IBMEC – SP

DISSERTAÇÃO DE MESTRADO

Aplicação de Metodologia de Dados em Painel em Modelos de Behaviour Score do Varejo

Edson Roberto da Silva

Dissertação apresentada ao Ibmec-SP para obtenção do título de Mestre em Finanças e Macroeconomia Aplicada.

Orientador: Prof. Dr. Naércio Aquino Menezes Filho

São Paulo

Dezembro/2006

Dedico a meus pais, Saulo e Nazareth, que me transmitiram as principais características de um vencedor: dedicação, bom caráter e coragem.

AGRADECIMENTOS

Agradeço muito a outras pessoas que de alguma forma fizeram esforços que somados tornaram possível a existência deste trabalho.

Primeiramente, à Associação Comercial de São Paulo, que através de seus Superintendentes Institucional e de Serviços, respectivamente Dr. Marcel Solimeo e Roberto Haidar me apoiaram no desenvolvimento desta dissertação. Nesta mesma empresa, também tive alguns apoios importantes tais como Elisângela Santos, que foi fundamental na geração da base de dados, e dos estatísticos Gina Mendonça e Sérgio Torres, que me apoiaram na compreensão dos resultados obtidos.

Também agradeço à Roseli Garcia, que foi a primeira pessoa a me incentivar para o desenvolvimento do tema escolhido, e que também me mostrou como é possível ser uma excelente profissional sem deixar de ser humana.

Este trabalho foi produzido no Ibmec-SP, e portanto também pertence à esta instituição, que tem professores da mais alta competência e aos quais tenho especial admiração. Entre eles destaco Andréa Minardi, Eduardo Carvalho, Eurilton Araújo, Ricardo Brito e Rinaldo Artes. Além destes ainda agradeço a Professora Regina Madalozzo e os monitores Márcio Laurini e Fábio Miessi, todos do IBMEC-SP, que me ajudaram a compreender e a utilizar o software utilizado neste trabalho.

Obviamente também agradeço ao meu orientador, Prof. Naércio Aquino Menezes Filho e ao Prof. Pedro Valls, que visualizaram antes de mim os passos que percorri arduamente.

Também devo muitos agradecimentos aos meus pais Saulo e Nazareth, e a meu irmão Edvaldo, que souberam compreender e me ajudar nos momentos especiais que um trabalho deste tipo demanda.

Por último, mas não menos importante, agradeço muito à minha esposa Josiane, que não apenas comemorou comigo cada etapa vencida, mas também me motivou nos momentos mais difíceis, sacrificando seus desejos em favor deste trabalho.

SUMÁRIO

LISTA DE TABELAS

LISTA DE GRÁFICOS

LISTA DE FIGURAS

LISTA DE ESQUEMAS

RESUMO

Esta dissertação aplica empiricamente a metodologia econométrica logito em painel com efeitos fixos em um modelo de ***behaviour score*** de varejo. Modelos desta natureza objetivam prever a inadimplência do consumidor em uma certa data através da utilização de informações contemporâneas ou passadas. A conclusão do trabalho é que a metodologia proposta não traz ganhos em relação à regressão logito modelada em ***cross-section***, que é a técnica usualmente utilizada no mercado financeiro.

Palavras-chave: crédito ao consumidor, empréstimos bancários, classificação.

ABSTRACT

This paper applies empirically the econometric methodology of fixed effects logit model in a retail behaviour score model. This kind of models target to preview consumer default using past or contemporaneous information. The paper conclusion is that the proposed methodology does not bring increase in performance compared to traditional cross-section logit regression, that is the most common tecnique used at financial market.

Keywords: consumer credit, consumer loans, credit scoring, classification.

1. INTRODUÇÃO

Este estudo tem o principal objetivo de apresentar e testar a utilização da metodologia logit com efeitos fixos no desenvolvimento de um modelo de previsão da inadimplência de consumidores, usualmente chamado de ***Behaviour Score***. Para isto será utilizada uma amostra com dados em painel e será avaliado o ganho comparativo de desempenho deste modelo com o seu similar gerado em uma amostra do tipo ***cross-section***.

É importante ressaltar que este trabalho não busca encontrar as causas da inadimplência, mas sim propor uma tecnologia que gere modelos de previsão mais eficazes. Muitos estudos têm sido realizados sobre o mercado de crédito ao consumidor brasileiro nos últimos anos, pois este segmento apresenta grandes índices de crescimento desde a implantação do Plano Real (Julho/94). Segundo CABRAL & PINHEIRO (1998) este efeito se observou como conseqüência da estabilização econômica, que provocou uma considerável queda na incerteza geral da economia, encorajando uma expansão substancial das linhas de crédito ao consumo. Com o aumento da oferta houve uma redução nas taxas de juros cobradas do consumidor e a expansão de outros meios de pagamento, tal como o cartão de crédito, que dobrou seu volume entre 93 e 97. Estes efeitos provocaram a estimulação de uma demanda latente até então inexplorada, e com isso houve aumento do investimento das empresas dos segmentos de financeiras, administradoras de cartão de crédito e bancário, entre outros, provocando inclusive o surgimento de novas empresas neste mercado.

Com a redução da demanda não atendida, houve aumento da competitividade, e a principal conseqüência foi que bancos e financeiras foram obrigados a evoluir tecnicamente, aperfeiçoando seu "modus operandis", principalmente em setores de monitoramento e controle do risco de crédito nas várias fases do ciclo de crédito, tais como na iniciação ao cliente, no gerenciamento de carteiras, na cobrança, e na prevenção à fraude, por exemplo. Uma das tecnologias fundamentais para medição, monitoração e controle do risco de crédito é a utilização das informações transacionais dos clientes para elaboração de modelos estatísticos com função de medir a expectativa de inadimplência do consumidor. A relevância de um modelo deste tipo é que ele permite determinar o risco do crédito ao consumidor, tornando possível segmentar a carteira de clientes conforme

probabilidade de inadimplência, e viabilizando o cálculo da lucratividade esperada e cálculo do ponto máximo de lucratividade.

Pela simplicidade, a metodologia mais utilizada no desenvolvimento de um modelo de previsão do tipo ***Behaviour Score*** é a regressão logito. Para isto, usualmente é selecionada uma amostra de indivíduos em que cada indivíduo é avaliado em um único momento "t" do tempo, ou seja amostra em ***cross-section***. Neste tipo de amostra, as variáveis independentes são obtidas utilizando a informação de perfil ou comportamento do cliente disponível até o tempo "t", e a variável dependente Inadimplência (Mau=1) ou Adimplência (Bom=0) é obtida utilizando informações do tempo "t+s", onde "s" é o tempo de desempenho.

É interessante notar que entre as técnicas alternativas propostas para modelar a probabilidade de inadimplência estão os algoritmos genéticos, redes neurais, análise de sobrevivência, programação linear, mas usualmente a conclusão é que não existe ganho significativo na previsão obtida a partir destas metodologias alternativas. Talvez estes resultados similares se devam ao fato das amostras de desenvolvimento serem geradas sempre sob a mesma perspectiva de ***cross-section***, não agregando informação aos dados.

Segundo MARQUES (2000), a modelagem utilizando dados em painel implica em maior quantidade de informação, e consequentemente maior eficiência na estimação, pois a amostra observa cada indivíduo sob uma perspectiva temporal, e não apenas como um corte no tempo. Então, a amostra contém uma análise longitudinal de cada indivíduo, permitindo avaliar não só a diversidade dos comportamentos individuais, mas também a existência de suas dinâmicas. Desta forma, considerando que a amostra para a construção do modelo com dados em painel contém mais informação, espera-se que este modelo de previsão da inadimplência seja mais eficaz que o modelo tradicionalmente desenvolvido.

2. REVISÃO BIBLIOGRÁFICA

Ao longo da história, algumas metodologias têm sido criadas e desenvolvidas com o objetivo de gerar modelos melhores para a previsão da inadimplência. Apesar deste trabalho não se propor a testar estas novas metodologias, será feita uma breve revisão histórica das tecnologias mais reconhecidas, assim como dos trabalhos que buscaram comparar a eficiência entre estas técnicas.

2.1 – Análise Discriminante

Ao longo dos anos, inúmeros trabalhos foram publicados a respeito do tema classificação de consumidor. FISHER (1936) foi o primeiro a trabalhar um modelo para diferenciar 2 grupos em uma população e DURAND (1941) foi o primeiro a aplicar a metodologia no segmento financeiro, diferenciando bons empréstimos dos maus empréstimos no National Bureau of Economic Research (EUA). ALTMAN (1968) também contribuiu muito incentivando a utilização deste tipo de técnica de decisão. Nesta época, as regras de bolso escritas por analistas experientes, denominadas sistemas especialistas, é que tomavam a decisão de conceder ou não o crédito, e dessa forma, permitiam que pessoas inexperientes conseguissem decidir acertadamente com um razoável grau de precisão para a época.

FISHER buscou identificar uma combinação linear das variáveis independentes que melhor separasse os dois grupos. Foi criada a Função Discriminante Linear[1], definida como:

$$Y = W_1 X_1 + W_2 X_2 + ... + W_p X_p$$

Onde W_i são os pesos atribuídos a cada uma das variáveis X_i, de forma a maximizar o poder discriminador de Y.

Uma vez estabelecida a função discriminante, é possível atribuir uma pontuação Y_i para cada indivíduo "i" da amostra, e restará determinar o ponto de corte Y_c, que deverá minimizar os erros Tipo I (aprovar crédito para indivíduos inadimplentes) e Tipo II (negar crédito para indivíduos adimplentes) .

[1] Descrição detalhada da técnica pode ser encontrada em SHARMA(1996) e JOHNSON & WICHERN (1982)

Desta forma:

- Se $Y_i < Y_c \Rightarrow$ crédito será rejeitado
- Se $Y_i \geq Y_c \Rightarrow$ crédito será aprovado

MARTELL & FITTS (1981) evoluíram para os modelos de análise discriminante quadrática, que pondera as diferentes variâncias das populações de adimplentes e inadimplentes. EISENBEIS (1977) criticou a utilização da análise discriminante devido à dificuldade de se separar a população em 2 grupos distintos e EISENBEIS (1978) ressaltou a necessidade de considerar aspectos de evolução do relacionamento do cliente com a instituição, afirmando que a ausência desta dinâmica leva à uma decisão menos eficaz. WIGINTON (1980) comparou o desempenho de um modelo de ***Credit Scoring*** utilizando análise discriminante com um modelo logito, concluindo que o último teve um desempenho ligeiramente superior. CAPON (1982) e ROSENBERG & GLEIT (1994) observaram que o tempo ocasionava mudanças frequentes nos modelos estimados. CAPON (1982), particularmente, acreditava que maior peso deveria ser dado ao comportamento de crédito do consumidor, ou seja, mais ênfase ao ***Behaviour Score.***

2.2 – Programação Linear

A programação linear também pode ser usada com o fim de gerar previsões para a inadimplência. Nesta técnica deseja-se construir uma pontuação através de um modelo linear que utiliza as variáveis independentes X de forma a deixar todos os N_G adimplentes acima de um valor predito "c" qualquer, e ao mesmo tempo deixar todos os N_B inadimplentes abaixo deste mesmo valor "c". Como isso não é possível, introduz-se a variável $|a_i|$, que contém o erro tipo I e tipo II, mas que deve ser minimizada. Desta forma, o modelo terá o seguinte formato:

Mín $|a_1| + |a_2| + ... + |a_{N_G+N_B}|$

Suj. a $w_1 x_{i,1} + w_2 x_{i,2} + ... + w_m x_{i,m} \geq c - a_i$ onde $1 \leq i \leq N_G$

$w_1 x_{i,1} + w_2 x_{i,2} + ... + w_m x_{i,m} \leq c + a_i$ onde $N_G + 1 \leq i \leq N_G + N_B$

Repare que $1 \leq i \leq N_G + N_B$ e que w_i = coeficiente da variável independente

MANGASARIAN (1965) foi o primeiro a abordar o problema desta forma, mas FREED & GLOVER (1981a,b) que fizeram a técnica ser de fato considerada para a previsão da inadimplência. NATH, JACKSON & JONES (1992) concluíram que o método logístico é melhor que aquele que utiliza PL, e HARDY & ADRIAN(1985) concluíram que PL classifica tão bem quanto a metodologia estatística tradicional.

GEHRLEIN & WAGNER (1997) desenvolveram um formulação de Programação Linear que incorpora na função objetivo o custo de inadimplência e o custo de oportunidade, dando um sentido mais moderno de gestão de carteira, e não simplesmente de previsão da probabilidade de inadimplência. A novidade deste modelo é incorporar a informação de política de juros e o custo de inadimplência, o que demanda um monitoramento mais freqüente, dado que mudanças de mercado podem modificar rapidamente o ponto "c" ótimo.

SCARPEL & MILIONI (2002) propuseram a utilização conjunta dos modelos de programação linear e de regressão logito, unindo num só modelo a probabilidade de inadimplência e a lucratividade (taxa de juros) do concedente, permitindo determinar o valor de empréstimo ótimo.

2.3 – Redes Neurais

Outra técnica que ganha cada vez mais força em modelos de risco de crédito são as Redes Neurais. A literatura[2] descreve esta técnica como sendo baseada no sistema nervoso central humano e uma rede neural pode ser classificada segundo suas características principais, ou seja, "topologia da rede neural", "forma de aprendizado" e "algoritmo de aprendizado".

A principal vantagem é que apesar do modelo de Redes Neurais ser fixo, o processo de aprendizado realizado constantemente permite modificação freqüente da fórmula. A desvantagem é que a mudança constante dificulta a observação dos fenômenos que geram a modificação no modelo.

[2] Foram utilizadas as seguintes referências: BARTH(2002), BERRY & LINOFF (1997), BIGUS (1996), GLUESNSTEIN(1998), PATERSON(1996) e THOMAS(2000).

Veja abaixo uma representação gráfica de uma Rede Neural, onde é possível observar sua topologia.

