

Backtesting para o Expected Shortfall do Trading Book: Uma Avaliação das Metodologias

(Backtesting for the Expected Shortfall of the Trading Book: An Assessment of Methodologies)

Leonardo Nascimento Castro*

Oswaldo Candido**

Resumo

Devido à Crise de 2008 o Comitê de Basileia acelerou o processo para atualização do Acordo e identificou algumas falhas como, por exemplo, a incapacidade do *VaR* em captar o risco de cauda. Posteriormente, recomendou-se substituir o *VaR*, uma medida não coerente de risco devido à ausência da subaditividade, pelo *CVaR*. Esta medida, apesar de não ser elicitável nem testável, é pelo menos condicionalmente elicitável e, portanto, também condicionalmente testável. Para validar os modelos do *CVaR*, foram feitas simulações com os três métodos de Acerbi e outro sugerido neste trabalho que adaptou a Aproximação dos Níveis de *VaR*. Destes quatro, nenhum foi perfeito. Os três de Acerbi tiveram falhas em alguns testes e o sugerido neste trabalho apresentou resultados melhores que o Backtesting do *VaR*.

Palavras-chave: Expected Shortfall. Elicitabilidade. Testabilidade. Backtesting.

Códigos JEL: C52, D81, G32.

Submetido em 23 de novembro de 2017. Reformulado em 12 de junho de 2018. Aceito em 12 de junho de 2018. Publicado on-line em 30 de junho de 2018. O artigo foi avaliado segundo o processo de duplo anonimato além de ser avaliado pelo editor. Editor responsável: Márcio Laurini.

* Universidade Católica de Brasília (UCB), Brasília, DF, Brasil. E-mail: leonardonc@gmail.com

**Universidade Católica de Brasília (UCB), Brasília, DF, Brasil. Este autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio. E-mail: candido.f@gmail.com

Rev. Bras. Finanças (Online), Rio de Janeiro, 16, No. 1, March 2018, pp. 123-155

ISSN 1679-0731, ISSN online 1984-5146

©2018 Sociedade Brasileira de Finanças, under a Creative Commons Attribution 3.0 license - <http://creativecommons.org/licenses/by/3.0>

Abstract

Due to the Crisis of 2008, the Basel Committee accelerated the process for update the Accord and identified some weaknesses such as the inability of VaR to capture the tail risk. Subsequently, it was recommended to substitute VaR , a non-coherent measure of risk due to the absence of sub-additivity, by $CVaR$. This measure, while not being testable or elicitable, is at least conditionally elicitable and therefore also conditionally testable. To validate the $CVaR$ models, simulations were made with the three Acerbi methods and another one suggested in this paper that adapted the Quantil Approximation. Of these four, none was perfect. The three Acerbi methods had failures in some tests and the one suggested in this paper presented better results than the VaR Backtesting.

Keywords: Elicitability, Testability, Backtesting.

1. Introdução

Após a Crise de 2008, o Comitê de Basileia acelerou o processo para atualização do Acordo e identificou algumas falhas como, por exemplo, a incapacidade do VaR em captar o risco de cauda. Posteriormente, recomendou-se substituir o VaR , uma medida não coerente de risco devido à ausência da subaditividade, pelo $CVaR$. O trabalho de Gneiting (2011) mostrou a ausência da elicibilidade para o $CVaR$ e isto induziu algumas pessoas a pensarem ser impossível realizar um backtesting para esta medida de risco.

A elicibilidade é uma propriedade matemática importante para a seleção de modelo (em escala relativa) e não para a validação (em escala absoluta). Apesar de a elicibilidade ter outra finalidade, a convexidade de sua função *scoring* é necessária para o backtesting.

Para entender melhor sobre o backtesting, torna-se necessário também um conhecimento sobre a identificabilidade e testabilidade, que possuem uma relação com a elicibilidade. De fato, o $CVaR$ não é elicível e, portanto, não-testável, porém é condicionalmente elicível, um conceito útil para previsão de algumas medidas de risco que não sejam elicíveis.

Além de alguns conceitos citados nesta seção, também torna-se necessário conhecer os termos nitidez e *ridge backtest*. Uma função é: nítida se for estritamente crescente e decrescente em relação às variáveis preditiva e realizada, respectivamente, isto é, melhora no resultado se aumentar o $CVaR$ ou reduzir a perda; e atende o re-

quisito de *ridge backtest* se depender o mínimo possível da medida de risco auxiliar VaR uma vez que o foco na regulação bancária para o requerimento de capital mínimo é o $CVaR$.

O objetivo desta pesquisa é verificar as possibilidades de submeter o $CVaR$ ao teste de aderência nas carteiras brasileiras atendendo à necessidade do Comitê de Basileia em relação ao número de exceções – perdas que ultrapassam a estimativa do VaR para o dia em análise – e magnitude das perdas e permitir que o $CVaR$ seja utilizado para calcular o requerimento de capital no Trading Book. Este artigo realiza simulações em várias situações possíveis para avaliar e analisar os quatro backtesting do $CVaR$: Z_1 e Z_{2s} de Acerbi and Szekely (2014), Z_{2c} de Acerbi and Szekely (2017) e a Aproximação dos Níveis de VaR (E^p).

Este artigo é composto por cinco seções incluindo Introdução e Considerações Finais. Nas seções 2 e 3 constam as teorias da Elicitabilidade, Identificabilidade e Testabilidade e Backtesting, respectivamente. A seção 4 dedica-se à avaliação e análise dos resultados nas simulações.

2. Elicitabilidade, Identificabilidade e Testabilidade

2.1 Elicitabilidade

Artzner et al. (1999) sugeriram o uso do $CVaR$ para resolver os problemas de coerência que o VaR apresentava no modelo de percentil (Simulação Histórica): perda além do nível de VaR desconsiderada e ausência da subaditividade. Entretanto, havia uma desvantagem que impedia a transição completa do VaR para o $CVaR$. A ausência de um backtesting convincente era o último obstáculo no caminho do $CVaR$ ao Comitê de Basileia (Acerbi; Szekely, 2014).

Enquanto isso, Gneiting (2011) contribuiu com o debate acadêmico mostrando que esta medida não é elicitável. Esta descoberta difundiu a crença (Chavez-Demoulin et al., 2016) de que o $CVaR$ não poderia ser submetido ao teste de aderência e despertou uma série de críticas à decisão do Comitê de adotá-lo no lugar do VaR .

A elicitabilidade é a propriedade de previsões serem avaliadas e classificadas. Em geral, uma medida de risco invariante a lei pega uma distribuição probabilística e transforma-a em uma previsão pontual de valor único (Wimmerstedt, 2015).

Uma estatística preditiva ψ de uma variável aleatória X é elicitável

se ela minimiza o valor esperado de uma função *scoring* S (Acerbi; Szekely, 2014).

$$\psi = \underset{m}{\operatorname{argmin}} \mathbb{E}[S(m, X)] \quad (1)$$

Dada uma história de previsões pontuais m_t para as estatísticas e realizações x_t da variável aleatória, a maneira de avaliar o modelo de previsão requer que o *scoring* médio

$$\bar{S} = \frac{1}{T} \sum_{t=1}^T S(m_t, x_t) \quad (2)$$

seja o mais baixo possível. A média e a mediana representam exemplos populares minimizando a média quadrática $S(m, X) = (m - X)^2$ e o erro absoluto $S(m, X) = |m - X|$, respectivamente. O α -quantil, e portanto o *VaR*, também é elicitável com uma função *scoring* $S(m, X) = (\mathbb{1}_{m \geq X} - \alpha)(m - X)$. Já o τ -expectil é elicitável sob a função $S(m, X) = |\mathbb{1}_{m \geq X} - \tau|(m - X)^2$.

Um exemplo da elicitabilidade para média:

$$\mathbb{E}[X] = \underset{m \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[(m - X)^2] \quad (3)$$

Para provar (3), minimiza-se o valor esperado da média quadrática derivando em relação a m em primeira ordem e igualando a zero.

$$\frac{d\mathbb{E}[(m - X)^2]}{dm} = 2m - 2\mathbb{E}[X] = 0 \quad \Rightarrow \quad m^* = \mathbb{E}[X]$$

Entretanto, existem algumas medidas que não são elicitáveis como a moda e a variância. Lambert et al. (2008) e Frongillo and Kash (2015) afirmam que esta por si só não é elicitável – devido a ausência de conjuntos de níveis convexos, uma condição necessária para a elicitabilidade. Entretanto, ela é considerada como 2-elicítavel através das duas médias.

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \underset{m_1 \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[(m_1 - X^2)^2] - \left\{ \underset{m_2 \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[(m_2 - X)^2] \right\}^2 \quad (4)$$

Ou de uma média dentro de outra.

$$\mathbb{V}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \underset{m_2 \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}\left[\left(m_2 - \left\{X - \underset{m_1 \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[(m_1 - X)^2]\right\}\right)^2\right]$$

Alguns termos encontrados na literatura para elicitabilidade de variância são 2-elicítavel (Gneiting, 2011), condicionalmente elicitável

ou elicitável de segunda ordem (Emmer et al., 2015), par (média e variância, por exemplo) elicitável e conjuntamente elicitável (Fissler; Ziegel, 2016).

Esta propriedade matemática tem importância porque permite comparações relativas de diferentes modelos que preveem uma estatística na mesma sequência exata de eventos e, através do *scoring* médio, determinar o melhor. Assim, se houver um único modelo, a função *scoring* não consegue passar uma informação sobre a qualidade de suas previsões.

