



UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

Risco de Crédito:
Estudo do Valor da Exposição no Momento de
Descumprimento

Thaís Almeida Nunes

Thiago Rodrigues Alencastro Veiga

Relatório Final

Orientador: Prof. Donald Matthew Pianto

Coorientadora: Edna Sousa Lima

Brasília

Junho 2011

Thaís Almeida Nunes

Thiago Rodrigues Alencastro Veiga

**Risco de Crédito:
Estudo do Valor da Exposição no Momento de
Descumprimento**

Relatório apresentado à disciplina Estágio
Supervisionado do curso de graduação em
Estatística, Departamento de Estatística,
Instituto de Exatas, Universidade de Brasília,
como parte dos requisitos necessários para
o grau de Bacharel em Estatística.

Orientador: Prof. Donald Matthew Pianto

Coorientadora: Edna Sousa Lima

Brasília

Junho 2011

Agradecimentos

Primeiramente agradeço a Deus pela paciência, sabedoria e disciplina que Ele depositou em mim proporcionando-me a felicidade de ter alcançado mais um objetivo em minha vida.

Agradeço aos meus pais, por sempre acreditarem que educação é a maior herança que os pais podem deixar para os filhos.

Agradeço ao meu namorado Rafael Mezzomo por todo companheirismo e paciência durante todo este tempo ao meu lado.

Ao meu irmão Rafael Almeida por sempre estar ao meu lado quando eu preciso.

Aos meus amigos e companheiros de curso que me orientaram e me ajudaram chegar até aqui. Em especial: Joyce Silvino, Larissa Sayuri, Maísa Cardoso e Taísa Costa. Obrigada por tudo, meninas!

Ao meu amigo e companheiro de monografia Thiago Rodrigues pela paciência e incansáveis horas de estudos nesses últimos anos.

Por fim, ao nosso orientador Professor Donald Pianto e a nossa coorientadora Edna Souza Lima por toda atenção e apoio.

Thaís Almeida Nunes

Agradecimentos

Agradeço primeiramente a Deus por dar-me saúde e disposição em cada dia de minha vida e por colocar sempre as pessoas certas ao meu lado.

Agradeço aos meus pais por acreditarem em minha capacidade e por terem abdicado de grande parte de suas vidas para me proporcionar educação pessoal e profissional.

Ao meu irmão Paulo César, por toda sua sincera amizade e companheirismo.

Agradeço aos meus grandes amigos pessoais por todo o apoio prestado e pelas palavras de amizade e de incentivo. Em especial: Carlos Fellipe, Helder Ribeiro, Iury Rodrigues, Joyce Silvino, Larissa Sayuri e Maísa Cardoso.

A minha grande amiga e companheira de monografia Thaís Almeida pela paciência e dedicação empregadas em nosso trabalho.

Finalmente agradeço ao nosso orientador Professor Donald Pianto e a nossa co-orientadora Edna Souza pelo apoio e pela paciência demonstrada por ambos.

Thiago Rodrigues Alencastro Veiga

Epígrafe

*A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original.*

Albert Einstein

Resumo

Este trabalho tem como intuito estudar técnicas de possível utilização na estimação do valor da exposição de cada operação no momento em que ocorre sua respectiva inadimplência (EAD). Restringimos este estudo somente para os casos em que possa ocorrer variabilidade do valor exposto em decorrência de possíveis retiradas adicionais amparadas em um limite pré-aprovado e não cancelável unilateralmente pela instituição que concede o crédito. Apresentaremos uma simulação que permita estimar o valor dessa exposição para as operações presentes num grupo homogêneo de risco.

Na simulação deste estudo, aleatoriamente geramos inadimplência usando o *behavior score* dos clientes e selecionamos aleatoriamente estimativas do parâmetro de estudo geradas por regressões quantílicas para 11 quantis diferentes. Em seguida a distribuição da EAD da simulação é comparada com o valor real observado na base de dados.

Palavras-chaves: Descumprimento, Inadimplência, Exposure at Default (EAD), FCC, Basileia II, Simulação, Estimação.

Sumário

1. Introdução.....	1
1.1 Objetivos	3
1.2 Metodologia.....	4
2. Tipos de fatores de conversão utilizados	5
3. Tipos de horizontes utilizados para cálculo dos fatores FCC observados.....	7
3.1 Método do Horizonte Fixo	8
3.2 Método de Cortes	10
3.3 Método de Horizonte Variável	11
3.4 Conclusão sobre os métodos	13
4. Base de dados para estimação	14
4.1 Estrutura do banco	14
4.2 Como lidar com fatores FCC observados negativos	14
4.3 Como lidar com fatores LEQ observados maiores que um	16
5. Estimação da EAD	18
5.1 Relação entre o banco de dados e a carteira atual	18
5.2 Equivalência entre estimações de EAD(f) e estimações de FCC(f)	18
5.3 Modelagem dos fatores de conversão a partir do banco de dados.....	19
5.4 Casos para quando LEQ é constante para observações similares a {f,t}	22
5.5 Método do Coeficiente de Utilização do Limite, com CCF Constante.....	23
6. Estudo de caso.....	25
6.1 Base de dados utilizada.....	25
6.2 Análise Exploratória	26
6.3 Conceito de Grupos homogêneos de risco.....	31
6.4 Utilizando a Arvore de decisão para criação dos grupos homogêneos.....	32
6.5 Metodologia de Simulação a ser utilizada	33
6.5.1 Primeira etapa: Geração de regressões quantílicas na Base de Estimação... 34	
6.5.2 Segunda etapa: Geração dos descumprimentos por meio do <i>Behavior Score</i>	35

6.5.3 Terceira etapa: Associação de fator LEQ por meio de seleção aleatória de um quantil	36
6.5.4 Quarta etapa: Cálculo da distribuição da EAD esperada para o Portfólio	36
6.6 Resultados.....	37
7. Conclusão.....	46
Referências Bibliográficas	47
Apêndice	49

1. Introdução

Com o objetivo de minimizar as desigualdades competitivas entre os bancos internacionalmente ativos bem como reforçar a solidez e a estabilidade do sistema bancário internacional, as autoridades de supervisão bancária no ano de 1988, após intenso processo de discussão, celebraram o Acordo de Basiléia. Tal Acordo definia mecanismos para mensuração do risco de crédito e estabelecia a exigência de capital mínimo para suportar riscos. Por ser o primeiro dentre os dois ocorridos até os dias de hoje, este acordo ficou conhecido como Basiléia I.

Entretanto Basiléia I não impediu a falência de inúmeras instituições financeiras ao longo da década de 90. Nesse sentido, o Comitê lançou em 2004 um novo acordo para regular o requerimento de capital. Basiléia II tinha os seguintes objetivos: promover a estabilidade financeira, fortalecer a estrutura de capital das instituições, favorecer a adoção das melhores práticas de gestão de riscos e estimular maior transparência e disciplina de mercado. O novo acordo é baseado em três grandes pilares: fortalecimento da estrutura de capitais das instituições, estímulo à adoção das melhores práticas de gestão de riscos e redução da assimetria de informação e favorecimento da disciplina de mercado.

Uma das grandes mudanças trazidas por “Basiléia II” foi justamente a possibilidade de que o requerimento de capital das instituições financeiras autorizados por seus respectivos órgãos controladores seja feito com base na estimação da perda esperada através de modelos de risco internos à própria instituição.

A possibilidade de criar modelos de crédito internos para embasar a alocação de capital vem tornando a gestão de crédito uma ferramenta cada vez mais importante dentro das instituições financeiras. Isto por que acredita-se que a instituição que conseguir realizar uma alocação de capital adequada a sua realidade estará fortemente preparada para lidar com perdas inesperadas sem que, por outro lado, tenha que alocar quantidades excessivas de capital, o que a forçaria a restringir o crédito tornando-o mais caro. “Basiléia II” reforça a visão de que um sistema de gestão de risco eficiente não representa apenas a segurança financeira da instituição, mas

também um instrumento capaz de possibilitar aumento da concessão de crédito e, conseqüentemente, uma ampliação dos lucros.

Para entender o que é a perda esperada é necessário entender o conceito do termo “descumprimento”. O termo “descumprimento” denota a qualidade de inadimplência e, segundo o Banco Central do Brasil (Comunicado 18.365 de 2009), o descumprimento se dará na ocorrência de pelo menos um dos eventos a seguir:

- A instituição financeira considera que o tomador ou contraparte não irá honrar integralmente a obrigação em questão sem que a instituição financeira recorra a ações tais como a execução de garantias prestadas;
- A obrigação está em atraso há mais de 180 dias, para as exposições de varejo relativas a financiamentos habitacionais, ou há mais de 90 dias para as demais exposições de varejo.

Após compreender esse conceito pode-se entender os parâmetros de Basiléia II utilizados para estimar a perda que o banco espera ter ao conceder determinado crédito, a dita perda esperada.

Inicialmente, para estimar a perda esperada é necessário que o banco conheça qual a probabilidade do cliente tornar-se inadimplente, ou descumprir. Essa probabilidade será dada pelo parâmetro PD (Probability of Default, ou Probabilidade de Descumprimento). Caso o cliente torne-se inadimplente a instituição muito provavelmente tentará recuperar parte do crédito concedido, enquanto a outra fração do valor não será recuperada. Esse percentual do valor exposto que não é recuperado é dado pelo parâmetro LGD (Loss Given Default, ou Perda dado o Descumprimento). Finalmente, para estimar a perda esperada, é necessário que a instituição conheça o montante referente a cada operação que entrou em descumprimento, esse montante será dado pelo terceiro parâmetro de Basiléia II, a EAD (Exposure at Default, ou Exposição no Momento de Descumprimento). Depois de calculados esses parâmetros a perda esperada será dada pelo produto dos três.

De acordo com Valvonis (2008) existe muito material escrito acerca de modelagem de PD e um pouco menos sobre modelagem da LGD. Já em relação à EAD existem poucos estudos publicados. Por essa razão esse trabalho será voltado ao estudo desse parâmetro.

A EAD é definida pelo Banco Central do Brasil (Comunicado 18.365 de 2009) como sendo o parâmetro que “corresponde ao valor da exposição da instituição financeira perante o tomador ou contraparte no momento da concretização do evento de descumprimento, bruto de provisões”. Estimar o valor dessa exposição em geral não é algo complexo já que para as operações registradas em balanço essa estimativa é o próprio saldo contábil da operação. Entretanto, segundo Moral (2006), o problema surge nas operações para as quais o valor exposto pode sofrer algum tipo de variação decorrente de um possível aumento da exposição por meio de retiradas adicionais amparadas em um limite pré-aprovado.

Para esse último caso a estimativa da EAD pode ser feita de duas maneiras. A primeira seria feita levando em consideração um componente chamado de fator de conversão de crédito, ou CCF (Credit Conversion Factor), que multiplicaria o valor do limite total disponível. A outra forma de estimar a EAD seria através de outro fator que se aplica a parte do limite disponível mas não utilizada, o LEQ (Loan Equivalent Exposure).

No Brasil as grandes instituições financeiras estão em fase de criação dos modelos de Basileia II. Portanto, parece de grande relevância que exista um estudo acadêmico apontando uma ou mais técnicas que possam ser utilizadas na estimativa da EAD. Tal estudo poderia certamente servir para auxiliar na estimativa deste parâmetro pelas instituições financeiras possibilitando um cálculo mais adequado da perda esperada.

1.1 Objetivos

O objetivo geral deste trabalho é estudar e apresentar algumas técnicas utilizadas na estimativa do valor da exposição de cada operação no momento em que ocorre sua respectiva inadimplência (EAD) bem como apresentar uma simulação que permita estimar o valor dessa exposição para as operações presentes num grupo homogêneo de risco. Esses estudos deverão ser feitos, mais especificamente, para os casos em que possa ocorrer variabilidade do valor exposto em decorrência de possíveis retiradas adicionais amparadas em um limite pré-aprovado e não cancelável unilateralmente pela instituição que concede o crédito.

Os objetivos específicos do trabalho são:

- Apresentar brevemente os dois fatores de conversão utilizados para estimar a EAD de operações fora do balanço, o CCF e o LEQ;
- Estudar os principais tipos de horizontes utilizados para selecionar as datas de referencia utilizadas no cálculo dos fatores CCF e LEQ observados, apresentando as principais vantagens e desvantagens de cada um;
- Fazer um breve estudo de como deveria ser montado um banco de dados utilizado para estimação da EAD;
- Desenvolver uma metodologia simples para segregar as operações em grupos homogêneos de risco;
- Estudar e apresentar uma metodologia de simulação que consiga estimar o valor da EAD para exposições de um portfólio e comparar as estimativas com os valores observados do parâmetro.

1.2 Metodologia

A base de dados a ser utilizada foi cedida por uma Instituição Financeira Brasileira de grande porte. Para manipulá-la e prepará-la para os estudos a serem realizados será utilizado o software SAS[®] 9.2.

No processo de criação dos grupos homogêneos de risco será desenvolvida uma árvore de decisão por meio do procedimento Decision Tree do SAS Enterprise Miner[®] 5.3.

Já para realizar a estimação da EAD será utilizado um processo de simulação para selecionar aleatoriamente estimativas do parâmetro de estudo geradas por regressões quantílicas para 11 quantis diferentes. Em seguida serão comparados os resultados da simulação com os valores reais observados na base de dados. Para essa etapa final do estudo será utilizado o software R[®] 2.13.0.

2. Tipos de fatores de conversão utilizados

As operações não registradas em balanço são aquelas nas quais o tomador de crédito pode realizar novos saques amparados em um limite pré-existente e, dessa forma, aumentar o valor da exposição ao longo da vida do contrato. A estimação da EAD para tais operações requer um tratamento diferenciado que considere justamente essa possível variabilidade, decorrente dos incrementos citados, no valor da exposição.

Existem basicamente dois fatores de conversão utilizados para tratar a estimação da EAD para uma operação ativa f não registrada em balanço:

- O primeiro considerará que a $EAD_{(f)}$ pode ser estimada através de um fator de conversão $CCF_{(f)}$ que multiplica o limite total $L_{(f)}$:

$$EAD_{(f)} = CCF_{(f)} * L_{(f)} \quad [1]$$

- O segundo considerará que a $EAD_{(f)}$ pode ser estimada através da soma do valor exposto $E_{(f)}$ com o produto entre um fator de conversão $LEQ_{(f)}$ e o limite disponível e não utilizado $LN_{(f)}$:

$$EAD_{(f)} = E_{(f)} + LEQ_{(f)} * LN_{(f)} \quad [2]$$

Para o escopo de operações tratadas neste estudo o processo de estimação da $EAD_{(f)}$ se resumirá basicamente à estimação dos fatores de conversão, seja o $LEQ_{(f)}$ ou o $CCF_{(f)}$.

Para realizar tal estimativa é necessário que se utilizem informações das operações f obtidas em determinadas datas de referencia tr anteriores à data do descumprimento. As diferentes técnicas utilizadas na escolha dessas datas de referencia tr serão mencionadas posteriormente.

