



IBMEC SÃO PAULO
Programa de Mestrado Profissional em Economia

Gustavo Liberali

**PREVISÃO DE RETORNOS INTRADIÁRIOS
ATRAVÉS DE REGRESSÕES USANDO FUNÇÕES-NÚCLEO**

São Paulo
2007

Gustavo Liberali

**Previsão de retornos intradiários
através de regressões usando funções-núcleo**

Dissertação apresentada no Mestrado Profissionalizante em Macroeconomia e Finanças, como requisito parcial para obtenção do grau de Mestre em Macroeconomia e Finanças do IBMEC SÃO PAULO.

Campo de conhecimento: Finanças

Orientador: Prof. Dr. Pedro Luiz Valls Pereira
(IBMEC SÃO PAULO)

São Paulo
2007

Liberali, Gustavo

Previsão de retornos intradiários através de regressões usando funções-núcleo / Gustavo Liberali. – São Paulo: IBMEC SÃO PAULO, 2007.

51 p.

Dissertação: (Mestrado – Programa de Mestrado Profissional em Economia. Área de concentração: Finanças) - IBMEC SÃO PAULO.

Orientador: Prof. Dr. Pedro Luiz Valls Pereira

1. Finanças de alta frequência
2. Regressões não-lineares
3. Indicadores Técnicos

FOLHA DE APROVAÇÃO

Gustavo Liberali

Previsão de retornos intradiários através de regressões usando funções-núcleo

Dissertação apresentada no Mestrado Profissionalizante em Macroeconomia e Finanças, como requisito parcial para obtenção do grau de Mestre em Macroeconomia e Finanças do Ibmec São Paulo.

Campo de conhecimento: Finanças

Aprovado em 21/Maio/2007

Orientador:

Prof. Dr. Pedro Luiz Valls Pereira

IBMEC SÃO PAULO

Banca Examinadora:

Prof. Dr. Pedro Luiz Valls Pereira
IBMEC – SÃO PAULO
Orientador

Prof. Dr. Rinaldo Artes
IBMEC – SÃO PAULO
Examinador

Prof. Dr. Luiz Koodi Hota
UNICAMP
Examinador

Agradecimentos

Os agradecimentos são muitos e todos merecidos.

A minha família, em primeiro lugar, agradeço o incondicional apoio e estímulo para a busca da realização profissional e acadêmica.

Ao Prof. Dr. Pedro Valls agradeço sua generosa atenção e orientação. Seus conhecimentos e experiência eu não encontraria em nenhum outro lugar.

Aos colegas da turma agradeço a amizade e companhia durante a jornada inteira. Tenho a certeza e a alegria de que nossos caminhos se cruzarão novamente muitas vezes.

Agradeço também a Deus pela oportunidade de me desenvolver como acadêmico e profissional de finanças.

Dedicatória

Dedico este trabalho a todos aqueles que acreditam na pesquisa e buscam aplicações práticas no conhecimento acadêmico de finanças.

“Only the clairvoyant could hope to predict with certainty.
Clairvoyant analysis have no need for the techniques of this monograph.”

Harry Markowitz

Resumo

Liberali, Gustavo. Previsão de retornos intradiários através de regressões usando funções-núcleo. São Paulo, 2007. 51 p. Dissertação – Programa de Mestrado Profissional de Economia do IBMEC SÃO PAULO.

São duas as contribuições deste trabalho. Primeiro, é discutido e aplicado um método de avaliação de regressões não lineares para a previsão de retornos intradiários de ações no mercado brasileiro, com o objetivo de maximizar o retorno de um portfólio simulado de compras e vendas. Segundo, são realizadas regressões usando funções-núcleo associadas ao particionamento da amostra por Vizinhos Mais Próximos. Algumas variáveis independentes utilizadas são indicadores técnicos, cujos parâmetros são otimizados dentro da amostra de estimação. Os resultados alcançados são positivos e superam, em uma análise quartil a quartil, os resultados produzidos por um modelo benchmark de autoregressão linear.

Palavras-chave: Retornos Intradiários, Regressão usando funções-núcleo, Vizinhos Mais Próximos, Indicadores Técnicos.

Abstract

Liberali, Gustavo. Forecasting Intraday returns using Kernel regression. São Paulo, 2007. 51 p. Dissertation – Professional Master in Economics IBMEC SÃO PAULO.

The contributions of this paper are twofold. First we discuss and apply a method for the evaluation of non linear regressions in forecasting intraday returns of Brazilian stocks, with the aim of maximizing the return of a simulated trading portfolio. Second, Kernel regressions associated with Nearest Neighbors sample partitioning are carried out. Some independent variables are technical indicators, which parameters are optimized in-the-sample. The results are positive as a trading strategy and outperformed the linear autoregression benchmark model in a quantile per quantile analysis.

Keywords: Intraday Returns, Kernel Regression, Nearest Neighbors, Technical Indicators.

Sumário

<i>Introdução</i>	12
1. Método	14
1.1. Base de Dados	15
1.1.1. Tratamento de Dados	16
1.1.2. Divisão da Amostra.....	17
1.2. Equacionamento Matemático das Regras	18
1.2.1. O Modelo Ingênuo	19
1.2.2. Indicadores Técnicos	19
1.2.3. Regressão Não Linear	21
1.2.4. Truncamento por Vizinhos Mais Próximos	24
2. Resultados Empíricos	26
2.1. Análise do Desempenho das Regras.....	26
2.2. Análise da distribuição dos resultados por negócio.....	31
2.3. Análise de sensibilidade à taxa de barreira.....	35
3. Considerações Finais	43
4. Conclusão	44
<i>Referências bibliográficas</i>	46
<i>Apêndice 1 - Marcas a superar para se efetivar ordem de compra</i>	48
<i>Apêndice 2 - Distribuições de retornos</i>	50

Lista de tabelas

Tabela 1 – Ações e amostragens de preços intradiários	16
Tabela 2 – Ganho médio por pregão na Regressão com Função-Núcleo de ordem 0	26
Tabela 3 – Média de negócios disparados por pregão na Regressão com Função-Núcleo de ordem 0	27
Tabela 4 – Ganho médio por negócio na Regressão com Função-Núcleo de ordem 0	28
Tabela 5 – Ganho médio por pregão na Regressão com Função-Núcleo de ordem 1	29
Tabela 6 – Média de negócios disparados por pregão na Regressão com Função-Núcleo de ordem 1	29
Tabela 7 – Ganho médio por negócio na Regressão com Função-Núcleo de ordem 1	30
Tabela 8 – Percentual de resultados negativos na Regressão com Função-Núcleo de ordem 0	31
Tabela 9 – Estatística da distribuição de ganho médio por negócio na Regressão com Função-Núcleo de ordem 0	32
Tabela 10 – Percentual de resultados negativos na Regressão com Função-Núcleo de ordem 1	33
Tabela 11 – Estatística da distribuição de ganho médio por negócio na Regressão com Função-Núcleo de ordem 1	33
Tabela 12 – Sensibilidade do ganho médio por pregão do Modelo Ingênuo	35
Tabela 13 – Sensibilidade da média de negócios realizados por pregão do Modelo Ingênuo	36
Tabela 14 – Sensibilidade do ganho médio por negócio do Modelo Ingênuo	37
Tabela 15 – Sensibilidade da estatística da distribuição do ganho médio por negócio do Modelo Ingênuo	37
Tabela 16 – Sensibilidade do ganho médio por pregão na Regressão com Função-Núcleo de ordem 0	38
Tabela 17 – Sensibilidade da média de negócios realizados por pregão na Regressão com Função-Núcleo de ordem 0	38

Tabela 18 – Sensibilidade do ganho médio por negócio na Regressão com Função-Núcleo de ordem 0	39
Tabela 19 – Sensibilidade da estatística da distribuição do ganho médio por negócio na Regressão com Função-Núcleo de ordem 0	40
Tabela 20 – Sensibilidade do ganho médio por pregão na Regressão com Função-Núcleo de ordem 1	40
Tabela 21 – Sensibilidade da média de negócios realizados por pregão na Regressão com Função-Núcleo de ordem 1	41
Tabela 22 – Sensibilidade do ganho médio por negócio na Regressão com Função-Núcleo de ordem 1	41
Tabela 23 – Sensibilidade da estatística da distribuição do ganho médio por negócio na Regressão com Função-Núcleo de ordem 1	42

Introdução

Este trabalho baseia-se principalmente em estudos de análise não linear de séries temporais. Especificamente, serão avaliados modelos para a previsão de retornos intradiários durante um pregão fora da amostra.

Mizrach (1992) desenvolve um modelo multivariado não paramétrico para projetar taxas de câmbio de três moedas europeias com amostragem diária, utilizando 3 lags de retornos como variáveis independentes. Seu modelo considerou 3063 observações para o período de estimação dentro da amostra e 757 observações para o período de validação fora da amostra, e para apenas uma das moedas os resultados encontrados foram significativamente superiores ao passeio aleatório (random walk). Nesta mesma linha, Dockner *et alli* (1997) usam regressões lineares e Vizinhos Mais Próximos para prever retornos futuros de um conjunto de ações do mercado austríaco, utilizando retornos recentes como variáveis independentes. Suas conclusões mostram que a regressão por Vizinhos Mas Próximos capturou a estrutura autorregressiva nos retornos e que o erro das previsões cresce muito rapidamente quanto maior for o horizonte de tempo previsto.

Lebaron (1992) encontra melhorias significativas nas regressões AR(1) se forem consideradas apenas observações Vizinhos Mais Próximos pelo critério de volatilidade condicional. Ele destaca que seu estudo não conclui pela superioridade do método em quaisquer circunstâncias, mas constitui uma forte evidência de que tal método não paramétrico tem muito a contribuir na exploração de conteúdo preditivo de séries de tempo, o que motivou em boa parte o presente trabalho.

Além de retornos defasados, decidiu-se usar neste estudo dois indicadores técnicos como variáveis independentes na regressão. Indicadores técnicos, embora já sejam de uso corrente há décadas, têm enfrentado avaliações muito criteriosas na literatura de previsão. Uma referência esclarecedora foi Brock *et alli* (1992), onde foram encontradas evidências de retornos significativamente positivos através do uso de duas famílias de regras elementares (Médias Móveis e Suporte/Resistência), mesmo quando testados por bootstrap. Sullivan *et alli* (1999) estendem tais estudos para uma amostra de retornos de ações no mercado norte-americano ainda mais representativa.

No Brasil, o trabalho de Baptista (2002), atualizado e estendido em Baptista e Valls Pereira (2006), testa 4 famílias de regras de Análise Técnica, encontrando retornos positivos significativos mesmo quando simulados negócios com custos de transação. Este trabalho é uma referência importante por ser um dos primeiros a tratar de dados intradiários no mercado brasileiro de ações. Através do White's Reality Check ele conclui que nenhuma regra individualmente tem performance superior às demais, porém o desempenho de alguns grupos foi sistematicamente superior à do benchmark.

Este trabalho está organizado em 3 seções. A próxima seção discute o tratamento da base de dados, o benchmark adotado e o método empregado nas regressões usando funções-núcleo. A seção seguinte analisa os retornos de cada modelo individualmente e em comparação com o benchmark, além de realizar algumas análises de sensibilidade ao parâmetro taxa de barreira. A última seção discute possíveis extensões deste estudo e traça conclusões com base nas evidências empíricas encontradas.

