Estimating Interest Rate Curves by Support Vector Regression

André d'Almeida Monteiro*

Gávea Investimentos

Abstract

A model that seeks to estimate an interest rate curve should have two desirable capabilities in addition to the usual characteristics required from any function-estimation model: it should incorporate the bid-ask spreads of the securities from which the curve is extracted and restrict the curve shape. The goal of this paper is to estimate interest rate curves by using Support Vector Regression (SVR), a method derived from the Statistical Learning Theory developed by Vapnik (1995). The motivation for this is that SVR features extra capabilities at a low estimation cost. The SVR is specified by a loss function, a kernel function and a smoothing parameter. SVR models the daily US dollar interest rate swap curves, from 1997 to 2001. As expected from previous analysis, the SVR equipped with the kernel generating a spline with an infinite number of nodes was the best performing SVR. Comparing this SVR with other models, it achieved the best cross-validation interpolation performance in controlling the bias-variance trade-off and generating the lowest error considering the desired accuracy fixed by the bid-ask spreads.

Key words: Interest rate curves, Bid-ask spread, Support Vector Regression, Interest rate swaps

JEL Classification: C14, E43

* Correspondence: Rua Dias Ferreira, 190, 701, Leblon. Rio de Janeiro, RJ, Brazil. 22431-050. Fax: + 55 21 3206-9001. email: amonteiro@gaveainvest.com.br

1. Introduction

An interest rate curve can be briefly described as a function that establishes a relationship of dependence between interest rates and their associated maturities. It is a fundamental tool in a market economy because it deeply impacts the investment and consumption decisions of firms, householders and governments. It plays a crucial role in asset pricing, in companies' hedge strategy and in the monetary policy set by the country's central bank.

The estimation of curves from the interest rates and maturities of fixed income securities traded on the markets is an essential and daily task for investors because of, for example, derivative pricing, risk monitoring and the search for arbitrage opportunities.

Durant (1942) was the first to estimate an interest rate curve. In hindsight, his work was artisanal: he tried to estimate the US curve by drawing it based on his subjective definition of reasonability. Since then, a considerable amount of research has been done on interest rate curve estimation. A large variety of parametric and non-parametric models has been applied to this problem.

A model that seeks to estimate an interest rate curve should have two desirable capabilities in addition to the usual characteristics required from any function-estimation model: the ability to incorporate the bid-ask spreads of the securities from which the curve is extracted and to restrict the curve shape in the estimation.

The transaction of a security is not fully described by the transaction price: the bid and the ask (or offer) prices also contain valuable information. The bid-ask spread (or bid-offer spread) is the difference between the ask price and the bid price. McCulloch (1987), on page 189, summarizes the role played by the bid-ask spread on asset prices: "(…) the value of a security is ambiguous within its bid-ask spread." Even if the economic value of a security is unchanged, transaction prices can fluctuate within the bid and ask prices. Some factors can explain the existence of bid-ask spreads, such as transaction cost and liquidity. The bid-ask spread causes some problems for curve estimation. Thus, they should be considered during the estimation.

A spot rate is the interest rate payable on a spot loan. A forward rate is the interest rate negotiated today for a loan in the future. An estimation model should also be able to generate a forward curve shape that fulfills two conditions. First, as claimed by Bliss (1997) and by Fisher, Nychka and Zervos (1995), among others, the forward rate curve should be smooth through maturity. Second, the forward curve should be asymptotically flat at long maturities. Livingston and Jain (1982) and Nelson and Siegel (1987) proved that this requirement is appropriate if forward rates are finite.

The available models usually focus on the restriction of the curve. The bid-ask spreads, however, have received less attention because of the optimization cost to incorporate them into estimation.

The goal of this paper is to estimate interest rate curves by using Support Vector Regression (SVR), a method derived from the Statistical Learning Theory developed by Vapnik (1995). The motivation is that SVR is able to simultaneously restrict the forward curve shape and incorporate the bid-ask spreads into the estimation at a very low optimization cost.

SVR is a non-parametric and distribution-free estimation model. It expands the target function in a linear combination of the explanatory variables after they have been preprocessed by a kernel function with special characteristics. The SVR cost function is comprised of two components: a loss function and a smooth regularization term. The estimation turns to be a quadratic optimization problem with linear restrictions. It thus has a global solution. On the other hand, SVR theory is not conclusive with regards to parameter selection: it relies upon re-sampling techniques.

SVR relates to kernel regression, like Nadarya-Wason, because of kernel functions. Nevertheless, the roles they play in these methods are very different. Less obvious, although stronger, is the linkage between SVR and other traditional methods such as smoothing splines, radial basis functions and Hodrick-Prescott filter.

Under a Bayesian framework, assuming Gaussian priors and considering an appropriate likelihood function, Sollich (1999) showed that the Gaussian SVR can be seen as a maximum *a posteriori* solution to the estimation problem. Following the same approach, Gao, Gunn and Harris (2002) provided the variance for the Gaussian-SVR predictions.

The SVR parameter selection requires the specification of a loss function, a kernel function and a smoothing parameter. We choose the ε-insensitive function, proposed by Vapnik (1995), as the loss function because it brings into the estimation the bid-ask spreads. The ε-insensitive function only penalizes the cost function if the absolute error is greater than ε. The ε-insensitive

SVR generates a sparse solution: only a fraction of the observations have non-null coefficients in the linear combination. These special observations are called support vectors.

Because securities in the same estimation might have different bid-ask spreads, a slight change is necessary in the ε-insensitive function: the parameter ε will be indexed to each observation. We will prove that this modification does not change any of the ε-insensitive-SVR properties. The bid-ask spreads will set the tolerance for deviation for each observation in the sample: the smaller the bid-ask spread, the more concerned with this rate the SVR should be by determining a tighter tolerance for deviation around the negotiated rate.

The kernel function and the smoothing parameter determine the shape of the implicit forward curve, while the SVR is estimating the spot curve. For instance, a second-degree-polynomial-kernel SVR generates a forward curve with the shape of a second degree polynomial. The kernel selection will be conducted on both theoretical and empirical basis.

Therefore, the task of SVR will be to estimate a function for the spot curve, restricting the shape of the implicit forward curve and trying to keep the observed rates inside the symmetrical intervals around the function determined by the associated bid-ask spreads.

SVR will estimate the spot curve US dollar-Libor interest rate swaps on a daily basis, from 1997 to 2001. Two reasons led to the choice for this security curve. First, it is the worldwide benchmark curve for derivative pricing that requires interest rates denominated in US dollars. The Bank for International Settlements estimated that swap notional principals amounted to US$55 trillion at the end of the first semester of 2005. Second, the swap bid-ask spreads are rich in information, such as liquidity, credit risk and transaction costs.

SVR was confronted against three other models typically applied to curve estimation. SVR slightly outperformed the others: it achieved the best cross-validation interpolation performance in controlling the bias-variance trade-off and generating the lowest error considering the desired accuracy fixed by the bid-ask spreads.

The present paper contributes to financial econometrics in the sense that it introduces SVR as a model for estimating interest curves that features the two desirable characteristics required for the problem. Empirically, SVR successfully estimated the US dollar swap. The paper also shows a new field of application for SVR, in which *a priori* information determines parameter selection, avoiding re-sampling techniques. Additionally, as far as we know, it is the first application of SVR on interest rate curves.

In addition to this introduction, this paper has seven sections. The second Section poses and discusses the problem of estimating the interest rate curve based on US dollar swaps. In the third Section, we introduce and derive the SVR. The way SVR handles the bid-ask spreads and the curve shape restrictions is discussed in Section 4, during the SVR parameter selection. The data are introduced in the fifth Section. In the sixth Section, we summarize how SVR models the problem. Section 7 investigates the SVR empirical performance. Finally, the eighth Section concludes this paper.

2. The Problem: To Estimate US Dollar-Libor Swap Curves

This Section breaks the problem down into two parts. The first Subsection describes the USD-Libor swaps, derivatives contracts on interest rates, and defines and discusses their bid-ask spreads. The second Subsection contains a brief survey of the literature concerning interest rate curve estimation: models, the role played by bid-ask spreads and the restrictions on the shape of forward curves.

2.1 Interest Rate Swaps

A fixed income security is an asset that promises to pay a cashflow in the future. The simplest type of fixed income asset is the one whose cashflow is completely known at issue date and makes a single payment at the end of its life. These assets are called zero-coupon bonds. A spot rate (or zero-coupon rate) is the interest rate payable on a spot loan. A forward rate is the interest rate negotiated today for a loan in the future. A security that pays coupon during its life is called a coupon bond. Its yield is the interest rate that equals its cashflow to its present price.

A forward rate can be understood as the price of money for a certain period in the future. Forward rates are implicit in spot rates. Conversely, a spot rate is the product of forward rates up to its maturity. A spot curve plots spot rates according to their maturities. It is also possible to define a forward curve in the same way. Indeed, the forward curve and the spot rate curve are equivalent ways of expressing the same term structure of interest rates. A yield curve plots yields according to the maturities of their bonds. One can extract the spot or forward curve from the yield curve. The focus of this paper is to directly model the spot curve, not to extract it from the yield curve.

An interest rate swap is a derivative on interest rate. The simplest type of interest rate swap is a contract under which two parties (called counterparties) agree to exchange the cashflows of interest payments without exchanging the underlying principal, called notional principal. The cashflows are calculated based on the notional principal by applying two different interest rates: a floating rate (T_{fl}) and a fixed rate (T_{fx}). Net payments are made periodically according to the net difference between the two interest rates multiplied by the notional principal. The contracts are negotiated at par rates, that is, the fixed rate defines the coupon rate.

While the counterparties agree on the value of the fixed rate in the beginning of the contract, the floating rate is only known close to each periodic payment. The floating rate may be an index, such as a US Treasury rate or Libor (the London Interbank Offered Rate). We will model the curve of swaps in which the floating rate is the US dollar-Libor: the rate of interest at which banks borrow US dollar-denominated funds from other banks in the London interbank market.

The US dollar-Libor interest rate swap (simply swap, henceforth) has rapidly increased since its appearance in 1981. The Bank for International Settlements estimated that swap notional principals amounted to US$55 trillion at the end of the first semester of 2005.