Vale ressaltar que a elicitabilidade não significa exatamente o backtesting direcional – em escala absoluta, validação do modelo que mostra se este subestima ou superestima a estatística –, mas somente se trata da seleção de modelo – em escala relativa. Ainda assim, para que uma estatística possa ser testável, é necessária que ela seja elicitável.

2.1.1 Elicitabilidade do VaR

A medida de risco $VaR_\alpha(X)$ pelo método de Simulação Histórica (Quantil) é elicitável através da função *scoring*

$$S(m, X) = (\mathbb{1}_{m \geq X} - \alpha)(m - X). \quad (5)$$

De acordo com (1), isto é verdadeiro se

$$VaR_\alpha(X) = \operatorname{argmin}_{m \in \mathbb{R}} \mathbb{E} [(\mathbb{1}_{m \geq X} - \alpha)(m - X)] \quad (6)$$

O valor esperado \mathbb{E} de (6) pode ser reescrito como

$$\int (\mathbb{1}_{m \geq X} - \alpha)(m - x) f_X(X) dX = \quad (7)$$

$$(1 - \alpha) \int_{-\infty}^m (m - x) f_X(X) dX - \alpha \int_m^{\infty} (m - x) f_X(X) dX$$

Wimmerstedt (2015) derivou (7) em primeira ordem pela regra de Leibnitz e igualou a zero encontrando o resultado ótimo m^* .

$$\frac{d}{dm} \mathbb{E}[S(m, X)] = (1 - \alpha) \int_{-\infty}^m f_X(X) dX - \alpha \int_m^{\infty} f_X(X) dX =$$

$$\int_{-\infty}^m f_X(X) dX - \alpha = 0$$

$$\int_{-\infty}^m f_X(X) dX = \alpha \Rightarrow m^* = F_X^{-1}(\alpha)$$

onde $F_X^{-1}(\alpha) = VaR_\alpha(X)$ e prova-se que esta medida é elicitável através de sua função *scoring* (5).

Para a medida de risco $VaR_\alpha(X)$ pelo método de Delta-Normal, é utilizada a elicitação das duas médias na variância conforme (4).

2.1.2 Elicitabilidade do CVaR

Osband (2011) e Gneiting (2011) afirmam que o $CVaR$ não é elicítavel, apesar de o VaR o ser. Este autor provou a ausência de conjuntos de níveis convexos, que é uma condição necessária para a elicitação de uma medida de risco. Entretanto, Emmer et al. (2015) mostram que o $CVaR$ é condicionalmente elicítavel, um conceito útil para previsão de algumas medidas de risco que não sejam elicítaveis.

A elicitação condicional do $CVaR$ divide o método de previsão em dois passos: 1º) prever $\hat{q}_\alpha(X)$, que é elicítavel da mesma forma em (6) e tornar este resultado como fixo (c); e 2º) então, prever $CVaR_\alpha(X)$, que é elicítavel com função *scoring* estritamente consistente

$$S(m, X) = (\phi(X) - \phi(m) - \phi'(m)(m - X)) \cdot \mathbb{1}_{X < c} \cdot (m)$$

onde $\phi(m) = \frac{m^2}{1+|m|}$.

De uma forma mais simples no segundo passo, o $CVaR_\alpha(X)$ pode ser previsto pela elicitação da média condicional:

$$CVaR_\alpha(X) = \operatorname{argmin}_{m_2 \in \mathbb{R}} \mathbb{E} [(m_2 - X)^2 | X < m_1]$$

onde m_1 prevê o $\hat{q}_\alpha(X)$, que é o VaR . Assim, o $CVaR$, apesar de não ser elicítavel, é considerado como 2-elicítavel.

2.2 Identificabilidade

Segundo Acerbi and Szekely (2017), uma estatística ψ é identificável se tiver uma função $I(m, X)$ tal que

$$\mathbb{E} [I(m, X)] = 0 \quad \text{quando} \quad m = \psi(X)$$

A média e a mediana são identificáveis pelas funções $I(m, X) = m - X$ e $I(m, X) = 1/2 - (X - m < 0)$, respectivamente. Assim, para a média, $\mathbb{E} [m - X]$ será igual zero quando $m = \bar{X}$. Já o τ -expectil é identificável sob a função $I(m, X) = 2|\mathbb{1}_{m \leq X} - \tau|(m - X)$. Vale

ressaltar que a mediana somente é identificável quando a sua função de distribuição acumulada for contínua.

Sob certas condições de regularidade, existe uma relação entre elicitabilidade e identificabilidade (Fissler; Ziegel, 2016):

$$S(m, x) = \int^m I(t, x) dt$$

cujo valor esperado $\mathbb{E}[S(m, X)]$ tem um mínimo global em m .

2.2.1 Identificabilidade do VaR

O α -quantil, e portanto o VaR , é identificável com uma função $I(m, X) = \alpha - (X - m < 0)$, desde que a sua fda seja contínua.

Para a medida de risco $VaR_\alpha(X)$ pelo método de Delta-Normal, utiliza-se a variância que é 2-identificável através das duas médias.

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = m_1 - m_2^2$$

onde $\mathbb{E}(m_1 - X^2) = 0$ e $\mathbb{E}(m_2 - X) = 0$.

Ou de uma média dentro de outra:

$$\mathbb{V}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = m_2$$

onde $\mathbb{E}[m_2 - (X - m_1)^2] = 0$ e $\mathbb{E}(m_1 - X) = 0$.

2.2.2 Identificabilidade do CVaR

O $CVaR$, assim como na elicitabilidade, também é condicionalmente identificável pela média condicional dos retornos piores que o VaR :

$$I(m_1, m_2, X) = (X + m_1 < 0) \cdot (m_2 + X)$$

onde m_1 prevê o $\hat{q}_\alpha(X)$, que é o VaR , também identificável, e m_2 , o $CVaR$. Assim, esta medida, apesar de não ser identificável, é considerado como 2-identificável.

Para o cálculo do $CVaR$ incondicional (13):

$$I(m_1, m_2, X) = m_2 + \frac{1}{\alpha} \{X(X + m_1 < 0) - m_1[\alpha - (x + m_1 < 0)]\}$$

2.3 Testabilidade

A testabilidade é a propriedade de previsões serem classificadas em uma escala onde os valores positivos e negativos indicam superestimação e subestimação, respectivamente, e rejeitadas se estiverem muito abaixo de zero (Acerbi; Szekely, 2017). Ou seja, uma estatística $\psi(X)$ é testável se existe uma função teste $Z(m, X)$ tal que $m \mapsto \mathbb{E}[Z(m, X)]$ seja estritamente crescente e $\mathbb{E}[Z(m, X)] \leq 0$ quando $m \leq \psi(X)$.

Em caso de duas previsões que subestimem a estatística, a pior deve gerar um resultado estritamente inferior.

$$\mathbb{E}[Z(m_1, X)] \leq \mathbb{E}[Z(m_2, X)] \leq 0 \quad \text{quando} \quad m_1 \leq m_2 \leq \psi(X)$$

Segundo Wang (2016), se uma medida de risco é testável, então esta é elicitável com a função *scoring* convexa:

$$S(m, x) = \int_x^m Z(t, x) dt$$

Ou, equivalentemente:

$$\frac{dS(m, x)}{dm} = Z(m, x)$$

Acerbi and Szekely (2017) afirmam que a função testável $\mathbb{E}[Z(m, X)]$ estritamente crescente em relação à variável preditiva m garante que a função elicitável $\mathbb{E}[S(m, X)]$ é convexa e tem um mínimo global em m . Ou seja, a convexidade da função *scoring* na elicitabilidade é uma condição necessária para a testabilidade. Ainda, a função testável deve ser estritamente decrescente em relação à variável realizada $m(F)$:

$$\mathbb{E}_{F_1}[Z(m, X)] > \mathbb{E}_{F_2}[Z(m, X)] \quad \text{sse} \quad m(F_1) < m(F_2) \quad , \quad \forall m$$

Assim, a função testável é dita ser nítida se for estritamente crescente e decrescente em relação às variáveis preditiva m e realizada $m(F)$, respectivamente. Isto parece óbvio! O problema é que o Teste de Basileia para o *VaR* não é nítido. Por exemplo, na distribuição Normal ($VaR_{0,01} = 2,33$) o banco A com 4 exceções no valor de 5 em cada perda tem seu modelo de risco classificado na zona verde enquanto o banco B, 5 exceções e valor de 2,5 em cada, tem seu mo-

delo classificado na zona amarela. Por não ser nítido, a posição da estatística verdadeira é incerta.

O $CVaR$, apesar de não ser elicitável, é 2-elicitável e, portanto, pode ser 2-testável. Se for 2-identificável com uma função identificação que seja crescente e $\mathbb{E}[I(m_1, m_2, X)] \lesssim 0$ quando $m_2 \lesssim CVaR(X)$, esta função pode ser usada para a testabilidade:

$$Z(m_1, m_2, X) = I(m_1, m_2, X).$$

Da mesma forma em elicitabilidade (2), a testabilidade também pode ser usada em escala relativa em vez de absoluta pela média

$$\bar{Z} = \frac{1}{T} \sum_{t=1}^T Z(m_t, x_t).$$

Assim, o teste é também direcional indicando um valor positivo e negativo como um sinal de superestimação e subestimação, respectivamente. Quanto mais próximo de zero é melhor porque os reguladores bancários se sentem seguros e os bancos não alocam mais capital que o necessário.