1. Método

Neste trabalho procurou-se comparar os ganhos e perdas produzidos por uma regra matemática simples (“ingênua”) com os de uma regra fundamentada em regressão não linear, através da sinalização de oportunidades de compra de ações para venda no período imediatamente seguinte.

Nos trabalhos de Baptista (2002) e Saffi (2003) o desempenho de regras baseadas em indicadores técnicos é testado de forma que as posições compradas ou vendidas podem ser encerradas muitos períodos depois. Makridakis (1990) avaliou regras cujas posições compradas ou vendidas poderiam durar muitos períodos. Tais regras foram baseadas em janelas móveis de observações passadas de tamanho proporcional ao número de períodos futuros para o qual se avalia uma compra ou venda. O presente trabalho, por sua vez, não avalia um conjunto de indicadores técnicos para abertura de posições com regras ad-hoc para sinalizar o respectivo fechamento. Regras ad-hoc estão sujeitas à calibração de uma variedade geralmente grande de parâmetros. Também não é objetivo da regressão não paramétrica estimar um processo fechado gerador de preços, como seria o caso nos modelos da classe ARMA.

A regressão usando funções-núcleo empregada, truncada por vizinhos mais próximos, pressupõe o comportamento dos preços como caótico e imprevisível no médio e longo prazo. Como descrevem Dackner *et alli* (1997), o método de vizinhos mais próximos determina que apenas uma fração da informação passada será usada para se fazer previsões. Esta fração é escolhida com base em um pequeno conjunto de condições correntes, selecionando-se apenas observações passadas onde tais condições haviam sido observadas. Este método de seleção de seções da amostra nos leva a esperar que a previsão seja muito eficiente para o período seguinte, e pouco eficiente para os demais períodos futuros.

Wolberg (2000) destaca ainda o fato de que quando as condições correntes não podem ser encontradas na amostra de observações passadas, a seleção de observações fica prejudicada, e o resultado comprometido. Segundo o autor, esta vulnerabilidade a “condições de bordas” também é encontrada em Redes Neurais Multilayer Perceptron. Portanto ela pode ser interpretada como uma deficiência

comum a estes dois métodos não-lineares na extrapolação da relação entre variável dependente e variáveis explicativas.

Uma limitação prática imposta a este estudo foi a não utilização de posições vendidas a descoberto, pois levar em conta as limitações e custos envolvidos nesta prática no mercado brasileiro tornaria o estudo empírico deveras mais complexo.

Desta forma, durante as simulações de compra e venda existiu apenas dois estados da carteira de investimentos: comprado ou neutro. Cada estado comprado dura apenas um período de 5 ou 20 minutos conforme a taxa de amostragem da série de preços, e não foi permitido se estar comprado duas vezes simultaneamente. Ao contrário das simulações realizadas em Dacorogna *et alli* (2001), onde o tamanho da posição em determinado ativo poderia variar em função da intensidade do sinal de compra ou venda, neste trabalho as compras foram realizadas com todo o recurso disponível e durante os períodos neutros o caixa não foi rentabilizado. Desta forma, o retorno sobre capital investido é acumulado ao longo das simulações, e o retorno logarítmico total é a soma dos retornos logarítmicos de cada negócio (compra e venda subsequente). Os resultados das simulações serão analisados por retorno log médio por negócio realizado.

1.1. Base de Dados

A base de dados compreende os preços intradiários de negócios fechados na Bovespa de 01-set-03 a 28-mar-05, tendo sido coletados na agência de informações Bloomberg. Destas séries de preços foram extraídas amostras regularmente espaçadas no tempo de 6 ações participantes do índice Ibovespa, conforme demonstrado na tabela 1.

Tabela 1 – Ações e amostragens de preços intradiários

Ação	Amostragem	Observações	Código
CSNA3	5 minutos	32.732	CSNA3_05min
CSNA3	20 minutos	9.351	CSNA3_20min
GGBR4	5 minutos	33.270	GGBR4_05min
GGBR4	20 minutos	9.351	GGBR4_20min
PETR4	5 minutos	33.515	PETR4_05min
TNLP4	5 minutos	33.363	TNLP4_05min
USIM5	5 minutos	33.556	USIM5_05min
USIM5	20 minutos	9.354	USIM5_20min
VALE5	5 minutos	32.615	VALE5_05min
VALE5	20 minutos	9.237	VALE5_20min

Os negócios destas 6 ações efetivamente ocorridos durante o dia são assíncronos, ou seja, os dados não são igualmente espaçados no tempo. Desta forma, pode haver muitos negócios em um período da amostra e nenhum no período seguinte, o que pode resultar também em diferentes números de observações encontrados em ações com a mesma taxa de amostragem.

1.1.1. Tratamento de Dados

Uma questão importante trata da real possibilidade de se negociar as ações no exato momento em que um sinal é disparado pela regra matemática. Similar à Baptista (2002), para se tornar o estudo mais aplicável à realidade operacional, foi considerado como preço efetivo de compra e venda à média simples, não ponderada, entre o preço no período no qual o sinal é emitido e o preço no período seguinte.

Outro aspecto operacional relevante se refere aos custos de transação. Alguns estudos anteriores, como em Baptista e Valls Pereira (2006), consideram um custo arbitrário por transação para se verificar a robustez dos resultados encontrados. Outros estudos de regras matemáticas de negociação de ações, como Saffi (2003) consideram tais custos como desprezíveis. Porém, este estudo trata de

regras matemáticas que podem emitir sinais de compra e venda intradiários, o que pode resultar em um grande volume de operações e um custo associado significativo. Além disto, compras e vendas intradiárias tendem a produzir retornos por negócio muito menores do que compras e vendas com frequência diária ou superior, ficando assim seu resultado líquido mais sensível a tais custos.

Neste trabalho não serão pré-estabelecidos custos arbitrários por transação. Em vez disto, será computada a estatística descritiva dos retornos médios por negócio de compra e venda, cabendo ao leitor julgar se tais retornos são compatíveis com os custos de transação à que ele está sujeito.

1.1.2. Divisão da Amostra

A amostra foi dividida em duas partes: os primeiros 252 dias de negócios (pregões) consecutivos foram considerados o período de estimação e o pregão seguinte foi considerado período de validação. Os negócios intradiários ocorridos em cada pregão foram amostrados às taxas de 5 ou 20 minutos. Considerando-se que cada pregão tem duração de 7 horas, existem aproximadamente 84 ou 21 observações por pregão, respectivamente.

A amostra completa cobre 391 pregões, o que permite repetir 139 vezes a simulação da realização de negócios intradiários durante o pregão de validação, com amostras de estimação e de validação diferentes. Para tanto, ao final de cada simulação, o pregão de validação foi incorporado à amostra de estimação e o pregão seguinte passou a ser a amostra de validação. Da mesma forma, o pregão de estimação mais antigo foi retirado da amostra para que ela se mantivesse sempre com 252 pregões.

Esta forma de divisão da amostra atinge dois objetivos metodológicos: a) certifica que toda compra ou venda realizada em um dado período t foi deflagrada pelo processamento apenas de informações disponíveis até o período $t-1$, e; b) permite um grande número de repetições do experimento, produzindo assim resultados cuja distribuição de frequência contém informações importantes sobre o desempenho das regras de trading empregadas.

1.2. Equacionamento Matemático das Regras

No momento t as regras de decisão pela compra, e posterior venda, do ativo precisam levar em conta apenas o conjunto de informações disponíveis até $t-1$. Embora tal escopo possa incluir uma variedade de informações tal como preços e quantidades negociadas de diferentes ativos e informação passada de indicadores econômicos, neste estudo foram considerados apenas os preços realizados do próprio ativo em questão.

Os modelos aqui estudados são regressões não lineares, conhecidos como regressões usando funções-núcleo. A variável dependente é o retorno logarítmico do período seguinte, e as variáveis explicativas são o próprio retorno logarítmico defasado um período e indicadores técnicos produzidos com a própria série de preços do ativo.

As regressões produziram uma projeção de retorno para o período seguinte. Nem toda projeção de retorno positivo resulta necessariamente em uma ordem de compra. Para que o sinal de compra seja disparado, a projeção de retorno no período de validação precisou superar uma determinada marca, previamente estabelecida. Esta marca foi calculada com base nas projeções de retorno calculadas na amostra de estimação. Um percentil dos retornos projetados dentro desta amostra foi empregado como tal marca. Este percentil é determinado pelo parâmetro aqui chamado de taxa de barreira, que varia entre 0 e 1. Uma taxa de barreira baixa, próxima a 0, é considerada restritiva por estabelecer uma marca a ser superada próxima ao maior retorno projetado na amostra de estimação. Da mesma forma, uma taxa de barreira alta, próxima a 1, estabelece uma marca a ser superada próxima ao menor retorno projetado na amostra de estimação, e, portanto tende a permitir que seja disparado um volume muito grande de ordens de compra.

Assumindo-se que as distribuições de retornos projetados no período de estimação e no período de validação sejam parecidas, espera-se que uma taxa de barreira próxima a 1 qualifique aproximadamente 100% das projeções como ordens de compras, enquanto uma taxa de barreira próxima a 0 evite que praticamente qualquer ordem de compra seja disparada.

1.2.1. O Modelo Ingênuo

Baptista e Valls Pereira (2006) e Lebaron (2002) destacam a significância das autocorrelações de primeira ordem nas séries de preços intradiários de ações. Dacorogna *et alli* (2001) no estudo sobre preços intradiários de taxas de câmbio observam que a autocorrelação de preços intradiários tende a ser negativa e muito significativa quando eles são amostrados com frequência de até 4 minutos.

Na amostra utilizada neste trabalho foram encontradas evidências de forte autocorrelação negativa de primeira ordem. Mesmo nos testes onde o poder explicativo dos indicadores técnicos individualmente foi baixo, a estrutura autoregressiva de ordem 1 muito raramente pode ser rejeitada, a uma significância de 5%.

Uma autoregressão de primeira ordem, portanto foi escolhida como o Modelo Ingênuo, ou benchmark, contra o qual os modelos não lineares serão avaliados neste trabalho.

1.2.2. Indicadores Técnicos

Foram utilizados dois indicadores técnicos amplamente empregados no mercado de ações: o Índice de Força Relativa (Relative Strength Index - RSI) e a Média Móvel Convergência/Divergência (Moving Average Convergence / Divergence - MACD). Para cada instante no tempo, cada indicador produz um índice, com base na informação passada da respectiva série de preços. Estes índices tendem a ser estacionários - motivo pelo qual eles são conhecidos popularmente como osciladores. Esta característica é de grande importância em uma regressão usando funções-núcleo, por dificilmente produzir valores no período de validação em uma faixa nunca antes observada no período de estimação. A observação de valores inéditos no índice exigiria do modelo a previsão em condições de extrapolação, onde é esperado do modelo de regressões usando funções-núcleo a menor eficiência.

