# Interest Rate Curve Estimation: a Financial Application for Support Vector Regression

**André d'Almeida Monteiro**[1]

Gavea Investments
Rio de Janeiro, Brazil
amonteiro@gaveainvest.com.br

First draft: 07-nov-2003

## Abstract

The Support Vector Regression (SVR) is applied on a key problem in financial economics: interest rate curve estimation. The SVR is able to introduce two important types of *a priori* information into the estimation. First, the kernel and the parameter $C$ can jointly assure an implicit quality requirement on the estimated curve. It is presented a proposition that reduces the search space for the optimal value of this parameter. Second, Vapnik's $\varepsilon$-insensitive cost function provides a natural channel to bring into the estimation information about liquidity and the price formation process of the assets from which the curve is extracted. It is captured by a variable observed during assets' transactions: bid-ask spreads (*BAS*). This cost function also allows for financial analysis of the assets selected as support vectors. The $\varepsilon$-insensitive function was modified in order to deal with different values for the parameter $\varepsilon$ in the same estimation, as imposed by the curve estimation problem. It is proved that this change keeps all the SVR properties. This paper models the dollar-Libor interest rate swap curves. It is a small-sample estimation: twelve contracts are available daily, from 1997 to 2001. The proper $C$-value ranges for some types of kernels were identified. The spline-with-an-infinite-number-of-nodes kernel provided the best out-of-sample specification for the SVR. On average, it required approximately 5 contracts to describe the curves under the desired accuracy fixed by the *BAS*. However, the support vectors cannot be used as sufficient statistics to analyze curve movements through time. Comparing this SVR with other three different approximation methods, it achieved the best control of the trade-off bias-variance for the interpolation problem. Its extrapolation performance, however, suggested that a new admissible kernel, incorporating long-term interest rate asymptotical behavior, would make the SVR more competitive.

**Keywords**: Support Vector Regression, Model Selection, Interest rate swaps, Interest rate curves, Bid-ask spread.

[1] This paper first draft was written when the author was visitinig the Department of Economics and The Bendheim Center for Finance at Princeton University.

## 1 - Introduction

The Support Vector Machine is a regression and pattern recognition tool derived from the Statistical Learning Theory, developed primarily by Vapnik.[2] Deep research has been done on the theoretical aspects of SV regression (SVR). However, this model has not been applied to a large number of real problems. This paper has two objectives: firstly, estimate interest rate curves by SVR, a cross-sectional financial real regression problem, and secondly, propose some new developments on its model selection.

An interest rate curve might be briefly described as a continuous function that establishes a dependence relation between an interest rate and its maturity. This function might be estimated from the observed data of interest rate and maturity of the negotiated contracts. In a normal economic and financial environment, one expects interest rate curve to be positively sloped and strongly convex, and the rate tends to an asymptotical value as the maturity approaches infinity. These are some stylized facts for interest rate curves, as shown by some authors, such as Campbell (2000). Figure 1 exhibits a typical interest rate curve. The rates are shown in annual basis (annualized). For instance, a borrower will pay 4.2% of interest per year for a one-year lending contract, whereas he or she will pay 6.0% per each one of ten years for a ten-year lending contract.

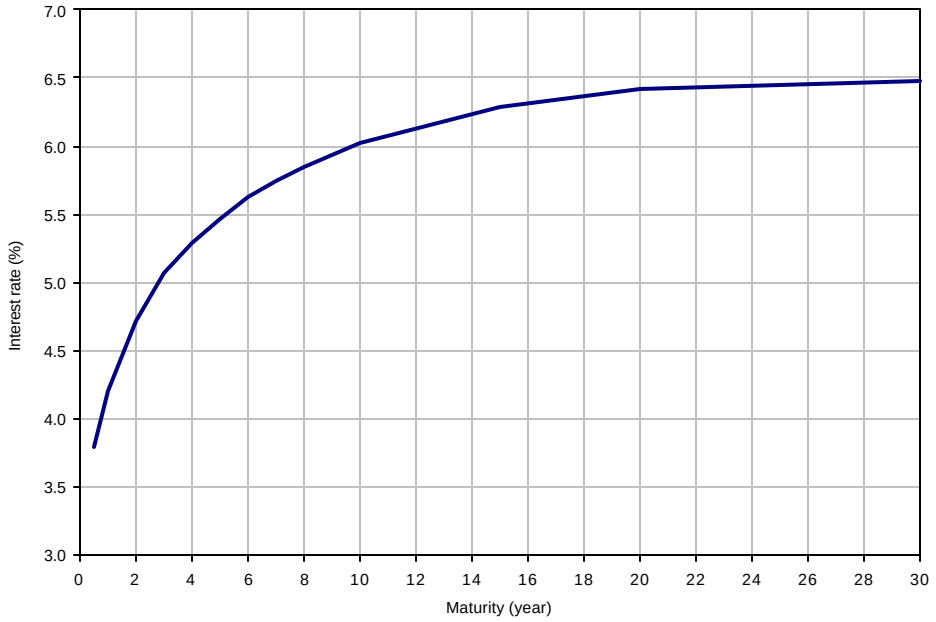



Figure 1: An illustrative typical interest rate curve

An interest rate curve is a fundamental tool in a market economy because it deeply impacts investment and consumption decisions of firms, householders and governments. A country might have more than one interest rate curve. However, economic agents usually choose one of them as the reference curve. Typically, it is built from the Federal Government debt. In the

[2] Vapnik (1979, 1982, 1995), Vapnik and Chervonenkis (1968, 1971, 1974).

United Stated (US), for example, the reference curve is derived from the securities issued by the United States Department of the Treasury. The reference curve plays a crucial rule for: pricing of securities issued by firms, states and municipality; commercial landings and mortgages; firm hedge strategy; and economic monitoring by the country's central bank. At the end of the 1990's two other curves got attention from financial researchers: agency mortgages and dollar-Libor interest rate swaps.

The SVR presents some attractive features in order to model interest rate curves. Firstly, the SVR is able to assure the estimated curve fulfill an implicit quality requirement. Secondly, it introduces into estimation important microstructure information of the assets from which the curve is extracted: their liquidity and price formation processes.

The two points above make this problem interesting from the SVR modeling point of view, since there is not a conclusive theory for SVR model selection. The biggest difficult in setting its generalization capacity resides in the interdependence relationship among its parameters. The quality requirement on the curve and the microstructure information about the assets will guide the decision of the three choices that specify a SVR: loss function, kernel function and parameter $C$.

The Vapnik $\varepsilon$-insensitive is chosen as the loss function. This decision was motivated by two reasons. Its parameters bring into the estimation process a variable observed during asset's negotiation: *bid-ask spread* (*BAS*). In short, the *BAS* is the difference between the *ask price*, the price at which the owner is willing to sell, and the *bid price*, the price at which the buyer is willing to buy. A modification on the $\varepsilon$-insensitive function is required in order to allow for different values of *BAS* in the same estimation. It will be proved by the Preposition 1 that this modification does not change any one of the $\varepsilon$-insensitive-SVR properties.

Additionally, this loss function leads to a sparse solution. It means that only a portion of observed contracts really matter for the estimation; they take place into the linear combination that expands the approached function with non-zero coefficients. In this sense, the support vectors set might be seen as a sufficient statistic for the curve. The sparsity allows for analysis of financial rule played by support vectors in the cross-sectional regressions. The support vectors capability to describe curve movements through time will be tested.

The kernel function and the parameter $C$ will jointly be established by estimation quality requirements used in financial literature. The effort applied on the selection of the parameter C will be lowed by the Proposition 2. It reduces the search space for the optimal value of this parameter by using its rule in the SVR coefficients estimation – an optimization process.

Under certain conditions, there are equivalence between SVR and regularization networks (RN) by Girosi, Jones and Poggio (1993). It is important because it builds a bridge between SVR and smoothing splines and radial basis functions, methods used in financial field. In addition, this connection allows RN to take advantage of the Preposition 2.

SVR will model the US interest rate curve built from dollar-Libor interest rate swaps. Two basic reasons motivate this choice. First, the importance of this financial asset has significatively increased since the 1998 hedge fund crisis. Second, swap bid-ask spreads

contain rich information set. The approach that will be presented can be modified to estimate other curves, for example, US treasuries. Estimation of interest rate curves is not a large dataset problem: it is a plain-vanilla quadratic programming.

In addition to this introduction, the paper has other eight sections. The second one briefly presents SVR and discusses the related literature, in particular its applications. In the third, swap interest rate curve estimation problem is presented and discussed step-by-step: basic concepts of fixed income assets, interest rate swap and it bid-ask spread, and interest rate curve estimation. SVR model selection questions are discussed in the fourth section, where the two propositions are presented and developed. The data is commented in the fifth section. The section six investigates the estimation (in-sample) performances of some different specifications of SVR, the financial meaning of support vectors and their capability as sufficient statistics for curve in cross-sectional and time series analysis. The out-of-the-sample performances of these SVR specifications are analyzed in the seventh section. The best SVR specification is then confronted against other three approximation methods. The ninth section concludes the paper.

## 2 – Support Vector Regression

Consider the problem of approximating the observed dataset, $\mathrm{A} \equiv \{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathrm{X} \otimes \Re$, where $x \in \Re^{\mathrm{d}}$, $X \equiv \{x_i\}$ is the set of training vectors (or explanatory or independent variables), and $y \in \Re$ the dependent variable. The SVR objective is to model a dependence relationship between $x$ and $y$ through (1):

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*).k(x_i, x) + b \tag{1}$$

where $k$ is a kernel that satisfies Mercer's condition (1909) and $b$ is a constant. Starting from kernels that satisfy Mercer's condition, Smola and Schölkopf (1998a) derive some rules for composition of kernels that also satisfy this condition. They called them admissible SV kernels.

The selection of the loss function is fundamental because the whole SVR construction depends on it. Muller, Smola, Ratsch, Scholkopf, Kohlmorgen and Vapnik (1997) and Smola, Scholkopf and Muller (1998b) show experimentally how loss-function-dependent is the SVR performance.

Let $\boldsymbol{x}_i$ be the deviation of estimated function from observed value: $\boldsymbol{x}_i = y_{x_i} - f(x_i)$, $i$= 1, …, $n$. I will use the ε-insensitive loss function, L, proposed by Vapnik (1995).

$$L = \sum_{i=1}^{n} |\boldsymbol{x}_i|_{\boldsymbol{e}}$$

$$|\boldsymbol{x}_i|_{\boldsymbol{e}} = \begin{cases} 0, if\ |\boldsymbol{x}_i| \le \boldsymbol{e} \\ |\boldsymbol{x}_i| - \boldsymbol{e}, otherwise \end{cases} \tag{2}$$

where ε is a fixed parameter. According to Vapnik (1995), the ε-insensitive-SVR coefficients $(\alpha, \alpha^*)$ may be estimated by its dual quadratic optimization problem:

$$\begin{aligned} &Maximize \quad -\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*).k(x_i, x_j) - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\ &sa \begin{cases} \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \tag{3}$$

where the constant $C>0$ controls the trade-off between the function flatness and the tolerance for deviations. The lower $C$, the flatter is the function $f$. Since the cost function is convex and the restrictions are linear, the SVR estimation is a convex quadratic programming problem. And, therefore, there exists one global solution. Burges and Crisp (1999) discuss this topic in detail.

Smola and Schölkopf (1998b) as well as Evgeniou, Pontil and Poggio (2000) examine the connection between the regularization networks by Girosi, Jones and Poggio (1993) and the SVR. The parameter $C$ might be understood as the inverse of a regularization parameter. For simplicity, I will call it *regularization parameter* henceforth. The constant $b$ in (1) can be computed by some different ways, according to optimization process. Once I will use a plain-vanilla quadratic programming, $b$ can be computed from equation (1) when the contract is a support vector. Muller, Smola, Ratsch, Scholkopf, Kohlmorgen and Vapnik (1997) suggest taking the mean over the values of $b$ calculated from each support vector. In this paper, I use the median, instead of the mean. The motivation is the same: numerical security.

Unfortunately, SVR theory does not provide a definitive model selection strategy. The biggest difficult in setting SVR generalization capacity resides in the interdependence relationship among the kernel, and the parameters $C$ and $\varepsilon$. The curve estimation problem offers important *a priori* information that can be naturally used in the model selection task. This current literature on SVR model selection and the choices made in this paper will be discussed in the section 4.

The reporting literature about application of SVR over real problems is not as rich as that one which approaches the pattern recognition problems. Some of the existing literature is commented in the following paragraphs.

Muller, Smola, Ratsch, Scholkopf, Kohlmorgen and Vapnik (1997) investigate the SVR performance over two synthetic time series. The performance is measured by *cross-validation* and the SVR is faced against radial basis neural networks. A multidimensional chaotic system is generated by the lagged Mackey-Glass differential equation. This series is then corrupted by additive noises originated from two different distributions: uniform and Gaussian. Some SVR specifications are tested: the values of the parameters $\varepsilon$ and $C$ are varied and two loss functions are used: Huber (1981) and ε-insensitive. There is not a conclusive outstanding method. The second time series approached is the D-data of the Santa Fe Competition. In this case, the SVR outperformed the neural networks.