Uma das medidas de risco que atende as propriedades de Invariância sob Translação, Subaditividade, Homogeneidade Positiva e Monotonicidade é o Expectil. Um dos problemas é que, como possui dependência nas duas caudas da distribuição, é possível que dois bancos com perdas equivalentes e lucros distintos alocam recursos diferentes para capital, o que não é bem visto para os reguladores. Outro é que esta medida de risco carece de uma interpretação intuitiva. Acerbi and Szekely (2017) mostram que, além de o backtesting para o quantil não ser nítido, também ocorre o mesmo problema para o expectil não fornecendo qualquer informação sobre o valor real da estatística.

3. Backtesting

Segundo Jorion (2003), “a *validação de um modelo* é o processo geral por meio do qual verifica-se se um modelo é ou não adequado”. O modelo pode ser validado através do backtesting, também chamado de teste de aderência, que é uma técnica estatística que consiste na comparação das perdas estimadas com as verificadas.

No documento consultivo *Fundamental review of the trading book: A revised market risk framework*, o Basel Committee on Banking

Supervision (2013) substituiu o VaR pelo $CVaR$ na abordagem interna devido à incapacidade de captar o risco de cauda no modelo de VaR e recomenda usar o nível de confiança de 97,5% para o $CVaR$, equivalente ao VaR de 99%.

Uma das vantagens no backtesting do $CVaR_{0,025}$ em relação ao do $VaR_{0,01}$ é que são esperadas mais exceções (6,25 contra 2,5). Quanto maior o tamanho da amostra, o teste torna-se mais forte estatisticamente.

Se o $CVaR$ não puder ser submetido ao teste de aderência, o Comitê de Basileia deve reconsiderar sua decisão sobre substituir o VaR pelo $CVaR$. Serão listados alguns métodos de avaliação dos modelos $CVaR$.

3.1 Teste de Basileia

No documento consultivo “Fundamental review of the trading book” (2012), o Comitê afirma que o backtesting baseado apenas no número de exceções de VaR é insuficiente para determinar a acurácia do modelo e recomenda um procedimento de backtesting condicional ou outras técnicas (como o tempo de exceções ou a magnitude das perdas) e propõe utilizar tanto no modelo interno quanto no padronizado o $CVaR$ por considerar o VaR como incapaz de captar o risco de cauda.

Apesar de reconhecer que a mudança do VaR para o $CVaR$ pode ter desafios operacionais, acredita que há compensação pelos benefícios do $CVaR$.

Noutro documento consultivo, “Fundamental review of the trading book: A revised market risk framework” (2013), foi reafirmada a proposta de substituir o VaR pelo $CVaR$ devido a diversas deficiências com uso do VaR para calcular o requerimento de capital. Recomendou-se o intervalo de confiança de 97,5% em relação a 99% do VaR .

Entretanto, o Comitê optou por manter o VaR de 99% e utilizar também o de 97,5%. Caso o modelo tenha mais que 12 exceções no intervalo de confiança de 99% ou 30 no de 97,5%, todas as suas posições devem ser capitalizadas pela abordagem padronizada.

3.2 Paramétrico

Existem diversos backtesting do Expected Shortfall que foram propostos por McNeil and Frey (2000), Berkowitz (2001), Kerkhof

and Melenberg (2004), Wong (2008) e Righi and Ceretta (2013). Entretanto, tais métodos exigem alguma forma de distribuição paramétrica, que não pode ser aplicada pelos reguladores. Ainda assim, os backtesting de Wong (2008) e Righi and Ceretta (2013) apresentam taxa de aceitação constante nas quantidades fixas de 1 a 10 exceções (Wimmerstedt, 2015).

3.3 Aproximação dos Níveis de VaR

3.3.1 Método de Emmer – E

Emmer et al. (2015) sugerem um método baseado na representação do $CVaR$ pelo VaR integrado (Acerbi; Tasche, 2002).

$$\begin{aligned} CVaR_\alpha(X) &= -\alpha^{-1} \int_0^\alpha q_u(X) du \\ &\approx -\frac{1}{4} [q_\alpha(X) + q_{0,75\alpha+0,0025}(X) + q_{0,5\alpha+0,005}(X) + q_{0,25\alpha+0,0075}(X)] \end{aligned} \quad (8)$$

onde $-q_\alpha(X) = VaR_\alpha(X)$.

Assim, se o VaR de cada nível for testado com sucesso, então a estimativa do $CVaR$ pode ser considerada confiável.

Este método aparentemente é uma aproximação simplista em comparação com outros de backtesting. A vantagem é que não são necessárias as simulações de Monte Carlo para a estatística de teste – o que acontece nos métodos de Acerbi citados nesta pesquisa. A desvantagem é que não há exatidão na magnitude das perdas uma vez que este método ainda baseia-se na quantidade de exceções para cada VaR e não no tamanho da perda em cada exceção.

3.3.2 Adaptação – E^p

Não há consenso a respeito de quantos e quais níveis de VaR devem ser utilizados para estimar o $CVaR$.

Caso se utilize o $CVaR$ original para calcular o requerimento de capital, este artigo considera ser inadequada a sugestão em (8) uma vez que a média dos quatro níveis de $VaR \sim N$ é subestimada em relação ao $CVaR$. Segue um exemplo $\sim N$ no nível de significância de 2,5%:

$$\widehat{CVaR}_{0,025} \approx \frac{1}{4} [1,96 + 2,03 + 2,11 + 2,20] = 2,08 < CVaR_{0,025} = 2,34$$

O $\widehat{CVaR}_{0,025}$ de Emmer et al. (2015) inferior ao $CVaR_{0,025}$ justifica-se pela inclusão de $q_\alpha(X)$ em (8). Este nível de significância não

deveria ser considerado uma vez que o $CVaR$ é o valor esperado das perdas piores – e não iguais ou piores – que o VaR conforme (10). Ainda que seja retirado $q_\alpha(X)$ e incluído $q_{0\alpha+0,01}(X)$ em (8), o $\widehat{CVaR}_{0,025}$ seria mais próximo do $CVaR_{0,025}$, mas ainda inferior:

$$\widehat{CVaR}_{0,025} \approx \frac{1}{4} [2,03 + 2,11 + 2,20 + 2,33] = 2,17 < CVaR_{0,025} = 2,34$$

Assim, o backtesting avaliaria o risco nas perdas mais leves, o que não atende o objetivo dos reguladores bancários. Se, por outro lado, a média for superior, o sucesso no backtesting não significaria que o $CVaR$ original – e menor – também fosse considerado como adequado. Neste caso deveriam retirar um nível de significância maior e/ou acrescentar um menor, a não ser que seja utilizada a média dos níveis de VaR para o requerimento de capital.

A tabela 4.1 de Wimmerstedt (2015) informa a taxa de aceitação em apenas 78% (onde o melhor desempenho é 95%) para o Backtesting Aproximação dos Níveis de VaR de Emmer et al. (2015).

Este artigo, visando o conservadorismo para os reguladores, adapta o método de Aproximação dos Níveis de VaR sem deixar de utilizar o $CVaR$ original para o requerimento de capital.

Considerando o período histórico de observações $T = 250$, são definidos os seguintes níveis com seis casas decimais para o $CVaR$ de 97,5%:

$$\widehat{CVaR}_{0,025} = \frac{1}{5} [VaR_{0,016021}(X) + VaR_{0,013214}(X) + VaR_{0,010503}(X) + VaR_{0,007913}(X) + VaR_{0,005484}(X)] \quad (9)$$

Assim, a média calculada torna-se equivalente ao $CVaR$ original nos métodos Delta-Normal, Normal Correlacionado, Simulação Monte Carlo e Quase-Monte Carlo. Na Normal(0, 1), por exemplo, a média é 2,326 (assim como o $VaR_{0,01}$) enquanto o $CVaR$ original é 2,338.

$$\begin{aligned} \widehat{CVaR}_{0,025}^{EP} &\approx \frac{1}{5} [2,14 + 2,22 + 2,31 + 2,41 + 2,54] = 2,326 \\ &= VaR_{0,01} \approx CVaR_{0,025} = 2,338 \end{aligned}$$

A dificuldade operacional deste método ocorre na Simulação Histórica, pois torna-se necessário procurar níveis em cada período histórico

Table 1Comparação entre VaR , $CVaR$, \widehat{CVaR}^E e \widehat{CVaR}^{EP}

Distribuição $c/\sigma^2 = 1$	$VaR_{0,01}$	$CVaR_{0,025}$	$\widehat{CVaR}_{0,025}^E$	$\widehat{CVaR}_{0,025}^{EP}$
N	2,33	2,34	2,08	2,33
t_{1000}	2,33	2,34	2,08	2,33
t_{100}	2,36	2,38	2,10	2,36
t_{10}	2,76	2,82	2,39	2,77

de observações para que a média dos níveis de VaR seja semelhante ao $CVaR$ original.

A tabela 1 mostra que a média dos cinco níveis de VaR em EP ficou mais próxima de $VaR_{0,01}$ e $CVaR_{0,025}$, se comparada com a média dos quatro níveis de Emmer et al. (2015).