Conforme Bloomberg Magazine (1995), o indicador RSI foi desenvolvido por J. Welles Wilder e é calculado da seguinte forma:

$$RSI_t = 100 - \frac{100}{1 + RS_t} \quad (1)$$

, onde:

$$RS_t = \frac{\sum_{i=t-N}^t \max(0, P_i - P_{i-1}) / N}{-\sum_{i=t-N}^t \min(0, P_i - P_{i-1}) / N} \quad (2)$$

Na equação 2, P_t é o preço do ativo no instante t e N é o número de observações consideradas na análise. O índice RSI, que oscila entre 0 e 100, foi criado originalmente para a análise de séries diárias de preços e o parâmetro N acima é calibrado normalmente para 14 dias. Os defensores deste indicador interpretam um RSI acima de 70 como o início de uma tendência de queda de preços e um resultado abaixo de 30 como o início de uma tendência de alta de preços.

Conforme StockCharts (2006), o indicador MACD foi desenvolvido originalmente por Gerald Appel através da observação da diferença entre duas médias móveis da mesma série de preços. Sua fórmula é a seguinte:

$$MACD_t = \frac{\sum_{i=t-N_1}^t P_i}{N_1} - \frac{\sum_{i=t-N_2}^t P_i}{N_2} \quad (3)$$

Este índice também foi criado originalmente com a intenção de se analisar séries diárias de preços e os parâmetros N_1 e N_2 são normalmente calibrados para 26 e 12 dias, respectivamente.

Ambos os indicadores têm a propriedade de serem calculados com relativa facilidade. Assim como as médias móveis do MACD, a soma das altas e baixas dos últimos N dias do RSI pode ser atualizada apenas adicionando-se, de forma ponderada, a informação mais recente. Evita-se assim o recálculo do indicador inteiro e acelera-se o processo computacional.

Antes da realização da regressão, para o indicador técnico RSI foi estimado o parâmetro N que produzisse o menor erro médio de previsão. Da mesma forma, para o indicador técnico MACD foi estimado o parâmetro N_1 , e de forma arbitrária foi considerado $N_2 = N_1 * 12 / 26$.

Especificamente, o erro médio de previsão calculado foi a média das diferenças entre as projeções de retornos positivos e os retornos efetivamente realizados. Este cálculo foi escolhido como critério de otimização porque leva em consideração as situações em que mais espera-se haver o disparo de ordens de compra, dado que elas são condicionais à ocorrência de projeções de retorno superiores à marca estabelecida pela taxa de barreira.

O procedimento de otimização foi realizado minimizando-se tal erro de projeção através do algoritmo FMINBND do Matlab versão 6.5, dentro apenas da amostra de estimação. O valor estimado para o parâmetro foi então utilizado para cálculo do respectivo indicador técnico na amostra inteira.

Para Baptista e Valls Pereira (2006) diferentes valores informados como parâmetro de um mesmo indicador técnico produziram diferentes regras de investimento que foram consideradas e avaliadas na amostra inteira como indicadores técnicos inteiramente novos. Neste trabalho, entretanto, a otimização dos parâmetros na amostra do período de estimação produz apenas um indicador técnico a ser empregado como variável independente nas regressões usando funções núcleo. Além disto, a cada nova amostra do período de estimação selecionada, os parâmetros são novamente estimados, relaxando-se parcialmente a hipótese de que eles sejam constantes no tempo.

1.2.3. Regressão Não Linear

A regressão não linear foi empregada para a realização da projeção de retornos do preço da ação com base na informação de preços realizados. As variáveis independentes escolhidas foram: 1) o termo autoregressivo de primeira ordem; 2) os indicadores técnicos RSI e MACD e 3) o horário em que o negócio foi realizado.

A regressão usando funções-núcleo é uma classe específica de métodos fundamentados em alisamento ou suavizamento. O propósito geral de um método de alisamento é encontrar uma linha, ou superfície, que descreva o comportamento geral de uma variável dependente como função das variáveis independentes, ou explicativas. Métodos de suavizamento não procuram descrever o processo gerador

da variável dependente através de equações matemáticas. Eles são métodos não paramétricos e, portanto, buscam diretamente na amostra a informação necessária para realizar uma projeção.

A característica específica do método aqui empregado é o uso de um kernel, ou seja, de uma função que confira pesos às observações passadas no momento em que forem ponderadas no cálculo de cada projeção. Neste estudo foi utilizado o kernel mais simples encontrado no trabalho de Wolberg (2000). A fórmula do peso da observação x_i na estimação do ponto x_j é a seguinte:

$$w(x_i, x_j, k) = e^{-kD_{ij}^2} \quad (4)$$

O parâmetro k controla o decaimento exponencial que o peso sofre em relação à distância euclidiana que separa x_i de x_j . A distância euclidiana D^{ij} foi a forma encontrada para se levar em conta a diferença entre os valores de cada variável independente nos pontos x_i e x_j . Segundo Wolberg (2000), a distância é a raiz quadrada da soma dos quadrados das diferenças dos valores das variáveis independentes em x_i e x_j , mas sem prejuízo relevante ao propósito do cálculo, a raiz quadrada pode ser retirada para tornar a computação mais veloz.

Com a Função-Núcleo definida, é preciso escolher a regressão onde ela será empregada. O estimador de Nadaraya-Watson é considerado como o algoritmo de regressão mais simples a empregar uma Função-Núcleo.

$$y_j = \frac{\sum_{i=1}^{nlrn} w(x_i, x_j, k) Y_i}{\sum_{i=1}^{nlrn} w(x_i, x_j, k)} \quad (5)$$

Nesta equação, y_j é a projeção que se está realizando e Y_i representa os valores já conhecidos da mesma variável. O tamanho da amostra no período de estimação está representado na equação acima por $nlrn$, que, no caso deste trabalho, trata-se do número de observações em 252 pregões. Por sua semelhança com um polinômio de ordem 0, esta equação será tratada como uma Regressão com Função-Núcleo de ordem 0, ou Regressão usando Função-Núcleo de ordem 0.

Uma Regressão com Função-Núcleo de ordem 1 e com p variáveis independentes recebe um termo adicional.

$$y_i = a_1 + a_2 x_1 + a_3 x_2 + \dots + a_{p+1} x_p + e \quad (6)$$

Segundo Wolberg (2000) a estimação dos p+1 coeficientes da regressão acima é tipicamente realizada através de Mínimos Quadrados Ponderados.

Como este estudo trabalha com séries históricas que contém informações com bastante ruído, não há sentido se trabalhar com algoritmos de ordem muito alta. Serão utilizadas apenas as ordens 0 e 1.

Da amostra no período de estimação foram selecionadas aleatoriamente observações de teste para se realizar a otimização do parâmetro k. Um parâmetro k igual a 0 faria com que todas as observações restantes na amostra no período de estimação tivessem o mesmo peso. Este efeito em uma regressão de ordem zero produziria apenas uma média aritmética não ponderada de todos os retornos da amostra; e em uma regressão de ordem um resultaria em uma regressão linear. Já um parâmetro k entre 0 e 8 faz com que as observações tenham pesos distintos na regressão conforme as similaridades entre x_i e x_j .

Para cada amostra de estimação, um parâmetro k inicial produziu um determinado erro de previsão das observações selecionadas aleatoriamente. Um parâmetro k diferente foi então escolhido para uma nova regressão das mesmas observações, na esperança de se reduzir o erro de regressão. Através do algoritmo FMINBND do Matlab v. 6.5 este processo foi repetido diversas vezes visando minimizar tal erro. Esta otimização de k foi realizada dentro do intervalo arbitrário de 0 a 300, e o parâmetro k encontrado nas amostras de estimação ficou tipicamente entre 1 e 30. O erro de regressão considerado neste processo otimização foi a soma das diferenças entre as projeções de retornos de alta e os retornos efetivamente realizados. Isto porque previsões de retornos de baixa tendem a não disparar ordens de compra e não há, em princípio, motivos para se levar em consideração desvios entre retornos projetados e realizados nestas ocasiões.

1.2.4. Truncamento por Vizinhos Mais Próximos

Para cada observação da amostra pode-se calcular um $w(x_i, x_j, k)$ e assim incluí-la na Regressão com Função-Núcleo. Dado que em um típico pregão, em séries amostradas de 5 em 5 minutos, existem aproximadamente 84 observações e que a amostra no período de estimação foi fixada em 252 pregões, podem existir mais de 21 mil observações a serem computadas a cada regressão.

O Truncamento por Vizinhos Mais Próximos visa reduzir a quantidade de observações computadas a cada regressão selecionando apenas determinadas seções da amostra através de um critério similar à distância euclidiana usada em w , mas de implementação mais rápida.

Segundo Wolberg (2000) este método permite encontrar rapidamente os pontos x_j mais próximos à x_i , levando-se em conta múltiplas variáveis independentes. Para tanto, uma vez determinada a amostra no período de estimação, ela foi particionada em duas conforme o valor observado na primeira variável independente. Para que ambas as partições tivessem o mesmo número de observações, uma delas recebeu as observações com valores acima da mediana da primeira variável independente e a outra com valores abaixo da mediana. Em seguida, cada uma das partições foi subdividida conforme o valor observado na segunda variável independente. Novamente, para que cada subpartição recebesse o mesmo número de observações, a mediana da segunda variável independente dentro de cada partição foi o critério de separação. Quando a última variável independente for usada como critério de partição, a primeira variável independente é usada outra vez e o processo se repete. Assim, sucessivos particionamentos foram realizados até que tivesse sido atingido o número mínimo arbitrário de 200 observações por partição.

Uma amostra assim particionada pode ser empregada em regressões lineares, como em Mizrach (1992). Neste estudo ela serviu para truncar a amostra de forma a preservar as 200 observações onde se espera encontrar os pesos w mais altos. Verificou-se, conforme esperado, um aumento de velocidade de processamento proporcional ao número de particionamentos da amostra. Por outro lado, um número excessivo de particionamentos deixa um número muito reduzido de observações por partição, podendo comprometer assim a eficiência da regressão.

Este particionamento reforça a tese de que o método empregado neste trabalho concentra-se em projeções para apenas um período à frente e leva em consideração apenas seções da informação passada onde se observou valores nas variáveis independentes o mais perto possível dos valores correntes das mesmas. Da mesma forma, nos casos onde os valores correntes das variáveis independentes não tenham sido observados simultaneamente na amostra de estimação, espera-se perda na eficiência da regressão e, portanto, um desempenho mais baixo nas simulações de compra e venda.

2. Resultados Empíricos

Os resultados empíricos serão apresentados em três seções: uma dedicada à análise comparativa do desempenho dos modelos empregados, outra à análise da distribuição dos retornos médios por negócio, e a última dedicada à análise de sensibilidade dos resultados à variações no parâmetro taxa de barreira.

2.1. Análise do Desempenho das Regras

Na tabela 2 é comparado o desempenho do modelo de Regressão usando Função-Núcleo de ordem 0 com o desempenho do modelo benchmark, ou Ingênuo. A coluna mais à esquerda mostra os resultados do Modelo Ingênuo e as demais colunas mostram os resultados de modelos de Regressão usando Função-Núcleo com diferentes conjuntos de variáveis independentes. Em negrito foram destacados os máximos de cada linha.

