

## **3 Metodologia**

### **3.1. Universo, amostra e fontes de dados**

Com o intuito de enriquecer o conhecimento sobre as recompras no mercado brasileiro, o estudo procura analisar todos os programas de recompra de ações anunciados pelas empresas listadas na BOVESPA. A amostra inicial de empresas é composta por todas as 363 empresas que têm ações listadas na BOVESPA na data de 30 de dezembro de 2013. Em complemento, foram incluídas empresas “deslistadas” no período de 2007 a 2013. Alguns critérios de seleção foram adotados com o intuito de garantir resultados mais robustos. O primeiro critério, de exclusão, está relacionado ao valor de mercado calculado pela multiplicação do número total de ações com base na cotação de fechamento do dia 30 de dezembro de 2013. Na análise, foram excluídas todas as empresas com valor de mercado inferior a R\$ 10 milhões.

Também foram excluídas todas as empresas que (i) não são domiciliadas no Brasil e que, portanto, não precisam obedecer a todas as regras exigidas pela legislação brasileira; (ii) por qualquer motivo, tenham deixado de apresentar os relatórios financeiros durante dois trimestres consecutivos. Não foram aplicados critérios restritivos tradicionais na literatura em relação à liquidez, pois um dos objetivos secundários da análise é observar o comportamento dessas ações.

Para a análise dos programas de recompra de ações, foi criada uma base de dados com informações sobre todos os programas de recompras anunciados pelas empresas selecionadas através dos critérios mencionados anteriormente. Assim como explicado, não serão incluídas na amostra qualquer tipo de recompra implementada via oferta pública de aquisição, sendo apenas objeto deste estudo as recompras implementadas via aquisição direta no mercado. Também será excluída da amostra qualquer recompra que: (i) foi lançada com o objetivo de fechar o capital

da empresa ou reduzir o *free-float* abaixo de 10%; (ii) qualquer recompra que por algum motivo não foi direcionada a todas as ações em circulação; (iii) anúncios que ocorreram no mesmo dia de outro evento corporativo significativo que possa distorcer o seu impacto isolado.

A introdução da Instrução 480 da CVM em 7 de dezembro de 2009 introduziu a obrigatoriedade da apresentação de dados detalhados sobre os programas de recompra anunciados nos Formulários de Referência. Estes formulários anuais, que precisam ser entregues ao órgão regulador no prazo máximo de 120 dias após o encerramento do ano fiscal, contém dados mais precisos e detalhados sobre os programas de recompra de ações anunciados no Brasil, tais como datas das aprovações nas reuniões dos conselhos de administração das empresas, número de ações que são objeto do programa, prazo para execução, dados referentes ao número total de ações que foram efetivamente recompradas ao longo do programa e o preço médio de execução das recompras. Apesar da entrega dos Formulários de Referência só ter se tornado uma obrigação a partir de 2010 (tendo como referência o ano fiscal de 2009), estes formulários apresentam dados referentes aos três últimos anos fiscais, o que possibilita acesso a informações padronizadas desde 2007.

Ainda que os Formulários de Referência forneçam dados sobre a execução dos programas de recompra, não existe a exigência de apresentar dados sobre a evolução da execução dos programas. Ou seja, as empresas fornecem apenas os dados referentes ao resultado final de cada programa de recompra de ações anunciado. Para permitir uma análise mais detalhada de como os programas de recompra são executados ao longo de sua existência, a base de dados também foi complementada com informações retiradas das notas explicativas dos formulários trimestrais, com base nos resultados das empresas que compõem a amostra. Dessa fonte foram extraídas informações sobre as recompras e alterações das ações em tesourarias ocorridas a cada trimestre, permitindo assim análise mais detalhada da execução individual de cada programa de recompra.

Em razão da restrição temporal imposta pela disponibilidade de dados padronizados nos Formulários de Referência e da periodicidade das notas explicativas dos resultados das empresas, o horizonte temporal de análise deste estudo é restrito a 28 trimestres ao longo de sete anos, tendo início no primeiro

trimestre de 2007 e se estendendo até o quarto trimestre de 2013, último período de dados publicados.