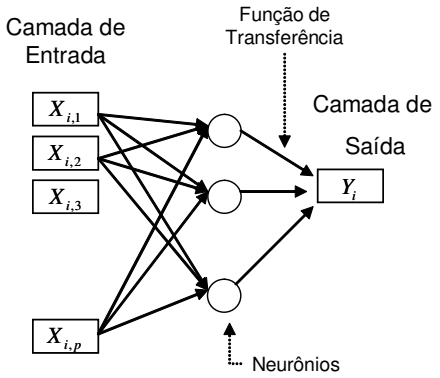


Figura 1: Representação Gráfica de uma Rede Neural

Inicialmente é importante conhecer a topologia de uma Rede Neural. Na camada de entrada estão todas as X_p variáveis independentes, que alimentam vários neurônios, e que por sua vez, calculam a função de transferência, que nada mais é que uma soma ponderada dessas variáveis independentes utilizando uma função que pode ser linear, sigmóide (também conhecida como logística) ou tangente hiperbólica. O caso da Figura 1 é uma rede neural do tipo ***one-layer,*** onde a saída recebe o input dos neurônios através da função de transferência e produz a pontuação final do modelo. Também é possível criar uma segunda camada de neurônios, que ficam posicionados entre os neurônios iniciais e a camada de saída, encadeando então vários neurônios e criando o caso ***two-layer***, que produz melhores estimativas em caso de relações fortemente não-lineares nos dados. Os neurônios posicionados nas camadas ocultas escondem a função de transferência.

Outra característica muito importante de uma rede neural é a forma de aprendizado, e a metodologia mais utilizada em crédito ao consumidor é o aprendizado supervisionado, que utiliza uma amostra de desenvolvimento com variáveis independentes e dependentes como referência (ou balizadora) para qualquer mudança sugerida pelo algoritmo de aprendizagem.

O algoritmo de aprendizagem, ou seja, a reponderação dos pesos da função de transferência é usualmente feita através do algoritmo de backpropagation, que identifica os

erros cometidos pelos neurônios de saída e corrige os pesos da(s) função(ões) de transferência das camadas anteriores. Este algoritmo é muito caro computacionalmente devido ao tempo de processamento, e está baseado em ajuste recursivo dos pesos da função de transferência no sentido de redução do erro na camada de saída. Este processo é chamado treinamento da rede e questões como condição de parada para não gerar ***overfitting*** e soluções sub-ótimas devem ser tratadas cuidadosamente, pois a rede neural não é estimada simultaneamente como nos outros métodos, mas através de uma aprendizagem recursiva e seqüencial que deve ser supervisionada.

HAIR (1998, pg. 685) destaca a inexistência de qualquer teste estatístico para avaliar a significância dos pesos da função de transferência, mas entre outros, ALTMAN (1994) comprovou a utilidade da aplicação desta metodologia em modelos de previsão de inadimplência, principalmente em casos onde existe não-linearidade nos dados da amostra.

2.4 – Algoritmos Genéticos

Algoritmo genético é uma técnica de otimização baseada no conceito de evolução de Darwin, onde os indivíduos mais aptos tendem a sobreviver através da seleção, reprodução e mutação. Este conceito foi criado por John Holland, um professor de Psicologia e Ciência da Computação da Universidade de Michigan, mas FOGARTY & IRESON (1993) e ALBRIGHT (1994) foram os primeiros a descrever esta tecnologia no campo da estatística[3].

Imagine então uma amostra de indivíduos para o desenvolvimento do modelo, onde estão disponíveis algumas variáveis independentes representando a história do indivíduo e uma variável dependente (adimplente/inadimplente) que se pretende prever. A pergunta a ser respondida é quais variáveis são melhores para discriminar a variável dependente e qual o peso de cada variável para atingir esta classificação maximizada?

A metodologia de Algoritmos Genéticos é descrita detalhadamente por BARTH (2002). Este processo de modelagem é todo baseado em rotinas de sorteios aleatórios, assim como nos processos genéticos, gerando inúmeras fórmulas (mais conhecidas como regras) que são testadas empiricamente quanto à eficácia na previsão da inadimplência. Obviamente a

regra escolhida será aquela onde as menores pontuações são dadas aos indivíduos que foram inadimplentes de fato.

O refinamento da capacidade preditiva do modelo de algoritmo genético é feito através da fase chamada "reprodução", onde novas fórmulas são geradas a partir da combinação das fórmulas mais eficazes da fase anterior do processo de sorteio. Este é um método recursivo conhecido como seleção de genitores, onde se deseja convergir para regras homogêneas.

Outro processo de refinamento ocorre na fase seguinte, chamada mutação, onde se modifica aleatoriamente uma variável independente escolhida também ao acaso, causando uma evolução que pode ter efeito positivo ou negativo na previsão da inadimplência. Este mecanismo tem o objetivo de evitar a convergência para regras não ótimas.

O procedimento recursivo destas 3 fases da técnica (estabelecimento de regras, reprodução e mutação) tendem a convergir para fórmulas semelhantes, todas igualmente aptas e com alto poder discriminador, que deverão ser estudadas pelo analista para validação. A condição de parada está relacionada à uma função objetivo relacionada ao número de erros (ou custo do erro), que deve ser minimizada. Entretanto, esta convergência não acontecerá sempre, pois as variáveis independentes podem ser inadequadas para prever o fenômeno desejado, ou podem ocorrer convergências rápidas, que podem denotar regras de discriminação fracas.

A metodologia de algoritmos genéticos demanda do analista 3 tipos de calibração, que são o número "N" de indivíduos iniciais, a taxa de reprodução e a taxa de mutação. A seleção destes parâmetros envolve a base de dados e a experimentação do analista, mas é importante salientar que os resultados obtidos dependem das escolhas destes parâmetros, e por isso é recomendável processar a rotinas mais de uma vez para comparar os resultados.

Há também que se considerar que esta metodologia gera modelos de fácil interpretação prática, pois o resultado é uma fórmula com variáveis e pesos, que determinarão diretamente a inadimplência esperada de cada indivíduo da amostra.

2.5 – Análise de Sobrevivência

[3] Outras literaturas importantes no campo de Algoritmos Genéticos são BAUER(1994), VARETTO(1998) e GOLDEMBERG(1989).

A análise de sobrevivência é outra técnica aplicada para estimar a probabilidade de inadimplência em um tempo pré-determinado e um dos primeiros propositores desta técnica foi NARAIN (1992). Nas técnicas descritas anteriormente, a atribuição de adimplente ou inadimplente para o indivíduo da amostra é executada a partir da análise da situação no tempo pré-determinado t+s. Entretanto, a análise de sobrevivência está preocupada com o momento da inadimplência, ou seja, modelará qual o tempo previsto da inadimplência, pressupondo que o evento de interesse ocorrerá, mesmo que em algum momento longínquo. Repare que este conceito é o mesmo utilizado para modelar tempo de vida de lâmpadas, máquinas ou pessoas. Entre outras, existem 2 vantagens deste tipo de modelagem:

- evitar a instabilidade advinda da escolha de um período de tempo pré-determinado para medição da performance.
- previsão dos índices de inadimplência em função do tempo, o que é muito útil para gestão de carteiras.

STEPANOVA E THOMAS (2001) trabalharam no conceito de utilização da técnica de análise de sobrevivência na elaboração de modelos de "***Behaviour Score"***. Neste artigo os autores enfatizam dois pontos muito importantes, a saber:

- A dinâmica do modelo precisa seguir a dinâmica da população, e em muitos casos a totalidade dos consumidores não seguem o mesmo processo estacionário. Neste caso é interessante definir subpopulações, e testar como os consumidores mudam de estado em cada grupo de indivíduos. É preciso estabelecer agrupamentos que sejam estáveis e internamente homogêneos.
- Este modelo está especialmente preocupado com a diferença na probabilidade de inadimplência entre os momentos t+1 e t+6. No artigo os autores calcularam estas predições nos momentos 0,1,2,3,4,5 e 6, utilizando apenas a informação disponível até o respectivo momento, e observaram que os coeficientes do modelo poderiam variar muito. Desta forma, poderia haver uma melhoria na eficácia do modelo quando se considera esta evolução. Alguns resultados

comparativos entre análise de sobrevivência e regressão logito já tinham sido apresentados em BANASIK, et al (1999).

2.7 – Comparação de Resultados

BARTH (2002) e THOMAS (2000) resumem em seus trabalhos os resultados de alguns estudos comparativos entre a utilização de diferentes técnicas na previsão da inadimplência.

BARTH (2002) relata em seu trabalho que:

- ALTMAN (1994) conclui que os resultados da metodologia de Redes Neurais se mostraram piores do que aqueles alcançados com Análise Discriminante e/ou Regressão Logística quando aplicados à amostra de validação.

- VARETTO (1998) concluiu que os resultados de Algoritmos Genéticos também se mostraram piores do que aqueles alcançados com Análise Discriminante e/ou Regressão Logística quando aplicados à amostra de validação.

- ADYA e COLOPPY (1998) estudaram vários trabalhos comparativos da técnica de Redes Neurais com outros métodos. Apesar de não definitivo, alguns dos trabalhos que receberam crédito dos autores posicionavam a metodologia de Redes Neurais como melhor comparativamente à Análise Discriminante/Regressão Logística em algumas situações específicas.

THOMAS (2000), por outro lado, apresenta uma tabela comparativa entre alguns estudos. Os indicadores apresentados se referem ao porcentual corretamente classificado pelos diferentes métodos. Repare que as comparações devem ser feitas na mesma linha, pois os estudos foram realizados em diferentes amostras.

Tabela 1 – Comparação entre técnicas em diferentes estudos

Autores	Regr. Linear	Regr. Logística	Program. Linear	Redes Neurais	Alg. Gen.
Henley (1995)	43,4	43,3	-	-	-
Boyle et al (1992)	77,5	-	74,7	-	-
Srinivisan & Kim (1987)	87,5	89,3	86,1	-	-
Yobas, Crook, Ross (1997)	68,4	-	-	62,0	64,5
Desai, Convay, Crook, Overstreet (1997)	66,5	67,3	-	64,0*	-

* O artigo apresenta 6,4; mas acredita-se que exista um erro de impressão.

Conclui-se desta tabela que:

- A Regressão Linear foi fracamente melhor em BOYLE et al (1992) e YOBAS et al (1997).
- Por outro lado, DESAY et al (1997) e SRINIVISAN & KIM (1987) apontaram a Regressão Logística como fracamente melhor.

Logo, observa-se na literatura existente que não existe uma técnica que seja unânime e apontada como a melhor em termos de eficácia na previsão. Muitas vezes, a aplicação de técnicas diferentes levam a resultados semelhantes, e portanto fica evidente que em um novo modelo, o melhor é testar todas as metodologias, escolhendo aquela que for mais conveniente, pois os melhores resultados parecem também depender do fenômeno e da base de dados estudada. Devido aos recursos computacionais disponíveis hoje, esta estratégia é perfeitamente factível.

Entretanto, THOMAS (2000) afirma que de antemão espera-se que Redes Neurais sejam melhores para estimar relações não-lineares, e que as Regressões Linear e Logística tenham a vantagem de gerar modelos mais robustos, dado que testes de significância são preliminarmente executados. Quanto à Programação Linear, a principal vantagem é a determinação à priori de propriedades desejadas através do estabelecimento de restrições, particularidade esta que os outros modelos não possuem.

3. METODOLOGIA ECONOMÉTRICA

3.1 – Modelo Linear de Efeitos Fixos

Conforme citado na introdução, os modelos para prever a probabilidade de inadimplência são usualmente feitos sob a perspectiva de ***cross-section***, mas segundo HSIAO (2003), modelos desenvolvidos desta forma tratam o agregado dos efeitos individuais e o efeito de variáveis omitidas (decorrente da assimetria de informação) como um evento aleatório único descrito nos erros ε.

Um exemplo claro disso é a característica desonestidade, que não está disponível para ajudar a prever a chance de inadimplência, mas é provável que os 50% mais desonestos da amostra tenham inadimplência significativamente maior que os outros 50% mais honestos. A característica honestidade é inerente ao indivíduo, com pouca variação no tempo, e desconhecida. Por isso trata-se de um efeito fixo C_i contido no erro aleatório e causado por uma variável omitida. No contexto de dados em ***cross-section***, este tipo de problema pode ser resolvido:

- Através de uma variável X_i que seja proxy para C_i
- Através de variáveis instrumentais X_i, que ajudam a estimar o efeito da variável omitida (no caso honestidade)

Outra possibilidade para estimar este efeito não-observável é utilizar dados em painel, ou seja, modelar observando as variáveis independentes e dependentes em vários períodos de tempo. Desta forma, modelos de previsão da inadimplência que antes eram feitos considerando-se apenas a dimensão "i" do indivíduo, também levarão a dimensão do tempo "t" de cada indivíduo em consideração.

Então, na presença de efeitos fixos inerentes ao indivíduo, a amostra longitudinal permitirá decompor o erro de modelagem de cada unidade amostral $\varepsilon_{i,t}$ em uma

componente fixa C_i, que é a porção do erro que é inerente ao indivíduo e invariante no tempo, e uma nova componente de erro aleatório $U_{i,t}$ de menor magnitude, gerando uma previsão que seja consistente por isolar o efeito de variáveis omitidas.

Matematicamente, o modelo estrutural de um painel de dados com efeitos fixos é definido como:

$$Y_{i,t} = \alpha + \beta.X_{i,t} + C_i + U_{i,t}$$

Onde:

$Y_{i,t}$ = informação da evolução temporal da variável dependente para todos os indivíduos da amostra.

$X_{i,t}$ = informação da evolução temporal das variáveis independentes para todos os indivíduos da amostra

C_i = efeito fixo no tempo e específico do indivíduo, ou seja, o modelo terá a presença de "n" efeitos fixos (um para cada indivíduo da amostra).