Tabela 2 – Ganho médio por pregão na Regressão com Função-Núcleo de ordem 0

Modelo	Ingênuo	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	+0,47%	+0,33%	+0,14%	+0,14%	+0,12%	+0,02%	+0,20%
csna3_20min	+0,79%	+0,82%	+0,28%	+0,36%	+0,53%	+0,00%	+0,46%
ggbr4_05min	+0,41%	+0,35%	+0,22%	+0,27%	+0,18%	+0,15%	+0,26%
ggbr4_20min	+0,82%	+0,65%	+0,36%	+0,49%	+0,26%	+0,05%	+0,44%
petr4_05min	+0,20%	+0,13%	+0,08%	+0,05%	+0,03%	-0,01%	+0,08%
tnlp4_05min	+0,60%	+0,56%	+0,49%	+0,50%	+0,42%	+0,43%	+0,50%
usim5_05min	+0,24%	+0,20%	+0,14%	+0,10%	+0,13%	+0,17%	+0,16%
usim5_20min	+0,22%	+0,18%	+0,05%	+0,13%	+0,10%	-0,08%	+0,10%
vale5_05min	+0,42%	+0,29%	+0,16%	+0,09%	+0,09%	+0,07%	+0,19%
vale5_20min	+0,62%	+0,54%	+0,23%	+0,34%	+0,18%	+0,06%	+0,33%
Média	+0,48%	+0,40%	+0,22%	+0,25%	+0,20%	+0,09%	

Entre 9 de 10 ações, o Modelo Ingênuo apresentou o maior ganho médio diário. Também é possível perceber uma grande diferença no ganho médio diário entre os resultados das simulações com o modelo de regressão usando a Função-Núcleo ordem 0. A simulação onde o retorno da própria ação defasado em um período não foi utilizado apresentou a pior performance, o que demonstra que esta variável independente contém informação relevante sobre o próximo retorno e que não pode ser substituída pelos indicadores técnicos. Também é importante notar que uma maior quantidade de variáveis explicativas não produziu necessariamente maiores ganhos diários.

Tabela 3 – Média de negócios disparados por pregão na Regressão com Função-Núcleo de ordem 0

Modelo	Ingênuo	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	17,62	3,76	4,44	3,58	3,74	4,23	6,23
csna3_20min	4,93	0,98	0,81	0,60	0,51	0,56	1,40
ggbr4_05min	19,36	3,72	3,60	3,09	3,16	3,96	6,15
ggbr4_20min	4,56	0,90	0,56	0,70	0,61	0,57	1,32
petr4_05min	26,12	3,13	3,71	2,74	2,71	4,13	7,09
tnlp4_05min	7,55	2,32	2,31	2,37	2,36	4,49	3,57
usim5_05min	18,54	2,01	2,30	2,17	3,02	5,00	5,51
usim5_20min	6,09	0,65	0,56	0,44	0,52	0,86	1,52
vale5_05min	24,00	6,15	5,64	4,03	3,79	4,86	8,08
vale5_20min	5,90	0,91	0,71	0,62	0,39	0,56	1,52
Média	13,47	2,45	2,46	2,03	2,08	2,92	

Conforme os resultados da tabela 3, Mesmo utilizando-se a mesma taxa de barreira em todas as simulações, o Modelo Ingênuo disparou uma quantidade muito maior de negócios que os modelos de Regressão com Função-Núcleo de ordem 0. Naturalmente esta característica dilui os ganhos por negócio realizado, o que torna o Modelo Ingênuo menos competitivo à medida que forem deduzidos os custos de transação.

Tabela 4 – Ganho médio por negócio na Regressão com Função-Núcleo de ordem 0

Modelo	Ingênuo	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	0,03%	0,09%	0,03%	0,04%	0,03%	0,01%	0,03%
csna3_20min	0,16%	0,84%	0,34%	0,60%	1,02%	0,00%	0,33%
ggbr4_05min	0,02%	0,09%	0,06%	0,09%	0,06%	0,04%	0,04%
ggbr4_20min	0,18%	0,72%	0,64%	0,70%	0,43%	0,09%	0,33%
petr4_05min	0,01%	0,04%	0,02%	0,02%	0,01%	0,00%	0,01%
tnlp4_05min	0,08%	0,24%	0,21%	0,21%	0,18%	0,09%	0,14%
usim5_05min	0,01%	0,10%	0,06%	0,05%	0,04%	0,03%	0,03%
usim5_20min	0,04%	0,28%	0,09%	0,30%	0,20%	-0,10%	0,07%
vale5_05min	0,02%	0,05%	0,03%	0,02%	0,02%	0,02%	0,02%
vale5_20min	0,11%	0,59%	0,32%	0,55%	0,45%	0,11%	0,22%
média	0,04%	0,16%	0,09%	0,12%	0,10%	0,03%	

Avaliando-se o ganho médio por negócio efetuado da tabela 4, o desempenho do Modelo Ingênuo fica aquém do desempenho dos modelos de Regressão com Função-Núcleo de ordem 0. A ordem relativa de desempenho entre os modelos de Regressão com Função-Núcleo não se altera significativamente.

O desempenho do modelo de Regressão com Função-Núcleo de ordem 1 com as mesmas combinações anteriores de variáveis explicativas é comparado com o desempenho do Modelo Ingênuo na Tabela 5. Novamente foram ressaltados em negrito os máximos de cada linha.

Tabela 5 – Ganho médio por pregão na Regressão com Função-Núcleo de ordem 1

Modelo	Ingênuo	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	+0,47%	+0,28%	+0,16%	+0,13%	+0,10%	-0,02%	+0,19%
csna3_20min	+0,79%	+0,29%	+0,57%	+0,05%	+0,12%	+0,05%	+0,31%
ggbr4_05min	+0,41%	+0,37%	+0,21%	+0,06%	+0,08%	+0,12%	+0,21%
ggbr4_20min	+0,82%	+0,64%	+0,30%	-0,01%	+0,01%	-0,14%	+0,27%
petr4_05min	+0,20%	+0,12%	+0,06%	+0,03%	+0,01%	+0,00%	+0,07%
tnlp4_05min	+0,60%	+0,55%	+0,48%	+0,17%	+0,11%	+0,26%	+0,36%
usim5_05min	+0,24%	+0,18%	+0,13%	-0,01%	-0,07%	+0,07%	+0,09%
usim5_20min	+0,22%	+0,14%	+0,09%	+0,03%	-0,01%	-0,06%	+0,07%
vale5_05min	+0,42%	+0,28%	+0,11%	+0,06%	+0,02%	+0,04%	+0,16%
vale5_20min	+0,62%	+0,52%	+0,24%	+0,05%	+0,01%	+0,13%	+0,26%
média	+0,48%	+0,34%	+0,23%	+0,06%	+0,04%	+0,04%	

O Modelo Ingênuo supera os demais modelos no critério de ganho médio por pregão para todas as ações. Ao contrário do modelo de Regressão com Função-Núcleo de ordem 0, os resultados acima mostram que o ganho médio diminui bastante quando são utilizados os indicadores técnicos como variáveis independentes isoladamente ou em conjunto com o retorno defasado em um período.

Tabela 6 – Média de negócios disparados por pregão na Regressão com Função-Núcleo de ordem 1

Modelo	Ingênuo	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	17,62	4,09	4,29	4,97	4,71	5,87	6,93
csna3_20min	4,93	1,23	1,01	0,79	0,79	0,67	1,57
ggbr4_05min	19,36	3,79	3,68	4,32	4,72	4,77	6,78
ggbr4_20min	4,56	0,77	0,77	0,83	0,70	0,58	1,37
petr4_05min	26,12	3,21	4,34	3,72	4,09	5,44	7,82

tnlp4_05min	7,55	2,38	2,32	4,78	4,88	4,29	4,36
usim5_05min	18,54	2,04	2,38	4,71	0,54	4,77	5,50
usim5_20min	6,09	0,90	0,63	0,49	0,73	0,58	1,57
vale5_05min	24,00	5,79	5,47	6,01	5,54	6,67	8,91
vale5_20min	5,90	0,89	0,87	0,84	0,65	0,84	1,67
média	13,47	2,51	2,57	3,15	2,73	3,45	

A quantidade de negócios disparados pelo modelo de Regressão com Função-Núcleo de ordem 1 é muito semelhante à do modelo de Regressão com Função-Núcleo de ordem 0. Vale observar que, como esperado, nas séries de amostragem mais freqüente a média de negócios disparados por dia é superior à média de negócios disparados por dia nas séries de amostragem menos freqüentes.

Tabela 7 – Ganho médio por negócio na Regressão com Função-Núcleo de ordem 1

Modelo	Ingênuo	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	Média
csna3_05min	0,03%	0,07%	0,04%	0,03%	0,02%	0,00%	0,03%
csna3_20min	0,16%	0,24%	0,57%	0,06%	0,15%	0,07%	0,20%
ggbr4_05min	0,02%	0,10%	0,06%	0,01%	0,02%	0,03%	0,03%
ggbr4_20min	0,18%	0,82%	0,39%	-0,01%	0,02%	-0,25%	0,20%
petr4_05min	0,01%	0,04%	0,01%	0,01%	0,00%	0,00%	0,01%
tnlp4_05min	0,08%	0,23%	0,21%	0,04%	0,02%	0,06%	0,08%
usim5_05min	0,01%	0,09%	0,05%	0,00%	-0,13%	0,01%	0,02%
usim5_20min	0,04%	0,15%	0,14%	0,06%	-0,01%	-0,11%	0,04%
vale5_05min	0,02%	0,05%	0,02%	0,01%	0,00%	0,01%	0,02%
vale5_20min	0,11%	0,59%	0,28%	0,06%	0,02%	0,15%	0,16%
Média	0,04%	0,13%	0,09%	0,02%	0,01%	0,01%	

Assim como nos modelos de Regressão com Função-Núcleo de ordem 0, os resultados médios por negócio nos modelos de Regressão com Função-Núcleo de ordem 1 foram claramente superiores ao benchmark. Comparando-se os resultados esperados dos modelos de regressões usando Função-Núcleo percebe-se uma superioridade dos ganhos produzidos pela Regressão com Função-Núcleo de ordem 0.

2.2. Análise da distribuição dos resultados por negócio

Nesta seção serão analisadas as distribuições dos ganhos médios por negócio de cada simulação. Esta análise se apóia na existência de uma grande quantidade de negócios disparados por simulação. Para cada ação, foram avaliados 139 diferentes períodos de validação de um pregão cada. Em cada pregão existem cerca de 84, ou 21, observações possíveis de disparar um negócio, conforme a frequência da amostragem.

As tabelas 8 e 10 demonstram o percentual de negócios que produziram resultados negativos em relação ao total de negócios disparados. Em ambas as tabelas serão contrastados os resultados do modelo de Regressão com Função-Núcleo com as 5 combinações de variáveis independentes utilizadas na seção anterior com os resultados do Modelo Ingênuo. Em negrito foram ressaltados os valores mínimos por linha.