Dada a inexistência de uma base de dados que consolide todas as informações disponíveis sobre os programas de recompra de ações para cada empresa que compõe a base de dados previamente definida, foi feito um processamento individual do Formulário de Referência arquivado anualmente para possibilitar o levantamento de todos os programas de recompras anunciados ao longo do horizonte temporal de análise do estudo.

Das 265 empresas analisadas, 115 anunciaram ao menos um programa de recompra de ações ao longo dos últimos sete anos. No período de análise, foram identificados um total de 503 programas de recompra que atendem aos critérios de seleção discutidos anteriormente, com uma duração média de 289 dias.

Para conseguir analisar o comportamento dos *insiders* nos períodos que antecedem e sucedem os anúncios dos programas de recompra de ações, também foi formada uma base de dados com informações relativas às operações de compra e venda de ações por parte dos executivos e demais *insiders*. Assim como no caso das informações referentes aos programas de recompra de ações, não existe atualmente uma base de dados detalhada de monitoramento do comportamento de compra e venda de ações abrangendo todo o período de análise desse estudo. Dessa forma, foram levantados os formulários mensais exigidos pela Instrução 358 da CVM de cada empresa, nos quais são apresentadas as negociações de ações realizadas pelos *insiders* durante o mês anterior.

Para efeito de padronização, serão considerados *insiders* na amostra os agentes que sejam classificados pela CVM como controladores, membros da diretoria, do Conselho de Administração e demais comitês internos. Investidores individuais ou em grupo que tenham grande relevância acionária, mas que não sejam classificados como *insiders* pelos critérios da CVM, não farão parte da amostra em análise.

Os dados até então mencionados foram extraídos diretamente dos relatórios arquivados na CVM, em especial o Formulário de Referência, o Formulário 358, e as notas explicativas dos formulários ITRs e DFPs. Os dados referentes às classificações setoriais e à segmentação dos mercados de listagem das ações na bolsa de valores foram obtidos da BOVESPA. Os demais dados financeiros históricos e as séries de preços e de volume das ações foram extraídos da base de

dados da *Bloomberg*. As descrições de todas as variáveis utilizadas nos modelos estão apresentadas no Apêndice B.

### **3.2. Método de Estudos de Eventos**

Com o intuito de analisar o comportamento de determinadas variáveis durante intervalos de tempo anteriores e posteriores às datas de anúncio dos programas de recompra, foi utilizada a metodologia de Estudos de Eventos.

Assim como apresentado por Kothari e Warner (2007) e Schilling, Diehl, e Macagnan (2011), o uso da metodologia de Estudos de Eventos tem se tornado cada vez mais presente na literatura. Inicialmente utilizado como forma de testar os níveis de eficiência do mercado nos trabalhos de Brown e Warner (1980) e Fama *et al* (1969), o uso de Estudos de Eventos também se popularizou nas demais áreas da Administração e Contabilidade, buscando avaliar o impacto nos preços das ações de eventos corporativos, como anúncios de dividendos e lucro, e nas novas regulações e decisões jurídicas.

No campo das Finanças, a utilização desta metodologia tem como foco avaliar o comportamento dos preços de determinados ativos em dias próximos ao evento (Brown e Warner, 1980). Apesar de não ter sido o primeiro a trabalhar com Estudo de Eventos no campo das Finanças, Fama *et al* (1969) é um dos grandes responsáveis por sua disseminação. Em seu estudo seminal, os autores implementam a metodologia por meio da análise do processo de ajuste dos preços das ações às novas informações sobre dividendos, implícitas em um desdobramento de ações. A conclusão é a de que o mercado acionário seria eficiente à medida que os preços das ações se ajustam rapidamente às novas informações para testar as hipóteses de eficiência de mercado desenvolvidas durante a década de 60. De maneira similar, Ball e Brown (1968) analisaram a resposta do mercado à divulgação de demonstrações contábeis pela variação dos preços de suas ações, identificando uma resposta antecipada do mercado à divulgação desses relatórios.

