



Marina Sequeiros Dias

**O uso de Máquina de Suporte Vetorial para
Regressão (SVR) na Estimação da Estrutura a
Termo da Taxa de Juros do Brasil**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática da PUC-Rio

Orientador : Prof. Hélio Côrtes Vieira Lopes

Co-Orientador: Prof. Luciano Vereda Oliveira

Rio de Janeiro
Fevereiro de 2007



Marina Sequeiros Dias

**O uso de Máquina de Suporte Vetorial para
Regressão (SVR) na Estimação da Estrutura a
Termo da Taxa de Juros do Brasil**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Hélio Côrtes Vieira Lopes

Orientador

Departamento de Matemática — PUC-Rio

Prof. Luciano Vereda Oliveira

Co-Orientador

IAPUC — PUC-Rio

Prof. Cristiano Augusto C. Fernandes

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Marco Antônio Grivet Mattoso Maia

Centro de Estudos em Telecomunicações – PUC-Rio

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 27 de Fevereiro de 2007

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marina Sequeiros Dias

Graduou-se em Licenciatura em Matemática na UERJ (Universidade do Estado do Rio de Janeiro) em 2004.

Ficha Catalográfica

Dias, Marina Sequeiros

O uso de Máquina de Suporte Vetorial para Regressão (SVR) na Estimação da Estrutura a Termo da Taxa de Juros do Brasil / Marina Sequeiros Dias; orientador: Hélio Côrtes Vieira Lopes; co-orientador: Luciano Vereda Oliveira. — Rio de Janeiro : PUC-Rio, Departamento de Matemática, 2007.

v., 77 f: il. ; 29,7 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática.

Inclui referências bibliográficas.

1. Matemática – Tese. 2. Séries Temporais. 3. Modelos de previsão. 4. Taxa de Juros. 5. Máquina de Suporte Vetorial para Regressão. 6. Aprendizagem estatística. I. Lopes, Hélio Côrtes Vieira. II. Oliveira, Luciano Vereda. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. IV. Título.

CDD: 510

Agradecimentos

À Deus, em primeiro lugar.

Aos professores Hélio Côrtes Vieira Lopes e Luciano Vereda Oliveira pela orientação.

Aos meus pais, Benjamin e Maria das Dores, por colaborarem e incentivarem os meus estudos.

Ao meu namorado Eduardo pelo seu companherismo e apoio aos meus estudos e projetos.

Ao meu irmão Adriano por acompanhar e ajudar nos meus estudos.

À Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) pela oportunidade de cursar o Mestrado e pela bolsa de isenção.

Ao Iapuc pelo apoio.

A CAPES e a FAPERJ pela concessão da bolsa de mestrado.

Aos meus amigos, por estarem sempre participando de todos os momentos de minha vida, em especial, aos amigos que fiz na PUC.

Às secretárias Creuza Nascimento e Katia Beatriz Aguiar pelo carinho e aos Auxiliares Administrativos do Departamento de Matemática da PUC-Rio pela colaboração de sempre.

Resumo

Dias, Marina Sequeiros; Lopes, Hélio Côrtes Vieira; Oliveira, Luciano Vereda. **O uso de Máquina de Suporte Vetorial para Regressão (SVR) na Estimação da Estrutura a Termo da Taxa de Juros do Brasil**. Rio de Janeiro, 2007. 77p. Dissertação de Mestrado — Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Nessa dissertação um novo método para previsão da Estrutura a Termo da Taxa de Juros Brasileira - ETTJ brasileira - conhecido como Máquina de Suporte Vetorial para Regressão é investigado, comparando-o com os métodos tradicionais, tais como modelos VAR (Vetor Auto-regressivo) e ECM (Modelos de Correção de Erros). Utiliza-se além dos retornos de títulos de renda fixa, algumas variáveis macro-econômicas, que conforme sugerido no artigo de Evans e Marshall (1998) e verificado para economia brasileira no artigo de Fukuda, Vereda e Lopes (2006) melhoram a previsão dos retornos de títulos de renda fixa no longo prazo. O experimento mostra uma melhora considerável do SVR sobre os modelos tradicionais mencionados no longo prazo, atuando ainda como ótimo indicador da direção das taxas em praticamente todos os horizontes de previsão. Para tal avaliação, foram utilizados os critérios de raiz do erro quadrado médio, erro absoluto médio, simetria direcional e simetria direcional ponderada, correta tendência para cima e correta tendência para baixo além do teste U de Theil, que faz uso da raiz do erro quadrado médio para verificar se ocorre uma melhora significativa de um modelo sobre outro. Uma vez que não existe uma maneira estruturada para escolha dos parâmetros livres do SVR, a escolha dos mesmos foi feita através de uma função do software R, que faz uma pesquisa em um domínio retangular fornecido pelo usuário. A análise dos resultados mostra que SVR é uma técnica promissora para previsão dos retornos de títulos de renda fixa, sugerindo-se ainda melhorar as escolhas dos parâmetros livres do SVR uma vez que os mesmos são meios poderosos de regularização e adaptação do ruído aos dados.