Os intervalos de confiança selecionados em (9) foram baseados naqueles que mais se aproximassem (sem ultrapassar) de 95% na distribuição binomial (pelo número de exceções x em $n = 250$ observações e nível de significância α) acumulando a seguinte função de zero até x exceções:

$$f(x, n, \alpha) = \frac{n!}{x!(n-x)!} \alpha^x (1-\alpha)^{n-x}.$$

O Basel Committee on Banking Supervision (1996) define as zonas verde, amarela e vermelha de acordo com a distribuição acumulada da distribuição binomial $F(x, 250, \alpha)$. Na zona verde enquadra-se o modelo de VaR que tiver distribuição acumulada inferior a 95%. Na amarela, distribuição acumulada no intervalo [95%; 99,99%). E na vermelha, igual ou superior a 99,99%.

Segue a tabela para esclarecer os critérios utilizados por este artigo:

Table 2
Regras de Basileia para outros níveis de VaR

Nº de Exceções	Distribuição Acumulada						
	$\alpha = 0,018907$	$\alpha = 0,016021$	$\alpha = 0,013214$	$\alpha = 0,010503$	$\alpha = 0,007913$	$\alpha = 0,005484$	$\alpha = 0,003279$
0	0,846%	1,764%	3,595%	7,139%	13,723%	25,290%	43,995%
1	4,924%	8,943%	15,632%	26,082%	41,086%	60,153%	80,178%
2	14,707%	23,497%	35,699%	51,116%	68,259%	84,088%	94,998%
3	30,293%	43,086%	57,912%	73,082%	86,175%	94,998%	99,029%
4	48,839%	62,781%	76,280%	87,480%	94,999%	98,713%	99,847%
5	66,425%	78,557%	88,382%	94,999%	98,463%	99,721%	99,980%
6	80,263%	89,047%	94,999%	98,258%	99,590%	99,948%	99,998%
7	89,558%	94,9995%	98,088%	99,464%	99,904%	99,991%	100,000%
8	94,9998%	97,944%	99,344%	99,852%	99,980%	99,999%	100,000%
9	97,819%	99,233%	99,796%	99,963%	99,996%	100,000%	100,000%
10	99,129%	99,738%	99,942%	99,992%	99,999%	100,000%	100,000%
11	99,680%	99,918%	99,985%	99,998%	100,000%	100,000%	100,000%
12	99,891%	99,976%	99,996%	100,000%	100,000%	100,000%	100,000%
13	99,965%	99,994%	99,999%	100,000%	100,000%	100,000%	100,000%
14	99,9898%	99,998%	100,000%	100,000%	100,000%	100,000%	100,000%
15	99,997%	100,000%	100,000%	100,000%	100,000%	100,000%	100,000%

Table 3

Posição das perdas em cada situação de exemplo

Quantidade de exceções	Posição das perdas (em $\alpha\%$)
Quatro no $VaR_{0,01}$	Quatro em $[0; 1)$
Duas no $VaR_{0,005}$ e quatro no $VaR_{0,01}$	Duas em $[0; 0,5)$ e duas em $[0,5; 1)$
Duas no $VaR_{0,005}$ e duas no $VaR_{0,01}$	Duas em $[0; 0,5)$ e nenhuma em $[0,5; 1)$
Zero no $VaR_{0,004}$, uma no $VaR_{0,007}$ e três no $VaR_{0,01}$	Nenhuma em $[0; 0,4)$, uma em $[0,4; 0,7)$ e duas em $[0,7; 1)$

A aproximação de 95% corrige uma das fraquezas no Teste de Basileia, que considerou a quantidade de cinco exceções como alerta para um $VaR_{0,01}$. Na distribuição binomial: a probabilidade de obter até quatro exceções é de 89,2%; até cinco, 95,9% (mais próximo do intervalo de confiança de 95%). É injusto porque o backtesting aplica para o $VaR_{0,01}$ a mesma penalidade para o $VaR_{0,008}$. A tabela 5 demonstra que o VaR de $\alpha = 1\%$ foi prejudicado (taxa de aceitação distante de 95% para $\phi = 5\%$). Uma solução para o Teste de Basileia seria a utilização do VaR no intervalo de confiança de 99,2% (transferindo uma exceção da zona amarela para vermelha) ou de 98,9% (uma da zona vermelha para amarela e uma desta para verde).

Quanto maior a quantidade de níveis do VaR , menor é a margem de erro no backtesting. A tabela 3 exemplifica como a posição das perdas se torna mais precisa conforme aumenta-se a quantidade de níveis.

Apesar de ser ideal o aumento na quantidade de níveis – incluindo $\alpha = 0,018907$ –, seria necessária a inclusão do outro, mais pesado ($\alpha = 0,003279$), para que o \widehat{CVaR} não se distancie do $CVaR$. Porém, este nível é inferior a $1/250$ (0,004) e pode não se enquadrar em algum modelo de Simulação Histórica, que utiliza o percentil.

Os níveis em (9) podem ser alterados desde que sua média não esteja distante do $CVaR$ original. Caso seja acrescentado um nível de significância alto, também terá de ser acrescentado um de baixo para a média dos níveis de VaR não se distanciar do $CVaR$ original.

Após o cálculo de cada VaR , calcula-se o coeficiente angular do $CVaR$ no \widehat{CVaR} . Se esta medida estatística não estiver no intervalo $[0,95; 1,05]$, define-se outra quantidade e valores de níveis.

Para submeter ao teste de aderência para cada nível de VaR , utiliza-se uma regra parecida com a do Teste de Basileia conforme a tabela 2.

Permite-se a transferência da(s) exceção(ões) de um nível para outro sem prejuízo de conservadorismo. Por exemplo, seja $VaR_A =$

$VaR_{0,005484}$, $VaR_B = VaR_{0,010503}$ e $VaR_C = VaR_{0,016021}$. Se o VaR_A obteve apenas duas exceções, acrescenta-se uma neste e retira-se uma do VaR_B . Se o VaR_A obteve quatro exceções, uma acima da zona verde, retira-se uma deste e acrescenta-se uma no VaR_B e outra no VaR_C , ou duas no VaR_C . O VaR de cada nível deve estar na zona verde para que o $CVaR$ possa ser considerado como adequado. Caso contrário, repete-se o procedimento entre as zonas amarela e vermelha da tabela 2.

Assim, desde que o coeficiente angular esteja dentro do intervalo $[0,95; 1,05]$ e que o VaR de cada nível, após a transferência de exceções, não se rejeite na zona verde, o modelo de $CVaR_\alpha$ considera-se como acurado.

Caso os reguladores bancários optem por utilizar a média dos níveis de VaR em vez do $CVaR$ para o requerimento de capital, não é necessário verificar o coeficiente angular nem há problema na Simulação Histórica.

3.4 Testando o CVaR Após o VaR

3.4.1 Método de Acerbi – Z_1

Este método é baseado na média das perdas condicionadas ao VaR :

$$CVaR_\alpha(X) = -\mathbb{E}[X|X < -VaR_\alpha(X)] \quad (10)$$

Então é esperado que:

$$CVaR_{\alpha,t} = -\mathbb{E}(X_t|X_t < -VaR_{\alpha,t})$$

$$\mathbb{E}\left[\frac{X_t}{CVaR_{\alpha,t}} + 1 \mid X_t < -VaR_{\alpha,t}\right] = 0 \quad (11)$$

onde X_t é uma variável aleatória representando o retorno no tempo t . Se o $VaR_{\alpha,t}$ foi testado com sucesso, Acerbi and Szekely (2014) sugerem realizar estatística de teste baseado em (11) para testar a magnitude das exceções realizadas contra o modelo previsto:

$$Z_1(\vec{X}) = \frac{\sum_{t=1}^T \frac{X_t I_t}{CVaR_{\alpha,t}}}{N} + 1$$

onde $I_t = \mathbb{1}_{X_t < -VaR_{\alpha,t}}$ desde que o número de exceções $N = \sum_{t=1}^T I_t > 0$.

É testada a hipótese nula

$$H_0 : P_t^\alpha = F_t^\alpha, \quad \forall t$$

onde

$P_t(x)$ é o retorno da distribuição prevista;

$F_t(x)$ é o retorno da distribuição observada; e

$P_t^\alpha(x) = \min\left(1, \frac{P_t(x)}{\alpha}\right)$ é a cauda da distribuição condicionada aos retornos piores que o quantil α contra a alternativa

$$H_1 : \begin{aligned} CVaR_{\alpha,t}^F &\geq CVaR_{\alpha,t}, & \forall t \text{ e } > \exists t \\ VaR_{\alpha,t}^F &= VaR_{\alpha,t}, & \forall t. \end{aligned}$$

É esperado que o valor de Z_1 seja zero (não se rejeita o $CVaR$), isto é, assume-se que a cauda realizada é a mesma que a cauda prevista, ou negativo (rejeita-se o $CVaR$ sem rejeitar o VaR) evidenciando-se a subestimação do risco. O $VaR_{\alpha,t}$ é aceito na hipótese alternativa porque este Teste é subordinado ao backtesting do VaR e avalia tão somente a magnitude da perda em cada exceção.

Para testar a significância neste método são feitas simulações independentes a partir da distribuição em H_0 :

$$\begin{aligned} \text{simular} & \quad X_t^i \sim P_t & \forall t, \forall i = 1, 2, \dots, M \\ \text{computar} & \quad Z^i = Z(\vec{X}^i) \\ \text{estimar} & \quad p = \sum_{i=1}^M (Z^i < Z(\vec{x})) / M \end{aligned} \quad (12)$$

onde M é o número de cenários para simulações independentes. Dado um nível de significância ϕ , o Teste é aceito ou rejeitado se $p \geq \phi$.