O formato estatístico básico dos estudos de eventos não se alterou significativamente ao longo do tempo (Kothari e Warner, 2007). O foco ainda é baseado no retorno anormal médio e retorno anormal médio acumulado. As

principais mudanças referem-se à periodicidade dos dados utilizados. Enquanto os trabalhos do início da segunda metade do século XX utilizavam periodicidade mensal, os mais recentes empregaram dados diários e intradiários, o que permite uma medição mais precisa dos eventos e dos respectivos retornos anormais.

Outra mudança importante que vem sendo observada é a utilização de métodos de estimação mais sofisticados, em particular os relacionados a eventos longos em decorrência de novos achados no final da década de 90, sobre as propriedades estatísticas de retornos de longo prazo de ações e novas metodologias de precificação de ativos.

De maneira geral, um estudo de eventos utiliza um modelo de geração de retorno de ações para determinar um retorno esperado para certo ativo. Independentemente do modelo utilizado, o retorno esperado calculado deve refletir o comportamento considerado normal para a evolução de preços do ativo e é tido como o retorno que deveria ser obtido na ausência de ocorrência de qualquer outro fator ou evento. Tendo como base esse retorno normal, pode ser realizada uma comparação com o retorno efetivo observado pelo mesmo ativo durante um determinado período e, através da diferença dos dois retornos, obter o que é considerado um retorno anormal.

Campbell, Lo e MacKinlay (1997) sintetizam grande parte dos procedimentos adotados pelos trabalhos até então desenvolvidos e descrevem as principais etapas para operacionalização de um Estudo de Eventos na Figura 1.

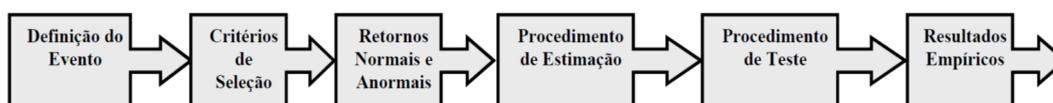


Figura 1 - Etapas de um Estudo de Eventos

Fonte: Camargos e Barbosa (2003)

A primeira etapa é a definição do evento e estimação da janela de análise. Nesta etapa, devem ser definidos especificamente quais são os eventos, bem como o período no qual os efeitos ocorrem, que também serão observados. E deve ser ainda estipulado o período no qual os efeitos dos eventos serão observados. Campbell, Lo e MacKinlay (1997) explicam que o tamanho da janela de análise depende diretamente da definição do evento e do período esperado de reação, podendo representar intervalos de segundos, minutos, horas, dias, meses ou até mesmo anos.

Apesar dessa determinação estar sujeita à subjetividade do pesquisador, o tamanho da janela de análise deve estar diretamente relacionado à duração do evento. Adicionalmente, para gerar robustez ao modelo e eliminar problemas de *framing*, pode-se utilizar duas ou mais janelas de análise para efetuar uma comparação dos resultados. De forma geral, em virtude da periodicidade diária e intradiária de cotações dos ativos, a maior parte dos estudos de Finanças adota um intervalo de dias ou meses.

Após a etapa inicial de definição do evento, o próximo passo é a definição dos critérios de seleção. Nesta fase, o pesquisador especifica exatamente quais empresas e ativos serão incluídos na base de dados, para além do tamanho do período de análise e quais as classes de ações que serão analisadas. O objetivo é gerar uma amostra padronizada e evitar a eventual presença de ativos que apresentem características indesejadas, como baixa liquidez e poucos dados, o que poderia enviesar a análise e comprometer os resultados.