The most important players in the swap market are companies, commercial banks and investment banks. The trades are oriented toward middle and long-term funding and hedging. These swaps are over-the-counter contracts, that is, they are not traded at organized exchanges. Some banks act as if they were market makers. They are always ready to quote two fixed interest rates for a swap contract: the bid and ask rates. The bid rate ($T_{fx,a}$) is the fixed rate at which the bank is willing to pay a cashflow tied to it and to receive cashflow tied to Libor. The ask rate ($T_{fx,b}$) is the fixed rate at which the bank is willing to receive a cashflow tied to it and to pay cashflow tied to Libor. The bid-ask spread (or bid-offer spread) of a contract i is then the difference between these two rates.

$$bid\text{ - }ask\ spread_i = T^i_{fx,b} - T^i_{fx,a} \tag{1}$$

The mean rate is defined by (2):

$$T_{M,i} = \frac{T^{i}_{fx,a} + T^{i}_{fx,b}}{2} \tag{2}$$

There are no theoretical models to explain the dynamics of intraday bid-ask spreads or their determinants for swap contracts.[1] By using regression, some authors have found empirical evidence that some variables have significant explanatory power for daily bid-ask spreads. All of them have concluded that bid-ask spreads are statistically different from zero. Sun, Sundersan and Wang (1993) noted that bid-ask spreads are sensitive to bank credit. Brooks and Malhotra (1994) showed that transaction costs, the maturity of the contract, the level of Treasury interest rates and payment frequency are relevant variables. Malhotra (1998) observed a positive relationship between bid-ask spreads and contract maturity. He attributed this behavior to the greater liquidity at low maturities.

2.2 Interest Rate Curve Estimation

A large amount of research has been conducted on interest rate curve estimation since the first attempt made by Durant (1942). Some authors have attempted to use parametric models, such as Cohen, Kramer and Waugh (1966), Echols and Elliott (1976), Dobson (1978) and Chambers, Carleton and Waldman (1984). Some of them applied polynomial regression and all of them imposed some kind of smoothness on the curve and generated unbounded interest rates for long-term maturity. Nelson and Siegel (1987) proposed a model that forces asymptotical long-term interest rates.

McCulloch (1971, 1975) introduced the cubic spline regression. Fisher, Nychka and Zervos (1995) added a smoothness parameter (or regularization parameter) to the cubic spline model that resulted in smoothing splines. The goal of this model was to minimize the trade-off between the mean square error and the smoothness of the estimated function. Waggoner (1997) increased this model's flexibility by allowing three different values for the regularization parameter over

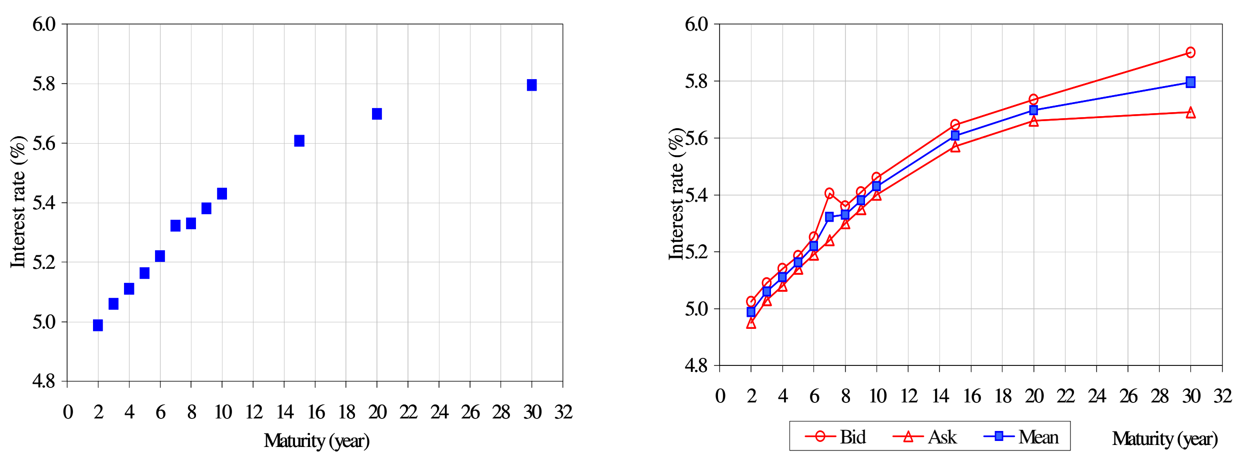
[1] Indeed, such models exist for stock bid-ask spreads. See, for example, O'Hara (1995).

the maturity domain. He reported that his model was able to outperform the original one in terms of data fitting.

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

The simplest way to deal with the bid-ask spread is to collapse it into the mean rate. Figure 1 illustrates one example of failure of the mean-rate solution. It exhibits the fixed rates of one day in the sample. Figure 1.a suggests that the seven-year mean rate is an outlier and it may damage the approximation. [2]

Figure 1.b reveals that, for some reason, the seven-year bid rate had a different behavior compared to the other bid rates, whereas its ask rate did not. It is possible to draw a curve inside the intervals defined by the bid and ask rates without treating the seven-year contract as an outlier. Thus, in order to handle more information about the security transaction, a model should fully incorporate the bid-ask spreads into the estimation.



(a) (b)

Figure 1: The interest rates of swap contracts in a sample day: (a) mean rates, (b) bid, ask, and mean rates.

[2] Throughout this paper, interest rates will be shown on an annual basis.

McCulloch (1971, 1975) went beyond the mean-rate solution. The author used the bid-ask spreads as weights in the weighted least-squared regression. However, this approach does not get to the core of problems generated by the bid-ask spreads.

A model should also be able to generate a forward curve that fulfills two conditions. First, as claimed by Bliss (1997) and by Fisher, Nychka and Zervos (1995), among others, the forward rate curve should be smooth through maturity. From the economic point of view, it is difficult to explain why the price of money (the interest rate) varies significantly for very close terms. In addition, the pricing of interest rate securities is sensitive to the stability of forward rates across maturity. Second, the forward curve should be asymptotically flat at long maturities. Livingston and Jain (1982) and Nelson and Siegel (1988) proved that this requirement is appropriate if forward rates are finite.

As pointed out by Abu-Mostafa (2001), forward curves implicit in similar spot curves might display very different shapes. As the cross-validation results will show in Section 6, keeping forward curves consistent with the two aforementioned conditions avoids bad performances.

3. Support Vector Regression

This Section is divided into four Subsections. In the first one, the linear SVR is introduced and derived. In the sequence, the nonlinearity is brought into the model. The SVR construction follows the usual approach in the support vector literature. The connection between SVR and other models in the estimation field is established in the third Subsection, where the SVR is reframed within a traditional mathematical framework. In the last Subsection, we focus on the Gaussian SVR in order to obtain a probabilistic interpretation for SVR and its prediction variance.

3.1 Linear SVR

Consider the problem of approximating the observed dataset composed by n observations, $A \equiv \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times \Re$, where X denotes the space of the explanatory or independent variables ($X = \Re^d$, for example) and $y \in \Re$ is the dependent variable. The SVR objective is to model a relationship of dependence between *x* and *y* through (1):

$$f(x) = \langle w, x \rangle + b \tag{3}$$

where *b* is a constant, $w \in \Re^{\mathrm{d}}$ and $\langle w, x \rangle$ denotes de inner product between *w* and *x* in X. Let ξ_i be the deviation of estimated value $f(x_i)$ from observed value y_i.

$$y_i = f(x_i) + \xi_i \tag{4}$$

The SVR loss function is selected in order to introduce the bid-ask spread into the estimation. Usually, there is a trade-off in the selection of the loss function. On one hand, we should select the loss function that best suits the problem. On the other hand, we should avoid loss functions that lead to a very difficult estimation process. The loss function that most suits the curve estimation problem, allowing for the bid-ask spread in the estimation, is exactly the one that leads to the simplest optimization problem: the ε-insensitive function proposed by Vapnik (1995).[3] The ε-insensitive loss function l_ε is defined by:

$$l_\varepsilon = \sum_{i=1}^{n} |\xi_i|_\varepsilon$$

$$|\xi|_\varepsilon \equiv \begin{cases} 0, \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon, \text{if } |\xi| > \varepsilon \end{cases} \tag{5}$$

The loss function in (5) only penalizes the cost function when the absolute deviation is greater than ε. Everything happens as if there were a ε-radius tube around the estimated function *f* (see Figure 2). If an observed value is inside the tube, even if not exactly over the estimated

[3] Muller, Smola, Ratsch, Scholkopf, Kohlmorgen and Vapnik (1997) and Smola, Scholkopf and Muller (1998b) show experimentally how loss-function-dependent SVR performance is.

value, the deviation is considered null. The parameter ε is understood as the desired accuracy for estimation. It will be set as the half of the bid-ask spread.

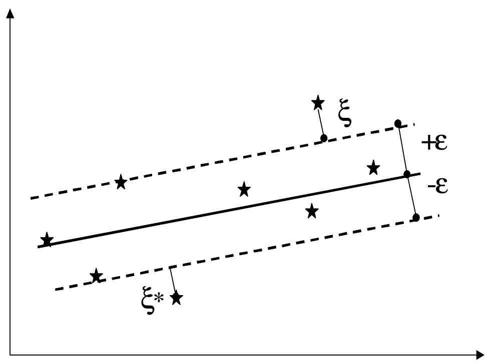


Figure 2: SVR ε-tube.

The main target of the ε-insensitive SVR is to find a function *f* that simultaneously seeks a maximum deviation equal to ε and maximum flatness. Flatness in the case of (3) means a small *w*. It can be achieved by minimizing the norm of *w*. Therefore, the SVR objective can be written as the following optimization problem. As suggested by Figure 2, the superscript * identifies the deviations below the tube.

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)$$
$$\text{subjected to}\begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{6}$$

where the constant $C > 0$ controls the trade-off between the function flatness and the tolerance for deviations. As it will be discussed in Subsection 3.3, this constant might be understood as the inverse of a smoothing or regularization parameter. For simplicity, we will call it the regularization parameter. The lower the C, the flatter the *f* function. Since the optimization in (6) has a convex cost function and the restrictions are linear, the SVR estimation is a convex quadratic programming problem. And, therefore, there is one global solution. Burges and Crisp (1999) discussed this issue in detail.