3.5 Testando Diretamente o CVaR

3.5.1 Método de Acerbi – Z_{2s}

Este método, também proposto por Acerbi and Szekely (2014), é baseado na média incondicional:

$$CVaR_\alpha(X) = -\frac{1}{\alpha} \mathbb{E} \{X (X + q < 0) - q [\alpha - (X + q < 0)]\} \quad (13)$$

onde $q = VaR_\alpha(X)$.

Porém, a (13) foi simplificada retirando-se o termo $q[\alpha - (X + q < 0)]$:

$$CVaR_{\alpha,t} = -\frac{1}{\alpha} \mathbb{E} [X (X + q < 0)] = -\mathbb{E} \left(\frac{X_t I_t}{\alpha} \right) \quad (14)$$

$$\mathbb{E} \left[\frac{X_t I_t}{\alpha CVaR_{\alpha,t}} + 1 \right] = 0$$

Define-se a estatística de teste

$$Z_{2s}(\tilde{X}) = \sum_{t=1}^T \frac{X_t I_t}{T \alpha CVaR_{\alpha,t}} + 1$$

Enquanto o Teste de Z_1 avalia somente a magnitude das exceções, o de Z_{2s} avalia tanto a magnitude quanto a frequência das exceções e possui uma relação com Z_1 :

$$Z_{2s} = 1 - (1 - Z_1) \frac{N}{T\alpha}$$

com $\mathbb{E}_{H_0}[N] = T\alpha$.

As hipóteses são

$$H_0 : P_t^\alpha = F_t^\alpha, \quad \forall t$$

$$H_1 : \begin{matrix} CVaR_{\alpha,t}^F \geq CVaR_{\alpha,t}, & \forall t \text{ e } > \exists t \\ VaR_{\alpha,t}^F \geq VaR_{\alpha,t}, & \forall t. \end{matrix}$$

Assim como no Teste de Z_1 , também é esperado que $\mathbb{E}_{H_0}[Z_{2s}] = 0$ e $\mathbb{E}_{H_1}[Z_{2s}] < 0$ e utiliza-se o mesmo procedimento em (12) para testar a significância.

3.5.2 Método Corrigido de Acerbi – Z_{2c}

O Teste de Z_{2s} tem uma crítica (Acerbi; Szekely, 2017) na qual o seu valor esperado depende muito do VaR e não somente do $CVaR$. Isto é um problema porque há possibilidade de os bancos manipularem o valor do VaR para melhorar o Backtesting do $CVaR$ – o desempenho indesejado encontra-se na figura 1. O próprio autor afirma que não foi uma boa ideia simplificar em (14) e sugere substituir esta expressão por uma citada em Acerbi and Tasche (2001), que é a mesma da (13):

$$CVaR_{\alpha,t} = -\frac{1}{\alpha} \mathbb{E} [X_t I_t - VaR_{\alpha,t} (\alpha - I_t)]$$

$$= VaR_{\alpha,t} - \frac{1}{\alpha} \mathbb{E} [(X_t + VaR_{\alpha,t}) I_t]$$

A função 2-identificável, e também 2-testável, do $CVaR$ passa a ser:

$$Z(v, e, x) = I(v, e, x) = \alpha(e - v) + (x + v)(x + v < 0) \quad (15)$$

Assim, tem-se a estatística de teste dada por Acerbi (2016)

$$Z_{2c}(\vec{X}) = \sum_{t=1}^T \frac{Z(v_t, e_t, x_t)}{T\alpha e_t}$$

e as hipóteses são

$$H_0 : P_t^\alpha = F_t^\alpha, \quad \forall t$$

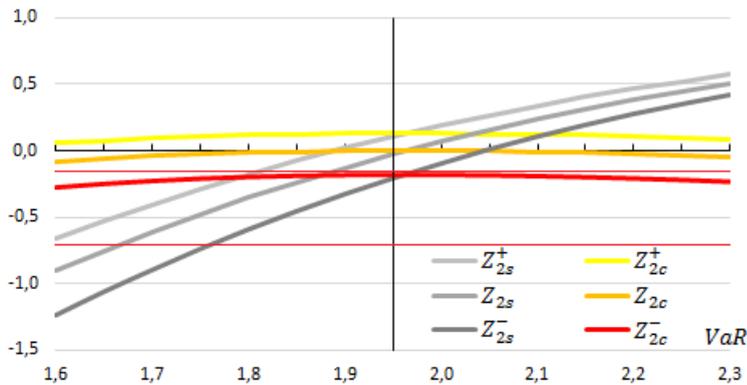
$$H_1 : \begin{aligned} & CVaR_{\alpha,t}^F \geq CVaR_{\alpha,t}, \quad \forall t \text{ e } > \exists t \\ & VaR_{\alpha,t}^F \sim VaR_{\alpha,t}. \end{aligned}$$

Espera-se que $\mathbb{E}_{H_0}[Z_{2c}] = 0$ e $\mathbb{E}_{H_1}[Z_{2c}] < 0$ e utiliza-se o mesmo procedimento em (12) para testar a significância.

Observa-se na figura 1 que, para as simulações com o VaR subestimado, o Teste Z_{2s} rejeita qualquer valor do $CVaR$, mesmo que este seja superestimado. Assim como aceita o $CVaR$ subestimado quando o VaR é superestimado. O Teste Z_{2c} corrige as falhas do Z_{2s} aceitando e rejeitando os $CVaR$ superestimado e subestimado, respectivamente, para qualquer valor do VaR .

Figure 1

Valores dos Testes Z_{2s} e $Z_{2c} \sim N(0,1)$ com $CVaR$ subestimado (-), preciso e superestimado (+), alterações do VaR e $\phi = 5\%$



Região de rejeição: abaixo de $-0,7$ para Z_{2s} e $-0,16$ para Z_{2c} conforme a tabela 4

Nas simulações com o $CVaR$ preciso e o VaR impreciso, a estatística de teste do Z_{2c} mostra um sinal negativo, apesar de ser pequeno. É o que Acerbi and Szekely (2017) chamam de *Ridge Backtest*: quando alguém está no topo de uma montanha e se perde em qualquer caminho, isto significa que ele se encontra abaixo de onde deveria estar. Isto é bom na regulação bancária porque está sendo adotado o $CVaR$ como uma medida de risco para calcular o requerimento de capital e não permite uma manipulação do VaR para melhorar os backtesting do $CVaR$. Portanto, é fundamental que sejam otimizados os resultados do Backtesting do VaR .

4. Análise dos Resultados

Nesta seção são avaliados e analisados os três Backtesting de Acerbi (Z_1 : testando o $CVaR$ após o VaR ; Z_{2s} : testando diretamente o $CVaR$; e Z_{2c} : correção do Z_{2s}) e Aproximação dos Níveis de VaR de Emmer et al. (2015) adaptado por este artigo (E^p). Em alguns momentos é incluído o Backtesting do $VaR_{0,01}$ para comparação.

Os níveis de significância são de 5% e 0,01%, que representam as zonas verde e amarela, respectivamente, do Teste de Basileia conforme Basel Committee on Banking Supervision (1996). Os gráficos com $\phi = 0,01\%$ foram omitidos porque mostram resultados próximos aos com $\phi = 5\%$.

Para calcular a estatística de teste em cada distribuição (t_3 , t_6 , t_9 , t_{100} e N) na tabela 4, foi utilizado o seguinte procedimento:

1. Gera-se uma série de 250 retornos (X), que são as funções inversas da distribuição dos números aleatórios $(0, 1)$.
2. Calcula-se o VaR , que é a função inversa da distribuição (nível de significância $\alpha = 2,5\%$) com variância 1 e média zero.
3. Calcula-se o $CVaR$ para distribuições Normal e t de Student conforme as fórmulas em (16) e (17), respectivamente, também com variância 1 e média zero:

$$CVaR = \frac{\phi [\Phi^{-1}(\alpha)]}{\alpha} \quad (16)$$

$$CVaR = \frac{g_\nu [t_\nu^{-1}(1 - \alpha)]}{\alpha} \cdot \frac{\nu + [t_\nu^{-1}(\alpha)]^2}{\nu - 1} \quad (17)$$

onde

ϕ é a função densidade de probabilidade normal padronizada,

$\Phi^{-1}(\alpha)$ é a função inversa da distribuição normal padronizada,

ν é o número de graus de liberdades,

g_ν é a função densidade de probabilidade t de Student e

$t_\nu^{-1}(1 - \alpha)$ é a função inversa da distribuição t de Student.

4. Calcula-se a estatística de teste para Z_1 , Z_{2s} e Z_{2c} conforme as subseções 3.4 e 3.5.
5. Após 10 milhões simulações conforme os passos acima, retira-se o percentil 5% (assim como 0,01%). Será rejeitada aquela estatística de teste que ficar abaixo do percentil calculado.
6. Repetem-se os passos 1-5 acrescentando e subtraindo em X , VaR e $CVaR$ a média (que era zero) por 1,5. Os resultados mostram que é necessário incluir μ nos passos 2-3.

O Teste Z_{2s} apresenta resultados semelhantes em tabela 4 de Acerbi and Szekely (2014). Coincidência ou não, em uma situação na qual os retornos $\sim N$ piores que o VaR sejam idênticos ao $CVaR$, o valor $-0,7$ da estatística deste Teste fica entre 10 e 11 exceções, exatamente como no Teste de Basileia ($\phi = 2,5\%$) entre as zonas verde e amarela. E o valor $-1,82$ fica entre 17 e 18 exceções, próximo ao Teste de Basileia (entre 16 e 17) para as zonas amarela e vermelha.

A partir daqui, a taxa de aceitação será o percentual das estatísticas de teste que ficarem iguais ou acima do valor mínimo descrito na tabela 4. Na situação de VaR e $CVaR$ precisos, quanto mais próxima de 95% e 99,99% para os níveis de significância $\phi = 5\%$ e 0,01%, respectivamente, melhor o desempenho. Em outras situações, a taxa de aceitação deve ser inferior para o $CVaR$ subestimado ou superior para superestimado.

A tabela 5 mostra que o poder de teste obteve o mesmo resultado em todas as distribuições. Os Testes Z_1 , Z_{2s} e Z_{2c} apresentaram resultados conforme esperado nos níveis de significância de 5% e 0,01%.

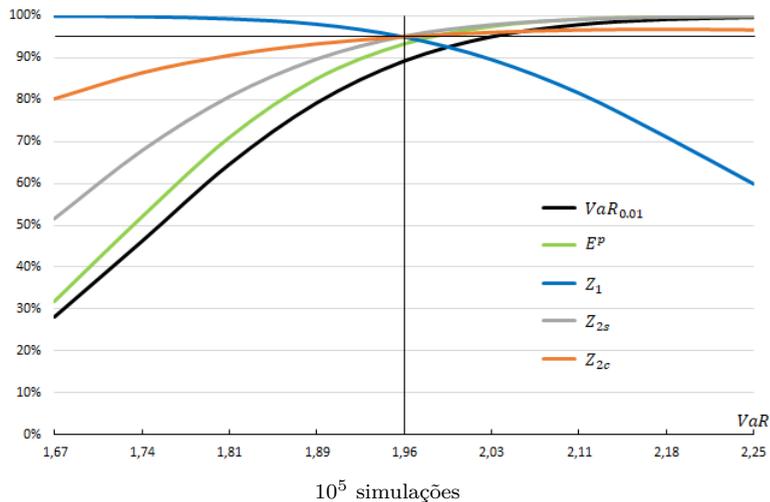
Table 4
Valores dos Testes para o $CVaR_{0,025}$

Teste	Distribuição $c/\sigma^2 = 1$	$\phi = 5\%$			$\phi = 0,01\%$		
		$\mu = -1,5$	$\mu = 0$	$\mu = +1,5$	$\mu = -1,5$	$\mu = 0$	$\mu = +1,5$
Z_1	t_3	0,33	0,43	0,61	4,13	5,36	7,63
	t_6	0,15	0,22	0,42	1,03	1,50	2,78
	t_9	0,12	0,18	0,37	0,67	1,02	2,12
	t_{100}	0,07	0,12	0,32	0,32	0,52	1,41
	N	0,07	0,11	0,32	0,30	0,50	1,39
Z_{2s}	t_3	0,77	0,82	0,92	3,76	4,80	6,57
	t_6	0,71	0,72	0,80	1,82	1,90	2,33
	t_9	0,70	0,71	0,78	1,80	1,83	2,11
	t_{100}	0,70	0,70	0,76	1,80	1,82	1,99
	N	0,70	0,70	0,76	1,79	1,82	1,99
Z_{2c}	t_3	0,38	0,50	0,71	3,36	4,36	6,20
	t_6	0,19	0,28	0,52	0,71	1,04	1,93
	t_9	0,15	0,23	0,48	0,47	0,72	1,50
	t_{100}	0,10	0,16	0,44	0,28	0,46	1,23
	N	0,10	0,16	0,44	0,27	0,44	1,23

Table 5
Taxa de aceitação nos Backtesting do $CVaR_{0,025}$

Teste	$\phi = 5\%$					$\phi = 0,01\%$				
	t_3	t_6	t_9	t_{100}	N	t_3	t_6	t_9	t_{100}	N
$VaR_{0,01}$	89,243	89,225	89,237	89,235	89,264	99,973	99,974	99,973	99,974	99,976
E^p	93,350	93,360	93,351	93,372	93,376	99,980	99,979	99,981	99,979	99,979
Z_1	95,000	95,000	95,003	95,002	95,049	99,990	99,990	99,990	99,990	99,991
Z_{2s}	95,000	95,001	95,000	94,997	95,025	99,990	99,989	99,990	99,990	99,986
Z_{2c}	95,000	95,000	95,002	94,999	95,016	99,990	99,990	99,990	99,990	99,984

10^6 simulações para o $VaR_{0,01}$ e E^p e 10^7 para os demais. Em %

Figure 2Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com alterações do VaR e $\phi = 5\%$ 

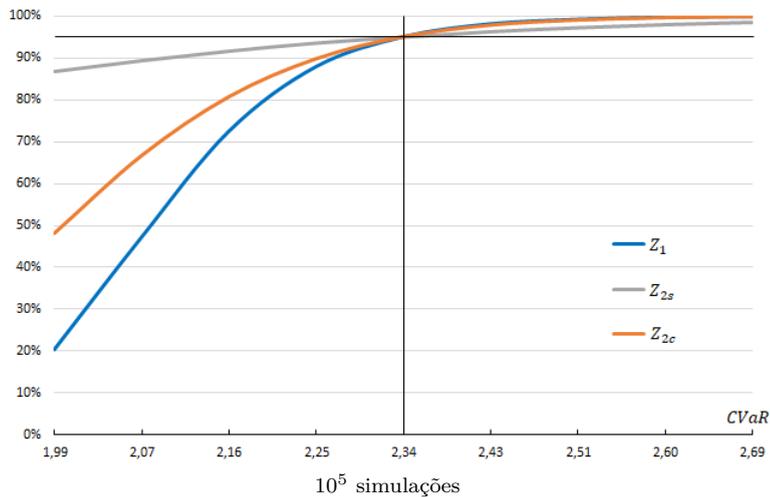
O $VaR_{0,01}$ no Teste de Basileia obteve apenas 89,2% em aceitação, inferior ao limite da zona verde (95%). Isto porque, na distribuição binomial, a probabilidade de obter até quatro exceções é de 89,2% conforme a tabela 2 de Basel Committee on Banking Supervision (1996) evidenciando-se uma das fraquezas no VaR . Até cinco, é de 95,9%, acima do intervalo de confiança. Da mesma forma na zona amarela: 99,97% para nove exceções contra 99,995% para dez.

A tabela 4.1 de Wimmerstedt (2015) informa taxas de aceitação entre 94% e 96% para Z_1 e Z_{2s} , que estão de acordo com este artigo em $\phi = 5\%$. Entretanto, para o Backtesting Aproximação dos Níveis de VaR de Emmer et al. (2015), a taxa é de apenas 78% enquanto este artigo mostra 93,4% para o Teste de Emmer adaptado (E^p). Dois fatores explicam esta diferença: 1) foram procurados os intervalos de confiança, em até seis casas decimais, que se aproximam de 95% na distribuição binomial; e 2) permitiu-se a transferência de exceção(ões) de um nível para outro sem prejuízo de conservadorismo. Com $\phi = 0,01\%$, esta adaptação obteve 99,97% em taxa de aceitação, semelhante ao $VaR_{0,01}$.

A figura 2 mostra um comportamento inesperado em Z_1 . Conforme aumenta-se o valor do VaR , este Teste passa a aceitar menos os modelos de risco. O motivo está na distribuição de probabili-

Figure 3

Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com alterações do $CVaR$ e $\phi = 5\%$



dade abaixo do nível de significância. Por exemplo, com o VaR e $CVaR \sim N(0,1)$ precisos ($= 1,96$ e $2,34$, respectivamente), tem-se $\alpha[VaR] = 0,025$ e $\alpha[CVaR] \approx 0,01$. Assim, quando ocorre uma exceção, existe a probabilidade de 61% de uma variável aleatória ficar entre 0,025 e 0,01 em α . Com o VaR menor ($= 1,67$) sem alteração do $CVaR$, o nível $\alpha[VaR]$ passa a ser maior ($\approx 0,048$) e a probabilidade de aceitação também aumenta. Assim como esta diminui quando o VaR é superestimado.

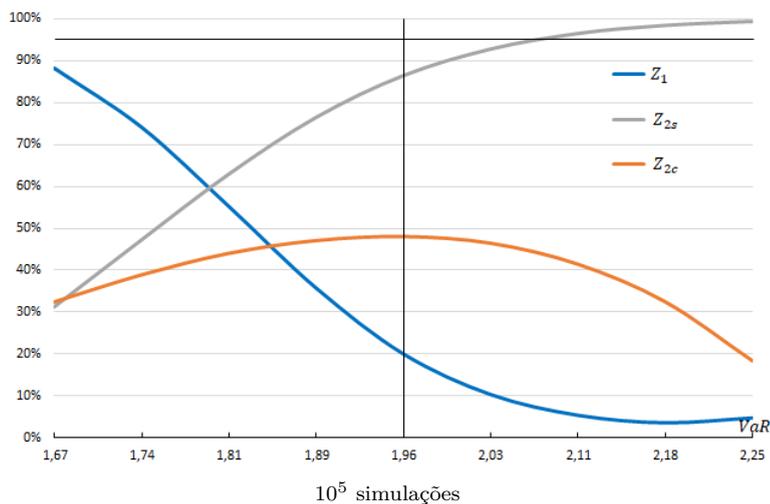
Nas simulações com o VaR preciso, os Testes EP , Z_{2s} e Z_{2c} tiveram taxas de aceitação próximas ao intervalo de confiança enquanto o $VaR_{0,01}$ ficou aquém de 95%.

O que menos dependeu do VaR foi o Z_{2c} tanto nos valores inferiores ao preciso quanto nos superiores. Isto significa que este Teste apresentou o melhor desempenho uma vez que o foco é o $CVaR$. Em outros Testes, existe a possibilidade de os bancos manipularem reduzindo o VaR em Z_1 ou aumentando em Z_{2s} para ter mais chances no backtesting do $CVaR$. As figuras 4, 5 e 6 também mostram esta possibilidade.

No Teste Z_{2c} , que teve uma leve melhora quando o VaR é superestimado, a diferença entre a taxa de aceitação e 95% não é significativa quando o VaR aproxima-se do $CVaR$.

Figure 4

Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com $CVaR$ subestimado ($= 1,99$), alterações do VaR e $\phi = 5\%$



A figura 3 mostra que os Testes apresentam desempenho esperado: aceitam menos os modelos de risco quando reduz-se o $CVaR$ e mais quando contrário. O poder de rejeição mais fraco quando o $CVaR$ é subestimado ($= 1,99$ contra $VaR = 1,96$) atribui-se ao Z_{2s} , que aceitou em 87% das simulações. O mais forte foi o Z_1 . Quando o $CVaR$ é superestimado, o Teste Z_{2s} é o menos encorajador.

Nas simulações com o $CVaR$ subestimado, o Teste Z_{2c} foi o único que teve alta rejeição e a ser rigoroso quando o VaR se distanciava da precisão, tanto para cima quanto para baixo (ver a figura 4, com $\phi = 5\%$). Novamente, Z_1 tem uma alta taxa de aceitação com o VaR inferior e Z_{2s} com o VaR superior, mesmo que o $CVaR$ seja subestimado. Quando o VaR era superestimado, o Z_{2s} chegou até a aceitar mais de 95% das simulações, um desempenho indesejado.

Mesmo quando o $CVaR$ é superestimado, o Z_{2s} ainda apresenta rejeição exagerada quando o VaR é subestimado. Mais uma vez o Teste Z_{2c} se comporta conforme desejado por este artigo.

Nas simulações com quantidade fixa de n exceções, a figura 6 mostra que os Testes E^p e Z_{2c} começam a rejeitar a partir de 6 exceções, um valor próximo à quantidade esperada ($250\alpha = 6,25$) e, a partir de 8, aceitam com menos de 95%. O Z_{2s} demora a reagir e depois tem uma queda acentuada na taxa de aceitação.

Figure 5

Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com $CVaR$ superestimado (= 2,69), alterações do VaR e $\phi = 5\%$

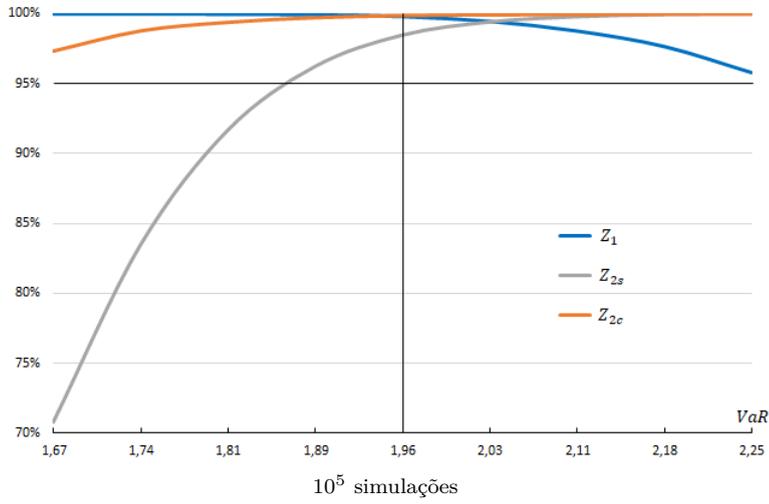
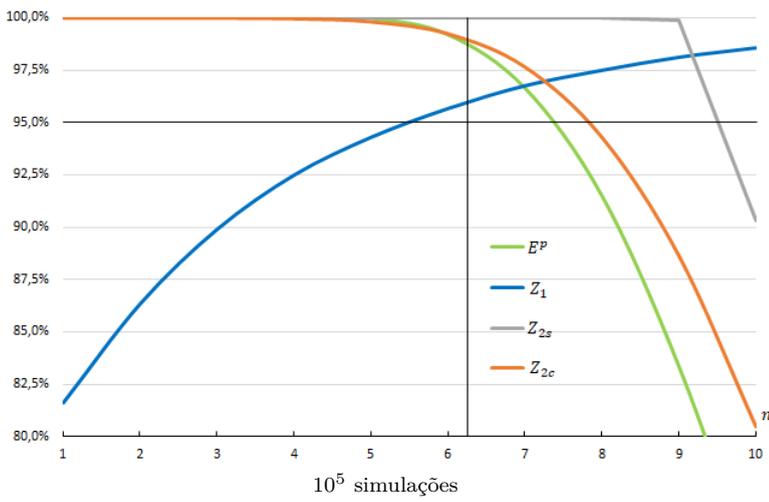


Figure 6

Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com quantidade fixa de n exceções e $\phi = 5\%$



O Teste Z_1 tem baixa taxa de aceitação quando ocorrem poucas exceções e passam a aceitar na mesma direção que a quantidade de exceções. Mesmo que o objetivo seja avaliar tão somente a magnitude da perda em cada exceção, este comportamento é considerado inadequado por este artigo porque os reguladores bancários não devem considerar o modelo de risco como inadequado quando, por exemplo, os retornos excedem o VaR apenas uma vez.

Vale ressaltar que as exceções nestes métodos, apesar de serem aceitas mais que no $VaR_{0,01}$, referem-se ao $VaR_{0,025}$ e não ao $CVaR_{0,025}$.

Após a avaliação e análise dos métodos de backtesting do $CVaR$, este artigo desclassificou os Testes Z_1 e Z_{2s} pelos comportamentos altamente indesejados conforme demonstrado nesta seção.

Apesar de o E^p ter apresentado resultados esperados nas simulações, ainda existem dificuldades devido à Simulação Histórica e ao backtesting do VaR .

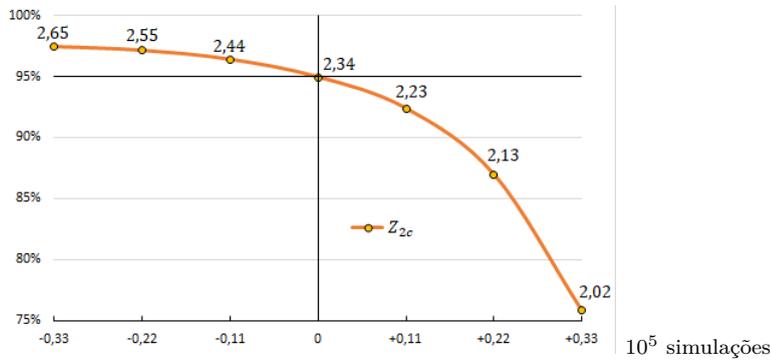
Outro método que apresentou desempenhos desejados até aqui foi o Teste Z_{2c} de Acerbi atendendo os requisitos de nitidez e *ridge backtest*. Porém, este método ainda apresenta alguns resultados estranhos. A equação em (15) mostra que depende não somente quando o VaR é excedido, mas também dos demais dias. Assim, se um banco estiver com o $CVaR$ superestimado nos dias das exceções e subestimado no restante do período, seu modelo de risco pode ter resultado < 0 no Backtesting. Isto não deveria ser permitido uma vez que o banco previu perda maior que a realizada em todo t .

A figura 7 mostra a taxa de aceitação com 10^5 simulações para cada alteração dinâmica do $CVaR$. No eixo da abscissa, o valor $+0,11$ significa $CVaR = 2,34 + 0,11$ quando o VaR é excedido e $CVaR = 2,34 - 0,11$ no restante do período. Por outro lado, o valor $-0,11$ significa $CVaR$ subestimado nos dias das exceções e $CVaR$ superestimado nos demais. O valor exibido em cada ponto refere-se à média do $CVaR$.

Na prática, tal resultado estranho é difícil acontecer na distribuição normal, porém na Simulação Histórica, principalmente no método de estimação da volatilidade EWMA, podem ser gerados com frequência os resultados positivos no modelo subestimado. Isto porque quando há exceção o $CVaR$ pode subir mais que o VaR .

Figure 7

Taxa de aceitação nos Backtesting do $CVaR_{0,025} \sim N$ com alteração dinâmica do $CVaR$ e $\phi = 5\%$



5. Considerações Finais

Foram avaliados através das simulações e analisados os seguintes backtesting do $CVaR$: os três de Acerbi [Z_1 e Z_{2s} de Acerbi and Szekely (2014) e Z_{2c} de Acerbi and Szekely (2017)] e um método deste artigo que é a Aproximação dos Níveis de VaR [E^p , adaptação de Emmer et al. (2015)].

Z_1 : nas situações normais, o Teste se comportou conforme desejado, porém, quando o VaR era superestimado ou subestimado, o backtesting do $CVaR$ foi fortemente afetado violando o requisito de *ridge backtest*. Pior, também não é nítido: quanto mais exceções, maior a aceitação dos resultados. É injusto um Backtesting condenar aquele modelo que tiver apenas uma exceção no $VaR_{0,025}$. O Teste deveria ser estritamente decrescente em relação às perdas. Além disto, este somente pode ser utilizado caso o $VaR_{0,025}$ tenha sucesso no Teste de Basileia (até 10 exceções).

Z_{2s} : o Teste também se comportou conforme desejado nas situações normais e indesejado noutras apresentando problemas de *ridge backtest*. Quanto maior o VaR , maior a aceitação mesmo quando o $CVaR$ era subestimado. E quando o $CVaR$ era superestimado, tinha alta rejeição quando o VaR era menor.

Z_{2c} : este método de testabilidade apresentou desempenhos desejados na maioria das simulações. No nível de significância $\phi = 5\%$, a taxa de aceitação ficou próxima de 95% nas situações normais e não foi fortemente afetada quando se alterava o VaR . Apesar de a

estatística de teste Z_{2c} demonstrar sinal negativo quando o VaR era superestimado, a taxa de aceitação era maior porque é pequena a probabilidade das exceções, que contribui positivamente para as simulações. Quando o $CVaR$ era superestimado, o Teste aceitou mais que 95% independentemente do VaR . Assim como aceitou menos quando o $CVaR$ era subestimado. Entretanto, as simulações foram feitas com VaR_t e $CVaR_t$ igual em todo t . Nas simulações com estas medidas de risco diferentes em cada t podem ser gerados resultados estranhos. Se um banco prever uma perda média menor que todas as realizadas nas exceções e uma perda média alta no restante do período – no qual o VaR não é excedido –, seu modelo de risco pode ser considerado como adequado quando deveria ser rejeitado. Este problema pode facilmente ser encontrado na Simulação Histórica pelo método de estimação da volatilidade EWMA.

E^p : em relação à Aproximação dos Níveis de VaR adaptada por este artigo, apesar de este método apresentar resultados esperados nas simulações e melhores que os Testes de Acerbi, existem ainda duas desvantagens: 1) dificuldade operacional de aplicar na Simulação Histórica caso seja utilizado o $CVaR$ para calcular o requerimento de capital; e 2) o backtesting do VaR não é nítido. No entanto, a ausência da nitidez não é um grande problema no Teste E^p uma vez que neste a margem de erro é menor. Por exemplo, se um modelo de risco tiver 5 exceções no $VaR_{0,010503}$ e 4 no $VaR_{0,007913}$, não se conhece a posição exata da quinta perda, mas esta certamente situa-se entre os dois níveis que tem um intervalo menor que o backtesting do VaR ($< \alpha$).

O apêndice mostra o resumo de cada método com a fórmula do $CVaR$ para Backtesting, desvantagem e consequência.

Apesar de não ser perfeito, o Teste E^p apresenta resultados mais confiáveis e seguros do que o Backtesting do VaR . Atende parcialmente nas regulações bancárias à necessidade do Comitê de Basileia em relação ao número de exceções e magnitude das perdas permitindo que o $CVaR$ ou a média dos níveis de VaR seja utilizado para calcular o requerimento de capital. Além disto, na medida de risco $CVaR_{0,025}$ são esperadas mais exceções do que no $VaR_{0,01}$ tornando o teste mais forte estatisticamente.

Além de corrigir uma falha no Teste Z_{2c} , este artigo também sugere para os trabalhos futuros a avaliação da alocação de capital

baseada no documento de Basileia que utiliza o $CVaR$ /média dos níveis de VaR ou a proposta de outros métodos para o requerimento de capital.

References

- Acerbi, C. (2016). L'es est mort, vive l'es! on backtesting es (and backtesting more in general). Working paper - MSCI Inc.
- Acerbi, C. and Szekely, B. (2014). Backtesting expected shortfall. *Risk*, pages 3–8.
- Acerbi, C. and Szekely, B. (2017). General properties of backtestable statistics. Working paper - MSCI Inc.
- Acerbi, C. and Tasche, D. (2001). Expected shortfall: a natural coherent alternative to value at risk. *arXiv:1506.07212v1*.
- Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26:1487–1503.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Basel Committee on Banking Supervision (1996). *Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*. BIS, Basel, Switzerland.
- Basel Committee on Banking Supervision (2013). *Fundamental Review of the Trading Book*. BIS, Basel, Switzerland.
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.
- Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *The Journal of Risk and Insurance*, 83(3):735–776.
- Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? *Journal of Risk*, 18(2):31–60.

- Fissler, T. and Ziegel, J. F. (2016). Higher order elicibility and osband's principle. *Annals of Statistics*, 44(4):1680–1707.
- Frongillo, R. and Kash, I. A. (2015). On elicitation complexity and conditional elicitation. *arXiv:1506.07212v1*.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.
- Jorion, P. (2003). *Value at Risk*. Bolsa de Mercadorias & Futuros, São Paulo, 2 ed. rev. e ampl. edition.
- Kerkhof, J. and Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance*, 28(8):1845–1865.
- Lambert, N., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. *ACM SIGecom Exchanges*, 7(3).
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series. *Journal of Empirical Finance*, 7:271–300.
- Nolde, N. and Ziegel, J. F. (2016). Elicibility and backtesting. *arXiv:1608.05498v1*.
- Osband, K. (2011). *Pandora's Risk*. Columbia Business School, New York Chichester.
- Righi, M. B. and Ceretta, P. S. (2013). Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation*, 7(3):3–20.
- Schmidt, P. (2016). Interpretation of point forecasts with unknown directive. *arXiv:1506.01917v2*.
- Wang, R. (2016). Theory and practice of risk measurement - part 4 - selected topics and recent developments on risk measures.
- Wimmerstedt, L. (2015). Backtesting expected shortfall. Master's thesis, School of Engineering Sciences.

Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance*, 32:1404–1415.

Apêndice

Table 6
Resumo dos Backtesting

Teste	$CVaR =$	Desvantagem	Consequência
Emmer (E)	$-\alpha^{-1} \int_0^\alpha q_u(X) du \approx$ $-\frac{1}{4}[q_\alpha(X) + q_{0,75\alpha+0,0025}(X) +$ $q_{0,5\alpha+0,005}(X) + q_{0,25\alpha+0,0075}(X)]$	Média dos quatro níveis de VaR distinto do $CVaR$.	Um backtesting não deve aceitar o $CVaR$ através do \overline{CVaR} se os dois forem distintos. Uma das soluções é incluir o nível mais pesado, além de retirar o $q_\alpha(X)$.
		Margem de erro, apesar de esta ser menor que no Teste de Basileia.	Posição imprecisa das perdas. Uma das soluções para reduzir a margem de erro é aumentar a quantidade de níveis.
		Sem aproximação de 95% na distribuição binominal.	O mesmo problema no Teste de Basileia para o $VaR_{0,01}$ no qual aplica-se a mesma penalidade para o $VaR_{0,008}$.
Emmer adaptado (E^p)	$\frac{1}{5}[VaR_{0,016021}(X) +$ $VaR_{0,013214}(X) + VaR_{0,010503}(X) +$ $VaR_{0,007913}(X) + VaR_{0,005484}(X)]$	Dificuldade operacional em encontrar níveis adequados em cada período histórico de observações para a Simulação Histórica.	Um backtesting não deve aceitar o $CVaR$ através do \overline{CVaR} se os dois forem distintos.
		Margem de erro, apesar de esta ser menor que no Teste de Basileia e de Emmer (E).	Posição imprecisa das perdas.
Acerbi (Z_1)	$-\mathbb{E}(X_t X_t < -VaR_{\alpha,t})$	Dependência do intervalo entre VaR e $CVaR$.	Quando o VaR é subestimado (com o $CVaR$ preciso), aumenta-se o intervalo entre VaR e $CVaR$ e, conseqüentemente, também a probabilidade de aceitação indevida. O foco deveria ser o $CVaR$ e não o VaR .
		Correlação indesejada entre o desempenho do Teste e a quantidade de exceções.	Em um exemplo com modelos de risco A (1 exceção média no VaR) e B (20 exceções leves no VaR), o Teste considera B como o melhor uma vez que preocupa-se com apenas a média. O foco deveria ser a magnitude das perdas e não a média das perdas pela quantidade de exceções.
Acerbi (Z_{2s})	$-\frac{1}{\alpha} \mathbb{E}[X(X+q < 0)] = -\mathbb{E}\left(\frac{X_t I_t}{\alpha}\right)$	Dependência do VaR .	O VaR impacta mais que o $CVaR$, que deveria ser o foco. O Teste aceita o modelo com $CVaR$ subestimado se o VaR for superestimado. O termo $q[\alpha - (X+q < 0)]$ não deveria ter sido retirado da fórmula.
Acerbi (Z_{2c})	$-\frac{1}{\alpha} \mathbb{E}[X_t I_t - VaR_{\alpha,t}(\alpha - I_t)] =$ $VaR_{\alpha,t} - \frac{1}{\alpha} \mathbb{E}[(X_t + VaR_{\alpha,t}) I_t]$	Dependência do período sem exceções do VaR .	Em um exemplo com modelos de risco A ($CVaR$ superestimado no período com exceções do VaR e subestimado no restante) e B ($CVaR$ subestimado no período com exceções do VaR e superestimado no restante), o Teste considera B como o melhor uma vez que preocupa-se com os $CVaR$ do período todo. O foco deveria ser apenas os $CVaR$ dos dias com exceção de VaR . Este problema pode facilmente ocorrer na Simulação Histórica pelo método de estimação da volatilidade EWMA, que aumenta o $CVaR$ nos dias seguintes à exceção alta.