

**Validação de sistemas internos de
classificação de risco de crédito
sob o arcabouço prudencial
de Basileia**

Guilherme Fernandes

Validação de sistemas internos de classificação de risco de crédito sob o arcabouço prudencial de Basileia

Guilherme Fernandes*

Resumo

O objetivo deste artigo é introduzir algumas ferramentas utilizadas na validação de sistemas internos de classificação de risco de crédito sob o âmbito das disposições prudenciais de Basileia. São descritos alguns testes para validação do poder discriminatório do sistema de classificação de risco de crédito, como curva Receiver Operating Characteristic (ROC), taxa de erro bayesiana, entropia condicional, valor informacional, estatísticas de Kendall e Somers, escore de Brier, estatística de Kolmogorov-Smirnov e ROC não binário. São realizados três estudos de caso nos quais é possível observar a utilização de tais ferramentas de validação em situações reais por que passam as áreas de validação independente dos bancos.

* Economista do BNDES. Este artigo é de exclusiva responsabilidade do autor, não refletindo, necessariamente, a opinião do BNDES.

Abstract

The goal of this paper is to introduce a few tools used in the validation of internal ratings-based systems under the Basel's prudential recommendations. Some validation tests are described, as ROC Curve, Bayesian Error Rate, Conditional Entropy, Informational Value, Kendall and Somers Statistics, Brier Score, Kolmogorov-Smirnov Statistic and non-binary ROC. Three case studies are performed in order to exemplify the use of such tools in real situations for banks' validation areas.

Introdução

Segundo Meirelles (2010), a regulação e a supervisão do sistema financeiro – sob responsabilidade do Conselho Monetário Nacional (CMN) e do Banco Central do Brasil, respectivamente – passaram de um regime altamente intervencionista e focado na solução de problemas para uma crescente preocupação com estabilidade financeira e aspectos prudenciais.

Como parte das recomendações de Basileia II referentes ao pilar 2, BCBS (2010) estipula que os bancos devem acessar sua própria adequação de capital por meio do processo interno de avaliação da adequação de capital (ICAAP, na sigla em inglês). Como esses processos são definidos internamente, eles agregam risco de várias formas diferentes. O objetivo do ICAAP é determinar a necessidade de alocação de capital adicional acima do disposto no pilar 1. Enquanto o pilar 1 é responsável pelo estabelecimento de requerimentos mínimos de capital de uma forma geral, o pilar 2 tem o objetivo de adicionar alguma necessidade extra de capital em decorrência de avaliação da própria instituição financeira que não esteja contemplada no pilar 1.

Dessa forma, Bacen (2011) estipula que o ICAAP deve analisar a suficiência do capital mantido pela instituição, devendo abranger avaliação e cálculo da necessidade de capital para cobertura dos riscos de crédito, de mercado e operacional, taxa de juros das operações não classificadas na carteira de negociação, crédito de contraparte e concentração. Também deve avaliar a necessidade de capital para cobertura dos riscos de liquidez, estratégia e reputação. Bacen (2011) disciplina ainda que o ICAAP deve ser submetido a um processo de validação independente do processo de desenvolvimento. Nesse sentido, a responsabilidade da validação independente é avaliar, no mínimo, as metodologias e premissas utilizadas

nas estimativas de necessidade de capital; as estimativas de correlação, quando utilizada; a inclusão de todos os riscos relevantes; a abrangência, a consistência, a integridade e a confiabilidade dos dados de entrada, bem como a independência de suas fontes; a adequação dos testes de estresse; a consistência e a confiabilidade das informações que compõem o relatório de ICAAP.

De acordo com o Comitê de Supervisão Bancária de Basileia [BCBS (2005b)], os sistemas internos de classificação de risco de crédito representam um ponto-chave para o cálculo do capital regulatório em abordagens *internal ratings-based systems* (IRB) porque são a base para a determinação do parâmetro probabilidade de descumprimento (PD). A PD e outros dois componentes de risco de crédito – perda, dado o descumprimento (LGD), e exposição no momento do descumprimento (EAD) – são os parâmetros de entrada da modelagem para o cálculo do capital regulatório e para a gestão do risco de crédito com base em modelos internos. Consequentemente, a validação desses três parâmetros e do sistema de classificação de risco subjacente é fundamental no processo de supervisão de instituições financeiras.

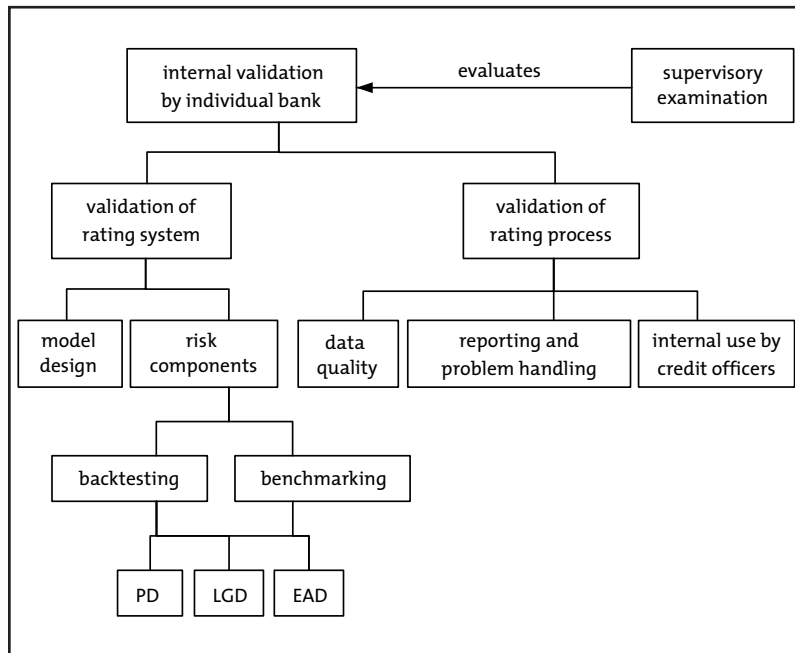
Para BCBS (2005b), os bancos devem demonstrar a *performance* de seus sistemas internos de classificação de maneira eficiente e conclusiva. Devem ser utilizadas diferentes metodologias quantitativas de validação para assegurar o bom desempenho de seus sistemas.

O objetivo deste artigo é introduzir algumas ferramentas estatísticas utilizadas na validação de sistemas internos de classificação de risco de crédito sob o âmbito das disposições prudenciais de Basileia e demonstrar como elas são capazes de auxiliar na escolha de modelos que estimem o risco da forma mais acurada possível do ponto de vista estatístico.

A próxima seção trata de princípios definidos pelo Comitê de Supervisão Bancária de Basileia [BCBS (2005a)] que devem guiar as atividades de validação, além de realizar uma breve revisão da literatura sobre o tema. Em seguida, são apresentadas algumas ferramentas estatísticas capazes de medir o poder discriminatório de sistemas internos de classificação de risco de crédito. São realizados, nas seções posteriores, três estudos de caso em que é possível observar a aplicação dessas ferramentas estatísticas na validação de modelos de classificação de risco. A última seção fecha o artigo com conclusão a respeito da importância e da complexidade da atividade de validação.