The Lagrange function from (6) is written in (7). The variables $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$ are the Lagrange multipliers related to restrictions.

$$L \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{n}\alpha_i\left(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b\right) - \sum_{i=1}^{n}\alpha_i^*\left(\varepsilon + \xi^*{}_i + y_i - \langle w, x_i\rangle - b\right) - \sum_{i=1}^{n}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right) \quad (7)$$

Following Vapnik (1995), the SVR estimation is formulated as the dual of (6).

$$\begin{aligned} &\text{Maximize}: \; -\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_{\mathrm{j}}^*\right)\left\langle x_i, x_j\right\rangle - \varepsilon\sum_{i=1}^{n}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n} y_i\left(\alpha_i - \alpha_i^*\right) \\ &\text{subject to} \begin{cases} \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \quad (8)$$

where the vectors α and α^* become dual Lagrangean multipliers. Notice that the primal Lagrange multipliers (w and the deviations) vanish from (6) to (8). The solution of (8) completely describes w as a linear combination of x_i.

$$w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) x_i \quad (9)$$

Replacing w by equation (9) in (3), we achieve the SVR expansion for f.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \quad (10)$$

The above expansion is sparse: only a portion of the n observations have non-zero coefficients, $(\alpha_i - \alpha_i^*) \neq 0$, in the linear combination. These observations are called support vectors. The support vectors are only located on the boundary of the ε-tube and outside it. The support vectors outside the tube might be seen as outliers given the desired accuracy fixed by ε.

In a sense, the number of support vectors is a measure of the complexity of the target function's representation in (10).

The constant b is computed from equation (1) when the observation is a support vector. Muller, Smola, Ratsch, Scholkopf, Kohlmorgen and Vapnik (1997) suggested taking the mean over the values of b calculated from each support vector.[4] We use the median, instead of the mean.

The selection of the ε-insensitive as the loss function has other advantages besides the introduction of the bid-ask spread into the estimation. This loss function determines the sparseness in the linear combination in (10). One can prove that the support vector coefficients are the same if the model is re-estimated excluding all non-support vector observations. Then, the support vector set might be seen as a sufficient statistic for set A (the interest rate curve in the case of this paper). This property offers an opportunity to investigate the role played by each one of the contracts in curve construction. The ε-insensitive makes the SVR estimates robust in the same that the influence of a single observation is bounded (Vapnik, 1995 and Smola and Schölkopf, 2004). Finally, this loss function leads to an efficient implementation of the associated optimization problem as shown above.[5]

3.2 Non-linear SVR

The nonlinearity is introduced following Aizerman, Braverman and Rozonoér (1964) and Nilsson (1965). The explanatory variables x_i are preprocessed by a map $\Phi : \mathrm{X} \rightarrow \Im$ into a feature space. The standard SVR is then applied on $\{(\Phi(x_1), y_1), \ldots, (\Phi(x_n), y_n)\} \subset \Im \times \Re$. The resulting model is still linear, but in the feature space $\Im$.

The major problem of the above construction is that the dimensionality of $\Im$ might make the estimation computationally unfeasible. Boser, Guyon and Vapnik (1992) use a trick to overcome this problem: the inner product of functions might be represented by a kernel function if it satisfies Mercer's condition (Mercer, 1909).

4 Theoretically the values of b calculated from each support vector should be exactly the same. However, there might be some small numerical differences.

5 In the case of a ε-insensitive loss function at the power $p > 1$, the robustness is lost because its derivative grows without bounds. The optimization is no longer a convex problem if $p < 1$.

$$k(x,x') \equiv \langle \Phi(x), \Phi(x') \rangle \tag{11}$$

Informally speaking, all $f \in L_2(\mathrm{X})$ must satisfy (12) for the kernel to be written as an inner product in some feature space. It means that the kernel must be positive definite.

$$\int_{X \times X} k(x,x') f(x) f(x') \geq 0 \tag{12}$$

Starting from kernels that satisfy Mercer's condition, Schölkopf, Burges and Smola (1999) derived some rules for composition of kernels that also satisfy this condition. They called them admissible SVR kernels.

The expression (11) can replace the inner products between explanatory variables in (8) and we can restate the SVR optimization problem as follows.

$$\begin{aligned} &\text{Maximize } \; -\sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) \\ &\text{subject to} \begin{cases} \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \tag{13}$$

Notice that it was not necessary to explicitly know Φ to restate the optimization problem. Thus, the SVR expansion of *f* becomes:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) . k(x_i, x) + b \tag{14}$$

The nonlinear SVR is equivalent to the linear one except for two points. First, the vector *w* is no longer available in the nonlinear case. However, the Fisher-Riesz theorem ensures that *w* is still uniquely defined in the weak sense by $\langle w, \Phi(x) \rangle$. Second, the SVR seeks the flatness of function *f* in the space $\Im$ generated by the kernel.

Schölkopf, Bartlett, Smola, and Williamson (1998) proposed the ν-SVR. In this approach, an accuracy level ν is endogenously adjusted by SVR in order to keep a pre-established number of support vectors.

SVR theory does not provide a definitive parameter selection strategy. The biggest difficulty lies in the relationship of interdependence between the kernel and the hyper-parameters C and ε. The parameter selection is typically based on re-sampling methods (cross-validation and bootstrap).

There are some applications of SVR in finance reported by the support vector literature. Chapados et al. (2001) model car insurance premia. Gestel et al. (2001), Tay and Cao (2003), Kim (2003) and Yang et al. (2004) applied SVR to predict financial time series. Perez-Cruz, Afonso-Rodriguez and Giner (2003) used SVR to estimate conditional volatility models for financial returns. As far as we know, this present paper is the first to approach curve estimation with SVR.

3.3 SVR and Other Models

SVR relates to kernel regression, like Nadarya-Wason (1964), because of kernel functions. Nevertheless, the roles they play in these models are very different. Its linkage with other traditional methods, some of them applied in financial problems, is less obvious, albeit stronger. SVR is written in a traditional mathematical framework in order to observe these linkages.

Assume that sample A is composed by independent and identically distributed observations, drawn from an unknown distribution $P(x, y)$. According to Vapnik (1982), the goal of an estimation method should be to find a function f that minimizes a functional called expected risk, R.

$$R[f] \equiv \int l(x, y, f(x))dP(x, y) \tag{15}$$

where $l(x, y, f(x))$ denotes a loss function determining how the estimation errors will be penalized. Because the distribution is unknown, we can only count on the available sample A to estimate f. The empirical risk is achieved if we replace the integral as follows.

$$R_{emp}[f] \equiv \frac{1}{n}\sum_{i=1}^{n} l(x_i, y_i, f(x_i)) \tag{16}$$

The minimization of (16) may lead to overfitting and bad out-of-sample performance if *f* belongs to a class of very flexible functions. Thus, it is necessary to consider a flexibility penalty in the estimation process. In the linear SVR case, the penalty term $\|w\|^2$ is added to the empirical risk. It leads to the regularized risk functional, R_{reg}, in the sense of Tikhonov and Arsenin (1977), Morozov (1984) and Vapnik (1982).

$$R_{reg}[f] \equiv R_{emp}[f] + \frac{\lambda}{2}\|w\|^2 \tag{17}$$

where $\lambda > 0$ is called the regularization constant.

By fixing $C = \frac{1}{\lambda n}$ and l as the ε-insensitive function in (5), it is easy to demonstrate that minimizing (17) is equivalent to minimizing (6). The non-linear-SVR analysis demands a regularization operator *P* in (17). Instead of seeking flatness in the feature space (recall that $w \in \Im$), the following functional seeks to optimize some smoothness criteria in the space of the explanatory variables.

$$R_{reg}[f] \equiv R_{emp}[f] + \frac{\lambda}{2}\|Pf\|^2 \tag{18}$$

where *P* is a regularization operator, according to Tikhonov and Arsenin (1977). Let *P* be a positive semidefinite operator such that $P : H \to D$, where *H* is the Hilbert space of the functions *f* under consideration, and *D* is an inner product space in which $\langle Pf, Pg \rangle$ is well defined for all $f \in H$. Smola, Schölkopf and Müller (1998a) explore the relationship between *P* and an SVR kernel that achieves the equivalence between (18) and (13).

Evgeniou, Pontil and Poggio (2000) investigated these linkages under another framework. Consider the functional in (19).

$$R_{reg}[f] \equiv R_{emp}[f] + \frac{\lambda}{2}\|f\|_k^2 \tag{19}$$

where $\|.\|_k^2$ is a norm in the reproducing kernel Hilbert space defined by a positive definite kernel *k*. A reproducing kernel Hilbert space is a Hilbert space *H* of functions defined over some bounded domain $X \subset \Re^d$ such that the functional $F_X[f] = f(x)$, for all $f \in H$, is linear and bounded. It is thus possible to show that to every *H*, there is a corresponding unique positive definite kernel function such that $f(x) = f(x) \bullet k(x, y)$ for all $f \in H$, where $\bullet$ denotes the scalar product in *H*. The function *k* is called the reproducing kernel of *H*. Indeed, Evgeniou, Pontil and Poggio (2000) showed that (19) might generalize (18).

Under the classical regularization theory, the solution of (19) is similar to (14). This result, however, is based on functional analysis arguments, which are only valid in asymptotical terms. Vapnik (1995) was the first to provide a theory – built on function analysis, probability and statistics – that solves (19) as (14) for finite sets.

Many traditional models, some of them applied in financial econometrics, have cost functions similar to (18). That is the case of radial basis functions, neural networks with decay networks (Bishop, 1995), the Hodrick-Prescott filter (Hodrick and Prescott, 1980) and smoothing splines (Fisher, Nychka and Zervos, 1995, for example).[6]

Despite of the similarity in cost function, the ε-insensitive SVR has an important advantage over smoothing spline. In a smoothing spline one needs to choose the knots, whereas in the SVR, this job is done by the bid-ask spreads through the parameter ε. The SVR approach also relates to Linton, Mammen, Nielsen and Tanggaard (2000) in the sense that the authors proposed a non-parametric approach based on a kernel smoothing model to estimate interest rate curve.