Drucker, Burges, Kaufman, Smola and Vapnik (1997) examine four linear problems, three synthetic and one real. SVR is always faced against *bagging* method by Breiman (1994). Three Friedman (1991) equations generate three different artificial series. The values of parameters $C$ and $\varepsilon$ are selected by *cross-validation*. The MVS solutions show high sensitive to these parameters. Some polynomial kernel functions are tried. The real problem is the estimation of 506 house prices in Boston by 20 explanatory variables. The SVR outperforms the *bagging* in the four cases.

Collobert and Bengio (2001) examine three problems in order to illustrate their new optimization procedure. The first works on a simulated forward dynamics of an 8-link all-revolute robot arm. The SVR job is to predict the distance of end-effectors from a target given some features. The last two cases deal with number of sunspots: a real series and an artificial one. In all experiments, the authors use Gaussian kernel and $C$=100. The kernel parameter and $\varepsilon$ are fixed *per* each problem.

Chapados et al. (2001) investigate car insurance premium from a large dataset. They apply SVR and other methods. They report that SVR does not achieve a good performance, along with those approximating methods that do not minimize a quadratic loss function.

Gestel et al. (2001) apply Bayesian evidence framework to Least Square-SVM by Suykens (2000) to infer non-linear models for predicting financial time series and related volatility. The model is tested on US Treasury Bill and on a Germany stock index (DAX30). The one-step-ahead performance measure shows that the model is able to capture a significant portion of out-of-sample volatility.

## 3 – The Problem: Estimation of Swap Interest Rate Curves

This section states and discusses the problem of estimating interest rate swap curves, taking into account the bid-ask spreads from the contracts used in the estimation and some quality requirements over the forward rate curves. It is composed by three subsections. The first one presents the basic concepts of fixed-income assets. The second one describes the interest rate swap contracts and its bid-ask spreads. The final subsection contains a brief survey on financial literature concerning interest rate curve estimation and the effects of bid-ask spreads on it.

### 3.1 – Fixed Income Asset Basics

A fixed income security is an asset that promises to pay a pre-established cash flow. The simplest fixed income asset is the one whose cash flow is completely known at issue date and makes a unique payment at the end of asset life. These assets are called zero-coupon bonds.

The payment day is called maturity day. The maturity is measure in days, months or years. Consider a bond that promises to return $P$ dollars $x$ periods ahead. $P$ is called principal. If an investor pays $p_x$ dollars for it at issue, he or she will lend money for the issuer requiring interest of $y_x$ per unity of time. This rate is called spot rate (or zero-coupon rate).

$$p_x = \frac{P}{(1+y_x)^x} \tag{4}$$

If a bond promises to return \$100 two years ahead and an investor buys it for \$94.3, the equation (4) calculates the interest rate required by this investor as 3% per year. In other words, the investor (buyer) is lending money for the bond issuer (borrower) and requiring interest of 3% per year.

Another fundamental concept is the forward rate. A forward rate is a rate at which a lender or a borrower can contract a lending today starting $x$ periods of time ahead to be paid back at $x$+1. In fact, it is a future interest rate negotiated today. Let $y_{x-1}$ and $y_x$ be, respectively, (x-1)-zero-coupon bond and $x$-zero-coupon bond. The forward rate $Fx$ relative to period $x$ is:

$$F_x = \frac{(1+y_x)^x}{(1+y_{x-1})^{x-1}} - 1 \tag{5}$$

Indeed, forward rates and spot rates are equivalent ways of expressing the same information about fixed-income asset prices. It is possible to see each one of them embedded in the others. Equation (5) says how the forward rates are implicit into spot rates. Conversely, a spot rate can be seen as a product of forward rates up to its maturity.

$$(1+y_x)^x = (1+F_1)(1+F_2)\ldots(1+F_x) \tag{6}$$

For instance, consider spot interest rates of 4.6%, 5% and 5.3% per year related to 2, 3 and 4 year maturities, respectively. The 3-year forward rate is 5.8% per year, following equation (6). In the same way, the 4-year forward rate is 6.2%. Now, the 4-year spot rate con be computed by (6):

$$(1+0.045)^2(1+0.058)(1+0.062) = (1+0.053)^4$$

The formulas (5) and (6) are capitalization-regime-dependent. The diagram showed in Figure 3 illustrates the relation between forward and spot rates.

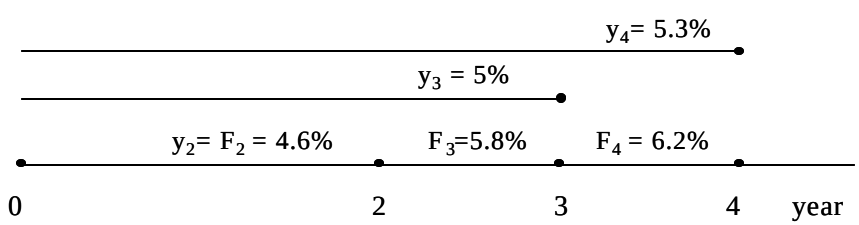



Figure 3: Forward and spot rates and their validity periods.

A yield curve is the plot of spot rates in function of their maturities. It is also possible to define a forward interest rate curve in the same way. The equivalence between spot and forward rates is extended to their curves.

Despite their convertibility, forward rates exhibit more singular economic meaning than spot rates, since the former composes the later. Therefore, the financial literature imposes some restrictions on the forward rate curve even when the spot rate curve has been estimated. This point will be discussed in the subsection 3.3.

The negotiation of an asset is not fully described by the price at which the deal is done. In fact, three other prices are important: bid price, ask price and mean price. The bid price is the price at which the buyer is willing to buy the asset $i$ at the beginning of negotiation. The ask price is the price at which the seller is willing to sell asset $i$ at the beginning of negotiation. The bid-ask spread could have been defined here. Nevertheless, its formal definition will be discussed in the next section to take advantage of the swap context.

The existence of bid-ask spread and its determining factors compose one of the topics covered by a subfield of financial economy called *market microstructure*. It investigates the dynamics of price formation and the liquidity conditions, fixed an information set. It also studies operational and institutional aspects that conduct daily market activities and that are fundamental factors to understand and model short-run financial phenomenon. For instance, the following two topics are under interest of this subfield: assets are not traded at evenly spaced intervals through the day; and their prices are usually denominated in fixed increments – they are not continuous. [3]

The *BAS* is one of the most important indicators of asset liquidity. Gouriéroux, Jasiek and Fol (1999) separate the usual liquidity measures into two groups: direct (traded volume and time required to trade a pre-determined volume of the asset) and indirect (bid-ask spread and volatilities). In a nutshell, one can say the more an asset is traded, the smaller tends to be the uncertainty involving its price and, consequently, its bid-ask spread. McCulloch (1987) resumes the rule played by BAS on asset price (page 189): "The problem is that the value of a security is ambiguous within its bid-ask spread."

### 3.2 – The Interest Rate Swaps

The simplest interest rate swap is a contract under which two parties (called counterparties) agree in exchanging cash flows of interest payments without exchanging the underlying principal, called notional principal (*NP*). The cash flows are calculated over the *NP* by applying two different interest rates: a floating ($T_{fl}$) and a fixed ($T_f$). Net payments are done periodically according to the net difference between interest rates times notional principal.

$$Net\ payment = NP\left(T_{fl} - T_f\right)$$

While the counterparties agree on the value of the fixed interest rate in beginning of the contract, the floating interest rate is only known close to each periodic payment. The floating rate might be an index, such as an US Treasury rate or the Libor. This paper focuses on swaps

[3] The following works cover the basic issues and models of this field: Campbell, Lo and Mackinlay (1997), Cohen, Mair, Schwartz and Whitcomb (1986), Davis and Holt (1993), Frankel, Galli and Giovannini (1996), Kagel and Roth (1995), Lo (1995) and O'Hara (1995).

that the floating interest rate is the dollar-Libor. The dollar-Libor (London Interbank Offered Rate) is the rate of interest at which banks borrow dollar-denominated funds from other banks in the London interbank market.

The dollar-Libor interest rate swap (only swap, henceforth) has rapidly increased since its beginning in 1981. The BIS estimates at the end of 1999 that swap notional principals summed $43.9 trillions, while US Treasury face value summed $5.7 trillions.

An important research issue related to swap contracts is the modeling and the analysis of the positive swap spreads over equivalent Treasury rates. Cooper and Mello (1991), Bollier and Sorensen (1994), Duffie and Huang (1996) and Duffie and Singleton (1997) are some examples. However, empirical evidences are divergent about the rule played by credit quality in swap spreads. McNulty (1990) claims there is small evidence of credit quality premium in the market rates. Brooks and Malhotra (1994), on the other hand, find out evidence of this premium in bid and ask swap rates.

The most important players at swap market are companies, commercial and investment banks. The trades are oriented to middle and long-term funding and hedging. These swaps are over-the-counter contracts, that is, they are not traded on organized exchanges. These banks are always ready to quote two fixed interest rates. The bid interest rate ($T_{f,a}$) is the fixed rate at which the bank is willing to pay a cash flow tied to it and to receive cash flow tied to Libor. The ask interest rate ($T_{f,b}$) is the fixed rate at which the bank is willing to receive a cash flow tied to it and to pay cash flow tied to Libor. The bid-ask spread of a contract $i$ is then the difference between the two rates.

$$BAS_i = T_{f,b,i} - T_{f,a,i} \tag{7}$$

The mean rate is straightforward defined by (13):

$$T_{M,i} = \frac{T_{f,b,i} + T_{f,a,i}}{2} \tag{8}$$

There are no theoretical models to explain the *BAS* dynamic and its determinants for swap contracts.[4] By using regression, some researches find empirical evidence that some variables have significant explanatory power on *BAS*. All of them conclude that *BAS* are statistically different from zero. Sun, Sundersan and Wang (1993) point that *BAS* are sensitive to the bank credit. Brooks and Malhotra (1994) show that the transaction costs, the maturity of the contract, the level of Treasury interest rates and the payment frequency are relevant variables. However, their explanatory power on BAS is not clear because of multicolinearity. Malhotra (1998) also observes a positive relationship between BAS and contract maturity. He attributes this behavior to the greatest liquidity in low maturities.

[4] Indeed, such models exist for stock bid-ask spreads.

### 3.3 – Interest Rate Curve Models: BAS and Forward Rates

A large amount of research has been done on interest rate curve estimation from observed rates since the first attempt made by Durant (1942). Since then, different models have been applied on this job.

Some authors attempt to use parametric models, such as Fisher (1966), Cohen, Kramer and Waugh (1966), Echols and Elliott (1976), Dobson (1978) and Chambers, Carleton and Waldman (1984). Some of them apply polynomial regression and all of them generate unbounded interest rates for long-term maturity. Nelson and Siegel (1987) propose a parsimonious model and force asymptotical long-term interest rates. However, as pointed out by the authors, the model still presents difficulty in estimating long-term rates.

Because parametric models were not entirely successful in estimating curves, non-parametric models and regularization theory were requested to this job. McCulloch (1971, 1975) introduces the cubic spline regression. Vasicek and Fong (1982) suggest an exponential transformation on data before the spline treatment. Nevertheless, Shea (1985) shows this last spline does not outperform traditional splines and fails in producing more stabilized forward curves.

Fisher, Nychka and Zervos (1995) add a smoothness parameter (or regularization parameter) into cubic spline model. It results in the smoothing splines (or splines with regularization). This spline seeks to minimize the trade-off between squared mean error (data fitting) and smoothness of estimated function. Waggoner (1997) increases the last model flexibility by allowing for three different values for the regularization parameter over the maturity domain. He reports that his model was able to outperform the original in terms of data fitting.

The majority of financial economic research has been focused on the US interest rate extracted from the securities issued by the US Department of the Treasury. Only in the end of the 1990's, the curves based on the agency mortgages and swap contracts gained some attention. For example, Nielsen and Ronn (1996), Collin-Dufresne and Solnik (1999), Grinblatt (1999) and Duffie and Singleton (1997).

The existence of bid-ask spread (*BAS*) causes some problems for curve estimation. McCulloch (1971) argues that (page 27) "(…) the term structure cannot be measure exactly" because of the existence of *BAS*. Dermody and Prisman (1988) show that, even under no-arbitrage assumption, the transactions costs (BAS and fees) are the responsible for the existence of an infinitely countable, convex and compact set of term structures for each class of investors. Therefore, the *BAS* has to be considered during the estimation of a model.

The simplest way to deal with *BAS* in curve estimation is to collapse it into the mean rate. Figure 4 illustrates the impacts that the *BAS* may cause on the curve estimation. It shows the bid, ask and mean rates of one day in the sample. In Figure 4.a, one can note that the 7-year ask rate is out of the trace jointly produced by the other contracts. Looking at bid rates (Figure 4.b), the 7-year and the 30-year contracts seem to be outliers. Figure 4.c shows that mean rates smooth the differences among these two contracts and the others. But the 7-year-contract mean rate seems to be out of the curve done by the other contract's rates. Finally, Figure 4.d exhibits

the reason why these contracts differ from the others: the uncertainties of their rates, measured by their bid-ask spreads, are greater than the uncertainty present in the other contracts.