Figura 1

Metodologia de validação



Fonte: BCBS (2005b).

Princípios da validação e revisão da literatura

BCBS (2005a) define validação de sistemas internos de classificação de risco de crédito como o grupo de processos e atividades que contribuem para uma avaliação se os níveis de *rating* diferenciam o risco adequadamente e se seus componentes (PD, LGD, EAD) caracterizam apropriadamente os aspectos relevantes do risco de crédito. O grupo de Basileia responsável por essa validação define seis princípios básicos de tal processo.

O primeiro princípio trata da validação como forma de avaliar as estimativas de risco e o uso dos *ratings* no processo de concessão de crédito. Isto é, por mais que as estimativas de risco sejam calculadas no passado, elas devem ser capazes de prever o futuro em alguma medida.

O segundo princípio define a responsabilidade primária da validação como do próprio banco e não de sua entidade supervisora – no caso do Brasil, o Banco Central do Brasil. Cabe ao supervisor avaliar se a validação está adequada, podendo realizar novos testes e procedimentos a fim de respaldar sua análise.

O terceiro princípio foca no aspecto iterativo da validação. Isto é, bancos e supervisores precisam redefinir, periodicamente, ferramentas de validação em resposta a mudanças nas condições de mercado ou de operação. Por isso, o diálogo entre supervisores e bancos é muito importante para pôr em prática esse princípio.

O quarto princípio estipula que não há um único método de validação. Não existe prática de validação que possa ser utilizada em todas as carteiras para todos os bancos. A ferramenta de *backtesting*, por exemplo, é muito difícil de ser utilizada em carteiras que apresentem baixo nível histórico de descumprimento.

O quinto princípio diz que a validação deve abordar aspectos quantitativos e qualitativos do sistema de classificação de risco. Embora as ferramentas matemáticas e estatísticas constituam aspecto

central no processo de validação, na avaliação da *performance* de um sistema de classificação é necessário verificar também os elementos qualitativos presentes nele, como controle, documentação, uso interno etc.

O sexto e último princípio estipula que o processo de validação seja analisado de forma independente. Em muitos casos, a auditoria interna constitui a área competente para tal revisão, sendo a responsável por assegurar que o processo de validação seja implementado da forma como foi planejado.

De acordo com BCBS (2005b), a análise estatística de sistemas internos de classificação de risco de crédito está predominantemente calcada na separação das contrapartes em dois tipos: instituições que entrarão em descumprimento em um horizonte de tempo determinado (más) e instituições que não entrarão em descumprimento em tal horizonte de tempo (boas). Desconsidera-se, neste trabalho, a definição de instituições intermediárias. Será utilizado o conceito binário representado por contrapartes boas ou más.

A ideia principal de qualquer sistema de classificação de risco é a de que, quanto melhor a classificação de determinada contraparte, menor é a proporção de instituições más e maior é a proporção de instituições boas presentes no conjunto de instituições que possuem determinado *rating*. O poder de discriminação de um sistema de classificação de risco denota, portanto, sua habilidade de distinguir, de maneira *ex ante*, contrapartes que entrarão em descumprimento de contrapartes que não entrarão. O poder discriminatório do sistema deve ser testado em amostra segregada (*out-of-sample*). Caso contrário, corre-se o risco de o poder discriminatório ser superestimado.

Cada nível de classificação de risco deve estar associado a uma PD. Uma instituição financeira que calcule o capital necessário para a cobertura de exposições a risco de crédito com base em modelos internos (IRB *approach* ou abordagem IRB) utilizará os parâmetros

de PD, LGD e EAD nesse cálculo. Por isso, é importante que tais parâmetros estejam devidamente calibrados, para que a estimativa de capital seja a mais acurada possível.

Erlenmaier (2006) cita algumas formas de validação para a classificação de risco e atribuição de PDs, como: cálculo do coeficiente de correlação linear entre PDs internas e externas; análise das companhias com maior diferença entre *ratings* internos e externos; comparação de PDs médias de pequenos grupos definidos em termos geográficos, setoriais etc.; coeficiente de Gini para os níveis do sistema de classificação; comparação dos resultados do novo sistema de classificação de risco com o sistema anterior, quando disponível.

Engelmann (2006) define o poder de discriminação de um sistema interno de classificação de risco de crédito como a habilidade de separar créditos de boa qualidade de créditos de qualidade inferior. Questões semelhantes são abordadas em outras áreas do conhecimento. Na medicina, por exemplo, a qualidade de um teste de diagnóstico é predominantemente determinada por sua habilidade de distinguir pessoas saudáveis e doentes. Aplicações análogas são vistas em outras áreas como biologia, tecnologia da informação e engenharia.

É importante destacar que poder discriminatório e calibração representam conceitos diferentes. Enquanto o poder discriminatório depende da diferença de frequência de observações de instituições más e boas, a correta calibração das PDs foca na acurácia dessas estimativas para diferentes níveis de classificação de risco.

Validação do poder discriminatório do sistema de classificação de risco de crédito

BCBS (2005b) lista uma série de testes relativos ao poder discriminatório dos sistemas internos de classificação de risco de crédito. Comenta-se a respeito de alguns deles a seguir.

ROC e o índice de Pietra

A curva ROC é construída por meio de duas amostras representativas de contrapartes boas e más. Espera-se que a curva ROC seja côncava, o que significa dizer que há maior concentração de instituições más nos piores escores e de instituições boas nos melhores escores. O mais importante índice derivado da análise de ROC é a área sob a curva ROC (**AUROC**). É muito simples construir intervalo de confiança para a estimativa de AUROC a fim de verificar se o valor 0,5 estaria presente nesse intervalo. Assim, é possível testar a hipótese de a AUROC ser estatisticamente diferente de 0,5, valor sob o qual o modelo não teria nenhum poder discriminatório. Vale destacar que as propriedades estatísticas da curva ROC coincidem com a estatística de Mann-Whitney. Andrade e Oliveira (2012) consideram que o poder discriminatório do modelo é aceitável para valor de AUROC acima de 70%, como mostra Quadro 1.

Quadro 1

Avaliação de AUROC

AUROC	Capacidade de discriminação
$ROC = 0.5$	Não existe discriminação
$0.7 \leq ROC \leq 0.8$	Discriminação aceitável
$0.8 \leq ROC < 0.9$	Excelente discriminação
$ROC \geq 0.9$	Discriminação acima do comum

Fonte: Andrade e Oliveira (2012).