Palavras-chave

Séries Temporais. Modelos de previsão. Taxa de Juros. Máquina de Suporte Vetorial para Regressão. Aprendizagem estatística.

Abstract

Dias, Marina Sequeiros; Lopes, Hélio Côrtes Vieira; Oliveira, Luciano Vereda. **the use of Support Vector Regression (SVR) in Estimating the Brazilian Term Structure of Interest Rates.** Rio de Janeiro, 2007. 77p. MsC Thesis — Department of Mathematics, Pontifícia Universidade Católica do Rio de Janeiro.

In this dissertation a new method for the prediction of the Brazilian Term Structure of Interest Rates - Brazilian ETTJ - known as Support Vector Regression is investigated. This is compared with the traditional methods used in this set up, such as VAR models (Vector Autoregressive) and ECM (Error Correction Models). Besides the interest rates, some macroeconomic variables are also used, as it was suggested in a work from Evans and Marshall(1998) and verified for brazilian economy in a work from Fukuda, Vereda and Lopes (2006), the inclusion of macroeconomic variables can improve the prediction of the interest rates in long term forecasts. The experiment show some improvements in using SVR in the long term in relation to the traditional methods mentioned, acting like a really good predictor of the direction of the interest rates along the short and long term forecasts. To make these assertions, we make use of some tests like the root mean squared error, mean absolute error, directional symmetry and weighted directional symmetry, Correct Up trend and Corret Down trend besides Theil U test, which uses the root mean squared error to verify if there is some significant improvement between two models. As there is not a structured way to choose the free parameters of SVR, a function in the R software was used in order to make a grid search over a supplied parameter ranges. The analysis of the results demonstrate that SVR is a promising technique to prediction of interest rates, suggestions are also made in order to get better the choices of the free SVR parameters once they are powerful means of regularization and adaptation to the noise in the data.

Keywords

Time series. Prediction models. Interest rates. Support Vector Regression. Statistical Learning.

Sumário

1	Introdução	11
1.1	Motivação	13
1.2	Objetivos	14
1.3	Estrutura da dissertação	15
2	Um pouco da teoria de otimização	16
2.1	Teoremas da teoria de otimização	17
3	Máquinas de Suporte Vetorial para regressão	20
3.1	Introdução à teoria de aprendizagem supervisionada	20
3.2	O Desenvolvimento da teoria SVM	22
3.3	Teoria de SVMs na aproximação por regressão	24
3.4	Como obter b	33
3.5	Os Núcleos	33
3.6	Arquitetura de um SVR	38
3.7	O SVR no <i>software</i> R	39
4	Descrição dos dados	40
5	Modelos de previsão	42
5.1	Previsão de séries temporais	42
5.2	Modelo de previsão usando SVR	42
5.3	Modelo de previsão usando Vetor Auto-regressivo (VAR) e Modelo de Correção de Erros (ECM)	45
5.4	Critérios de desempenho	45
6	Resultados experimentais	48
7	Conclusão e trabalhos futuros	62
	Referências Bibliográficas	75

Lista de figuras

1.1	Modelo de aprendizagem supervisionada a partir de exemplos (Vapnik, 2000(30)).	12
3.1	A máquina de vetores de suporte (Vapnik, 1998 (29)).	23
3.2	Hiperplano linear de separação ótima para dados linearmente separáveis (esquerda) e dados não separáveis (direita). Os vetores suporte estão circulados (Burgess (1998) (3)).	23
3.3	Função de custo ε -insensível linear (Schölkopf, Smola, 2001(18)).	27
3.4	Hiperplano separador linear com a função de custo ε -insensível (linear)(Schölkopf, Smola, 2001(18)).	28
3.5	A função de custo ε -insensível quadrática (Cristianini, Shawe-Taylor, 2000 (7)).	29
3.6	O mapeamento dos dados de entrada do espaço de dimensão R^n para um espaço característico com alta dimensionalidade R^N , onde uma regressão linear é efetuada (Smola, 1996 (22)).	34
3.7	Arquitetura de máquina de vetor suporte (Schölkopf, Smola, 2001 (18)).	38
7.1	Valores previstos nos dados de teste de $\log_s 30$ (1 mês).	63
7.2	Valores previstos nos dados de teste de $\log_s 60$ (1 mês).	63
7.3	Valores previstos nos dados de teste de $\log_s 90$ (1 mês).	63
7.4	Valores previstos nos dados de teste de $\log_s 120$ (1 mês).	64
7.5	Valores previstos nos dados de teste de $\log_s 180$ (1 mês).	64
7.6	Valores previstos nos dados de teste de $\log_s 360$ (1 mês).	64
7.7	Valores previstos nos dados de teste de $\log_s 720$ (1 mês).	65
7.8	Valores previstos nos dados de teste de $\log_s 30$ (3 meses).	65
7.9	Valores previstos nos dados de teste de $\log_s 60$ (3 meses).	65
7.10	Valores previstos nos dados de teste de $\log_s 90$ (3 meses).	66
7.11	Valores previstos nos dados de teste de $\log_s 120$ (3 meses).	66
7.12	Valores previstos nos dados de teste de $\log_s 180$ (3 meses).	66
7.13	Valores previstos nos dados de teste de $\log_s 360$ (3 meses).	67
7.14	Valores previstos nos dados de teste de $\log_s 720$ (3 meses).	67
7.15	Valores previstos nos dados de teste de $\log_s 30$ (6 meses).	67
7.16	Valores previstos nos dados de teste de $\log_s 60$ (6 meses).	68
7.17	Valores previstos nos dados de teste de $\log_s 90$ (6 meses).	68
7.18	Valores previstos nos dados de teste de $\log_s 120$ (6 meses).	69
7.19	Valores previstos nos dados de teste de $\log_s 180$ (6 meses).	69
7.20	Valores previstos nos dados de teste de $\log_s 360$ (6 meses).	69
7.21	Valores previstos nos dados de teste de $\log_s 720$ (6 meses).	70
7.22	Valores previstos nos dados de teste de $\log_s 30$ (9 meses).	70
7.23	Valores previstos nos dados de teste de $\log_s 60$ (9 meses).	70
7.24	Valores previstos nos dados de teste de $\log_s 90$ (9 meses).	71
7.25	Valores previstos nos dados de teste de $\log_s 120$ (9 meses).	71
7.26	Valores previstos nos dados de teste de $\log_s 180$ (9 meses).	71
7.27	Valores previstos nos dados de teste de $\log_s 360$ (9 meses).	72

7.28	Valores previstos nos dados de teste de $\log_s 720$ (9 meses).	72
7.29	Valores previstos nos dados de teste de $\log_s 30$ (12 meses).	72
7.30	Valores previstos nos dados de teste de $\log_s 60$ (12 meses).	73
7.31	Valores previstos nos dados de teste de $\log_s 90$ (12 meses).	73
7.32	Valores previstos nos dados de teste de $\log_s 120$ (12 meses).	73
7.33	Valores previstos nos dados de teste de $\log_s 180$ (12 meses).	74
7.34	Valores previstos nos dados de teste de $\log_s 360$ (12 meses).	74
7.35	Valores previstos nos dados de teste de $\log_s 720$ (12 meses).	74

Lista de tabelas

3.1	Funções de custo comumente usadas.	29
5.1	Valores escolhidos para cada parâmetro em cada um dos modelos.	44
5.2	Métricas utilizadas (a_i e p_i representam os valores atuais e previstos, respectivamente).	46
6.1	Resultados para previsões 1 passo à frente.	53
6.2	Teste de Theil para previsões 1 passo à frente.	53
6.3	Resultados para previsões 3 passos à frente.	55
6.4	Teste de Theil para previsões 3 passos à frente.	55
6.5	Resultados para previsões 6 passos à frente.	57
6.6	Teste de Theil para previsões 6 passos à frente.	57
6.7	Resultados para previsões 9 passos à frente.	59
6.8	Teste de Theil para previsões 9 passos à frente.	59
6.9	Resultados para previsões 12 passos à frente.	61
6.10	Teste de Theil para previsões 12 passos à frente.	61