$U_{i,t}$ = Erro aleatório do modelo, que tem distribuição i.i.d.(0, σ^2).

Observe que o subscrito "i" se refere à unidade amostral (consumidor) e que o subscrito "t" se refere à evolução temporal. A metodologia de dados em painel tratará simultaneamente toda esta amostra, que é composta de "i" séries de tempo com "t" observações para cada uma delas.

Repare que o efeito fixo C_i não é estimado em um modelo tradicional do tipo ***cross-section***, mas que supor sua presença é como admitir que os erros $\varepsilon_{i,t}$ e $\varepsilon_{i,t-1}$ são correlacionados, pois ambos os erros $\varepsilon_{i,t} = C_i + U_{i,t}$ e $\varepsilon_{i,t-1} = C_i + U_{i,t-1}$ tem uma porção comum C_i.

Também é importante perceber que neste novo modelo é necessário estimar "n" efeitos fixos C_i, causando a perda de muitos graus de liberdade, devido ao número muito maior de variáveis a serem estimadas. Entretanto, estes efeitos fixos C_i podem ser eliminados em

um modelo de regressão linear através de uma transformação onde primeiramente se obtém a média de cada variável para cada indivíduo "i" da seguinte forma:

$$\overline{Y_i} = \frac{\sum_{t=1}^{T} Y_{i,t}}{T} \quad , \qquad \overline{X_i} = \frac{\sum_{t=1}^{T} X_{i,t}}{T} \qquad \text{e} \qquad \overline{U_i} = \frac{\sum_{t=1}^{T} U_{i,t}}{T}$$

E logo após obtém-se os desvios com relação à média para cada observação através do seguinte cálculo:

$$Y_{i,t} - \overline{Y_i} = \beta.\left(X_{i,t} - \overline{X_i}\right) + \left(C_i - C_i\right) + \left(U_{i,t} - \overline{U_i}\right)$$

Esta transformação remove os efeitos específicos individuais, e uma regressão OLS é aplicada em $\ddot{Y}_{i,t} = \beta.\ddot{X}_{i,t} + \ddot{U}_{i,t}$ onde $\ddot{Y}_{i,t} = Y_{i,t} - \overline{Y_i}$, $\ddot{X}_{i,t} = X_{i,t} - \overline{X_i}$ e $\ddot{U}_{i,t} = U_{i,t} - \overline{U_i}$

Portanto:

$$\hat{\beta}_{Efeito\ Fixo} = \left(\sum_{i=1}^{N} \ddot{X}_i^{'}.\ddot{X}_i\right)^{-1}\left(\sum_{i=1}^{N} X_i^{'}.\ddot{Y}_i\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{X}_{it}^{'}.\ddot{X}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{X}_{it}^{'}.\ddot{Y}_{it}\right)$$

3.2 – Modelos Logito

3.2.1 – Modelo Logito em ***Cross-Section***

Neste trabalho, o modelo para estimação da probabilidade de inadimplência será construído utilizando variáveis dependentes discretas do tipo adimpliu/inadimpliu, que precisam de um tratamento teórico especial, pois:

- A variável dependente é dicotômica (0 e 1).
- A previsão deve estar necessariamente entre "0" e "1".

- Os erros NÃO são normalmente distribuídos

Então para calcular a $P[Y=1 \mid X]=E[Y \mid X]$ pode-se utilizar o seguinte modelo linear:

$$P[Y_i=1 \mid X_i]=\beta_0+\beta_1.X_1+\beta_2.X_2+...+\beta_k.X_k$$

Onde Y_i é a variável dependente, que só pode assumir os valores "0" ou "1". Repare que neste modelo, o interesse é explicar os efeitos de X_i na probabilidade de inadimplência, mas na prática não se pode observar o efeito da variável X_i sobre a probabilidade de inadimplência, que é uma média condicional não observável. A alternativa é utilizar a variável inadimplência observada Y_i para calcular o efeito de mudança na inadimplência causada pela aumento/redução de uma unidade em X_i, ou seja, $\beta_1=\frac{\partial P(Y_i=1 \mid X_i)}{\partial X_i}$. WOOLDRIDGE (2002, pg. 458) afirma que a direção dos efeitos da variável X_i na variável observável Inadimplência/Adimplência efetiva (representado por Y_i) é a mesma que sobre a média condicional não-observável probabilidade de inadimplência.

Mas a variável inadimplência, representada por $P[Y_i=1 \mid X_i]$, está definida apenas nos valores "0" e "1", enquanto o cálculo $\beta_0+\beta_1.X_1+...+\beta_k.X_k$ pode variar entre $-\infty$ e $+\infty$. Logo, a função logística G(X) representada pela fórmula abaixo pode ser utilizada para representar este fenômeno, pois além de estar limitada ao intervalo "0" e "1", ainda tem uma característica assintótica[4] que permite modelar melhor os indivíduos adimplentes e inadimplentes representados na Figura 2.

$$E[Y \mid X]=G(X\beta)=\frac{\exp(\beta_0+\beta_1.X)}{1+\exp(\beta_0+\beta_1.X)}$$

[4] A função converge "rapidamente" a "0" quando X diminui e converge "rapidamente" a "1" quando X aumenta

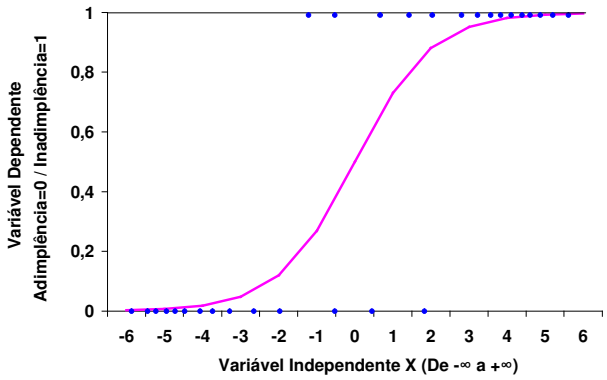


Figura 2: Exemplo da função logística modelando chance de inadimplência

Observe na figura acima que a chance do consumidor ser adimplente diminui conforme a variável independente decresce, e por outro lado, a chance de inadimplência aumenta conforme a variável dependente aumenta.

HOSMER & LEMESHOW (2000, pg. 47-48) relatam de forma especial que nos modelos do tipo logito, os coeficientes β´s devem mensurar a taxa de mudança da variável dependente Y por unidade de mudança na variável independente X. Mas para isto é necessário estabelecer uma relação funcional linear entre a variável dependente e a variável independente. São as chamadas "Funções de Ligação".

Como podemos observar na Figura 2, a função logística G(X) não é linear, mas dividindo-se numerador e denominador de $G(X\beta)$ por $\exp(\beta_0 + \beta_1.X_i)$, obtém-se:

$$G(X\beta) = \frac{1}{1+\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]}$$

$$\text{Logo } 1 - G(X\beta) = \frac{\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]}{1+\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]}$$

HOSMER & LEMESHOW (2000, pg. 6) apresentam então a ***transformação logito,*** que é linear nos seus parâmetros, e obtida a partir da razão de chances da função logística, ou seja:

$$\frac{G(X\beta)}{1-G(X\beta)} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]} . \frac{1+\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]}{\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]} =$$

$$= \frac{1}{\exp[-(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)]} = \exp[(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k)].$$

$$\Rightarrow \ln\left[\frac{G(X\beta)}{1-G(X\beta)}\right] = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$$

Repare que a estimação de β_0 e β_1 pode ser realizada a partir da existência de N observações i.i.d. através da função de verossimilhança condicional, que pode ser definida como:

$$L(\beta) = \prod_{i=1}^{n} [G(X\beta)]^{y_i} [1 - G(X\beta)]^{1-y_i}$$

Aplicando-se a função logaritmo a ambos os lados da função de verossimilhança condicional e derivando-se, obtém-se a função de máxima log-verossimilhança:

$$l(\beta) = \sum_{i=1}^{n} y_i . \ln[G(X.\beta)] + (1 - y_i) \ln[1 - G(X.\beta)]$$

A estimação do vetor $\hat{\beta}$ que maximiza $l(\beta)$ envolve derivar esta função em relação a cada um dos β_i´s e igualar a zero, encontrando o valor que maximiza a função. Estes cálculos estão amplamente implementados em diversos pacotes estatísticos através de procedimentos computacionais de análise numérica, porque apesar da derivada da função logística G(X) existir, só pode ser encontrada através de métodos iterativos de otimização. Veja em VASCONCELLOS (2002) os passos necessários para o desenvolvimento de um modelo para previsão de inadimplência utilizando esta tecnologia.

Como a otimização da função de verossimilhança se dá iterativamente, ainda é possível escolher um algoritmo de seleção de variáveis que escolha apenas as variáveis significativas para compor o modelo logito final. Os métodos mais conhecidos são o backward selection, backward hierarchical selection, backward stepwise, forward selection, forward hierarchical selection e forward stepwise. VASCONCELLOS (2002) não apenas

afirma que os métodos stepwise são adequados por permitir examinar com rapidez um grande número de variáveis, com ainda faz uma descrição detalhada de todos os passos do procedimento Stepwise, que é muito útil para compreensão do algoritmo.

Entretanto, na estimação do modelo logito deste trabalho não foi utilizado nenhum método de seleção, pois como o objetivo principal foi comparar as técnicas de regressão logito em ***cross-section*** e em painel, os 2 modelos utilizaram o mesmo número de variáveis independentes, isolando então, efeitos de assimetria de informação entre as técnicas.

3.2.2 – Modelo Logito com Efeitos Fixos

A técnica alternativa proposta neste trabalho é a regressão logito utilizando amostra com dados em painel, ou seja, com informações de "N" indivíduos em "T" tempos. A estimação de β sem considerar a existência dos efeitos específicos $C_{i's}$ descritos em 3.1 é realizada considerando todos os dados como uma grande e independente cross-section de tamanho NT, ou seja, a função de máxima verossimilhança a ser maximizada é:

$$l(\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\{Y_{it}.\ln G(X_{it}.\beta) + (1-Y_{it}).\ln[1-G(X_{it}.\beta)]\}$$

Entretanto, a presença do efeito específico não observável C_i no modelo logito torna a solução do "Incidental Parameter Problem" muito mais difícil, pois a não-linearidade da variável dependente inviabiliza a aplicação da solução de desvios com relação à média apresentada no modelo linear de efeitos fixos.

Segundo HSIAO (2003) este problema foi tratado por ANDERSEN (1970,1973), que demonstrou como maximizar densidades condicionais; por MCFADDEN (1974), que apresentou o logit condicional em painel; e por CHAMBERLAIN (1980), que demonstrou que a matriz de covariância condicionada à $\sum y_{i,t}$ converge assintoticamente para β quando N tende ao infinito. Desta forma, condicionando a distribuição de $Y_{i,t}$ à C_i e à $n_i = \sum_{t=1}^{T} Y_{i,t}$ (com T=número de períodos avaliados), o efeito específico C_i é eliminado do estimador, possibilitando o cálculo sem o "Incidental Parameter Problem". No problema de

prever a probabilidade de inadimplência, condicionar em n_i significa calcular a estatística apenas para os consumidores que mudaram de estado, ou seja, que inadimpliram pelo menos 1 vez nos "T" períodos estudados, pois quando n_i=0, a contribuição da observação para a verossimilhança condicional é nula.

WOOLDRIDGE (2002) apresenta um exemplo simples em que T=2, $Y_{i,t}$ é a variável resposta dicotômica do indivíduo "i" no tempo "t", e portanto n_i={0,1,2}. Nos casos de indivíduos em que n_i=0 e n_i=2, a variável resposta $Y_{i,t}$ não mudou de estado, e portanto estes indivíduos não contribuem na estimação de β, pois sob a perspectiva de painel, não existe meio de observar como $X_{i,t}$ está influenciando $Y_{i,t}$ nestes casos. Mas observe que se n_i=1, então houve mudança de estado, e a função de máxima verossimilhança condicional é calculada em duas partes.

A probabilidade de inadimplir em T=1 é:

$$P\left[Y_{i,2}=1 \mid X_{i,2}, C_i, n_i=1\right]=\frac{P\left[Y_{i,2}=1, n_i=1 \mid X_i, C_i\right]}{P\left[n_i=1 \mid X_i, C_i\right]}=$$

$$=\frac{P\left[Y_{i,2}=1 \mid X_{i,2}, C_i\right].P\left[Y_{i,1}=0 \mid X_{i,2}, C_i\right]}{P\left[Y_{i,1}=0, Y_{i,2}=1 \mid X_{i,2}, C_i\right]+P\left[Y_{i,1}=1, Y_{i,2}=0 \mid X_{i,2}, C_i\right]}=$$

$$=\frac{\Lambda.\left(X_{i,2}.\beta+C_i\right).\left[1-\Lambda.\left(X_{i,1}.\beta+C_i\right)\right]}{\left[1-\Lambda.\left(X_{i,1}.\beta+C_i\right)\right].\Lambda.\left(X_{i,2}.\beta+C_i\right)+\Lambda.\left(X_{i,1}.\beta+C_i\right).\left[1-\Lambda.\left(X_{i,2}.\beta+C_i\right)\right]}$$

Lembrando que a função logística Λ é definida como $\Lambda.\left(X.\beta\right)=\frac{\exp\left(X.\beta\right)}{1+\exp\left(X.\beta\right)}$

Conclui-se que $P\left[Y_{i,2}=1 \mid X_{i,2}, C_i, n_i=1\right]=\frac{\exp\left[C_i+X_{i,2}.\beta\right]}{\exp\left[C_i+X_{i,2}.\beta\right]+\exp\left[C_i+X_{i,1}.\beta\right]}=$

$$= \frac{\exp[C_i].\exp[X_{i,2}.\beta]}{\exp[C_i].\exp[X_{i,2}.\beta] + \exp[C_i].\exp[X_{i,1}.\beta]} = \frac{\exp[X_{i,2}.\beta]}{\exp[X_{i,2}.\beta] + \exp[X_{i,1}.\beta]} = \Lambda[(X_{i,2} - X_{i,1})\beta]$$

Similarmente, a probabilidade de inadimplir em T=1 é:

$$P[Y_{i,1} = 1 \mid X_{i,1}, C_i, n_i = 1] = \Lambda[-(X_{i,2} - X_{i,1})\beta] = 1 - \Lambda[(X_{i,2} - X_{i,1})\beta]$$

Então, a log-verossimilhança condicional neste caso é:

$$l(\beta) = \sum_{i=1}^{n} 1[n_i = 1].\{w_i.\ln \Lambda[(X_{i,2} - X_{i,1}).\beta]\} + (1 - w_i).\ln\{1 - \Lambda[(X_{i,2} - X_{i,1}).\beta]\}$$

Onde:

- $w_i = 1$ se $(Y_{i,1} = 0, Y_{i,2} = 1)$ e $w_i = 0$ se $(Y_{i,1} = 1, Y_{i,2} = 0)$
- função indicadora $1[n_i = 1]$ seleciona apenas os indivíduos que mudam o estado nos "T" períodos. Portanto, consumidores com $1[n_i = 0]$ e $1[n_i = 2]$ não estão contribuindo para a verossimilhança.