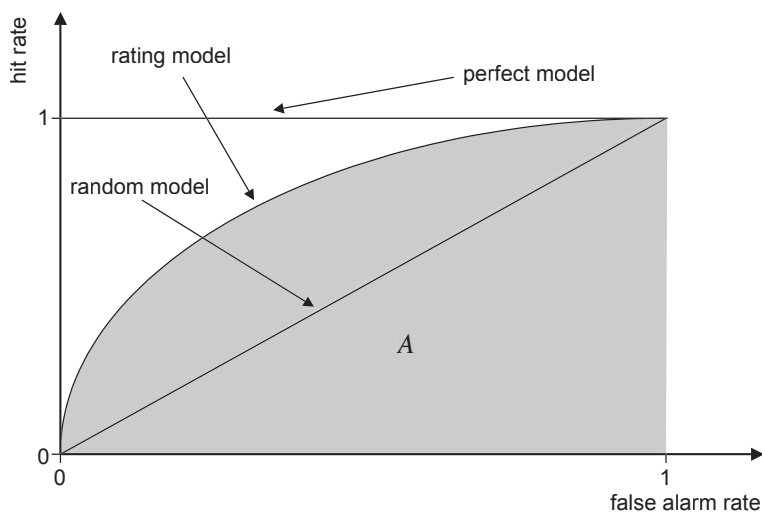
O uso da curva ROC como ferramenta de análise para sistemas de classificação de risco é indicado por diversos autores, entre eles BCBS (2005b), que demonstra preferência pelo método ROC para a verificação do poder discriminante de sistemas de classificação de risco de crédito graças a suas propriedades estatísticas. Segundo Annibal (2008), a possibilidade de facilmente calcular os intervalos

de confiança da área sob a curva ROC é uma propriedade bastante desejada, pois, considerando que os valores desse índice são, via de regra, obtidos por meio da seleção de amostras, torna-se necessária a realização de testes de hipótese estatísticos para que se possam comparar os valores dos índices AUROC encontrados com o valor que significa a inexistência de poder discriminante de um sistema de classificação, qual seja, 0,5. Além disso, a determinação de intervalos de confiança também é fundamental para que se possam comparar os desempenhos de valores distintos de AUROC provenientes de diferentes modelos de classificação de risco.

O índice de Pietra é outra importante medida derivada da análise ROC. Ele é definido como a metade da distância máxima entre a curva ROC e a reta de 45°. O Gráfico 1 constitui exemplo ilustrativo de construção da curva ROC.

Gráfico 1

Área sob a curva ROC



Fonte: BCBS (2005b).

Define-se o *hit rate* (HR) como:

$$HR(C) = \frac{H(C)}{N_M}$$

Em que:

$H(C)$ = número de instituições más previstas corretamente para um corte C ; e

N_M = número total de instituições más na amostra.

A taxa de alarme falso (TAF) é definida como:

$$TAF(C) = \frac{F(C)}{N_B}$$

Em que:

$F(C)$ = número de instituições boas classificadas incorretamente como más para um corte C ; e

N_B = número total de instituições boas na amostra.

Define-se, então, a forma de cálculo para a AUROC:

$$AUROC = \int_0^1 HR(TAF)d(TAF)$$

Sendo:

S_M = variável aleatória representando o escore das instituições más; e

S_B = variável aleatória representando o escore das instituições boas, então, podem-se redefinir $HR(C)$ e $TAF(C)$ da seguinte forma:

$$HR(C) = P(S_M < C)$$

$$TAF(C) = P(S_B < C)$$

$$\begin{aligned} AUROC &= \int_0^1 HR(TAF)d(TAF) \\ &= \int_0^1 P(S_M < C)dP(S_B < C) = \int_{-\infty}^{+\infty} P(S_M < C)f_{S_B}(C)dC = P(S_M < S_B) \end{aligned}$$

Em que:

f_{S_B} = função de densidade de probabilidade de S_B .

Engelmann (2006) demonstra a relação existente entre a AUROC e a estatística de Mann-Whitney [Mann e Whitney (1947)]; sejam s_M e s_B realizações de S_M e S_B , respectivamente.

$$u_{M,B} = \begin{cases} 1, & \text{se } s_M < s_B \\ \frac{1}{2}, & \text{se } s_M = s_B \\ 0, & \text{se } s_M > s_B \end{cases}$$

A estatística \hat{U} de Mann-Whitney é definida como:

$$\hat{U} = \frac{1}{N_M \cdot N_B} \sum_{(M,B)} u_{M,B}$$

Assim, a AUROC pode ser vista como um estimador não viesado de \hat{U} :

$$AUROC = E(\hat{U}) = P(S_M < S_B) + \frac{1}{2}P(S_M = S_B)$$

A variância de \hat{U} é definida da seguinte forma:

$$\begin{aligned} \hat{\sigma}_{\hat{U}}^2 &= \frac{1}{4(N_M - 1)(N_B - 1)} \left[\hat{P}_{M \neq B} + (N_M - 1)\hat{P}_{M,M,B} + (N_B - 1)\hat{P}_{B,B,M} \right. \\ &\quad \left. - 4(N_M + N_B - 1) \left(\hat{U} - \frac{1}{2} \right)^2 \right] \end{aligned}$$

$\hat{P}_{M,M,B}$ e $\hat{P}_{B,B,M}$ são estimadores para $P_{M,M,B}$ e $P_{B,B,M}$, respectivamente.

$$P_{M,M,B} = P(S_{M,1}, S_{M,2} < S_B) + P(S_B < S_{M,1}, S_{M,2}) - P(S_{M,1} < S_B < S_{M,2}) - P(S_{M,2} < S_B < S_{M,1})$$

$$P_{B,B,M} = P(S_{B,1}, S_{B,2} < S_M) + P(S_M < S_{B,1}, S_{B,2}) - P(S_{B,1} < S_M < S_{B,2}) - P(S_{B,2} < S_M < S_{B,1})$$

$S_{M,1}, S_{M,2}$ e $S_{B,1}, S_{B,2}$ são realizações independentes de S_M e S_B respectivamente.

O termo $\left[\frac{(AUROC - \hat{U})}{\hat{\sigma}_{\hat{U}}} \right]$ converge em distribuição para uma normal-padrão:

$$\left[\frac{(AUROC - \hat{U})}{\hat{\sigma}_{\hat{U}}} \right] \xrightarrow{D} N(0,1)$$

Então, obtém-se:

$$P\left(\hat{U} - \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{1+\alpha}{2}\right) \leq AUROC \leq \hat{U} + \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{1+\alpha}{2}\right)\right) \approx \alpha$$

Em que Φ denota a função de distribuição acumulada da normal-padrão. Assim, pode-se definir o intervalo de confiança para AUROC:

$$I_\alpha = \left[\hat{U} - \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{1+\alpha}{2}\right), \hat{U} + \hat{\sigma}_{\hat{U}}\Phi^{-1}\left(\frac{1+\alpha}{2}\right) \right]$$

Para todo auroc pertencente ao intervalo I_α , a hipótese (AUROC = auroc) não pode ser rejeitada sob nível de significância α .