A terceira etapa no processo é a definição do modelo e cálculo dos retornos normais e anormais para todos os ativos da amostra. Esta parte do processo é seguida pela estimação do modelo dentro da janela temporal. Assim como explicado por MacKinlay (1997), para cada ativo  $i$  o retorno observado no tempo  $t$  é dado por:

$$R_{it} = E(R_{it}|X_t) + AR_{it} \quad (1)$$

no qual  $E(R_{it} | X_t)$  é o valor estimado do retorno normal  $R_{it}$ , condicionado ao grupo de informações disponíveis no tempo  $t$   $X_t$ , e  $AR_{it}$  é o componente anormal ou inesperado do retorno. Ou seja, assim como apresentado na equação (2),  $AR_{it}$  é a diferença entre o retorno observado e o retorno normal esperado.

$$AR_{it} = R_{it} - E(R_{it}|X_t) \quad (2)$$

Dessa forma, assim como observado por Kothari e Warner (2007), o retorno anormal pode ser entendido como uma medida direta da variação de riqueza do proprietário do ativo devido à ocorrência do evento. Antes de calcular o retorno anormal, contudo, é necessário especificar um modelo e calcular os retornos normais esperados. Assim como descrito na seção 2.1, apesar da existência de

diversos modelos, grande parte dos trabalhos em Finanças utiliza os modelos de mercado de Sharpe (1964) e Lintner (1965) e o de fatores de Fama e French (1992).

A quarta etapa do processo proposto por Campbell, Lo e MacKinlay (1997) está relacionada com a estimação dos parâmetros do modelo que será utilizado para o cálculo dos retornos normais esperados. Trata-se da definição da janela temporal para a estimação dos parâmetros do modelo. Esta janela deve ser definida de acordo com o modelo escolhido e encerrada antes da ocorrência do evento para evitar que as estimativas sejam poluídas com os efeitos gerados pelo evento.

Assim como apresentado em MacKinlay (1997), a janela temporal de um estudo de eventos pode ser representada pela Figura 2, na qual  $\tau = 0$  é a data de ocorrência do evento, o intervalo entre  $\tau = T_1 + 1$  e  $\tau = T_2$  representa a janela do evento, e o período entre  $\tau = T_0 + 1$  e  $\tau = T_1$  representa a janela de estimação explicada acima.

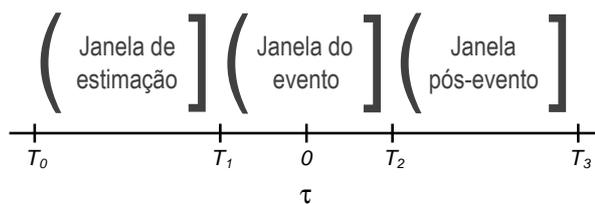


Figura 2 - Janela temporal de um estudo de eventos  
Fonte: MacKinlay (1997)

Uma vez estimados os parâmetros necessários dos modelos de retorno, os retornos anormais podem ser calculados com base na equação (2), permitindo a execução dos testes para verificar as hipóteses propostas. Assim como exposto em Kothari e Warner (2007), dado o objetivo de verificar se a ocorrência do evento está, na média, associada a uma variação no valor do ativo, deve-se estabelecer uma hipótese nula que testa se o retorno anormal médio no tempo  $t$ ,  $\overline{AR}_t$  é igual a zero. Assim, analisa-se o primeiro momento da distribuição do retorno, que será dado por:

$$\overline{AR}_t = \frac{1}{N} \sum_{i=1}^N \widehat{AR}_{it} \quad (3)$$

no qual  $\overline{AR}_t$  é o retorno *cross-sectional* anormal médio no período  $t$ ,  $N$  é o número de firmas dentro da amostra e  $\widehat{AR}_{it}$  é o retorno anormal estimado para firma  $i$  no período  $t$ . Além do primeiro momento, também é necessário avaliar o segundo

momento da distribuição dos retornos anormais, afinal a hipótese nula do teste busca mostrar que a ocorrência do evento não produz impacto no comportamento do retorno (nem em sua média, nem na variância). Com isso, assim como expresso em MacKinlay (1997), sob a hipótese nula, a distribuição dos retornos anormais de uma dada observação dentro da janela do evento deve ser:

$$\widehat{AR}_{it} \sim N(0, \sigma^2(\widehat{AR}_{it})) \quad (4)$$

Para que seja possível fazer inferências sobre o evento observado é necessário que os retornos anormais sejam acumulados durante o período da janela de tempo pós-evento apresentada na Figura 2. Ou seja, no intervalo de tempo entre  $\tau_1 = T_2 + 1$  e  $\tau_2 = T_3$ . Esta análise deve ser realizada através do cálculo do retorno anormal médio acumulado no período,  $\overline{CAR}$ . Trata-se de uma acumulação de valores com duas dimensões, sendo a primeira relacionada aos diferentes ativos da amostra e a segunda ao intervalo de tempo (MacKinlay, 1997).