Repare que a necessidade de variação da variável dependente em cada indivíduo da amostra é uma limitação do modelo logito em painel, e que ocasionará a eliminação de muitos indivíduos do cálculo da verossimilhança da amostra. A mesma exigência de variação não existe no modelo logito em cross-section, inclusive porque não se toma mais que uma variável dependente para um mesmo indivíduo na amostra.

Outra importante limitação da metodologia de dados em painel é a impossibilidade de utilizar variáveis independentes que não variam no tempo no cálculo da verossimilhança[5]. Desta forma não é possível aplicar esta metodologia em variáveis independentes de perfil, tais como renda, profissão ou grau de instrução, que muitas vezes tem variação nula. Por isto, este trabalho foi empiricamente testado em um modelo de ***Behaviour Score***, que considera apenas informações de comportamento dos consumidores.

3.3 – Medidas de Comparação entre Modelos

Uma questão fundamental neste trabalho é como medir o desempenho dos modelos de predição para escolher o melhor entre as 2 metodologias propostas. Entre outras, foram escolhidas 2 medidas para executar estes testes comparativos.

3.3.1 - Teste de KS (Kolmogorov-Smirnov)

A estatística de Kolmogorov-Smirnov (KS) é uma das medidas de desempenho mais conhecidas e utilizadas no mercado de crédito ao consumo. A hipótese deste teste é supor que indivíduos com alta chance de serem inadimplentes devem estar concentrados no baixo valor predito, e indivíduos com alta chance de serem adimplentes devem estar concentrados no alto valor predito. Desta forma, o melhor modelo deverá prover a maior separação entre consumidores adimplentes e inadimplentes ao longo do valor predito, e isto sendo verdade, a curva de distribuição acumulada da população de bons pagadores cresce antes que a curva de distribuição acumulada da população de maus pagadores. Repare que quanto maior for esta separação, maior será a máxima diferença entre estas duas curvas acumuladas, pois a estatística KS é definida como:

$$\text{KS} = \max \left| F_{Bons}(Score) - F_{Maus}(Score) \right|$$

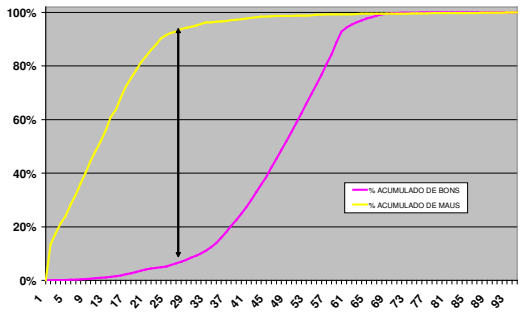
Onde: F_{Bons} = Distribuição Acumulada de Adimplentes

F_{Maus} = Distribuição Acumulada de Inadimplentes

Observe um exemplo no gráfico abaixo.

Gráfico 1 – Exemplo de Teste KS

[5] Observe que o estimador de β é baseado na variação da variável independente $X_{i,t}$.



Valor Predito*100

Repare que no exemplo do Gráfico 1, o ponto de maior distância entre as curvas acumuladas ocorre na ordenada 29. Este é o ponto em que as distribuições acumuladas de inadimplentes e adimplentes tem a maior separação, e portanto, este é o ponto em que a estatística KS é calculada.

Concluindo, um bom modelo deverá gerar valores preditos acertados, que diferenciem os consumidores conforme a chance de se tornarem inadimplentes**.** Repare que o modelo perfeito separaria *totalmente* os Adimplentes dos Inadimplentes em algum ponto do eixo da ordenada, ou seja, a estatística KS seria de 100%.

3.3.2 - Medida da Curva ROC

A curva ROC[6] (Receiver Operating Characteristic) ou Diagrama de Lorentz[7] também é bastante utilizada para medir a eficiência de um modelo de previsão. O mesmo conceito de taxa de acerto da estatística KS também é válido na Curva ROC. Para o cálculo desta medida é necessário apresentar as seguintes definições:

Sensitividade da predição: proporção de "inadimplentes classificados como adimplentes".

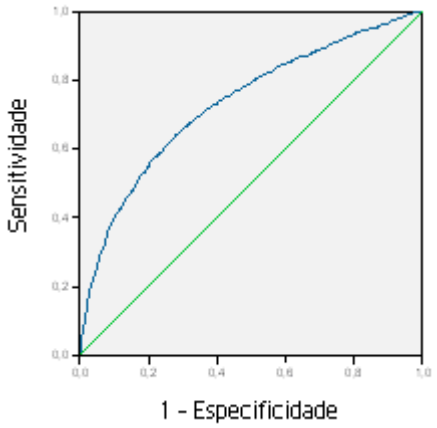
Especificidade da predição: proporção de "adimplentes bem classificados".

[6] Segundo TAPE(2001), a denominação ROC – Receiver Operating Characteristic se originou da utilização desta técnica na 2ª Guerra Mundial para fins de análise de imagens de radar.

[7] A denominação Diagrama de Lorenz ocorre em função da semelhança com a Curva de Lorenz, desenvolvida por Max O. Lorenz para descrever a desigualdade social (DRISLANE e PARKINSON, 2001)

A partir da determinação da especificidade e da sensitividade esperada em cada valor predito do modelo, é possível construir o Gráfico 2, em que a Sensitividade da predição é observado no eixo das abcissas e o índice (1-Especificidade da predição) é observado no eixo das ordenadas para cada um dos valores preditos possíveis do modelo.

Gráfico 2: Exemplo de Curva ROC



Observe que a partir do conhecimento do erro (sensitividade) e do acerto (especificidade) esperado em cada valor predito, e adicionando-se a informação do ganho médio do acerto e do custo médio do erro, é possível determinar a lucratividade esperada da população classificada em cada valor predito.

A partir da determinação da lucratividade e do prejuízo esperado em cada valor predito é possível impedir a aceitação de créditos a consumidores que sejam classificados em valores preditos com expectativa de prejuízo. Um bom modelo tem os resultados financeiros crescentes conforme o valor predito e desta forma é possível estabelecer um valor predito abaixo do qual os resultados financeiros esperados são negativos e acima do qual os resultados financeiros esperados são positivos. Este é o chamado de Ponto de Corte, que permite maximizar a lucratividade esperada de uma carteira a partir da rejeição de crédito a consumidores que geram expectativa de resultados financeiros negativos. Portanto, cada ponto da curva ROC corresponde a um candidato à ponto de corte.

A estatística é obtida a partir do cálculo da área abaixo da curva ROC (linha azul), que pode variar de 0 a 1. Quanto melhor o poder de discriminação do modelo, mais

próxima de 1 será a área abaixo da curva ROC. A diagonal representa uma curva ROC=0,50; que não tem poder discriminatório, visto que a Sensitividade e a Especificidade é igual em todos os pontos de valor predito, ou seja, o % de adimplentes classificados como adimplentes é sempre igual ao % de inadimplentes classificados como adimplentes, independentemente do ponto estudado.

4. DESCRIÇÃO DOS DADOS

O modelo de ***behaviour score*** proposto ***busca prever no momento da concessão do crédito*** qual é ***a probabilidade de inadimplência do consumidor***. É importante notar que a principal aplicação deste tipo de modelo ocorre em empresas que precisam decidir a respeito da concessão de algum tipo de crédito a clientes, e que tem disponibilidade do comportamento passado do consumidor na própria empresa, ou no mercado, através de um ***bureau*** de crédito. Este fenômeno é muito comum em instituições financeiras, pois usualmente o correntista (ou possuidor de cartão de crédito) solicita outros tipos de financiamento às empresas da qual é cliente, tais como o crédito pessoal, crédito para compra de veículos, CDC's ou outros.

Neste estudo, a amostra foi selecionada a partir de um bureau de crédito, que é consultado pelas instituições financeiras ou redes de varejo no momento da decisão de concessão do crédito, ou seja, ***todos os indivíduos da amostra estão solicitando algum tipo de crédito no momento da compra.*** A captação desta informação é possível porque as empresas concedentes buscam observar neste momento, se o referido cliente possui no mercado algum tipo de dívida vencida e não paga. Usualmente o consumidor não recebe o empréstimo solicitado caso seja apontado como inadimplente por alguma outra empresa do mercado, e neste caso o indivíduo é retirado da amostra, pois na verdade não é necessário conhecer a probabilidade de inadimplência, uma vez que o indivíduo já é inadimplente. Por isso, ***a medição do desempenho comparativo dos 2 modelos propostos foi realizado apenas entre consumidores sem nenhum apontamento de dívida no momento da compra.***

O procedimento de seleção dos indivíduos da amostra em um modelo para previsão da inadimplência é usualmente executado através da seleção de consumidores que realizam compra em uma certa data, denominada ***Data de Referência.*** A partir desta data, calcula-se o ***desempenho***, que é a variável endogêna/dependente Y_i que se deseja prever

(Inadimplência=1 e Adimplência=0) utilizando apenas informações de débitos ocorridos após a Data de Referência.

É intuitivo que o problema do viés de seleção descrito por AVERY (1977) e usualmente apontado em estudos desta natureza está reduzido nesta amostra devido à sua origem, pois apesar de não ter sido testado empiricamente, supõe-se que créditos rejeitados em algumas instituições possam ser obtidos em outras, de forma que a informação de inadimplência será obtida nesta situação. Por outro lado, calcula-se as variáveis exógenas/independentes do consumidor a partir do ***Histórico*** anterior à Data de Referência correspondente.

4.1 – Tratamento do Efeito Sazonal

Empiricamente, observa-se que o fenômeno inadimplência tem um efeito sazonal anual que pode ser justificado a partir do comportamento do consumidor, que se endivida em época de acúmulo de obrigações (Jan/Fev) e quita dívidas em época de final de ano, em que a economia lhe propícia maiores rendimentos. Devido a isso, uma amostra de consumidores selecionada em uma única data (ou mês) poderia conter vícios sazonais que tornam inviável a utilização permanente do modelo em uma instituição financeira. Por outro lado, ***controlar o efeito sazonal no modelo demandaria seleção de consumidores com compra em no mínimo 24 meses diferentes, de forma a controlar a sazonalidade da inadimplência com um mínimo de rigor estatístico.***

Para tratar este problema, e com o objetivo de simplificar o modelo, o mais usual é compor a amostra de forma balanceada de indivíduos que compraram em cada um dos meses do ano, ou seja, seleciona-se aleatoriamente e de forma proporcional à representatividade do mês. Por exemplo, para a ***Data de Referência Inicial*** em Janeiro seleciona-se de forma proporciona às vendas deste mês, uma quantidade de indivíduos que compraram neste mês, e assim sucessivamente para todos os meses do ano. Como são selecionados inúmeros indivíduos com compra em cada uma das diferentes Datas de Referência (que pode ser em qualquer um dos 12 meses do intervalo de tempo de seleção), espera-se que o efeito sazonal esteja considerado, mas não controlado no modelo final.

4.2 – Seleção da Amostra

A amostra foi composta através da seleção aleatória de 20.080 indivíduos adimplentes que estavam sofrendo algum tipo de avaliação de crédito em alguma instituição financeira ou de varejo, logo estavam comprando a crédito na respectiva Data de Referência Inicial.

Portanto, o modelo resultante desta amostra deverá predizer qual a probabilidade de inadimplência dado que o consumidor está buscando crédito e dado que não está inadimplente no mercado no momento da compra, situação freqüente nas operações de crédito ao consumidor. Lembre-se que este modelo é verossímil ao momento da compra, e é justamente neste momento que o decisor do crédito precisa da informação da chance de inadimplência para tomar sua decisão de aprovação ou rejeição da proposta.

Devido à sazonalidade discutida na seção anterior, as datas de referência iniciais foram escolhidas no período de 1 ano, ou seja, entre Nov/2003 e Out/2004.

Tabela 2: Distribuição da Amostra conforme Dt_Referência Inicial

ANOMES	#	%	%_AC
200311	1.736	8.65	8.65
200312	1.910	9.51	18.16
200401	1.399	6.97	25.12
200402	1.259	6.27	31.39
200403	1.493	7.44	38.83
200404	1.426	7.10	45.93
200405	1.603	7.98	53.91
200406	1.682	8.38	62.29
200407	1.674	8.34	70.63
200408	1.968	9.80	80.43
200409	2.110	10.51	90.94
200410	1.820	9.06	100.00
Total	20.080	100.00	

Primeiramente é importante ressaltar que nesta amostra não foi feito nenhum balanceamento da amostra, ou seja, composição da amostra com 50% de Inadimplentes e 50% de Adimplentes, apesar das considerações de ROSA (2000), que afirmou *"...quando se utiliza uma amostra proporcional à população, como a quantidade de clientes bons é sempre muito maior do que a de ruins, o modelo final acaba sendo excelente para discriminar os clientes bons, porém, ineficiente para discriminar os ruins"*.