Taxa de erro bayesiana

A taxa de erro bayesiana especifica a probabilidade mínima de ocorrência de erro para um sistema interno de classificação de risco de crédito no âmbito de uma decisão binária relativa à ocorrência ou não de *default*. O erro pode ser estimado de forma paramétrica, assumindo

normalidade para a distribuição dos respectivos eventos, ou de forma não paramétrica, por meio de estimativas de densidade Kernel.

$$Taxa\ de\ erro\ bayesiana = \begin{cases} \min_c (p_D(1 - HR(C)) + (1 - p_D)TAF(C)), & se\ p_D > 50\% \\ \frac{1}{2} - \frac{1}{2} \max_c |HR(C) - TAF(C)|, & se\ p_D \leq 50\% \end{cases}$$

Entropia condicional, distância de Kullback-Leibler e razão de entropia informacional condicional (CIER)

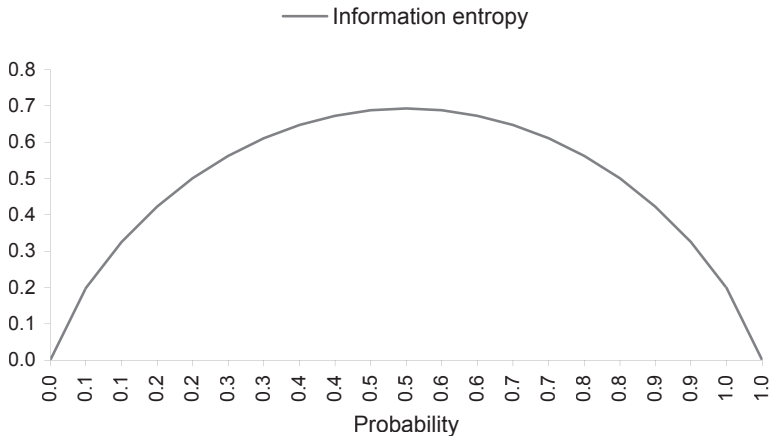
A entropia informacional $H(p)$ de um evento com probabilidade p é definida como:

$$H(p) = -(p \cdot \log(p) + (1 - p) \cdot \log(1 - p))$$

O Gráfico 2 exemplifica construção de curva de entropia informacional.

Gráfico 2

Entropia informacional



Fonte: BCBS (2005b).

Sejam:

S = escore;

D = evento de descumprimento; e

\bar{D} = evento complementar a D – isto é, não há descumprimento, então:

$$H(P(D|S)) = -(P(D|S) \cdot \log(P(D|S)) + P(\bar{D}|S) \cdot \log(P(\bar{D}|S)))$$

A entropia condicional é definida como:

$$H_S = E[H(P(D|S))] = -E[P(D|S) \cdot \log(P(D|S)) + P(\bar{D}|S) \cdot \log(P(\bar{D}|S))]$$

A distância de Kullback-Leibler é definida como:

$$\text{Distância de Kullback – Leibler} = H(p) - H_S$$

A distância de Kullback-Leibler é, então, normalizada pela entropia informacional incondicional, gerando a medida conhecida como CIER:

$$CIER = \frac{H(p) - H_S}{H(p)}$$

Valor informacional

Outra medida de poder discriminatório baseada no conceito de entropia é o chamado valor informacional, que se trata de uma medida da diferença entre as distribuições de escore de instituições boas e más. Sejam:

f_D = função de densidade de probabilidade de escore das instituições más; e

$f_{\bar{D}}$ = função de densidade de probabilidade de escore das instituições boas.

Então, o valor informacional (IV) é definido como:

$$IV = E \left[\log \frac{f_D(S)}{f_{\bar{D}}(S)} | D \right] + E \left[\log \frac{f_{\bar{D}}(S)}{f_D(S)} | \bar{D} \right] = \int (f_D(S) - f_{\bar{D}}(S)) \log \frac{f_D(S)}{f_{\bar{D}}(S)} ds$$

Testes de Kendall e Somers (*shadow rating*)

Os testes de Kendall e Somers são indicados por BCBS (2005b) para a validação de modelos de *shadow rating*. Esse modelo foi criado com o objetivo de replicar classificações de risco externas considerando aspectos econômico-financeiros das instituições, quando informações a respeito de suas frequências de descumprimento não estão disponíveis.

O teste de Kendall é um teste estatístico cujo objetivo é medir o grau de dependência entre duas variáveis aleatórias. A noção de dependência comonotônica presente nele generaliza o conceito de dependência linear, expresso pelo coeficiente de correlação. Na literatura de ciências atuariais, a dependência comonotônica é considerada a forma mais forte de dependência entre variáveis aleatórias.

Seja (X,Y) um par de variáveis aleatórias. A estatística τ de Kendall é definida como:

$$\tau_{XY} = P(X_1 < X_2, Y_1 < Y_2) + P(X_1 > X_2, Y_1 > Y_2) - P(X_1 < X_2, Y_1 > Y_2) - P(X_1 > X_2, Y_1 < Y_2)$$

(X_1, Y_1) e (X_2, Y_2) são realizações independentes de (X,Y) . A estatística τ_{XY} de Kendall pode ser vista como uma diferença entre duas probabilidades: a probabilidade de o maior valor de X estar associado ao maior valor de Y , e a probabilidade de o maior valor de X estar associado ao menor de valor de Y .

A estatística D de Somers é definida em função da estatística τ de Kendall:

$$D_{XY} = \frac{\tau_{XY}}{\tau_{YY}}$$

Ela também pode ser vista a partir da equação abaixo:

$$D = P(S_M < S_B) - P(S_M > S_B)$$

Escore de Brier

O escore de Brier (1950) é um método de avaliação da qualidade da previsão de uma probabilidade. Sejam p_0, p_1, \dots, p_k PDs estimadas para contrapartes em K níveis de classificação de risco. Então, o escore de Brier é definido como:

$$B = \frac{1}{n} \sum_{j=1}^n (p_j - \theta_j)^2$$

Em que:

$$\theta_j = \left\{ \begin{array}{l} 1, \text{ se contraparte } j \text{ entra em descumprimento} \\ 0, \text{ caso contrário} \end{array} \right\}$$

Vale destacar que o escore de Brier apresenta fraca *performance* para baixos valores de PD. Suponha, por exemplo, que todas as contrapartes de determinada amostra tenham PD no valor de p . Então, o Escore de Brier, nesse caso, será igual a:

$$B = (1 - p)p$$

$$\lim_{p \rightarrow 0} B = 0$$

Estatística de Kolmogorov-Smirnov

A estatística de Kolmogorov-Smirnov (KS) mede a capacidade do *rating* em distinguir bons e maus clientes. Quanto maior a estatística KS, maior a separação entre bons e maus clientes. Essa estatística é definida como o valor máximo da diferença entre as distribuições

acumuladas de boas e más IFs. Andrade e Oliveira (2012) estipulam níveis de discriminação para diferentes valores de KS.

Quadro 2

Avaliação da estatística KS

Valores de KS	Nível de discriminação
Abaixo de 20%	Baixa discriminação
De 20% a 30%	Discriminação aceitável
De 30% a 40%	Boa discriminação
De 40% a 50%	Excelente discriminação
Acima de 50%	Não são muito comuns

Fonte: Andrade e Oliveira (2012).

ROC não binário

Obuchowski (2005) desenvolve metodologia não paramétrica para estimação de AUROC não binário e cria estimadores para AUROC não binário discreto (*ordinal*) e contínuo (*continuous*), de acordo com o Quadro 3.