A primeira dimensão desse processo foi abordada na equação (3) ao calcular o  $\overline{AR}_t$ . Similarmente, a segunda dimensão pode ser individualmente acumulada através do processo apresentado pela equação (5), ao calcular o retorno acumulado para cada ativo  $i$  no intervalo de tempo entre  $\tau_1$  e  $\tau_2$ . O acúmulo total, incluindo ambas as dimensões, é apresentado na equação (6).

$$\widehat{CAR}_i(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \widehat{AR}_{it} \quad (5)$$

$$\overline{CAR}(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} \overline{AR}_t \quad (6)$$

Inferências sobre o impacto do evento podem ser obtidas por meio de testes para verificar se a hipótese nula dos retornos anormais médio,  $\overline{CAR}$ , tiver valor igual a zero, de tal forma que:

$$\overline{CAR}(\tau_1, \tau_2) \sim N(0, var(\overline{CAR}(\tau_1, \tau_2))) \quad (7)$$

Para o  $\overline{CAR}$  apresentado na equação (6), a hipótese nula será rejeitada se a estatística do teste exceder o valor crítico correspondente a 10%, 5% ou 1% de significância. Assim como apresentado por MacKinlay (1997) e Kothari e Warner (2007), a estatística padrão do teste é dada pela equação (8).

$$\theta_1 = \frac{\overline{CAR}(\tau_1, \tau_2)}{\text{var}(\overline{CAR}(\tau_1, \tau_2))^{1/2}} \sim N(0,1) \quad (8)$$

### 3.2.1. Modelos de Retornos Normais e Anormais

Como já explicitado, o primeiro passo para medir o retorno anormal de um ativo é estimar qual deveria ser o seu retorno normal. Desde meados da década de 60, diversos modelos de precificação de ativos foram propostos e a literatura sobre o tema ainda não convergiu para uma escolha uniforme. Entre os principais modelos de precificação de ativos estão o *Capital Asset Pricing Model* (CAPM), introduzido por Sharpe (1964) e Lintner (1965), e o modelo de três fatores elaborado por Fama e French (1992).

Cabe ressaltar que, apesar de todos os modelos terem por objetivo estimar o retorno esperado dos ativos, cada um possui suas premissas e vieses que, por consequência, podem afetar diretamente a precisão das estimativas subsequentes do retorno anormal, gerando possíveis vieses indesejados (Kothari e Warner, 2007).

O modelo do CAPM propõe a utilização da equação (9), na qual  $E(R_i)$  é o retorno esperado do ativo  $i$ ,  $R_f$  é a taxa livre de risco,  $\beta_{im}$  é o coeficiente de sensibilidade do retorno do ativo  $i$  em relação ao retorno do mercado e  $E(R_m)$  é o retorno esperado do mercado.

$$E(R_i) = R_f + \beta_{im}(E(R_m) - R_f) \quad (9)$$

Por se tratar de um modelo com um único fator explicativo, prêmio de risco do mercado, o CAPM torna-se muito dependente da escolha do indicador do mercado que será utilizado. No Brasil, a dificuldade em determinar um indicador

que contemple as premissas que atendam ao modelo de Sharpe e Litner torna a sua implementação suscetível à formação de vieses indesejados.