Outro ponto a ser ressaltado é que o objetivo deste trabalho é avaliar o indivíduo com a perspectiva de dados em painel, e por isso, ***a partir da Data de Referência Inicial***

atribuída, observou-se o mesmo consumidor a cada 2 meses, avaliando outras Datas de Referência, a saber:

***Data de Referência* Inicial: Data de Referência 1** = Momento da Compra à Crédito

Data de Referência 2: Exatamente 2 meses após a Data de Referência 1

Data de Referência 3: Exatamente 4 meses após a Data de Referência 1

Data de Referência 4: Exatamente 6 meses após a Data de Referência 1

Data de Referência 5: Exatamente 8 meses após a Data de Referência 1

Repare que desta forma, a amostra conterá 100.400 registros, ou seja 20.080 indivíduos avaliados em 5 datas diferentes. Para maior compreensão, observe a tabela abaixo:

Tabela 3: Distribuição da Amostra em perspectiva de painel

	DT REFERÊNCIA					
ANOMES	**1**	**2**	**3**	**4**	**5**	**Total**
200311	1.736	0	0	0	0	1.736
200312	1.910	0	0	0	0	1.910
200401	1.399	1.736	0	0	0	3.135
200402	1.259	1.910	0	0	0	3.169
200403	1.493	1.399	1.736	0	0	4.628
200404	1.426	1.259	1.910	0	0	4.595
200405	1.603	1.493	1.399	1.736	0	6.231
200406	1.682	1.426	1.259	1.910	0	6.277
200407	1.674	1.603	1.493	1.399	1.736	7.905
200408	1.968	1.682	1.426	1.259	1.910	8.245
200409	2.110	1.674	1.603	1.493	1.399	8.279
200410	1.820	1.968	1.682	1.426	1.259	8.155
200411	0	2.110	1.674	1.603	1.493	6.880
200412	0	1.820	1.968	1.682	1.426	6.896
200501	0	0	2.110	1.674	1.603	5.387
200502	0	0	1.820	1.968	1.682	5.470
200503	0	0	0	2.110	1.674	3.784
200504	0	0	0	1.820	1.968	3.788
200505	0	0	0	0	2.110	2.110
200506	0	0	0	0	1.820	1.820
Total	*20.080*	*20.080*	*20.080*	*20.080*	*20.080*	*100.400*

4.3 – Definição de Desempenho

É fundamental ressaltar a importância da regra de elaboração da variável de desempenho na construção de um modelo de predição. Neste trabalho, a variável

dependente (ou variável resposta) é dicotômica (do tipo 1 / 0) e indicadora da inadimplência do consumidor no período seguinte. Existem muitas formas de definir a regra de elaboração desta variável para construção de um modelo de ***Behaviour Score*** verossímil à realidade e que seja consistente e eficaz na predição da chance de inadimplir. Por exemplo, é possível que não se deseje considerar como inadimplente um consumidor que teve um título em aberto por apenas 5 dias, pois provavelmente não seja justo atribuir como inadimplente um consumidor que simplesmente esqueceu de pagar uma dívida. Note que este detalhe pode mudar completamente os resultados finais do modelo.

Em um modelo logito tradicional pode-se avaliar apenas a situação do consumidor ao final do período de desempenho, mas em um modelo com tecnologia de dados em painel esta definição se torna mais complexa, pois o consumidor deve ser avaliado em sub-intervalos. Uma possibilidade é observar a existência de débitos não pagos em cada um dos períodos entre 2 datas de referência quaisquer. A principal conseqüência desta metodologia é que um consumidor que fosse incluído e excluído dentro do mesmo intervalo seria considerado inadimplente, mesmo que tenha quitado sua dívida rapidamente. Por isso existem argumentos razoáveis para considerar um caso como este como adimplente. Outra dúvida ocorreria sempre que um consumidor tivesse 2 dívidas em uma Data de Referência e apenas 1 dívida na Data de Referência seguinte, pois este consumidor poderia ser considerado adimplente porque pagou uma dívida ou inadimplente porque ainda tem uma dívida em aberto. Segundo THOMAS (2000), o ponto mais importante na definição do período de análise para definição do desempenho é a curva de inadimplência em função do tempo após a Data de Referência inicial, pois é fundamental observar após quanto tempo existe a estabilização do índice de inadimplência. Repare que definir horizontes de tempo muito curtos para classificação em adimplente/inadimplente pode afetar a eficácia do modelo à medida que uma parcela significativa de consumidores pode atingir o estado de inadimplência em horizontes de tempo mais longos. Por outro lado, horizontes muito longos de previsão deixam o modelo muito vulnerável a mudanças da distribuição das características da população.

Neste trabalho, a regra de construção da variável de desempenho GBI_t será:

$$GBI_t = \begin{cases} 0 \ \ se\ o\ consumidor\ for\ "Adimplente - Não\ Default"\ \ na\ Data\ de\ Referência(t+s) \\ 1 \ \ se\ o\ consumidor\ for\ "Inadimplente - Default"\ \ na\ Data\ de\ Referência(t+s) \end{cases}$$

Portanto, a metodologia de cálculo é observar na ***Data de Referência (t+s)*** se o consumidor tem algum débito não pago. Caso houver algum débito em aberto, GBI_t=1=M=Mau=Inadimplente, senão GBI_t=0=B=Bom=Adimplente. No desempenho do modelo em cross-section, o indivíduo será avaliado no 8º. Mês após Dt_Referência, portanto s=8. No modelo em painel, os consumidores serão avaliados a cada 2 meses, ou seja, s=2.

Desta forma, a amostra deste estudo apresenta as seguintes características:

- A variável de desempenho GBI_{t+s} está emparelhada com as variáveis independentes X_t.

- Uma particularidade do desempenho na amostra em painel é que os débitos em aberto na Data de Referência (t+s) poderão ser as mesmas dívidas já consideradas na Data de Referência (t), e que por ainda não terem sido pagas, continuam deixando o indivíduo com GBI_{t+s}=1. Outra possibilidade é que os débitos da Data de Referência (t) já tenham sido pagos, mas novas dívidas continuam deixando o consumidor com GBI_{t+s}=1. Repare que não existirá diferenciação entre estes dois casos.

Utilizando a regra de construção de desempenho descrita, o índice de inadimplência observado em cada Data de Referência pode ser avaliado na Tabela 4 apresentada abaixo.

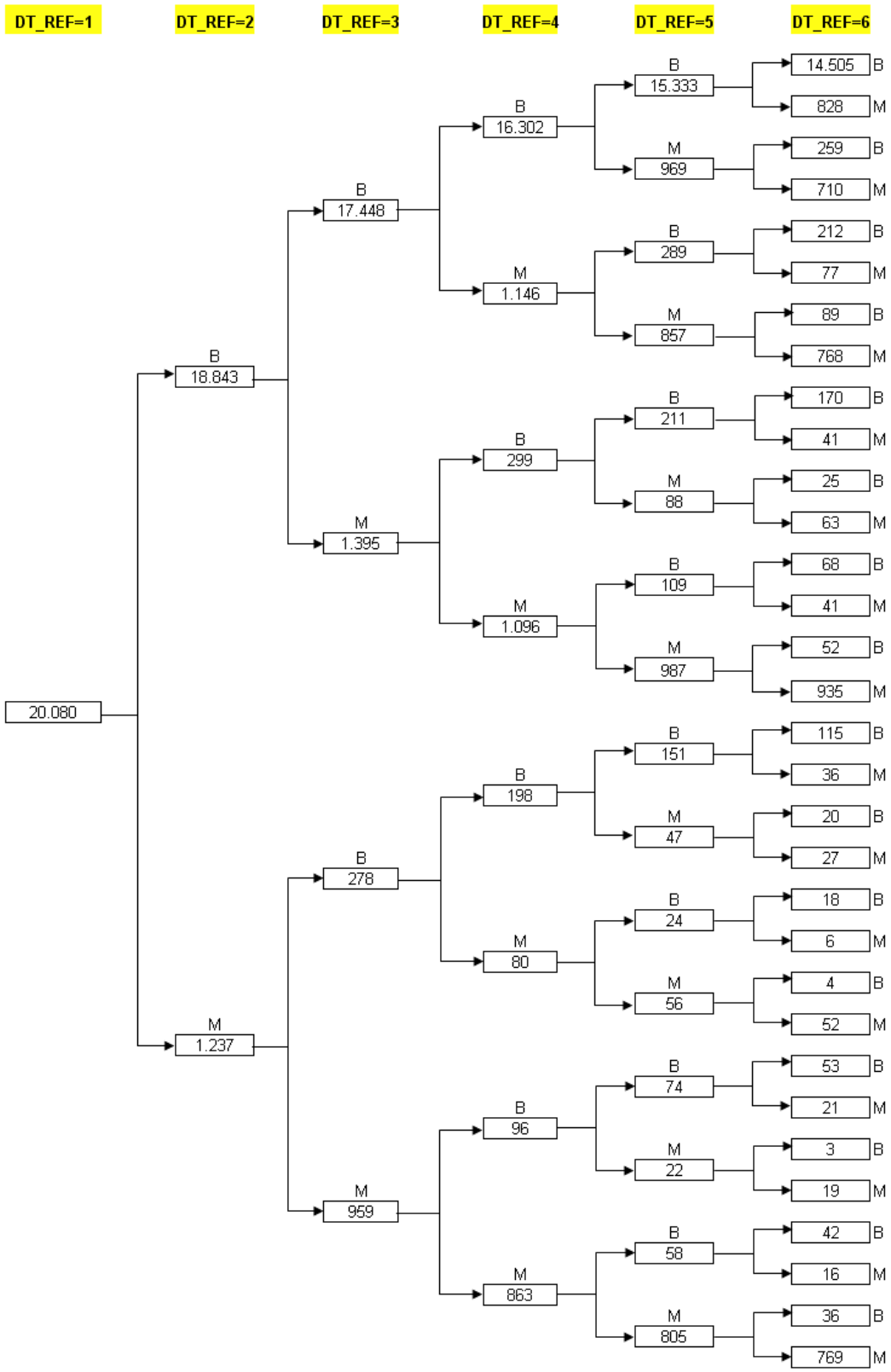
Tabela 4: Evolução do % Inadimplência na Amostra

DT_REFER "T"	# TOTAL	# INADIMPLENTES DT_REFER "T+s"	% INADIMPL
DT_REF_1	20.080	1.237	6,16%
DT_REF_2	20.080	2.354	11,72%
DT_REF_3	20.080	3.185	15,86%
DT_REF_4	20.080	3.831	19,08%

DT_REF_5	20.080	4.409	21,96%

Dado esta amostra, é importante lembrar que o Modelo ***Logito em Cross-Section será o Controle*** na comparação com o Modelo ***Logito utilizando Dados em Painel, que será o teste*** na predição dos 4.409 consumidores que estão classificados como Inadimplentes ("M") na Data de Referência 5. Observe no esquema abaixo, ***a evolução do estado de Inadimplência ("M") para a população da amostra em todas as Datas de Referência.***

Esquema 1 : Evolução da Amostra conforme Inadimplência



Poderia-se então, resumir o esquema anterior da seguinte forma:

Tabela 5: Descrição dos Consumidores que mudam de Estado

		Situação Final	
Evolução	**Qtde Inicial**	**BOM**	**MAU**
Não Se Alteram	15.274	14.505	769
Se Alteram	4.806	1.166	3.640
Total	*20.080*	*15.671*	*4.409*

A interpretação para a Tabela 5 pode ser descrita como:

1. 15.274 indivíduos não mudaram de estado durante todo o período.
   - 14.505 foram adimplentes durante todo o período.
   - 769 inadimpliram no 1ª. Período e permaneceram neste estado até o final.

   Observação: Repare que estes 15.274 indivíduos serão excluídos do cálculo do modelo logito utilizando dados em painel.

2. 4.806 indivíduos mudaram de estado pelo menos uma vez durante o período.
   - 3.640 estavam na última data de referência como Inadimplentes
   - 1.166 inadimpliram em algum período, mas estavam adimplentes ao final.

4.4 – Montagem das Variáveis Independentes

É importante ressaltar que as 20 variáveis preditoras (ou independentes) foram calculadas para todos os indivíduos nas datas de referência 1, 2, 3, 4 e 5. Elas são basicamente de 2 tipos: são 14 variáveis que descrevem o histórico de débitos do consumidor e 6 variáveis que descrevem o histórico de compras do consumidor.