Quadro 3

Estimadores de AUROC não binários

Gold Standard	Estimator of Accuracy	
Binary	$\hat{\theta} = \frac{1}{n_t n_s} \sum_{i=1}^{n_t} \sum_{j=1}^{n_s} \Psi(X_{it}, X_{js})$	where: $\Psi = 1$ if $X_{it} > X_{js}$ $\Psi = 0.5$ if $X_{it} = X_{js}$ $\Psi = 0$ if $X_{it} < X_{js}$.
Continuous	$\hat{\theta}' = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \Psi'(X_{it}, X_{js})$	where: $\Psi' = 1$ if $t > s$ and $X_{it} > X_{js}$ or $s > t$ and $X_{js} > X_{it}$ $\Psi' = 0.5$ if $t = s$ or $X_{it} = X_{js}$ $\Psi' = 0$ otherwise.
Ordinal	$\hat{\theta}'' = 1.0 - \sum_{(t=1)}^T \sum_{(s>t)}^T w_{ts} \cdot L(t,s) \cdot (1 - \hat{\theta}_{ts})$	where: is the binary-scale estimator
Nominal	SAME AS ORDINAL	where $\hat{\theta}_{ts} = \frac{1}{(n_t \cdot n_s)} \sum_{i=1}^{n_t} \sum_{j=1}^{n_s} \Psi(D_{(t-s)ij}, D_{(t-s)isk})$ $\Psi(D_{(t-s)ij}, D_{(t-s)isk}) = 1$ if $D_{(t-s)ij} > D_{(t-s)isk}$ $\Psi(D_{(t-s)ij}, D_{(t-s)isk}) = 1/2$ if $D_{(t-s)ij} = D_{(t-s)isk}$ $\Psi(D_{(t-s)ij}, D_{(t-s)isk}) = 0$ if $D_{(t-s)ij} < D_{(t-s)isk}$.

Fonte: Obuchowski (2005).

ROC não binário contínuo

O ROC não binário contínuo é interessante para experimentos em que o suporte da função é um intervalo – como no caso de PDs, que são números reais presentes no intervalo $[0,1]$.

ROC não binário discreto

O ROC não binário discreto é semelhante ao caso binário, com a diferença de que, neste último, pretende-se avaliar determinado grau de acerto de experimento, enquanto o primeiro avalia a adequação a uma variável discreta que possa assumir mais valores além de zero ou um.

Estudo de Caso I: Modelo de *shadow rating* para instituições não financeiras

BCBS (2005b) define *shadow rating* como uma metodologia cujo objetivo é replicar uma classificação de risco externa para contrapartes que não têm tal classificação. Modelos de *shadow rating* são construídos com base em informações econômico-financeiras e probabilidades de descumprimento de empresas que têm *ratings* externos. Tais relações são, então, extrapoladas para empresas que não têm esses *ratings* e, assim, suas classificações de risco são geradas com base nos parâmetros estimados pelo modelo.

Cardoso *et al.* (2013) propõem um modelo de *shadow rating* para instituições não financeiras que inclua características ligadas a lucro, alavancagem, liquidez, porte, setor e cobertura de dívida. Assim, foram selecionadas as seguintes informações econômico-financeiras: Lucro líquido/Ebitda, Cobertura de juros, ROA, *Dummy* para o

setor *Utilities*, Passivo total/Ativo total e (logaritmo neperiano de) Ativo total. O modelo de regressão resultante é:

$$\begin{aligned}
 \text{score} = & \alpha + \beta_1 \cdot \frac{\text{Lucro líquido}}{\text{Ebitda}} + \beta_2 \cdot \text{ROA} + \beta_3 \cdot \text{Cobertura de juros} + \beta_4 \cdot \ln(\text{Ativo total}) \\
 & + \beta_5 \cdot \frac{\text{Passivo total}}{\text{Ativo total}} + \beta_6 \cdot \text{Dummy Utilities} + \varepsilon \\
 PD = & \frac{1}{1 + \exp(-\text{score})}
 \end{aligned}$$

Visualiza-se, na Tabela 1, relação entre PDs e classificações de risco utilizada por Cardoso *et al.* (2013) e no presente artigo.

Tabela 1

Ratings corporativos e PDs de cinco anos – 1983-2009 (em %)

<i>Rating</i> Moody's	<i>Rating</i> Standard & Poors	PD Moody's
Aaa	AAA	0,086
Aa1	AA+	0,141
Aa2	AA	0,195
Aa3	AA-	0,324
A1	A+	0,492*
A2	A	0,746
A3	A-	0,830
Baa1	BBB+	1,180
Baa2	BBB	2,024
Baa3	BBB-	3,081
Ba1	BB+	7,289
Ba2	BB	8,084
Ba3	BB-	16,948
B1	B+	20,077

(*Continua*)

(Continuação)

Rating Moody's	Rating Standard & Poors	PD Moody's
B2	B	25,211
B3	B-	36,907
Caa1	CCC+	47,262
Caa2	CCC	49,868
Caa3	CCC-	66,96
Ca-C	CC-SD	70,176

Fonte: Cardoso *et al.* (2013).

* Com ajuste de monotonicidade (diferente do original).

Com base em informações econômico-financeiras relativas ao exercício de 2013 presentes em uma amostra de 500 empresas, reestimou-se o modelo de Cardoso *et al.* (2013) e obtiveram-se os resultados expostos na Tabela 2.

Tabela 2

Estimação dos parâmetros

Parâmetro	Valor estimado	Estatística t	P-Valor
Intercepto	10,188439	14,422	0,00%
Lucro líquido / Ebitda	0,1005165	3,87	0,01%
ROA	-3,2173693	-3,177	0,16%
Cobertura de juros	-0,0003267	-1,56	11,93%
ln(Ativo total)	-0,6048249	-20,973	0,00%
Passivo total / Ativo total	0,5517481	1,675	9,45%
<i>Dummy Utilities</i>	-2,0232181	-8,463	0,00%
R2	52,93%	n.a.	n.a.
R2 ajustado	52,36%	n.a.	n.a.

Fonte: Elaboração própria, com base em valores estimados em *R*.

Destaca-se que o modelo de Cardoso *et al.* (2013) tem o objetivo de replicar uma PD global com base em informações econômico-financeiras. Isto é, não leva em consideração aspectos geográficos das empresas. Adicionou-se, então, uma variável independente representando a PD do respectivo país a fim de verificar se há alguma melhora na previsão da classificação de risco estimada. Com isso, o modelo de regressão passa a ser:

$$\begin{aligned}
 \text{score} = & \alpha + \beta_1 \cdot \frac{\text{Lucro líquido}}{\text{Ebitda}} + \beta_2 \cdot \text{ROA} + \beta_3 \cdot \text{Cobertura de juros} + \beta_4 \cdot \ln(\text{Ativo total}) \\
 & + \beta_5 \cdot \frac{\text{Passivo total}}{\text{Ativo total}} + \beta_6 \cdot \text{Dummy Utilities} + \beta_7 \cdot \text{PD}_{\text{país}} + \varepsilon \\
 \text{PD} = & \frac{1}{1 + \exp(-\text{score})}
 \end{aligned}$$

Visualizam-se na Tabela 3 os resultados da estimação.