3.4 Probabilistic Framework for Gaussian SVR

Following Sollich (2000) and Gao, Gunn and Harris (2002), a probabilistic interpretation for SVR is available if we restrict the attention to Gaussian distributions. Reconsider equation (4) as

[6] See, for example, Campbell, Lo and Mackinley (1997)

follows: let $f(x)$ be the true function underlying the sample A and ξ the additive noise corrupting it. Assuming that $P(\xi) \propto \exp(-Cl(\xi))$, the conditional probability of A given $f(X)$ and ε-insensitive loss function is:

$$P[A / f(X)] = \left[\frac{C}{2(\varepsilon C + 1)}\right]^n \exp\left(-C\sum_{i=1}^{n} l\left(|\xi|_\varepsilon\right)\right) \tag{20}$$

Let $P[f(X)]$ be the *a priori* probability of the unknown function $f(X)$. It should give high probability only to those functions for which the regularization term in (19) is small. Thus,

$$P[f(X)] \propto \exp\left(-\frac{1}{2}\|f\|_k^2\right) \tag{21}$$

Define a Gaussian process as a stochastic process specified by a mean vector and a covariance matrix for any finite sample. Consider the Gaussian process with zero mean and a covariance matrix *K* evaluated on observations *X* through the kernel *k*: $K_{X,X} = [k(x, x')]$.[7]

Let *w* be the weight vector in the feature space defined by the kernel *k*. Smola (1998) showed that $f(X)^T K_{X,X}^{-1} f(X) = \|w\|_k$. Then, $P[f(X)]$ is written as:

$$P[f(X)] = \frac{1}{\sqrt{\det 2\pi K_{X,X}}} \exp\left(-\frac{1}{2} f(X)^T K_{X,X}^{-1} f(X)\right) \tag{22}$$

Relating equations (19) and (21) through the Bayesian Rule, the *a posteriori* probability of $f(X)$ turns to be:

$$P[f(X) / A] = \frac{G(C, \varepsilon)^n}{P(A)\sqrt{\det 2\pi K_{X,X}}} \exp\left(-C\sum_{i=1}^{n} l_\varepsilon(y_i - f(x_i)) - \frac{1}{2} f(X)^T K_{X,X}^{-1} f(X)\right) \tag{23}$$

[7] For the relationship between Gaussian Processes and kernel functions, see Rasmussen and Williams (2006).

where $G(C,\varepsilon)=C/2(\varepsilon C+1)$. It can be shown that the function that maximizes the *a posteriori* probability in (22), through the Maximum A Posteriori (MAP) estimative, has (14) as solution with the Gaussian kernel.

Gao, Gunn and Harris (2002) showed that the variance for predictions made by the Gaussian ε-insensitive SVR has two independent additive terms. The first one captures the uncertainty in the function due to the finite amount of data available. The authors found out an approximation and an upper bound for it.

$$\operatorname{var}\left(\hat{f}(z)\right)_1 \leq K_{z,z} - K_{X,z}^T K^{-1} K_{X,z} \tag{24}$$

where $z \notin X$ and $\hat{f}$ is the SVR model estimated over *A*. The second term can be seen as the uncertainty generated by the parameters *C* and ε. This term has an exact close form.

$$\operatorname{var}\left(\hat{f}(z)\right)_2 = \frac{2}{C^2} + \frac{\varepsilon^2(\varepsilon C+3)}{3(\varepsilon C+1)} \tag{25}$$

4. SVR Parameter Selection

This section addresses SVR parameter selection for the swap curve estimations. It is divided into two Subsections. The first Subsection works on the connection between the loss function and the bid-ask spread. The second Subsection investigates how the parameter *C* and the kernel jointly restrict the shape of the forward curve.

4.1 Bid-ask Spreads: Loss function

In order to apply SVR to curve estimation, it is necessary to allow for different bid-ask spreads in the same estimation. Thus, a modification in the ε-insensitive loss function is required: the parameter ε should be indexed to each sample observation. Equation (26) introduces the ε_i-insensitive loss function.

$$l_{\varepsilon_i} = \sum_{i=1}^{n} \left|\xi_i\right|_{\varepsilon_i}$$

$$\left|\xi_i\right|_{\varepsilon_i} \equiv \begin{cases} 0, \; if \; \left|\xi_i\right| \le \varepsilon_i \\ \left|\xi_i\right| - \varepsilon_i, \; if \; \left|\xi_i\right| > \varepsilon_i \end{cases} \tag{26}$$

This change in Vapnik's loss function does not damage any one of the SVR properties according the following proposition.

Proposition 1 *The SVR equipped with the ε_i-insensitive loss function (26) maintains all the optimization properties of an equivalent SVR with the ε-insensitive loss function: a convex quadratic programming problem with linear restriction that might generate a sparse solution.*

Proof: see the Appendix A.

The ε_i-insensitive SVR coincides with the Non-fixed and Symmetrical Margin SVR proposed by Yang, Chan and King (2002). One way to understand the estimation is to view the bid-ask spread as a measure of the "trustfulness" of the negotiated rate. The lower the *bid-ask spread*, the more trustful the rate is, and the more concerned with this rate the SVR should be by determining a tighter tolerance for deviation around the observed data – and vice-versa.

The rate at which the swap deal is made is not available, so we have to use the mean rate as the *y* variable. The parameter ε will be fixed as half of the bid-ask spread of each contract.

$$\varepsilon_{i,t} = \frac{bid\text{ - }ask\ spread_{i,t}}{2}, \; i = 1, \ldots, 12 \text{ and } t = 1, \ldots T. \tag{27}$$

Because of the lack of the traded swap rates, the tolerance intervals are symmetrical around the estimated function. However, if these rates were available, it would have been possible to built an SVR to deal with different and asymmetrical tolerance interval.

4.2 Curve Shape Restriction: Parameter *C* and Kernel

Because of the relationship between spot and forward rates, when SVR is estimating a spot curve, the kernel and the parameter *C* jointly determine the implicit forward curve. In fact, the kernel (affected by *C*) "gives its shape" to the implicit forward curve. For instance, an SVR equipped with the second degree polynomial as the kernel generates a forward curve with the shape of a second degree polynomial. Thus, the restriction on the shape of the forward curve will be determined by the selection of the kernel and the value of *C*.

The selection process will be conducted in two stages. In the first stage, we theoretically evaluate if a kernel function properly imposes the conditions on the forward rate. Because the parameter *C* impacts the kernel shape and the impact is data-dependent, the theoretical analysis is not conclusive. Thus, we need a second stage: empirical sensibility analysis of the kernel to the value of *C*.

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 (*rbfgua*)

$$k(x_i, x_j) = exp\left(\frac{(x_i - x_j)^2}{2\omega^2\sigma^2}\right),$$

where σ is the standard deviation of the maturities of the contracts and ω is a constant.

- Exponential radial-basis function (*rdfabs*)

$$k(x_i, x_j) = exp\left(\frac{|x_i - x_j|}{2\omega^2\sigma^2}\right)$$

where σ is the standard deviation of the maturities of the contracts and ω is a constant.

We arbitrarily selected two values for the ω parameter: 1 and 0.5. We are not concerned with setting optimal values for it, as we are only interested in illustrating its impacts on the forward curve.

It is possible to partially verify *a priori* how each kernel handles the requirements on forward curves. All the three polynomial kernel functions generate smooth curves. However, the second degree polynomial cannot generate the asymptotic trend for long-term rates. The Vovk's real infinite and the third degree polynomials may generate this trend depending on *C*. The curve generated by the spline with an infinite number of nodes is smooth and asymptotically flat at long-term rates if this trend is present in the sample. The radial-basis functions are not able to produce a forward curve that is asymptotically flat at long-term rates. Depending on *C* and kernel parameters, the Gaussian radial-basis function may generate a smooth curve. The exponential radial basis function produces a non-smooth forward curve independent of *C* because of the discontinuity point at the support vector maturities.

The empirical analysis of the sensitivity of the kernel to the parameter *C* will enlarge the above conclusions. The following proposition, which reduces the search space for the optimal value of *C*, will be very helpful for this analysis.

Proposition 2 *Given the fixed values of* ε_i*, i=1,.., n and the kernel, there exists only one value of C, called* C_S*, such that any* $C \geq C_S$ *generates the same SVR coefficients* α_i, α_i^*, $i = 1, \ldots, n$*, and that a* $C = C_S - \delta < C_S$, $\delta > 0$, *generates at least one different SVR coefficient everything else being kept constant. Having estimated SVR with a high C value, called C', in the sense that no inequality restriction in (13) achieves the upper limit,* $\forall\, \alpha_i, \alpha_i^* < C', i = 1, \ldots, n$, C_S *is defined as the maximum value among the estimated dual coefficients* α_i, α_i^*, $i = 1, \ldots, n$*:*