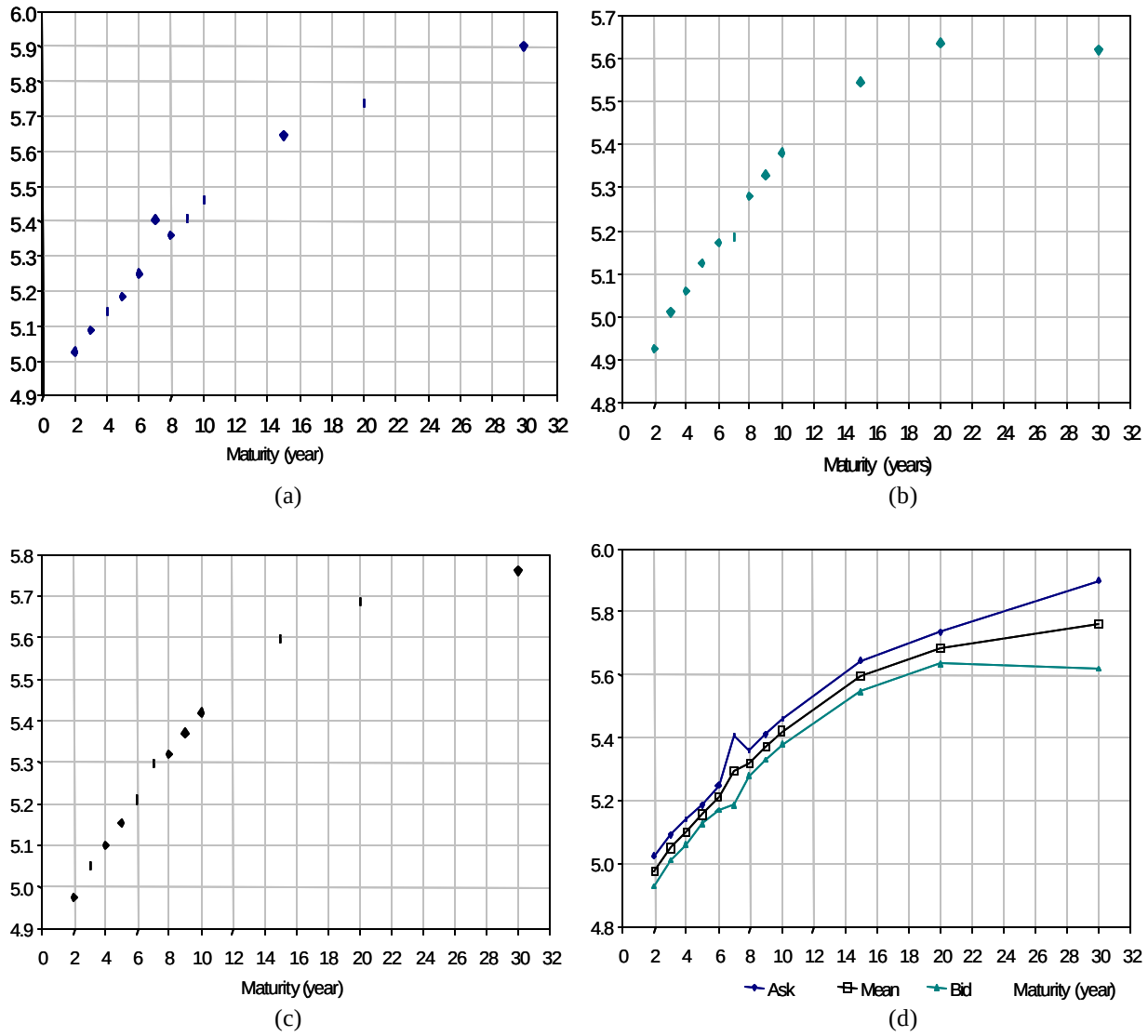



Figure 4: The rates of swap contracts of a sample day: (a) ask rates, (b) bid rates, (c) mean rates, (d) ask, bid and mean rates.

The mean-rate solution may be useful when both bid and ask rates differ from the other rates symmetrically, as it was the case of the 30-year contract in Figure 4.d. However, the same figure shows it is not a good solution if the differences among the bid and ask rates are asymmetric, as it occurred with the 7-year contract.

In the first subsection, attention was required to the singularity of forward rates. Its stability is an extremely important criterion of the quality of any statistical model that wishes to estimate interest rate curve. The forward rate curve has to be continuous and smooth through maturity. There are two reasons for this requirement. First, from economic point of view it is hard to

explain why the price of money significantly varies over maturities. And second, a non-smooth term structure may cause problems for interest rate derivative pricing.

Now it is possible to summarize the problem faced by SV regression: to estimate the swap spot rate curve, taking into account the swap contracts' bid-ask spreads, which implicit forward rate curve is continuos and smooth.

## 4 – SVR Model Selection

After the problem has been established in the last section, this one specifies the SVR parameters in accordance with the estimation objective. This section is divided into three subsections; each one of them is dedicated to each of the three SVR parameters. They first discuss suggestions of the literature for parameter selection and, in the sequence, comment how this paper effectively made the choices.

The selection of the parameter $\varepsilon$ will be discussed in the first subsection. In order to introduce this variable into the estimation, the formulation of the SVR model will be slightly modified. In particular, the Vapnik loss function will be changed. However, the Proposition 1 will assure that all the SVR properties are kept after this modification is done. The second one focuses on the regularization parameter $C$. A proposition (Proposition 2) that reduces the search space for its optimal value will be presented. The selection of the kernel will be restricted by some requirements on forward rates, in the last subsection.

### 4.1 – Parameter e

The parameter $\varepsilon$ is the focus of the greatest amount of research so far. It can be interpreted as additive noise corrupting the function's true value ($f(x)$) or as the desired accuracy in estimation. Smola, Murata, Scholkopf and Muller (1998) show that fixing $\varepsilon$=0 is not a good choice, because some experiments suggest the existence of a linear relationship between the level of additive noise and the parameter $\varepsilon$. Schölkopf, Bartlett, Smola, and Williamson (1998) propose the $\nu$-SVR. In this alternative approach, $\varepsilon$ is endogenously adjusted by the SVR in order to keep fixed a pre-established portion of observations as support vectors (SV's). Their objective is to directly control the number of SV's. They assume that the number of SV's falling exactly on the tube boundary is asymptotically insignificant. If this hypothesis fails, the $\nu$-SVR controls only the number of SV's located out of the tube. Smola and Scholkopf (1998a) propose a selection criterion for $\nu$ based on the noise distribution.

One purpose of this paper is to take in consideration these misbehaviors of the asset prices by introducing the bid-ask spreads into the estimation process as discussed in the end of the last section. Unfortunately, the rate at which the deal is done (if it really is) is not available, so I have to use the mean rate as $y$ variable.

The parameter $\varepsilon$ will be fixed as the half of the BAS of each one the available contracts. Once introducing the "ambiguity" of the rates into the estimation process, the model capacity to

represent the approached problem should increase. The parameter ε will be fixed as the half of the observed bid-ask spread.

$$\boldsymbol{e}_{i,t} = \frac{BAS_{i,t}}{2}, \; i=1, ..., 11 \text{ and } t=1,... T. \tag{9}$$

In order to apply SVR on the curve estimation and incorporate into estimation process the greates possible amount of information contained within the bid-ask spreads, a change in the traditional SVR specification is required. In fact, the parameter ε needs to present different values in the same estimation, as suggested by equation (9). Therefore, I indexed the parameter ε according to the observation. Equation (10) presents the $\varepsilon_i$-insensitive loss function.

$$L = \sum_{i=1}^{n} \left| x_i \right|_{e_i}$$

$$\left|\xi_i\right|_{\varepsilon_i} \equiv \begin{cases} 0, \; if \; \left|\xi_i\right| \le \varepsilon_i \\ \left|\xi_i\right| - \varepsilon_i, \; if \; \left|\xi_i\right| > \varepsilon_i \end{cases}, \; i = 1, ..., n \tag{10}$$

This change in Vapnik's loss function does not damage any one of the SVR properties as proved by the next proposition.

<u>Proposition 1</u>

*A SVR equipped with the $\boldsymbol{e}_i$-insensitive loss function (10) keeps all the optimization properties of an equivalent SVR with the $\boldsymbol{e}$-insensitive loss function: a quadratic-programming problem with linear restriction that might generate a sparse solution.*

Proof: see the Appendix.

One can see the determents of *BAS*, commented in subsection 3.2, distorting the true value of the swap rate. In this sense, the *BAS* can be understood as an additive noise. However, the true value is unknown. On the other hand, as the bid-ask spread defines an interval containing "ambiguous" rates, the cost function should not be penalized by any estimation which value is inside this interval. In other words, the parameter ε reflects in some sense additive noise, and it might be understood as the desired accuracy as well.

One way to understand this estimation process is to understand the *BAS* as a measure of the trustful information presents in the swap rate. The smaller (greater) the *BAS*, the more (less) trustful is the information contained in the rate, and more (less) concerning with this rate the SVR should be by determining a smaller (greater) tolerance for deviation around the observed data.

Smola and Schölkopf's (1998b) result allows for a theoretical comparison between the SVR and the smoothing splines. These last methods require the number and the location of knots *a*

*priori*. By its turn, the SVR endogenously selects the number of the "knots" (support vectors) and their location conditional to the selected kernel, *C* and ε. What makes this selection endogenous is the ε-insensitive loss function (or the $\varepsilon_i$-insensitive loss-function). Theoretically, it is one advantage of SVR over splines.

### 4.2 – Parameter *C*

In general, the value for parameter *C* is chosen by re-sampling methods (*cross-validation* and *bootstrap*). The selection process of parameter *C* will be lead by the following proposition. It reduces the search space for the optimal value of *C*.

<u>Proposition 2</u>

*Fixed the values of* $\boldsymbol{e}_i$*, i=1,.., n and the kernel, there exists only one value of C, called Cs, such that any C ³* $C_S$ *generates the same SVR coefficients* $\boldsymbol{a}_i, \boldsymbol{a}_i^*, \mathrm{i} = 1, ..., \mathrm{n}$*, everything else kept constant.*

*Having estimated SVR with a high C value, called C', in the sense that no inequality restriction in (2) have achieved the upper limit, that is* $\forall \alpha_i, \alpha_i^* < C', \mathrm{i} = 1, ..., \mathrm{n}$*,* $C_S$ *is defined as the maximum value among the estimated dual coefficients* $\boldsymbol{a}_i, \boldsymbol{a}_i^*, \mathrm{i} = 1, ..., \mathrm{n}$.

$$\mathrm{Cs} \equiv \max_{\mathrm{i}} (\alpha_{\mathrm{i}}, \alpha_{\mathrm{i}}^*), \mathrm{i} = 1, ..., \mathrm{n} \tag{11}$$

The proof is trivial. It is valid for the linear and the non-linear case of SVR. The parameter *C* is the upper bound in the dual inequality restrictions (2) over the coefficients α and α*. Once these restrictions are not active, an increase in the value of *C* starting from *Cs* does not change the values of the SVR coefficients. Therefore the search space for its optimal value reduces to ]0, Cs]. Once *Cs* has been chosen as the value of *C*, the resulting model is the most flexible SVR that can be achieved, everything else kept equal. In other words, the minimal regularization model is the one with *Cs*.

Essentially, the Proposition 2 does not come from *a priori* information neither from a re-sampling strategy. Only one estimation is enough to determine the value of *Cs*. In this sense, it is not found out by a re-sampling method. However, this value depends on the data and on the remaining SVR parameters. Thus, it is not *a priori* information. In the next subsection, this proposition will be the instrumental in determining comprehensive intervals for the parameter C according to the selected kernel. The regularization networks (RN) can also take advantage of the Proposition 1 since the conditions under which RN are equivalent to SVR are achieved.

### 4.3 –Kernel Function

The kernel choice is typically based on some re-sampling method as well as the parameter C (cross-validation and bootstrap). Some authors investigate the introduction of a priori information in the kernel-selection process in the SV pattern recognition, such as Burges

(1998) and Scholkopf, Simard, Smola and Vapnik (1998). There is, nevertheless, a lack of this kind of research for regression problems.

The quality required over forward rate curves addresses to kernel selection, because the kernel determines its shape, jointly with the parameter C. This section investigates the characteristics of the forward curves generated by different kernels and their sensibility to the value of C, taking advantage of Proposition 2. The following kernels will be considered:

- Spline with an infinite number of nodes (*splinf*)
  $k(x_i, x_j) = 1 + x_i.x_j.\min(x_i, x_j) - \frac{x_i + x_j}{2}\left(\min(x_i, x_j)\right)^2 + \frac{(\min(x_i, x_j))^3}{3}$, where $x_i$, $x_j$ <1.

- Vovk's real infinite polynomial (*polvok*)
  $k(x_i, x_j) = \frac{1}{1 - (x_i.x_j)}$, where $-1 < (x_i.\ x_j) < 1$.