Dessa forma, seja g uma operação que entrou em descumprimento e tr a data de referência selecionada para a obtenção das variáveis explicativas necessárias à estimação. Deste modo sabe-se que a $EAD_{(g, tr)}$ será dada por:

$$EAD_{(g, tr)} = E_{(tr)} + LEQ_{(g, tr)} * (L_{(tr)} - E_{(tr)}) \quad [3]$$

Onde:

$E_{(tr)}$: valor da exposição na data de referencia tr ;

$L_{(tr)}$: Limite total da operação observado em tr ;

Dessa forma o fator $LEQ_{(g, tr)}$ pode ser dado por:

$$LEQ_{(g, tr)} = \frac{E_{(td)} - E_{(tr)}}{L_{(tr)} - E_{(tr)}} \quad [4]$$

Onde: $E(td) = EAD(g, tr)$ é a exposição no momento de descumprimento.

Ou ainda:

$$LEQ_{(g, tr)} = \frac{ead_{(g, tr)} - e_{(tr)}}{1 - e_{(tr)}} \quad [5]$$

Onde:

$ead_{(g, tr)}$: exposição percentual no momento do descumprimento;

$e_{(tr)}$: exposição percentual no momento de referência tr ;

Utilizando essa fórmula é possível obter sem grandes esforços o fator $LEQ_{(g, tr)}$ observado, entretanto a utilização dessa expressão pode se mostrar inconsistente nos casos em que $L_{(tr)} = E_{(tr)}$ ou ainda quando $L_{(tr)} \approx E_{(tr)}$.

Por isso algumas instituições utilizam para o cálculo do valor da EAD um fator de conversão baseado no limite total da operação g dado por:

$$CCF_{(g, tr)} = \frac{EAD_{(g, tr)}}{L_{(tr)}} = ead_{(g, tr)} \quad [6]$$

Embora esse fator de conversão seja menos utilizado ele possui a vantagem de estar bem definido tanto quando $L_{(tr)} = E_{(tr)}$ e quando $L_{(tr)} \approx E_{(tr)}$.

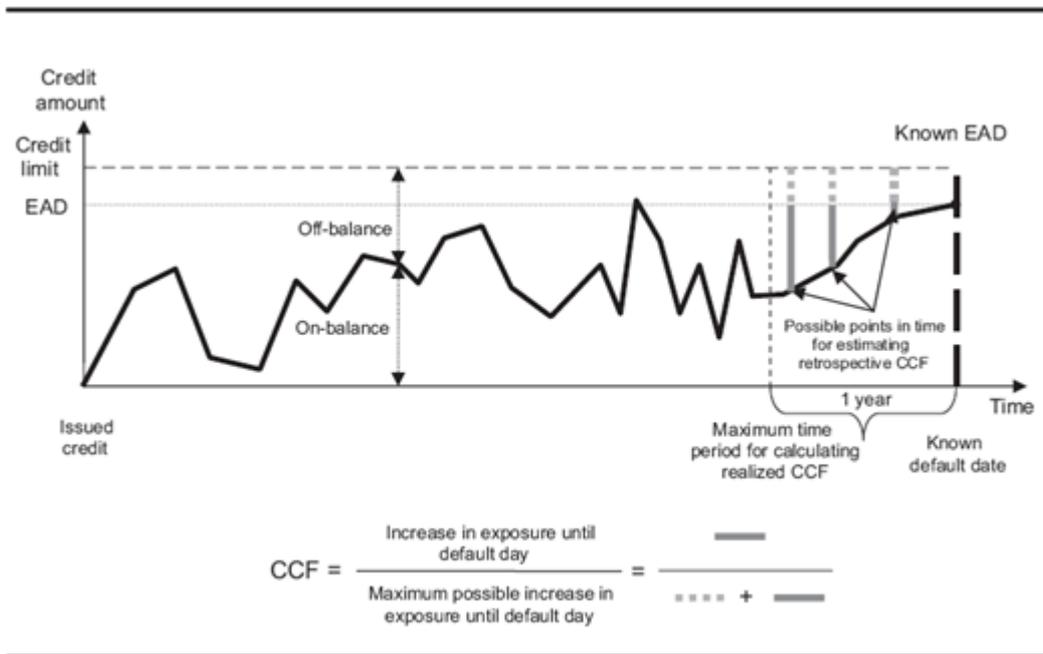
O Novo Acordo de Capitais evidencia no parágrafo 310 que o CCF deve multiplicar o valor comprometido, mas não sacado. Entretanto a norma brasileira não explicita que o fator calculado deve necessariamente multiplicar o limite não utilizado. Portanto, para evitar eventuais confusões com a nomenclatura não será feita distinção entre o fator LEQ e CCF na maior parte do estudo, e trataremos o fator de conversão de uma forma generalizada chamando-o pela sigla em português FCC (Fator de Conversão de Crédito).

3. Tipos de horizontes utilizados para cálculo dos fatores FCC observados

Evidentemente se analisarmos o FCC com base na própria data de descumprimento, o valor deste será zero, pois nessa data $EAD = E (tr)$. O que interessa é analisar o incremento do valor da exposição ao longo do tempo até a data de descumprimento. Pode-se, por exemplo, analisar a evolução desse valor na semana que antecede ao evento de descumprimento, ou no mês que antecede tal evento, por exemplo.

Quando analisamos a figura 1 fica claro que a barra tracejada, que representa a diferença entre o Limite Total e a EAD, é constante para todo o horizonte de tempo que sucede a data de descumprimento. Por outro lado a barra contínua, que representa a utilização adicional de limite de crédito não utilizadas em relação à EAD final, apresenta variação ao longo do tempo e sua altura depende do momento em que o FCC é estimado. Dessa forma, de acordo com Valvonis (2008), essa variabilidade representa justamente o fato de que o FCC calculado pode ser diferente para cada horizonte de tempo utilizado no cálculo. Portanto é de fundamental importância estabelecer uma técnica adequada para selecionar tais horizontes de tempo. As principais técnicas existentes são apresentadas a seguir.

Figura 1 – Estimação da CCF realizada para a EAD



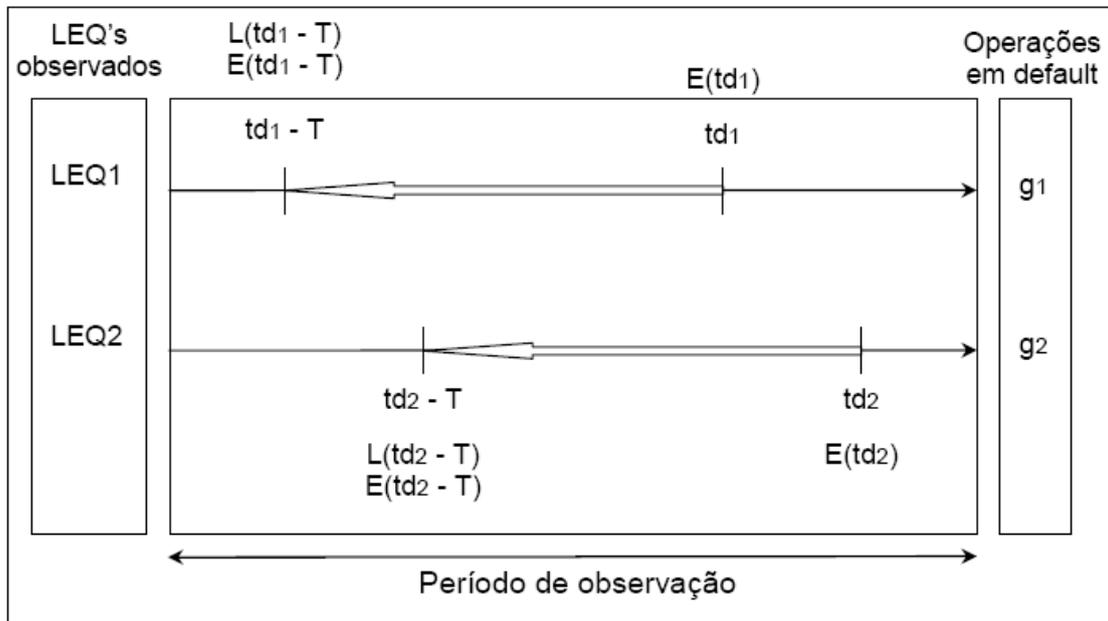
Fonte: Valvonis, 2008.

3.1 Método do Horizonte Fixo

Primeiro selecionamos um horizonte T , normalmente de um ano, e em seguida para cada operação que entra em descumprimento para a qual é verificada que $L(td - T) \neq E(td - T)$ calcula-se um fator CCF. Neste caso, usar-se como data de referência $TR = td - T$, como é observado a seguir:

$$CCF(td - T) = \frac{E(td) - E(td - T)}{L(td - T) - E(td - T)} \quad [7]$$

Figura 2 – Fatores LEQ observados com o enfoque de horizonte fixo



Fonte: Moral, 2006

Essa metodologia possui as seguintes vantagens em relação às demais:

- Existe boa dispersão entre as datas de referência
- Maior homogeneidade dos fatores FCC calculados em consequência do uso de um horizonte comum para todas as operações, neste caso $T = td - TR$
- É o método mais simples de executar.

Por outro lado o método tem as seguintes desvantagens:

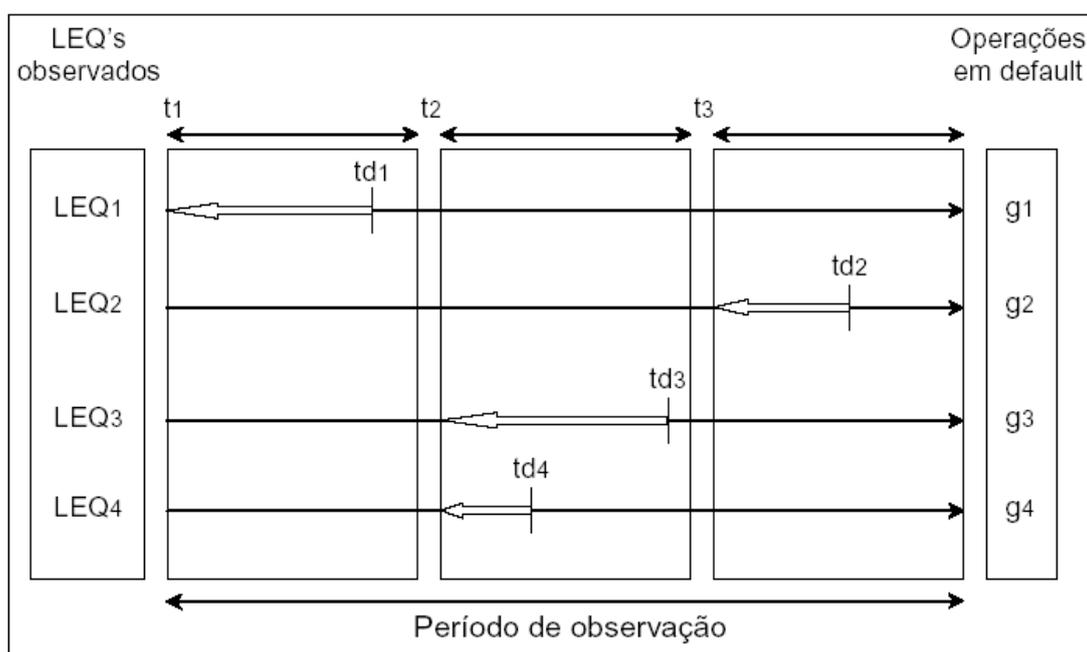
- Não é possível trabalhar com as operações cuja maturidade na data de descumprimento seja inferior ao horizonte T ;
- Não é levada em consideração toda a informação relevante, já que, para cada operação g que entra em descumprimento durante o período de observação, só se utiliza uma observação dada por $\{g, td - T\}$.

3.2 Método de Cortes

Nesta abordagem divide-se primeiramente o período de observação em janelas de tempo fixas, geralmente de um ano, chamadas de cortes. Cada operação é agrupada no corte que contém sua respectiva data de descumprimento. Para calcular um fator FCC_i observado associado a cada operação, utiliza-se como data de referência o início da janela de tempo para a qual está associada a operação, $\{t_1, t_2, \dots, t_i, \dots, t_n\}$:

$$FCC(ti) = \frac{E(td) - E(ti)}{L(ti) - E(ti)} \quad [8]$$

Figura 3 – Fatores LEQ observados com o enfoque de cortes



Fonte: Moral, 2006.

As principais desvantagens deste método são as seguintes:

- As informações relevantes não são totalmente utilizadas, já que para cada operação g , que entra em descumprimento durante o período de

observação, só é utilizada a informação dada por $\{g, t_j\}$, onde j depende do corte que contém td

- As datas de referência não são dispersas ao longo do horizonte de tempo analisado;
- Os fatores FCC_i observados são menos homogêneos que os que seriam obtidos utilizando o método de horizonte.

A principal vantagem desse método é:

- Considera a possibilidade de que as exposições podem entrar em descumprimento em qualquer momento durante o ano em questão.

3.3 Método de Horizonte Variável

Primeiramente é fixado um intervalo para o horizonte em que se quer calcular os fatores FCC_i , (normalmente de 1 ano). Depois, para cada operação que entra em descumprimento, são calculados os fatores FCC_i observados associados a um conjunto de datas de referência anteriores ao descumprimento.

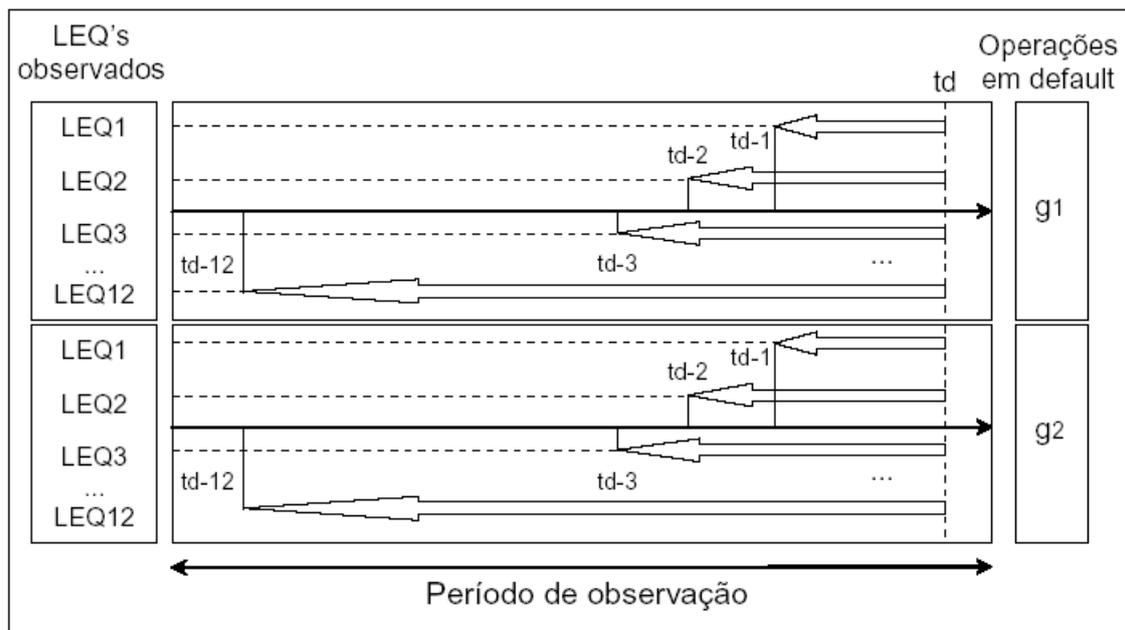
Neste método, leva-se em consideração um conjunto amplo de possíveis intervalos de descumprimento quando se vai estimar o FCC. Considerando o intervalo de um ano, ou seja, ($j = 1, 2, \dots, 12$), tem-se:

$$\{FCC(td - j) = \frac{E(td) - E(td - j)}{L(td - j) - E(td - j)}, j=1, \dots, 12 \text{ meses}\} \quad [9]$$

Desse modo pode-se associar 12 fatores FCC_i observados a cada operação que entra em descumprimento. No entanto, estes fatores FCC_i são claramente não homogêneos, já que alguns desses valores foram calculados utilizando observações muito próximas da data de descumprimento ($i = \{g, td - 1\}$) e outros estão baseadas nas observações de um ano antes do descumprimento ($j = \{g, td - 12\}$).

