1}{1 - \beta} \quad (29)$$

em que β é a probabilidade de se cometer o erro do Tipo II, ou seja, a probabilidade de que uma sinalização não ocorra mesmo que o processo esteja fora de controle.

Quando os parâmetros das amostras observadas forem desconhecidos, os gráficos de controle passam por um procedimento dividido em duas fases. A Fase I as amostras históricas são utilizadas para definir o estado de controle do processo. O gráfico de controle só pode ser empregado em um monitoramento de qualidade quando os parâmetros em questão são conhecidos e mantidos sob controle. Depois de conhecidos os parâmetros, os limites de controle podem ser estimados para o uso na Fase II. Na Fase II as amostras são monitoradas prospectivamente por desvios do estado de controle. Caso as estatísticas obtidas das amostras estiverem dentro dos limites, significa que processo está sob controle, caso contrário, a evidência é de que o processo está fora de controle (JENSEN *et al.*, 2006).

Uma suposição muito importante para a validade dos gráficos de controle é a de que as observações amostrais são independentes e identicamente distribuídas. Entretanto, a hipótese de independência é violada em alguns problemas, tal como na indústria química, por exemplo, em que as amostras são coletadas em uma frequência curta em relação ao período constante do processo, fazendo com que as observações sejam autocorrelacionadas (MONTGOMERY; MASTRANGELO, 1991).

Os gráficos de controle CUSUM e EWMA são abordagens que minimizam o problema de autocorrelação das observações e conseguem detectar pequenas mudanças no processo de forma mais eficiente.

2.1.2 Gráficos de controle CUSUM

Preconizados por Page (1954), os gráficos de soma acumulada (CUSUM) levam em consideração os testes sequenciais de Wald para construção de limites inferiores e superiores. A soma acumulada é computada da seguinte forma:

$$Z_n = y + \sum_{i=1}^{i=n} X_i \quad (30)$$

em que y é a diferença entre a média das amostras observadas e a média global do processo ($X_i - \mu_0$) e $\sum_{i=1}^{i=n} X_i$ é o somatório das médias amostrais.

Quanto maior for o distanciamento da média amostral com respeito à média global do processo, maior será o valor da soma acumulada, aumentando assim a evidência de que ocorram desvios no processo monitorado.

Mingoti e Yassukawa (2008) apresentam as estatísticas de limites inferiores (desvios negativos abaixo da média) e superiores (desvios positivos acima da média) do gráfico de controle CUSUM, respectivamente, como:

$$z_i^- = \max [0, (\mu_0 - K) - x_i + z_{i-1}^-] \quad (31)$$

$$z_i^+ = \max[0, x_i - (\mu_0 + K) + z_{i-1}^+] \quad (32)$$

sendo K o valor de tolerância, definido de acordo com o coeficiente de confiança de interesse.

Assim como nos gráficos de controle de Shewhart, se os limites definidos por z_i^- e z_i^+ ultrapassarem um limite pré-estabelecido H , o processo está fora de controle.

2.1.3 Gráficos de controle EWMA

Os gráficos de controle EWMA levam em consideração médias móveis exponencialmente ponderadas que detectam com eficiência até mesmo pequenas mudanças na média de um processo. Idealizado por Roberts (1959), esses gráficos atribuem maior peso às amostras observadas mais recentes, no intuito de captar melhor as variações no processo.

A média calculada exponencialmente é definida pela seguinte estatística:

$$Z_i = \lambda X_i + (1 - \lambda)Z_{i-1} \quad (33)$$

em que Z_{i-1} é a média do período anterior multiplicada pelo parâmetro de ponderação λ , ($0 \leq \lambda \leq 1$). X_i representa a média amostral no período recente. Assim, é possível deduzir que a ponderação da média cai exponencialmente nos períodos anteriores até que, no limite, assume o valor zero. Isso significa que os dados mais antigos possuem peso menor comparativamente aos dados mais recentes.

Segundo Samohyl (2009), o desvio-padrão de Z_i é dado por:

$$\sigma_{z_i} = \sigma_x \sqrt{\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i})} \quad (34)$$

A partir da equação (34), e considerando que a distribuição das amostras segue uma distribuição normal, é possível construir os limites inferiores e superiores dos gráficos de controle EWMA.

$$LIC = \mu_0 - K\sigma_x \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i})}, \quad (35)$$

$$LC = \mu_0, \quad (36)$$

$$LSC = \mu_0 + K\sigma_x \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i})}, \quad (37)$$

onde que K e λ são os parâmetros de referência, podendo ser escolhidos conforme um desempenho pré-estabelecido, como, por exemplo, a maximização do ARL. A escolha de λ pode se dar de forma a aumentar a sensibilidade dos gráficos de controle; quanto menor for o valor de λ , mais eficiente será a detecção de pequenas mudanças no processo.

2.1.4 Aplicação dos gráficos de controle para o VaR

Pena, Rivera e Mata (2006) propõem um *backtesting* alternativo ao recomendado pelo Comitê de Basileia, utilizando um controle de qualidade das medidas de risco (QCRM²³). A nova proposta se baseia na inversão das hipóteses nula e alternativa para diminuir a probabilidade do erro do Tipo II; a elaboração de uma função poder menos conservadora no controle da probabilidade do erro do Tipo II, fornecendo um equilíbrio entre os erros do Tipo I e Tipo II e construção de intervalos de confiança para definição das regiões de aceitação e rejeição.

Dessa forma, o QCRM define as hipóteses nula e alternativa da seguinte forma:

$$H_0^Q = \text{o modelo VaR é incorreto vs } H_1^Q = \text{o modelo VaR é correto}$$

A Tabela 2 demonstra o erro do Tipo I e o erro do Tipo II para a abordagem adotada por Pena, Rivera e Mata (2006).

²³ QCRM: Quality control of risk measures

Tabela 2 – Erro do Tipo I e Tipo II no QCRM

		Decisão	
		Rejeitar modelo VaR	Aceitar modelo VaR
Modelo incorreto	Ho	Decisão correta	Erro do Tipo I
Modelo correto	H1	Erro do Tipo II	Decisão correta

Fonte: Pena, Rivera e Mata (2006), adaptado

A hipótese nula e alternativa do *backtesting*, de acordo com as diretrizes de Basileia, são as seguintes:

$$H_0^B: p = p_0 \quad vs \quad H_1^B: p > p_0 \quad (38)$$

em que p é a probabilidade condicional de uma exceção no dia i e p_0 representa a o nível de significância desejado no *backtesting*, geralmente no valor de 0,01.

A nova hipótese nula e alternativa do *backtesting* a partir do QCRM serão:

$$H_0^Q: p > p_0 \quad vs \quad H_1^Q: p \leq p_0 \quad (39)$$

em que p é uma probabilidade desconhecida de uma exceção e p_0 é a probabilidade de uma exceção quando o VaR está correto.

Dessa forma, Pena, Rivera e Mata (2006) argumentam que o novo teste de hipóteses pode ser visto como um problema de controle de qualidade, na medida em que o VaR pode ser rejeitado se a proporção de falhas do teste for maior que o limite aceitável, como $p_0 = 0,01$.

O estudo de Abbasi e Guilen (2013) vai além dos tradicionais gráficos de controle e emprega o uso do método de *bootstrap* com tamanho da amostra variável para monitoramento do VaR para empresas de seguros. O VaR é estimado através do método *bootstrap* paramétrico a partir de distribuições de Weibull, Burr II, Pareto e Birnbaum-Saunders para construir gráficos de controle para detectar rapidamente qualquer mudança nos modelos VaR.

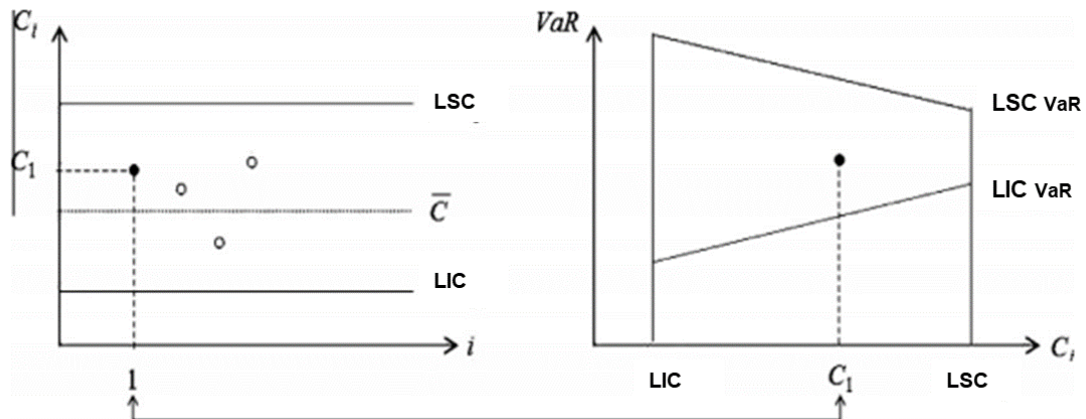
Os autores delineiam alguns passos principais para construir os gráficos de controle de *bootstrap* para um determinado quantil:

- (1) Seleção de uma amostra grande de variáveis aleatórias;
- (2) Encontrar a melhor distribuição paramétrica da amostra;
- (3) Gerar uma amostra por *bootstrap* do melhor modelo ajustado a estimar o quantil da distribuição com os parâmetros estimados;

- (4) Repetir o passo 3 diversas vezes, obtendo N amostras *bootstrap* e realizar o passo 2 para as amostras encontradas;
- (5) Construir o intervalo de confiança²⁴ com base no quantil das amostras *bootstrap* encontradas;
- (6) Denotar os intervalos de confiança encontrados como os limites inferiores e superiores do gráfico de controle ($LIC = L_p$ e $LSC = U_p$).

O valor do *ARL* será calculado da mesma forma a qual é tratado nas equações (28) e (28), assim como nos gráficos de controle de Shewhart. A figura 2 apresenta a construção do gráfico de controle a partir do *bootstrap*:

Figura 2 - Gráfico de controle *bootstrap*



Fonte: Abbasi e Guillen (2013), adaptado.

Mihailescu (2004) propõe um *backtesting* sequencial a partir de gráficos de controle, indo além dos métodos de *backtesting* tradicional como o teste de Kupiec, teste de Christoffersen e a abordagem binomial de Basileia. Dessa forma, são apresentados gráficos de controle de Shewhart, CUSUM e EWMA para verificar qual método é mais apropriado para avaliação do VaR.

A construção das hipóteses nula e alternativa pode ser descrita como abaixo:

$$H_0: E(D_i) = \frac{p_0}{1 - p_0} \quad vs \quad H_1: E(D_i) \neq \frac{p_0}{1 - p_0} \quad (40)$$

²⁴ O intervalo de confiança é descrito no artigo.

em que D_i são os valores observados do VaR e p_0 o nível de significância estabelecido. Se o modelo VaR estiver correto, H_0 não deve ser rejeitado.

Para o problema dos gráficos de controle, as decisões giram em torno do ARL, que será representado pelo D_i plotados no gráfico antes de alguma amostra ultrapassar os limites inferiores e superiores. Quando o processo está sob controle, a probabilidade do erro do Tipo I será menor quando o ARL for grande. Caso contrário, se o processo está fora de controle, o ARL deve ser pequeno. A proposta para um novo ARL levará em consideração o número médio inspecionado (ANI), representado pelo número de observações do VaR até que um sinal fora de controle ocorra.

$$ANI_0 = ARL_0 + \frac{p_0}{1 - p_0} \quad (41)$$

Dessa forma, Mihailescu (2004) constrói os gráficos de controle de Shewhart, CUSUM e EWMA a partir das definições apresentadas e chegam a conclusão de que o gráfico de controle EWMA é a abordagem mais apropriada para detectar alterações na magnitude da medida de risco e identificar de maneira adequada os modelos VaR que são precisos e imprecisos.