### **3.2.2. Limitações Metodológicas**

A metodologia de Estudos de Eventos apresenta algumas limitações importantes. A primeira delas está relacionada com a validade de uma das premissas básicas do estudo de eventos. Devido à ineficiência dos mercados financeiros, os preços observados podem não refletir completamente e imediatamente toda a informação disponível. Além disso, é importante ressaltar que eventos podem ter sido antecipados em alguns casos ou podem estar coexistindo sem que sejam percebidos. Essa coexistência pode causar uma dúvida sobre qual dos eventos está efetivamente impactando os preços e, conseqüentemente, levar a vieses nos retornos esperados dos ativos de tal forma que os retornos anormais podem não ser apenas resultado do evento analisado.

Outra limitação do modelo está na subjetividade na qual as janelas de observação e de estimação do evento são determinadas. Conseguir precisar onde o evento ocorre e o período no qual o seu efeito será analisado é crucial para resultados eficientes e conclusivos.

A escolha do modelo de estimação dos retornos normais tem grande impacto na magnitude e significância dos retornos anormais. Se os retornos esperados forem calculados impropriamente e outros fatores não forem propriamente controlados, a qualidade do resultado do estudo de eventos pode ser comprometida.

O estudo de eventos se propõe a generalizar os resultados obtidos com base em uma amostra limitada tanto no horizonte temporal como no número de ativos e eventos. Portanto, a escolha dos ativos, eventos e do espaço temporal que compõe a amostra pode afetar o resultado do estudo e o seu poder de generalização.

Por fim, é importante ter em mente o cuidado com possíveis *clusters* temporais que podem causar um problema de dependência *cross-sectional* nos dados.

### 3.3. Método de Diferenças-em-Diferenças

Adicionalmente à referida metodologia de Estudo de Eventos, a verificação da existência de retornos anormais em torno das datas de anúncio dos programas de recompra de ações também foi feita por meio da utilização do método conhecido como diferenças-em-diferenças (*Diff-in-Diff*), que se baseia em princípios econométricos mais rigorosos e, conseqüentemente, confere maior robustez às análises propostas.

Introduzido por Heckman *et al* (1998), o método procura evidenciar o efeito de um determinado evento externo sobre dois grupos amostrais com características similares durante um determinado período de tempo. Para isso, aplica uma dupla subtração que apresenta a diferença das médias da variável de resultado entre o período anterior e posterior ao evento para dois diferentes grupos equivalentes: de controle e tratamento, para os quais seja possível assumir que todas as características são as mesmas, com exceção da exposição à variável a ser estudada ou testada pelo modelo.

O produto final da dupla subtração proposta pelo método é conhecido como estimador DD (Meyer, 1995), que como já sublinhado representa o efeito de um determinado evento exógeno na comparação entre grupos de controle e de tratamento. Na ausência de um tratamento, a diferença entre os grupos de controle (B) e o de tratamento (A) deveria ser constante ao longo do tempo. O estimador DD compara a diferença entre as médias dos grupos nos períodos anterior e posterior ao tratamento. Quanto maior a diferença no período pós-tratamento, maior o efeito do tratamento, assim como observado na Figura 3.

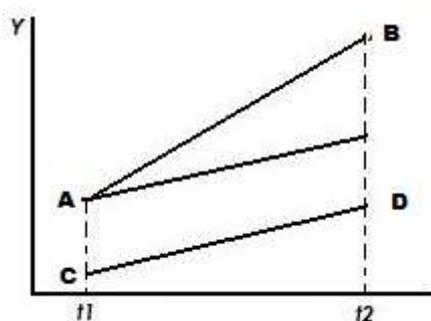


Figura 3 - Diferenças-em-Diferenças

|            | Antes | Depois | Diferenças  |
|------------|-------|--------|-------------|
| Tratamento | A     | B      | A-B         |
| Controle   | C     | D      | C-D         |
| Diferenças | A-C   | B-D    | (C-D)-(A-B) |

O método selecionado controla as possibilidades de apresentar vieses a partir do pressuposto de que a diferença não observada entre a média do produto dos grupos de controle e de tratamento não varia com o tempo. Sendo assim, o viés de seleção, caso presente, se anula, fazendo com que o estimador DD não seja enviesado (Heckman *et al*, 1998).