De acordo com Moral (2006) essa falta de homogeneidade seria mais problemática apenas nas operações para as quais a instituição financeira tem a política de acompanhar e fazer algum tipo de marcação apontando uma situação de anormalidade. Essa marcação poderia acontecer porque quando algumas operações atingem determinada quantidade de dias de atraso, mesmo que essa quantidade seja inferior à quantidade de dias estabelecida pelo órgão controlador como critério de descumprimento, a instituição pode atuar, à critério de suas políticas internas de gestão de risco, de modo a impedir retiradas adicionais de crédito. Nesses casos só deve ser realizado o cálculo dos fatores FCC associados às datas de referência anteriores à referida marcação.

Figura 4 – Fatores LEQ observados com o enfoque de horizonte variável



Fonte: Moral, 2006.

A principal vantagem desse método é que:

- Ele possui a vantagem de levar em consideração um número maior de observações. Portanto, acredita-se que os procedimentos de estimação baseados neste enfoque possam produzir estimadores mais estáveis e mais precisos, desde que observada a possibilidade de que a instituição atue, como mencionado no parágrafo anterior, de modo a impedir novas retiradas de crédito.

As desvantagens do método são que:

- Ele exige que a instituição financeira armazene maior quantidade de informações.
- A estimação deve levar em conta a não independência entre as diferentes observações geradas pelo mesmo contrato.

3.4 Conclusão sobre os métodos

Sabe-se que a estimação do FCC, e conseqüentemente da EAD, deve ser feita em consonância com a estimação da PD. Como a estimativa da PD é feita para um horizonte de tempo de um ano a janela de tempo para a estimativa da FCC não deve ser superior a esse período. Por isso sugere-se um ano como o horizonte T no primeiro método, janelas de um ano para os cortes do segundo método e o cálculo de 12 fatores FCC para cada operação no terceiro método.

Se partirmos do pressuposto de que quanto mais próximo da data de descumprimento maior deverá ser o incremento na exposição realizado pelo cliente pode-se concluir que o método de Horizonte fixo será o mais conservador. Isso porque a janela de tempo média nesse método será maior que no Método de Cortes e no Método de Horizontes Variáveis. Logo o FCC tenderá a ser maior quando calculado pelo primeiro método, o que pode se mostrar adequado no Brasil que é um país bastante conservador em suas políticas de crédito.

4. Base de dados para estimação

4.1 Estrutura do banco

Caso deseje-se testar o cálculo dos fatores FCC por meio do Método de Horizonte Fixo a estrutura de banco de dados adequada aos procedimentos deverá ser construída de forma a atender à necessidade de utilização da faixa de referencia tr como uma das variáveis chave para cada observação. Dessa forma seria interessante construir uma estrutura de dados onde cada observação tenha como chave o par $i=(g, tr)$. Tais observações deveriam conter:

- As principais características da operação g , aquelas que independem da faixa de referencia tr (tipo de operação, código de identificação do cliente, código de identificação do contrato, etc);
- As variáveis com provável poder de predição relacionadas à g e à tr , inclusive as variáveis utilizadas para o calculo do fator de conversão (data de referencia tr , data de descumprimento td , exposição na data de referencia $E(tr)$, limite na faixa de referencia $L(tr)$, renda $R(tr)$, score $S(tr)$, etc);
- O valor observado da EAD

Além disso, seria necessário que houvesse uma definição clara do conceito de descumprimento utilizado no banco construído, tal conceito deve ser coerente ao conceito utilizado nas estimações dos parâmetros PD e LGD.

4.2 Como lidar com fatores FCC observados negativos

É comum que ocorram valores negativos do fator FCC observado. Isso acontecerá sempre que $E(td) < E(tr)$. Tal situação é especialmente comum quando observa-se que o valor percentual da exposição de crédito na data de referencia é próximo a um. Para mostrar isso matematicamente pode-se usar a equação 4, de onde segue que:

$$\frac{\partial LEQ_i}{\partial e_{(tr)}} = \frac{(-1)(1 - e_{(tr)}) - (e_{(td)} - e_{(tr)})(-1)}{(1 - e_{(tr)})^2} = -\frac{1 - e_{(td)}}{(1 - e_{(tr)})^2} \quad [10]$$

Observa-se que a derivada do LEQ, ou FCC, com respeito à $e_{(tr)}$ cresce à medida que $(1 - e_{(tr)})^2$ diminui. Ou seja, a distribuição empírica dos fatores FCC_i terá sua variabilidade aumentada para valores de $e_{(tr)}$ próximos a um.

Os bancos utilizam-se da média do fator FCC_i para estimar $FCC_{(f)}$, e tais médias, seriam substancialmente afetadas pela ocorrência de fatores negativos e grandes em valor absoluto. Dessa forma é necessário que a instituição financeira adote alguma medida para lidar com o referido problema. As soluções mais comuns são:

- Censura de dados (definindo o fator FCC_i como sendo $FCC_i^+ = \max[0, FCC_i]$);
- Eliminação das observações para as quais o fator FCC_i observado é negativo;
- Manter os fatores FCC_i negativos inalterados, mas fixar um valor mínimo para as estimações, $FCC_{(f)} \geq 0$.

A eliminação não seria muito recomendada uma vez que é difícil de ser justificada. Já a censura dos dados pode ser justificada argumentando-se que introduziria um viés conservador do ponto de vista regulamentar. Entretanto, a opção mais natural e comum é a de manter os fatores FCC_i negativos inalterados fixando-se o zero como valor mínimo para a estimação de $FCC_{(f)}$. No caso de adotar-se a ultima medida deve-se assegurar que os fatores FCC_i negativos não são excessivamente grandes em valor absoluto de forma a subestimar as estimativas de $FCC_{(f)}$.

4.3 Como lidar com fatores LEQ observados maiores que um

Observando a definição do fator FCC imagina-se que seja impossível a ocorrência de valores maiores que um. No entanto essa ocorrência é relativamente comum. Isso se deve a dois principais motivos.

O primeiro deles é a existência de alteração de limite do contrato ao longo do período analisado. Nesse caso, segundo Valvonis (2008), qualquer alteração no valor do limite deve ser excluída do cálculo dos fatores FCC observados ou tratada como uma nova operação. Segundo ele, caso ocorra alteração no limite de determinada operação e, posteriormente, esta entre em descumprimento, deve-se considerar que a primeira operação, antes da alteração do limite, não entrou em descumprimento, enquanto a segunda, após a alteração, deverá ser considerada como inadimplente. Isso porque a estimativa da EAD deverá ser feita para as exposições atuais do Banco, desconsiderando eventuais possibilidades de alteração de limites. Caso o limite do cliente seja alterado este deverá ser reavaliado gerando uma outra estimativa de EAD.

Outra causa comum para a ocorrência de fatores FCC_i observados maiores que um é o incremento da taxa de juros no valor da dívida total do cliente. No ponto de vista regulamentar, em geral, o contrato só será considerado como inadimplente após completar 90 dias em atraso. No entanto é muito possível que o cliente esteja usando todo o seu limite no momento em que se inicia o atraso e, dessa forma, será também muito comum que, ao completar os 90 dias, momento do qual se extrai a informação do valor exposto $E_{(td)}$, a soma dos valores de encargos e juros ao valor da dívida torne a exposição $E_{(td)}$ maior que o próprio limite $L_{(tr)}$ e, conseqüentemente, torne o fator FCC_i maior que um.

O tratamento dos fatores FCC_i maiores que um é mais delicado que o tratamento dos fatores negativos. Segundo Moral (2006), embora algumas instituições fixem o valor máximo do LEQ_i a tal procedimento não é muito recomendável, nem para fins internos nem para fins regulamentares. O mais importante nesses casos seria

identificar de forma clara a origem das distorções ocorridas para que a instituição possa tomar uma decisão coerente e justificável.

5. Estimação da EAD

5.1 Relação entre o banco de dados e a carteira atual

Serão apresentados aqui alguns métodos comumente utilizados para a estimação da EAD de operações f que não estão inadimplente no momento t . Tais métodos baseiam-se em um subconjunto de RDS (Reference data set) que agrupa observações de operações que entraram em descumprimento e são parecidas com a operação associada a f . Esse subconjunto pode ser entendido como um grupo homogêneo de risco, e as estimativas geradas podem variar de um grupo homogêneo para outro mesmo nos casos em que esses grupos contenham o mesmo tipo de operação.

O conceito de grupo homogêneo de risco, bem como uma possível técnica de criação deste serão melhor apresentados na seção 6.

5.2 Equivalência entre estimações de EAD(f) e estimações de FCC(f)

Para melhor compreensão dessa seção em diante voltaremos a fazer distinção entre os fatores de conversão CCF e FCC, que deverão ser entendidos como foi mostrado na seção 2.

Considere uma operação f da carteira atual que não entrou em descumprimento. Dado um estimador de EAD(f), se $L(f) \neq E(f)$ a estimação anterior pode ser construída em termos de LEQ(f), como:

$$EAD(f) = E(f) + LEQ(f) * (L(f) - E(f)) \quad [11]$$

Onde LEQ(f) é:

$$LEQ(f) = \frac{EAD(f) - E(f)}{L(f) - E(f)} = \frac{ead(f) - e(f)}{1 - e(f)} \quad [12]$$

Para estimar a EAD(f) de operações onde $L(f) \neq E(f)$ basta estimar os fatores de conversão LEQ(f) adequadamente baseado no RDS(f) e, em seguida, empregar a equação [11] para atribuir uma estimativa individual da EAD. Para algumas operações os bancos atribuem um valor de EAD a f utilizando a formula [13] através de um CCF(f) estimado:

$$EAD(f) = CCF(f) \cdot L(f) \quad [13]$$

Logo pode-se escrever:

$$LEQ(f) = \frac{CCF(f) \cdot L(f) - E(f)}{L(f) - E(f)} = \frac{CCF(f) - e(f)}{1 - e(f)} \quad [14]$$

E obter CCF(f) em função do LEQ(f):

$$CCF(f) = LEQ(f) \cdot (1 - e(f)) + e(f) \quad [15]$$

Ou seja, estimar a EAD(f) baseado em fatores LEQ(f) ou baseado em fatores CCF(f) é equivalente, exceto nos casos em que $e(f)=1$.

5.3 Modelagem dos fatores de conversão a partir do banco de dados

Existem basicamente 3 métodos de estimar os fatores de conversão a partir da seguinte equação básica:

$$EAD(f) - E(f) = LEQ(f) \cdot (L(f) - E(f)) \quad [16]$$

Esses métodos se diferenciam na maneira como medem o incremento de crédito entre a data t_r e t_d : como porcentagem do limite disponível (ênfase no fator LEQ observados), como porcentagem do limite observado (ênfase na utilização do limite) e, finalmente, em valor absoluto (ênfase na exposição).

Método 1 – Ênfase nos fatores LEQ observados

Dividindo a equação [16] por $L(f) - E(f)$ tem-se:

$$\frac{EAD(f) - E(f)}{L(f) - E(f)} = LEQ(f) \quad [17]$$

A lógica aqui é determinar uma função das variáveis explicativas $LEQ(RD)$ para explicar os fatores LEQ_i observados dados por $LEQ_i = (EAD_i - E_i)/(L_i - E_i)$ em função de $LEQ(RD_i)$. Isso pode ser feito fixando uma expressão para o erro associado à diferença $LEQ_i - LEQ(RD_i)$ e resolvendo um problema de minimização desse erro. Atualmente usa-se uma função quadrática e simétrica para o erro em quase todos os casos. Portanto o problema de minimização (P.1) que se deve resolver resume-se por:

$$Mín_{LEQ} \left\{ \sum_i (LEQ_i - LEQ(RD_i))^2 \right\} = Mín_{LEQ} \left\{ \sum_i \left(\frac{EAD_i - E_i}{L_i - E_i} - LEQ(RD_i) \right)^2 \right\} \quad [18]$$

Ou:

$$\frac{1}{LEQ(f)} = Mín_{LEQ} \left\{ \sum_i \frac{1}{(L_i - E_i)^2} \cdot (EAD_i - E_i - LEQ(RD_i) \cdot (L_i - E_i))^2 \right\} \quad [19]$$

A primeira vista este método parece ser o mais natural já que foca diretamente a estimação do fator de conversão.