$$Cs \equiv \max_i (\alpha_i, \alpha_i^*),\ i = 1, \ldots, n \qquad (28)$$

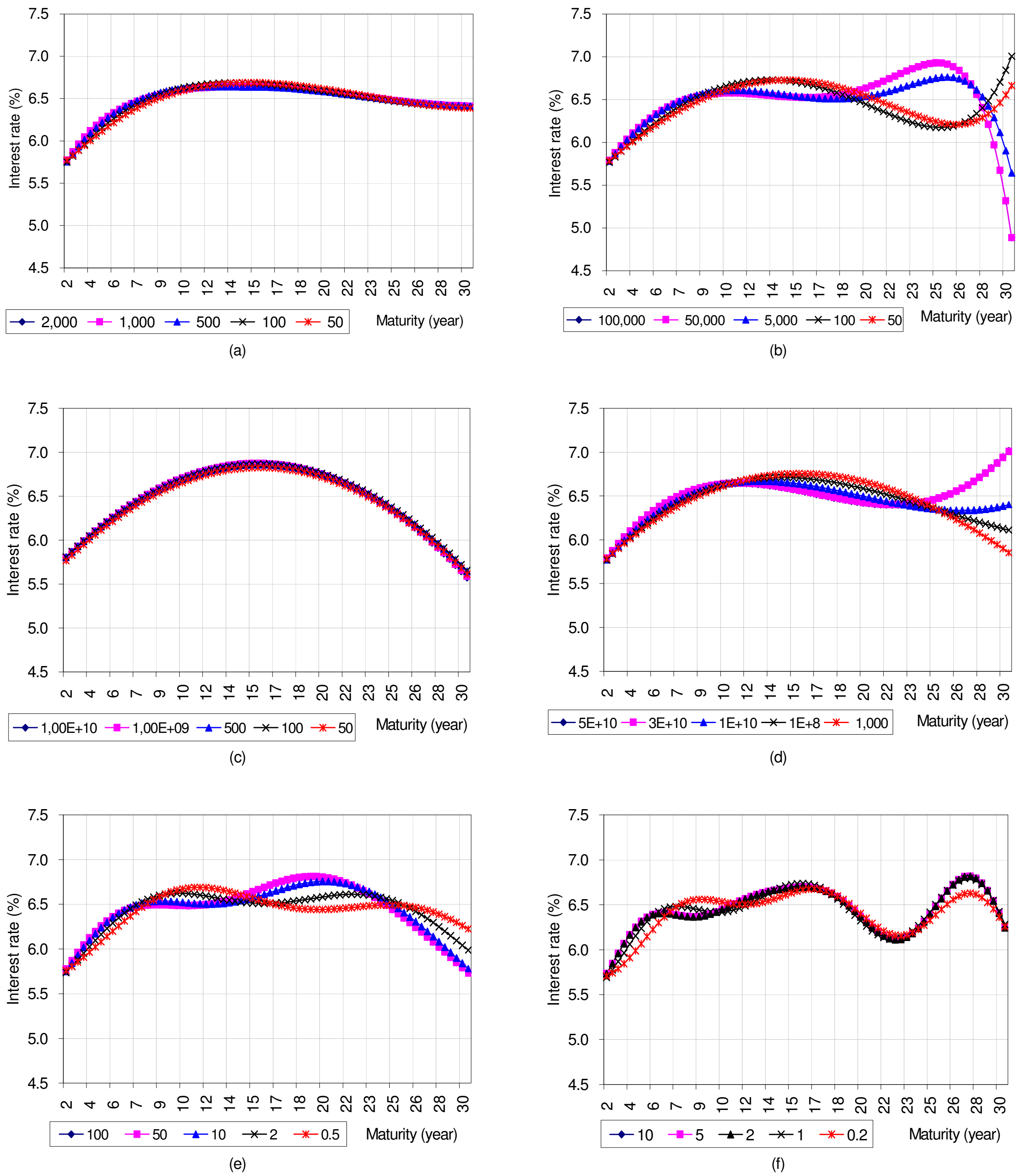
The proof is trivial. The search space for its optimal value is then reduced to]0, *C*s]. As *Cs* has been chosen as the value of *C*, the resulting model is the most flexible SVR that can be achieved, everything else being kept equal. Proposition 2 is not based on pure *a priori* information nor on a re-sampling strategy. A single estimation is enough to determine the value of *Cs*. In the next Subsection, this proposition will be crucial in determining feasible intervals for the parameter *C* according to each kernel.

Employing Proposition 2 and equation (28), we computed the values of C_S for SVRs equipped with the above eight kernels for each sample day. Then, the percentiles of C_S were calculated for each kernel. Based on these percentiles, we decided upon five values of *C* which are shown in Table 1. The 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). These ranges may be used as *a priori* information for feasible values of *C* per kernel.

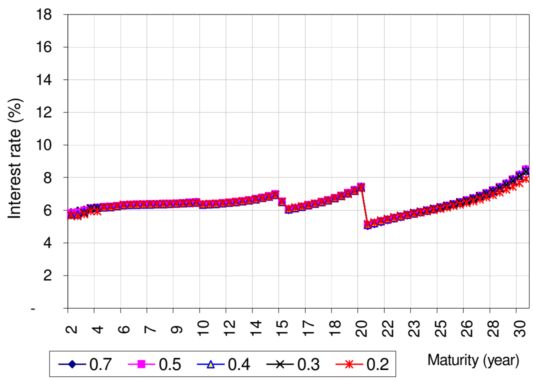
Kernel	*C1*	*C2*	*C3*	*C4*	*C5*
rbfabs(0.5)	0.6	0.4	0.3	0.2	0.1
rbfabs(1)	0.7	0.5	0.4	0.3	0.2
rbfgua(0.5)	10.0	5.0	2.0	1.0	0.2
rbfgua(1)	100.0	50.0	10.0	2.0	0.5
polin2	5.0E+10	3.0E+10	1.0E+10	1.0E+08	1000.0
polin3	1.0E+10	1.0E+09	500.0	100.0	50.0
polvok	100,000.0	50,000.0	5,000.0	100.0	50.0
splinf	2,000.0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

Table 1: Different values for *C* per kernel.

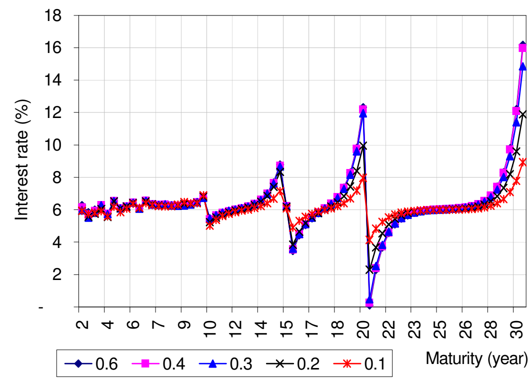
Some SVR specifications will model one typical curve selected from the sample. The values of *C* will be varied from *C1* to *C5*, totaling 40 SVR specifications. The SVR coefficients were estimated for the spot curves and we computed the respective forward curves. Figure 3 displays the five forward curves generated by each of the eight kernels.[8]



[8] Some of the kernels impose the normalization of *x*, so we arbitrarily selected a value to suit all of them: the maturities were divided by 40. Figure 3 exhibits non-normalized values for term to maturities.



(g)



(h)

Figure 3. 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) *rbfabs*(1); (h) *rbfabs*(0,5).

The shapes of the curves generated by the Vovk's real infinite (Figure 3.b) and the third degree polynomials (Figure 3.d) were highly sensitive to *C*. In order for these kernel functions to produce the asymptotic trend, a very careful selection of *C* is necessary. However, we cannot ensure that this value will work for all days. On the other hand, the curve shape produced by the spline with an infinite number of nodes was almost *C*-insensitive (Figure 3.a).

The kernel generating a spline with an infinite number of nodes is the one that produces a forward curve that fulfills the conditions placed upon it. Nevertheless, this kernel does not enforce asymptotic behavior for long-maturity rates: it only uses the information present in the data. The development of an admissible SVR kernel that imposes an asymptotic trend might be an interesting research project. Finally, these qualitative conclusions can apply for interest rate curves in general, not only for swap curves.

Despite the fact that we concluded for a most proper kernel, we will not rule out the others from the rest of the paper. The empirical results ahead will illustrate how relevant the conditions on the forward curve are for estimation performance.

5. Data

The data source was Bloomberg. It provided annualized bid and ask *spot rates* for the 12 standard maturities (in years): 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 and 30. These rates are the means of the rates at which banks, acting as market makers, were willing to make a deal, quoted at the end of each day. Therefore, each estimation process will be based on 12 observations.

The precision of the rates is fixed in accordance with the minimum quoting unit: 0.001%. The bid-ask spread and the mean rates are computed following equations (1) and (2), respectively. The sample begins on March 3, 1997, and finishes on April 30, 2001. Initially, the sample was composed of 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 is then composed of 1,024 days. Relevant international financial crises took place during the considered period, namely the Asian (1997), the Russian and Long Term Capital hedge fund (1998), the Brazilian (1999) and the burst of the NASDAQ bubble (2000). Despite these financial crises, the swap curves exhibited well-behaved shapes, which do not mean that the dataset is composed of near-identical data. However, their bid-ask spreads displayed a different pattern after the 1998 crisis.

Table 2 presents the bid-ask spread descriptive statistics over the sample. Two bid-ask spread values, 0.030% and 0.040%, showed high sample frequencies: 26.7% and 58.9%, respectively, whereas 95% of the whole sample's bid-ask spread values were between 0.030% and 0.050%.

Statistics	***bid-ask spread$_i$* (%)**	***bid-ask spread$_i$* / $T_{M,i}$**
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 2: Bid-ask spread descriptive statistics.

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

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 3: Bid-ask spread descriptive statistics per year (%).

On average, the three longest maturity contracts showed the greatest mean bid-ask spreads. However, when the annual mean bid-ask spread is divided by its annual mean rate, the difference vanishes. This result is in accordance with the empirical finding that the bid-ask spread value depends on the rate level. Volatility, on the other hand, is greater for the shortest maturity contracts. Malhotra (1998) remarked that this behavior is due to liquidity concentration on these contracts.

6. SVR Modeling

This section summarizes how SVR models the estimation of the swap curve. The task of SVR is to estimate a function for the spot curve, restricting the shape of the implicit forward curve and trying to keep the observed rates inside the symmetrical intervals around the function determined by the contracts' bid-ask spreads.

SVR will estimate the spot curve from the 12 standard contracts with fixed maturities. The dependent variable *y* is the mean spot rate as defined by (2) and the explanatory variable *x* is the maturity of the contracts preprocessed by a kernel function *k*.

$$y = f(x) = \sum_{i=1}^{12} (\alpha_i - \alpha_i^*).k(x_i, x) + b$$

The kernel and the parameter *C* jointly determine the implicit forward curve as discussed in the previous section. The loss function is the ε_i-insensitive loss function defined by (26). The values of ε_i will be fixed as half of the bid-ask spread of each contract according to (27).

7. Results

The result section is divided into two Subsections. The first analyzes the in-sample and cross-validation interpolation performances of different SVR specifications and the role played by the contracts in curve estimation. The second compares the best SVR with three other models by using cross-validation. All the computations in the present paper were performed in MatLab.

In the case of financial assets, the performance metric should take the holding period into consideration, that is, the period of time that investors, on average, keep the position. The point is that the holding period allows for the correct assessment of the economic meaning of different performances. The performance metric frequency theoretically should equal the holding period. For instance, if the holding period is a month, the performance metric should be based on monthly errors. According to Section 3.2, the majority of US dollar swap trades focus on middle and long terms: a reasonable holding period is one year. Because the available sample contains only four years, we base the performance metric on daily errors and consider the holding period in the analysis.