Tabela 8 – Percentual de resultados negativos na Regressão com Função-Núcleo de ordem 0

Modelo	Ingênuo	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	41%	16%	37%	28%	37%	51%	37%
csna3_20min	68%	29%	39%	51%	18%	51%	55%
ggbr4_05min	44%	16%	27%	17%	27%	28%	34%
ggbr4_20min	42%	21%	34%	31%	48%	53%	40%
petr4_05min	32%	30%	39%	36%	38%	51%	35%
tnlp4_05min	16%	5%	6%	11%	9%	19%	13%
usim5_05min	24%	21%	27%	38%	32%	35%	27%
usim5_20min	55%	26%	48%	39%	40%	56%	51%
vale5_05min	47%	23%	31%	31%	29%	30%	38%
vale5_20min	57%	28%	28%	24%	23%	48%	47%
média	39%	20%	31%	28%	30%	37%	

A proporção de negócios disparados pelo Modelo Ingênuo com resultados negativos foi maior em comparação aos resultados do modelo de Regressão com

Função-Núcleo de ordem zero. Esta constatação era esperada pelo fato de se observar um ganho médio por negócio superior nos resultados do modelo não linear.

Tabela 9 – Estatística da distribuição de ganho médio por negócio na Regressão com Função-Núcleo de ordem 0

Modelo	Ingênuo	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0	Função Núcleo ordem 0
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário
Max	5,27%	5,14%	6,43%	6,31%	5,27%	4,18%
Q3	0,02%	0,16%	0,11%	0,12%	0,11%	0,07%
Mediana	0,00%	0,07%	0,04%	0,04%	0,03%	0,02%
Q1	-0,01%	0,01%	-0,01%	-0,01%	-0,01%	-0,02%
Min	-1,66%	-2,22%	-2,15%	-4,50%	-4,50%	-5,58%
Qtd negócios	18676	3390	3408	2815	2886	4046
Média	0,04%	0,17%	0,09%	0,12%	0,10%	0,03%
Desv pad	0,25%	0,49%	0,34%	0,49%	0,42%	0,26%
Assimetria	1319	574	727	603	700	-193
Curtose	20820	4567	9878	6132	7723	15054

A descrição das distribuições destaca a forte assimetria dos resultados de quase todas as simulações. A simulação do modelo de Regressão com Função-Núcleo com apenas um retorno defasado como variável independente teve sucesso em termos de média de retorno, mas também produziu variância de retornos superior aos demais modelos.

Foi encontrada forte curtose na distribuição de retornos em todas as simulações. A inspeção visual dos histogramas de retorno mostra uma anormal concentração de observações próximas a mediana e algumas observações a grande distância do centro da distribuição, normalmente superiores a ele (apêndice II). Em uma análise quarti a quartil, e desprezando-se os limites superior e inferior, o modelo não linear com variável AR(1) apresentou o melhor desempenho, superando os quartis 1 e 2 por uma boa margem em relação ao Modelo Ingênuo.

A seguir, serão contrastados os resultados do modelo de Regressão com Função-Núcleo de ordem 1 com as 5 combinações de variáveis independentes utilizadas na última seção com os resultados do Modelo Ingênuo.

Tabela 10 – Percentual de resultados negativos na Regressão com Função-Núcleo de ordem 1

Modelo	Ingênuo	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Média
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05	
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário	
csna3_05min	41%	24%	33%	35%	43%	55%	40%
csna3_20min	68%	49%	32%	53%	44%	46%	57%
ggbr4_05min	44%	17%	28%	45%	38%	36%	39%
ggbr4_20min	42%	17%	40%	44%	48%	67%	42%
petr4_05min	32%	30%	40%	42%	51%	50%	37%
tnlp4_05min	16%	4%	6%	35%	37%	31%	24%
usim5_05min	24%	25%	27%	54%	65%	43%	32%
usim5_20min	55%	39%	48%	48%	56%	54%	52%
vale5_05min	47%	19%	37%	45%	42%	43%	42%
vale5_20min	57%	28%	46%	46%	42%	38%	50%
média	39%	22%	32%	43%	43%	44%	

Os resultados acima evidenciam que menos de $\frac{1}{4}$ dos negócios realizados pelo modelo de Regressão com Função-Núcleo de ordem 1 com um AR(1) como variável independente produziram retornos negativos.

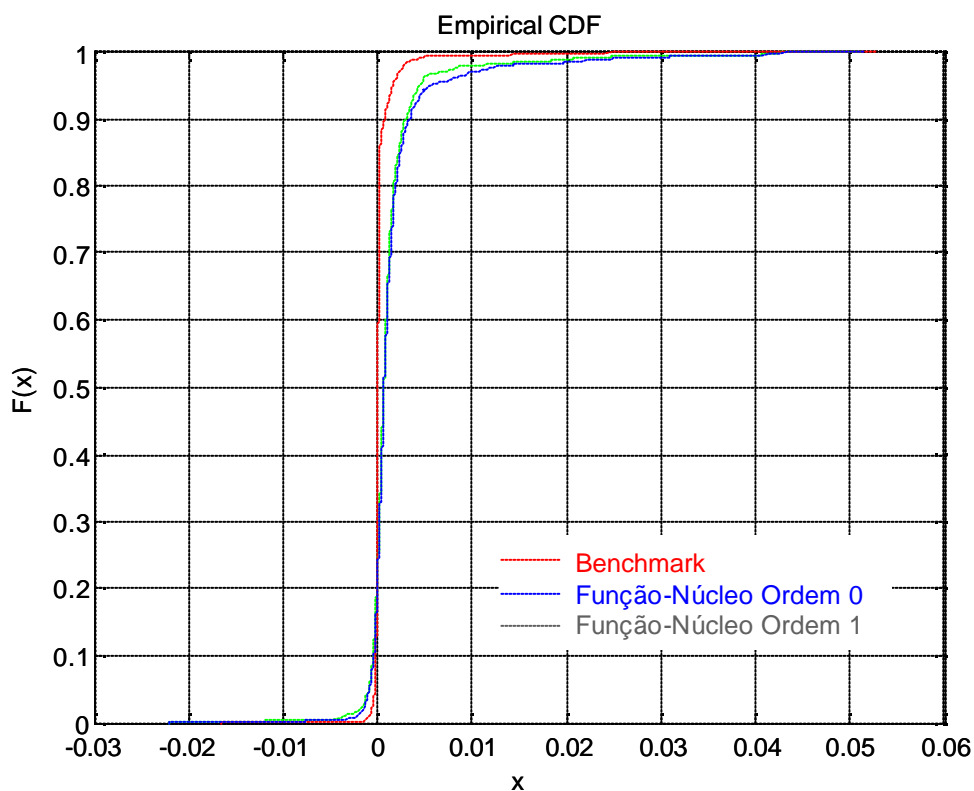
Tabela 11 – Estatística da distribuição de ganho médio por negócio na Regressão por Função-Núcleo de ordem 1

Modelo	Ingênuo	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1	Função Núcleo ordem 1
Taxa de barreira	0,05	0,05	0,05	0,05	0,05	0,05
Variáveis Independ.	AR(1)	AR(1)	AR(1) horário	AR(1) MACD RSI	AR(1) MACD RSI horário	MACD RSI horário
Max	5,27%	5,14%	5,14%	10,27%	4,51%	4,35%
Q3	0,02%	0,14%	0,10%	0,05%	0,05%	0,04%
Mediana	0,00%	0,07%	0,03%	0,01%	0,01%	0,01%
Q1	-0,01%	0,01%	-0,02%	-0,03%	-0,03%	-0,02%
Min	-1,66%	-2,22%	-3,05%	-5,08%	-5,08%	-3,92%
Qtd negócios	18676	3470	3564	4352	3778	4773
Média	0,04%	0,13%	0,09%	0,02%	0,01%	0,01%
Desv pad	0,25%	0,45%	0,43%	0,36%	0,32%	0,23%
Assimetria	1319	638	507	592	41	93
Curtose	20820	5865	5265	24883	8982	11901

As distribuições de ganho médio do modelo de Regressão com Função-Núcleo de ordem 1 em geral assemelham-se às distribuições de resultados do modelo de Regressão com Função-Núcleo de ordem 0. A curtose nos retornos de todos os modelos foi bastante alta, sugerindo uma grande concentração de resultados próximos a zero e a existência de observações extremas com mais frequência do que seria esperado para uma distribuição normal. Esta característica associada a uma forte assimetria positiva explicou o retorno esperado, ou médio, superior à sua mediana.

Na Figura 1, são demonstradas as distribuições cumulativas empíricas de retornos dos modelos com taxa de barreira de 0.05. A comparação quantil a quantil destas distribuições permite concluir pela superioridade, em mais de 90% do espectro da distribuição, dos resultados do modelo de Regressão com Função-Núcleo ordem 0 em relação aos resultados do modelo de Regressão com Função-Núcleo ordem 1, e mais ainda em relação aos resultados do modelo linear.

Figura 1 – Distribuições cumulativas empíricas de retornos.



2.3. Análise de sensibilidade à taxa de barreira

As fortes assimetria e curtose presentes na maioria dos resultados apresentados sugerem que o uso de diferentes taxa de barreira possa produzir resultados significativamente diferentes.

Portanto, foram selecionados os modelos ingênuo, Regressão com Função-Núcleo de ordem 0 com AR(1), e Regressão com Função-Núcleo de ordem 1 com AR(1) para a realização de simulações com taxa de barreiras diferentes. Em princípio, taxas de barreira tão baixas como 0 e altas como 1 não fariam muito sentido para a realização de estratégias de compra e venda de ações, pois ocasionariam praticamente nenhuma ou uma quantidade excessiva de ordens de compra. Estes extremos poderão ser verificados na medida em que a distribuição do retorno projetado na amostra de validação for semelhante à distribuição de retornos projetados na amostra de estimação. Assim sendo, taxa de barreiras extremas foram simulados e seus resultados colocados nas tabelas de comparação apenas com o objetivo de se avaliar a sensibilidade a este parâmetro em todo seu espectro.

Tabela 12 – Sensibilidade do ganho médio por pregão do Modelo Ingênuo

Modelo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Média
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	
csna3_05min	+0,06%	+0,47%	+0,58%	+0,56%	+0,44%	+0,35%	+0,06%	+0,36%
csna3_20min	-0,07%	+0,79%	+0,97%	+1,33%	+1,27%	+1,29%	-0,18%	+0,77%
ggbr4_05min	+0,04%	+0,41%	+0,50%	+0,52%	+0,42%	+0,30%	-0,05%	+0,30%
ggbr4_20min	+0,04%	+0,82%	+1,13%	+1,24%	+1,55%	+1,58%	-0,14%	+0,89%
petr4_05min	+0,05%	+0,20%	+0,26%	+0,27%	+0,23%	+0,20%	+0,01%	+0,17%
tnlp4_05min	+0,01%	+0,60%	+0,78%	+0,84%	+0,65%	+0,36%	+0,11%	+0,48%
usim5_05min	+0,06%	+0,24%	+0,44%	+0,62%	+0,49%	+0,41%	+0,08%	+0,34%
usim5_20min	+0,00%	+0,22%	+0,33%	+0,54%	+0,55%	+0,52%	-0,17%	+0,28%
vale5_05min	+0,01%	+0,42%	+0,47%	+0,42%	+0,34%	+0,22%	-0,04%	+0,26%
vale5_20min	-0,10%	+0,62%	+0,80%	+0,98%	+1,05%	+1,00%	-0,26%	+0,58%
média	+0,01%	+0,48%	+0,63%	+0,73%	+0,70%	+0,62%	-0,06%	

Conforme a tabela 12 demonstra, a taxa de barreira de 25% foi a medida que produziu, na média, os melhores resultados médios por pregão. Como era esperado, o desempenho cai à medida que taxa de barreira ultrapassam 50%, pois

nestas condições espera-se um número de ordens de compra superior ao número esperado de retornos positivos no mercado, que seria próximo a 50%.