No objeto de estudo desta tese, para comprovar os efeitos dos anúncios dos programas de recompra nos retornos anormais apresentados por diferentes grupos de empresas com características similares, foi utilizada uma análise de regressão com variáveis anterior e posterior ao anúncio dos programas de recompra. A interação entre as duas nos grupos selecionados apresenta a variável que nos dá o efeito do evento em análise.

Para tal finalidade, os grupos foram selecionados de acordo com classificação setorial da BOVESPA, adicionalmente à análise da descrição de atividades de cada uma delas. Foram utilizados como grupo de tratamento as empresas que anunciaram programas de recompra na janela de tempo estudadas e como grupos de controle as companhias com características similares às aquelas que anunciaram programas de recompra e que não dispunham de programas próprios de recompra ativos no mesmo período de análise.

### **3.4. Modelo Tobit**

Uma das principais variáveis sujeita à análise ao longo do desenvolvimento da tese é o percentual de execução do programa de recompras de ação. Como trata-se de uma variável que, por definição, não terá valores abaixo de 0%, é necessário utilizar um modelo de estimação que acomode este tipo de característica censurada.

A metodologia utilizada foi o modelo Tobit, desenvolvido por Tobin (1998), cujo nome faz referência à nomenclatura e às semelhanças aos modelos Probit e Logit. Também conhecido como regressão Tobit, o modelo utiliza dados censurados para uma determinada variável dependente, isto é, disponíveis somente em determinado intervalo de valores.

O modelo Tobit é um processo de estimação baseado na máxima verossimilhança que trata as informações censuradas de forma distinta das

restantes. Matematicamente, o método pode ser representado pela seguinte equação:

$$\begin{aligned}
 y^* &= \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon \\
 y_i^* &= \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i, \\
 y_i &= 0 \quad \text{if } y_i^* \leq 0, \\
 y_i &= y_i^* \quad \text{if } y_i^* > 0.
 \end{aligned} \tag{10}$$

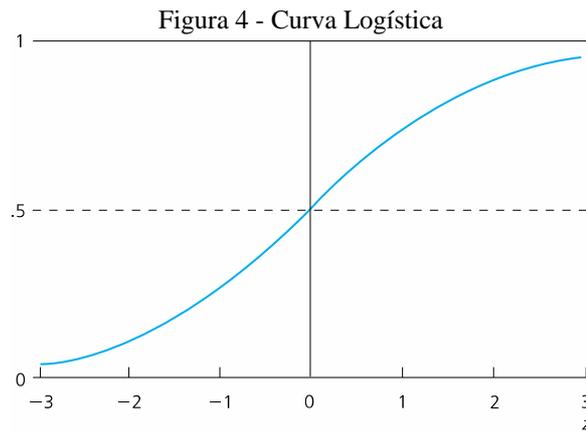
A sua função de log-verossimilhança é dada por:

$$\ln L = \sum_{y_i > 0} -\frac{1}{2} \left[ \log(2\pi) + \ln \sigma^2 + \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{\sigma^2} \right] + \sum_{y_i = 0} \ln \left[ 1 - \Phi \left( \frac{\mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right) \right] \tag{11}$$

### 3.5. Modelo Logístico

Para testar as hipóteses relacionadas e trabalhar com as variáveis das diferentes categorias nas quais as empresas serão classificadas em relação à execução ou não dos programas de recompra autorizados, é necessária a utilização do modelo logístico. Originada na década de 1960, a regressão logística busca descrever a relação entre uma variável dependente binária e as variáveis explicativas (Fávero *et al.*, 2009). Segundo Hair *et al* (1999), a regressão logística é particularmente útil, pois não pressupõe a existência de homogeneidade de variância e normalidade dos resíduos, características necessárias às regressões múltiplas.

O logístico é assim chamado porque utiliza uma distribuição cumulativa logística para transformar a regressão de modo que as probabilidades seguem a forma de S, estando limitado assintoticamente entre 0 e 1, como apresentado na Figura 4 (Brooks, 2008).