Método 2 – Ênfase no incremento da exposição medido como porcentagem do limite

Dividindo a equação [16] por $L(f)$ temos:

$$\frac{EAD(f) - E(f)}{L(f)} = LEQ(f) \cdot \frac{L(f) - E(f)}{L(f)} \quad [20]$$

Nesse caso deseja-se explicar as quantidades observadas $(EAD_i - E_i)/L_i$ e os valores explicativos são $LEQ(RD_i) \cdot (L_i - E_i)/L_i$. Logo tem-se o seguinte problema de minimização (P.2):

$$Mín_{LEQ} \left\{ \sum_i \left(\frac{EAD_i - E_i}{L_i} - LEQ(RD_i) \cdot \frac{(L_i - E_i)}{L_i} \right)^2 \right\} \quad [21]$$

Ou:

$$\overline{LEQ(f)} = M\acute{m}_{LEQ} \left\{ \sum_i \frac{1}{L_i^2} * (EAD_i - E_i - LEQ(RD_i) \cdot (L_i - E_i))^2 \right\} \quad [22]$$

Método 3 – Ênfase no incremento da exposiç o

Utilizando diretamente a equaç o base deseja-se explicar $EAD_i - E_i$ atrav s da vari vel explicativa $LEQ(RD_i) \cdot (L_i - E_i)$. O problema de minimizaç o   (P.3):

$$\overline{LEQ(f)} = M\acute{m}_{LEQ} \left\{ \sum_i (EAD_i - E_i - LEQ(RD_i) \cdot (L_i - E_i))^2 \right\} \quad [23]$$

Atrav s das equa es [18], [21] e [23] deduz-se um problema de minimizaç o geral (P.4):

$$M\acute{m}_{LEQ} \left\{ \sum_i \left(\frac{EAD_i - E_i}{\omega_i} - LEQ(RD_i) \cdot \frac{(L_i - E_i)}{\omega_i} \right)^2 \right\} \quad [24]$$

Onde ω_i representa $L_i - E_i$ no M todo 1, L_i no M todo 2 e 1 no M todo 3. Denotaremos por F^* a funç o de distribuiç o emp rica de $(EAD - E)/\omega$ associada com as observa es inclu das no RDS(f). Podemos denotar o problema de minimizaç o geral (p.4) como:

$$\overline{LEQ}(RD) = M\acute{m}_{LEQ} \left\{ E_{F^*} \left\langle \left(\frac{EAD - E}{\omega} - LEQ(RD) \cdot \frac{(L - E)}{\omega} \right)^2 \right\rangle \right\} \quad [25]$$

Supondo que $(L - E)/\omega$   constante para as observa es em RDS(f), portanto, a soluç o da equa o anterior   dada por:

$$\overline{LEQ}(f) = E_{F^*} \left\langle \left(\frac{EAD - E}{\omega} \right) | RD(f) \right\rangle \cdot \frac{\omega(f)}{L(f) - E(f)} \quad [26]$$

Na pr tica, o problema se resume em encontrar m todos para se aproximar essas esperanç as condicionais. Se assumirmos uma forma param trica para o fator LEQ, o problema se converte em:

$$\overline{LEQ}(f) = LEQ(\hat{a}, \dots)$$

$$\{\hat{a}, \dots\} = M\acute{m}_{\{\hat{a}, \dots\}} \left\{ E_{F^*} \left\langle \left(\frac{EAD - E}{\omega} - LEQ(a, b, \dots) \cdot \frac{L - E}{\omega} \right)^2 \right\rangle \right\} \quad [27]$$

Assim, a forma funcional é linear nos parâmetros e o problema se reduz a uma regressão linear. Este método pode apresentar vantagens quando as estimativas devem ser utilizadas no cálculo do capital regulamentar devido a relação entre os requisitos de capital e a EAD.

Para resumir, pode-se afirmar que os métodos tradicionais podem ser classificados como modelos de regressão que focam na minimização do erro quadrático e nas predições de: fatores LEQ_i's, das EAD_i's com porcentagem de limite, ou das EAD_i's.

5.4 Casos para quando LEQ é constante para observações similares a {f, t}

Problema 1 – A média amostral

É comum que algumas instituições financeiras, de acordo com Moral (2006), utilizem a média amostral dos fatores LEQ observados (LEQ_i) como estimador de LEQ(f) em t, desde que sejam restringidas aquelas observações que são similares a {f, t, RD}. Assim, supondo que o fator de conversão é constante para observações similares a {f,t} $LEQ(f) = LEQ$, e resolvendo o problema 1, obtém-se:

$$\overline{LEQ} = \text{Mín}_{LEQ \in R} \left\{ \sum_i \left(\frac{EAD_i - E_i}{(L_i - E_i)} - LEQ \right)^2 \right\} = \frac{1}{n} \sum \frac{EAD_i - E_i}{(L_i - E_i)} = \frac{1}{n} \sum LEQ_i \quad [28]$$

Em alguns casos é utilizada uma média amostral ponderada que busca considerar uma possível relação entre o tamanho das exposições (ou dos limites) e os fatores LEQ. Se introduzirmos um peso w_i no problema 1 e supusermos que o fator LEQ é constante para as observações similares a {f,t}, então teríamos:

$$\overline{LEQ} = \text{Mín}_{LEQ \in R} \left\{ \sum_i w_i \left(\frac{EAD_i - E_i}{(L_i - E_i)} - LEQ \right)^2 \right\} = \frac{\sum w_i \cdot LEQ_i}{\sum w_i} \quad [29]$$

No entanto quando a justificativa para incluir esses pesos for a de levar em consideração uma variável explicativa para o LEQ, este tratamento se mostrará

incoerente, pois a média ponderada seria uma solução ótima somente se supuséssemos que os fatores LEQ são constantes.

Problema 2 – Regressão sem constante

Outro método muito utilizado pelos bancos é utilizar o estimador da inclinação de uma regressão linear baseada no modelo 2 (apresentado anteriormente) e assumir que o fator LEQ é constante. A partir destas condições a expressão para o estimador de regressão seria dada por:

$$\begin{aligned} \overline{LEQ} &= \text{Mín}_{LEQ \in R} \left\{ \sum_i \left(\frac{EAD_i - E_i}{L_i} - LEQ \cdot \left(\frac{L_i - E_i}{L_i} \right) \right)^2 \right\} \\ &= \frac{\frac{\sum (EAD_i - E_i)(L_i - E_i)}{L_i^2}}{\sum \left(\frac{L_i - E_i}{L_i} \right)^2} = \frac{\sum (ead_i - e_i)(1 - e_i)}{\sum (1 - e_i)^2} \end{aligned} \quad [30]$$

Problema 3 – Média Ponderada Amostral

Neste caso, supondo que LEQ = constante, pode-se expressá-lo como:

$$\overline{LEQ(f)} = \text{Mín}_{LEQ \in R} \left\{ \sum_i (L_i - E_i)^2 \frac{EAD_i - E_i}{(L_i - E_i)} - LEQ^2 \right\} \quad [31]$$

E o ótimo seria dado por:

$$\overline{LEQ} = \frac{\sum w_i \cdot LEQ_i}{\sum w_i}, \text{ com } w_i = (L_i - E_i)^2 \quad [32]$$

5.5 Método do Coeficiente de Utilização do Limite, com CCF Constante

Ainda de acordo com Moral (2006) este método seria mais comumente utilizado por bancos para tentar evitar o uso explícito de fatores LEQ observados (ou

estimados) negativos ou para os tipos de operações para as quais a utilização atual do limite não tem capacidade preditiva para a EAD. Para essas operações a (EAD(f)) seria estimada utilizando diretamente a equação [13] e uma simples estimativa do CCF como a média amostral dos CCF observados associados a um conjunto de operações C que tenham descumprido:

$$\overline{EAD}(f) = \overline{CCF}(C).L(f) \quad [33]$$

A partir da equação [14] e supondo que o CCF é uma constante seria possível obter uma forma funcional específica para o fator LEQ(e(f)) dado por:

$$\overline{LEQ}(f) = \frac{\overline{CCF}.L(f) - e(f)}{L(f) - E(f)} = \frac{\overline{CCF} - e(f)}{1 - e(f)} \quad [34]$$

Segundo a equação anterior, duas operações com o mesmo CCF estimado e com diferentes valores de sua utilização percentual atual, e(t) terão associados diferentes fatores LEQ.

O principal problema com o método da equação [33] é que a prática mostra que, em geral, o limite utilizado e o limite disponível têm forte capacidade explicativa para a EAD. Por esta razão, o método não satisfaria, para muitos dos tipos de operações que aparecem na prática, o requisito imposto pelo órgão controlador de que seja utilizada toda a informação relevante e disponível para a estimativa dos parâmetros.

6 Estudo de caso

Para tornar esse estudo mais interessante optou-se por testar uma nova metodologia, diferente das estudadas e sugeridas no capítulo anterior, para a estimação da EAD de um portfólio.

Essa metodologia será aplicada a uma base de dados real e utilizará basicamente das seguintes técnicas estatísticas:

- Árvore de decisão para geração de grupos homogêneos de risco
- Regressões quantílicas aliadas a um processo de simulação para gerar os valores estimados dos fatores LEQ e da EAD de cada grupo homogêneo.

6.1 Base de dados utilizada

As principais características da base de dados utilizada são:

- Fonte dos dados: Instituição financeira brasileira de grande porte.
- Número de observações: uma amostra de 20.000 contratos.
- Período de observação: 5 anos.
- Descumprimento: considera-se que o descumprimento se dá no momento em que se acumulam 90 dias de atraso.
- Estrutura: para cada operação foram calculados 12 fatores LEQ observados, um para cada horizonte ($td - tr$), totalizando $20.000 * 12$ registros. Foram mantidas as variáveis Valor do limite Total, Salário bruto do Cliente em tr , Valor da dívida em tr , Valor do pagamento efetuado no mês anterior a tr , *Behavior Score* em tr , Valor da exposição na data de referencia e o valor da exposição no momento do descumprimento (EAD).

A base de dados é referente a uma operação de crédito destinada a pessoas físicas. Tal operação é caracterizada por possuir um valor de limite global que não necessariamente será utilizado pelo cliente em sua totalidade.

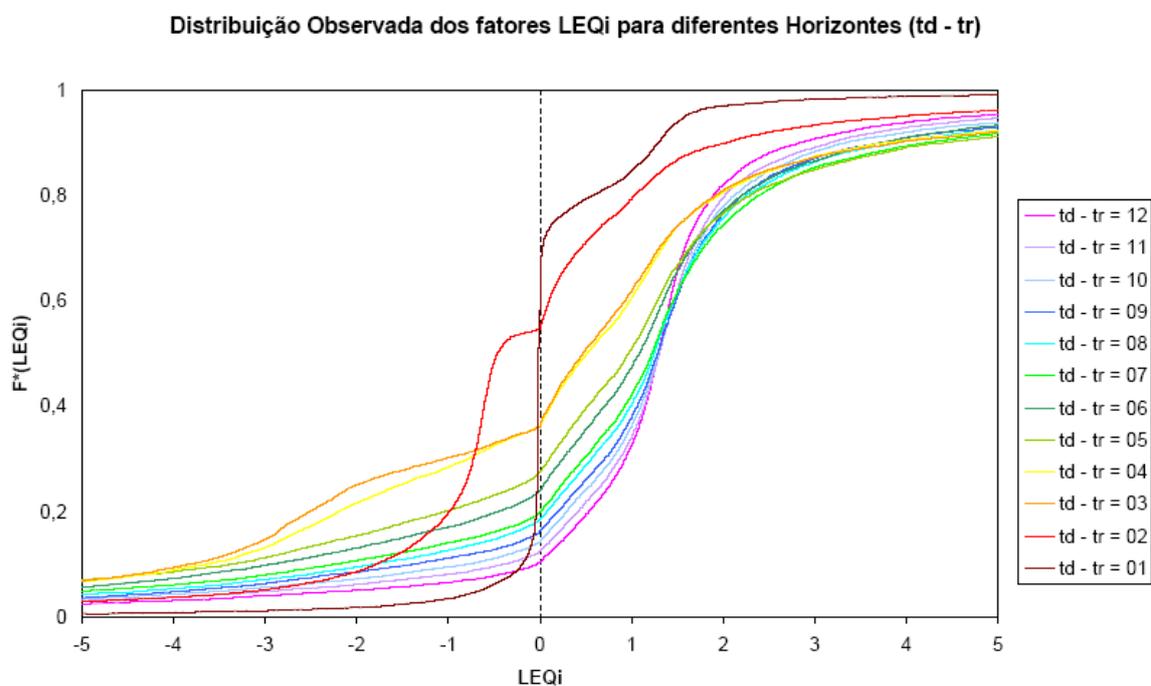
Ao utilizar parte do limite disponibilizado por essa linha de crédito o cliente deverá pagar no mês seguinte um montante não inferior ao valor dos juros incidentes sobre o valor utilizado. O atraso se dará caso o cliente não pague determinado valor mínimo desses juros no mês posterior à utilização do crédito.

6.2 Análise Exploratória

Antes de desenvolver uma técnica para criação dos grupos homogêneos e para o processo de simulação foi necessário realizar algumas análises exploratórias no banco de dados.

Inicialmente foram criados os gráficos com as distribuições empíricas dos valores LEQ observados associados com cada uma das 12 distâncias fixas entre as datas de referencia e as datas de descumprimento ($td - tr = 1, 2, \dots, 12$). O resultado é apresentado na figura 5:

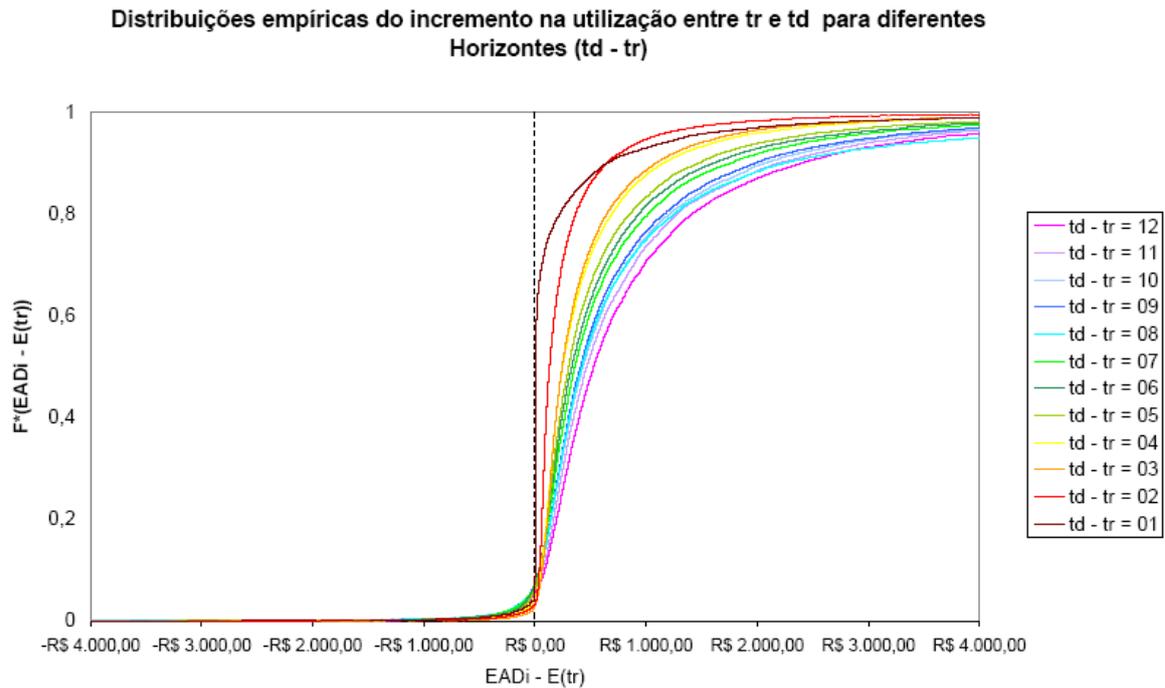
Figura 5 – Distribuição Observada dos fatores LEQ_i para diferentes Horizontes



Pode-se observar na figura 5 que a distribuição empírica dos valores do fator LEQ varia de acordo com o horizonte temporal selecionado.

Resolvemos fazer ainda uma segunda análise. Dessa vez observando a distribuição empírica do incremento na utilização ocorrido com cada uma das 12 distancias fixas entre as datas de referencia e as datas de descumprimento ($td - tr = 1, 2, \dots, 12$). O resultado é mostrado na figura 6.