Para o cálculo de cada variável independente foi considerada apenas a informação disponível até a respectiva DT_REFERÊNCIA, ou seja, informação "a posteriori" de cada data de referência não foi considerada no cálculo das variáveis preditoras Na Tabela 6, observe ***a descrição e o sinal esperado das variáveis independentes calculadas nas 5 Datas de Referência para os 20.080 consumidores.***

Tabela 6: Descritivo das Variáveis Independentes

Variável	Descrição	Categoria	Sinal Esperado	Justificativa
V1	Tt. de Segmentos com Débito no histórico	Débito	+	Se dívidas em muitos segmentos, então maior Prob de Inadimplir
V2	Idade do Ultimo Débito Pago	Débito	-	Se última Inadimplência há muito tempo, então menor chance Inadimplir
V3	Tt. Compras recentes	Compra	+	Se muitas compras recentes, então maior chance de inadimplir
V4	Tt_Compras no histórico	Compra	-	Se muitas compras no passado, então bom consumidor, e baixa chance de Inadimplir
V5	Maior Número de Compras no Mesmo Mês	Compra	+	Se muitas compras, então maior a chance de inadimplir
V6	Idade da 1a. Compra à Vista	Compra	-	Se compra com frequência, então primeira compra aconteceu a mais tempo, e menor risco de inadimplir
V7	Idade da 1a. Compra à Prazo	Compra	-	Se compra com frequência, então primeira compra aconteceu a mais tempo, e então menor o risco de Inadimplir
V8	Idade da Última Compra	Compra	+	Se compra pouco, então última compra há muito tempo, então maior chance de inadimplir
V9	Tt. Débitos à vista recentes não pagos	Débito	-	Se muitas dívidas recentes de compras à vista, então menor prob. Inadimplir
V10	Tt. Débitos à Vista pagos	Débito	-	Se muitas dívidas à vista na história, mas todas pagas, então menor Prob. Inadimplir
V11	Tempo Médio de Pagto da Dívida à Vista	Débito	+	Quanto maior tempo como devedor, maior Prob Inadimplir novamente
V12	Tt. Débitos à vista recentes pagos	Débito	-	Quanto mais pagamentos, menor a Prob de Inadimplir novamente
V13	Tt débitos à vista no 1o. mês c/ Inadimplência	Débito	+	Quanto mais débitos à vista no 1o. Mês c/ débito, maior Prob Inadimplir
V14	Tempo sem Inadimplência após pagto dívida à vista	Débito	-	Quanto maior o tempo sem Dívidas, menor a Prob de Inadimplir
V15	Tt débitos à vista não pagos	Débito	+	Quanto mais Cheques Inadimplentes, maior a Prob de Inadimplir
V16	Qtde de Reincidências após Pagto Débito à Vista	Débito	-	Quanto mais Reincidências, menor a Prob de Inadimplir novamente
V17	Tt. Débitos pagos	Débito	-	Quanto mais pagamentos, menor a Prob de Inadimplir novamente
V18	Vl dos Débitos incluídos e excluídos	Débito	-	Quanto maior o Vl dos Débitos recentes, maior a Prob Inadimplir
V19	Tt_Meios de Pagto c/ Histórico de Débito	Débito	+	Quanto maior o número tipos de débitos diferentes, maior a Prob Inadimplir
V20	Qtde débitos à vista e à prazo recentes	Débito	+	Quanto mais débitos recentes, maior Prob Inadimplir

4.5 - Correlação entre as Variáveis Independentes

Abaixo a Matriz de Correlação de Spearman das Variáveis Independentes para os 20.080 consumidores na Data de Referência Inicial. Observe que existem 10 pares de variáveis com correlação superior a 0.60, e que poderiam ser especialmente avaliados com relação à colinearidade. Como esta dissertação está estudando modelos de previsão, e retirar algumas variáveis do modelo significaria inevitavelmente perda do poder de predição, optou-se por não excluir variáveis que poderiam ser consideradas colineares, tais como as apontadas na Tabela 7 apresentada abaixo.

Tabela 7: Matriz de Correlação de Spearman das Variáveis Independentes

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
V1	1,00																			
V2	-0,74	1,00																		
V3	0,30	-0,24	1,00																	
V4	0,19	-0,17	0,55	1,00																
V5	0,32	-0,27	0,76	0,58	1,00															
V6	0,16	-0,20	0,04	0,13	0,07	1,00														
V7	0,05	-0,05	-0,07	0,01	-0,07	0,50	1,00													
V8	-0,05	0,02	-0,22	-0,15	-0,19	0,51	0,38	1,00												
V9	0,24	-0,20	0,13	0,14	0,16	0,05	0,02	-0,04	1,00											
V10	0,30	-0,23	0,14	0,13	0,16	0,06	0,02	-0,05	0,63	1,00										
V11	0,34	-0,27	0,17	0,08	0,17	0,01	-0,01	-0,02	0,27	0,29	1,00									
V12	0,29	-0,21	0,17	0,12	0,17	0,02	0,00	-0,04	0,47	0,58	0,50	1,00								
V13	0,19	-0,17	0,10	0,08	0,11	0,04	0,01	-0,03	0,27	0,29	0,41	0,39	1,00							
V14	0,01	0,00	0,00	0,02	0,01	-0,01	-0,01	-0,01	0,00	0,03	0,01	0,01	0,01	1,00						
V15	0,17	-0,13	0,13	0,07	0,12	0,01	-0,01	-0,02	0,16	0,08	0,47	0,23	0,31	0,15	1,00					
V16	0,31	-0,28	0,13	0,15	0,16	0,08	0,03	-0,06	0,60	0,65	0,30	0,38	0,30	-0,01	0,02	1,00				
V17	0,75	-0,62	0,26	0,21	0,28	0,15	0,05	-0,07	0,25	0,35	0,19	0,23	0,16	0,02	0,07	0,35	1,00			
V18	0,13	-0,11	0,08	0,09	0,07	0,03	0,01	-0,02	0,10	0,07	0,02	0,05	0,05	0,00	0,02	0,10	0,21	1,00		
V19	0,65	-0,78	0,25	0,21	0,28	0,19	0,06	-0,08	0,41	0,49	0,26	0,31	0,26	0,01	0,08	0,58	0,67	0,13	1,00	
V20	0,32	-0,27	0,17	0,17	0,18	0,06	0,02	-0,04	0,69	0,38	0,11	0,28	0,17	0,02	0,09	0,41	0,40	0,29	0,38	1,00

5. RESULTADOS

Nesta seção são apresentados os resultados empíricos da aplicação do modelo logito em ***cross-section*** e em painel na amostra de 20.080 consumidores, tornando possível comparar o efeito destas duas metodologias em um modelo de ***Behaviour Score*** construído para previsão de inadimplência.

Usualmente o modelo de regressão logito é aplicado em amostras do tipo ***cross-section*** avaliando a variável dependente em um horizonte de previsão (desempenho) de 6 meses, ou seja s=6. Entretanto, a aplicação da metodologia de dados em painel foi realizada observando a variação do comportamento do consumidor a cada 2 meses em 5 períodos. Logo, para comparação entre as técnicas foram construídas 2 variáveis dependentes para classificação dos 20.080 consumidores da Data de Referência Inicial, a saber:

GBI_2M: Esta variável dependente é construída observando o estado do consumidor (desempenho) 2 meses após a Data de Referência.

GBI_6M: Esta variável dependente é construída observando o estado do consumidor (desempenho) 6 meses após a Data de Referência.

5.1 – Modelo Completo

Nesta primeira comparação os modelos de predição foram obtidos utilizando todas as variáveis independentes, ou seja, tanto as variáveis de histórico de débitos quanto de compras.

5.1.1 – Análise Comparativa dos Coeficientes

Observe na Tabela 8 a comparação entre os coeficientes dos modelos logito em painel e em ***cross-section***. É importante lembrar que o modelo logito em ***cross-section*** foi desenvolvido prevendo as variáveis dependentes GBI_2M e GBI_6M, e que o modelo com efeitos fixos utilizou GBI_2M nas 5 datas de referência como variável dependente.

Na estimação do modelo logito com efeitos fixos foram excluídos 15.274 indivíduos que não mudaram o estado em nenhum dos 5 períodos estudados. Desta forma, foram considerados apenas 4.806 indivíduos na estimação, ou seja, 24.030 observações.

Tabela 8: Comparação de Coeficientes – Modelo com todas as Variáveis

Variável	Tipo Variável	Hipótese do Sinal do Coeficiente	CROSS-SECTION		PAINEL
			LOGITO GBI_2M	LOGITO GBI_6M	EF_FIXO
V1	Débito	+	0,146 3,9	0,214 7,3	-0,040 -0,9
V2	Débito	-	-0,004 -8,9	-0,004 -14,5	0,001 * 2,3
V3	Compra	+	0,043 4,5	0,090 10,7	-0,025 * -4,6
V4	Compra	-	-0,014 -2,5	-0,028 -4,6	0,184 * 15,5
V5	Compra	+	0,059 4,4	0,051 4,0	0,445 15,7
V6	Compra	-	-0,013 -2,1	-0,042 -12,4	0,033 * 4,7
V7	Compra	-	-0,014 -3,2	-0,005 -3,1	-0,018 -3,1
V8	Compra	+	-0,010 -0,6	0,021 3,5	0,009 1,4
V9	Débito	-	-0,063 -2,1	-0,016 -0,9	-0,055 -2,3
V10	Débito	-	-0,006 -0,5	-0,005 -0,6	-0,121 -4,6
V11	Débito	+	0,006 2,0	0,009 4,7	0,204 7,5
V12	Débito	-	-0,010 -1,0	0,003 0,4	-0,055 -3,0
V13	Débito	+	0,000 0,0	0,005 0,7	-0,020 -0,4
V14	Débito	-	0,013 1,3	0,000 0,1	0,035 * 2,1
V15	Débito	+	-0,002 -0,8	0,002 0,8	-0,005 -0,2
V16	Débito	-	0,075 1,8	0,081 2,6	0,640 * 8,8
V17	Débito	-	-0,014 -1,2	-0,010 -1,1	-0,089 -4,6
V18	Débito	-	0,000 -1,2	0,000 -1,1	0,000 0,2
V19	Débito	+	0,004 0,1	-0,084 -4,5	0,263 13,5
V20	Débito	+	0,164 5,6	0,118 5,4	0,123 10,4
Constante			-2,292 -14,3	-0,398 -3,8	
		Log Likelih.	-4.046,9	-9.059,2	-7.314,7
		N	20.080	20.080	24.030

* Coeficiente significativamente diferente do esperado
Abaixo dos coeficientes está a estatística t.
"N" refere-se ao Número de registros considerados no modelo.

Observe através da análise da Tabela 8 que *os coeficientes do Modelo de Efeitos Fixos apresentam resultados de sinal e/ou de significância muito diferentes dos Modelos em* ***cross-section*** *GBI_2M e GBI_6M, que tem resultados similares entre si.* Esta diferença está mais evidente nas variáveis V1, V3, V4, V5, V6, V10, V11, V12, V16, V17 e V19, e uma possível causa para esta diferença é a forte influência de multicolinearidade no modelo em painel.

5.1.2 – Comparação dos Resultados de Performance

As Tabelas 9 e 10 contém os testes de desempenho Kolmogorov-Smirnov e Curva ROC para a previsão atribuída à Data de Referência Inicial pelas 2 técnicas. O objetivo principal é observar ***se o modelo logito utilizando dados em painel*** gera melhores resultados de previsão no momento da compra (Data de Referência Inicial) ***que o modelo logito em cross-section***.

A justificativa para este procedimento é que o decisor de crédito precisa da chance de inadimplência do consumidor apenas neste momento, quando ocorre a decisão da venda à crédito.

Tabela 9: Comparação de KS do Modelo Completo

		DESEMPENHO		
	AVALIAÇÃO DE KS	PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,433	
	LOGITO - GBI6			0,409
Painel	LOGITO - Efeitos Fixos	0,349	0,310	0,215
	Perda/Ganho		-28,4%	-47,4%

Tabela 10: Comparação da Curva ROC do Modelo Completo

		DESEMPENHO		
	Avaliação Curva ROC	PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,784	
	LOGITO - GBI6			0,766
Painel	LOGITO - Efeitos Fixos	0,730	0,703	0,633
	Perda/Ganho		-10,4%	-17,4%

É possível concluir a partir da análise das tabelas acima que há perda significativa de desempenho de previsão quando se utiliza o modelo logito em painel comparativamente ao modelo tradicional logito em ***cross-section***. *Esta perda de poder de previsão ocorre tanto*

quando se avalia a estatística KS quanto a Curva ROC, e a perda é mais acentuada quando se deseja prever GBI_6M.

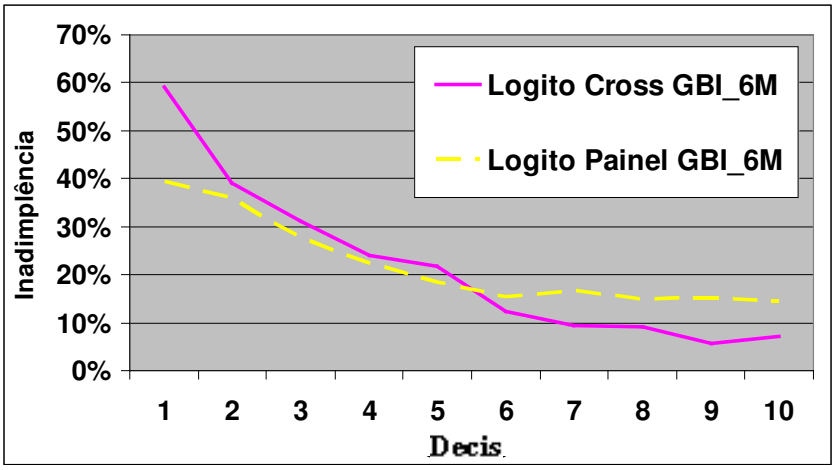
5.1.3 – Comparativo da Distribuição Populacional e do Desempenho

Observe na tabela abaixo a análise descritiva de distribuição da amostra conforme valor predito, que compara o desempenho de ambos os modelos quanto à inadimplência efetiva em cada grupo. Como as 2 técnicas geraram modelos com distribuições muito diferentes, definiu-se 10 decis para cada um dos 2 modelos, com 2.008 indivíduos em cada. A medida de inadimplência efetiva por decis em cada modelo está apresentada no Gráfico 3.