Tabela 13 – Sensibilidade da média de negócios realizados por pregão do Modelo Ingênuo

Modelo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	14,75	17,62	20,28	27,80	39,39	55,04	68,41	34,76
csna3_20min	4,25	4,93	5,32	7,23	10,06	14,75	17,49	9,15
ggbr4_05min	17,35	19,36	21,35	29,05	40,01	55,42	66,32	35,55
ggbr4_20min	4,08	4,56	4,96	7,02	10,30	15,21	17,66	9,11
petr4_05min	25,07	26,12	27,14	32,07	40,55	51,33	58,11	37,20
tnlp4_05min	5,60	7,55	10,43	20,85	39,59	62,92	78,11	32,15
usim5_05min	17,42	18,54	20,44	27,36	39,82	56,02	65,94	35,08
usim5_20min	5,76	6,09	6,46	7,82	10,46	13,88	15,96	9,49
vale5_05min	21,30	24,00	25,87	32,36	41,03	51,47	62,47	36,93
vale5_20min	5,34	5,90	6,23	7,70	10,13	13,76	16,41	9,35
média	12,09	13,47	14,85	19,93	28,13	38,98	46,69	

O número médio de negócios foi sensível, mas em pouca medida, à variação da taxa de barreira. Este efeito se deve à diferença na distribuição de previsões de retorno no período de estimação e de validação. Afinal, a marca a ser superada pela previsão de retorno no período de validação para que uma ordem de compra seja disparada é função da taxa de barreira e da distribuição de retornos previstos no período de estimação.

Vale destacar que a taxa de barreira mais restritiva possível permitiu um grande número de ordens disparadas e a taxa de barreira menos restritiva possível ainda conteve um grande número de ordens.

Tabela 14 – Sensibilidade do ganho médio por negócio do Modelo Ingênuo

Modelo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	0,00%	0,03%	0,03%	0,02%	0,01%	0,01%	0,00%	0,01%
csna3_20min	-0,02%	0,16%	0,18%	0,18%	0,13%	0,09%	-0,01%	0,08%
ggbr4_05min	0,00%	0,02%	0,02%	0,02%	0,01%	0,01%	0,00%	0,01%
ggbr4_20min	0,01%	0,18%	0,23%	0,18%	0,15%	0,10%	-0,01%	0,10%
petr4_05min	0,00%	0,01%	0,01%	0,01%	0,01%	0,00%	0,00%	0,00%
tnlp4_05min	0,00%	0,08%	0,08%	0,04%	0,02%	0,01%	0,00%	0,02%
usim5_05min	0,00%	0,01%	0,02%	0,02%	0,01%	0,01%	0,00%	0,01%
usim5_20min	0,00%	0,04%	0,05%	0,07%	0,05%	0,04%	-0,01%	0,03%
vale5_05min	0,00%	0,02%	0,02%	0,01%	0,01%	0,00%	0,00%	0,01%
vale5_20min	-0,02%	0,11%	0,13%	0,13%	0,10%	0,07%	-0,02%	0,06%
Média	0,00%	0,04%	0,04%	0,04%	0,02%	0,02%	0,00%	

No Modelo Ingênuo, as taxas de barreira menos restritivas, entre 0% e 25%, não penalizaram o ganho médio por negócio. Porém, conforme a Tabela 14, taxas de barreira maiores aumentam de sobremaneira a quantidade de negócios, diluindo o ganho médio por negócio.

Tabela 15 – Sensibilidade da estatística da distribuição do ganho médio por negócio do Modelo Ingênuo

Modelo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)
Max	0,19%	5,27%	4,51%	2,98%	1,49%	0,88%	0,45%
Q3	0,01%	0,02%	0,02%	0,04%	0,03%	0,02%	0,01%
Mediana	0,00%	0,00%	0,01%	0,01%	0,01%	0,01%	0,00%
Q1	-0,01%	-0,01%	0,00%	0,00%	0,00%	0,00%	-0,01%
Min	-0,13%	-1,66%	-1,66%	-0,60%	-1,11%	-0,71%	-0,46%
Qtd negócios	16775	18676	20583	27613	38974	53994	64657
Média	0,00%	0,04%	0,04%	0,04%	0,03%	0,02%	0,00%
Desv pad	0,02%	0,25%	0,23%	0,14%	0,10%	0,06%	0,03%
Assimetria	81	1319	1169	821	494	487	-236
Curtose	2017	20820	17592	9706	5490	4814	6212

Nesta análise percebe-se que a distribuição de resultados por negócio é sensível à variação da taxa de barreira. Apenas à exceção da simulação com taxa de barreira de 0, percebe-se que tanto a variância quanto a curtose e a assimetria decresceram em função do aumento da taxa de barreira.

Tabela 16 – Sensibilidade do ganho médio por pregão na Regressão com Função-Núcleo de ordem 0

Modelo	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	+0,13%	+0,33%	+0,37%	+0,35%	+0,28%	+0,21%	+0,15%	+0,26%
csna3_20min	+0,56%	+0,82%	+1,05%	+1,10%	+0,97%	+1,09%	+0,31%	+0,84%
ggbr4_05min	+0,12%	+0,35%	+0,41%	+0,37%	+0,21%	+0,10%	-0,01%	+0,22%
ggbr4_20min	+0,28%	+0,65%	+0,96%	+1,19%	+1,17%	+1,37%	+0,80%	+0,92%
petr4_05min	+0,03%	+0,13%	+0,15%	+0,14%	+0,12%	+0,07%	+0,06%	+0,10%
tnlp4_05min	+0,10%	+0,56%	+0,74%	+0,73%	+0,46%	+0,28%	+0,14%	+0,43%
usim5_05min	+0,02%	+0,20%	+0,36%	+0,43%	+0,32%	+0,22%	+0,14%	+0,24%
usim5_20min	+0,10%	+0,18%	+0,30%	+0,26%	+0,38%	+0,44%	+0,18%	+0,27%
vale5_05min	+0,12%	+0,29%	+0,32%	+0,25%	+0,26%	+0,18%	+0,00%	+0,20%
vale5_20min	+0,43%	+0,54%	+0,68%	+0,95%	+1,15%	+0,98%	+0,60%	+0,76%
média	+0,19%	+0,40%	+0,53%	+0,58%	+0,53%	+0,49%	+0,24%	

Segundo a Tabela 16, o ganho médio diário mais alto é verificado na taxa de barreira de 25%. Percebemos que os resultados continuaram positivos mesmo quando este parâmetro chegou à 100%. Isto indica que as poucas ordens de compra que deixaram de ser disparadas na simulação 7 seguramente contém uma grande concentração de retornos negativos.

Tabela 17 – Sensibilidade da média de negócios realizados por pregão na Regressão com Função-Núcleo de ordem 0

Modelo	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	0,81	3,76	8,44	21,46	41,53	61,51	82,45	31,42
csna3_20min	0,61	0,98	1,73	4,57	10,72	16,96	21,12	8,10
ggbr4_05min	0,81	3,72	8,12	20,23	41,67	62,63	83,10	31,47
ggbr4_20min	0,49	0,90	1,70	4,78	10,54	17,14	21,19	8,10
petr4_05min	0,52	3,13	6,94	19,79	41,29	64,19	82,71	31,22
tnlp4_05min	0,29	2,32	5,98	18,36	40,54	63,35	83,32	30,59
usim5_05min	0,22	2,01	5,21	17,65	39,36	63,65	83,03	30,16
usim5_20min	0,33	0,65	1,36	4,18	10,89	16,73	21,26	7,91
vale5_05min	1,44	6,15	11,24	23,08	42,95	62,47	82,56	32,84
vale5_20min	0,58	0,91	1,70	4,65	10,67	16,78	21,05	8,05
média	0,61	2,45	5,24	13,88	29,02	44,54	58,18	

O número médio de negócios segue com bastante aderência o volume esperado para cada taxa de barreira. Séries de preços amostradas de 5 em 5 minutos apresentam em média 84 períodos por pregão e séries amostradas de 20 em 20 minutos apresentam em média 21 períodos por pregão. Este leve descolamento residual entre a quantidade média de negócios esperada e a quantidade média de negócios efetivamente disparada não existiria se fosse possível utilizar uma marca de retorno mínimo em função da taxa de barreira e das projeções de retorno no período de validação, mas neste caso seria violada a premissa de usar apenas a informação disponível até $t-1$ para se disparar ou não ordens de compras em t .

Tabela 18 – Sensibilidade do ganho médio por negócio na Regressão com Função-Núcleo de ordem 0

Modelo	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Média
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	
csna3_05min	0,16%	0,09%	0,04%	0,02%	0,01%	0,00%	0,00%	0,01%
csna3_20min	0,92%	0,84%	0,61%	0,24%	0,09%	0,06%	0,01%	0,10%
ggbr4_05min	0,15%	0,09%	0,05%	0,02%	0,00%	0,00%	0,00%	0,01%
ggbr4_20min	0,56%	0,72%	0,56%	0,25%	0,11%	0,08%	0,04%	0,11%
petr4_05min	0,05%	0,04%	0,02%	0,01%	0,00%	0,00%	0,00%	0,00%
tnlp4_05min	0,35%	0,24%	0,12%	0,04%	0,01%	0,00%	0,00%	0,01%
usim5_05min	0,11%	0,10%	0,07%	0,02%	0,01%	0,00%	0,00%	0,01%
usim5_20min	0,30%	0,28%	0,22%	0,06%	0,04%	0,03%	0,01%	0,03%
vale5_05min	0,08%	0,05%	0,03%	0,01%	0,01%	0,00%	0,00%	0,01%
vale5_20min	0,75%	0,59%	0,40%	0,20%	0,11%	0,06%	0,03%	0,09%
média	0,31%	0,16%	0,10%	0,04%	0,02%	0,01%	0,00%	

De acordo com a Tabela 18, o ganho médio por negócio apresentou-se tão mais alto quanto menor a taxa de barreira. Esta característica é muito conveniente se quisermos calibrar um ganho médio em função de determinados custos de transação.

Uma taxa de barreira mais restritiva resultou em maiores ganhos médios por negócio, mas também provocou um aumento do desvio padrão da distribuição de resultados. Além disto, é grande a sensibilidade da curtose a este parâmetro. Este fato, associado à variação do resultado esperado, fez com que um aumento do

desvio padrão viesse acompanhado da redução da frequência de resultados negativos.