Figura 6 – Distribuições empíricas do incremento na utilização entre TR e td para diferentes horizontes



Novamente percebe-se que a distribuição do valor absoluto de incremento realizado até a data de descumprimento varia de acordo com a data de referencia tr que é utilizada, como já era esperado.

Com base na análise de ambas as distribuições empíricas optamos por trabalhar daqui em diante apenas com os dados cuja data de referencia seja dada pelo horizonte de tempo de um ano, isto é $td - tr = 12$. Dessa forma teremos informações homogêneas para realizar nossa estimativa e, ao mesmo tempo, teremos um cálculo da EAD compatível e coerente com as estimativas de PD, que devem obedecer a um horizonte temporal de um ano.

Para avaliar se as variáveis do banco de dados podiam explicar a variável de estudo LEQ utilizamos o software R[®] 2.13.0 para traçar linhas de ajuste para os pontos

de cada gráfico de dispersão bidimensional gerado pelo cruzamento da variável LEQ com outras variáveis do banco. As figuras 7, 8, 9, 10 e 11 mostram os resultados obtidos para os principais gráficos de dispersão testados:

Figura 7 – Gráfico de dispersão do Salário Bruto do cliente X LEQ

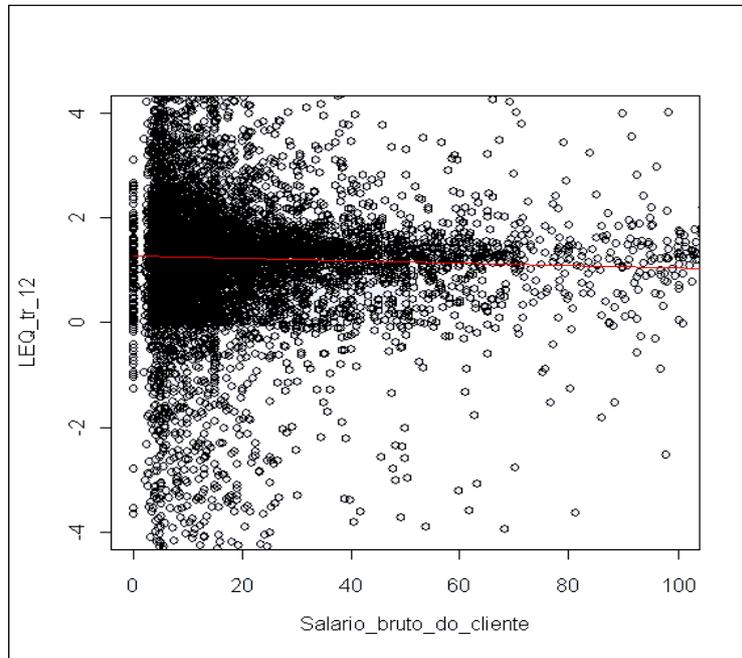


Figura 8 – Gráfico de dispersão do valor da dívida X LEQ

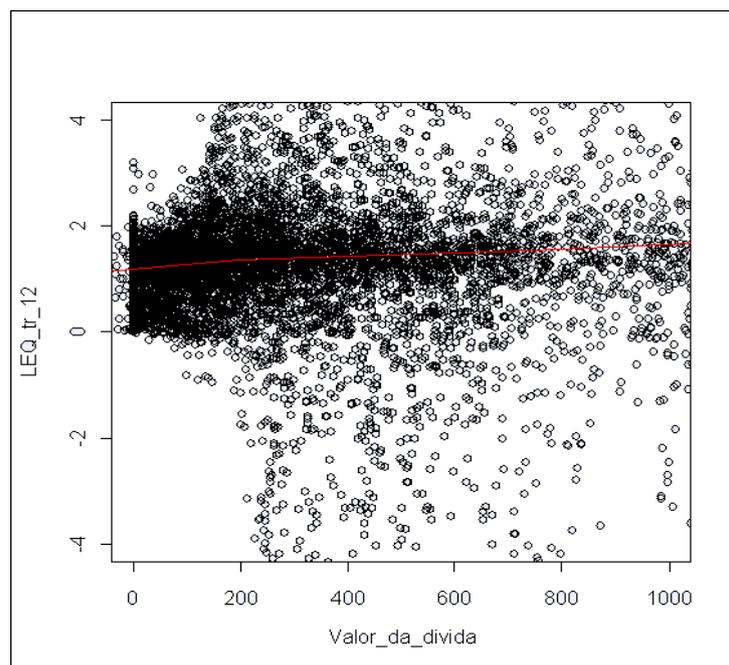


Figura 9 – Gráfico de dispersão do Limite total X LEQ

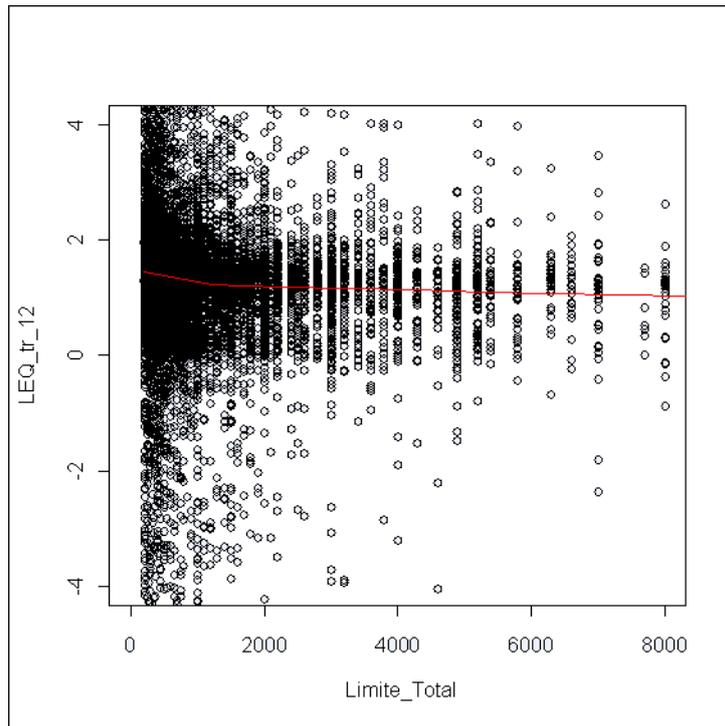


Figura 10 – Gráfico de dispersão do Pagamento efetuado no mês anterior X LEQ

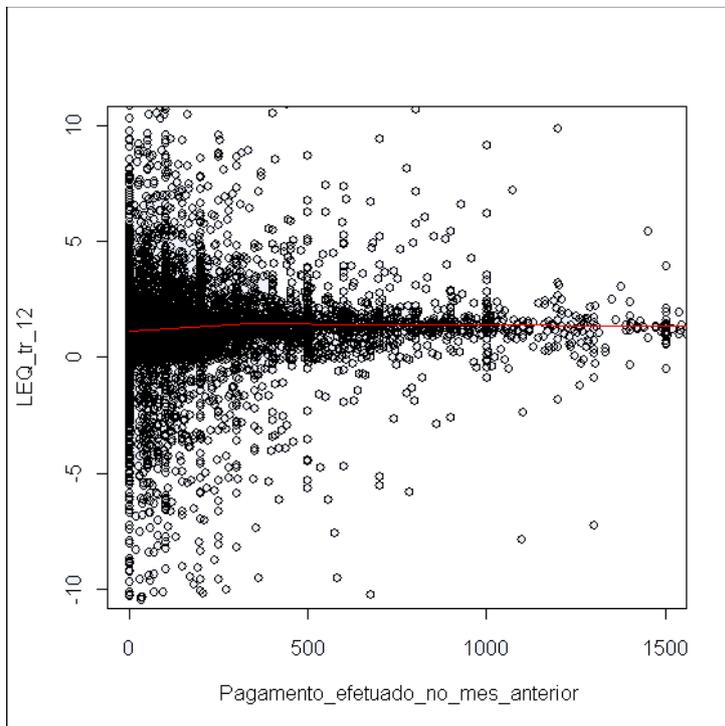


Figura 11 – Gráfico de dispersão da e(tr) percentual X LEQ

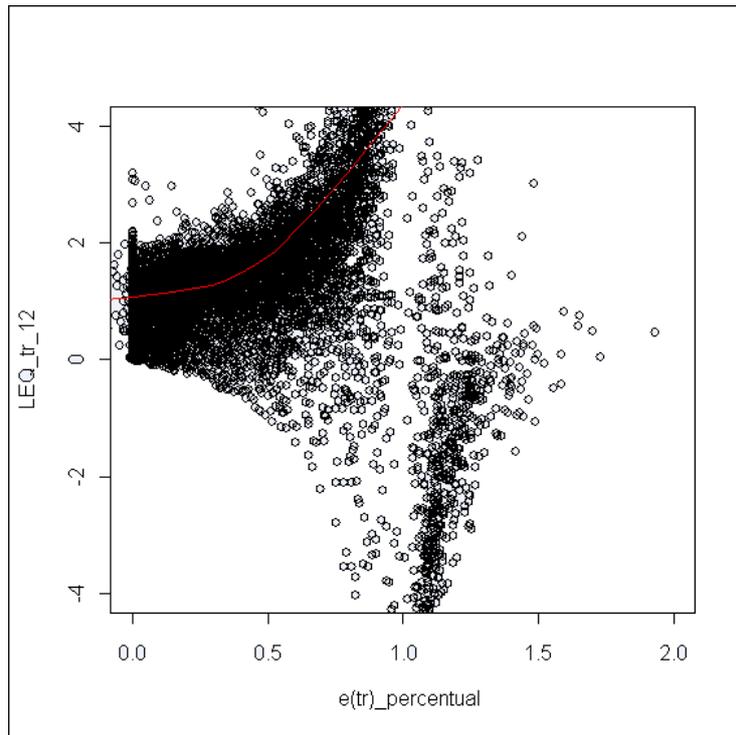
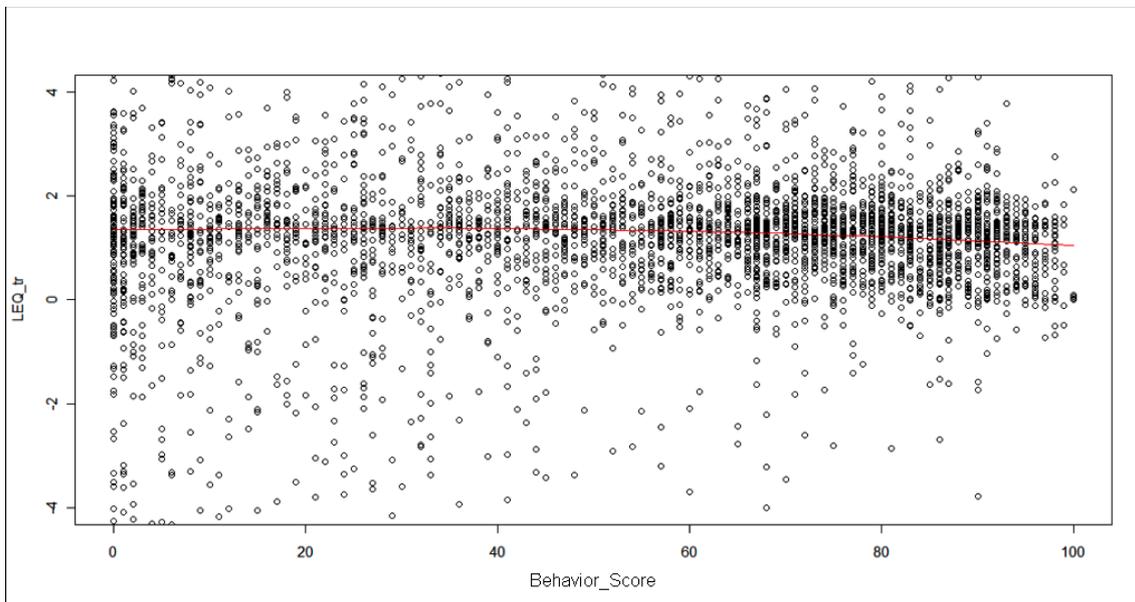


Figura 12 – Gráfico de dispersão do Behavior Score X LEQ



Observa-se por meio da análise dos gráficos gerados que a única variável com capacidade explicativa é o $e(tr)$ (figura 11), que representa o valor da utilização percentual, em relação ao limite total do cliente, observado um ano antes do descumprimento. O que se observa é que quanto maior o percentual utilizado pelo cliente um ano antes do descumprimento maior deverá ser o incremento percentual que esse cliente realizará até a data do descumprimento, ou seja, maior será o LEQ calculado para ele.

6.3 Conceito de Grupos homogêneos de risco

É importante esclarecer que o Banco Central do Brasil, seguindo à prática dos países que já adotaram os modelos avançados de Basiléia II, exige que a estimação dos parâmetros de risco (PD, LGD e EAD) seja feita para grupos homogêneos de risco criados pela instituição, e não para a carteira como um todo.

Mais especificamente o Banco Central do Brasil define em 2011 no Artigo 38 do Edital de Audiência Pública Nº 37 que “Para cada grupo homogêneo de risco devem ser estimados os valores dos parâmetros PD, LGD e EAD, admitida a ocorrência circunstancial de estimativas idênticas para diferentes grupos homogêneos”. No parágrafo único do Artigo 34 do mesmo documento um grupo homogêneo de risco é definido como: “define-se ‘grupo homogêneo de risco’ como o conjunto de exposições de varejo com características comuns para fins de avaliação e quantificação de risco de crédito”.

Dessa forma considera-se que a estimação da EAD não deve ser feita para cada portfólio como um todo. Ao contrário, cada portfólio deve ser dividido em grupos que apresentem uma diferenciação significativa de riscos entre si. Para esses grupos é que se deverá fazer a estimação da EAD.

Ressaltamos aqui que é muito importante que a criação desses grupos seja feita de maneira consistente e bem fundamentada. O processo de criação dos ditos grupos pode ser alvo de um estudo específico, mas o objetivo principal do estudo em questão

não é apresentar uma técnica ideal para criação desses grupos homogêneos, e sim apresentar uma técnica de estimação da EAD para cada um dos grupos já pré-existentes. Portanto, optou-se por utilizar uma metodologia estatística simples, mas que permitisse refletir os diferentes níveis de risco de cada grupo criado. Optamos por utilizar uma Árvore de Decisão estatística.

6.4 Utilizando a Arvore de decisão para criação dos grupos homogêneos

Dentre outras funções, uma das utilidades da Árvore de Decisão é justamente a criação de uma categorização de determinada variável com respeito a uma característica alvo.

No caso desse estudo optou-se por categorizar a variável *Behavior Score* tendo como característica alvo o percentual de descumprimentos observados. Para tanto utilizou-se o procedimento Decision Tree do SAS Enterprise Miner® 5.3 para realizar a criação dos grupos homogêneos. Dessa maneira o Decision Tree procurou basicamente, por meio de exaustivos processos de iterações, categorizar a variável Behavior Score levando em consideração o percentual de descumprimentos ocorrido dentro de cada categoria criada.

Como a base de dados a ser utilizada no estudo era limitada optou-se por restringir o procedimento Decision Tree para a criação de no máximo 5 grupos homogêneos. Após dada essa restrição o algoritmo do programa basicamente fará de forma iterativa todas as combinações de categorizações possíveis para a variável Behavior Score e calculará o teste Qui-quadrado para a comparação dos percentuais de descumprimento observados em cada uma das amostras criadas. Essa estatística é calculada comparando as amostras duas a duas e, por fim, é mantida aquela categorização da variável que proporciona um maior valor nos testes Qui-quadrado. Assim o procedimento entenderá que categorizou a variável *Behavior Score* de maneira a propiciar a máxima diferenciação dos percentuais de descumprimento dentro de cada categoria criada. O resultado obtido é mostrado no Apêndice I.

Observa-se no resultado que os Escores comportamentais intermediários não discriminam muito bem o risco de descumprimento uma vez que o intervalo de Behavior Score [75,5; 81,5] apresentou maior percentual de descumprimento que o intervalo [20,5; 75,5], o que não deveria ter ocorrido. Caso uma instituição financeira hipoteticamente desejasse utilizar essa técnica para a criação de seus grupos homogêneos de risco o problema da não discriminação de riscos dos Escores intermediários deveria ser estudado mais a fundo. Entretanto, ressaltando o que já foi citado anteriormente, como esse estudo não tem a pretensão de apresentar uma técnica ideal de criação dos grupos homogêneos, não nos prenderemos a esse problema e partiremos para a estimação da EAD a partir dos grupos criados pela categorização da variável *Behavior Score* como apresentado no Anexo I.

6.5 Metodologia de Simulação a ser utilizada

Os estudos de simulação visam obter aproximações de quantidades de interesse. De maneira bastante ampla, estes tentam reproduzir, em um ambiente controlado, o que se passa com um problema real. Uma simulação pode ser entendida como uma particular realização do modelo. Nesse sentido, os valores simulados muitas vezes podem ser entendidos como uma amostra.

Neste estudo o objetivo da simulação será estimar a distribuição da *Exposure at Default* (EAD). Para isso a base de dados original será dividida, por meio de amostragem aleatória simples, em duas bases de dados diferentes cada uma com 10.000 contratos. Denominaremos as bases como Base de Portfólio e Base de Estimação. A primeira simulará uma carteira com exposições que não entraram em descumprimento para as quais se deseja estimar o valor da EAD. Já a Base de Estimação será formada pelas observações que já entraram em inadimplência e para as quais os valores das variáveis de estudo já foram observados. A Base de Estimação será considerada o insumo para estimação da EAD na Base de Portfólio. Na verdade os valores das variáveis de estudo já terão sido observados para ambas as bases de dados a serem criadas, no entanto, para efeitos de estudo, partiremos do pressuposto de que

o valor dessas variáveis é desconhecido na Base de Portfólio. Os valores observados nessa base de dados só serão utilizados no final do procedimento para realizar uma comparação entre a EAD estimada pelo processo de simulação e a EAD observada

Após dividir a base de dados será realizado o processo de estimação de 11 quantis de FCC na base de estimação usando somente os contratos em descumprimento, pois não observamos um FCC para quem não descumpriu. Depois, na base de portfolio será feita uma simulação de descumprimentos considerando o valor da variável *Behavior Score* e, em seguida, para cada uma das operações marcadas como descumpridas, será atribuído um valor de FCC selecionado aleatoriamente entre um dos onze quantis para os quais foram gerados regressões.

Para melhor compreensão o processo será detalhado nas próximas seções.

6.5.1 Primeira etapa: Geração de regressões quantílicas na Base de Estimação

Para compreender a primeira etapa do processo de simulação utilizado faz-se necessário ter um pequeno entendimento do que é uma regressão quantílica.

A regressão quantílica, como qualquer regressão, nos permite explorar e inferir a relação de uma variável dependente (resposta) com variáveis independentes (explicativas). A análise da regressão pode ser usada como um método descritivo da análise de dados sem que sejam necessárias quaisquer suposições acerca dos processos que permitiram gerar os dados. Há regressões que produzem estimativas aproximadas da média condicional da variável dependente para determinados valores das variáveis preditoras. A regressão quantílica produz estimativas aproximadas da mediana ou de quaisquer outros quantis da variável dependente condicional no valor da variável preditora. Uma vantagem de usar a regressão quantílica para estimar a distribuição condicional é que o resultado da regressão quantílica vai ser mais robusto, em resposta aos outliers.

Estimar os coeficientes da função de regressão quantílica é simplesmente estabelecer a relação de regressão entre a variável dependente e as variáveis independentes. Minimiza-se os desvios entre os valores observados e os estimados para se obter os parâmetros da regressão. Assim, temos:

$$\text{Mín} \left(\sum_{i=1}^n \rho_{\tau}(y_i - x_i' b) \right) \quad [35]$$

Em que:

$$\rho_{\tau}(u) = u(\tau - I(u < 0)), \tau \in (0,1) \quad [36]$$

É a função perda quantil.

No estudo em questão, optou-se pela geração de 11 regressões quantílicas na Base de Estimação para cada um dos 5 grupos de risco regredindo para os quantis 0,01; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 0,99. Tais quantis foram selecionados de modo que fosse possível prever a distribuição da variável de estudo dentro de cada um dos grupos.

Como observamos que a única variável com capacidade explicativa seria o valor percentual da exposição na data de referência $E(tr)$ utilizamos somente esta variável como preditora para variável resposta, o fator de conversão LEQ.

6.5.2 Segunda etapa: Geração dos descumprimentos por meio do *Behavior Score*

Essa etapa serve basicamente para estimar quantos contratos dentro de um portfólio entrarão em descumprimento ao longo do ano seguinte. Para isso foi sorteado aleatoriamente um número entre 0 e 100 para cada exposição da Base Portfólio. Para cada numero gerado se aplicou a seguinte regra:

- se o número selecionado foi maior ou igual ao *Behavior Score* do contrato, assumiu-se que este entrará em inadimplência.

- se o número selecionado foi menor que o *Behavior Score* do contrato, assumiu-se que este não entrará em inadimplência.

Isso porque já se sabe de antemão que o *Behavior Score* da instituição é gerado por meio de uma regressão logística e seu resultado representa justamente a probabilidade do contrato entrar em descumprimento ao longo do ano seguinte. Portanto acredita-se que para a amostra de 10.000 contratos da Base de Portfólio a quantidade de descumprimentos gerada pelo processo aleatório para cada faixa de Score será aproximadamente a quantidade que ocorreria naturalmente em uma carteira com mesma distribuição de Scores.

6.5.3 Terceira etapa: Associação de fator LEQ por meio de seleção aleatória de um quantil

Na Base de Portfólio, para cada observação marcada como descumprida pelo passo anterior, foi selecionado aleatoriamente um dos seguintes quantis: 0,01; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 0,99.

Esse quantil foi utilizado para gerar um valor de LEQ para a respectiva operação utilizando-se dos parâmetros calculados pela regressão quantílica gerada a partir das observações da Base de Estimação.

Como estamos trabalhando com uma amostra considerável de 10.000 contratos e a distância entre os 11 quantis selecionados é praticamente constante, aumentando basicamente em 0,1 de uma para o outro, acredita-se que no final do procedimento teremos uma distribuição dos fatores LEQ gerados bem próxima do que seria na realidade.

6.5.4 Quarta etapa: Cálculo da distribuição da EAD esperada para o Portfólio

Após associar um valor de LEQ para cada operação marcada por descumprimento é calculado o valor da EAD de cada exposição do banco Portfólio aplicando a fórmula [2].

Finalmente, somando-se as EADs calculadas para cada operação, tem-se o primeiro valor da EAD gerado para cada um dos 5 grupos homogêneos de riscos.

O procedimento da simulação foi repetido por 500 vezes gerando assim as 5 distribuições do valor da EAD esperada.

Em seguida foi realizado todo o procedimento sem desagregar as operações em grupos homogêneos de risco, e foi gerada uma 6ª distribuição de valor da EAD esperada para o portfólio como um todo. Os resultados obtidos são apresentados na próxima seção.

6.6 Resultados

Inicialmente a simulação acima foi aplicada para cada um dos cinco grupos homogêneos de risco gerando as distribuições mostradas nas figuras 12, 13, 14, 15, 16 e 17. O traço vertical presente em cada uma das figuras representa o valor observado da EAD na Base Portfólio.