Tabela 11: Distribuição da Amostra por Decis dos Logitos Cross (GBI6) e Painel

		Decis do Logito em Painel										
		1	2	3	4	5	6	7	8	9	10	#
Decis do Logito Cross Section 6M	1	722	655	275	137	94	47	34	14	17	13	2.008
	2	388	336	303	236	208	126	127	82	104	98	2.008
	3	253	290	284	280	207	231	188	128	81	66	2.008
	4	167	197	223	210	218	256	235	188	261	53	2.008
	5	110	122	154	154	151	136	127	122	191	741	2.008
	6	108	143	197	198	211	221	268	207	192	263	2.008
	7	78	67	145	159	214	245	280	269	371	180	2.008
	8	78	89	162	236	172	192	206	315	300	258	2.008
	9	45	56	129	196	298	382	296	315	178	113	2.008
	10	59	53	136	202	235	172	247	368	313	223	2.008
	#	2.008	2.008	2.008	2.008	2.008	2.008	2.008	2.008	2.008	2.008	20.080

Gráfico3: Comparação 1 da Inadimplência por Decis entre Logito Cross-Section e Painel



As conclusões das tabelas 9 e 10 são confirmadas pelo Gráfico 3, pois um maior percentual de consumidores inadimplentes foram classificados no 1º. Decil no modelo Logito em ***cross-section***. *Lembre-se que este modelo tem consumidores Inadimplentes e Não-Inadimplentes, e que o modelo será tão melhor quanto menor o Valor Predito dos consumidores inadimplentes.*

Uma característica desejável de ambos os modelos é a declividade monotônica da inadimplência. Observe ainda a curva mais tênue de diminuição da inadimplência por decil no Logito em Painel, indicando que este modelo também tem poder preditivo, apesar de menor.

Resumindo, observa-se na Seção 5.1.1 que os coeficientes dos modelos logito em ***cross-section*** e em painel são significativamente diferentes, indicando a presença dos efeitos fixos advindo de variáveis omitidas, que pode por exemplo ser o efeito advindo de variáveis como honestidade e caráter, que foram descritas na Seção 3.1. Entretanto, o desempenho do modelo logito em ***cross-section*** é melhor que o desempenho do modelo logito em painel.

É importante ressaltar que o principal objetivo dos modelos logito com efeitos fixos é obter a contribuição marginal das variáveis independentes na previsão da inadimplência, excluindo-se os efeitos fixos e invariantes do indivíduo.

Na tentativa de encontrar as razões do pior desempenho do modelo logito em painel, será observado nas próximas seções se o tipo de variável independente (histórico de débitos ou de compras) contribui de forma diferente na eficiência de um modelo logito em painel e em ***cross-section***.

5.2 – Modelo utilizando somente Variáveis de Histórico de Débito

Nesta seção os mesmos modelos de predição da Seção 5.1 foram obtidos, entretanto aqui apenas as 14 variáveis que descrevem o histórico de débitos do consumidor foram utilizadas como variáveis independentes na determinação dos modelos de predição.

5.2.1 – Análise Comparativa dos Coeficientes

Tabela 12: Comparação de Coeficientes – Modelo com variáveis de débito

Variável	Tipo Variável	Hipótese do Sinal do Coeficiente	CROSS-SECTION		PAINEL
			LOGITO GBI_2M	LOGITO GBI_6M	EF_FIXO
V1	Débito	+	0,213 5,8	0,277 9,9	-0,009 -0,2
V2	Débito	-	-0,004 -8,5	-0,004 -14,1	0,001 * 2,2
V9	Débito	-	-0,075 -2,4	-0,034 -2,0	-0,021 -0,9
V10	Débito	-	-0,007 -0,6	-0,003 -0,4	-0,160 -6,3
V11	Débito	+	0,009 3,4	0,013 7,1	0,237 8,8
V12	Débito	-	0,000 0,0	0,012 1,7	-0,086 -4,9
V13	Débito	+	-0,003 -0,2	0,000 0,0	0,020 0,5
V14	Débito	-	0,009 1,0	-0,002 -0,3	0,038 * 2,2
V15	Débito	+	0,000 0,0	0,004 1,8	-0,062 * -2,4
V16	Débito	-	0,055 1,3	0,054 1,8	0,756 * 10,7
V17	Débito	-	-0,013 -1,0	-0,011 -1,3	-0,040 -2,2
V18	Débito	-	0,000 -1,1	0,000 -1,1	0,000 -0,3
V19	Débito	+	0,012 0,5	-0,092 -5,2	0,333 17,8
V20	Débito	+	0,181 6,0	0,145 6,7	0,145 12,5
Constante			-2,546 -19,7	-0,806 -9,5	
		Log Likelih.	-4.221,7	-9.603,7	-7.838,7
		N	20.080	20.080	24.030

* Coeficiente significativamente diferente do esperado
Abaixo dos coeficientes está a estatística t.
"N" refere-se ao Número de registros considerados no modelo.

Observe através da análise da Tabela12 que *os coeficientes do Modelo Logito em Painel apresentam resultados de sinal e/ou de significância muito diferentes dos Modelos Logito GBI_2M e GBI_6M, que tem resultados similares entre si.* A diferença está mais evidente nas variáveis V1, V10, V11, V12, V14, V15, V16, V17 e V19, e uma possível causa para esta diferença é a forte influência de multicolinearidade no modelo em painel.

5.2.2 – Comparação dos Resultados de Performance

Nesta seção, as mesmas considerações da seção 5.1.2 devem ser avaliadas. Para facilitar a análise comparativa, as tabelas abaixo incorporam os resultados já descritos nas tabelas 9 e 10, de forma a tornar possível a visualização do desempenho alcançado com os modelos em ***cross-section*** e em painel utilizando todas as variáveis disponíveis e apenas as variáveis de débito.

Tabela 13: Comparação do KS entre Modelo Completo e Modelo c/ Var. de Débito

		TODAS AS VARIÁVEIS			APENAS VAR. DE DÉBITO		
		DESEMPENHO			DESEMPENHO		
	AVALIAÇÃO DE KS	PAINEL	CROSS SECTION		PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,433			0,367	
	LOGITO - GBI6			0,409			0,338
Painel	LOGITO - Efeitos Fixos	0,349	0,310	0,215	0,515	0,333	0,255
	Perda/Ganho		-28,4%	-47,4%		-9,3%	-24,6%

Tabela 14: Comparação Curva ROC entre Mod. Completo e Modelo c/ Var. de Débito

		TODAS AS VARIÁVEIS			APENAS VAR. DE DÉBITO		
		DESEMPENHO			DESEMPENHO		
	CURVA ROC	PAINEL	CROSS SECTION		PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,784			0,725	
	LOGITO - GBI6			0,766			0,697
Painel	LOGITO - Efeitos Fixos	0,730	0,703	0,633	0,771	0,667	0,617
	Perda/Ganho		-10,4%	-17,4%		-8,1%	-11,5%

Através das Tabelas 13 e 14 observa-se que há *perda de desempenho no modelo logito em **cross-section** quando se utiliza apenas as variáveis de histórico de débito.* Nos modelos com GBI_2Meses, observa-se uma queda de desempenho de 0,433 para 0,367 no KS (-15%) e de 0,784 para 0,725 na Curva ROC (-8%) quando se compara o modelo logito em ***cross-section*** utilizando todas as variáveis independentes disponíveis com o mesmo modelo utilizando apenas variáveis independentes de histórico de débito. O mesmo comportamento se observa no modelo logito ***cross-section*** GBI_6M.

No modelo logito em painel pode ser melhor não utilizar todas as variáveis disponíveis, pois se observou situações em que o desempenho foi melhor quando se utilizou apenas as variáveis independentes de histórico de débito na construção do modelo logito em painel. Na variável dependente GBI_2M o desempenho do modelo em painel que utilizou todas as variáveis disponíveis foi 7,4% pior em KS (compara-se 0,310 com 0,333) e 5,1% melhor na Curva ROC (0,703 para 0,667). O mesmo padrão se observa nos modelos com GBI_6Meses. Este resultado indica que o tipo de variável independente considerada no modelo pode influenciar positivamente a favor do modelo logito em painel.

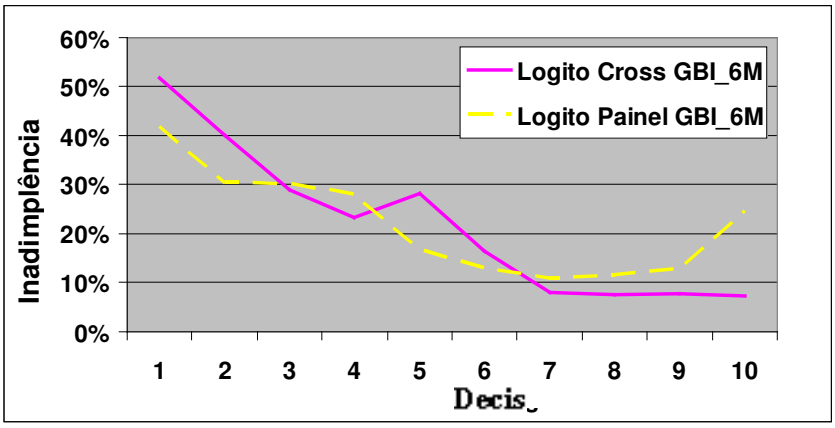
*Entretanto, utilizando apenas as variáveis independentes com histórico de débito, observou-se que o modelo logito em **cross-section** é melhor que o modelo logito em painel,* pois há uma perda significativa de desempenho da técnica de logito com efeitos fixos comparativamente à mesma técnica em ***cross-section.*** Repare que há uma variação de 0,367 para 0,333 (-9,3%) no KS e 0,725 para 0,667 (-8,1%) na Curva ROC. Nos modelos com GBI_6 Meses a perda é ainda maior.

*Até este ponto conclui-se que como o principal objetivo é alcançar o melhor desempenho de previsão, unanimemente (tanto por KS quanto por Curva ROC) o melhor modelo para prever GBI_2M e GBI_6M é o modelo logito em **cross-section**.*

5.2.3 – Comparativo da Distribuição Populacional e do Desempenho

Nesta seção também foi realizada uma análise descritiva comparando a distribuição da amostra conforme valor predito e suas respectivas inadimplências esperadas por decis em ambos os modelos (***cross-section*** e em painel).

Gráfico4: Comparação 2 de Inadimplência por Decis entre Logito Cross-Section e Painel



Os dois modelos descritos no Gráfico 4 utilizam *apenas as variáveis independentes de histórico de débito, e observa-se que a expectativa de inadimplência no modelo logito em* ***cross-section*** *apresenta um padrão decrescente, e que esta tendência é mais tênue no modelo logito em painel. Isto indica que o logito em* ***cross-section*** *qualifica pior os consumidores que de fato se tornam inadimplentes, e portanto, tem melhor eficiência que o modelo logito em painel. É interessante notar a inversão da tendência decrescente de inadimplência nos Decis 9º. e 10º. para o modelo logito Painel GBI_6M.*

5.3 – Modelo utilizando somente Variáveis de Histórico de Compras

Complementarmente à seção anterior, que utilizou apenas as variáveis independentes que descrevem o histórico de débitos, nesta seção os modelos foram calculados considerando-se apenas as variáveis independentes que se referem ao histórico de compras do consumidor antes de cada Data de Referência. Neste modelo também foram excluídos 15.274 indivíduos que não mudaram o estado em nenhum dos 5 períodos estudados.

5.3.1 – Análise Comparativa dos Coeficientes

Tabela 15: Comparação de Coeficientes – Modelo com variáveis de compra

Variável	Tipo Variável	Hipótese do Sinal do Coeficiente	CROSS-SECTION		PAINEL
			LOGITO GBI_2M	LOGITO GBI_6M	EFEITO FIXO
V3	Compra	+	0,057 5,8	0,109 13,3	-0,064 * -12,7
V4	Compra	-	-0,015 -2,0	-0,033 -5,8	0,265 * 23,1
V5	Compra	+	0,083 5,6	0,098 7,5	0,634 21,7
V6	Compra	-	0,008 1,5	-0,025 -8,0	0,038 * 5,5
V7	Compra	-	-0,011 -3,5	-0,004 -2,9	-0,020 -3,6
V8	Compra	+	-0,033 -1,9	0,011 1,6	0,031 4,3
Constante			-3,166 -28,9	-1,495 -23,4	
		Log Likelih.	-4.315,8	-9.694,8	-7.961,7
		N	20.080	20.080	24.030

* Coeficiente significativamente diferente do esperado
Abaixo dos coeficientes está a estatística t.
"N" refere-se ao Número de registros considerados no modelo.

Observa-se que os coeficientes do modelo logito em painel são significativamente diferentes dos coeficientes do modelo logito em ***cross-section***, principalmente as variáveis V3, V4, V5 e V6. Uma possível causa para esta diferença é a forte influência de multicolinearidade no modelo em painel.

5.3.2 – Comparação dos Resultados de Performance

Observe que os modelos em ***cross-section*** e em painel foram avaliados quanto ao desempenho no Teste de Kolmogorov-Smirnov (Tabela 16) e na Curva ROC (Tabela 17).

Tabela 16: Comparação de KS do Modelo de Histórico de Compras

		DESEMPENHO		
	AVALIAÇÃO DE KS	PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,362	
	LOGITO - GBI6			0,314
Painel	LOGITO - Efeitos Fixos	0,266	0,275	0,171
	Perda/Ganho		-12,4%	-45,5%

Tabela 17: Comparação de Curva ROC do Modelo de Histórico de Compras

		DESEMPENHO		
	Avaliação Curva ROC	PAINEL	CROSS SECTION	
Amostra	Técnica	GBI_2M	GBI_2M	GBI_6M
Cross	LOGITO - GBI2		0,734	
	LOGITO - GBI6			0,702
Painel	LOGITO - Efeitos Fixos	0,561	0,677	0,604
	Perda/Ganho		-7,7%	-14,0%

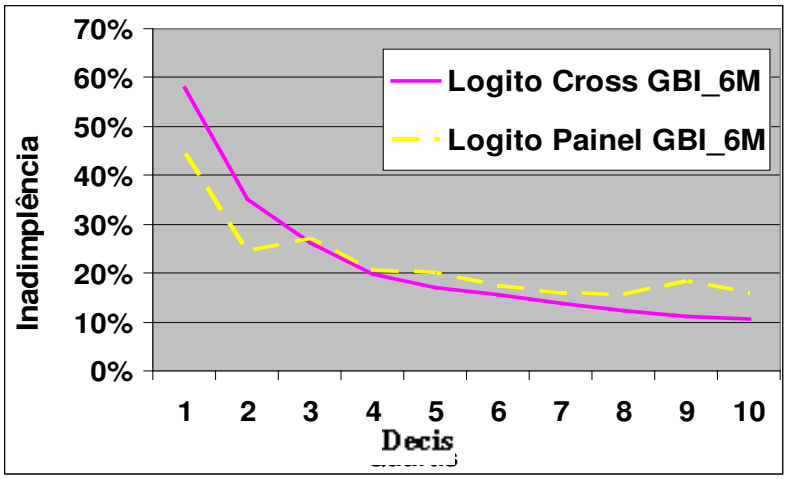
O modelo logito em painel desta seção apresentou o pior resultado de KS e Curva ROC entre todos os modelos estudados. No modelo logito em ***cross-section,*** *o valor predito do modelo que utiliza apenas variáveis independentes de histórico de compras apresentou desempenho de KS e Curva ROC semelhante ao valor predito do modelo que utiliza apenas variáveis independentes de histórico de débito.*

A diferença percentual de desempenho entre o modelo logito em ***cross-section*** *e em painel foi maior nos modelos que utilizam apenas variáveis independentes de compra, indicando que este tipo de variável causa mais fortemente a diferença de desempenho dos 2 modelos. Portanto, as variáveis independentes de compra tem menor poder de explicar os efeitos fixos propostos neste estudo.*

5.3.3 – Comparativo da Distribuição Populacional e do Desempenho

Esta análise é complementar à seção anterior, pois através do gráfico apresentado abaixo é possível observar como cada um dos modelos propostos classificou a amostra, assim como se de fato a inadimplência ocorrida depois apresentou um padrão de monotonicidade ao longo do valor previsto.