Tabela 3

Estimação dos parâmetros do modelo de Cardoso *et al.* (2013) com a inclusão da PD do país

Parâmetro	Valor estimado	Estatística t	P-Valor
Intercepto	9,5498537	14,025	0,00%
Lucro líquido / Ebitda	0,1082517	4,36	0,00%
ROA	-2,7986695	-2,888	0,41%
Cobertura de juros	-0,0003219	-1,61	10,81%
ln(Ativo total)	-0,5849545	-21,125	0,00%
Passivo total / Ativo total	0,6797796	2,157	3,15%
<i>Dummy Utilities</i>	-1,9579639	-8,567	0,00%
PD país	5,3463172	6,961	0,00%
R ₂	57,15%	n.a.	n.a.
R ₂ ajustado	56,54%	n.a.	n.a.

Fonte: Elaboração própria, com base em valores estimados em R.

Estimou-se, ainda, um modelo de regressão com o escore do país no lugar da PD do país:

$$\begin{aligned} \text{escore} = & \alpha + \beta_1 \cdot \frac{\text{Lucro líquido}}{\text{Ebitda}} + \beta_2 \cdot \text{ROA} + \beta_3 \cdot \text{Cobertura de juros} + \beta_4 \cdot \ln(\text{Ativo total}) \\ & + \beta_5 \cdot \frac{\text{Passivo total}}{\text{Ativo total}} + \beta_6 \cdot \text{Dummy Utilities} + \beta_7 \cdot \text{Escore}_{\text{país}} + \varepsilon \\ PD = & \frac{1}{1 + \exp(-\text{escore})} \end{aligned}$$

Tabela 4

Estimação dos parâmetros do modelo de Cardoso *et al.* (2013) com a inclusão do escore do país

Parâmetro	Valor estimado	Estatística t	P-Valor
Intercepto	10,8011144	15,335	0,00%
Lucro líquido / Ebitda	0,0955244	3,752	0,02%
ROA	-2,9325375	-2,95	0,33%
Cobertura de juros	-0,000284	-1,383	16,72%
ln(Ativo total)	-0,6086148	-21,536	0,00%
Passivo total / Ativo total	0,8400974	2,558	1,08%
Dummy Utilities	-1,9197627	-8,161	0,00%
Escore país	0,1078985	4,668	0,00%
R ₂	54,93%	n.a.	n.a.
R ₂ ajustado	54,29%	n.a.	n.a.

Fonte: Elaboração própria, com base em valores estimados em R.

Observa-se que as variáveis **PD país** (Tabela 3) e **escore país** (Tabela 4) são altamente significativas do ponto de vista estatístico, pois

o p-valor associado a suas estatísticas *t-student*, em que a hipótese nula reside na insignificância do parâmetro, é extremamente baixo. A Tabela 5 mostra os resultados para alguns testes na comparação entre as PDs estimadas pelos modelos de *shadow rating* e as PDs obtidas pelos *ratings* das agências externas. Os testes foram realizados para uma amostra de 479 empresas que contempla informações econômico-financeiras relativas ao exercício de 2013. Destaca-se que tal amostra não possui interseção com aquela utilizada na estimação do modelo de regressão.

Tabela 5

Testes de aderência – validação do modelo

Teste	Modelo original	Modelo acrescido da PD do país	Modelo acrescido do score do país	Modelo topado pela CR do país
AUROC contínuo	72,44%	72,89%	72,91%	72,96%
AUROC discreto	75,18%	75,68%	75,70%	75,76%
Estatística de Kendall (p-valor)	0,514 (0,00%)	0,525 (0,00%)	0,524 (0,00%)	0,527 (0,00%)
Estatística de Somers Dxy	0,525	0,536	0,534	0,539
Coefficiente de Correlação Linear	62,52%	61,08%	62,16%	62,74%

Fonte: Elaboração própria, com base em valores estimados em R.

O *modelo original* não faz qualquer tipo de tratamento em relação à classificação de risco do país onde fica a empresa, utilizando apenas

a PD de referência do resultado do modelo de regressão proposto por Cardoso *et al.* (2013). O **modelo acrescido da PD do país** inclui a PD do respectivo país na regressão. O **modelo acrescido do score do país** inclui o score do respectivo país, que nada mais é do que uma transformação não linear da PD do país, no modelo original. Já o **modelo topado pela CR do país** utiliza o valor máximo de PD entre a PD estimada pelo modelo de regressão original e a PD do respectivo país.

O Teste AUROC Contínuo de Obuchowski (2005) computa uma área abaixo da curva ROC (imaginária) não binária para valores contínuos das variáveis utilizadas – o que faz todo o sentido para análise de PDs, uma vez que probabilidades são variáveis contínuas contidas no intervalo [0,1]. O Teste AUROC Discreto de Obuchowski (2005) computa a mesma área, com a diferença de que, nesse caso, os valores das variáveis utilizadas são discretos, podendo ser interpretados como níveis de classificação de risco. Em ambos os casos, quanto maior o valor da AUROC, melhor é a discriminação do modelo. Observa-se que o modelo topado pela CR do país foi o vencedor para os testes AUROC não binários.

As estatísticas de Kendall e Somers medem o quanto duas variáveis são dependentes, não necessariamente do ponto de vista linear. Quanto maior o valor da estatística, mais dependentes são as variáveis. Para os três modelos, o p-valor de Kendall foi desprezível, indicando que em nenhum caso é possível aceitar a hipótese nula de independência, mas é possível comparar os valores das estatísticas e encontrar o modelo que mais fortemente rejeita a hipótese de independência – no caso, o modelo topado pela CR do país.

O coeficiente de correlação linear mede o quanto duas variáveis aleatórias são relacionadas do ponto de vista linear. O modelo que obteve melhor aderência foi aquele topado pela CR do país.

Estudo de caso II: modelo de *shadow rating* para países

Macedo *et al.* (2014) propõem um modelo de *shadow rating*, para a replicação das probabilidades de *default* de países que inclua características ligadas a fatores econômicos, políticos, fiscais, externos e monetários através de um modelo econométrico de dados em painel. Assim, foram selecionadas as seguintes variáveis explicativas: Saldo em conta corrente (*cab*), (log base 10) PIB *per capita* (*gdppc*), Resultado nominal do setor público (*gsd*), Índice WGI, Inflação, (log base 10) Reservas em moeda estrangeira (*trc*) e uma *dummy* para segregar países com índice WGI acima de 75%. O modelo de regressão resultante é o seguinte:

$$score = \alpha + \beta_1 \cdot cab + \beta_2 \cdot gdppc + \beta_3 \cdot gsd + \beta_4 \cdot wgi + \beta_5 \cdot inflação + \beta_6 \cdot trc + \beta_7 \cdot Dummy\ WGI + \varepsilon$$

$$PD = \frac{1}{1 + \exp(-score)}$$

A Tabela 6 apresenta os coeficientes estimados:

Tabela 6

Coeficientes do modelo proposto por Macedo *et al.* (2014)

Variável	Coeficiente	P-valor
Intercepto	5,75	o
Saldo em conta-corrente (<i>cab</i>)	-1,6467	o
(log base 10) PIB <i>per capita</i> (<i>gdppc</i>)	-0,6478	o
Resultado nominal do setor público (<i>gsd</i>)	-3,3652	o
Índice WGI	-6,418	o

(*Continua*)

(Continuação)

Variável	Coefficiente	P-valor
Inflação	2,4554	o
(log base 10) Reservas em moeda estrangeira (trc)	-0,3434	o
Dummy WGI	-2,0206	o
Número de observações	886	n.a.
R2 ajustado	0,892	n.a.

Fonte: Macedo *et al.* (2014).

Macedo *et al.* (2014) estimam o valor de AUROC contínuo proposto por Obuchowski (2005) igual a 88,28%. Calculou-se, no presente trabalho, o valor de AUROC contínuo e mais quatro outras medidas de validação para as probabilidades de *default* estimadas, utilizando-se as PDs de referência associadas aos *ratings* das agências de classificação de risco como *benchmark*, em uma amostra de 123 países e cinco anos (2008 a 2012), totalizando 609 observações:

Tabela 7

Validação do modelo de *shadow rating* para países

AUROC contínuo	0,8211801
AUROC discreto	0,8604042
Somers	0,6936
Coefficiente de correlação	0,6141144
Kendall (p-valor)	0,707 (10 ⁻¹⁴ %)

Fonte: Elaboração própria, com base em valores estimados em *R*.

Os elevados valores para AUROC contínuo, AUROC discreto, estatística de Somers, estatística de Kendall e coeficiente de correlação corroboram a correta especificação do modelo.

Apenas para efeito de exemplificação, reestimou-se o modelo proposto por Macedo *et al.* (2014) para a nova base de dados. Inicialmente, realizou-se o Teste de Chow (1960) em que a hipótese nula reside na assunção de que os mesmos coeficientes são aplicáveis para todos os indivíduos ou períodos. A hipótese alternativa reside na ideia de que existem efeitos fixos para os indivíduos ou períodos. Rejeitou-se a hipótese nula para os indivíduos e não se rejeitou para o período, conforme demonstrado na Tabela 8.

Tabela 8

Teste de Chow

Teste de Chow	Estatística Qui-Quadrado	P-Valor
Entre indivíduos	1,4917	0,47%
Período	2,0072	9,20%

Fonte: Elaboração própria, com base em valores estimados em *R*.

Em seguida, realizou-se o teste de Hausman (1978) para a verificação de efeitos fixos. O teste de Hausman é aplicado sobre as regressões com tratamento de efeitos fixos e aleatórios. As hipóteses nula e alternativa residem na existência de efeitos aleatórios e fixos, respectivamente. Aceitou-se a hipótese nula para a existência de efeitos aleatórios para o período e rejeitou-se para os indivíduos:

Tabela 9

Teste de Hausman

Teste de Hausman	Estatística Qui-Quadrado	Graus de liberdade	P-Valor
Entre indivíduos	278,2449	8	0,00%
Período	10,8651	7	14,46%

Fonte: Elaboração própria, com base em valores estimados em *R*.

Os últimos testes realizados antes da estimação do modelo de regressão foram os testes LM de Breusch-Pagan (1980) e Honda (1985). Em ambos a hipótese nula reside na assunção de que os mesmos coeficientes são aplicáveis para todos os indivíduos. A hipótese alternativa reside na existência de efeitos significativos não tratados. Rejeitou-se a hipótese nula tanto para os indivíduos, que possuem efeitos fixos, como para os períodos, que têm efeitos aleatórios:

Tabela 10

Teste de Breusch-Pagan

Teste LM de Breusch-Pagan	Estatística Qui-Quadrado	Graus de liberdade	P-Valor
Entre indivíduos	515,2925	1	0,00%
Período	515,2925	1	0,00%

Fonte: Elaboração própria, com base em valores estimados em *R*.

Tabela 11

Teste de Honda

Teste LM de Honda	Estatística Normal	P-Valor
Entre indivíduos	22,7001	0,00%
Período	22,7001	0,00%

Fonte: Elaboração própria, com base em valores estimados em *R*.

Estimou-se, então, o modelo econométrico de dados em painel com tratamento para efeitos fixos para indivíduos e obtiveram-se os resultados exibidos na Tabela 12.

Tabela 12

Reestimação dos parâmetros

Variável	Coefficiente	P-valor
Intercepto	5,774265	0,00%
Saldo em conta corrente (cab)	-3,576365	0,05%
(log base 10) PIB <i>per capita</i> (gdppc)	-0,339294	10,99%
Resultado nominal do setor público (gsd)	-2,962724	8,12%
Índice WGI	-6,540152	0,00%
Inflação	2,182832	31,02%
(log base 10) Reservas em moeda estrangeira (trc)	-0,445123	0,00%
<i>Dummy</i> WGI	-1,135342	0,00%
Número de observações	609	n.a.
R2 ajustado	0,83412	n.a.

Fonte: Elaboração própria, com base em valores estimados em R.

Estudo de Caso III: histórico de *default* de países

Através do histórico de eventos de descumprimento de países retirado do trabalho de Reinhart (2010), realizou-se estudo relativo a testes de aderência binários com vistas a verificar a adequação da classificação de risco atual de agências externas ao histórico de *default* de países. De um total de 123 países, a 26 foi atribuído o

valor 1 (um) – indicando elevado histórico de *default* – e ao restante o valor 0 (zero) – indicando baixo ou nenhum histórico de *default*. A classificação de risco de cada país foi obtida por meio da mediana dos *ratings* das agências Fitch, Moody's e Standard & Poors (S&P). Aos níveis de classificação de risco, foram atribuídos os valores inteiros de 1 a 20.

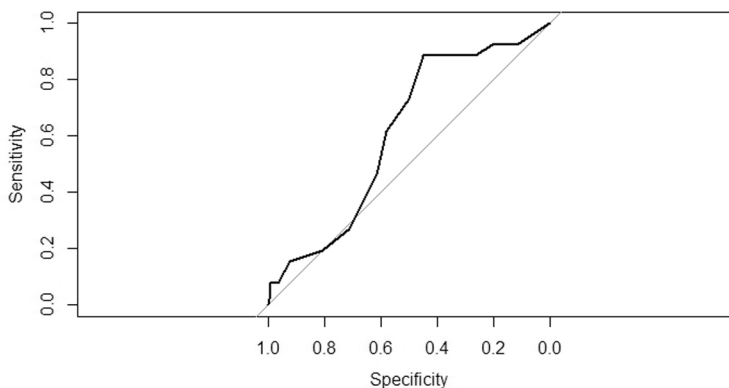
De acordo com Reinhart (2010), os 26 países que apresentam maior número de eventos de descumprimento desde 1800 são: Equador (10), Venezuela (10), Uruguai (9), Costa Rica (9), Brasil (9), Chile (9), Argentina (8), Peru (8), México (8), Turquia (8), Grécia (7), República Dominicana (7), Nicarágua (7), Paraguai (7), Guatemala (7), Áustria (7), Colômbia (7), Espanha (6), Nigéria (5), Rússia (5), Bolívia (5), Gana (5), Tunísia (5), El Salvador (5), Alemanha (4) e Portugal (4).

O objetivo principal desse estudo é, mais do que checar se as PDs atuais refletem o histórico observado de *default*, exemplificar o uso de ferramentas estatísticas para testes binários de forma simples e didática.

Foi encontrado valor baixíssimo de AUROC: 58,66%. O intervalo de confiança de 95% foi igual a [0,4715; 0,7018]. Como o valor de 0,5 está dentro desse intervalo, conclui-se que não é possível rejeitar a hipótese nula de que o modelo não discrimina bons e maus países adequadamente, isto é, que AUROC seja igual a 0,5.