Figura 12 – Densidade da variável EAD estimada para o grupo homogêneo 1

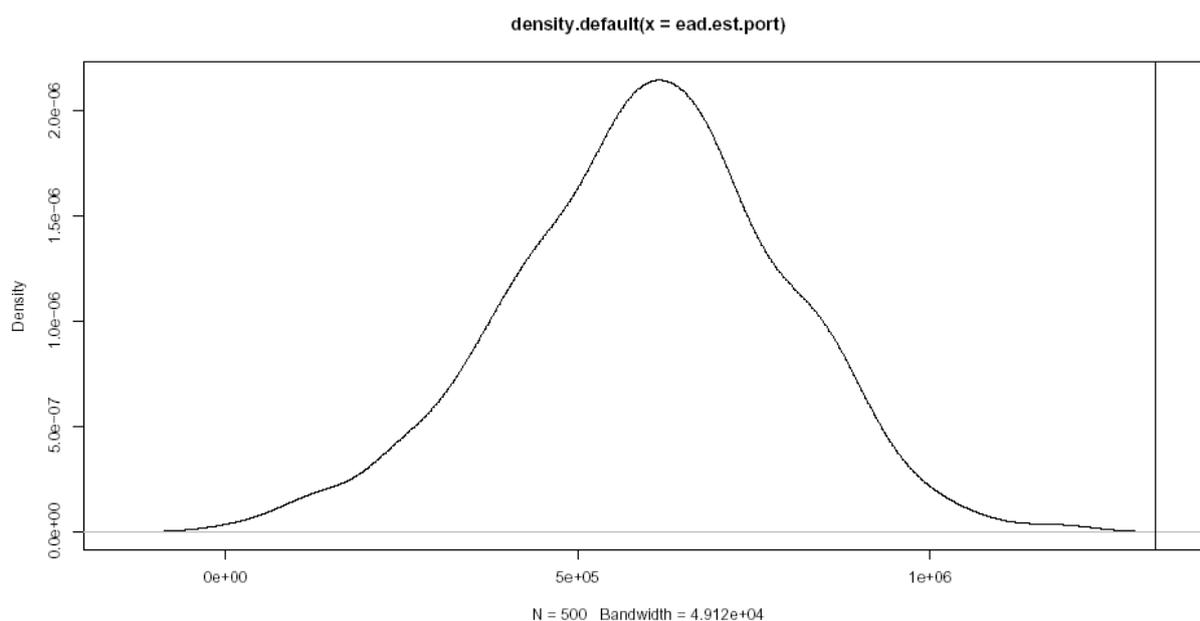


Figura 13 - Densidade da variável EAD estimada para o grupo homogêneo 2

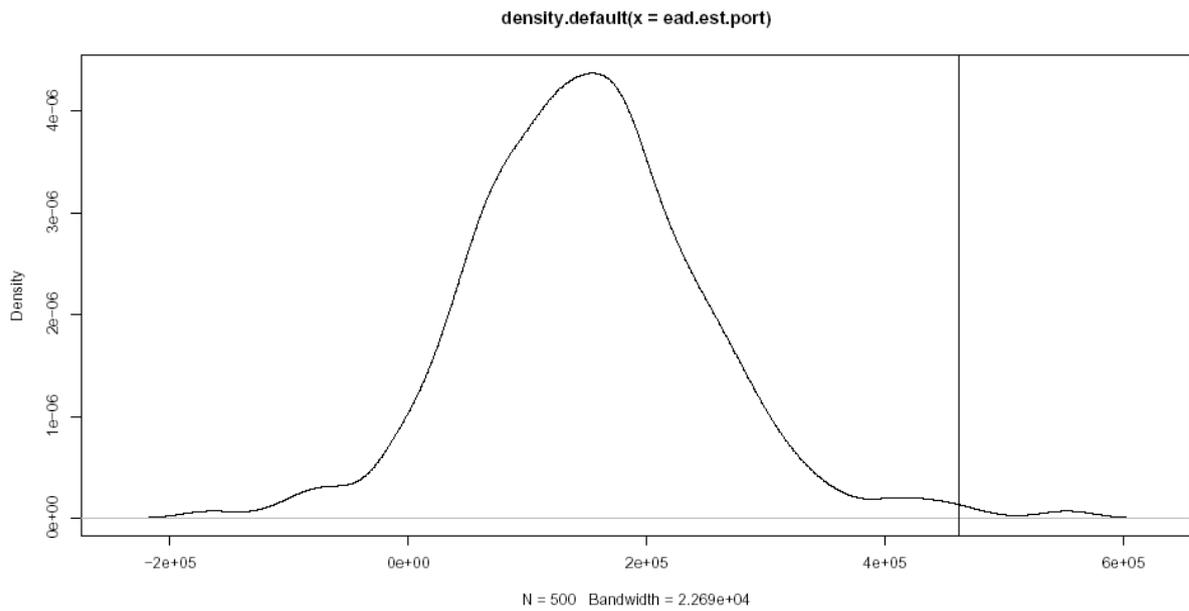


Figura 14 – Densidade da variável EAD estimada para o grupo homogêneo 3

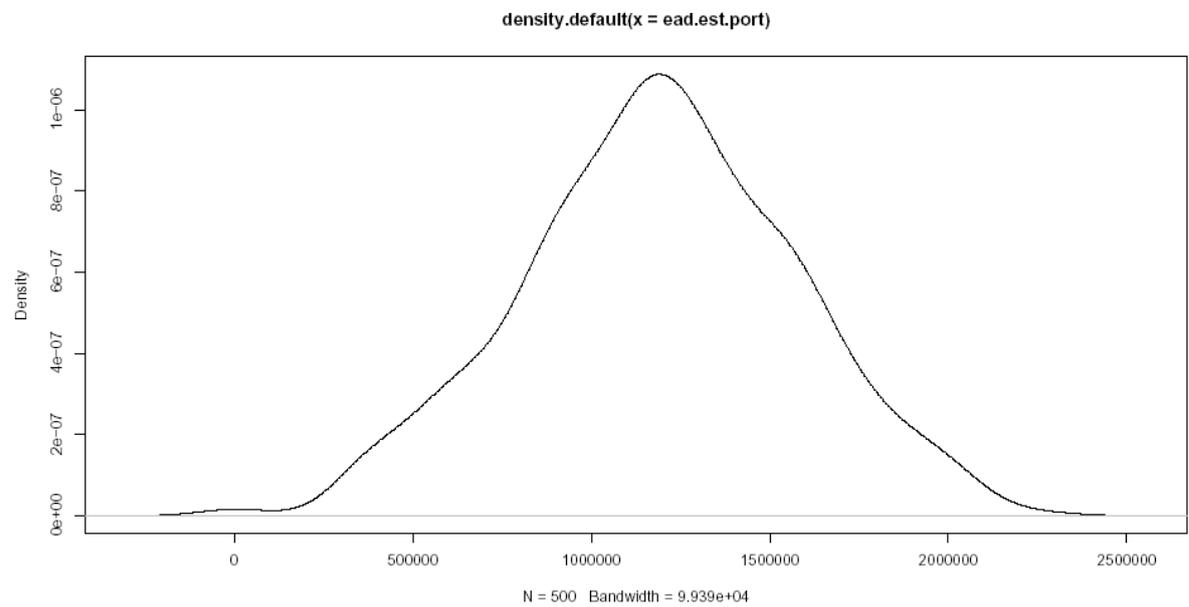


Figura 15 – Densidade da variável EAD estimada para o grupo homogêneo 4

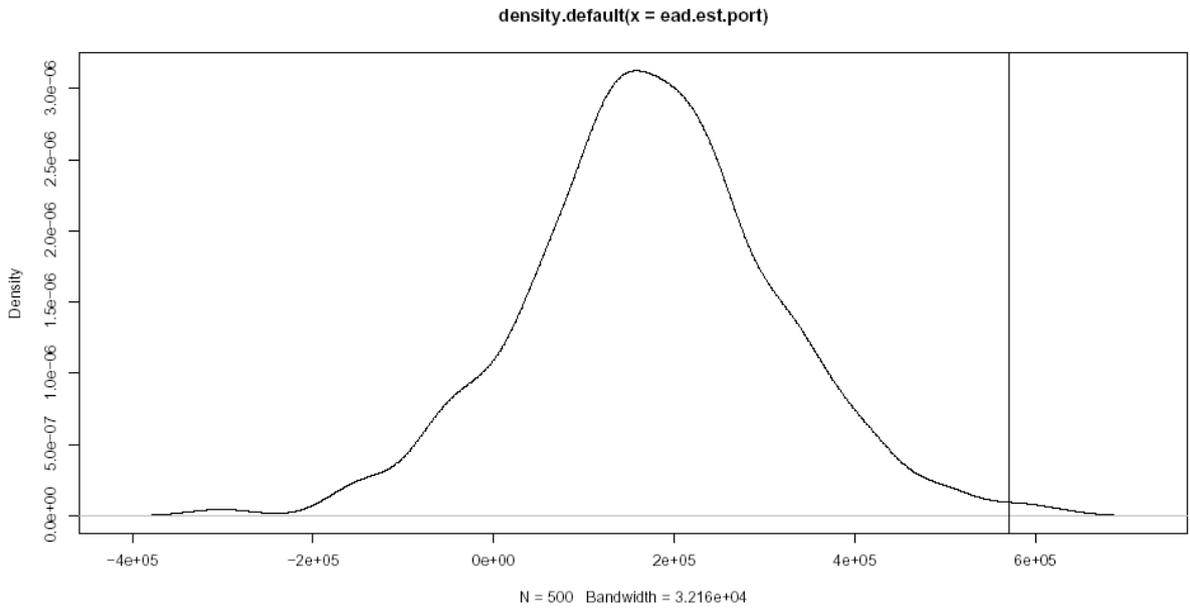


Figura 16 – Densidade da variável EAD estimada para o grupo homogêneo 5

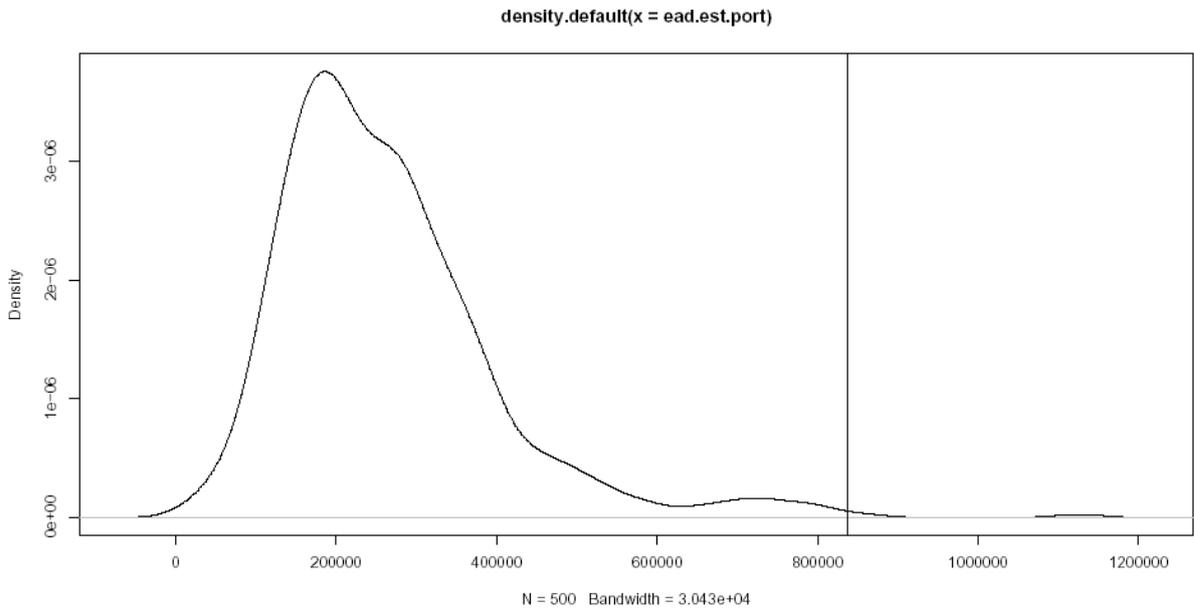
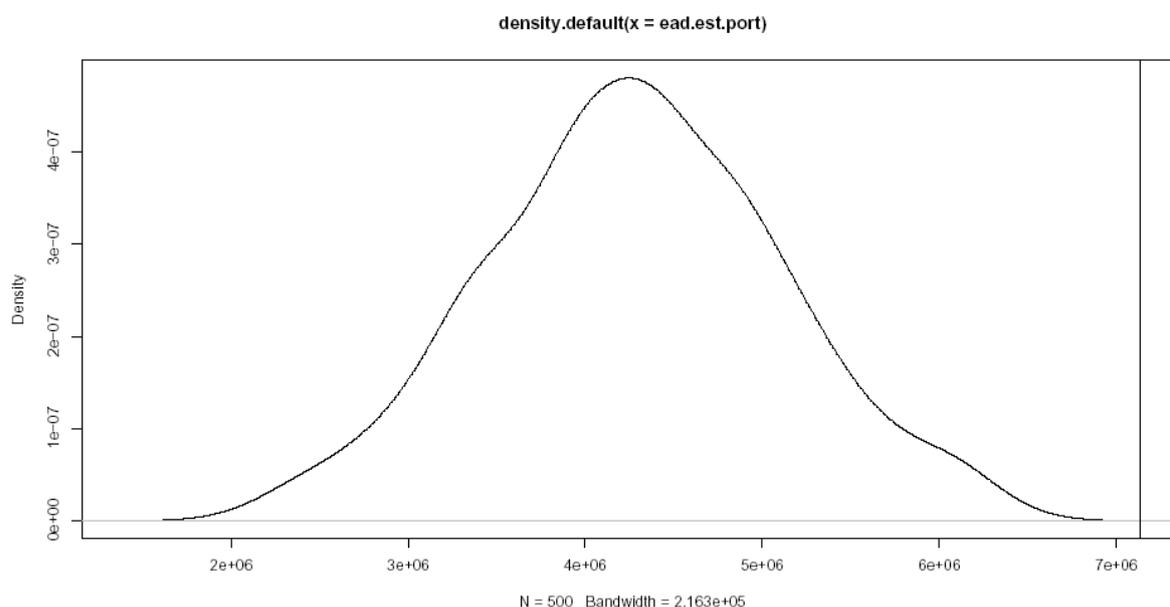


Figura 17 – Densidade da variável EAD da Base Portfólio sem segregação em grupos de risco.



Observa-se em todas as figuras apresentadas acima que o valor observado não está bem ajustado à densidade gerada pelo processo de simulação. A tabela 1 mostra o valor da EAD observada na Base Portfólio bem como o valor dos quantis 0,95 e 0,99 de cada distribuição gerada para cada grupo de risco.

Tabela 1 – Resultados por grupo de risco sem pesos

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Portfólio Total
Ead Observada	1.319.622,00	462.374,50	2.758.926,00	571.083,80	836.838,40	7.136.831,00
Var (95%)	917.710,84	317.667,76	1.848.402,40	396.434,48	496.521,59	5.706.912,00
Var (99%)	1.077.971,43	407.913,35	2.109.626,40	496.794,44	661.398,17	6.169.362,00

Observa-se que o valor da EAD observada para cada grupo é sempre maior que o percentil 0,99 da densidade gerada para o mesmo. Esse problema de subestimação do valor da EAD ocorreu também quando avaliamos o portfólio como um todo (figura 17).

Entretanto tal resultado não é totalmente inesperado. Acontece que a base de dados foi construída com 10.000 contratos que entraram em descumprimento e 10.000 contratos que não entraram em descumprimento. Essa metodologia de estimação a partir de um banco de dados dividido em duas partes iguais, uma de contratos descumpridos e outra de contratos que não descumpriram, foi adotada porque é muito utilizada na construção de modelos de escoragem. Entretanto, ao estimar os descumprimentos baseado no valor da variável *Behavior Score* a quantidade de descumprimentos será bem inferior à quantidade realmente observada na base de dados. Como a base original tem 50% de contratos descumpridos é totalmente esperado que ela apresente um maior valor para a EAD observada. No entanto nunca se espera que ocorra inadimplência em metade dos contratos de uma carteira normal.

Para resolver esse problema de diferença do número de descumprimentos gerados e o número real optamos por ponderar os valores de cada uma das 500 EAD's geradas por um peso. Desse modo, para cada uma das 500 iterações realizadas foi calculada uma razão entre a quantidade de descumprimentos observados na carteira e a quantidade de descumprimentos gerada pela metodologia descrita na seção 6.4.2. Esse valor, no final de cada repetição do procedimento, foi multiplicado pelo valor da EAD gerando então uma nova densidade para cada um dos grupos. Os resultados obtidos seguem nas figuras 18, 19, 20, 21, 22 e 23.

Assim temos:

$$P_{i,j} = \frac{NO_j}{NG_{i,j}} \quad [37]$$

Onde:

NO_j : Número de descumprimentos observados no grupo j

$NG_{i,j}$: Número de descumprimentos gerados na iteração i para o grupo j

Esse valor, no final de cada repetição do procedimento, foi multiplicado pelo valor da EAD gerando então uma nova densidade para cada um dos grupos dada por:

$$EAD_{ponderada_{i,j}} = EAD_{i,j} \cdot P_{i,j} \quad [38]$$

Os resultados das densidades ponderadas obtidas seguem nas figuras 18, 19, 20, 21, 22 e 23.

Figura 18 – Densidade ponderada da variável EAD estimada para o grupo homogêneo 1

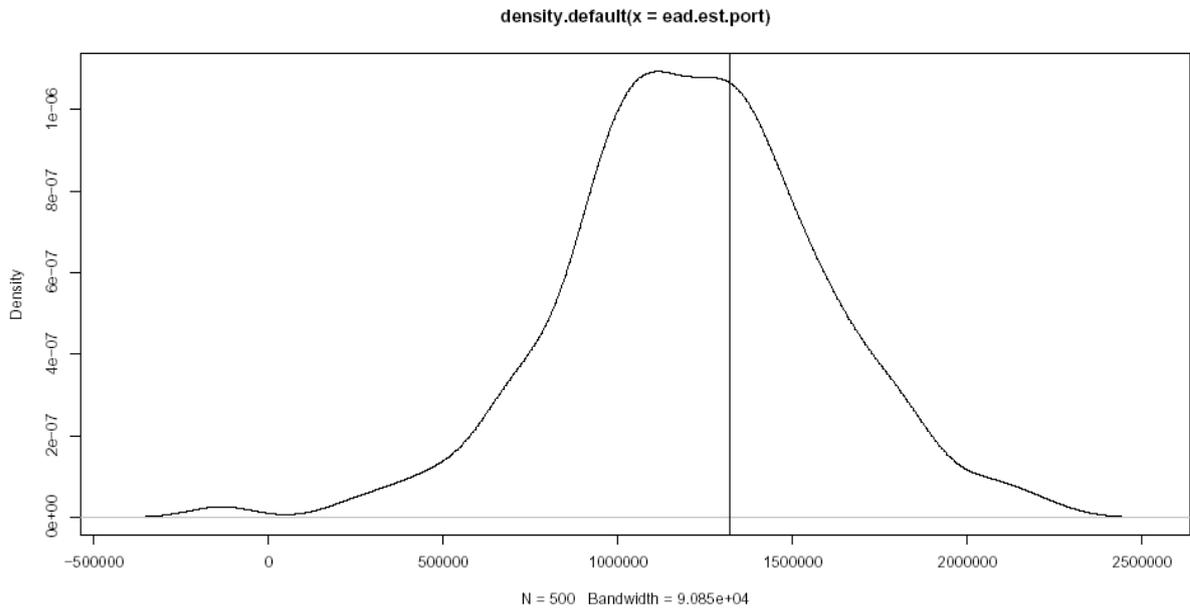


Figura 19 – Densidade ponderada da variável EAD estimada para o grupo homogêneo 2

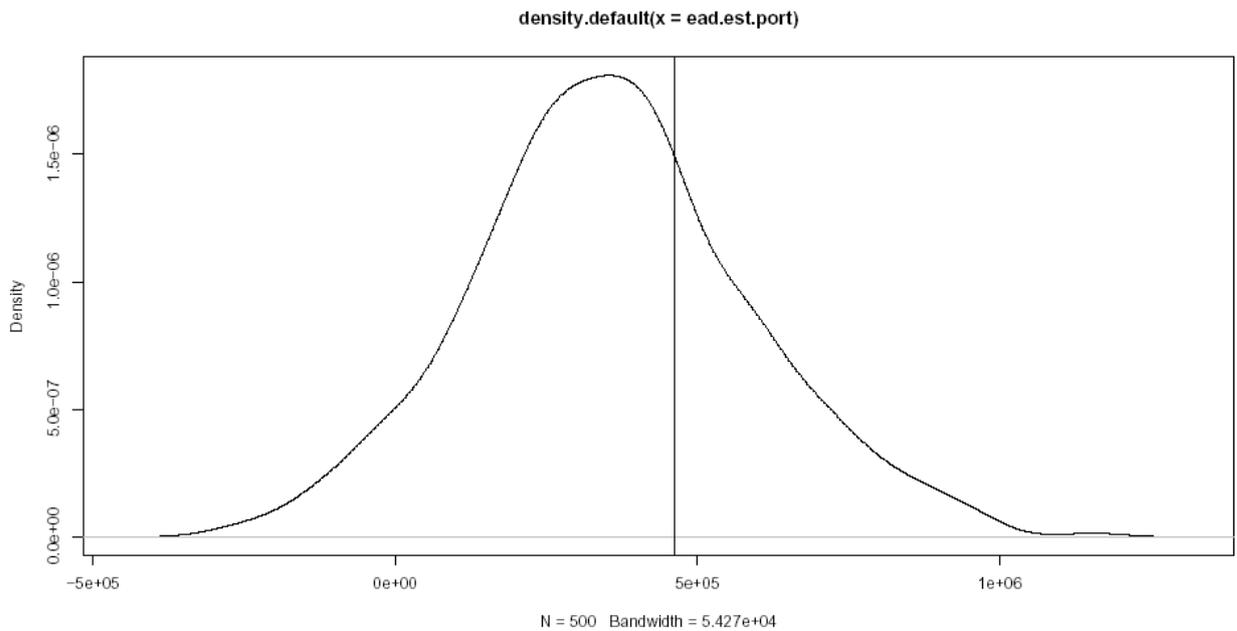


Figura 20 - Densidade ponderada da variável EAD estimada para o grupo homogêneo 3

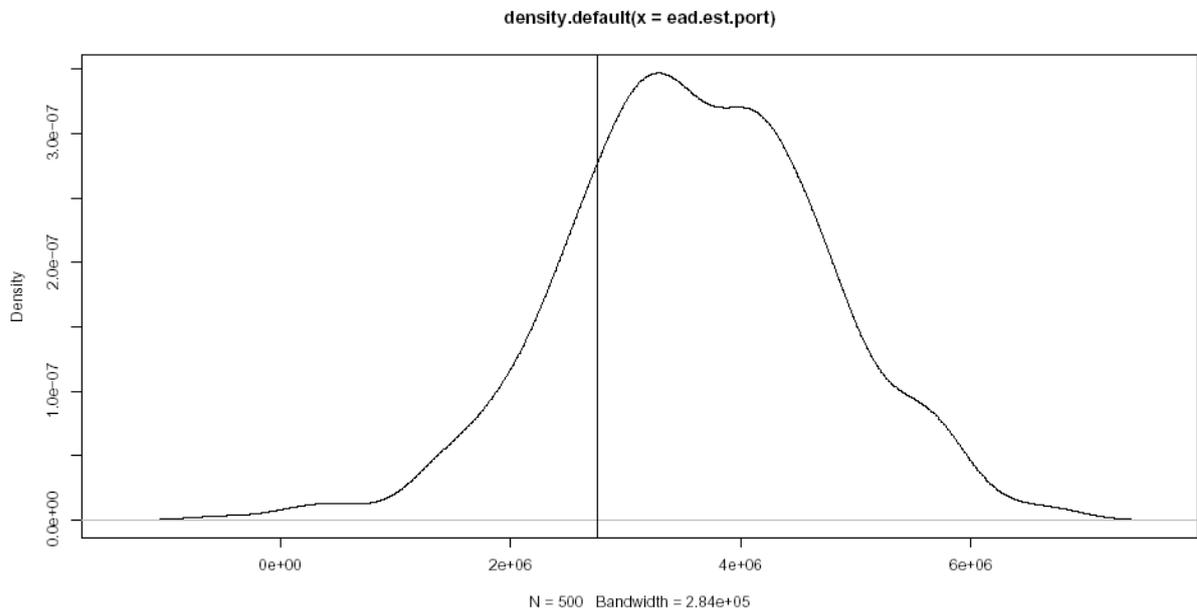


Figura 21 - Densidade ponderada da variável EAD estimada para o grupo homogêneo 4

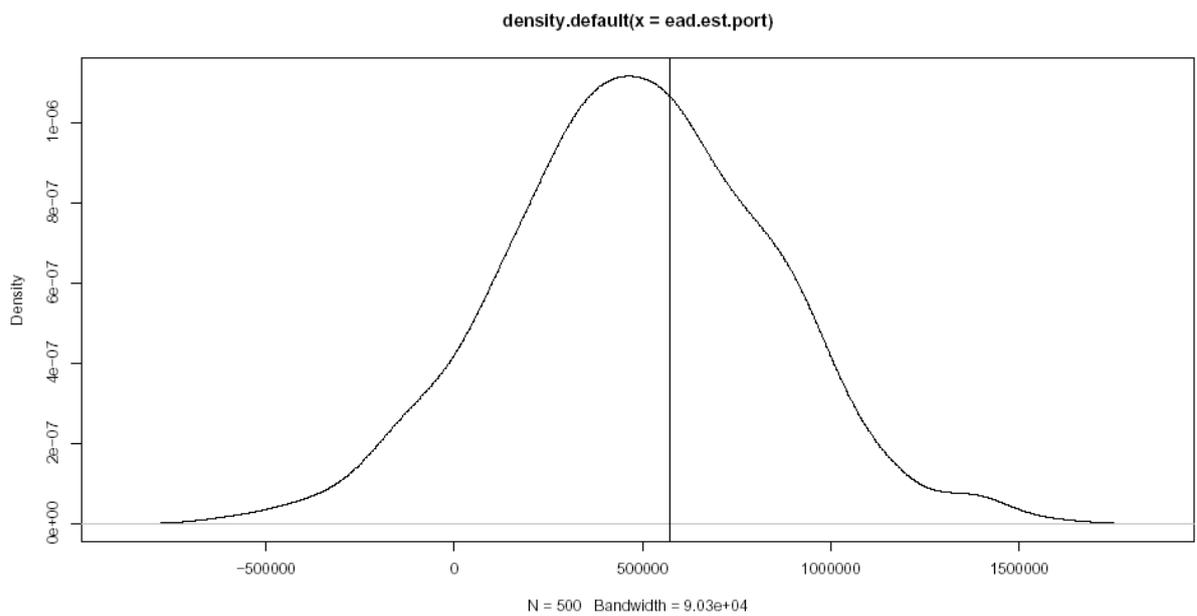


Figura 22 - Densidade ponderada da variável EAD estimada para o grupo homogêneo 5

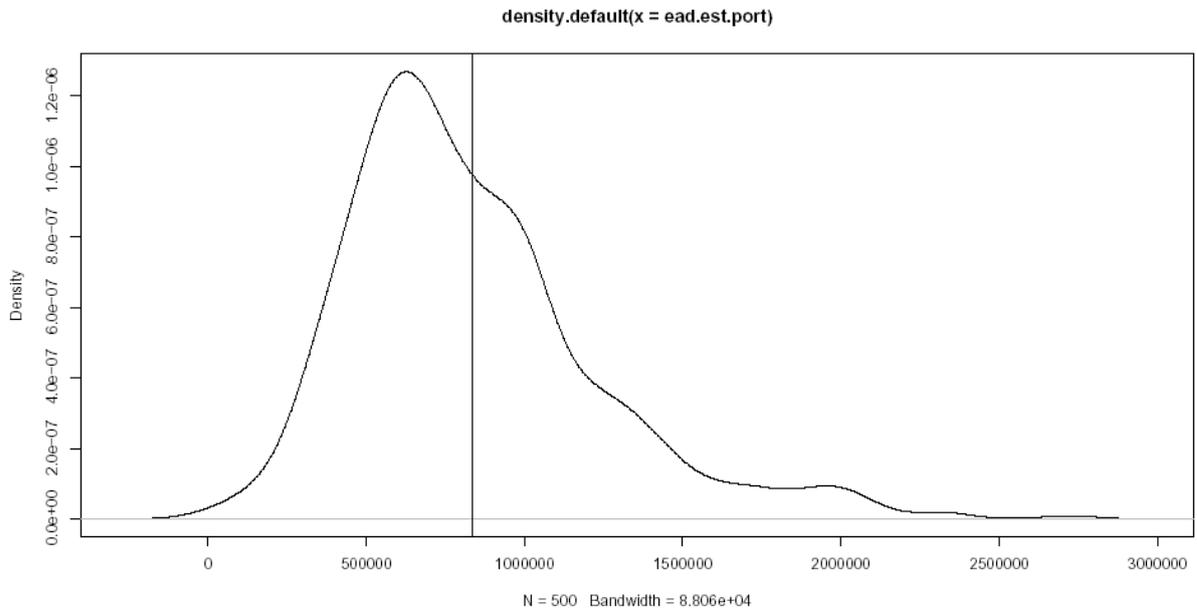
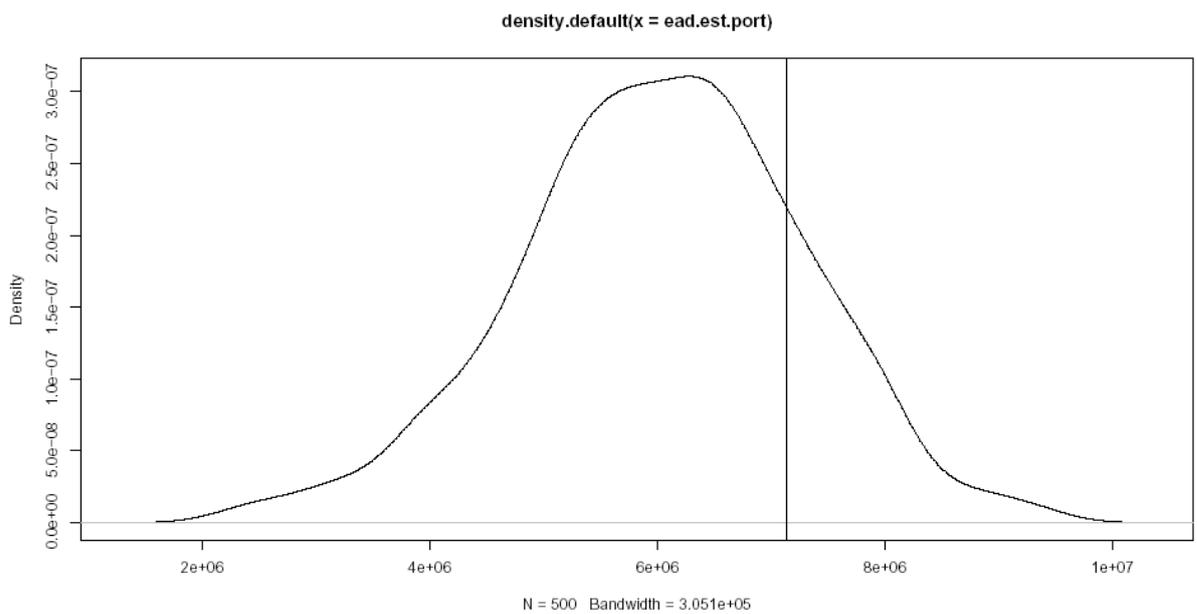


Figura 23 – Densidade ponderada da variável EAD da Base Portfólio sem segregação em grupos de risco



Pode-se notar agora que os valores da EAD observada na Base Portfólio estão aderentes com os valores das densidades geradas pela simulação. Pode-se atribuir esse melhor ajuste à ponderação criada entre número de descumprimentos observados e o número de descumprimentos gerados.

A Tabela 2 mostra o valor da EAD observada na Base Portfólio bem como o valor dos quantis 0,95 e 0,99 de cada distribuição gerada para cada grupo de risco utilizando-se da ponderação criada.

Tabela 2 – Resultados por grupo de risco com pesos

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Portfólio Total
EAD observada	1.319.622,00	462.374,50	2.758.926,00	571.083,80	836.838,40	7.136.831,00
Var (95%)	1.829.085,40	764.725,26	5.546.541,50	1.048.480,38	1.614.748,54	8.008.127,00
Var (99%)	2.100.029,90	929.518,19	6.044.867,50	1.373.821,04	2.042.965,28	8.893.338,00

Ao contrário do que foi observado na Tabela 1, após a ponderação do valor da EAD pelos pesos criados caso observa-se que o valor da EAD observada para cada grupo é sempre menor que o percentil 0,95 da densidade gerada para a mesma. Nessa nova situação o modelo parece ser adequado. Diante de uma situação como essa uma Instituição Financeira poderia defender a utilização de algum percentil elevado, por exemplo, o percentil 0,95, como uma estimativa conservadora para o valor da EAD.

7 Conclusão

Analisando visualmente os resultados obtidos observa-se que inicialmente, quando os valores da EAD não são ponderados pela quantidade de descumprimentos gerados pelo processo de simulação, a distribuição gerada não se ajusta bem ao valor observado. Isso aparentemente ocorreu graças ao alto percentual de descumprimentos observados para a Base Portfólio. Tal percentual elevado é fruto da metodologia empregada na construção da base de dados e não reflete os verdadeiros percentuais de descumprimento observados normalmente em carteiras de crédito.

Esse problema foi facilmente resolvido criando uma variável para ponderar a quantidade de descumprimentos observados na Base de Portfólio pela quantidade de descumprimentos gerados pelo processo de simulação. Esse parâmetro de ponderação poderia ser facilmente estimado pela Instituição Financeira através da observação da média histórica de descumprimentos da carteira, por exemplo.

Alternativamente a isso seria possível reconstruir uma nova base de dados, dessa vez a partir de uma amostragem aleatória simples de toda a carteira estudada, e processar a simulação desenvolvida. Isso provavelmente permitiria que a simulação fosse utilizada sem fazer jus ao parâmetro de ponderação utilizado. Se nesse caso a distribuição gerada pelo modelo se ajustasse bem aos valores observados, então faria sentido utilizar a técnica de simulação desenvolvida nesse estudo para a estimação do valor da EAD de um portfólio de contratos que não entraram em descumprimento.

Referências Bibliográficas

- [1] MORAL, G. EAD estimates for facilities with explicit limits. In: _____. *The Basel II parameters: estimation, validation, and stress testing*. Frankfurt: Springer, 2006. p. 197-242.
- [2] Sicsú, A. L. (2010). *Credit Scoring Desenvolvimento Implantação Acompanhamento*. Blucher.
- [3] BANCO CENTRAL DO BRASIL. *Comunicado 18.365. Orientações preliminares para a implantação de modelos internos de risco de crédito – Basileia II*. Brasília, abril de 2009.
- [4] BASEL COMMITTEE ON BANKING SUPERVISION. “International convergence of capital measurement and capital standards”. Disponível em: <www.bis.org>, 2006.
- [5] BANCO CENTRAL DO BRASIL. “Edital de Audiência pública nº 37 - Minuta de Circular dispondo sobre utilização de sistemas internos de risco de crédito segundo abordagem IRB – Basileia II”. Disponível em: <https://www3.bcb.gov.br/audpub/edital/lista_editais.jsp?edt=Publicados>, 2011.
- [6] KOENKER, R. *Quantile Regression*. 1 ed. Illinois: Cambridge University Press, 2005. 349 p.
- [7] VALVONIS, V. *Estimating EAD for retail exposures for Basel II purposes*. *Journal of Credit Risk* (79–109), Lithuania, p.79–109, Volume 4/ Number 1, Spring 2008.
- [8] R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

[9] SAS Institute Inc. 2011. SAS[®] 9.2 Intelligence Platform: System Administration Guide, Second Edition. Cary, NC, USA: SAS Institute Inc.

Apêndice