Weib e Testik (2019) apresentam um estudo com uma abordagem um pouco diferente do que demonstrado pelos outros autores supracitados. Os autores fazem um “caminho inverso” no qual monitoram os gráficos de controle a partir da mensuração do VaR. A crítica levantada por eles é que o ARL tradicional pode ser muito limitado para avaliar a performance dos gráficos de controle, propondo assim utilizar as métricas de risco as quais utilizam de quantis exatos para monitorar os gráficos de controle de Shewhart, CUSUM e EWMA.

Dessa forma, Weib e Testik (2019) propõem o TCE (*Tail conditional expectation*) o qual representa a duração média condicional de execução nas exceções do VaR, substituindo o ARL *in control* e o ARL *out of control*:

$$TCE_{\alpha}^{ic} = E [L \mid L < VaR_{\alpha}^{ic}] \quad (42)$$

$$TCE_{\alpha}^{oot} = E [L \mid L > VaR_{\alpha}^{oot}] \quad (43)$$

em que:

$$VaR_{\alpha}^{ic} = \min\{x \in \mathbb{N} \mid P(L \leq x) \geq \alpha\} \quad (44)$$

$$VaR_{\alpha}^{oot} = \min\{x \in \mathbb{N} \mid P(L \leq x) \geq 1 - \alpha\} \quad (45)$$

sendo L o comprimento de execução representado por:

$$L = \min \{t \in \mathbb{N} \mid Z_t \notin [l, u]\} \quad (46)$$

e Z_t representando as estatísticas monitoradas.

Se o gráfico estiver sob controle, o risco representa a ocorrência precoce de um alarme falso, e, portanto, refere-se a um *ARL* baixo. Caso o gráfico esteja fora do controle, o risco pode ser detectado tardiamente, representando um *ARL* alto.

2.2 Análise Sequencial

A análise sequencial pode ser definida como uma área da inferência estatística que se preocupa em realizar inferências estatísticas usando dados cumulativos no tempo, isto é, o tamanho da amostra é uma variável aleatória e não um número pré-determinado antes do início de um monitoramento (WALD, 1945). No teste de hipóteses sequencial, por exemplo, o tamanho da amostra depende do resultado das observações, ao contrário do teste de hipóteses não sequencial usual, no qual o tamanho da amostra é pré-estabelecido.

A abordagem sequencial começou a ser desenvolvida mais fortemente a partir da década de 40, quando Wald (1945) propõe o teste da razão de probabilidade sequencial²⁵ (SPRT). O SPRT foi inspirado no teste da razão de verossimilhança e tem sido amplamente utilizado em diversas áreas da ciência, tais como em experimentos clínicos e no monitoramento da segurança de medicamentos e vacinas recém-comercializadas. Trabalhos como o de Kulldorff *et al.* (2011) e Kulldorff e Silva (2017) utilizam o MaxSPRT, uma extensão do SPRT.

Wald (1945) descreve o seguinte processo de decisão posterior a um teste de hipóteses: (1) aceitar a hipótese nula a ser testada; (2) rejeitar a hipótese nula; (3) continuar o experimento utilizando informações adicionais. Se a primeira ou a segunda decisão ocorrerem, a análise sequencial termina. Se a terceira decisão é tomada, a análise sequencial é mantida. A abordagem de Wald deixa claro que o tamanho da amostra é uma variável aleatória na análise sequencial.

Assim como no teste de hipóteses usual, o teste sequencial também se preocupa em otimizar as probabilidades dos erros do Tipo I e o erro do Tipo II. Na prática, o que se faz é estabelecer uma tolerância para a probabilidade global de ocorrência do erro do Tipo I, geralmente denotado por α . Wald propõe que o erro do Tipo II também seja controlado sob um valor pré-definido, denotado por β . Para tanto, diferente dos testes de hipóteses usuais, com o

²⁵ Teste de razão de probabilidade sequencial: tradução da expressão *sequential probability ratio test* (SPRT).

SPRT de Wald (1945), se a razão de verossimilhança ultrapassar um limite superior, ou inferior, pré-determinados, então a análise sequencial é finalizada para a tomada de decisão.

Kulldorff *et al.* (2011) explora o SPRT proposto por Wald (1945) e desenvolve o teste de razão de probabilidade sequencial maximizado (MaxSPRT), e o aplica a estudos clínicos envolvendo vigilância de segurança de medicamentos e vacinas recém-comercializadas. Os autores demonstram que o SPRT é muito sensível à escolha do risco relativo utilizado na construção da hipótese alternativa relacionado aos experimentos clínicos. Diante disso, propõem um teste em que a hipótese alternativa é composta, e o resolve adequadamente, pois não demanda que o usuário especifique um risco relativo alvo.

O MaxSPRT possibilita a construção de um limite superior, valor crítico para se rejeitar a hipótese nula quando é detectado um excesso de risco, e isto é feito por se estabelecer um limite superior para o tamanho máximo da amostra até o final do monitoramento sequencial.

Kulldorff e Silva (2017) ressaltam que além da análise sequencial garantir um nível *alpha* (cota superior para a probabilidade do erro do Tipo I) correto, é importante considerar o poder estatístico de se rejeitar a hipótese nula, o tempo médio de sinalização até que a hipótese nula seja rejeitada, e o tamanho final da amostra quando a hipótese nula não é rejeitada. Além disso, essa abordagem pode ser aplicada a dados com diferentes distribuições de probabilidade e desenhos sequenciais, tais como as distribuições de *Poisson* e a binomial.

Silva (2018) destaca que a abordagem do MaxSPRT proposta por Kulldorff *et al.* (2011) é fundamentado no monitoramento do valor da estatística de teste comparado a um limiar constante de sinalização. Jennison e Turnbull (2000) apresentam a abordagem chamada “*gasto de alpha*” ou “gasto do erro do Tipo I”, que estabelece uma função não decrescente no intervalo $[0, \alpha]$, em que α é o nível de significância global até o término do experimento. A escolha do formato da função “*gasto de alpha*” pode ser feita de modo a otimizar medidas de desempenho pré-experimentais, tal como minimizar o tempo médio de sinalização do teste.

2.2.1 Teste da Razão de Probabilidade Maximizado (MaxSPRT) para dados Binomiais

Kulldorff *et al.* (2011) apresenta o método MaxSPRT para dados binomiais a partir do fato de que nem sempre é possível obter estimativas confiáveis antes do início de um processo. Visto isso, o autor orienta a coleta de informações de eventos adversos seguindo a premissa de que as amostras seguem uma distribuição binomial. Isso significa, por exemplo, na observação de exceções do VaR (se um retorno excedeu ou não o valor do VaR).

O MaxSPRT para dados binomiais foi inspirado no problema de se monitorar o número de eventos adversos associados a uma nova droga/vacina colocada à disposição da população. Isso significa que o monitoramento de um evento continuaria até que hipótese nula fosse rejeitada, ou quando um total de N eventos adversos fossem observados.

A construção do MaxSPRT para dados binomiais permite a utilização do método a quaisquer problemas nos quais os dados seguem uma distribuição de probabilidade binomial, incluindo o *backtesting* do VaR. O teste de proporção de falhas de Kupiec, por exemplo, testa a hipótese de que a probabilidade determinada empiricamente corresponde ao nível de tolerância do VaR, em que p geralmente é 0,01 ou 0,05.

Usando a mesma notação de Kulldorff *et al.* (2011), seja n o número de eventos adversos observados até o momento de um dos testes sequenciais, e seja $c_n \leq n$ o número de eventos adversos, dentre os n provenientes do grupo de tratamento (exposto à vacina/droga). Seja z a razão entre o número de indivíduos do grupo controle e o número de indivíduos do grupo tratamento. O parâmetro z é o termo de pareamento entre os indivíduos dos grupos controle e tratamento, do inglês *matching ratio*. Assim, pode-se construir a razão de verossimilhança para dados binomiais da seguinte forma:

$$LR_n = \frac{P(C_n = c_n | H_A)}{P(C_n = c_n | H_0)} = \frac{[RR/(z + RR)]^{c_n} [z/(z + RR)]^{n-c_n}}{[1/(z + 1)]^{c_n} [z/(z + 1)]^{n-c_n}} \quad (47)$$

O estimador de máxima verossimilhança do risco relativo (RR) é $zc_n/(n - c_n)$. Assim, a equação (47) pode ser reescrita da seguinte forma:

$$LR_n = \frac{(c_n/n)^{c_n} [(n - c_n)/n]^{n-c_n}}{[1/(z + 1)]^{c_n} [z/(z + 1)]^{n-c_n}} \quad (48)$$

onde $zc_n/(n - c_n) > 1$ e $LR_n = 1$ caso contrário. O log da razão de verossimilhança será definido como:

$$LLR_n = \ln(LR_n) = c_n \ln\left(\frac{c_n}{n}\right) + (n - c_n) \ln\left(\frac{n - c_n}{n}\right) - c_n \ln\left(\frac{1}{z + 1}\right) - (n - c_n) \ln\left(\frac{z}{z + 1}\right) \quad (49)$$

quando $zc_n/(n - c_n) > 1$ e 0 caso contrário.

A razão de verossimilhanças empregada por Kulldorff *et al.* (2011) é uma estatística de teste em que o numerador corresponde à hipótese alternativa (H_A) e o denominador é a probabilidade sob a hipótese nula (H_0). No MaxSPRT, a hipótese alternativa é composta,

portanto o risco relativo sob a hipótese nula não é apenas um ponto, mas sim valores maiores que 1 ($H_A: RR > 1$).

O MaxSPRT fornece uma equação geral para a razão de verossimilhanças, pois a escolha sobre os formatos das hipóteses nula e alternativa ficam a critério do gestor e/ou responsável pelo experimento, levando em consideração o risco relativo relacionado ao experimento.

Kulldorff *et al.* (2011) calcula os valores críticos exatos para o log da razão de verossimilhanças para dados binomiais, no caso de vigilância de segurança de medicamentos e vacinas pós-comercializadas. Os autores explicam que os valores críticos foram calculados analiticamente usando um procedimento de uma cadeia de Markov iterada, dessa forma é possível obter valores críticos exatos no lugar da utilização de simulações de Monte Carlo ou de cálculos assintóticos.

Kulldorff *et al.* (2011) explica que uma cadeia de Markov é construída para cada um dos valores de verossimilhança l , tendo como espaço de estado (n, c_n) , onde $n > 0$ e $0 \leq c_n \leq n$. Para o valor n , a probabilidade de cada estado pode ser facilmente calculada iterativamente a partir das probabilidades para o valor $n - 1$, com a condição inicial de que $P[(0,0)] = 1$. Estados para os quais $LLR_n(c_n) \geq l$ são estados absorventes. Somando as probabilidades dos estados absorventes, obtemos o nível α quando o valor de verossimilhança l é usado como valor crítico. A partir disso, é possível calcular valores críticos para qualquer problema de nível α , N observações e um valor p de interesse.