7.1 SVR Performance

Although the bid-ask spreads are rarely presented in model estimation, because of estimation cost, they are commonly used in the performance metric: if the estimated rate is within the bid-ask spread, its error is null. Thus, the mean absolute error will be adjusted to consider the bid-ask spread: the ε_i-adjusted mean absolute error (MAE-ε_i) will evaluate the interpolation performance. It does not consider the errors from two and 30-year contracts.

$$MAE-\varepsilon_i = \frac{1}{T(n-2)}\sum_{t=1}^{T}\sum_{i=2}^{n-1}\left|y_{t,i}-\hat{y}_{t,i}\right|_{\varepsilon_{t,i}} \quad , \tag{29}$$

where n=12, T=1024 and $\left|\,.\,\right|_{\varepsilon_i}$ is defined by (26).

Table 4 shows the error measures for each SVR. All SVR with smoothness parameter equal to *Cs*-SVR (the most flexible curves) achieved null *MAE*-ε_i. For all of them, the less flexible curves underperformed the more flexible curves.

Kernel	*splinf*	*polvok*	*polin2*	*polin3*	*rbfgua(1)*	*rbfgua(0.5)*	*rbfabs(1)*	*rbfabs(0.5)*
C	MAE-ε_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

Table 4: In-sample performance metric per kernel and *C*.

At first sight, the performances of *splinf*-SVR with *C4* and with *C5* are very close. Do they really matter in economic terms? The proper answer to this question involves the one-year holding period. A daily difference in the *MAE*-ε_i of 0.001% is indeed small. However, a persistent daily difference of 0.001% is relevant over one year. Suppose that an investor has a US$ 15 million notional position in the 20-year swap contract, which is the average ticket in this market. If he holds this position for one year, this daily difference in the rate will generate a loss of approximately 3% of the notional, which is economically significant in interest rate market positions.

Table 5 illustrates the average number of support vectors (total, on the boundary and outside the tube) required by each SVR specification across the sample. The *splinf*-SVR demanded the smallest number of support vectors in order to describe the curves: on average, between four and five. This means that 33% to 42% of the swap contracts were necessary to estimate the curves with the desired accuracy established by the bid-ask spreads. Nonetheless, this somewhat high percentage was expected since one needs at least three points to describe curvature (25%). The variability around the mean number of support vectors required by *splinf*-SVR specifications was low. The fraction of support vectors outside the tube is small on average, which suggests a small number of outliers considering the tolerance intervals fixed by the bid-ask spreads.

Only *polin2* and *polin3*-SVR showed a relatively high variability in the number of support vectors according to *C*. On the other hand, the radial-basis-function SVRs were the less *C*-sensitive models in terms of the number of required support vectors.

Kernel	*splinf*	*polvok*	*polin2*	*polin3*	*rdfgua(1)*	*rdfgua(0.5)*	*rdfabs(1)*	*rdfabs(0.5)*
C	Average number of support vectors							
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	Average number of support vectors on the boundary of the tube							
Cs	4.6	5.1	8.7	7.1	5.3	6.2	7.3	9.9
C1	4.4	4.6	8.6	6.9	5.0	6.0	7.4	9.9
C2	4.3	4.5	8.4	6.3	5.0	6.0	7.2	9.6
C3	4.3	4.3	7.6	3.7	4.7	5.8	7.2	9.1
C4	4.0	3.9	6.0	3.5	4.5	5.5	7.0	7.7
C5	3.9	3.9	3.6	3.4	3.9	4.6	6.6	5.7
C	Average number of support vectors outside the tube							
Cs	0	0	0	0	0	0	0	0
C1	0.2	0.3	0.1	0.2	0.2	0.1	0.1	0.0
C2	0.4	0.4	0.3	0.9	0.3	0.2	0.2	0.4
C3	0.6	0.7	1.2	1.0	0.7	0.5	0.4	0.8
C4	1.1	1.7	2.7	1.9	1.2	0.7	0.7	2.2
C5	1.7	2.3	2.4	2.4	2.1	1.7	1.2	4.2

Table 5: Average number of support vectors required per kernel and *C*.

The figures in Table 5 are not conclusive regarding the relationship between *C* and the number of support vectors. However, the analysis by type of support vector is revealing. For all the kernels in Table 5, the average number of support vectors on the boundary of the tube decreased with the increase of *C* whereas the number of support vectors outside the tube increased. This effect reflects the smoothness degree required from the estimated function. The lower the value of C, the more flexible the function is. More flexible functions tend to adjust better to in-sample data, which requires more support vectors on the boundary. Less flexible functions tend to identify more observations as outliers – which are support vectors outside the tube. The increasing rate of the number of support vectors outside the tube and the decreasing rate of the number of support vectors on the boundary with *C* is kernel-dependent.

The *splinf*-SVR achieved both the best in-sample performance and the lowest number of support vectors.

Figure 4 depicts the frequency in which the contracts were selected as support vectors by the *C3*-SVR (we chose *C3* because it represents an intermediate regularization). For example, *splinf*-SVR required the 30-year contract as a support vector in 98% of the days. One can identify a common U-shape for all SVR specifications: the percentages decrease up to the seven or eight-year contracts and begin to increase up to the longest maturity contract. The *rbfabs(0.5)* and *polin2* SVRs differed from the average shape. The most demanded contracts were those in the ends of the curves, reflecting the curve shapes in the sample: positively sloped and convex.

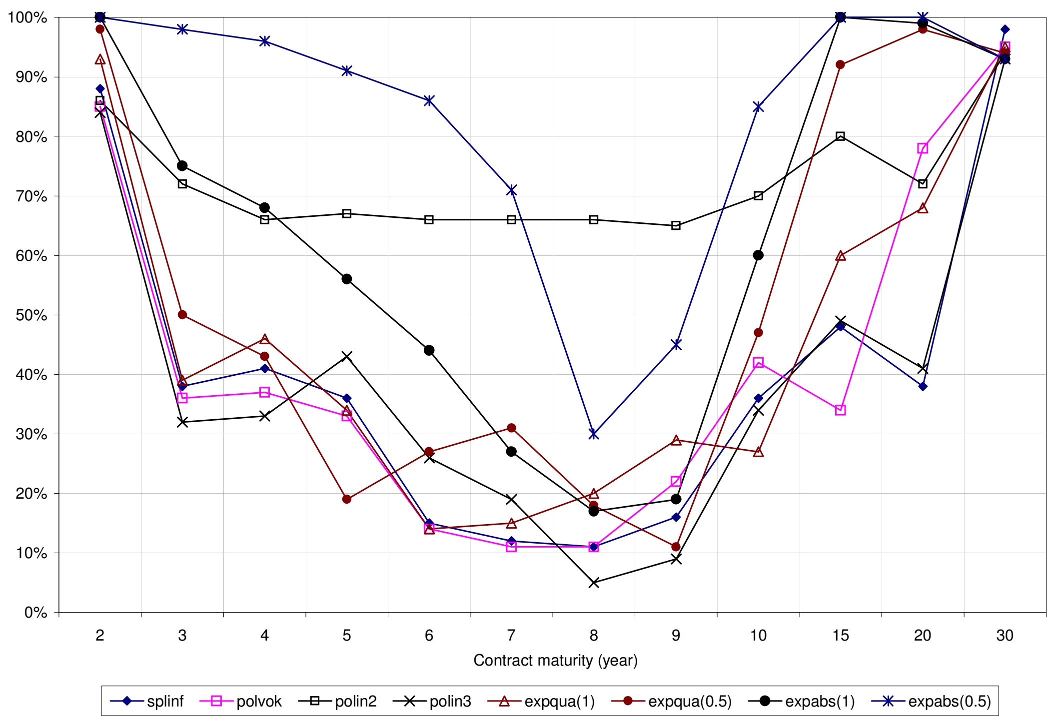


Figure 4: The sample frequencies in which contracts were chosen as support vectors.

Data reduction is an important tool commonly used to represent a curve in the cross-section and to describe its movements through time. Nelson and Siegel's (1987) model, for example, represents a curve through three parameters: level, slope and curvature.

The support vector set is a sufficient statistic for a curve. It means that the contracts selected as support vectors are able to completely represent the curve given the desired accuracy. This property is very attractive for a practitioner and a trader: only a small fraction of the negotiated contracts are enough to determine the others. In fact, the support vectors relate to the Nelson and Siegel's parameters. Level is the asymptotical rate when the maturity goes to infinity. This rate is close to rate of the 30-year contract – selected as a support vector by all SVR specifications in more than 94% of the days. Slope is the difference between level and the zero-maturity rate. The contract at the short-end of the curve is chosen in almost 90% of the days. The difference between the rates of these two support vectors also defines a slope for the curve, which is close to Nelson and Siegel's slope. The remaining support vectors determine the curvature.

Litterman and Scheinkman (1991) have applied principal component analysis to empirically show that two or three independent factors are needed to describe curve movements through time. The components are interpreted as level, slope and curvature.

At first glance, the support vectors could track the curve's movements through time. Initially, ν-SVR could be used to fix the number of support vectors. However, it is not possible to guarantee that the support vector set is the same for all the sample days. Since it does not make financial sense to monitor a time series composed by rates with different maturities, the support vectors are not able to perform this task.

The cross-validation analysis was performed using leave-one-out (Stone, 1974). Table 6 presents the values for MAE-ε_i. [9] Again, as expected, *splinf*-SVR achieved both the best results and the lowest sensitivity to the C. The performances of the SVRs with *rbfabs* were very poor compared to the others. The best *splinf*-SVR was obtained using $C1$. According to the previous analysis of the holding period, the 0.001% difference from the second best models (those with MAE-ε_i equal to 0.003%) is economically significant.

[9] Some SVRs required a long time to be concluded or simply reached unfeasible solutions. Thus, such days do not take part in the error statistics. Since this happened an insignificant number of times (0,2% to 3,5%, depending on

Kernel	*splinf*	*polvok*	*polin2*	*polin3*	*rbfgua(1)*	*rbfgua(0.5)*	*rbfabs(1)*	*rdfabs(0.5)*
C	MAE-ε_i (%)							
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

Table 6: Cross-validation performance metric per kernel and *C*.

7.2 SVR versus Other Models

The following models were selected in order to offer a large variety of estimation approaches applied to interest rate curves: four types of cubic splines, Nelson and Siegel's (1987) parametric regression model and Nadaraya-Watson's (1964) non-parametric regression model.