- Second degree polynomial (*polin*2)
  $k(x_i, x_j) = ((x_i.x_j) + 1)^2$

- Third degree polynomial (*polin*3)
  $k(x_i, x_j) = ((x_i.x_j) + 1)^3$

- Gaussian radial basis function (*rdfgua*)
  $k(x_i, x_j) = exp\left(\frac{(x_i - x_j)^2}{2\omega^2\sigma^2}\right)$, where $\sigma$ is the sample standard deviation of the training vectors *x* (contract maturities) and $\omega$ is a constant.

- Exponential radial basis function (*rdfabs*)
  $k(x_i, x_j) = exp\left(\frac{|x_i - x_j|}{2\omega^2\sigma^2}\right)$

I arbitrarily selected two values for the $w$ parameter: 1 and 0.5. There is no wish in setting optimal values for it. I am only interested in illustrating its impacts on the forward rate curve.

A remark is required at this point. As the next section will show, the spot rate curves estimated by SVR equipped with different kernels might be very close. In fact, visual inspection on estimated spot curves is not able to point strong differences among performances by varying the kernel function. The bottom line is that the forward rate curves implicit in similar spot rate curves might exhibit completely different shapes and *C*-sensitivity. One advantage of modeling spot rate curves with SVR is that kernel set up the shape of the forward curve. As out-of-sample results will show in the section 7, keeping consistence on forward curves avoids bad performances.

Following Proposition 2 and equation (11), I computed the values of $C_S$ for the SVR equipped with the above eight kernels for each sample day. Then, the percentiles are calculated for each kernel. Based on them, we decided for 5 values of $C$ for each kernel. They are exhibited in Table 1. The required regularization increases from *C1* to *C5* as the value of $C$ decreases (one should remember that in fact the parameter $C$ might be seen as the inverse of the regularization parameter). The ranges might be used as *a priori* information for feasible values of $C$ per kernel.

Table 1: Different values for *C* per kernel

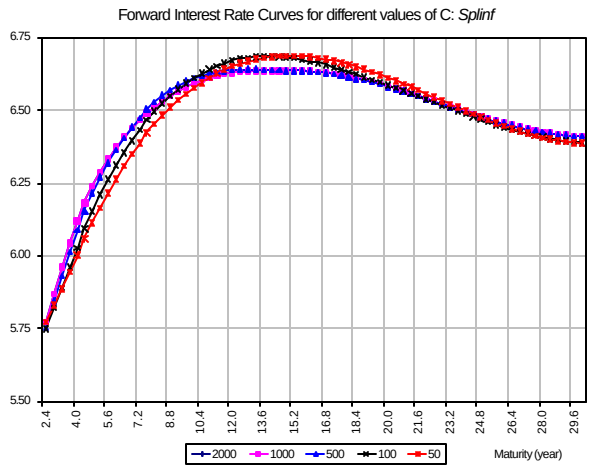
| Kernel | C1 | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| rdfabs(0.5) | 0.6 | 0.4 | 0.3 | 0.2 | 0.1 |
| rdfabs(1) | 0.7 | 0.5 | 0.4 | 0.3 | 0.2 |
| rdfgua(0.5) | 10.0 | 5.0 | 2.0 | 1.0 | 0.2 |
| rdfgua(1) | 100.0 | 50.0 | 10.0 | 2.0 | 0.5 |
| polin2 | 5.0E+10 | 3.0E+10 | 1.0E+10 | 1.0E+08 | 1000.0 |
| polin3 | 1.0E+10 | 1.0E+09 | 500.0 | 100.0 | 50.0 |
| polvok | 100,000.0 | 50,000.0 | 5,000.0 | 100.0 | 50.0 |
| splinf | 2,000.0 | 1,000.0 | 500.0 | 100.0 | 50.0 |

The investigation of implicit forward curve is based on one typical curve (and indeed a representative curve of the sample) picked up from the sample (Jan 1st 2001). The values of $C$ will be varied from *C1* to *C5*. Therefore, 40 SVR specifications will be tested.

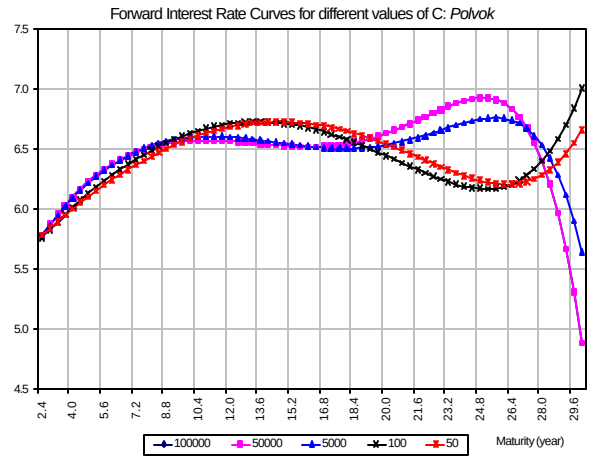
The SVR coefficients are estimated for the spot curve. As the swap capitalization regime is linear assuming 360 days in a year, the annualized forward rates are then computed for 71 discrete maturities following the expression (12), instead of (5).

$$F_j = \left[ x_j\left(\frac{y_j}{360}\right) - x_l\left(\frac{y_l}{360}\right)\right](x_j - x_l)360\,,\ x_j > x_l \tag{12}$$

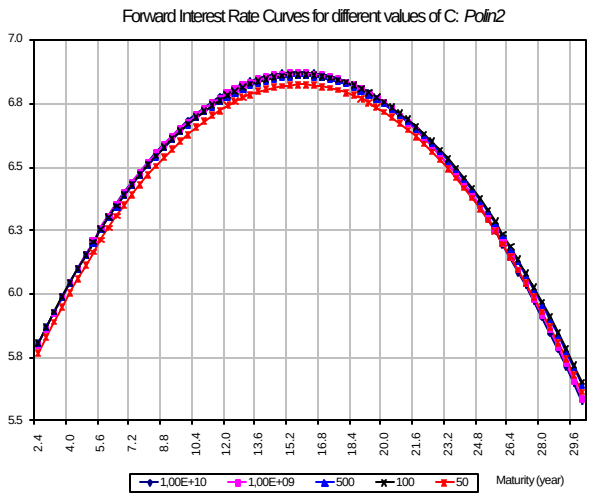
Figure 5 exhibits the five forward curves generated by each one of the eight kernels. One can see that the forward rate curve takes the shape of the kernel, controlled by the parameter $C$.


Forward Interest Rate Curves for different values of C: *Splinf*
2000
1000
500
100
50
Maturity (year)

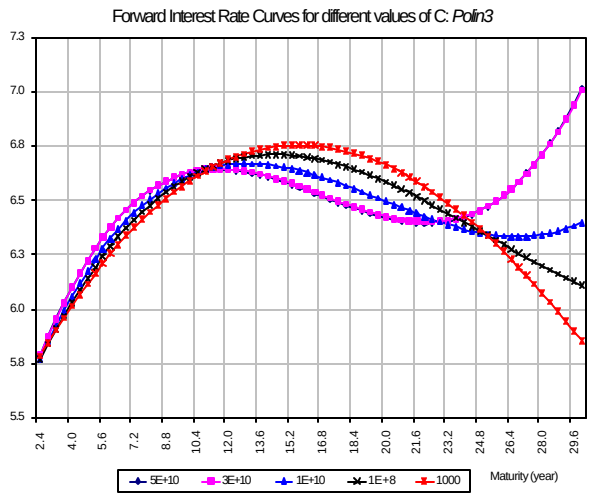
(a)


Forward Interest Rate Curves for different values of C: *Polvok*
100000
50000
5000
100
50
Maturity (year)

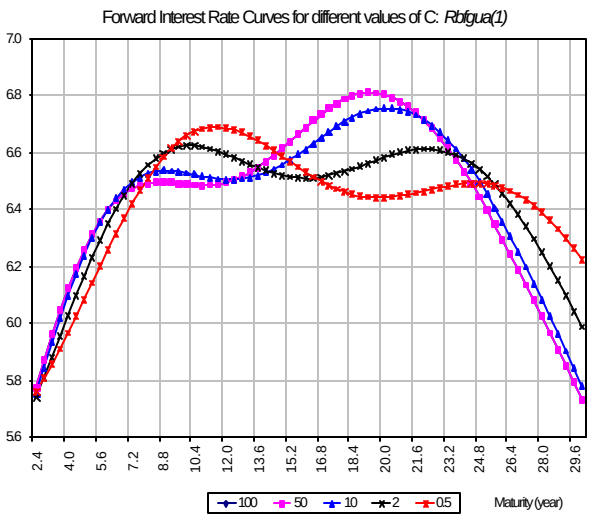
(b)


Forward Interest Rate Curves for different values of C: *Polin2*
1,00E+10
1,00E+09
500
100
50
Maturity (year)

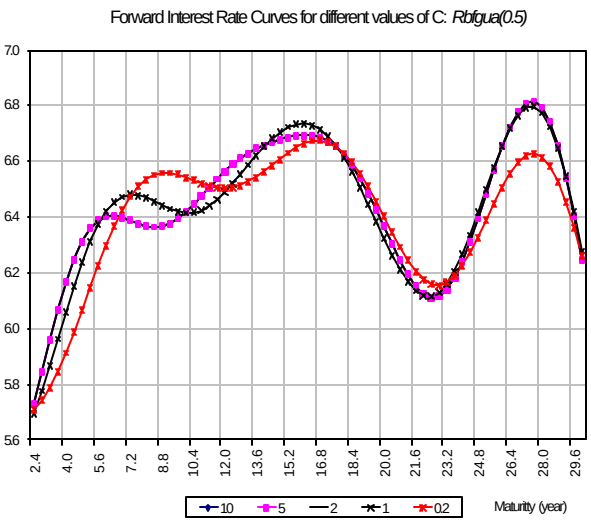
(c)


Forward Interest Rate Curves for different values of C: *Polin3*
5E+10
3E+10
1E+10
1E+8
1000
Maturity (year)

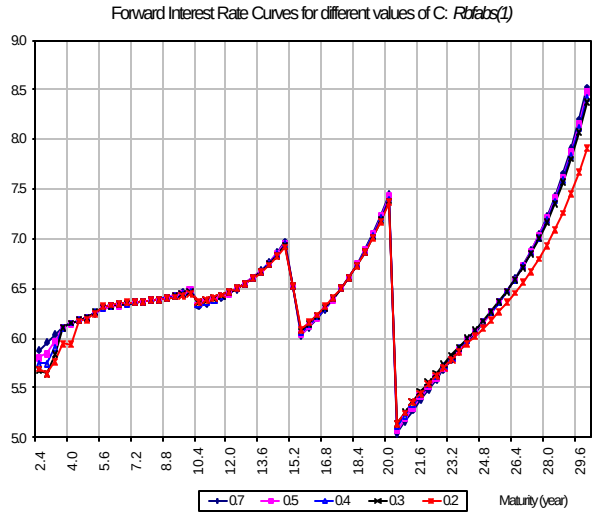
(d)


Forward Interest Rate Curves for different values of C: *Rbfgua(1)*
100
50
10
2
0.5
Maturity (year)

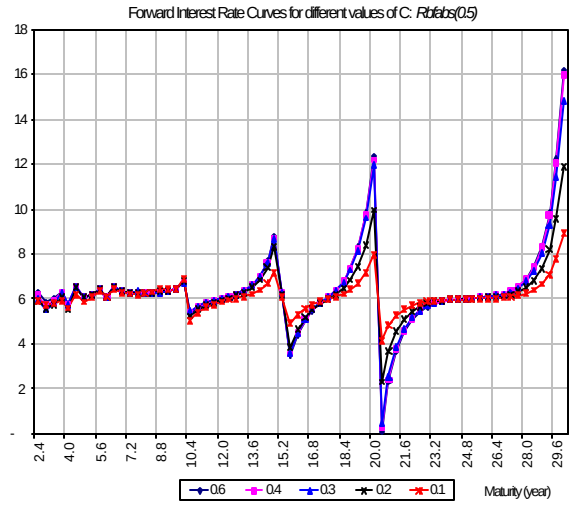
(e)


Forward Interest Rate Curves for different values of C: *Rbfgua(0.5)*
10
5
2
1
0.2
Maturity (year)

(f)




(g)




(h)

Figure 5: Forward curves generated by different kernels varying the value of the parameter *C*: (a) *Splinf*; (b) *Polvok*; (c) *Polin*2; (d) *Polin*3; (e) Rbfgua (1); (f) Rbfgua(0,5); (g) *Rdfabs*(1); (h) *Rdfabs*(0,5).

The only model that generates a discontinuous implicit forward curve is the *rdfabs*-SVR. The shapes of the curves generated by the *polvok, polin3 and rdfgua*-SVR are very sensitive to *C*. For the SVR equipped with radial-basis-function kernels the greater the parameter ω is, the more sensitive to *C* the model becomes. The kernels of the type *splinf* and *polin2* were able to generate almost *C*-insensitive implicit forward curves.

Therefore, only one kernel, *splinf*, was able to model the spot interest rate curve generating a proper forward rate curve. Nevertheless, this kernel does not present constructive asymptotical behavior for long-maturity rates: it only uses the information presents in the data. The composition of an admissible SVR kernel that imposes asymptotical trend for long-term rates might be an interesting research point in the future.

Those qualitative conclusions might be valid for generic interest rate curve, not only for swaps curves.

## 5 – Data

The dataset was extracted from the Bloomberg system. Its primary database is composed by bid and ask rates input by some swap-market dealers for the standard maturities.[5] These are the rates at which these dealers were willing to make a deal, quoted at the end of each day. Bloomberg, subsequently, runs some liquidity filter and releases only one bid rate and one ask rate per maturity. The use of different values of the parameter ε is reinforced because the SVR has to deal with different dealer credit risks in the same estimation.

[5] Such as, Greenwich Cap. Markets, Prebon Yamane US and Tullett & Tokyo.

The dataset consists of annualized bid and ask interest rates for the 12 standard maturities (in years): 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 and 30. Thus, every estimation will count on twelve observed points.

The precision of the rates is fixed in accordance with the minimum quoting unit: 0.001%. The BAS and the mean rates are computed following the equations (7) and (8), respectively. The sample begins in March 3rd, 1997 and finishes on April 30th, 2001. Initially, the sample was composed by 1,073 days. Due to data inconsistency, 26 days were left out. Three other days were eliminated because of outlier bid-ask spreads. The final sample, therefore, is composed by 1,024 days. Relevant international financial crises took place during the considered time interval: Asian (1997), Russian and Long Term Capital hedge fund (1998) and Brazilian (1999). Despite these financial crises, the swap curves maintained a well-behaved shape. However, their bid-ask spreads changed their pattern after the 1998 crisis.

Table 2 presents the BAS descriptive statistics over the sample. Two *BAS* values, 0.030% and 0.040%, showed high sample frequencies: 26.7% and 58.9%, respectively. Between 0.030% and 0.050% are 95% of the whole *BAS* values.

Table 2: *BAS* descriptive statistics

| Statistics | $BAS_i$ (%) | $BAS_i / T_{M,i}$ |
|---|---|---|
| Mean | 0.039 | 0.6% |
| Minimum | 0.006 | 0.1% |
| Maximum | 0.150 | 3.0% |
| Median | 0.040 | 0.6% |
| Mode | 0.040 | 0.5% |
| Standard Deviation | 0.010 | 0.2% |

Table 3 shows the descriptive statistics per year. One can identify two significant movements: increase of the mean and the volatility (1997-1999); mean stabilization with volatility reduction (2000-2001).

Table 3: *BAS* descriptive statistics per year (%)

| Statistics / Year | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|
| Mean | 0.033 | 0.039 | 0.042 | 0.040 | 0.040 |
| Minimum | 0.010 | 0.010 | 0.010 | 0.006 | 0.022 |
| Maximum | 0.140 | 0.150 | 0.150 | 0.094 | 0.060 |
| Median | 0.030 | 0.040 | 0.040 | 0.040 | 0.040 |
| Mode | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 |
| Standard Deviation | 0.007 | 0.013 | 0.014 | 0.004 | 0.001 |

Table 4 exhibits the same statistics from the above table, but comprising each one of twelve maturities. On average, the longest three contracts show the greatest mean bid-ask spreads. However, as indicated by the last row, when the annual mean *BAS* is divided by its relative annual mean rate the difference vanishes. This result is in accordance with findings commented in the subsection 3.2 that the BAS depends on the rate level. The volatility, on the other hand,

is greater for the shortest maturity contracts. Malhotra (1998) remarks that this behavior is due to liquidity concentration on these contracts.

Table 4: *BAS* descriptive statistics per contract maturity (%)

| Statistics / Maturity (year) | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean (1) | 0.038 | 0.038 | 0.038 | 0.038 | 0.037 | 0.039 | 0.039 | 0.037 | 0.038 | 0.042 | 0.041 | 0.041 |
| Minimum | 0.006 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| Maximum | 0.150 | 0.150 | 0.150 | 0.150 | 0.130 | 0.120 | 0.120 | 0.130 | 0.150 | 0.140 | 0.150 | 0.140 |
| Median | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 | 0.040 |
| Mode | 0.011 | 0.011 | 0.011 | 0.011 | 0.008 | 0.013 | 0.012 | 0.007 | 0.011 | 0.007 | 0.008 | 0.007 |
| (1) / $T_{M, year}$ | 0.63% | 0.62% | 0.61% | 0.61% | 0.59% | 0.62% | 0.61% | 0.58% | 0.59% | 0.63% | 0.63% | 0.63% |

Some of the eight kernels impose normalization for the training vectors I arbitrarly selected a value to suit all of them: the maturities were divided by 40.

## 6 – The SVR Estimation Performance and Descriptive Capability

This section is dedicated to estimate swap curves and to analyze the estimation (in-sample) results and the SVR capacity of describing time series evolution of the curves by using support vectors as a statistic sufficient set.

The parameters $\varepsilon_i$ are fixed to match the half of the bid-ask spread of the swap contracts. Table 5 exhibits descriptive statistics for the numbers of support vectors (SV's) required by each SVR specification across the sample. The *splinf*-SVR demanded the smallest numbers of SV's in order to describe the curves and achieved low variability around these means. On average, the *splinf*-SVR required around 4 to 5 SV's. It means that 33% to 42% of swap contracts were necessary to estimate the curves with the desired accuracy established by the contract bid-ask spreads. This somewhat high percentage, however, was expected since one needs at least 3 points to describe a curvature (25%).

Only the *polin2* and *ploin3*-SVR showed relative high variability in the SV figures according to the *C* value and the largest standard deviation values. On the other hand, the radial-basis-function SVR were the less *C*-sensitive models in terms of demanded number of SV's.

Table 6 presents the means of the SV's located out of the tube. The first line tells that no SV was out of the tube when *C* took the most flexible-data-fitting value (*Cs*). As one expects, the number of these SV increased as the value of *C* decreased (from *C1* to *C5*). The most important conclusion drawn from this table is that the number of SV`s outside the tube, on average, is smaller compared with the total number of SV‘s. Then, the number of SV`s on the boundary is relevant.

Table 5: Descriptive statistics for the number of support vectors required per kernel and *C* value

| Kernel | *Splinf* | *polvok* | *polin2* | *polin3* | *rdfgua(1)* | *rdfgua(0.5)* | *rdfabs(1)* | *rdfabs(0.5)* |
|---|---|---|---|---|---|---|---|---|
| *C* | | | | Mean | | | | |
| *Cs* | 4.6 | 5.1 | 8.7 | 7.1 | 5.3 | 6.2 | 7.3 | 9.9 |
| *C1* | 4.6 | 4.9 | 8.7 | 7.1 | 5.2 | 6.1 | 7.4 | 9.9 |
| *C2* | 4.7 | 4.9 | 8.7 | 7.1 | 5.3 | 6.2 | 7.4 | 9.9 |
| *C3* | 4.8 | 5.0 | 8.7 | 4.7 | 5.4 | 6.3 | 7.6 | 9.9 |
| *C4* | 5.1 | 5.6 | 8.7 | 5.4 | 5.6 | 6.2 | 7.7 | 9.9 |
| *C5* | 5.6 | 6.2 | 6.0 | 5.8 | 6.0 | 6.3 | 7.8 | 9.9 |
| *C* | | | | Standard Deviation | | | | |
| *Cs* | 1.3 | 2.0 | 4.3 | 4.2 | 1.5 | 1.3 | 1.6 | 1.3 |
| *C1* | 1.3 | 1.4 | 4.3 | 4.2 | 1.3 | 1.2 | 1.6 | 1.3 |
| *C2* | 1.3 | 1.4 | 4.3 | 4.2 | 1.3 | 1.3 | 1.6 | 1.3 |
| *C3* | 1.4 | 1.5 | 4.3 | 1.6 | 1.4 | 1.3 | 1.5 | 1.3 |
| *C4* | 1.7 | 2.1 | 4.3 | 2.3 | 1.6 | 1.3 | 1.5 | 1.3 |
| *C5* | 2.0 | 2.4 | 2.9 | 2.7 | 1.9 | 1.5 | 1.5 | 1.3 |

Table 6: Mean of the number of support vectors outside the tube

| C / Kernel | *splinf* | *polvok* | *polin2* | *polin3* | *rdfgua(1)* | *rdfgua(0.5)* | *rdfabs(1)* | *rdfabs(0.5)* |
|---|---|---|---|---|---|---|---|---|
| *Cs* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *C1* | 0.18 | 0.29 | 0.08 | 0.17 | 0.18 | 0.12 | 0.05 | 0.04 |
| *C2* | 0.39 | 0.39 | 0.28 | 0.85 | 0.26 | 0.23 | 0.23 | 0.35 |
| *C3* | 0.55 | 0.74 | 1.15 | 0.96 | 0.69 | 0.50 | 0.41 | 0.84 |
| *C4* | 1.09 | 1.72 | 2.70 | 1.86 | 1.15 | 0.66 | 0.71 | 2.21 |
| *C5* | 1.66 | 2.28 | 2.38 | 2.38 | 2.07 | 1.66 | 1.17 | 4.16 |

Because of the reduced number of SV for relevant kernels, the algorithm proposed by Downs, Gates and Masters (2001) to reduce the number of SV has limited use in swap curves. However, stylized facts about interest rate curves suggest that linear combination among rate is more likely observed in shortest and longest maturities. Thus, this algorithm might be useful for modeling interest rate curves extracted from high number of securities.

In order to be consistent with the SVR estimation, the in-sample performance will be evaluated by two measures close related to $\varepsilon_i$-insensitive loss function: $\varepsilon_i$-corrected mean absolute error (MAE-$\varepsilon_i$) and $\varepsilon_i$-corrected interpolation MAE (IMAE-$\varepsilon_i$). The last measure does not consider the erros from 2 and 30-year contracts.

$$MAE - \boldsymbol{e}_i = \frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|y_{t,i} - \hat{y}_{t,i}\right|_{e_{t,i}}$$

$$IMAE - \boldsymbol{e}_i = \frac{1}{T(n-2)}\sum_{t=1}^{T}\sum_{i=2}^{n-1}\left|y_{t,i} - \hat{y}_{t,i}\right|_{e_{t,i}}$$

Where $n$=12, $T$=1024 and $\left|\,.\,\right|_{e_i}$ is defined by equation (10).

Table 7 shows the values of the error measures for each SVR. On average, all the SVR face difficulties in the estimation of the extreme-position contracts in the curves. Allowing for the most flexible curves (*Cs*-SVR), the mean absolute errors are zero.

Table 7: The in-sample performance measures per kernel and *C* value.

| Kernel | *splinf* | *polvok* | *polin2* | *polin3* | *rdfgua(1)* | *rdfgua(0.5)* | *rdfabs(1)* | *rdfabs(0.5)* |
|---|---|---|---|---|---|---|---|---|
| *C* | | | | MAE-$\varepsilon_i$ (%) | | | | |
| *Cs* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *C1* | 0 | 0 | 0 | 0 | 0 | 0 | 0,001 | 0,001 |
| *C2* | 0.001 | 0 | 0 | 0 | 0 | 0 | 0.001 | 0.003 |
| *C3* | 0.001 | 0.001 | 0 | 0 | 0.001 | 0.001 | 0.003 | 0.008 |
| *C4* | 0.004 | 0.005 | 0.001 | 0.001 | 0.003 | 0.003 | 0.006 | 0.021 |
| *C5* | 0.006 | 0.007 | 0.007 | 0.005 | 0.009 | 0.013 | 0.011 | 0.051 |
| *C* | | | | IMAE-$\varepsilon_i$ (%) | | | | |
| *Cs* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *C1* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *C2* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.001 |
| *C3* | 0 | 0 | 0 | 0.001 | 0 | 0 | 0 | 0.002 |
| *C4* | 0.001 | 0.002 | 0.001 | 0.002 | 0 | 0.001 | 0 | 0.009 |
| *C5* | 0.002 | 0.003 | 0.005 | 0.003 | 0.004 | 0.003 | 0.004 | 0.031 |

The *splinf*-SVR was the best model: it managered in the most effective way the trade-off between the model complexity (number of support vectors) and the data fitting (in-sample performance measures).

Data reduction is an important tool commonly used in finance to describe interest rate curve movements. Litterman and Scheinkman (1993), for instance, apply principal components analysis to empirically show that two or three independent factors are needed to describe the movements of the curve. Another alternative is to monitor estimated parameters of parametric model.

At first glance, the support vectors could be used in this task, since they are sufficient statistics of the curve in the cross-section. The first challenge would be to assure that the number of SV's was kept unchanged through the sample. One could take advantage of MVS-ν proposed by Schölkopf, Bartlett, Smola, and Williamson (1998). However, Tables 4 and 5 weaken the hypotheses behind this model. Besides, one cannot enforce the same contract to be a SV for the entire sample. The problem arises because it does not make financial sense to monitor different maturity rates under the same label. Therefore, I could not find out a way to enforce the SVR be a useful tool in data reduction targeting the monitoring of the curves movements through time.

Figure 6 exhibits the frequency in which each one of the twelve contracts was appointed as a SV by the *C3*-SVR (I chose *C3* because it represents an intermediate regularization). For example, the *splinf*-SVR took the 30-year contract as a SV in 98% of the days. One can identify a common U-shape for all the SVR specifications: the percentages decrease up to the 7 or 8-year contracts and begin to increase till the longest maturity contract. The *rdfabs(0.5)* and *polin2* SVR`s got away from the mean shape. The most demanded contracts were those located

in the extreme positions on the curve as expected. The analysis of the outside-the-tube-SV frequency reveals that this kind of SV concentrates on the curve beginning.

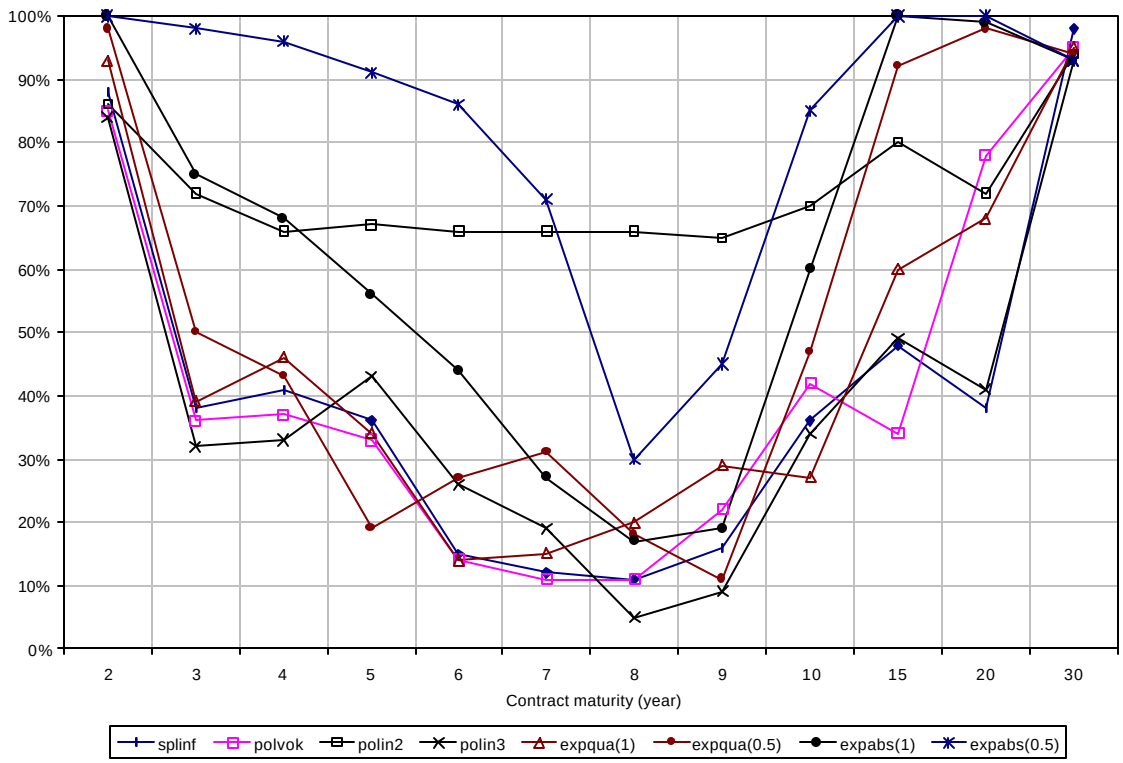



Figure 6: The sample frequencies in which the contract was chosen as a support vector

## 7 – The SVR Out-of-sample Performance

This section examines the out-of-sample performance of the SVR specifications used in the last section. The performance will be measured by the leave-one-out cross-validation strategy proposed by Stone (1974). Table 8 presents MAE-$\varepsilon_i$ and IMAE-$\varepsilon_i$. [6]

On average, extrapolation was a more difficult job than interpolation for all the SVR specifications. Again, the *splinf*-SVR achieved both the best results and the lowest sensitivity to the choice of the parameter *C*. The performances of the *rdfabs*-SVR's were really poor compared to the others. These results are consistent with the in-sample analysis made in the previous section. From Table 8 the best *splinf*-SVR was obtained by using *C1*. However, the other specifications achieved similar results.

[6] Some SVM required a long time for be concluded or simply faced unfeaseble solution. Thus, these days do not take part in the error statistics. Since the numbers of the times that it happened were not significant (0,2% to 3,5%, depending on the kernel), the results are still valid. The number of days that suffer from these problems are in the following per kernel: *rdfgua*(1): 17; *rdfgua*(0,5): 36; *rdfabs*(1): 17; *rdfabs*(0,5): 2.

Table 8: Out-of-sample performance measures per kernel and C value.

| Kernel | *splinf* | *polvok* | *polin2* | *polin3* | *rdfgua(1)* | *rdfgua(0.5)* | *rdfabs(1)* | *rdfabs(0.5)* |
|---|---|---|---|---|---|---|---|---|
| *C* | | | | EAM-ε (%) | | | | |
| *Cs* | 0.017 | 0.612 | 0.038 | 0.038 | 0.904 | 0.040 | 0.056 | 0.105 |
| *C1* | 0.017 | 0.070 | 0.038 | 0.038 | 0.030 | 0.038 | 0.056 | 0.105 |
| *C2* | 0.017 | 0.053 | 0.036 | 0.036 | 0.030 | 0.038 | 0.056 | 0.107 |
| *C3* | 0.017 | 0.047 | 0.020 | 0.020 | 0.029 | 0.038 | 0.057 | 0.109 |
| *C4* | 0.020 | 0.036 | 0.030 | 0.030 | 0.029 | 0.038 | 0.060 | 0.114 |
| *C5* | 0.023 | 0.037 | 0.031 | 0.031 | 0.033 | 0.049 | 0.067 | 0.122 |
| *C* | | | | EAMI-ε (%) | | | | |
| *Cs* | 0.003 | 0.027 | 0.007 | 0.007 | 0.143 | 0.019 | 0.027 | 0.072 |
| *C1* | 0.002 | 0.009 | 0.006 | 0.006 | 0.007 | 0.017 | 0.027 | 0.072 |
| *C2* | 0.003 | 0.008 | 0.006 | 0.006 | 0.007 | 0.017 | 0.028 | 0.074 |
| *C3* | 0.003 | 0.006 | 0.003 | 0.003 | 0.006 | 0.017 | 0.029 | 0.076 |
| *C4* | 0.004 | 0.007 | 0.007 | 0.007 | 0.006 | 0.017 | 0.031 | 0.081 |
| *C5* | 0.005 | 0.010 | 0.009 | 0.009 | 0.010 | 0.023 | 0.036 | 0.090 |

The swap-curve-estimation problem strongly shows how important the choice of the kernel is. This point is made not only by the empirical results (out-of-sample performance and number of demanded SV), but also by the quality requirements on the implicit forward rate curve.

## 8 – The SVR vs Other Methods: Out-of-sample Performance

In this section, I will compare the best SVR specification (*splinf*-SVR) with three other approximation methods. The following methods were selected in order to offer a large variety of approximation approaches: four types of cubic splines, the Nelson and Siegel (1987) parametric regression model and the Nadaraya-Watson (1964) non-parametric regression model.

The most commonly used spline in finance is the cubic splines because it is able to generate a differentiable forward rate. Four different cubic splines will be used: not-a-knot condition, periodic, complete and second. For further details about splines, the reader may consult de Boor (1978). Typically, choices concerning knots are made based on re-sampling. Because swap curves are well behaved on average and they offer few training vectors, I set all of them as knots. This means that splines will act as a perfect-data-fitting model.

The parsimonious parametric model by Nelson and Siegel (1987) is a widely used model for interest rate curve estimation. Barret, Gosnell and Heuson (1995), Willner (1996) and Ferguson and Raymar (1998) are some examples. It was specially develop to perform this task. This model was specifically developt to estimate interest rate curves. It explains a curve with 3 comprehensive parameters: level (*l*), slope (*s*) and curvature (*c*).

$$y(x) = l + (-s + c)\left[1 - \exp\left(\frac{-x}{\boldsymbol{t}}\right)\right]\frac{\boldsymbol{t}}{x} - c\exp\left(\frac{-x}{\boldsymbol{t}}\right),$$

where $\tau$ is a constant that controls the rate of exponential decay. Fixing a value for $\tau$, ordinary least squares can estimate the three parameters of the model. In fact, $\tau$ might also become a parameter to be estimated. The four-parameter model, however, requires a non-linear estimation process. Some empirical results, such as Nelson and Siegel (1987) and Barret, Gosnell and Heuson (1995), claim that the it causes overfitting problems for the US Treasury curve. Here, the value of $\tau$ will be predetermined.

Nadaraya (1964) and Watson (1964) proposed simultaneously and independently the following non-parametric regression:

$$f(x) = \frac{\sum_{i=1}^{n} K(x_i, x) y_i}{\sum_{i=1}^{n} K(x_i, x)},$$

where k is a kernel. One common choice for k is the Gaussian kernel:

$$K(x_i, x_j) = \frac{1}{h\sqrt{2\pi}} \exp\left( -\frac{(\mathrm{x}_i - x_j)^2}{2\mathrm{h}^2} \right),$$

where, $h$ is the bandwidth parameter. Small values for $h$ introduces noises into the estimated function $f$. The greater the parameter $h$, the smoother the function $f$ is and the closer to the sample mean of $y$ it is. As usual, $h=\omega\sigma$, where the parameter $w$ will be pre-established and $s$ is sample standard deviation of the training vectors $x$.

I call attention to the different rules played by kernel functions in the Nadaraya-Watson and the SVR's. The selection of both the Nadaraya-Watson's bandwidth parameter $h$ and the Nelson-Siegel's parameter $\tau$ will be based on cross-validation as discussed below.

A consistent comparison of these six different methods is not straightforward. The basic difficulty arises from the different cost functions used by each one of them. The SVR seeks to minimize a cost function composed by the $\varepsilon_i$-insensitive loss function and a regularization restriction. The four splines are like perfect-data-fitting models. And, finally, the Nelson-Siegel model tries to minimize the mean squared error. Therefore, a error measure might generate unfair comparison and lead to wrong conclusions.

One way to keep the imbalances under control is to make use of different error measures to cover the wide spread of the models' target. Thus, the following measures will be added to MAE-$\varepsilon_i$ and IMAE-$\varepsilon_i$: root of mean squared error (RMSE), interpolation root of mean squared error (IRMSE), percent mean absolute error (PMAE) and interpolation percent mean absolute error (IPMAE). Again, the errors $(y - \hat{y})$ will be generated by leave-one-out.

$$RMSE = \sqrt{\frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(y_{t,i} - \hat{y}_{t,i}\right)^2}$$

$$IRMSE = \sqrt{\frac{1}{T(n-2)}\sum_{t=1}^{T}\sum_{i=2}^{n-1}\left(y_{t,i} - \hat{y}_{t,i}\right)^2}$$

$$PMAE = \frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|\left(\frac{y_{t,i} - \hat{y}_{t,i}}{y_{t,i}}\right) - 1\right| 100$$

$$IPMAE = \frac{1}{T(n-2)}\sum_{t=1}^{T}\sum_{i=2}^{n-1}\left|\left(\frac{y_{t,i} - \hat{y}_{t,i}}{y_{t,i}}\right) - 1\right| 100,$$

where, $n$=12 and $T$=1024.

The following three tables (9 to 11) present the error statistics for the four splines, the Nelson-Siegel model and the Nadaraya-Watson regression model. For the last two models, six different values were tested for the parameters $h$ and $\tau$.

Table 9: The out-of-sample performance measures of the cubic splines

| Statistics / Splines | not-a-knot | periodic | complete | second |
|---|---|---|---|---|
| MAE | 0.074 | 0.530 | 0.058 | 0.028 |
| PMAE (%) | 1.12 | 7.91 | 0.87 | 0.43 |
| MAE-$\varepsilon_i$ | 0.064 | 0.519 | 0.047 | 0.018 |
| RMSE | 0.364 | 2.188 | 0.265 | 0.070 |
| IMAE | 0.015 | 0.024 | 0.014 | 0.011 |
| IPMAE (%) | 0.23 | 0.37 | 0.21 | 0.17 |
| IMAE-$\varepsilon_i$ | 0.006 | 0.015 | 0.005 | 0.003 |
| IRMSE | 0.039 | 0.056 | 0.034 | 0.021 |

Table 10: The out-of-sample performance measures of Nelson-Siegel

| Statistics / $\tau$ value | 0.06 | 0.08 | 0.1 | 0.12 | 0.14 | 0.5 |
|---|---|---|---|---|---|---|
| MAE | 0.020 | 0.019 | 0.020 | 0.020 | 0.020 | 0.028 |
| PMAE (%) | 0.31 | 0.30 | 0.31 | 0.32 | 0.32 | 0.44 |
| MAE-$\varepsilon_i$ | 0.007 | 0.007 | 0.008 | 0.008 | 0.008 | 0.016 |
| RMSE | 0.028 | 0.029 | 0.030 | 0.031 | 0.031 | 0.055 |
| IMAE | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.016 |
| IPMAE (%) | 0.24 | 0.23 | 0.23 | 0.23 | 0.23 | 0.26 |
| IMAE-$\varepsilon_i$ | 0.004 | 0.003 | 0.003 | 0.003 | 0.003 | 0.005 |
| IRMSE | 0.021 | 0.020 | 0.020 | 0.020 | 0.020 | 0.024 |

Table 11: The out-of-sample performance measures of Nadaraya-Watson

| Statistics / $\lambda$ value | 1.00 | 0.50 | 0.30 | 0.20 | 0.10 | 0.05 |
|---|---|---|---|---|---|---|
| MAE | 0.111 | 0.081 | 0.055 | 0.038 | 0.026 | 0.025 |
| PMAE (%) | 1.80 | 1.31 | 0.89 | 0.61 | 0.42 | 0.40 |
| MAE-$\varepsilon_i$ | 0.093 | 0.063 | 0.038 | 0.023 | 0.014 | 0.014 |
| RMSE | 0.153 | 0.112 | 0.079 | 0.059 | 0.045 | 0.044 |
| IMAE | 0.103 | 0.072 | 0.046 | 0.029 | 0.019 | 0.018 |
| IPMAE (%) | 1.65 | 1.16 | 0.74 | 0.46 | 0.29 | 0.29 |
| IMAE-$\varepsilon_i$ | 0.085 | 0.054 | 0.029 | 0.014 | 0.008 | 0.008 |
| IRMSE | 0.134 | 0.091 | 0.057 | 0.039 | 0.029 | 0.029 |

The second spline achieved the best performance among the splines. As they are perfect-data-fitting models, there is no natural error measure. The second cubic spline was selected because it was the one that achieved the lowest error according to the majority of the measures. In the same way, the best specification for the Nadaraya-Watson regression was the one with $\lambda$=5. On the other hand, as OLS seeks to minimize the squared error, the RMSE is the natural criterion to select the best value for the Nelson-Siegel's parameter $\tau$. The model with $\tau$=0.06 was selected. In order to ease the visual comparison, the best specification of each method has its performance measures consolidated in Table 12.

Table 12: Out-of-sample performance measures of best specifications of each method

| Statistics / Methods | SVR (splinf, C1) | Second cubic spline | Nadaraya-Watson ($\lambda$ = 0.05) | Nelson-Siegel ($\tau$ = 0.06) |
|---|---|---|---|---|
| MAE | 0.030 | 0.028 | 0.025 | 0.020 |
| PMAE (%) | 0.47 | 0.43 | 0.40 | 0.31 |
| MAE-$\varepsilon_i$ | 0.017 | 0.018 | 0.014 | 0.007 |
| RMSE | 0.054 | 0.070 | 0.044 | 0.028 |
| IMAE | 0.015 | 0.011 | 0.018 | 0.015 |
| IPMAE (%) | 0.23 | 0.17 | 0.29 | 0.24 |
| IMAE-$\varepsilon_i$ | 0.002 | 0.003 | 0.008 | 0.004 |
| IRMSE | 0.019 | 0.021 | 0.029 | 0.021 |

The first four lines of Table 12 point Nelson-Siegel as the best performing method. Its MAE achieved 0.020. It is a good performance since the minimum quoting unit is 0.001. When this number becomes a percentage of the sought interest rates, the PMAE shows that the mean absolute error was 0.31% of the rates. When considering the bid-ask spreads in the error measure (MAE-$\varepsilon_i$), this model was able the produce a 0.007 mean absolute error.

However, when the analysis is restricted to interpolation performance, the Nelson-Siegel is no longer the best performing method according to all the measures. It suggests that the outstanding performance of this method was due to better extrapolation, in particular at the longest maturity, this model enforces asymptotical behavior on the long-term interest rates.

Reducing the analysis to interpolation job, the second cubic spline exhibits the best performance in terms of mean (the smallest IMAE and IPMAE). The *splinf*-SVR achieved the best performance measured by IMAE-$\varepsilon_i$ and IRMSE. Two conclusions arise from these figures. First, when considering only the errors outside the interval defined by the bid and ask rates around the mean rates, the *splinf*-SVR was the best model. In this sense, the BAS's were useful

information for the estimation process. Second, it also achieves the best performance in controlling the trade-off bias-variance. Its extrapolation performance, however, recommend the construction of an admissible SVR kernel function that enforces an asymptotical trend for long maturities, as the Nelson-Siegel's model does. Finally, these results are conditional on the re-sampling strategy.

## 9 – Conclusion

This paper used the Support Vector Regression (SVR) to model interest rate curve, a very important financial economic problem. Despite the numerous advances in the theory, the number of applications of SVR to real problems is small. Curve estimation offers some natural hints for SVR model selection (loss function, parameter C and kernel), but requires a modification of its loss function.

The SVR was able to introduce in the estimation process two kinds of *a priori* information about the problem: the smoothness and stability of the implicit forward rate curve, and the liquidity and the price formation process of the assets from which the curve is extracted. The bid-ask spreads of the contracts measured the latter kind of information. McCulloch (1987) summarizes the role played by BAS on asset prices (page 189): "The problem is that the value of a security is ambiguous within its bid-ask spread."

The Vapnik $\varepsilon$-insensitive loss function was chosen to allow for both the introduction of bid-ask spreads and for sparse solutions. The parameter $\varepsilon$ was then fixed as half of bid-ask spread. Because contracts in the same estimation might present different values for BAS, a change was made to Vapnik's loss function: each training pattern might have its own value for $\varepsilon$. It was showed that this indexing does not change any $\varepsilon$-insensitive SVR properties. The sparsity opened a window to analyze the financial meaning of those contracts selected as the support vectors – they are sufficient statistics for the curve.

The selection of the kernel and the parameter $C$ are guided by requirements over the forward rate curve. They both define the smoothness and stability of the implicit forward rate curve. Even without empirical results, one can notice that kernel selection is a key aspect in SVR on interest rate curves. A proposition was presented to reduce the search space for the optimal value of the parameter C based on its role in quadratic programming optimization. One can extend this proposition for regularization network model selection. Proper ranges of C values were identified for different kernels.

In particular, this paper worked on the dollar-Libor swap interest curve. This curve was selected for its increasing importance in financial markets, especially after the hedge fund crisis in 1998, and for the rich set of information contained in the bid-ask spreads of swap contracts. Empirical research points out that the maturity and liquidity of the contract, the bank credit risk, the level of Treasuries interest rates and the payment frequency have explanatory power on the swap-contract bid-ask spreads.

The dataset was collected from Bloomberg. The sample frequency is daily, from March 3, 1997 to April 30, 2001. Eight kernel functions were used to estimate the curves. The kernel

generating spline with an infinite number of nodes was able to build a proper forward rate, in accordance with observed mean rates and bid-ask spreads. It described the curves with the smallest number of support vectors (SV's). On average, it demanded 4 to 5 contracts to describe the curve with desired accuracy measured by bid-ask spreads. The selection of contracts as SV's was dominated by their position on the curve. Despite the fact that the support vector set is a sufficient statistic of the curve in cross-section, they cannot be used as a tool to monitor curve movements through time.

This kernel also achieved the best leave-one-out performance. The shortest and longest maturity contracts were chosen as SV's most frequently. Then, I compared the best SVR to four kinds of cubic splines, the Nadaraya-Watson non-parametric regression model and the Nelson-Siegel parametric model. To accommodate the different objectives of these approximating methods, a variety of error measures were considered. The Nelson-Siegel model achieved the best global performance.

When extrapolated rates are not taken into consideration, the natural cubic spline got the smallest bias performance. The SVR achieved the best control of the trade-off bias-variance and the smallest bias when we consider as null the estimation errors inside the interval defined by the BAS around the mean rate. These results suggest that a new admissible SVR kernel enforcing asymptotical behavior over long maturities would probably increase SVR performance in estimating interest rate curves.

## References


Barret, B., Gosnell, T. Jr. and Heuson, A. (1995), "Yield Curves Shifts and the Selection of Immunization Strategies", Journal of Fixed Income, September, 53-64.

Bliss, R. (1997), "Testing Term Structure Estimation Methods", *Advances in Futures and Options Research*.

Bollier, T. and Sorensen, E. (1994), "Pricing Swap Default Risk', *Financial Analyst Journal*, May/June, 23-33.

de Boor, C. (1978). "A practical guide to splines", *Appl. Math. Sci*., v. 27, Springer, New York.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992), "A training algorithm for optimal margin classifiers". In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144-152, Pittsburgh, PA, ACM Press.

Breiman, L., (1994), "Bagging Predictors", *Technical Report 421*, Department of Statistics, University of California Berkeley, CA.

Brooks, R. and Malhotra, D. K. (1994), Components of the Bid-Ask Spread in Default-Risky Interest Rate Swap, *Advances in Futures and Options Research*, 7, 237-249.

Burges C. J. C. and Crisp, D. (1999), Uniqueness of the SVM Solution, *Annals of Neural Information Processing Systems.*

Burges, C. J. C. (1998), Geometry and Invariance in Kernel Based Methods. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods --- Support Vector Learning*, 89-116, Cambridge, MA, 1999. MIT Press.

Campbell, J. (2000), Asset Pricing at the Millennium, *Journal of Finance*, 55, 1515-1567.

Campbell, J., Lo, A. and Mackinley, A. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Chapados, N., Bengio, Y., Vicent, P., Ghosn, J., Dugas, C., Takeuchi, I, Meng, L., (2001), Estimating Car Insurance Premia: a Case Study in High-Dimensional Data Inference, *Annals of Neural Information Processing Systems Conference*.

Chambers, D. R., Carleton, W. T. and Waldman, D. W. (1984), "A New Approuch to the Estimation of the Term Strucutre of Interest Rates", Journal of Financial and Quantitative Analysis, 19, September, 233-252.

Cohen, K., Maier, S., Schwartz, R. and Whitcomb, D. *(1986)*, The Microstructure of Securities Markets, Prentice-Hall, Englewood Cliffs, NJ.

Cohen, K. J., Kramer, R. L. and Waugh, W. H. (1966), "Regression Yield Curves for US Government Securities", Management Science, 13, Dezember, B-168-B175.

Collin-Dufresne, P. and Solnik, B. (1999), On the Term Structure of Default Premia in the Swap and LIBOR Markets, *Working paper*, Carnegie Mellon University.

Collobert R., and Bengio S. (2001), SVM torch: Support Vector Machines for Large – Scale Regression Problems, *Journal of Machine Learning Research*, 1.

Cooper, I. and Mello, A. (1991), "The Default Risk of Swaps", *Journal of Finance*, 46, 597-620.

Dai, Q and Singleton, K. (2000), "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, 55, 6, 1943-1978.

Davis, D. and Holt, C. (1993), "Experimental Economics", Princeton University Press, Princeton, NJ.

Dermody, J. and Prisman, E. (1988), "Term Structure Multiplicity and Clientele in Markets with Transactions Costs and Taxes", The Journal of Finance, XLII, 4.

Dobson, S. W. (1978), "Estimating Term Structure Equations with Individual Bond Data", Journal of Finance, 33, March, 75-92.

Downs, T., Gates, K. and Masters, A. (2001), "Exact Simplification of Support Vector Solutions", Journal of Machine Learning Research, vol. 2, 293-297.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. and Vapnik, V. (1997), "Support vector regression machines", Advances in Neural Information Processing Systems, 9, 155–161, Cambridge, MA. MIT Press.

Duffie, D. and Huang, M. (1996), "Swap Rates and Credit Quality", Journal of Finance, 51, 921-949.

Duffie, D. and Singleton, K. (1997), "An Economic Model of the Term Structure of Interest-Rate Swap Yields", Journal of Finance, Vol. LII, No. 4, 1287-1321.

Durand, D. (1942), "Basic Yields of Corporate Bonds", Technical Paper nº 3, NBER, Cambridge, Mass.

Echols, M. E. and Elliott, J. W. (1976), "A Quantitative Yield Curve Model for Estimating the Term Structure of Interest Rates", Journal of Financial and Quantitative Analysis, 11, March, 87-114.

Evgeniou, T., Pontil, M. and Poggio, T. (2000), "Regularization Networks and Support Vector Machines", Advances in Computational Mathematics, 13, 1, 1-50.

Ferguson, R. and Raymar, R. (1998), "A Comparative Analysis of Several Popular Term

Structure Estimation Models", Journal of Fixed Income, March, 17-33.

Fischer, D. (1966), "Expectation, the Term Structure of Interest Rates, and Recent British Experience", Economics 33, August, 319-329.

Fisher, M., Nychka, D. and Zervos (1995), "Fitting the Term Structure of Interest Rates with Smoothing Splines", Working Paper 95-1, Finance and Economics Discussion Series, Federal Reserve Board.

Frankel, J., Galli, G., and Giovannini, A. (1996), "The Microstructure of Foreign Exchange Markets" (NBER Conference Report), University of Chicago Press, Chicago, IL.

Friedman, J. (1991), "Multivariate Adaptive Regression Splines", Annal of Statistics, 19, 1-141.

Gestal, T., Suykens, J., Baestaens, D., Lambrechets, A., Lanckiet, G., Vandaele, B., Moor, B., Vandewalle, J. (2001), "Financial Time Series Prediction usinh Least Squares Support Vector Machines within the Evidence Framework", IEEE Transactions on Neural Networks, Special Issue on Neural Networks in Financial Engineering, vol. 12, 4, pp. 809-821.

Girosi, F., Jones, M. J. and Poggio, T. (1993), "Priors, Stabilizers and Basis Functions: from regularization to radial, tensor and additive splines", AI Memo 1430, CBCL Paper 75. June.

Gouriéroux, C., Jasiak, J. and Fol, G. (1999), "Intra-Day Market Activity", Journal of Financial Markets, 2, 193-226.

Grinblatt, M. (1999), "An Analytical Solution for Interest Rate Swap Spreads", Reivew of International Finance.

He, H., (2000), "Modeling Term Structures of Swap Spread", working paper 00-16, Yale School of Management.

Huber, P. J. (1981), "Robust Statistics", John Wiley and sons, New York.

Litterman, R. and Scheinkman, J. (1993), "Common Factors Affecting Bond Returns", Journal of Fixed Income, 1, 51-61.

Liu, J., Longstaff, F. A. and Mandell, R. E. (2000), "The Market Price of Credit Risk: An Empirical Analysis of Interest Rate Swap Spreads", SSRN Working Paper.

Lo, A. (1995), "The Industrial Organization and Regulation of the Securities Industry", NBER Conference Report, University of Chicago Press, Chicago, IL.

Malhotra, D. K. (1998), "The Impact of Interest Rate Reset Period on the Bid-Offer Rates in a Interest rate swap contract – an Empirical Investigation", Journal of Multinational Financial Management, 8, 79-88.

Mattera, D. and Haykin, S. (1999), "Support vector machines for dynamic reconstruction of a chaotic system", Advances in Kernel Methods – Support Vector learning, 211 – 242,

Cambridge, MA. MIT press.

McCulloch, J. H. (1971), “Measuring the Term Structure of Interest Rates”, Journal of Business, 34, January, 19-31.

McCulloch, J. H. (1975), “The Tax-Adjusted Yield Curve”, Journal of Finance, June, 811-830.

McCulloch, J. H. (1987), “The Monotonicity of the Term Premium – A Closer Look”, Journal of Financial Economics, 18, 185-192.

McNulty, J. (1990), “Pricing of Interest Rate Swaps”, Journal of Financial Services Research, March, 53-63.

Mercer, J. (1909), “Fundations of positive and negative typeand their connection with the theory of integral equations”. Philos. Trans. Roy. Roy. Soc. London, A 209:415-446.

Muller, K. R., Smola, A. J., Ratsch, G., Scholkopf, B., Kohlmorgen, J. and Vapnik. V. N. (1997), “Predicting Time Series with Support Vector Machines”. Artificial Neural Networks- ICANN 97, 999 – 1004, Berlin. Springer Lecture Notes in Computer Science.

Nadaraya, E. A. (1964), “On estimating regression”. Theory Prob. Appl., 10, 186-190.

Nelson, C. R., and Siegel, A. F. (1987), “Parsimonious Modeling of Yield Curves.” Journal of Business, 60, 473-489.

Nielsen, S. and Ronn, E. (1996), “The Valuation of Default Risk in Corporate Bonds and Interest Rate Swaps”, Working paper, University of Texas at Austin.

Nilsson, N. J. (1965), “Learning machines: Foundations of trainable Pattern Classifying Systems”. McGraw-Hill.

O´Hara, M. (1995), “Market Microstructure Theory”, Blackwell Publishers, Cambridge, MA.

Schölkopf, B., Bartlett, P., Smola, A. J., and Williamson, R. (1998). “Support vector regression with automatic accuracy control”. L. Niklasson, M. Bodén, and T. Ziemke, editors, Proceedings of ICANN'98, Perspectives in Neural Computing, 111-116, Berlin. Springer Verlag.

Scholkopf, B., Simard P. Y., Smola, A. J., and Vapnik, V. N. (1998), “Prior Knowledge in support vector kernels”. Advances in Neural Information Processing Systems, 10, 640 – 646, Cambridge, MA. MIT Press.

Shea, G. (1985), “Interest Rate Term Structure Estimation with Exponential Splines: A Note”, The Journal of Finance, XL, 1.

Smola, A. J. and Scholpkopf, B. (1998a), “A Tutorial on Support Vector Regression”,

Technical Report Series, NeuroCOLT2.

Smola, A. J. and Schölkopf, B. (1998b), "From regularization operators to support vector kernels", Advances in Neural Information Processing Systems, 10, 343-349, San Mateo, CA, 1998.

Smola, A. J., Schölkopf, B. and Muller, K. R. (1998a), "The connection between regularization operators and support vector kernels". Neural Networks, 11, 637-649.

Smola, A. J., Schölkopf, B. and Müller, K. R. (1998b). "General cost functions for support vector regression". In T. Downs, M. Frean, and M. Gallagher, editors, Proc. of the Ninth Australian Conf. on Neural Networks, 79 - 83, Brisbane, Australia.

Smola, A. J., Murata, N., Scholkopf, B. and Muller, K. R. (1998), "Asymptotically optimal choice of $\varepsilon$-loss for support vector machines", 8ª International Conference on Artificial Neural Networks, Perspectives in Neural Computing, pages 105 – 110. Berlin. Springer Verlag.

Stone, M. (1974), "Cross-validatory choice and assessment of statistical predictions (with discussion)", Journal of the Royal Statistical Society, 36, 111-147.

Sun, T., S. Sundaresan, and Wang, C. (1993), "Interest Rate Swaps: An Empirical Investigation", Journal of Financial Economics, 34, 77-99.

Suykens, J.A.K. (2000), "Least Square Support Vector Machine for Classification and Non-linear Modelling", Neural Network World, vol.10, pp. 29-48.

Vapnik, V. N. (1979), "Estimation of Dependencies Based on Empirical Data", Nauka, Moscow (in Russian). (in english: Vapnik, V. N. (1982), "Estimation of Dependences Based on Empirical Data". Springer, NY)

Vapnik, V. N. (1995), "The Nature of Statistical Learning Theory", Springer, New York.

Vapnik, V. N. and Chervonenkis, A. J. (1968), "On the uniform convergence of relative frequencies of events to their probabilities", Doklady Akademii Nauk USSR 181 (4).

Vapnik, V. N. and Chervonenkis, A. J. (1971), "On the uniform convergence of relative frequencies of events to their probabilities", Theory Probab. Apl., 16, 264-280.

Vapnik V. N. e Chervonenkis, A. J. (1974), "Theory of Pattern Recognition", Nauka, Moscow (in Russian).

Vasicek, O. A. and Fong, H. G. (1982). "Term Structure Modeling Using Exponential Splines", Journal of Finance, 37, pp. 339-356.

Waggoner, D. (1997), "Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices", Federal Reserve Bank of Atlanta, working paper 97-10.

Watson, G. S. (1964). “Smooth Regression Analysis”. Sankhya, Ser. A, 26, 101-116.

Willner, R. (1996), “A New Tool for Portfolio Managers: Level, Slope and Curvature Duration”, Journal of Fixed Income, June, 48-59.

## Appendix

### A - The $\varepsilon_i$-insensitive Loss Function and SVM Properties

This appendix contains the proof that the $\varepsilon_i$-insensitive loss function does preserve the $\varepsilon$-insensitive SVR properties: convex quadratic programming problem with linear restrictions and sparsity of solution. In the following is presented the complete proof for the linear SVR regression. The proof for the non-linear case is analogous.

Consider the traditional SVR primal formulation.

$$\begin{aligned} &Minimize \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \\ &sa: \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \ , i = 1,\ldots, n \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \ , i = 1,\ldots, n \\ \xi_i, \xi_i^* \ge 0 \ , i = 1,\ldots, n \end{cases} \end{aligned}$$

The starting point is to replace the $\varepsilon$-insensitive loss function above by $\varepsilon_i$-insensitive defined in (8). The only difference appears in indexed variables $\varepsilon_i$.

$$\begin{aligned} &Minimize \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \\ &sa: \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon_i + \xi_i \ , i = 1,\ldots, n \\ \langle w, x_i \rangle + b - y_i \le \varepsilon_i + \xi_i^* \ , i = 1,\ldots, n \\ \xi_i, \xi_i^* \ge 0 \ , i = 1,\ldots, n \end{cases} \end{aligned}$$

Next, the associated Lagrangean is written with the same multipliers $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \ge 0$.

$$L \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{n}\alpha_i\left(\varepsilon_i + \xi_i - y_i + \langle w, x_i \rangle + b\right) - \sum_{i=1}^{n}\alpha_i^*\left(\varepsilon_i + \xi_i^* + y_i - \langle w, x_i \rangle - b\right) - \sum_{i=1}^{n}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right) \qquad \text{(A1)}$$

Once partial derivatives of L with respected to primal variables ($w$, $b$, $\xi_i$, $\xi_i$*) must be zero, four equations are written.

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\left(\boldsymbol{a}_i^* - \boldsymbol{a}_i\right) = 0 \qquad \text{(A2)}$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n}\left(\boldsymbol{a}_i - \boldsymbol{a}_i^*\right)x_i \qquad \text{(A3)}$$

$$\frac{\partial L}{\partial \boldsymbol{x}_i} = 0 \Rightarrow \boldsymbol{h}_i = C - \boldsymbol{a}_i, \text{ i} = 1,..., \text{ n} \tag{A4}$$

$$\frac{\partial L}{\partial \boldsymbol{x}_i^*} = 0 \Rightarrow \boldsymbol{h}_i^* = C - \boldsymbol{a}_i^*, \text{ i} = 1,..., \text{ n} \tag{A5}$$

The introduction of (A4) and (A5) into the last right-side term of equation (A1) vanishes with the multipliers $\eta_i$, $\eta_i$*, $i$=1, ..., $n$. Then, this term becomes

$$-C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) + \sum_{i=1}^{n}\xi_i\left(\alpha_i + \alpha\right)_i^*$$

By returning the above modified term to (A1), the slack variables $\xi_i$ and $\xi_i^*$ disappear.

$$L \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i\left(\varepsilon_i - y_i + \langle w, x_i\rangle + b\right) - \sum_{i=1}^{n}\alpha_i^*\left(\varepsilon_i + y_i - \langle w, x_i\rangle - b\right)$$

The last equation is re-arranged into:

$$L \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\varepsilon_i\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n}y_i\left(\alpha_i - \alpha_i^*\right) - \sum_{i=1}^{n}\langle w, x_i\rangle\left(\alpha_i - \alpha_i^*\right) + b\sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right)$$

The last term of the right side vanishes because of (A2). The vector $w$ is replaced by (A3) and making use of $\|w\|^2 = \langle w, w\rangle$, the Lagrangean is modified to the following expression.

$$L \equiv \frac{1}{2}\left\langle \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i, \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i \right\rangle - \sum_{i=1}^{n}\varepsilon_i\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n}y_i\left(\alpha_i - \alpha_i^*\right) - \sum_{i=1}^{n}\left\langle \sum_{j=1}^{n}\left(\alpha_j - \alpha_j^*\right)x_j, x_i \right\rangle\left(\alpha_i - \alpha_i^*\right)$$

The properties that define a mapping as an inner product assure the next relation.

$$\left\langle \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i, \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i \right\rangle = \sum_{i=1}^{n}\left\langle \sum_{j=1}^{n}\left(\alpha_j - \alpha_j^*\right)x_j, x_i \right\rangle\left(\alpha_i - \alpha_i^*\right) = \sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j\rangle$$

Replacing the first term of the Lagrangean by the above equation, it becomes

$$L \equiv -\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j\rangle - \sum_{i=1}^{n}\varepsilon_i\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n}y_i\left(\alpha_i - \alpha_i^*\right)$$

The preservation of non-negativity of the dual Lagrangean multipliers $(\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0)$ implies the following groups of inequalities.

$$0 \le \alpha_i \le C, \text{ i} = 1,..., \text{ n}$$
$$0 \le \alpha_i^* \le C, \text{ i} = 1,..., \text{ n}$$

Now, the dual formulation can be expressed by (A6).

$$\begin{aligned} &Maximize: -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle - \sum_{i=1}^{n} e_i(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\ &subjected\,\text{to}\begin{cases}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, \text{C}]\end{cases} \end{aligned} \tag{A6}$$

As in the traditional SVR formulation, the primal variables ($\xi$, $\xi^*$ and $w$) and the dual variables ($\eta$ and $\eta^*$) do not appear in (A6). (A6) is still a quadratic programming problem with linear restrictions. Thus, it presents one global solution. The comparison between (A6) and (6) shows that the unique difference is that the parameter $\varepsilon$ receives the index $i$ and it appears inside the second sum of the cost function.

The SVR expansion of function $f$ is written as in the original formulation:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b$$

The following Karush-Kuhn-Tucker conditions came from (A6):

$$\alpha_i(e_i + x_i - y_i + \langle w, x_i\rangle + b) = 0$$

$$\alpha_i^*(e_i + x_i^* + y_i - \langle w, x_i\rangle - b) = 0$$

$$(C - \alpha_i)x_i = 0$$

$$(C - \alpha_i^*)x_i^* = 0$$

The last two equations keep unchanged from traditional formulation. Therefore, the sparsity property of the $\varepsilon$-insensitive SVR is preserved. The first two equations are changed only by indexing the parameter $\varepsilon$. Hence, the computation of the parameter $b$ is unchanged ÿ