Tabela 19 – Sensibilidade da estatística da distribuição do ganho médio por negócio na Regressão com Função-Núcleo de ordem 0

Modelo	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0	Função-núcleo ordem 0
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)
Max	5,14%	5,14%	4,59%	2,80%	2,22%	1,34%	0,76%
Q3	0,30%	0,16%	0,10%	0,04%	0,02%	0,01%	0,01%
Mediana	0,13%	0,07%	0,04%	0,02%	0,01%	0,00%	0,00%
Q1	0,02%	0,01%	0,00%	0,00%	-0,01%	-0,01%	-0,01%
Min	-2,22%	-2,22%	-2,22%	-1,66%	-0,70%	-0,42%	-0,46%
Qtd negócios	843	3390	7249	19214	40194	61704	80597
Média	0,31%	0,17%	0,10%	0,04%	0,02%	0,01%	0,00%
Desv pad	0,79%	0,49%	0,36%	0,17%	0,10%	0,06%	0,04%
Assimetria	322	574	698	622	613	538	569
Curtose	1710	4567	6996	6199	8023	7203	8675

Na comparação quartil a quartil encontramos dominância da simulação com taxa de barreira 0,00 sobre as demais.

Tabela 20 – Sensibilidade do ganho médio por pregão na Regressão com Função-Núcleo de ordem 1

Modelo	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	+0,13%	+0,28%	+0,31%	+0,33%	+0,26%	+0,25%	+0,15%	+0,25%
csna3_20min	+0,08%	+0,29%	+0,62%	+1,06%	+1,18%	+1,24%	+0,33%	+0,69%
ggbr4_05min	+0,11%	+0,37%	+0,40%	+0,36%	+0,22%	+0,20%	-0,01%	+0,23%
ggbr4_20min	+0,28%	+0,64%	+0,97%	+1,19%	+1,01%	+1,47%	+0,80%	+0,91%
petr4_05min	+0,03%	+0,12%	+0,14%	+0,13%	+0,10%	+0,08%	+0,05%	+0,09%
tnlp4_05min	+0,11%	+0,55%	+0,74%	+0,68%	+0,48%	+0,29%	+0,14%	+0,43%
usim5_05min	+0,01%	+0,18%	+0,31%	+0,38%	+0,30%	+0,19%	+0,14%	+0,22%
usim5_20min	+0,04%	+0,14%	+0,30%	+0,45%	+0,48%	+0,52%	+0,21%	+0,31%
vale5_05min	+0,09%	+0,28%	+0,29%	+0,27%	+0,25%	+0,12%	-0,02%	+0,18%
vale5_20min	+0,43%	+0,52%	+0,68%	+0,94%	+0,92%	+0,85%	+0,43%	+0,68%
média	+0,13%	+0,34%	+0,48%	+0,58%	+0,52%	+0,52%	+0,22%	

Tal como nos resultados do modelo de Regressão com Função-Núcleo de ordem 0, os resultados do modelo de Regressão com Função-Núcleo de ordem 1 foram mais altos quando a taxa de barreira utilizada foi 25%, e apresentaram apenas leve decaimento para taxa de barreiras ainda maiores.

Tabela 21 – Sensibilidade da média de negócios realizados por pregão na Regressão com Função-Núcleo de ordem 1

Modelo	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	0,84	4,09	8,46	21,02	41,42	62,16	82,42	31,49
csna3_20min	0,66	1,23	2,17	5,06	10,51	17,10	21,13	8,27
ggbr4_05min	0,80	3,79	7,96	19,81	42,05	62,31	83,11	31,41
ggbr4_20min	0,50	0,77	1,57	4,79	10,66	16,91	21,18	8,06
petr4_05min	0,59	3,21	6,59	19,14	41,29	62,97	82,57	30,91
tnlp4_05min	0,32	2,38	5,75	18,63	40,28	62,93	83,25	30,50
usim5_05min	0,24	2,04	5,09	17,21	40,02	62,61	82,94	30,02
usim5_20min	0,53	0,90	2,16	4,73	10,48	16,86	21,25	8,13
vale5_05min	1,42	5,79	10,92	22,77	42,53	62,13	82,61	32,59
vale5_20min	0,59	0,89	1,63	4,61	10,52	16,75	20,86	7,98
média	0,65	2,51	5,23	13,78	28,98	44,27	58,13	

Similar aos resultados da Regressão com Função-Núcleo de ordem 0, o número médio de negócios realizados por pregão aumenta na proporção em que se utiliza uma taxa de barreira menos restritiva.

Tabela 22 – Sensibilidade do ganho médio por negócio na Regressão com Função-Núcleo de ordem 1

Modelo	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	Média
csna3_05min	0,15%	0,07%	0,04%	0,02%	0,01%	0,00%	0,00%	0,01%
csna3_20min	0,12%	0,24%	0,29%	0,21%	0,11%	0,07%	0,02%	0,08%
ggbr4_05min	0,13%	0,10%	0,05%	0,02%	0,01%	0,00%	0,00%	0,01%
ggbr4_20min	0,56%	0,82%	0,61%	0,25%	0,10%	0,09%	0,04%	0,11%
Petr4_05min	0,05%	0,04%	0,02%	0,01%	0,00%	0,00%	0,00%	0,00%
Tnlp4_05min	0,35%	0,23%	0,13%	0,04%	0,01%	0,00%	0,00%	0,01%
usim5_05min	0,06%	0,09%	0,06%	0,02%	0,01%	0,00%	0,00%	0,01%
usim5_20min	0,08%	0,15%	0,14%	0,09%	0,05%	0,03%	0,01%	0,04%

Vale5_05min	0,06%	0,05%	0,03%	0,01%	0,01%	0,00%	0,00%	0,01%
Vale5_20min	0,74%	0,59%	0,42%	0,20%	0,09%	0,05%	0,02%	0,09%
média	0,20%	0,13%	0,09%	0,04%	0,02%	0,01%	0,00%	

Também como na Função-Núcleo de ordem 0, na Tabela 22 o ganho médio por negócio aumenta quanto mais restritiva for a taxa de barreira. Isto mostra que as ordens de compra que são mantidas à medida que se aumenta a marca a ser atingida na previsão de retorno, são na média, as ordens que produzem os melhores retornos para a carteira.

Tabela 23 – Sensibilidade da estatística da distribuição do ganho médio por negócio na Regressão com Função-Núcleo de ordem 1

Modelo	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1	Função-núcleo ordem 1
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00
Variáveis Independ.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)
Max	5,14%	5,14%	4,59%	4,40%	1,94%	1,07%	0,76%
Q3	0,24%	0,14%	0,09%	0,04%	0,02%	0,01%	0,01%
Mediana	0,09%	0,07%	0,04%	0,02%	0,01%	0,00%	0,00%
Q1	-0,01%	0,01%	0,00%	0,00%	-0,01%	-0,01%	-0,01%
Min	-2,41%	-2,22%	-2,22%	-1,41%	-0,70%	-0,41%	-0,46%
qtd negócios	896	3470	7238	19079	40140	61337	80534
Média	0,20%	0,13%	0,09%	0,04%	0,02%	0,01%	0,00%
desv pad	0,67%	0,45%	0,35%	0,18%	0,10%	0,07%	0,04%
Assimetria	375	638	756	777	626	613	550
Curtose	2542	5865	8142	10889	8663	6832	8615

Conforme esperado, a variação das medidas de dispersão de segundo e quarto momentos do resultado médio por negócio revelou-se proporcional à variação do parâmetro taxa de barreira.

A escolha deste parâmetro pode ser orientada tanto pelo retorno mínimo necessário para cobrir custos de transação, como pelo objetivo de risco a ser perseguido.

3. Considerações Finais

Este estudo aplicou regressões usando funções-núcleo no mercado intradiário de ações no Brasil. Seus resultados comparados a um benchmark tão citado na literatura são animadores e estimulam tanto o seu uso prático quanto a realização de pesquisas complementares.

O método de cálculo e avaliação de resultado aqui empregado é robusto para aplicação a outros ativos do mercado brasileiro, e também a outras taxas de amostragem intradiária. Uma eventual aplicação a séries ainda mais longas de tempo permitiria observar e tirar conclusões interessantes sobre o comportamento dos parâmetros otimizados k e N ao longo do tempo.

Muitos trabalhos promissores vêm sendo desenvolvido com séries intradiárias no sentido de estabelecer escalas de tempo não cronológicas, mas de outra natureza, como de volume negócios e afins. Zumbach (1998) demonstra o potencial desta técnica para se melhorar a eficiência das regressões e também estimula novos trabalhos nesta área.

Outra pesquisa complementar que pode ser desenvolvida à medida que mais dados estejam disponíveis é o uso de informações intradiárias tais como volume e quantidade de negócios, ou operações, observados no mercado. Isto possibilitaria o uso de tais informações diretamente nas regressões ou através de outros dos muitos indicadores técnicos hoje em uso no mercado.

Por fim, a extensão deste estudo de mais longo alcance seguramente passa por incluir variáveis independentes calculadas a partir de indicadores econômicos e outros ativos que não a própria variável dependente realizada. O uso de regressões usando funções-núcleo em muito poderia contribuir com Vetores Autoregressivos e outras técnicas econométricas já muito conhecidas e em uso tanto na academia como no mercado.

4. Conclusão

Não foram encontradas evidências da utilização prática e sistemática de regras de compra e venda de ações no mercado intradiário brasileiro segundo as recomendações de um modelo de autoregressão ou de regressão de indicadores técnicos. Espera-se que à medida que isto aconteça, oportunidades de ganhos como as identificadas neste trabalho tendam a desaparecer. Naturalmente, os custos de transação são um obstáculo permanente nesta trajetória, e a magnitude deles é proporcional à probabilidade de rejeição da hipótese de mercados eficientes.

Como as evidências empíricas aqui produzidas não se apoiaram em premissas de custos de transação, podemos comparar os modelos sem a influência delas.

Em uma comparação dos quartis 2 e 3 da distribuição de ganhos médios por negócio simulado, foi encontrada dominância dos modelos não lineares que utilizaram apenas o retorno defasado em um período como variável independente. Entre os modelos não lineares, o modelo de Regressão usando Função-Núcleo de ordem 0 teve performance ligeiramente melhor que o de ordem 1.

Alguma similaridade de resultados entre o modelo linear benchmark, ou ingênuo, e o modelo de Regressão usando Função-Núcleo de ordem 0 com AR1 era esperada, pois a variável independente de ambos era a mesma. No entanto, um volume muito mais parcimonioso de ordens de compra disparadas pelo modelo não linear permitiu que seus retornos fossem negativos em menos de $\frac{1}{4}$ dos negócios disparados.

A sensibilidade à taxa de barreira foi diferente entre os modelos linear e não linear. O ganho médio por negócio e suas medidas de dispersão do Modelo Ingênuo apresentaram sensibilidade relevante, mas não necessariamente linear, a este parâmetro.

Por outro lado, os modelos não lineares reagiram de forma relativamente previsível à mudança da taxa de barreira. Um parâmetro mais baixo produziu um

volume menor de negócios, mas com uma concentração maior de negócios lucrativos. A variância nominal dos resultados por negócio destas simulações também aumentou de forma proporcional. Esta calibragem representou portanto um uma escolha entre risco e retorno para o portfólio, e a reação “bem comportada” dos resultados do modelo faz dela uma ferramenta muito conveniente para uma eventual aplicação real no mercado financeiro.