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

The parametric model provided by Nelson and Siegel (1987) is a popular model for interest rate curve estimation. Indeed, it was developed for this task. It describes a curve with three parameters: level (l), slope (s) and curvature (c).

$$y(x) = l + (-s + c)\left[1 - \exp\left(\frac{-x}{\tau}\right)\right]\frac{\tau}{x} - c\exp\left(\frac{-x}{\tau}\right) \tag{30}$$

The rate of exponential decay is controlled by τ. Fixing τ, ordinary least squares (OLS) can estimate the other three parameters. The four-parameter model requires a non-linear estimation. Nelson and Siegel (1987) and Barret, Gosnell and Heuson (1995) pointed out that the four-

the kernel), the results are still valid. The number of days that encountered these problems is the following, per kernel: *rbfgua*(1): 17; *rbfgua*(0,5): 36; *rbfabs*(1): 17; *rbfabs*(0,5): 2.

parameter model causes over-fitting problems for US Treasury curves. Here, the value of τ will be predetermined.

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

$$f(x)=\frac{\sum_{i=1}^{n}k(x_i,x)y_i}{\sum_{i=1}^{n}k(x_i,x)} \tag{31}$$

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

$$k(x_i,x_j)=\frac{1}{h\sqrt{2\pi}}\exp\left(-\frac{(\mathrm{x_i}-x_j)^2}{2\mathrm{h}^2}\right) \tag{32}$$

where h is the bandwidth parameter. Small values for h introduce noises in the estimated function *f*. The greater the parameter h, the smoother f is and the closer it is to the sample mean of y. As usual, $h=\lambda\sigma$, where the parameter λ is pre-established and σ is the standard deviation of x.

The selection of both Nadaraya-Watson's bandwidth parameter h and Nelson and Siegel's parameter τ will be based on cross-validation, as discussed below.

A consistent comparison of these four different models is not straightforward because they possess different cost functions. SVR seeks to minimize a cost function composed by the ε_i-insensitive loss function and a regularization term. The four splines are perfect-data-fitting models. Nelson and Siegel and Nadaray-Watson models minimize the mean square error. Therefore, to apply only the MAE-ε_i might generate unfair comparisons and lead to wrong conclusions. One way to keep the imbalances under control is to make use of different error measures selected in accordance with the model cost functions. Thus, the root of mean square error (*RMSE*) will also be calculated in addition to MAE-ε_i. Again, the cross-validation errors will be generated by leave-one-out:

$$RMSE = \sqrt{\frac{1}{T(n-2)} \sum_{t=1}^{T} \sum_{i=2}^{n-1} \left(y_{t,i} - \hat{y}_{t,i} \right)^2} \,, \tag{33}$$

where, n=12 and T=1024.

Tables 7, 8 and 9 illustrate the performance metrics for the four splines, Nelson and Siegel and Nadaraya-Watson, respectively. Six different values were tested for parameters τ and λ.

Statistics / Splines	not-a-knot	periodic	complete	second
***MAE*-ε_i**	0.006	0.015	0.005	0.003
RMSE	0.039	0.056	0.034	0.021

Table 7: Cross-validation performance metrics of cubic splines.

Statistics / τ	0.06	0.08	0.10	0.12	0.14	0.50
***MAE*-ε_i**	0.004	0.003	0.003	0.003	0.003	0.005
RMSE	0.021	0.020	0.020	0.020	0.020	0.024

Table 8: Cross-validation performance metrics of Nelson and Siegel.

Statistics / λ	1.00	0.50	0.30	0.20	0.10	0.05
***MAE*-ε_i**	0.085	0.054	0.029	0.014	0.008	0.008
RMSE	0.134	0.091	0.057	0.039	0.029	0.029

Table 9: Cross-validation performance metrics of Nadaraya-Watson.

The second cubic spline achieved the lowest error according to the two performance metrics. As the OLS seeks to minimize the square error, *RMSE* is the natural criterion to select the best specification of Nelson and Siegel. Four specifications of the Nelson and Siegel model (τ = 0.08, 0.10, 0.12, 0.14) achieved the same performance measured not only by *RMSE* but also by *MAE*-ε_i. Following the same criterion, the best specification for Nadaraya-Watson was the one with λ = 0.10 or λ = 0.05. For the sake of visual comparison, the best specification of each model has its performance measures consolidated in Table 10.

Statistics / Models	SVR (splinf, *C1*)	Second cubic spline	Nadaraya-Watson (λ = 0.05)	Nelson and Siegel (τ = 0.08)
MAE-ε_i	0.002	0.003	0.008	0.003
RMSE	0.019	0.021	0.029	0.020

Table 10: Cross-validation performance metrics of each model's best specifications.

The results in Table 10 demonstrate how important is the use of *a priori* information for modeling the USD swap interest rate curves. The Nadaraya-Watson regression, the model which does not count on this kind of information, performed poorly in comparison with the others.

The *splinf*-SVR was the best model when the performance metric was the *MAE*-ε_i: 0.002%. The second cubic spline and Nelson and Siegel achieved close values for their *MAE*-ε_i: 0.003%. Since the observed rates fluctuated between 4% and 8% within the sample, these three models achieved a very good performance. However, it is important to point out, as it was shown in the last Subsection, that the 0.001% difference in the daily *MAE*-ε_i is economically relevant when the one-year holding period is taken into consideration. The *splinf*-SVR also achieved the best performance according to *RMSE*, followed by the Nelson and Siegel and the second cubic spline.

Therefore, SVR slightly outperformed the other models for simultaneously considering the bid-ask spreads as desired accuracy for the cross-validation error and controlling the bias-variance trade-off.

8. Conclusion

The present paper has shown that Support Vector Regression (SVR) is a suitable model for approaching the estimation of interest rate curves. It is simultaneously able to introduce the bid-ask spreads of the securities from which the curve is extracted and to restrict the curve shape into the estimation. These pieces of *a priori* information guided the selection of the SVR parameter selection: loss function, a regularization (or smoothing) parameter (*C*) and a kernel function. They are introduced into the estimation at a very low optimization cost: the SVR estimation is a plain-vanilla convex quadratic programming problem.

We explored the linkages among SVR and many other traditional models, some of them applied to financial econometrics, such as radial basis functions, neural networks, the Hodrick-Prescott filter and smoothing splines. Once we work only with Gaussian distributions, it is possible to interpret SVR under a probabilistic framework.

The loss function that best suits the curve estimation problem, allowing for the bid-ask spread in the estimation, is exactly the one that leads to the simplest optimization problem: the ε-insensitive function proposed by Vapnik (1995). In order to allow for different bid-ask spreads in the same estimation, we proposed the ε_i-insensitive function: the parameter ε was indexed to each sample observation. We proved that the ε_i-insensitive SVR has the same properties as an equivalent ε-insensitive SVR.

Because of the relationship between spot and forward rates, when SVR is estimating a spot curve, the kernel and the parameter *C* jointly determine the implicit forward curve. It was possible to partially verify *a priori* how each kernel function handles the requirements on forward curves. Because the parameter *C* impacts the kernel shape and this impact is data-dependent, the theoretical analysis is not conclusive. Thus, we needed to run an empirical sensibility analysis of the kernel to the value of *C*.

We considered eight different kernel functions and the conclusions about how their properness to equip the SVR are valid for interest rate curves in general, not only for swap curves. The kernel generating a spline with an infinite number of nodes was the one that produced a forward curve that best fulfilled the conditions placed upon it. This kernel was almost *C*-insensitive and it does not enforce asymptotic behavior for long-maturity rates: it only uses the information present in the data. The development of an admissible SVR kernel that imposes an asymptotic trend might be an interesting research project.

We modeled the spot curve from US dollar-Libor swaps. These securities were selected because of the important role they play in financial markets and for the rich set of information contained in their bid-ask spreads. Empirical research shows that the maturity and liquidity of the contract, bank credit risk, the level of interest rates and 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. As expected from the previous analysis, the SVR equipped with spline with an infinite number of nodes was the best in-sample and cross-validation performing SVR according to the ε_i-adjusted mean absolute error (MAE-ε_i). This kernel described the curves with the smallest number of support vectors considering the accuracy required by bid-ask spreads: on average, four to five contracts. The shortest and longest maturity contracts were the most frequently demanded.

The best SVR was compared to four kinds of cubic splines, Nadaraya-Watson's non-parametric model and Nelson and Siegel's parametric model. To accommodate the different purposes of these models, another performance metric was considered: the root of mean square error (*RMSE*). SVR reached the best control of the bias-variance trade-off (the lowest *RMSE*) and the lowest bias, when we consider accuracy as defined by the bid-ask spreads (the lowest *MAE*-ε_i). These results suggest that the bid-ask spreads and the restriction on the forward curve were useful *a priori* information manipulated by SVR in order to estimate interest rate curves.

Acknowledgements

The author is indebted to Carlos Pedreira and Paulo Blikstein for valuable discussions. He thanks the comments of the participants in the Machine Learning in Finance Workshop at 2005 Neural Information Processing Systems (NIPS), Vancouver, Canada.

Appendix A

The proof is presented for the linear SVR. The proof for the non-linear case is analogous. Consider the primal formulation of SVR equipped with ε_i-insensitive loss function:

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)$$

$$\text{subject to} \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon_i + \xi_i \text{ , } i = 1,..., n \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon_i + \xi_i^* \text{ , } i = 1,..., n \\ \xi_i, \xi_i^* \geq 0 \text{ , } i = 1,..., n \end{cases} \tag{A.1}$$

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

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

The partial derivatives of L with respect to primal variables (w, b, ξ_i, ξ_i^*) must be zero. Thus, we have four equations.

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\left(\alpha_i^* - \alpha_i\right) = 0 \tag{A.3}$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i \tag{A.4}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \eta_i = C - \alpha_i \text{, i = 1,..., n} \tag{A.5}$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \Rightarrow \eta_i^* = C - \alpha_i^* \text{, i = 1,..., n} \tag{A.6}$$

The introduction of (A.5) and (A.6) in the last right-side term of equation (A.2) vanishes with the multipliers η_i, η_i^*, $i = 1, ..., n$. Then, this term becomes:

$$-C\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right)+\sum_{i=1}^{n}\xi_i\left(\alpha_i+\alpha\right)_i^* \tag{A.7}$$

By returning the above modified term to (A.2), the deviations ξ_i and ξ_i^* disappear.

$$L\equiv\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i\left(\varepsilon_i-y_i+\langle w,x_i\rangle+b\right)-\sum_{i=1}^{n}\alpha_i^*\left(\varepsilon_i+y_i-\langle w,x_i\rangle-b\right) \tag{A.8}$$

The above equation is re-arranged into:

$$L\equiv\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\varepsilon_i\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{n}y_i\left(\alpha_i-\alpha_i^*\right)-\sum_{i=1}^{n}\langle w,x_i\rangle\left(\alpha_i-\alpha_i^*\right)+b\sum_{i=1}^{n}\left(\alpha_i^*-\alpha_i\right) \tag{A.9}$$

The last term of the right side vanishes because of (A.3). Vector w is replaced by (A.4) and, replacing $\|w\|^2$ by $\langle w,w\rangle$, (A.9) becomes (A.10).

$$L\equiv\frac{1}{2}\left\langle\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)x_i,\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)x_i\right\rangle-\sum_{i=1}^{n}\varepsilon_i\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{n}y_i\left(\alpha_i-\alpha_i^*\right)-\sum_{i=1}^{n}\left\langle\sum_{j=1}^{n}\left(\alpha_j-\alpha_j^*\right)x_j,x_i\right\rangle\left(\alpha_i-\alpha_i^*\right) \tag{A.10}$$

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

$$\left\langle\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)x_i,\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)x_i\right\rangle=\sum_{i=1}^{n}\left\langle\sum_{j=1}^{n}\left(\alpha_j-\alpha_j^*\right)x_j,x_i\right\rangle\left(\alpha_i-\alpha_i^*\right)=\sum_{i,j=1}^{n}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\langle x_i,x_j\rangle \tag{A.11}$$

Replacing the first term of (A.10) by (A.11), the Lagrangean becomes:

$$L\equiv-\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\langle x_i,x_j\rangle-\sum_{i=1}^{n}\varepsilon_i\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{n}y_i\left(\alpha_i-\alpha_i^*\right) \tag{A.11}$$

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.

$$0 \leq \alpha_i \leq C,\ i = 1,\ldots, n$$
$$0 \leq \alpha_i^* \leq C,\ i = 1,\ldots, n$$

Now, the dual formulation can be expressed by (A.12).

$$\begin{aligned}
&\text{Maximize } -\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\left\langle x_i, x_j\right\rangle - \sum_{i=1}^{n}\varepsilon_i\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{n} y_i\left(\alpha_i - \alpha_i^*\right) \\
&\text{subjected to} \begin{cases} \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) = 0 \\ \alpha_i, \alpha_i^* \in [0, \mathrm{C}] \end{cases}
\end{aligned} \tag{A.12}$$

The primal variables (ξ, ξ^* and w) and the dual variables (η and η^*) do not appear in (A.12). (A.12) is a quadratic programming problem with linear restrictions. Thus, it presents one global solution. The SVR expansion of function f is written as in the original formulation:

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

The following Karush-Kuhn-Tucker conditions came from (A.12):

$$\alpha_i\left(\varepsilon_i + \xi_i - y_i + \left\langle w, x_i\right\rangle + b\right) = 0 \tag{A.13}$$
$$\alpha_i^*\left(\varepsilon_i + \xi_i^* + y_i - \left\langle w, x_i\right\rangle - b\right) = 0 \tag{A.14}$$
$$\left(C - \alpha_i\right)\xi_i = 0 \tag{A.15}$$
$$\left(C - \alpha_i^*\right)\xi_i^* = 0 \tag{A.16}$$

The equations (A.15) and (A.16) assure that the sparseness property of the ε-insensitive SVR is preserved in the ε_i-insensitive SVR. The equations (A.13) and (A.14) are changed only by the indexing of ε. Hence, the computation of the parameter *b* is unchanged from the traditional formulation.■

References

Abu-Mostafa, Y. S. (2001). Financial model calibration using consistency hints. *IEEE Trans. on Neural Networks*, 12(4):791–808.

Aizerman, M.A., Braverman, É.M., Rozonoér, L.I. (1964). Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. *Automation and Remote Control*, 25, 821-37.

Barret, B., Gosnell, T., 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*, 9, 197–231.

Bishop, C.M. (1995). Neural Networks for Pattern Recognition. Oxford: Claredon Press.

de Boor, C. (1978). *A practical guide to splines*. New York, NY: Springer.

Boser, B.E., Guyon, I.M., Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. In Haussler, D. (Ed), Proceedings of the Annual Conference on Computational Learning Theory, 144-152. Pittsburg, PA: ACM.

Brooks, R., 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., Crisp, D. (1999). Uniqueness of the SVM Solution. In *Annals of Neural Information Processing Systems*.

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

Cao, L. J., Tay, F.E.H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 14(6), 1506-1518.

Chapados, N., Bengio, Y., Vicent, P., Ghosn, J., Dugas, C., Takeuchi, I., Meng, L. (2002). Estimating Car Insurance Premia: a Case Study in High-Dimensional Data Inference. In *Annals of Neural Information Processing Systems Conference*, 14, 1369-1376. Cambridge, MA: MIT Press.

Chambers, D. R., Carleton, W. T., Waldman, D. W. (1984). A New Approach to the Estimation of the Term Structure of Interest Rates. *Journal of Financial and Quantitative Analysis*, 19(3), 233-252.

Cohen, K. J., Kramer, R. L., Waugh, W. H. (1966). Regression Yield Curves for US Government Securities. *Management Science*, 13(4), 168-175.

Dermody, J., Prisman, E. (1988). Term Structure Multiplicity and Clientele in Markets with Transactions Costs and Taxes. *The Journal of Finance*, 41(4), 893-911.

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

Durand, D. (1942). Basic Yields of Corporate Bonds. Technical Paper nº 3, National Bureau of Economic Research.

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

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

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

Gao, J.B, Gunn, S.R., Harris, C.J. (2002). A Probabilistic Framework for SVM Regression and Error Bar Estimation. *Machine Learning*, 46, 71–89.

Gestel, T. V., Suykens, J. A. K., Baestaens, D., Lambrechts, A., Lanckriet, G., Vandaele, B., De Moor, B., Vandewalle, J. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural*

Networks, Special Issue on Neural Networks in Financial Engineering, 12(4), 809-821.

Hodrick, R.J., Prescott, E.C. (1980). Postwar U.S. Business Cycles: An Empirical Investigation. Carnegie-Mellon University, Discussion Paper Nº 451.

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

Kim, K.J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307-319.

Linton, O., Mammen, E., Nielsen, J., Tanggaard, C. (2000). Estimating the Yield Curve by Kernel Smoothing Methods. *Journal of Econometrics*, 105, 185-223.

Litterman, R. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *Journal of Fixed Income*, 1:51-61.

Livingston, M. B., Jain S.K. (1982). Flattening of Bond Yield Curves for Long Maturities. *Journal of Finance,* 37, 157-167.

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(1), 79-88.

McCulloch, J. H. (1971). Measuring the Term Structure of Interest Rates. *Journal of Business*, 34(1), 19-31.

McCulloch, J. H. (1975). The Tax-Adjusted Yield Curve. *The 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.

Mercer, J. (1909). Foundations of positive and negative type and their connection with the theory of integral equations. Philos. Trans. Roy. Soc. London, Serie A, 209, 415-446.

Morozov, V.A. (1984). *Methods for Solving Incorrectly Posed Problems*. Springer.

Müller, K. R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., Vapnik, V. (1999). Predicting Time Series with Support Vector Machines. In B. Schölkopf, C. J. C. Burges and A. J. Smola (Eds), *Advances in Kernel Methods - Support Vector Learning*: 243-254.

Cambridge, MA: MIT Press.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9, 141-142.

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

Nilsson, N.J. (1965). Learning Machine: *Foundations of Trainable Pattern Classifying Systems*. McGraw-Hill.

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

Perez-Cruz, F., Afonso-Rodriguez, J.A., Giner, J. (2003). Estimating GARCH models using support vector machines. *Quantitative Finance*, 1, 1-10.

Rasmussen, C.E., Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, MIT Press.

Schölkopf, B., Bartlett, P., Smola, A. and Williamson, R. (1998). Support Vector Regression with Automatic Accuracy Control. In L. Niklasson, M. Bodén, and T. Ziemke (Eds), *Proceedings of ICANN'98, Perspectives in Neural Computing*, 111-116, Springer-Verlag.

Schölkopf, B., Burges, C.J.C., Smola, A. J., (1999). *Advances in Kernel Methods – Support Vector Learning*. Cambridge, MA: MIT Press.

Smola, A. (1998). Learning with kernels. Ph.D. Thesis. Technischen Universit at Berlin, Berlin, Germany.

Smola, A. J., Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*, 14(3), 199-222.

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

Smola, A. J., Schölkopf, B., Müller, K. R. (1998b). General cost functions for support vector regression. In T. Downs, M. Frean, and M. Gallagher (Eds), *Proceedings of the Ninth Australian Conference on Neural Networks,* 79-83.

Sollich, P. (2000). Probabilistic methods for Support Vector Machines. In Solla, A., Leen, T.K., Müller, K. (Eds), *Advances in Neural Information Processing Systems*, 12, 349-55. MIT Press: Cambridge, MA.

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

Sun, T., Sundaresan, S., 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*, 10, 29-48.

Tikhonov, A.N., Arsenin, V.Y. (1977). *Solution of Ill-posed Problems*. V.H. Winston and Sons.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*. Berlin: Springer.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York, NY: Springer.

Waggoner, D. (1997). Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices. Working paper, 97-10, Fed of Atlanta.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhaya: The Indian Journal of Statistics*, A(26), 359-372.

Yang, H., Chan, L., King, I. (2002). Support Vector Machine Regression for Volatile Stock Market Prediction. In H. Yin, N. Allison, R. Freeman, J. Keane and S. Hubbard (Eds), *Intelligent Data Engineering and Automated Leraning – IDEAL 2002*, 2412 of LNCS, 391-396. Springer.

Yang, H., Kim, I., Chan, L., Huang, K. (2004). Financial Time Series Prediction Using Non-fixed and Asymmetric Margin Setting with Momentum in Support Vector Regression. In J.C. Rajapakse and L. Wang (Eds), *Neural Information Processing: Research and Development*, 334-350. Springer.