2.2.2 Função “gasto de alpha”

Como descrito na seção anterior, os testes de hipóteses no contexto da análise sequencial são aplicados aos dados acumulados no tempo, e isso permite ao analista antecipar decisões estatisticamente significativas, ainda que as amostras sejam pequenas. No teste sequencial convencional, como o SPRT de Wald (1945) e o MaxSPRT de Kulldorff *et al.* (2011), o monitoramento da estatística de teste se dá pela comparação de um limiar de sinalização constante em cada teste individual. A análise é interrompida assim que a hipótese nula é rejeitada, caso em que a estatística de teste se torna maior que o limiar definido (SILVA, 2018).

Silva (2018) argumenta que o teste de hipóteses sequencial pode ser arquitetado por meio do planejamento do “gasto do erro Tipo I”, ou simplesmente “gasto de alpha”, no lugar de um limiar constante dado na escala da estatística de teste, como proposto pelos testes

sequenciais convencionais. A função “*gasto de alpha*” é não-decrescente no intervalo $[0, \alpha]$, onde α é o nível de significância global escolhido pelo analista. Uma das formas mais utilizadas para a função “*gasto de alpha*” é a do tipo potência:

$$S(n) = \alpha \times \left(\frac{n}{N}\right)^\rho, \rho > 0, n \in (1, N]. \quad (50)$$

onde n é o índice de tempo definido como fração do tempo total de N e ρ é um parâmetro de forma especificada pelo analista. Assim, $S(n)$ determina o “gasto” da probabilidade do erro Tipo I no decorrer dos testes sequenciais.

Silva (2018) destaca ainda que a função “*gasto de alpha*” é útil para se otimizar o *trade-off* entre o poder estatístico do teste e o tamanho da amostra, ou do primeiro com o “tempo esperado de vigilância”. O tempo esperado de vigilância é uma medida de desempenho que pode ser traduzida como a média do tamanho da amostra no momento em que a vigilância é interrompida, independentemente de a hipótese nula ser rejeitada ou não. Existe ainda o “tempo esperado para sinalização”, que é a esperança condicional do tamanho médio da amostra quando a hipótese nula é rejeitada.

Silva (2018) demonstra que uma função “*gasto de alpha*” do tipo côncava, ou seja, $\rho < 1$ em (50), é mais indicada para minimizar o tempo esperado de sinalização, e que uma função “*gasto de alpha*” convexa, quando $\rho > 1$, é mais indicada para minimizar o tempo esperado de vigilância.

2.2.3 Medidas de desempenho: Função Poder, tempo médio de sinalização e tempo esperado de vigilância

Seja C_n um processo estocástico não negativo discreto. Assuma que C_n seja uma variável aleatória que representa o número de exceções do VaR com um n fixo. Sendo assim, pode-se escrever:

$$Y_n = C_n - C_{n-1} \quad (51)$$

em que Y_n segue uma distribuição de Bernoulli com probabilidade de sucesso $p_{n,R}$ para $n = 1, 2, \dots$ e $C_0 = 0$. Assuma que Y_1, Y_2, \dots , sejam independentes, ou seja:

$$Pr[Y_{n+1} = 1 | Y_1 = y_1, \dots, Y_n = y_n] = p_{n,R} \quad (52)$$

para qualquer sequência de observações 0's e 1's, y_1, \dots, y_n .

Silva (2018) destaca que a análise sequencial é dita “contínua” se os testes sequencias puderem rejeitar a hipótese nula para qualquer $C_n \geq b_n$ e, $1 \leq n \leq N$, em que b_n é uma sequência não decrescente de números nos naturais.

Seja n_1 o número mínimo de eventos exigidos antes que seja permitido que se rejeite a hipótese nula, ou seja, n_1 é um valor mínimo para n tal que:

$$Pr[C_n \geq b_n | R = 1] > 0 \quad (53)$$

Sendo assim, a função poder do teste sequencial que representa a probabilidade de se rejeitar a hipótese nula será:

$$\begin{aligned} \beta(R) &= Pr[C_{n_1} \geq b_{n_1} \cup \dots \cup C_N \geq b_N | R] \\ &= Pr[C_{n_1} \geq b_{n_1} | R] + Pr[C_{n_1} < b_{n_1} \cap C_{n_1+1} \geq b_{n_1+1} | R] + \dots + \\ &\quad Pr[\cap_{n=n_1}^{N-1} \{C_n < b_n\} \cap C_N \geq b_N | R] \\ &= \pi_{n_1}(R) + \dots + \pi_N(R) \end{aligned} \quad (54)$$

Para $n = n_1$, tem-se:

$$\pi_{n_1}(R) = Pr[C_{n_1} \geq b_{n_1} | R] = \sum_{c=b_1}^{n_1} \binom{n_1}{c} [p_{n,R}]^c [1 - p_{n,R}]^{n_1-c} \quad (55)$$

Portanto, o nível de significância global pode ser calculado a partir da equação (54), considerando o risco relativo igual ao especificado pela hipótese nula. No mesmo sentido, a probabilidade do erro tipo I gasta até o enésimo evento é dada por:

$$\alpha_n = \sum_{t=1}^n \pi_t(R = 1) \quad (56)$$

Assim, (56) é o “gasto de alpha” até o n-ésimo teste.

Tal como definido em Silva (2018), o tempo médio de sinalização é dado por:

$$\mathbb{E}[T | \text{rejeitar } H_0, R] = \frac{\sum_{n=1}^N n \times \pi_n(R)}{\beta(R)} \quad (57)$$

Analogamente, o tempo esperado de vigilância, denotado por $\mathbb{E}[T | R]$, é dado por:

$$\mathbb{E}[T | R] = \sum_{n=1}^N n \times \pi_n(R) + N \times [1 - \beta(R)] \quad (58)$$

Silva (2018) ressalta o fato de que, apesar das equações do tempo médio de sinalização e do tempo esperado de vigilância estarem relacionadas entre si, a melhor forma da função “*gasto de alpha*” será distinta para cada caso, considerando um poder fixo, pois a forma ótima depende das medidas estatística de desempenho a serem otimizadas em cada problema.

2.2.4 Aplicações da análise sequencial na economia

Algumas abordagens sequenciais são empregadas para problemas econômicos, como detecção de pontos de viragem nos ciclos reais de negócio (FRISEN, 2009) e acompanhamento das volatilidades de ativos financeiros (SCHIPPER; SCHMID, 2001). Entretanto, o uso da abordagem do MaxSPRT na economia ainda é pouco empregado, destacando o artigo de Silva e Barros (2016). Os autores utilizam o método MaxSPRT para mensurar o risco de crédito, bem como delinear os valores críticos do modelo.

Silva e Barros (2016) usam os seguintes pares de hipóteses:

H_0 : o modelo de crédito gera perdas aceitáveis

H_1 : o modelo de crédito gera perdas maiores que uma taxa aceitável

De acordo com a abordagem dos autores, o número de inadimplentes é um processo estocástico com distribuição de *Poisson* como segue:

$$P_r[X_l(m) = x | \lambda_l, \theta_l] = \begin{cases} \frac{e^{-\lambda_l m} (\lambda_l m)^x}{x!} & \text{se } m > 0 \\ \frac{e^{-\theta_l} \theta_l^x}{x!} & \text{se } m = 0 \end{cases} \quad (59)$$

em que θ_l representa o número esperado de clientes que causam perda de todo o contrato, λ_l a taxa de bons e maus pagadores e m é o tempo contínuo, com $x = 0, 1, 2, \dots$

Os autores utilizam da função “*gasto de alpha*” para especificar a quantidade da probabilidade do erro do Tipo I a ser gasta em cada teste sequencial:

$$F_\alpha(t_i) = \sum_{j=1}^i \alpha_j \leq \alpha \quad (60)$$

onde α é o nível de significância global do teste.

Os valores críticos do teste são construídos implicitamente a partir da definição da função “*gasto de alpha*”, uma vez que o valor crítico é usado como limite de sinalização. A função poder do teste mostra que até para amostras pequenas o teste tem um bom desempenho. Os autores concluem então que o método sequencial proposto detecta de forma rápidas possíveis perdas monetárias.

Considerações sobre o capítulo

Este capítulo apresentou as definições e pressupostos matemáticos dos gráficos de controle, empregados no controle estatístico de qualidade, bem como a metodologia relacionada à análise estatística sequencial. Comparando as duas abordagens, nota-se algumas semelhanças pontuais no que tange a avaliação da performance dos testes: o ARL (*in-control e out-of-control*) dos gráficos de controle e o tempo médio de sinalização e tempo esperado de vigilância utilizado na análise sequencial.

Entretanto, algumas diferenças entre as duas técnicas levantam observações importantes a serem discutidas. Ao utilizar os gráficos de controle plotando as amostras observadas, o processo em questão já deve estar sob controle para que o monitoramento seja realizado e qualquer alteração no processo deve ser detectado o quanto antes, ou seja, média e variância devem ser conhecidas. Caso as amostras observadas ultrapassem os limites definidos, o processo é tido como fora de controle.

Por outro lado, na análise sequencial, nos testes de hipótese não se conhece o verdadeiro valor dos parâmetros e a ideia do processo não passa pela terminologia de “controle”. Pressupõe-se que o processo tenha um valor intrínseco, como média e variância verdadeiras.

De forma sucinta, os gráficos de controle querem monitorar para detectar uma possível mudança em um processo que está sob controle, e o teste de hipótese na análise sequencial quer testar o real estado dos parâmetros os quais geraram o processo desde o início.

Sob a ótica dos gráficos de controle, o *backtesting* trata sobre o monitoramento de um modelo de risco que está sob controle e o que se quer é identificar o momento em que, de repente, há uma mudança no processo indicando que o VaR está mal especificado, ou melhor, que a distribuição dos dados não é a mesma do início do monitoramento. O objetivo, então, é detectar rapidamente a mudança no processo que define os parâmetros relacionados com o retorno da série, de modo a reajustar o cálculo do VaR.

De outra forma, o *backtesting* também pode ser interpretado como um problema da análise sequencial, uma vez que se queira testar se o modelo VaR está bem estimado desde o início do monitoramento, ou seja, queira-se testar uma hipótese sob a condição do processo intrínseco, desde o primeiro momento.

CAPÍTULO 3 - O PROBLEMA DOS TESTES MÚLTIPLOS: DEMONSTRAÇÃO TEÓRICA E ESTUDO DE SIMULAÇÕES

Apresentadas as principais abordagens do *backtesting* no Capítulo 1, e demonstrada a construção dos gráficos de controle no Capítulo 2, pode-se dizer que as metodologias para o *backtesting* atuais são arquitetadas segundo a ótica dos gráficos de controle, pois os valores críticos em cada teste são calculados de acordo com um nível de significância válido apenas para cada teste individual.

Sob a ótica sequencial, fixar o valor crítico de cada teste igual ao nível de significância global, não levando em consideração os testes antecedentes, gera uma inflação da probabilidade do erro do Tipo I global ao final de um período de testes consecutivos. Ou seja, erroneamente, pode-se julgar que um modelo VaR é inadequado quando na realidade ele está bem ajustado, e isto pode acontecer com uma probabilidade bem maior que α .

O objetivo desse capítulo é apresentar os primeiros resultados desse trabalho, demonstrando que um *backtesting* realizado sucessivas vezes, e que não leva em consideração o fato de que os dados chegam sequencialmente, pode inflar a probabilidade do Erro do Tipo I, ocasionando uma alta probabilidade de decisões equivocadas quanto a validade do VaR. Para tanto, a primeira seção demonstrará o problema geral teórico de testes consecutivos, levando em conta a ocorrência ou não de uma exceção do VaR, caracterizando um problema de natureza binomial, e a segunda seção apresentará uma investigação de alguns dos *backtesting* descritos no Capítulo I, aplicados sucessivas vezes para dados simulados, apresentando estimativas para o verdadeiro tamanho destes testes, seus valores críticos, e os poderes estatísticos associados.

3.1 O problema dos testes múltiplos para o *backtesting*

Suponha que se queira monitorar o VaR realizando testes até que a hipótese nula seja rejeitada (quando ocorrer alguma exceção) para um determinado nível de significância global α pré-estabelecido, ou até que o número de testes atinja N , o que ocorrer primeiro. Seja $P_t = Pr [VaR_t < R_t]$, onde R_t é o retorno do ativo. Assim, pode-se formular as seguintes hipóteses:

$$\begin{aligned} H_0: P_t &\leq 1 - \gamma \\ H_1: P_t &> 1 - \gamma \end{aligned} \quad (61)$$

onde γ é o coeficiente de confiança utilizado no cálculo do VaR.

Sendo assim, a probabilidade do erro do Tipo I será a seguinte:

$$Pr(X_t > c_t | H_0) = \alpha \quad (62)$$

em que c_t é o valor crítico no t-ésimo teste.

Seja Z a varável que conta o número de backtests realizados, consecutivamente, até que o modelo VaR seja rejeitado. Os eventos ocorrerão da seguinte forma:

$$\begin{aligned} Pr(\{Z = 1\} \cup \{Z = 2\} \cup \{Z = 3\} \cup \{Z = 4\} \cup \dots \cup \{Z = N\}) &= \alpha \\ \alpha &= \alpha^* + (1 - \alpha^*)\alpha^* + (1 - \alpha^*)^2\alpha^* + (1 - \alpha^*)^3\alpha^* + \dots + (1 - \alpha^*)^{N-1}\alpha^* \\ \alpha &= [\alpha^* + (1 - \alpha^*)^{N-1}] \end{aligned} \quad (63)$$

em que α^* é o “*gasto de alpha*” em cada teste.

Veja que cada teste está associado a um *alpha*. Ao final de N testes, o nível de significância global será maior que o nível de significância individual, α^* , usado no cálculo dos valores críticos para cada teste.

3.1.1 Exemplo prático de múltiplos testes para o *backtesting*

A aplicação de múltiplos testes pode inflacionar a probabilidade do erro do Tipo I global. Vejamos o exemplo a seguir.

Seja Y_n o número de exceções observados em um *backtesting* do VaR após n retornos diários. Supondo que as exceções ocorrem de forma independente, temos então que $Y_n \sim Binomial(n, p)$. Sendo assim, queira-se testar:

$$H_0: p \leq 0,02$$

$$H_1: p > 0,02$$

Primeiramente, será realizado um *backtesting* após 200 dias de monitoramento, e um segundo *backtesting* será executado 300 dias após o primeiro, ambos considerando um nível de significância $\alpha = 0,05$.

O primeiro teste terá como valor crítico $c_1 = 7$, pois $Pr(Y_{200} \geq 7 | p \leq 0,02) \leq Pr(Y_{200} \geq 7 | p = 0,02) = 0,0493$. Isto é, considerando os 200 primeiros dias observados, se 7 ou mais exceções forem observadas, então pode-se inferir que o VaR está mal ajustado. Ao contrário, se 6 ou menos exceções forem observadas, então pode-se continuar o monitoramento até que se observe mais 300 dias. O valor crítico do segundo teste será $c_2 = 16$, pois $Pr(Y_{500} \geq 16 | p = 0,02) = 0,047$.

Se considerarmos apenas o segundo *backtesting*, a probabilidade de se rejeitar a hipótese nula será de 0,047, se apenas este teste for realizado. Entretanto, um primeiro *backtesting* fora calculado e não incorporado ao segundo. Visto isso, a máxima probabilidade de se rejeitar a hipótese nula será ligeiramente maior, de modo a ultrapassar o nível de significância desejado de 0,05:

$$\begin{aligned} Pr(rej. H_0 | p = 0,02) &= Pr(Y_{200} \geq 7 | p = 0,02) + Pr(Y_{200} < 7, Y_{500} \geq 16 | p = 0,02) \\ &= 0,0493 + \sum_{y=0}^6 [Pr(Y_{300} > 15 - y | p = 0,02) Pr(Y_{200} = y | p = 0,02)] \end{aligned} \quad (64)$$

$$Pr(rej. H_0 | p = 0,02) = 0,0792$$

Observe-se então que a probabilidade do erro do Tipo I associada aos dois testes é próxima de 0,08, maior que o nível de significância desejado. Se repetidos testes forem realizados ao longo do tempo, a probabilidade do erro do Tipo I será consideravelmente muito maior, o que tornará o *backtesting* inadequado para verificação da adequabilidade do VaR.

3.2 Estudo de simulação

As simulações apresentadas nessa seção têm o objetivo de investigar o comportamento do *backtesting* aplicado sucessivas vezes e sem levar em consideração o fato de que os dados chegam sequencialmente, isto é, que os dados vão se acumulando ao longo do tempo.

Mais especificamente, a ideia principal desse experimento é testar um *backtesting* considerando duas situações hipotéticas: a primeira em que a distribuição dos retornos de um ativo fictício segue uma normal (hipótese nula) e a segunda em que a distribuição dos retornos do ativo em questão segue uma *t-student* ou que os retornos apresentem uma variância condicional no tempo, modelada por um GARCH (hipóteses alternativas).

A estimação do VaR é realizada por meio de uma abordagem paramétrica, considerando o percentil 1% ($p = 0,01$) da distribuição normal.

Dado o contexto, o processo pelo qual a hipótese nula está submetida encontra-se sob controle. Isso significa dizer que a média dos retornos do ativo fictício bem como sua variância são tidos como verdadeiros. Supostamente, o VaR e o *backtesting* não teriam problemas de estimativa, apresentando assim modelos bem ajustados. Isso significa que se todos os testes são válidos, eles serão testes de nível α .

No entanto, a hipótese a ser verificada é se o *backtesting* sob a hipótese nula realmente está sob controle e não apresenta problemas de estimação, como a inflação da probabilidade do erro do Tipo I, mesmo sob a circunstância de que os dados vão se acumulando no tempo.

Assim sendo, esta seção está subdividida em duas partes: a primeira referente aos procedimentos metodológicos necessários para a realização das simulações e a segunda parte no que tange os resultados obtidos, computados através do tamanho e nível do teste, construção de valores críticos e análise da função poder dos testes.

3.2.1 Procedimentos metodológicos

Os procedimentos metodológicos adotados para a execução das simulações estão descritos a seguir:

(i) **Criando retornos de um ativo financeiro**

O primeiro passo para a realização da simulação é construir uma série de retornos estimados de um ativo financeiro. Conforme Morettin (2011), um retorno é definido como a variação relativa de preços. Para a avaliação de risco de uma série financeira é importante realizar algumas suposições sobre a distribuição dos retornos, visto que na prática é muito difícil conhecer completamente as distribuições finito-dimensionais de tais retornos.

Diante isso, optou-se em gerar 3 amostras aleatórias de tamanho 1000 que simulam retornos (R_t) que seguem uma distribuição normal, uma distribuição *t-student* assimétrica e uma modelagem de volatilidade a partir de um modelo GARCH, assumindo que os retornos apresentam uma variância condicional evoluindo no tempo, assim sendo:

- (1) Os retornos do ativo seguem uma distribuição normal: $R_t \sim N(\mu, \sigma^2)$, onde a média (μ) é próxima de 0 e desvio-padrão (σ^2) próximo a 1;
- (2) Os retornos do ativo seguem uma distribuição *t-student* assimétrica: $R_t \sim st_d(\nu, \mu, \sigma^2, \gamma)$, com 3 graus de liberdade (ν), média (μ) é próxima de 0, desvio padrão (σ^2) próximo de 1 e assimetria (γ) de $-0,5$.
- (3) A volatilidade dos retornos do ativo, pode ser modelada a partir de um GARCH (r, s):

$$R_t = \sqrt{h_t} \varepsilon_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i R_{t-1}^2 + \sum_{j=1}^s \beta_j h_{t-j} \quad (65)$$

onde ε_t segue uma distribuição *t-student* assimétrica, com assimetria de -0.1 e 10 graus de liberdade.

A escolha do tamanho da amostra é baseada em alguns estudos empíricos (CHAN *et al.*, 2007; NIETO; RUIZ, 2016; ESCANCIANO; OLMO, 2010), em que os autores argumentam que para a estimação do VaR é importante que seja considerado um longo período de tempo, isto é, o número de observações deve ser suficientemente grande para uma previsão de VaR bem ajustada.

(ii) Cálculo do VaR

O VaR nada mais é do que o quantil condicional da distribuição de perda dos retornos de um ativo. Abad *et al.* (2014) descreve os retornos como um processo estocástico:

$$\begin{aligned} R_t &= \mu + \varepsilon_t \\ \varepsilon_t &= z_t \sigma_t, \quad z_t \sim iid(0,1) \end{aligned} \quad (66)$$

Dado um nível de significância $\alpha \in (0,1)$, o $VaR(\alpha)$ é definido como um quantil α da distribuição de probabilidade dos retornos do ativo em que $F(VaR(\alpha)) = \Pr(R_t < VaR(\alpha)) = \alpha$. O cálculo do VaR, portanto, pode se dar pela seguinte forma:

$$VaR(\alpha) = F^{-1}(\alpha) = \mu + \sigma_t G^{-1}(\alpha) \quad (67)$$

onde μ é a média dos retornos, σ_t o desvio padrão e $G^{-1}(\alpha)$ é uma função de distribuição condicional relacionado ao nível de significância. Importante ressaltar que, por simplicidade²⁶, optou-se em trabalhar com o desvio padrão dado pela distribuição dos próprios retornos, bem como estabelecer $G^{-1}(\alpha)$ como o quantil da distribuição normal para um nível de significância de 0,01, ou seja, -2.326348.

Considerando que cada uma das amostras delineadas no item 1 tem 1000 observações, foi realizada uma divisão entre duas partes: uma amostra *in-sample* de tamanho 750, utilizada para construir a estimativa do VaR e outra amostra *out-of-sample* de tamanho 250, empregada no cálculo do *backtesting* para validação do modelo VaR.

(iii) Realização do *backtesting*

De modo a validar o modelo VaR construído no passo 2, foi realizado o *backtesting* sob três óticas: o teste de Kupiec, teste de Christoffersen e Quantil Dinâmico. A escolha desses três tipos de *backtests* se dá pela ampla utilização desses testes nos estudos empíricos, como elencado no quadro 1 na revisão de literatura. Dessa forma será possível analisar diferentes propriedades das exceções observadas, como já fora elencado anteriormente: cobertura incondicional da proporção de exceções, a independência das exceções e se as exceções são correlacionadas com o conjunto de informações disponíveis quando o VaR fora calculado. A

²⁶ A escolha de não utilizar nenhum modelo para estimar a volatilidade dos retornos, se dá na premissa de que não importa como o VaR fora calculado, o objetivo é analisar como é a performance do *backtesting*. Já que o foco não era estimar o melhor modelo VaR, optou-se por usar o cálculo em sua forma mais simples e direta.

partir desse conjunto de propriedades já será factível delinear uma análise capaz de trazer discussões interessantes.

Para o cálculo do *backtesting* foi considerada a amostra *out-of-sample* desenhada no passo 2, considerando uma janela com os últimos 250 retornos observados, conforme recomendações do Comitê de Basileia e literatura empírica.

O nível de significância empregado no *backtesting* pode ser escolhido conforme objetivo do analista que geralmente pode ser fixado em 0,01, 0,025 ou 0,05. Desse modo, nesse trabalho foi utilizado um nível de significância de 0,05, para diferenciar o nível de significância usado no cálculo do VaR e não gerar equívocos de interpretação dos resultados, visto que o VaR e o *backtesting* apresentam duas estatísticas de teste distintas.

(iv) **Delineando uma situação hipotética**

Imagine a seguinte situação hipotética em que um analista irá calcular, no dia 1, um VaR levando-se em consideração a amostra *in-sample* de 750 retornos, bem como o seu *backtesting* utilizando-se da amostra *out-of-sample* com as últimas 250 observações. Passados 10 dias do primeiro teste, o analista irá calcular novamente um VaR, mas agora considerando as novas observações que chegaram, somando assim 760 retornos na amostra *in-sample* e o *backtesting* continua levando em conta as últimas 250 observações. Passados mais 10 dias, um novo VaR é calculado, considerando 770 observações na amostra *in-sample* e o *backtesting* ainda considera os últimos 250 retornos para a amostra *out-sample*. Esse processo é repetido sucessivamente a cada 10 dias, até completar 30 testes realizados.

A intenção desse experimento é mostrar que a cada dia um retorno do ativo é adicionado à amostra *in-sample* do teste, indicando que os dados vão se acumulando ao longo do tempo. Em contra partida, o *backtesting* não considera a chegada dos novos dados, considerando apenas a janela dos últimos 250 dias na amostra *out-sample*.

Esse experimento foi realizado 10.000 vezes, para cada uma das distribuições traçadas no passo 1, e para cada um dos *backtests* delineados no passo 2. O intuito de repetir o experimento 10.000 vezes é de obter robustez nos resultados, permitindo uma análise mais verossímil. Dessa forma, no final do experimento foi obtido uma matriz de 10.000 linhas (sendo cada linha uma simulação) e 30 colunas (número total de *backtests* realizados, considerando um intervalo de 10 dias entre cada *backtesting*).

3.2.2 Resultados e discussões

Traçado o procedimento metodológico do estudo de simulações, esta seção terá como objetivo apresentar os resultados obtidos no experimento, no que tange a identificação do tamanho e nível dos testes, a delimitação dos valores críticos construção da função poder dos *backtests*.

3.2.2.1 Tamanho do teste

Para discutir os primeiros resultados do experimento, se faz necessário apresentar dois conceitos importantes intrinsecamente ligados aos testes de hipóteses, que são o tamanho do teste e o nível do teste.

Na definição de Casella e Berger (2014), para $0 \leq \alpha \leq 1$, um teste com função poder $\beta(\theta)$ é um teste de tamanho α se $\sup_{\theta \in \theta_0} \beta(\theta) = \alpha$. Isso significa que o tamanho do teste é o valor máximo (supremo) da probabilidade de se rejeitar H_0 , dentro do espaço paramétrico sob a hipótese nula.

Diferentemente, para $0 \leq \alpha \leq 1$, um teste com função poder $\beta(\theta)$ é um teste de nível α se $\sup_{\theta \in \theta_0} \beta(\theta) \leq \alpha$. Isso quer dizer que um teste é de nível alfa se o valor máximo da probabilidade de se rejeitar H_0 for menor ou igual a α . Um teste então é tido como válido, se o teste for de nível α .

Visto isso, a Tabela 3 apresenta o tamanho dos testes realizados na simulação:

Tabela 3 - Tamanho dos testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$

Distribuição dos retornos	Teste Kupiec	Teste Christoffersen	DQ teste
Normal	0,3874	0,0409	0,3554
<i>t-student</i>	0,6868	0,5092	0,8497
Garch	0,9973	0,0244	0,0667

Fonte: Elaborado pelo autor a partir dos resultados

Analisando a Tabela 3, nota-se que apenas dois testes são de nível *alpha*, ou seja, são iguais ou menores que o nível de significância estabelecido de 0,05: o teste de Christoffersen no cenário em que a distribuição dos retornos segue uma normal, apresentando um tamanho 0,0409 e o teste de Christoffersen em que a variância dos retornos foi modelada através de um GARCH, com tamanho 0,0244.

É possível dizer que os outros testes não são válidos, visto que o tamanho desses testes foi superior ao *alpha*. Significa dizer que a probabilidade do erro do Tipo é, na verdade, maior do que o nível de significância desejado.

Em uma situação em que todos os testes sejam de nível *alpha*, ou seja, testes válidos e bem construídos, o nível de significância de 0,05 poderia ser utilizado na escala da estatística de teste (*valor-p*) como o valor crítico do teste. Em outras palavras, os *valores-p* dos testes deveriam coincidir com o nível de significância desejado.

Dado o contexto dos testes não serem todos válidos, o próximo passo é construir valores críticos na escala dos *valores-p* que deveriam ser utilizados para que todos os testes sejam de nível *alpha*.

3.2.2.2 Valores críticos

Os valores críticos foram calculados conforme o procedimento descrito na seção 2.2.1, em que é possível calcular valores críticos exatos para qualquer teste de nível α . A Tabela 4 sumariza tais resultados:

Tabela 4 - Valores críticos estimados, na escala do valor-p, para os testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$

	Teste Kupiec	Teste Christoffersen	DQ Teste
Valores críticos	0,01904923	0,05032401	0,0000000057

Fonte: Elaborado pelo autor a partir dos resultados

Examinando os resultados da Tabela 4, pode-se verificar que os valores críticos para o teste de Kupiec e para o DQ teste são substancialmente menores que o nível de significância desejado. O teste de Christoffersen foi o teste em que o valor crítico mais se aproxima do nível de significância.

Sendo assim, para que os testes sejam de nível *alpha*, os novos valores críticos a serem utilizados devem ser de 0,01904923, 0,05032401 e 0,0000000057 para os testes de Kupiec, teste de Christoffersen e DQ teste, respectivamente. Dessa forma, a probabilidade do erro do Tipo I será controlada.

Uma vez controlada a probabilidade do erro do Tipo I, será possível arquitetar a função poder dos testes, para comparar qual *backtesting* apresenta melhor desempenho.

3.2.2.3 Função Poder

A função poder é a probabilidade de se rejeitar a hipótese nula. Matematicamente, Casella e Berger (2014) definem a função poder de um teste de hipótese com região de rejeição R como $\beta(\theta) = P_{\theta}(X \in R)$. Um bom teste tem função poder próximo de 1 para $\theta \in \Theta_0^c$, em outras palavras, um teste considerado ideal tem função poder próximo de 1 para os parâmetros sob a hipótese alternativa.

Visto esse conceito, a Tabela 5 sintetiza a função poder para os testes:

Tabela 5 - Poderes estimados para os testes de Kupiec, Christoffersen e DQ para $\alpha = 0,05$

	Teste Kupiec	Teste Christoffersen	DQ teste
Normal	0,9777	0,362	0,9505
<i>t-student</i>	0,5283	0,0668	0,7445
Garch	0,9992	0,0012	0,9456

Fonte: Elaborado pelo autor a partir dos resultados

Sob a hipótese alternativa, situação em que o modelo VaR deveria ser rejeitado, os testes de Kupiec e DQ foram os que apresentaram melhores desempenhos. Considerando o cenário em que a variância dos retornos é modelada através de um GARCH, o teste de Kupiec rejeita 99,92% das vezes um modelo VaR mal especificado e o DQ teste rejeita 94,56%. Para o contexto em que a distribuição dos retornos segue uma *t-student*, o teste de Kupiec rejeita um VaR mal especificado 52,83% das vezes e o DQ teste rejeita 74,45% das vezes.

Com esse resultado, pode-se inferir que na maior parte das vezes os testes de Kupiec e DQ conseguem detectar de forma eficiente um modelo VaR mal especificado. Já o teste de Christoffersen apresentou poder muito baixo, indicando que o teste tem maior dificuldade em identificar e rejeitar um VaR mal especificado.

Considerações sobre o capítulo

Este capítulo buscou exemplificar, de maneira prática, o problema dos *backtests* aplicados consecutivamente ao longo do tempo, sem levar em consideração os dados acumulados no tempo.

Para evidenciar esse problema de maneira completa, primeiramente foi apresentado um exemplo teórico em que se queira monitorar um VaR até que a hipótese nula seja rejeitada, considerando um nível de significância global desejado α . O fato é que até a hipótese nula seja rejeitada, acontece n testes aleatórios e independentes, todos associados a um *alpha*. Dessa

forma, foi possível corroborar que ao final do teste onde a hipótese nula seja rejeitada, o α associado é bem maior que o α determinado no início dos testes, isto é, há uma inflação da probabilidade do erro do Tipo I.

Posteriormente, é demonstrado um exemplo de testes consecutivos levando em consideração um problema de natureza binomial, em que um *backtesting* é realizado sucessivamente sem levar em consideração a realização dos testes anteriores. Em um teste analisado de maneira independente, a probabilidade do erro do Tipo I condiz com o nível de significância estipulado no início do experimento. Entretanto, quando os testes são analisados de forma conjunta, considerando que os dados vão se acumulando no tempo, a probabilidade do erro do Tipo I ultrapassa o nível de significância desejado.

Em um segundo momento, foi realizado um estudo de simulações cujo ambiente controlado fosse possível testar a hipótese de que os *backtests* de Kupiec, Christoffersen e Quantil Dinâmico são testes válidos e de nível α .

Os resultados demonstraram que os *backtests* não são de nível α , apresentando uma probabilidade do erro do Tipo I maior do que o nível de significância desejado no início do experimento. Visto isso, houve a necessidade de nivelar os testes na escala da estatística de teste (*valor-p*) para que fosse possível compará-los entre si e que todos estivessem sob o mesmo nível. Para tanto, foram calculados valores críticos para cada um dos *backtests*, apresentando valores substancialmente inferiores ao nível de significância desejado.

Depois de nivelados os *backtests*, foi calculado o poder dos testes para que fosse possível comparar a medida de desempenho dos *backtests*. Os resultados indicaram maior poder para o teste de Kupiec e Quantil Dinâmico, isto é, esses *backtests* possuem alto poder em rejeitar a hipótese nula quando um modelo VaR realmente é mal especificado.

O capítulo a seguir será aplicado o conceito de “*gasto de alpha*” a partir da estatística de teste do MaxSPRT especificado no Capítulo 2, utilizando o pacote *Sequential* do *software R*, de modo a apresentar uma nova proposta de *backtesting* considerando os dados ao longo do tempo.

CAPÍTULO 4 - “GASTO DE ALPHA”: UMA NOVA ABORDAGEM PARA O BACKTESTING

O desenvolvimento do trabalho construído até então permite inferir que existe algumas lacunas a serem preenchidas acerca das metodologias de cálculo de alguns *backtests* que a literatura se encarrega em discutir, principalmente no que fixar os valores críticos na escala do nível de significância global, fazendo com que os *backtests* não sejam de nível α e, portanto, testes não válidos.

A proposta desse capítulo é propor uma nova abordagem de *backtesting*, através do conceito de “*gasto de alpha*” apresentado na seção 2.2.2, propondo, assim, o rompimento do paradigma que permeia os *backtests* atuais.

O novo método será executado com o uso do pacote *Sequential* do *software R*, em que será possível calcular o “*gasto de alpha*” para cada *backtesting* realizado ao longo do tempo, bem como traçar os novos valores críticos associados. A novidade dessa abordagem terá como medidas de desempenho o conceito de risco relativo (*RR*), levando-se em consideração o tempo médio de sinalização, tempo esperado de vigilância como também a comparação entre o poder dos testes.

4.1 Consolidando a função no *software R*

Para aplicar a nova abordagem de *backtesting*, primeiramente será apresentado a função do pacote *Sequential*, o qual possibilita calcular exatamente o poder, o tempo médio de sinalização, o tempo esperado de vigilância e os valores críticos (*threshold*) associados à função “*gasto de alpha*” para dados binomiais.

A função utilizada para a construção da nova proposta de *backtesting* é delineada da seguinte forma:

$$\text{Performance.AlphaSpend.Binomial}(N = 550, \alpha = 0,05, \text{AlphaSpend} = 1, \\ p = 0,01, \text{GroupSizes} = c(250, \text{rep}(10,30)), \text{Tailed} = \text{“upper”}, \\ \rho = 0,5, \text{RR} = \text{seq}(1,4), \text{Statistic} = c(\text{“MaxSPRT”}))$$

onde,

N = tamanho da amostra;

α = nível de significância desejado para o *backtesting*;

AlphaSpend = fórmula do “*gasto de alpha*”, esboçada na equação (51);

p = nível de significância desejado para o VaR;

$GroupSizes$ = o número de *backtesting* realizados ao longo do tempo;

$Tailed$ = Limite superior do *backtesting* (*threshold*);

ρ = parâmetro positivo utilizado na função “*gasto de alpha*”, equação (50);

RR = Sequência de vetor para diferentes riscos relativos;

$Statistic$ = Estatística de teste em que se origina os valores críticos, teste MaxSPRT.

O cenário em que o *backtesting* foi construído sob a abordagem do “*gasto de alpha*” segue as premissas delineadas no estudo de simulação do capítulo anterior. Imagine um VaR estimado a um nível de significância de 1%, e deseja-se realizar um *backtesting* considerando um nível de significância global de 5%. O primeiro *backtesting* será realizado depois de observados 250 retornos²⁷. O segundo *backtesting* será realizado após 10 dias, considerando agora 260 retornos. O terceiro *backtesting* será realizado após 10 dias do segundo teste, levando em conta 270 retornos e assim sucessivamente até chegar no 30º teste, completando assim 550 observações.

Aplicando a função “*gasto de alpha*”, equação (50), será possível observar o gasto da probabilidade do erro do Tipo I em cada teste ao longo do tempo. É interessante observar que cada *backtesting* terá um nível de significância menor que 5%, ou seja, o α será utilizado a medida em que os *backtests* irão acontecendo, garantindo que todos os testes sejam válidos e, portanto, de nível α .

Silva, Kulldorff e Yih (2020) ressaltam que o RR é um parâmetro muito importante na estimação do modelo, cujo valor está intrinsicamente ligado ao valor de p sob a hipótese nula. Se a hipótese nula for verdadeira, o RR será 1. Caso contrário, quanto mais a hipótese nula se distanciar de seu verdadeiro valor, maior será o RR . Para nosso caso, verificaremos um RR de até 4, quando o cálculo do VaR se distancia até 4 vezes mais da hipótese nula, para que seja possível comparar o desempenho dos quatro tipos de riscos relativos.

Para o parâmetro ρ da função “*gasto de alpha*”, equação (51), atribuiu-se o valor de 0,5 conforme utilizado por Silva (2018). O autor afirma que, para o “*gasto de alpha*” do tipo potência, um ρ igual a 0,5 favorece um equilíbrio entre o poder do teste e o tempo médio de sinalização.

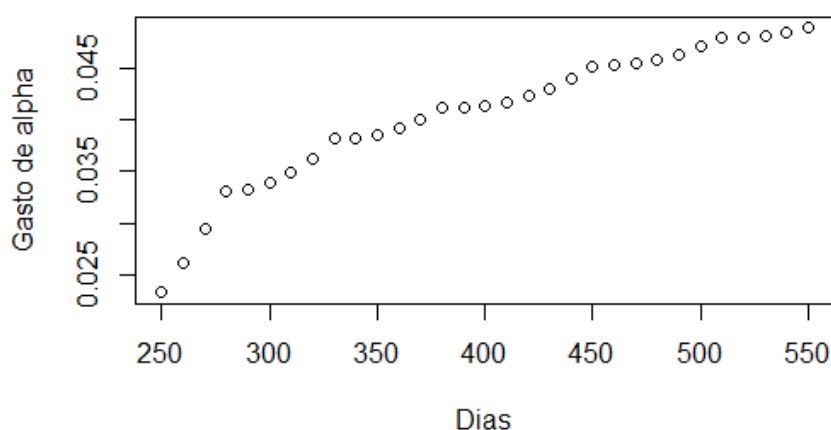
²⁷ Importante destacar que para o *backtesting* tratado nesse trabalho, 250 retornos é o mesmo que dizer 250 dias, pois a hipótese formulada estima um VaR diário. Sendo assim, quando se observa uma janela de 250 dias, tem-se 250 valores de VaR. O mesmo vale para todas as demais janelas utilizadas nesse texto.

Os valores críticos a serem calculados levam em consideração a estatística de teste do MaxSPRT já delineada na seção 2.2.1, em que uma cadeia de Markov é construída para cada estado absorvente, retornando assim valores críticos exatos para um teste de nível α .

4.1.1 Resultados e discussões

O primeiro resultado que a ser apresentado é o “*gasto de alpha*” para cada *backtesting* realizado ao longo do tempo, esboçado pela Figura 3:

Figura 3 - “Gasto de alpha” para cada *backtesting* realizado

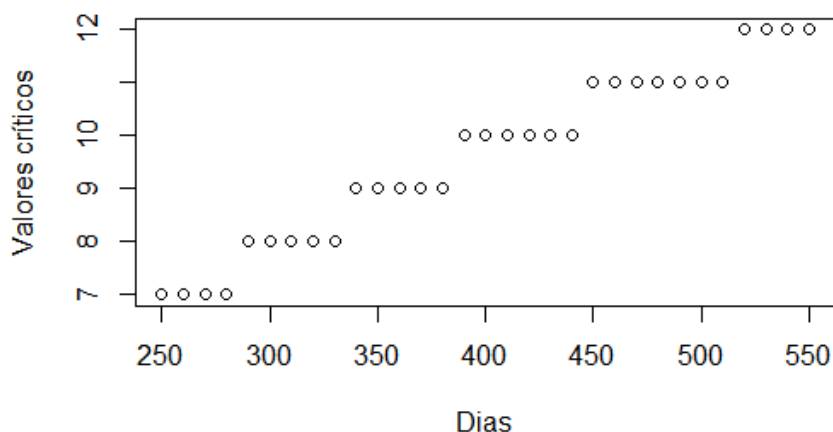


Fonte: Elaborado pelo autor a partir dos resultados

A Figura 3 evidencia que o “*gasto de alpha*” no primeiro *backtesting* realizado no dia 250 é de 0,02329, o segundo *backtesting* realizado no dia 260 é de 0,2615, no terceiro *backtesting* no dia 270 é 0,02942 e assim sucessivamente até chegar no último *backtesting* realizado no dia 550 com “*gasto de alpha*” de 0,0488. Isso garante que todos os *backtests* são de nível α , pois apresentaram valores menores que o nível de significância global estipulado de 0,05.

Posteriormente, podemos delinear os novos valores críticos na escala do número de exceções observadas, como mostra a Figura 4:

Figura 4 - Novos valores críticos



Fonte: Elaborado pelo autor a partir dos resultados

Como mostra a Figura 4, os valores críticos seguem o tipo escada, isto é, ficam constantes em determinado período de tempo e depois “saltam” conforme os testes foram sendo realizados. Nota-se que do 1º teste até o 4º, ou seja, para os *backtests* que consideraram uma janela de 250, 260, 270 e 280 dias, respectivamente, o valor crítico é de 7 exceções observadas. Isso significa que 7 exceções de VaR podem ser observadas até que a hipótese nula seja rejeitada.

O mesmo acontece do 5º ao 9º teste, ou seja, para os *backtests* que consideraram uma janela de 290, 300, 310, 320 e 330 dias, respectivamente, o valor crítico é de 8 exceções observadas.

Para o 10º *backtesting* até o 14º, considerando assim uma janela de 340, 350, 360, 370 e 380 dias, respectivamente, o valor crítico é de 9 exceções observadas.

A lógica segue sucessivamente até chegar no valor crítico máximo de 12 exceções de VaR observadas, para os *backtests* que consideraram uma janela de 520, 530, 540 e 550 dias.

O próximo resultado se refere a medida de desempenho do experimento, sumarizado na Tabela 6, apresentando os valores do RR, Poder, Tempo médio de sinalização e Tempo esperado de vigilância:

Tabela 6 - Medida de desempenho do teste sequencial com função “gasto de alpha” do tipo potência e $\rho=0.5$

RR	Poder	Tempo Médio de Sinalização	Tempo Esperado de Vigilância
1	0,009716	302,28	548,58
2	0,319472	328,68	479,97
3	0,818176	303,3	348,34
4	0,978693	271,69	277,64

Fonte: Elaborado pelo autor a partir dos resultados

A Tabela 6 oferece resultados interessantes, no que tange a análise do risco relativo. Sob o pior cenário da hipótese alternativa quando o risco relativo é igual a 4, o novo *backtesting* consegue rejeitar um modelo VaR mal especificado em 97,86% das vezes. Isso significa que se o verdadeiro valor de VaR for muito maior que a hipótese nula de 0,01, o teste tem alto poder em detectar esse modelo mal estimado e rejeitar a hipótese nula. Concomitantemente, o tempo médio de sinalização, ou seja, o tempo médio para que a hipótese nula seja rejeitada é de 271 dias. O tempo esperado de vigilância para o risco relativo igual a 4 é de 277,64, ou seja, o tamanho total da amostra no momento que o *backtesting* é interrompido é de 277 dias. Em outras palavras, o *backtesting* pode ser interrompido no 277º dia, ou melhor, no 4º teste, caso seja detectado alguma mudança no processo. Dessa forma, é possível realizar ajustes na estimação do VaR de forma rápida, não havendo a necessidade de esperar o final do monitoramento para tomar alguma decisão.

Quando a hipótese nula é realmente verdadeira, ($RR = 1$), o *backtesting* tem baixa probabilidade de rejeitar H_0 , indicando que não há problemas na estimação do modelo de risco e que não há indícios de rejeição da hipótese nula. À medida que o risco relativo aumenta, o poder do teste também aumenta, minimizando o tempo médio de sinalização e o tempo esperado de vigilância.

Considerações sobre o capítulo

Este capítulo teve a finalidade de propor uma nova abordagem para o *backtesting* do VaR a partir da abordagem do “gasto de alpha”. Utilizando o pacote *Sequential* do software R, foi possível calcular valores críticos exatos na escala do número de exceções observadas, associados ao gasto da probabilidade do erro do Tipo I em cada *backtesting* realizado.

Primeiramente foi calculado o “gasto de alpha” para cada *backtesting* à medida que os dados iam se acumulando longo do tempo. O resultado mostra que o primeiro teste tem um

“*gasto de alpha*” pequeno e conforme os dados vão se acumulando no tempo, o “*gasto de alpha*” vai aumentando até chegar no limite máximo de 0,0488. Isso significa que todos os testes são válidos e de nível α , pois o tamanho do teste foi menor do que o nível de significância global desejado de 0,05.

Posteriormente, foi apresentado os valores críticos exatos na escala do número de exceções observadas. A título de comparação, tomando os testes de Kupiec, Christoffersen e DQ teste como referência, o valor crítico calculado para esses testes considerando uma janela de 250 dias era de no máximo 2 exceções observadas, obedecendo o nível de significância de 1% utilizado no cálculo do VaR. Na abordagem do “*gasto de alpha*”, para uma janela de 250 dias, podem ser observadas até 7 exceções sem que a hipótese nula seja rejeitada. Isso corrobora com a hipótese de que os testes de Kupiec, Christoffersen e DQ teste podem rejeitar um modelo VaR mesmo que esse esteja bem ajustado, isto é, esses *backtests* podem rejeitar um modelo VaR cedo demais.

Essa situação ilustra que o modelo VaR na realidade pode estar bem especificado, porém o *backtesting* utilizado para o teste de aderência do modelo de risco pode estar sendo construído de maneira equivocada, atrapalhando assim a tomada de decisão por parte do analista.

Por fim, foi especificado a função poder do novo *backtesting* constatando que a abordagem do “*gasto de alpha*” tem alto poder em rejeitar a hipótese nula quando o risco relativo relacionado ao VaR for muito alto, minimizando também o tempo médio de sinalização e o tempo esperado de vigilância.

CONSIDERAÇÕES FINAIS

O objetivo dessa dissertação foi propor uma mudança de paradigma acerca do cálculo do *backtesting* do *Value-at-Risk* e apresentar uma nova proposta para o *backtesting* utilizando-se da abordagem do “*gasto de alpha*”. A mudança de paradigma proposta se debruça em duas frentes principais: apresentar o contexto estatístico sobre o qual o *backtesting* é construído e salientar sobre o cenário em que o novo *backtesting* poderia ser arquitetado.

Sendo assim, o Capítulo 1 encarregou-se em trazer os conceitos fundamentais e estudos empíricos sobre o VaR e o *backtesting*. As normas internacionais exigem que as instituições financeiras expostas ao risco de mercado calculem o VaR como medida de risco, bem como o *backtesting* para verificar a aderência do modelo.

Para elucidar esses conceitos, foram abordadas discussões sobre os principais modelos de VaR utilizados na literatura, a fim de tornar claro que existem inúmeras metodologias de cálculo do VaR. A partir dos estudos apresentados, foi possível observar que a literatura emprega abordagens paramétricas, semi-paramétricas e não paramétricas para a mensuração do modelo de risco e que não há uma abordagem a qual se sobressai mais do que outra, existe na verdade uma metodologia de VaR mais ajustada para determinadas séries temporais financeiras de acordo com suas especificidades.

No que se refere ao *backtesting*, foi apresentado os principais testes abordados na literatura empírica, em que foi demonstrado os pressupostos matemáticos, bem como suas estatísticas de teste. A partir do estudo bibliográfico levantado, pode-se observar que os *backtests* mais utilizados são os testes de Kupiec, Christoffersen e Quantil Dinâmico.

O teste de Kupiec é o mais difundido e o mais simples de ser aplicado, cujo método de avaliação consiste em verificar se a proporção de falhas observadas no VaR corresponde ao nível de significância estipulado. Apesar de existir algumas limitações como baixo poder estatístico para amostras pequenas e não considerar a hipótese de independência das exceções, esse é o método mais utilizado e que ainda apresenta resultados satisfatórios como mostra alguns estudos empíricos.

O teste de Christoffersen é construído sobre a hipótese de independência das exceções do VaR, diferentemente do teste de Kupiec o qual não considera esse tipo de hipótese. A partir da construção de uma cadeia de Markov binária, o teste verifica se uma exceção observada no tempo t está relacionada com uma exceção no tempo $t - 1$.

O teste do Quantil Dinâmico se utiliza de uma abordagem distinta do que fora proposto no teste de Kupiec e Christoffersen. O DQ teste baseia-se em uma regressão linear cujos

parâmetros incluem uma variável relacionada à exceção observada do VaR, como também todo o conjunto de informações disponíveis no tempo $t-1$. Arquitetado para avaliar a aderência da metodologia CAViaR, o DQ teste pode ser aplicado para qualquer cálculo de VaR, muitas vezes apresentando melhor desempenho se comparado ao teste de Kupiec e Christoffersen por utilizar de mais informações para se rejeitar a hipótese a nula.

Após apresentar os conceitos fundamentais os quais transpassam o VaR e o *backtesting*, o Capítulo 2 buscou fazer um paralelo entre as duas abordagens estatísticas que o *backtesting* poderia ser idealizado, fundamentando o discurso acerca da ruptura de paradigma. Dessa forma, foi levantado a comparação entre as áreas de controle estatístico de qualidade e a análise sequencial estatística, ambos campos capazes de fornecer mecanismos para o monitoramento de processos e que a literatura ainda não se incumbiu de levantar esse debate.

No controle estatístico de qualidade, o intuito é monitorar um processo que esteja sob controle, detectando de forma diligente qualquer alteração no processo, isto é, qualquer mudança na média e variância. Para tanto, é utilizado os gráficos de controle, os quais se utilizam de limites inferiores e superiores para monitorar um procedimento e caso o parâmetro de interesse ultrapasse qualquer desses limites, o processo é interrompido. Como medida de desempenho utilizam do *ARL*, ou comprimento médio de corrida, traduzido como o tamanho médio de observações esperadas até que algum sinal de alteração ocorra.

De forma similar, a análise estatística sequencial se utiliza do monitoramento contínuo do processo, considerando as informações no momento em que elas vão chegando no tempo. O objetivo é avaliar se a verdadeira média e variância do processo são as mesmas desde o início do monitoramento. A medida de desempenho utilizada pela análise sequencial é o tempo médio de sinalização quando a hipótese nula é rejeitada, e o tempo esperado de vigilância, similar a ideia do *ARL* dos gráficos de controle. Diferentemente dos gráficos de controle, o processo não necessariamente precisa estar sob controle.

No que tange o escopo do *backtesting*, os gráficos de controle assumem valores críticos associados ao nível de significância considerando apenas cada teste em particular, sem levar em consideração a chegada dos dados ao longo do tempo. No entanto, a análise sequencial encara essa concepção como equivocada, uma vez que valores críticos associados ao nível de significância sem levar em consideração os testes antecedentes, pode incorrer na inflação da probabilidade do Erro do Tipo I. Visto isso, com o intuito de verificar essa premissa da análise sequencial, foi realizado demonstrações teóricas bem como um estudo de simulação no Capítulo 3, de forma a ilustrar o problema da inflação da probabilidade do erro do Tipo I.

Sendo assim, o Capítulo 3 demonstrou para testes binomiais que a probabilidade do erro do Tipo I pode ser maior que o nível de significância desejado, na perspectiva de verificar cada teste individualmente, sem levar em consideração os testes anteriores. Esse problema foi evidenciado no estudo de simulações, em que um *backtesting* aplicado sucessivas vezes ao longo do tempo apresentou um tamanho superior ao nível de significância desejado, mostrando que o *backtesting* não é de nível α e, portanto, um teste não válido. Também foi possível verificar que os testes de Kupiec e Quantil Dinâmico apresentaram maior poder em rejeitar a hipótese nula quando essa realmente for falsa, isto é, denotam maior poder em identificar um VaR mal especificado. Surpreendentemente, o teste de Kupiec apresentou desempenho superior ao teste de Christoffersen, mesmo que esse último formula hipóteses mais robustas do que o teste de Kupiec.

De forma a cumprir com todos os objetivos traçados por essa pesquisa, mas sem a intenção de esgotar o tema, foi proposto no Capítulo 4 uma nova abordagem de *backtesting* do VaR, utilizando-se o conceito do “*gasto de alpha*”, pertencente ao campo da análise estatística sequencial. O novo modelo permite construir valores críticos exatos associados ao nível de significância global de um *backtesting*, sem deixar de considerar os testes antecedentes. Reforçando assim, o discurso sobre a mudança de paradigma sobre o qual o *backtesting* está inserido, de forma a evidenciar que é possível fornecer um teste de monitoramento contínuo com alto poder em rejeitar um modelo VaR que possui um risco relativo acima do que o nível de significância desejado.

Por fim, como continuidade da agenda de pesquisa, propõe-se que sejam realizadas aplicações do *backtesting* via “*gasto de alpha*” para um portfólio real, considerando diferentes metodologias de VaR, com o intuito de avaliar o poder do novo *backtesting* em identificar os modelos de risco mal especificados, analisando conjuntamente o tempo médio de sinalização.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBASI, B; GUILLEN, M. Bootstrap control charts in monitoring value at risk in insurance. **Expert Systems with Applications**. 40. 6125-6135. 2013.

ALEXANDER, C. **Market risk analysis: value-at-risk models**. Chichester, England: John Wiley & Sons, 2008.

BALI, G; MO, H; TANG, Y. The role of autoregressive conditional skewness and kurtosis in the estimation of VaR. **Journal of Banking and Finance**, 32(2), 269–282. 2008.

BANK FOR INTERNATIONAL SETTLEMENTS (BIS). **87th Annual Report**. Chapter VI - Understanding Globalization: Trade and financial openness are intertwined and the evolution of globalization, 2017. Disponível em: <https://www.bis.org/publ/arpdf/ar2017e.htm> Acesso em 25 de Março de 2019.

_____. Basel Committee on Banking Supervision. **Minimum capital requirements for market risk**. 2019. Disponível em <https://www.bis.org/bcbs/publ/d457.pdf>. Acesso em 15 de Julho de 2019.

_____. Basel Committee on Banking Supervision. **International Convergence of Capital Measurement and Capital Standards**. 1988. Disponível em <https://www.bis.org/publ/bcbs04a.pdf>. Acesso em 08 de Novembro de 2019.

_____. Basel Committee on Banking Supervision. **Revisions to the Basel II market risk framework**. 2011. Disponível em <https://www.bis.org/publ/bcbs193.pdf>. Acesso em 08 de Novembro de 2019.

_____. Basel Committee on Banking Supervision. **Supervisory Framework For The Use Of "Backtesting" In Conjunction With The Internal Models Approach To Market Risk Capital Requirements**, 1996. Disponível em: <https://www.bis.org/publ/bcbs22.pdf>. Acesso em 12 de Abril de 2019.

_____. Basel Committee on Banking Supervision. **Amendment to the Capital Accord to incorporate market risks**, 2005. Disponível em <https://www.bis.org/publ/bcbs119.pdf>. Acesso em 14 de Abril de 2019.

_____. **History of Basel Committee and Its Memberships**, 2001. Disponível em: <https://www.bis.org/publ/bcbssc101.pdf>. Acesso em 13 de Abril de 2019.

BARNARD, G. Control Charts and Stochastic Processes. **Journal of the Royal Statistical Society**. Series B (Methodological), 21(2), 239-271. 1959.

BERKOWITZ, J; CHRISTOFFERSEN, P; PELLETIER, D. Evaluating Value-at-Risk models with desk-level data. **Management Science**, 57(2), 2213–2227. 2011.

BERKOWITZ, J; O'BRIEN, J. How accurate are value-at-risk models at commercial banks? **Journal of Finance**, 57, 1093–1111. 2002.

- BILIR, H. Value at Risk (VaR) Measurement on a Diversified Portfolio: Decomposition of Idiosyncratic Risk in a Pharmaceutical Industry. *European Journal of Business and Management*. Vol 8, No 6. 35-40. 2016
- BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307–327. 1986.
- BONTEMPS, C; MEDDAHI, N. Testing Normality: A GMM Approach. *Journal of Econometrics* 124: 149–186. 2005.
- BOUCHER, C; DANIELSSON, J; KOUONCHOU, P; MAILLET, B. Risk models-at-risk. *Journal of Banking and Finance*, 44, 72–92. 2014.
- BROWNLEES, C; GALLO, G. Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, 8, 29–56. 2010.
- CAMPBELL, S. A review of backtesting and backtesting procedure. *Journal of Risk*, 9(2), 1–17. 2007.
- CANDELON, B; COLLETAZ, G; HURLIN, C; TOKPAVI, S. Backtesting value-at-risk: A GMM duration-based approach. *Journal of Financial Econometrics*, 9(2), 314–343. 2011.
- CASELLA, G; BERGER, R. *Inferência Estatística*. 2ed. São Paulo: Cengage Learning, 2014.
- CHAN, K; GRAY, P. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22(2), 283–300. 2006.
- CHAN, N; DENG, S; PENG, L; XIA; Z. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, 137(2), 556–576. 2007.
- CHEN, C; GERLACH, R; LIN, E; LEE, W. Bayesian forecasting for financial risk management, pre and post the global financial crisis. *Journal of Forecasting*, 31, 661–687. 2011.
- CHRISTOFFERSEN, P. Evaluating Interval Forecasts. *International Economic Review*, Vol. 39, No. 4, pp. 841–862. 1998.
- CHRISTOFFERSEN, P; PELLETIER, D. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, 2(1), 84–108. 2004.
- DANIELSSON, J; JORGENSEN, B; SAMORODNITSKY, G; SARMA, M; DE VRIES, C. Fat tails, VaR and subadditivity. *Journal of Econometrics*, 172, 283–291. 2013.
- DIAS, A. Market capitalization and value-at-risk. *Journal of Banking and Finance*, 37(12), 5248–5260. 2013.
- EMBRECHTS, P. Linear correlation and EVT: properties and caveats. *Journal of Financial Econometrics*, 7(1), 30–39. 2009.

- EMMER, S; KRATZ, M; TASCHE, D. What is the best risk measure in practice? A comparison of standard measures. **Journal of Risk**, 18(2), 31-60. 2015.
- ENGLE, F; MANGANELLI, S. CAViaR: Conditional autoregressive value at risk by regression quantiles. **Journal of Business and Economic Statistics**, 22(4), 367–381. 2004.
- ERGEN, I. Two-step methods in VaR prediction and the importance of fat tails. **Quantitative Finance**, 15(6), 1013–1030. 2015.
- ESCANCIANO, J; OLMO, J. Backtesting parametric value-at-risk with estimation risk. **Journal of Business and Economic Statistics**, 28(1), 36–51. 2010.
- ESCANCIANO, J; OLMO, J. Robust backtesting tests for value-at-risk models. **Journal of Financial Econometrics**, 9, 132–161. 2011.
- ESCANCIANO, J; PEI, P. Pitfalls in backtesting historical simulation VaR models. **Journal of Banking and Finance**, 36, 2233–2244. 2012.
- FRISEN, M. Optimal Sequential Surveillance for Finance, Public Health, and Other Areas. **Sequential Analysis: Design Methods and Applications**, 28:3, 310-337. 2009.
- GAGLIANONE, W; LIMA, L; LINTON, O; SMITH, D. Evaluating Value-at-Risk models via quantile regression. **Journal of Business and Economic Statistics**, 29(1), 150–160. 2011.
- GERLACH, R; WANG, C. Forecasting risk via realized GARCH, incorporating the realized range. **Quantitative Finance**, 16:4, 501-511. 2016.
- GEYER, C. Practical Markov Chain Monte Carlo. **Statistical Science**, 7(4), 473-483. 1992.
- GLASSERMAN, P; HEIDELBERGER, P; SHAHABUDDIN, P. Portfolio Value-at-Risk with Heavy-Tailed Risk Factors. **Mathematical Finance**, No 3, 239-269. 2002.
- GOURIEROUX, C; JASIAK, J. Dynamic quantile models. **Journal of Econometrics**, 147, 198–205. 2008.
- GRIGOLETTO, M; LISI, F. Looking for skewness in financial time series. **Econometrics Journal**, 12(2), 310–323. 2009.
- GRIGOLETTO, M; LISI, F. Practical implications of higher moments in risk management. **Statistical Methods and Applications**, 20, 487–506. 2011.
- HAAS, M. Improved duration-based backtesting of value-at-risk. **Journal of Risk**, 8(2), 17–38. 2005.
- HUANG, A. Y. An optimization process in value-at-risk estimation. **Review of Financial Economics**, p. 109-116, 2010.
- JENSEN, W; JONES-FARMER, A; CHAMP, C; WOODALL, W. Effects of Parameter Estimation on Control Chart Properties: A Literature Review. **Journal of Quality Technology**, 38:4, 349-364. 2006.

JENNISON, C.; TURNBULL, B. W. Group Sequential Methods with Applications to Clinical Trials. Boca Raton, London, FL: Chapman and Hall/CRC, 2000.

JORION, P. **Value at risk: The New Benchmark for Managing Financial Risk**. 3rd ed. New York: The McGraw-Hill Companies, 2006.

KING, M; WADHWANI, S. Transmission of Volatility between Stock Markets. **Review of Financial Studies**, vol 3(1), 5-33. 1990.

KOENKER, R; BASSET, G. Regression quantiles. **Econometrica**, 46, 33–50. 1978.

KOENKER, R; XIAO, Z. Quantile Autorregression. **Journal of the American Statistical Association**, 101:475, 980-990. 2006.

KOUTRAS, M; BERSIMIS, S; MARAVELAKIS, P. Statistical Process Control using Shewhart Control Charts with Supplementary Runs Rules. **Methodol Comput Appl Probab**, 9, 207-224. 2007.

KULLDORFF, M. *et al.* A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance. **Sequential Analysis**, 30:1, 58-78. 2011.

KULLDORFF, M; SILVA, I. Continuous Post-Market Sequential Safety Surveillance with Minimum Events to Signal. **Statistical Journal**. Volume 15. Number 3. pg 373-394. 2017.

KUPIEC, P. Techniques for Verifying the Accuracy of Risk Measurement Models. **The Journal of Derivatives**. Vol. 3, No. 2, pp. 73–84. 1995.

LONGIN, F. The asymptotic distribution of extreme stock market returns. **Journal of Business**, 69, 383 – 408. 1996.

LOPEZ, J. Methods for evaluating value-at-risk estimates. **Economic Policy Review**. No. August 1996, pp. 119–124.1998.

MIHAILESCU, L. A sequential method for the evaluation of the VaR model based on the run between exceedances. **Allgemeines Statistisches Archiv** 88, 51–72. 2004.

MINGOTI, S; YASSUKAWA, F. Uma comparação de gráficos de controle para a média de processos autocorrelacionados. **Revista Eletrônica Sistemas & Gestão**. v. 3, n. 1, p. 55–73. 2008.

MONTGOMERY, D; MASTRANGELO, C. Some Statistical Process Control Methods for Autocorrelated Data, **Journal of Quality Technology**, 23:3, 179-193. 1991.

MORETTIN, P. **Econometria Financeira: Um curso em séries temporais financeiras**. 2^a ed. São Paulo. 2011.

NELSON, D. Conditional heteroskedasticity in asset returns: A new approach. **Econometrica**, 59, 347–370. 1991.

- NIETO, M; RUIZ, E. Frontiers in VaR forecasting and backtesting. **International Journal of Forecasting**, 32, 475-201. 2016.
- PAGE, E. Continuous inspection schemes. **Biometrika**. vol. 41 pp. 100–115. 1954.
- PENA, V.H DE LA; RIVERA, R; MATA, J. Quality control of risk measures: Backtesting VaR models. **Journal of Risk**, 9(1), 39-54. 2006.
- PÉRIGNON, C; DENG, Y; WANG, Y. Do banks overstate their value-at-risk? **Journal of Banking and Finance**, 32, 783–794. 2008.
- PÉRIGNON, C; SMITH, D. The level and quality of value-at-risk disclosure by commercial banks. **Journal of Banking and Finance**, 34 (2), 362–377. 2010.
- PRITSKER, M. The hidden dangers of historical simulation. **Journal of Banking and Finance**, 30(2), 561–582. 2006.
- RISKMETRICS. Technical Document. J.P Morgan. 4^a Ed. Nova York, 1996.
- ROBERTS, S. Control charts based on geometric moving averages. **Technometrics** vol. 1 pp. 239–250. 1959.
- RUBIA, A; SANCHIS-MARCO, L. On downside risk predictability through liquidity and trading activity: A dynamic quantile approach. **International Journal of Forecasting**. Elsevier. vol. 29(1), 202-219. 2013
- SAMOHYL, Robert. **Controle estatístico de qualidade**. Rio de Janeiro: Elsevier, 2009.
- SCHIPPER, S; SCHMID, W. Control charts for GARCH Processes. **Nonlinear Analysis**. 2049-2060. 2001.
- SHEWHART, W. **Economic control of quality of manufactured product**. New York: Van-
Nostrand Reinhold, Princeton. 1931.
- SILVA, I. Análise Sequencial usando R para Monitoramento de Vacinas e Drogas Pós-Comercializadas. **Sigmae**, Alfenas. v.3, n.2, pg. 68-87. 2014
- SILVA, I. Type I error probability spending for post-market drug and vaccine safety surveillance with binomial data. **Statistics in Medicine**, 37:107–118. 2018.
- SILVA, I; BARROS, V. Monetary loss surveillance for credit models. **Sequential Analysis**, 35:3, 347-357. 2016.
- SILVA, I; KULLDORFF, M; YIN, W. Optimal alpha spending for sequential analysis with binomial data. **Journal of the Royal Statistical Society Series B, Royal Statistical Society**, vol. 82(4), pages 1141-1164. 2020.
- TRUCÍOS, C; TIWARI, A; ALQAHTANI, F. Value-at-risk and expected shortfall in cryptocurrencies' portfolio: a vine copula-based approach. **Applied Economics**. 2019.

VIRDI, N. A Review of backtesting Methods for Evaluating Value-at-Risk. **International Review of Business Research Papers**. Vol 7, No 4. 14-24. 2011.

WALD, A. Sequential Tests of Statistical Hypotheses. **Annals of Mathematical Statistics** 16: 117–186. 1945.

WEIB, C; TESTIK, M. Risk-based metrics for performance evaluation of control charts. **Qual Reliab Engng Int**. 35:280–291. 2019.

WONG, W. Backtesting value-at-risk based on tail losses. **Journal of Empirical Finance**, 17(3), 526–538. 2010.

ZIGGEL, D; BERENS, T; WEI, G; WIED, D. A new set of improved value-at-risk backtests. **Journal of Banking and Finance**, 14, 29–41. 2014.