Referências bibliográficas

BAPTISTA, R. F. de F. **Avaliação da Performance de Regras de Análise Técnica no Mercado Intradiário do Futuro de Índice Bovespa**. São Paulo, 2002. Dissertação de Mestrado em Estatística, IME-USP, 2002

BAPTISTA, R. F. de F. e VALLS PEREIRA, P.L., **Avaliação da Performance de Regras de Análise Técnica no Mercado Intradiário do Futuro de Índice Bovespa**. 2006

Bloomberg Magazine. **Measuring the Market With the Relative Strength Index**. Jan 1995

BROCK, W.; LAKONISHOK, J.; LEBARON, B. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, **Journal of Finance**, 48 , 1731-1764, 1992

DACOROGNA, M. M.; GENÇAY, R.; MÜLLER, U.; OLSEN, R. B.; PICTET, V. P. **An Introduction do High Frequency Finance**, Academic Press, 2001

DOCKNER, E. J.; PRSKAWETZ, A.; FEICHTINGER, G. Non-linear Dynamics and Predictability in the Austrian Stock Market. In: **System Dynamics in Economic and Financial Models**. John Wiley & Sons Inc, 45-71, 1997

LEBARON, B. Forecast Improvements using a Volatility Index. **Journal of Applied Econometrics**, Vol 7, Supplement, S137-S149, 1992

MAKRIDAKIS, S. Sliding Simulation: A New Approach to Time Series Forecasting. **Management Science**, Vol. 36, No. 4, 505-512, 1990

MIZRACH, B. Multivariate Nearest-Neighbour Forecasts for EMS Exchange Rates. **Journal of Applied Econometrics**, 7, S151-S163, 1992

_____. Moving **Average Convergence / Divergence**, em
<www.StockCharts.com>. Acessado em Nov 2006

SAFFI, P.A.C. Análise Técnica: Sorte ou Realidade? **Revista Brasileira de Economia**, 4, 953-974, 2003

SULLIVAN, R.; TIMMERMANN, A.; WHITE, H. Data-Snooping, Technical Trading Rules and the Bootstrap, **Journal of Finance**, 54, 1647-1692, 1999

WHITE, H. A Reality Check for Data Snooping, **Econometrica** 68, 1097-1126. 2000

WOLBERG, J. R. **Expert Trading Systems: Modeling Financial Markets with Kernel Regressions**. John Wiley & Sons Inc, 2000

ZUMBACH, G. O. Considering Time as the Random Variable. **Neutral Network World**, 243-253, 1998

Apêndice 1 - Marcas a superar para se efetivar ordem de compra

A marca de retorno a ser superada para que seja disparada uma ordem de compra é função da taxa de barreira e da distribuição de projeções de retorno produzidas no período de estimação. As tabelas abaixo mostram tais marcas para cada uma das diferentes simulações.

Simulações com Modelo Ingênuo:

Modelo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	Ingênuo	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Var. Indep.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	média
csna3_05min	0,15%	0,02%	0,01%	0,01%	0,00%	0,00%	-0,11%	-0,02%
csna3_20min	4,47%	0,59%	0,23%	0,03%	-0,02%	-0,07%	-3,89%	-0,72%
ggbr4_05min	0,32%	0,04%	0,03%	0,01%	0,00%	-0,01%	-0,36%	-0,07%
ggbr4_20min	3,14%	0,71%	0,30%	0,04%	-0,01%	-0,06%	-3,01%	-0,57%
petr4_05min	0,17%	0,02%	0,01%	0,01%	0,00%	0,00%	-0,17%	-0,02%
tnlp4_05min	0,53%	0,09%	0,05%	0,02%	0,00%	-0,02%	-0,64%	-0,19%
usim5_05min	0,54%	0,08%	0,05%	0,02%	0,00%	-0,02%	-0,49%	-0,08%
usim5_20min	2,36%	0,40%	0,18%	0,04%	0,00%	-0,06%	-2,07%	-0,24%
vale5_05min	0,14%	0,02%	0,01%	0,00%	0,00%	0,00%	-0,12%	-0,01%
vale5_20min	2,78%	0,40%	0,19%	0,03%	-0,01%	-0,04%	-2,68%	-0,40%
Média	0,73%	0,11%	0,06%	0,01%	0,00%	-0,02%	-0,71%	

Simulações com modelo de Regressão com Função-Núcleo de ordem 0:

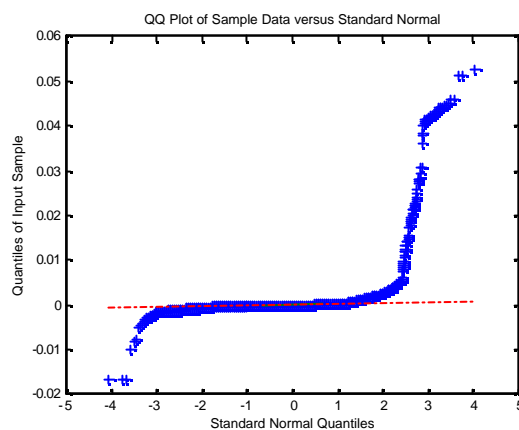
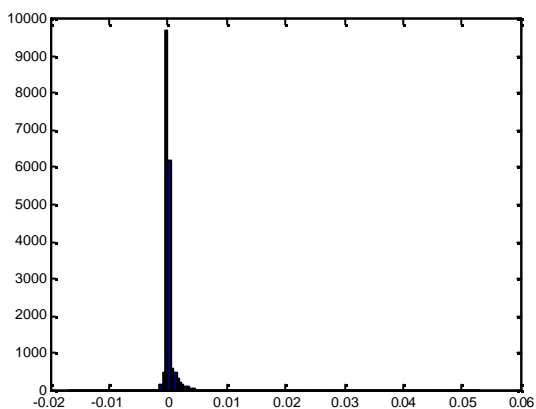
Modelo	Função- Núcleo ordem 0	Função- Núcleo ordem 0	Função- Núcleo ordem 0	Função- Núcleo ordem 0	Função- Núcleo ordem 0	Função- Núcleo ordem 0	Função- Núcleo ordem 0	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Var. Indep.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	média
csna3_05min	0,07%	0,03%	0,01%	0,01%	0,00%	-0,01%	-0,04%	-0,02%
csna3_20min	0,19%	0,17%	0,13%	0,02%	-0,01%	-0,03%	-0,08%	-0,03%
ggbr4_05min	0,13%	0,05%	0,02%	0,00%	0,00%	-0,01%	-0,08%	-0,03%
ggbr4_20min	0,23%	0,20%	0,15%	0,03%	0,00%	-0,04%	-0,16%	-0,06%
petr4_05min	0,06%	0,02%	0,01%	0,00%	0,00%	-0,01%	-0,04%	-0,02%
tnlp4_05min	0,31%	0,09%	0,02%	-0,01%	-0,01%	-0,02%	-0,09%	-0,04%
usim5_05min	0,20%	0,08%	0,03%	0,00%	-0,01%	-0,02%	-0,12%	-0,05%
usim5_20min	0,24%	0,22%	0,16%	0,03%	-0,01%	-0,02%	-0,17%	-0,06%
vale5_05min	0,04%	0,02%	0,01%	0,00%	0,00%	-0,01%	-0,05%	-0,02%
vale5_20min	0,15%	0,13%	0,09%	0,02%	0,00%	-0,01%	-0,07%	-0,02%
Média	0,13%	0,06%	0,03%	0,01%	-0,01%	-0,02%	-0,08%	

Simulações com modelo de Regressão com Função-Núcleo de ordem 1:

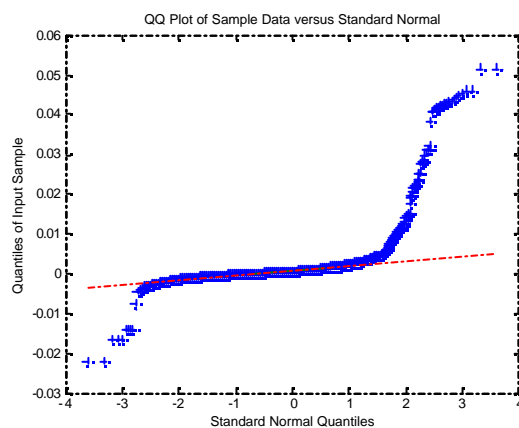
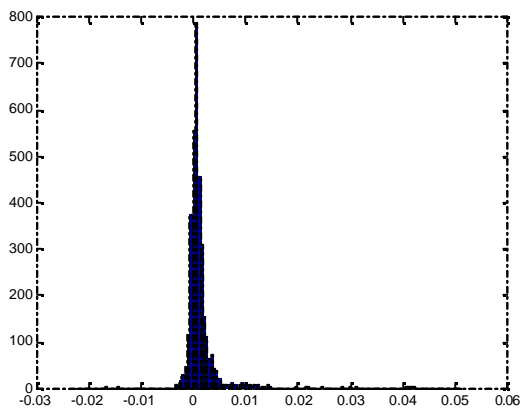
Modelo	Função- Núcleo ordem 1	Função- Núcleo ordem 1	Função- Núcleo ordem 1	Função- Núcleo ordem 1	Função- Núcleo ordem 1	Função- Núcleo ordem 1	Função- Núcleo ordem 1	
Taxa de barreira	0,00	0,05	0,10	0,25	0,50	0,75	1,00	
Var. Indep.	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	AR(1)	média
csna3_05min	0,07%	0,03%	0,02%	0,01%	0,00%	-0,01%	-0,06%	-0,02%
csna3_20min	0,32%	0,29%	0,23%	0,10%	0,00%	-0,02%	-0,47%	-0,15%
ggbr4_05min	0,13%	0,05%	0,02%	0,01%	0,00%	-0,01%	-0,09%	-0,03%
ggbr4_20min	1,37%	1,15%	0,40%	0,06%	0,00%	-0,04%	-0,87%	-0,29%
petr4_05min	0,06%	0,02%	0,01%	0,01%	0,00%	-0,01%	-0,07%	-0,03%
tnlp4_05min	0,31%	0,10%	0,02%	0,00%	-0,01%	-0,03%	-0,10%	-0,05%
usim5_05min	0,20%	0,08%	0,03%	0,01%	-0,01%	-0,02%	-0,14%	-0,06%
usim5_20min	0,54%	0,49%	0,38%	0,17%	0,00%	-0,04%	-0,67%	-0,22%
vale5_05min	0,04%	0,02%	0,01%	0,01%	0,00%	-0,01%	-0,06%	-0,02%
vale5_20min	0,89%	0,65%	0,21%	0,03%	0,00%	-0,03%	-0,32%	-0,10%
Média	0,33%	0,13%	0,06%	0,02%	0,00%	-0,02%	-0,16%	

Apêndice 2 - Distribuições de retornos

Distribuição de retornos produzidos pelo Modelo Ingênuo com taxa de barreira de 0,05.



Distribuição de retornos produzidos pelo modelo de Regressão com Função-Núcleo de ordem 0 com AR(1) como variável independente e taxa de barreira de 0,05.



Distribuição de retornos produzidos pelo modelo de Regressão com Função-Núcleo de ordem 1 com AR(1) como variável independente e taxa de barreira de 0,05.