Fonte: Wooldridge (2013)

A função logística é expressa pela equação (12) e assume valores entre 0 e 1. Assim, a funcionalidade do modelo não é justificada apenas pela possibilidade de prever a ocorrência do evento analisado, mas também por sua capacidade de apresentar a probabilidade de ocorrência (Fávero *et al.*, 2009).

$$f(z) = \frac{1}{1 + e^{-z}} \quad (12)$$

Onde  $z$  é dado por:

$$z = \ln\left(\frac{p}{1-p}\right) = \alpha + \sum_{k=1}^K \beta_k x_k \quad (13)$$

Em que  $p$  é a probabilidade de ocorrência do evento,  $x$  são variáveis explicativas e  $\alpha$  e  $\beta$  são os parâmetros do modelo. O termo  $\ln\left(\frac{p}{1-p}\right)$  é chamado de *logit* e o termo  $\frac{p}{1-p}$  representa a chance de ocorrência do evento.

Substituindo (13) em (12), temos que a probabilidade do evento  $i$  ocorrer é dada por:

$$P_i = f(Y = 1 | X_1, X_1, \dots, X_k) = \frac{1}{1 + e^{-\alpha + \sum_{k=1}^K \beta_k x_k}} \quad (14)$$

Como pode ser observado, o modelo logístico não é linear e, portanto, não pode ser estimado pelo método de mínimos quadrados ordinários. Em vez disso, utiliza-se a estimação através do método de máxima verossimilhança, que estimará parâmetros  $\beta$  que maximizem a probabilidade de se obter o conjunto observado de dados (Hosmer e Lemeshow, 1989) através da função de log-verossimilhança (LLF).

Para assegurar a significância do impacto das variáveis independentes na função de log-verossimilhança, o teste-t é aplicado de forma similar ao modelo de mínimos quadrados ordinários, tendo como hipótese nula o valor do coeficiente ser igual a zero, ou seja, não ser significativo. Assim como os testes do Método de Mínimos Quadrados Ordinários utilizam a soma do quadrado dos resíduos como critério para seleção dos parâmetros, o método de regressão logística utiliza a log-verossimilhança (multiplicada por -2), -2LLF, como critério de seleção dos parâmetros (Hair *et al.*, 2005). De forma geral, quanto maior o valor, pior a capacidade preditiva da variável dependente (Menard, 2001).

Menard (2001) também sugere o uso do teste qui-quadrado, que basicamente testa as diferenças entre o esperado e o observado. Brooks (2008) sugere duas medidas para testar o *goodness-of-fit*:

1. Percentual de previsões corretas: uma estatística intuitiva que, assim como o nome diz, busca comparar o número de estimativas corretamente previstas com o número total de estimativas totais. Deve ser ajustada e calculada assim como mostrado em (15). Quanto maior o valor da medida, melhor será o *fit* do modelo.

$$\% \text{ de previsões corretas} = 100 \times \left[ \frac{\sum y_i I(\hat{P}_i)}{\sum y_i} + \frac{\sum (1 - y_i) (1 - I(\hat{P}_i))}{N - \sum y_i} \right] \quad (15)$$

$$\text{onde } y_i = \begin{cases} 1, & \text{caso afirmativo} \\ 0, & \text{caso contrário} \end{cases} \text{ e } I(\hat{y}_i) = \begin{cases} 1, & \text{se } \hat{y}_i > \bar{y} \\ 0, & \text{caso contrário} \end{cases}$$

2. Pseudo-R<sup>2</sup>: também conhecido como R<sup>2</sup> McFadden, busca verificar o nível de significância do modelo. Está restrito entre 0 e 1, atingindo o valor de 1 apenas em caso de estimativas perfeitas. Diferentemente do R<sup>2</sup> clássico, não se pode fazer uma interpretação sobre a proporção da variação na variável dependente que está sendo explicada pelo modelo.

$$\text{Pseudo-R}^2 = 1 - \frac{LLF}{LLF_0} \quad (16)$$

No qual LLF é o valor maximizado da função log-verossimilhança para o modelo logístico e LLF<sub>0</sub> é o valor da função log-verossimilhança de um modelo restrito, em que todos os coeficientes dos parâmetros são iguais a 0.