O valor estimado do escore de Brier aplicado sobre a PD de referência da Moody's foi igual a 18,44%. O valor da estatística de Kolmogorov-Smirnov foi de 29,7%, indicando fraco poder discriminatório da classificação de risco em relação à definição de bons e maus países.

Gráfico 3
ROC Soberano



Fonte: Elaboração própria, com base em valores estimados em *R*.

Conclusão

Sem perda de generalidade, viu-se o quão importante é o processo de validação dos sistemas internos de classificação de risco de crédito para as instituições financeiras. As ferramentas estatísticas de validação de sistemas internos de classificação de risco de crédito apresentadas neste trabalho têm o objetivo de medir o poder discriminatório de tais sistemas em relação a contrapartes boas e más e de avaliar sua aderência a classificações de risco de agências externas.

No Estudo de Caso I, relativo ao modelo de *shadow rating* para instituições não financeiras, a utilização das ferramentas estatísticas propostas permitiu avaliar a melhor forma de incorporar a classificação de risco do país à análise. O modelo “topado” pela classificação de risco do país foi aquele que apresentou melhor aderência aos dados de agências externas, de acordo com todas as ferramentas estatísticas utilizadas.

O Estudo de Caso II, responsável pela validação do modelo de *shadow rating* para países, permitiu concluir que o modelo proposto por Macedo *et al.* (2014) tem elevada aderência às classificações de risco externas.

No Estudo de Caso III, em que é avaliado o histórico de *default* de países, os resultados evidenciam que a classificação de risco atual das agências não tem ligação com o longo histórico de *default* dos países relatado por Reinhart (2010). As análises estatísticas realizadas permitiram concluir que a classificação de risco atual não ordena de forma satisfatória boas e más contrapartes, quando tal definição leva em conta os 26 países com maior número de eventos de descumprimento desde 1800. Tal conclusão não é de nenhuma forma inesperada, uma vez que muitos desses eventos são muito antigos e podem não apresentar relevância para a solvência atual das nações.

Na visão de BCBS (2005b), os sistemas internos de classificação de risco de crédito formam a base para a construção de mecanismos de precificação, prêmios de risco e requerimentos de capital. Cada nível de classificação de risco deve estar associado a uma determinada PD. Nesse sentido, uma interessante extensão do presente trabalho reside no estudo de métricas quantitativas capazes de avaliar a acurácia das estimativas de PD para diferentes níveis de *rating*.

A atividade de validação passa por aspectos computacionais, matemáticos e estatísticos extremamente sofisticados e, simultaneamente, pelo entendimento de questões contábeis e financeiras – no limite, envolvendo até a análise de aspectos externos às instituições financeiras, como a probabilidade de ocorrência de eventos adversos relativos a seus fundos de pensão. A tendência é que as áreas de validação independente, em sua maioria criadas para analisar e centralizar as discussões em torno do ICAAP, sejam, cada vez mais, um importante pilar na gestão de risco das instituições financeiras.

Referências

- ANDRADE, F. W. M.; OLIVEIRA, J. G. C. Comparação entre medidas de performance de modelos de credit scoring. *Revista Tecnologia de Crédito*, Serasa Experian, 33 ed., 2012. Disponível em: <http://www.serasaexperian.com.br/serasaexperian/publicacoes/revistas/2002/33/revista_0180.htm>. Acesso em: 18 jul. 2014.
- ANNIBAL, C. A. O poder discriminante das operações de crédito das instituições financeiras brasileiras. *Trabalhos para Discussão*, 167. Banco Central do Brasil, 2008. Disponível em: <<https://www.bcb.gov.br/pec/wps/port/wps167.pdf>>. Acesso em: 24 jul. 2014.
- BACEN – BANCO CENTRAL DO BRASIL. *Circular 3.547*, de 7 de julho de 2011. Estabelece procedimentos e parâmetros relativos ao Processo Interno de Avaliação da Adequação de Capital (ICAAP).
- BCBS – BASEL COMMITTEE ON BANKING SUPERVISION. Update on work of the Accord Implementation Group related to validation under the Basel II Framework. *Newsletter*, 4. Jan. 2005a. Disponível em: <http://www.bis.org/publ/bcbs_nl4.htm>. Acesso em: 24 jul. 2014.
- _____. Studies on the validation of internal rating systems. *Working paper*, 14. Bank for International Settlements, May 2005b. Disponível em: <http://www.bis.org/publ/bcbs_wp14.pdf>. Acesso em: 24 jul. 2014.
- _____. *Developments in modelling risk aggregation*. Bank for International Settlements, 2010.
- BREUSCH, T.; PAGAN, A. The Lagrange Multiplier Test and its applications to model specification in econometrics. *Review of Economic Studies*, 47, p. 239-253, 1980.
- BRIER, G. W. *Verification of forecast expressed in terms of probability*. 1950.

- CARDOSO, V. S. *et al.* Assessing corporate risk: a PD model based on credit ratings. *ACRN Journal of Finance and Risk Perspectives*, vol. 2, Issue 1, p. 51-58, 2013.
- CHOW, G. C. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28 (3): 591-605, 1960.
- ENGELMANN, B. Measures of a rating's discriminative power: applications and limitations. In: ENGELMANN, B.; RAUHMEIER, R. *The Basel II risk parameters*. Springer, 2006, cap. IV, p. 39-77.
- ERLENMAIER, U. The shadow rating approach – experience from banking practice. In: ENGELMANN, B.; RAUHMEIER, R. *The Basel II risk parameters*. Springer, 2006, cap. IV, p. 39-77.
- GRÜN, B. *et al.* Deriving consensus ratings of the big three rating agencies. *Research Report Series*, 99. Department of Statistics and Mathematics, Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna, 2010.
- HAUSMAN, J. A. Specification tests in econometrics. *Econometrica*, 46 (6), p. 1.251-1.271, 1978.
- HONDA, Y. Testing the error components model with non-normal disturbances. *Review of Economic Studies*, 52, p. 681-690, 1985.
- KENDALL, M.G. *Rank correlation methods*. 4th ed. Griffin, 1976.
- MACEDO, H. F. *et al.* Estimando o risco-país: um modelo de probabilidade de *default* baseado em *ratings*. *Revista do BNDES*, Rio de Janeiro, BNDES, n. 41, p. 415-434, 2014.
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18 (1): 50-60, 1947.

MEIRELLES, A. M. *Regulação e supervisão do sistema financeiro num contexto de crise internacional*. Cabo Verde, Banco Central do Brasil: Cabo Verde, set. 2010. Disponível em: <http://www.bcb.gov.br/pec/appron/apres/Regula%E7%E3o_financeira_na_crise_Cabo_Verde.pdf>. Acesso em: 24 set. 2014.

OBUCHOWSKI, N. A. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Academic Radiology*, Elsevier, n. 12, p. 1.198-1.204, 2005.

REINHART, C. M. This time is different chartbook: country histories on debt, default, and financial crises. *Working Paper*, 15.815. National Bureau of Economic Research, 2010.