Gráfico5: Comparação 3 de Inadimplência por Decis entre Logito Cross-Section e Painel



Conclui-se do Gráfico 5 acima que o Modelo Logito em Painel utilizando apenas variáveis de histórico de compras apresenta menor poder de diferenciação dos inadimplentes que o Modelo Logito em ***cross-section***.

A curva de inadimplência do Modelo Logito em Painel também apresenta padrão decrescente de inadimplência, porém, com alguma instabilidade desta tendência no 3º. Decil.

Repare que no modelo Logito em Painel desta seção também foi observado, mesmo que de forma mais tênue, um pico no 9º. Decil, que já tinha sido observado no Gráfico4 (Descrição da Inadimplência por Decis no modelo logito em painel que utiliza apenas variáveis independentes de débito). Portanto, a hipótese de que as variáveis de histórico de débito geraram este pico é provável.

6. CONCLUSÃO

A teoria de Modelos Logito com Efeitos Fixos apresentada na Seção 3.2.2 é adequada para estimar de forma consistente o efeito de cada variável independente sobre a probabilidade de inadimplência, excluindo o efeito fixo decorrente de variáveis independentes não-observadas. Como o resultado das estimações dos coeficientes na modelagem em ***cross-section*** e em painel foram significativamente diferentes, é possível afirmar que estes efeitos fixos de fato existem no fenômeno inadimplência. Entretanto na amostra de estudo foi observado que o desempenho do modelo logito em ***cross-section*** foi melhor que o modelo logito em painel nas medidas de desempenho KS e Curva ROC, o que é contra-intuitivo, pois o modelo em painel contém mais informação, e portanto deveria produzir um modelo de ***Behavior Score*** mais eficaz.

É possível que a exclusão do efeito individual de cada consumidor na obtenção dos valores preditos tenha causado a perda de poder de predição, pois a estimação final não incorporou o coeficiente linear, contrariamente ao resultado do modelo logito em ***cross-section***, que estima este parâmetro no modelo final. Por isso, há que se estudar com mais profundidade o efeito causado pelos diferentes tratamentos amostrais dos modelos logito em ***cross-section*** e em painel, e neste caso, a pior eficiência do modelo logito em painel poderia ser decorrente do fato de que estes efeitos invariantes e inerentes ao indivíduo existem e são diferentes o suficiente entre os indivíduos da amostra para gerarem perda de poder de predição. Para comprovar esta afirmação é importante obter uma metodologia para recuperar os $C_{i's}$ e incorporá-los no valor predito, obtendo então uma fórmula mais eficiente de prever a probabilidade de inadimplência.

Apesar da hipótese apresentada acima ser plausível, ela ainda não é definitiva, e outros trabalhos estudando este fenômeno também podem realizar outras modificações em algumas definições do modelo, testando variações que podem indicar empiricamente outras explicações para este resultado de eficiência da aplicação da metodologia de dados em painel para previsão da inadimplência. Destaca-se entre elas aumentar o tamanho da amostra de estudo, aumentar o número de períodos "T" observados, utilizar uma amostra de estudo balanceada (mesmo número de adimplentes e inadimplentes) e ter outras variáveis independentes.

7. BIBLIOGRAFIA

ADYA, M., COLOPPY, F. "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation", Journal of Forecasting, 17, 1998.

ALBRIGHT, H.T. "Construction of a polynomial classifier for consumer loan applications using genetic algorithms", Working Paper, Department of Systems Engineering, University of Virginia, 1994.

ALMEIDA, Fernando C., Dumontier P. "O uso de Redes Neurais em Avaliação de Riscos de Inadimplência", Revista de Administração, São Paulo, V.31, no. 1, p. 52-63, 1996.

ALTMAN, E.I. "Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy". Journal of Finance 23, p. 589-609, 1968.

ALTMAN, E.I.; Marco Giancarlo, Varetto Franco. "Corporate Distress Diagnosis: Comparisons using Linear Discriminant Analysis and Neural Networks (The Italian Experience)", Journal of Banking and Finance18, pgs. 505-529, 1994.

ANDERSEN, E.B., "Asymptotic Properties of Conditional Maximum Likelihood Estimators", Journal of the Royal Statistical Society, Series B32:283-301, 1970.

ANDERSEN, E.B., " Conditional Inference and Models for Measuring", Kobenharn: Mentalhygiejnish Farlag, 1973.

AVERY, R.B. "Credit scoring models with discriminant analysis and truncated samples", 1977.

BANASIK, J; CROOK,J.N; THOMAS,L.C. "Not if but when borrowers default", J. Operational Research Society 50, pp. 1.185 – 1.190 , 1999.

BARTH, Nelson L. – "Métodos de Discriminação entre Grupos – Aplicação ao Problema da Concessão do Crédito" – Dissertação de Mestrado, FGV, São Paulo, 2002.

BAUER Jr., Richard J. "Genetic Algorithms and Investment Strategies". John Wiley & Sons, USA, 1994.

BERRY, M.J.A.; LINOFF, Gordon. "Data Mining Techniques", John Wiley & Sons, USA, 1997

BIGUS, Joseph P. "Data Mining With Neural Networks", McGraw-Hill, USA, 1996.

BOYLE, M.; CROOK, J.N.; HAMILTON, R.; THOMAS, L.C. "Methods for credit scoring applied to slow payers", Credit Scoring and Credit Control, Oxford University Press, Oxford, p. 75-90, 1992.

BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A.; STONE, C.J. "Classification and Regression Trees" Wadsworth, Belmont, CA, 1984.

CABRAL,C.; PINHEIRO,A.C. "Mercado de Crédito no Brasil: O papel do Judiciário e de outras Instituições. Ensaios BNDES, no. 9, Rio de Janeiro, Dez, 1998.

CAPON N. "Credit Scoring Systems: a critical analysis". Journal of Marketing, 46, 82-91,1982.

CHAMBERLAIN, G. "Analysis of Covariance with Qualitative Data". Review of Economic Studies 47: 225-38, 1980.

DESAI,V.S; CONVAY,D.G.;CROOK, J.N.;OVERSTREET,G.A. "Credit scoring models in the credit union environment using neural networks and genetic algorithms", IMA J. Mathematics applied in Business and Industry 8, p. 323-346, 1997.

DRISLANE, R., PARKINSON, G. "On line Dictionary of the Social Sciences". Canada´s Open University. Endereço Eletrônico: http://datadump.icaap.org/cgi-bin/glossary/SocialDict/SocialDict?term=LORENZ%20CURVE, 2001.

DURAND, D. "Risk elements in consumer instalment financing", National Bureau of Economic Research, New York, 1941.

EISENBEIS R.A. "Pitfalls in the application of discriminat Analysis in business, finance and economics". Journal of Finance 32, 975-900,1977

EISENBEIS R.A. "Problems in applying discriminat analysis in credit scoring models" Journal of Banking and Finance 2, p. 205-219 – 1978.

FISHER, R.A. "The use of multiple measurements in taxonomic problems" – Annals of Eugenics 7 , p.179-188, 1936.

FOGARTY T.C., IRESON N.S. "Evolving Bayesian classifiers for credit control", IMA J. Mathematics Applied in Business and Industry 5, 65-76, 1993.

FREED,N.; GLOVER,F. (a) "A linear programming approach to the discriminant problem", Decision Sciences 12, p. 68-74, 1981.

FREED,N.; GLOVER,F. (b) "Simple but powerful goal programming formulations for the discriminant problem", European Journal of Operational Research 7, 44-60, 1981.

GEHRLEIN, W.V.; WAGNER, B.J. "A two-stage least cost credit scoring model. Annals of Operations Research 74, p.159-171, 1997.

GLUENSTEIN, John M.L. "Optimal Use of Statistical Techniques in Model Building". In Credit Risk Modelling", Fitzroy Dearborn Publishers, USA, 1998.

GOLDENBERG, David E. “Genetic Algorithms in Search, Optimization & Machine Learning”, Addison-Wesley, EUA, 1989.

HAIR Jr., J.F. et al. “Multivariate Data Analysis”. 5th Edition, Prentice Hall, USA, 1998.

HAND, D.J. – Discrimination and Classification, Wiley, Chichester, 1981.

HARDY, W.E.; ADRIAN, J.L. “A linear programming alternative to discriminant analysis, Abribus1, p. 285-292, 1985.

HENLEY, W.E. “Statistical aspects of Credit Scoring”. Ph.D. thesis, Open University, 1995.

HOSMER, D.W.J; LEMESHOW,S “Applied Logistic Regression” – Second Edition – John Wiley & Series, 2000.

HSIAO, C. “Benefits and Limitations of Panel Data”, Econometric Reviews, 4 – 121-174, 1985.

HSIAO, C. “Analysis of Panel Data”, Cambridge University Press - 2a_edicao. – Caps.7-8, 2003.

JOHNSON, R.A.; WICHERN, D.W. “Applied Multivariate Statistical Analysis”. Prentice-Hall, 1982

MANGASARIAN, O.L. – “Linear and non-linear separation of patterns by linear programming”. Operations Research 13, p. 444-452, 1965.

MCFADDEN, D. “Conditional Logit Analysis of Qualitative Choice Behavior”, Frontiers in Econometrics, editado por P. Zarembka, pgs. 105-42. New York; Academic Press, 1974.

MARQUES, L. D. “Modelos Dinâmicos com Dados em Painel: revisão de literatura”,2000. www.fep.up.pt/investigacao/workingpapers/wp100.PDF às 15:51h do dia 01/Dez/2006.

MARTELL, T.F.; FITTS, R.L. “A quadratic discriminant analysis of bank credit card user characteristics. Journal of Economics and Business 33, p.153-159, 1981.

NARAIN, B. “Survival Analysis and the credit granting decision”, in Credit Scoring and Credit Control, Oxford University Press, Oxford, pp. 109-122, 1992.

NATH R., JACKSON W.M., JONES T.W. “A comparison of the classical and the linear programming approaches to the classification problem in discriminant analysis”. J. Statistical Computation and Simulation 41, p.73-93, 1992.

PATERSON, Dan W. “Artificial Neural Networks, Theory and Applications”, Prentice-Hall Inc., USA, 1996.

ROSA, P.T.M. “Modelos de Credit Scoring: Regressão Logística, Chaid e Real”, Dissertação de Mestrado, IME-USP, São Paulo, 2000.

ROSENBERG, Eric, GLEIT, Alan, “Quantitative Methods in Credit Management: a survey”. Operations Research 42, p. 589-613, 1994.

SCARPEL, R.A. e MILIONI, A.Z. “Utilização Conjunta de Modelagem Econométrica e Otimização em Decisões de Concessão de Crédito” http://www.scielo.br/pdf/pope/v22n1/a04v22n1.pdf às 15:52h do dia 01/Dez/2006.

SHARMA, Subhash. “Applied Multivariate Techniques”. Wiley, 1996.

SILVA, Edson R. – Diagnósticos em Regressão – Resumos 12º. SINAPE – Associação Brasileira de Estatística, 1996.

SRINIVASAN,V.; KIM,Y.H. “Credit granting: a comparative analysis of classification procedures”. J. Of Finance, 42, p. 665-683, 1987.

STEPANOVA,M.;THOMAS,L.C. “PHAB Scores; Proportional Hazards analysis Behavioural Scores”, in J. Operational Research Soc. 52, p. 1.007 – 1.016, 2001

TAPE, Thomas G. “Interpreting Diagnostic Tests”. University of Nebraska Medical Center. http://www.gim.unmc.edu/dxtestes/default.htm, 2001.

THOMAS, Lyn C. “A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers”. International Journal of Forecasting 16, p.149-172, 2000.

VARETTO, Franco. “Genetic Algorithms Appications in the Analysis of Insolvency Risk”, Journal of Banking and Finance, 22, 1998.

VASCONCELLOS, M.S. “Proposta de Método para Análise de Concessões de Crédito a Pessoas Físicas” Dissertação de Mestrado, FEA/USP, 2002.

YOBAS, M.B; CROOK, J.N.; ROSS, P. “Credit scoring using neural and evolutionary techniques’, Working Paper 97/2, Credit Research Centre, University of Edinburgh, 1997.

WIGINTON, J.C. “A note on the comparison of logit and discriminant models of consumer credit behavior”. Journal on Finances and Quantitative Analysis 15, p. 757-768, 1980.

WOOLDRIDGE, J.M. – Econometric Analysis of Cross Section and Panel Data – MIT Press, 2002

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo