5
Modelos de previsão

5.1
Previsão de séries temporais

Um método comum para escrever o problema de previsão de séries temporais em uma estrutura de estimação por regressão é considerar as séries temporais como sistemas dinâmicos. Seja uma série temporal $x(t)$, se $N \in \mathbb{N}$ e $\tau > 0$ são escolhidos adequadamente, então podemos prever $x(t)$ a partir de $(x(t-\tau), ..., x(t-N_\tau)) \in \mathbb{R}^N$. Portanto, pode-se considerar um problema de regressão onde o conjunto de treinamento consiste de entradas $(x(t-\tau), ..., x(t-N_\tau))$ e saídas $x(t)$ para diferentes valores de t. Porém, conforme destacado por Schölkopf e Smola (2001), várias características de previsão de séries temporais tornam o problema difícil para esta aproximação ingênua para regressão. Dentre elas:

– Séries temporais são freqüentemente não estacionárias, ou seja, a distribuição de séries temporais muda com o tempo e como conseqüência, exemplos de treinamento que são gerados como descrito acima se tornam menos úteis se eles são tomados de um passado distante.

– Os diferentes exemplos de treinamento não são iid, que é uma das hipóteses assumidas por modelos de aprendizagem estatística, como o SVR.

Em contrapartida, excelentes resultados têm sido obtidos usando SVR em problemas de previsão de séries temporais (25, 11) e tais resultados não podem ser desconsiderados.

5.2
Modelo de previsão usando SVR

Foram utilzadas uma e três defasagens para cada um dos modelos propostos e cada modelo foi identificado com o número de defasagens colocado após o nome dos mesmos. Assim, SVR1 corresponde a Máquina de Suporte Vetorial para Regressão com uma defasagem enquanto que SVR3 corresponde a três defasagens.

O primeiro passo para utilizar o SVR é definir qual núcleo será utilizado. Nesse estudo, opta-se por usar uma função núcleo Gaussiano, uma vez que funções núcleo Gaussianos, ou seja, $K(x, y) = \exp(-\gamma \parallel x - y \parallel^2)$, são úteis quando não se tem nenhum conhecimento dos dados disponíveis [11]. Também estima-se SVRL1 e SVRL3 que correspondem a uma regressão usando o núcleo linear, ou seja, $K(x, y) = \langle x, y \rangle$. O objetivo dessa estimação é verificar a contribuição da não-linearidade para o modelo. Os números 1 e 3 depois do nome do método têm o mesmo significado anterior. Uma das vantagens do SVR linear é que não existe parâmetro para ser ajustado além da constante C e do parâmetro ε. Para o SVR com núcleo Gaussiano existe um parâmetro adicional a ser ajustado, o parâmetro do núcleo γ.

No modelo SVR, cada entrada corresponde a um vetor com os valores de cada uma das sete taxas *swap* e cada uma das três variáveis macro-econômicas no instante (t-1), para o caso de uma defasagem. No caso de três defasagens, além dos valores em (t-1), observa-se ainda os valores de cada uma das variáveis em (t-2) e (t-3), ou seja, existem 10 e 30 valores de entrada para uma e três defasagens, respectivamente. Já os valores alvos desejados corresponderão a um único valor da taxa de *swap* a qual se quer determinar no instante t.

Veja alguns exemplos.

1. Caso o objetivo seja prever a taxa de *swap* 30 um passo à frente usando uma defasagem, os dados de entrada devem estar representados da seguinte maneira:

   $x_{t-1} = (S30_{(t-1)}, S60_{(t-1)}, S90_{(t-1)}, S120_{(t-1)}, S180_{(t-1)}, S360_{(t-1)}, S720_{(t-1)}, Selic_{(t-1)}, hiato_{(t-1)}, igpdi_{(t-1)})$. Com esses dados o valor estimado $f(x_{t-1}) = S30^*_t$ é obtido, o qual poderá ser comparado com o valor alvo desejado $y_t = S30_t$.

2. E se o objetivo for usar três defasagens?

   A única diferença estaria no vetor de entrada: $x_{t-1} = (S30_{(t-1)}, S30_{(t-2)}, S30_{(t-3)}, S60_{(t-1)}, S60_{(t-2)}, S60_{(t-3)}, S90_{(t-1)}, S90_{(t-2)}, S90_{(t-3)}, S120_{(t-1)}, S120_{(t-2)}, S120_{(t-3)}, S180_{(t-1)}, S180_{(t-2)}, S180_{(t-3)}, S360_{(t-1)}, S360_{(t-2)}, S360_{(t-3)}, S720_{(t-1)}, S720_{(t-2)}, S720_{(t-3)}, Selic_{(t-1)}, Selic_{(t-2)}, Selic_{(t-3)}, hiato_{(t-1)}, hiato_{(t-2)}, hiato_{(t-3)}, igpdi_{(t-1)}, igpdi_{(t-2)}, igpdi_{(t-3)})$.

3. E se o interesse está em fazer previsões três passos à frente usando uma defasagem?

   Seria preciso calcular as estimativas de cada uma das dez variáveis para o instante t, como no exemplo 1. Em seguida, essas estimativas seriam usadas como elementos do vetor de entrada para o cálculo da previsão de cada uma das variáveis no instante seguinte (t+1), logo uma previsão dois passos à frente seria obtida. Por último, as estimativas das variáveis obtidas no último passo seriam utilizadas como elementos do vetor de entrada para se obter a estimativa procurada, ou seja, a previsão três passos à frente da taxa *swap* desejada. É importante notar que embora o vetor de entrada seja o mesmo para cada etapa de previsão, o modelo SVR mudará, posto que cada previsão de uma taxa *swap* diferente terá valores diferentes para os parâmetros ε, γ e C.

Para se fazer previsões com seis, nove e doze passos à frente, o procedimento é análogo ao anterior, e para usar três defasagens, basta modificar o vetor de entrada.

A escolha dos parâmetros C, ε e γ (para a função núcleo Gaussiano) é feita baseada num conjunto de validação através de uma pesquisa em um domínio retangular utilizando um conjunto de valores fornecidos pelo usuário. Tal pesquisa é feita utilizando a função *tune* do pacote e1071 do *software* R. Para cada uma das dez variáveis do modelo, para cada um dos modelos com uma e três defasagens e para ambas funções núcleo é feita essa busca dos parâmetros.

Os valores usados para busca dos parâmetros são exibidos a seguir.

γ(função núcleo gaussiano): (0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001, 0.003, 0.005, 0.007, 0.009, 0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.3, 0.5).

ε: (0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001, 0.003, 0.005, 0.007, 0.009, 0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.3, 0.5).

C: 1-200 (de 5 em 5).

A tabela 5.1 exibe os parâmetros escolhidos.

PARÂMETROS SVR1

	log_s30	*log_s60*	*log_s90*	*log_s120*	*log_s180*	*log_s360*	*log_s720*	*log_selic*	*log_hiato*	*log_igpdi*
γ	0.01	0.007	0.0009	0.003	0.005	0.003	0.005	0.005	0.01	0.0001
C	26	191	161	146	21	41	21	101	196	191
ϵ	0.1	0.0001	0.09	0.07	0.1	0.3	0.3	0.03	0.3	0.3

PARÂMETROS SVR3

	log_s30	*log_s60*	*log_s90*	*log_s120*	*log_s180*	*log_s360*	*log_s720*	*log_selic*	*log_hiato*	*log_igpdi*
γ	0.0005	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.005	0.1	0.1
C	136	196	196	176	196	96	156	16	6	11
ϵ	0.0001	0.3	0.0001	0.0001	0.01	0.007	0.009	0.07	0.3	0.07

PARÂMETROS SVRL1

	log_s30	*log_s60*	*log_s90*	*log_s120*	*log_s180*	*log_s360*	*log_s720*	*log_selic*	*log_hiato*	*log_igpdi*
C	1	6	9	11	76	41	16	146	1	1
ϵ	0.1	0.1	0.0007	0.01	0.03	0.1	0.1	0.0003	0.05	0.1

PARÂMETROS SVRL3

	log_s30	*log_s60*	*log_s90*	*log_s120*	*log_s180*	*log_s360*	*log_s720*	*log_selic*	*log_hiato*	*log_igpdi*
C	1	11	1	56	91	11	1	1	1	36
ϵ	0.05	0.03	0.0001	0.05	0.0007	0.007	0.03	0.0003	0.1	0.005

Tabela 5.1: Valores escolhidos para cada parâmetro em cada um dos modelos.

5.2.1 Metodologia de aprendizagem

Todos os 7 retornos de contratos *swap* DI-pré e todas as variáveis macro-econômicas são divididas em três partes de acordo com a seqüência no tempo. A primeira parte (dados de junho de 1999 a janeiro de 2004) é usada para treinamento, a segunda parte para validação é usada para selecionar os parâmetros ótimos para SVRs (dados de fevereiro de 2004 a janeiro de 2005) e a última parte é usada para testes (dados de fevereiro de 2005 a junho de 2006).Cada uma das partes possui respectivamente 65%,15% e 20% dos dados.

5.3 Modelo de previsão usando Vetor Auto-regressivo (VAR) e Modelo de Correção de Erros (ECM)

Emprega-se o *software* EVIEWS 4, um programa de análise estatística e econométrica, que possui ferramentas para realizar previsões, a fim de modelarmos dois modelos de vetores auto-regressivos com uma e três defasagens.

A partir da observação dos gráficos de cada série e da realização de testes da raiz unitária de *Dickey-Fuller Augmented* foi verificado que as séries possuíam raiz unitária. Logo, também serão empregados dois modelos de correção de erros com uma e três defasagens. Testes de cointegração de *Johansen* indicaram a existência de 6 relações cointegrantes para uma defasagem e quatro relações cointegrantes para três defasagens.

Ambos os modelos, VAR e ECM, foram estimados e previsões foram realizadas a fim de serem comparados com os modelos de aprendizagem SVR.

Mais informações do *software* podem ser encontradas em http://www.eviews.com.

5.4 Critérios de desempenho

A performance de predição baseia-se nas métricas estatísticas seguintes:

- RMSE (Raiz do erro quadrado médio), MAE (erro absoluto médio): medem os desvios entre os valores previstos e atuais. Quanto menor esses valores, mais próximos estão os valores das séries temporais previstas dos valores atuais.
- DS (simetria direcional): fornece uma indicação de que a direção prevista está correta, dado na forma percentual e um valor grande sugere um preditor melhor.

$RMSE = \sqrt{\sum_{i=1}^{n} \frac{1}{n}(a_i - p_i)^2}$
$MAE = \frac{1}{n} * \sum_{i=1}^{n} \mid a_i - p_i \mid$
$DS = \frac{100}{n} * \sum_{i=1}^{n} d_i$ $d_i = \begin{cases} 1, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$\text{WDS}=\frac{\sum_{i=1}^{n} d'_i \lvert a_i - p_i \rvert}{\sum_{i=1}^{n} d_i \lvert a_i - p_i \rvert}$ $d'_i = \begin{cases} 0, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 1, & \text{caso contrário} \end{cases}$ $d_i = \begin{cases} 1, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$\text{CP}=\frac{100}{n} \sum_{i=1}^{n} d_i$ $d_i = \begin{cases} 1, & se(p_i - p_{i-1}) > 0, (a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$CD = \frac{100}{n} \sum_{i=1}^{n} d_i$ $d_i = \begin{cases} 1, & se(p_i - p_{i-1}) < 0, (a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$U = \frac{RMSE_i}{RMSE_j}$

Tabela 5.2: Métricas utilizadas (a_i e p_i representam os valores atuais e previstos, respectivamente).

- WDS (simetria direcional ponderada): mede a magnitude do erro e da direção de previsão, penalizando os erros relacionados à direção prevista incorretamente e recompensa aqueles associados com as direções previstas corretamente. Quanto menor o valor de WDS, melhor a performance de previsão tanto em magnitude quanto na direção.
- CP e CD (tendência para cima ou para baixo correta, respectivamente): medem a correta previsão de tendência para cima e para baixo, respectivamente, e em termos percentuais.
- Estatística de Theil-U : mede o desempenho de previsão relativa entre dois modelos, utilizando o RMSE como métrica.

As definições desses critérios estão ilustradas na tabela 5.2.

Em relação a estatística de Theil-U, $RMSE_{i(j)}$ representa a raiz do erro quadrado médio do método $i(j)$; i =SVR1 ou SVR3 e j =SVR1, SVR3, VAR1, VAR3, ECM1, ECM3, SVRL1, SVRL3 ($i \neq j$). Se $U < 1$ indica que a primeira previsão tem um RMSE menor do que o da segunda.

Os critérios DS, WDS, CP e CD fornecem uma boa medida da consistência na previsão da direção das taxas.