

WAGES FORECASTING USING SV REGRESSION APPROACH

Dušan Marček^{1,2}

¹Institute of Computer Science, Faculty of Philosophy and Science, The Silesian University Opava

²The Faculty of Management Science and Informatics, University of Žilina

Abstract

SVM's modelling approaches are used for automated specification of a functional form of the model. We provide the fit of the average nominal wages time series by SVM (Support Vector Machine) model over the period January 1,1991 to December 31, 2006 in the Slovak Republic, and use them as a tool to compare their forecasting abilities with those obtained using Box- Jenkins methodology [1]. Some methodological contributions are made to dynamic and SVM's modelling approaches in economics and to their use in time series modelling.

Keywords: Support vector machines, learning machines, time series analysis and forecasting,

Introduction

The wages time series are in fact stochastic in which successive observations are dependent and can be represented by a linear combination of independent random variables $\varepsilon_t, \varepsilon_{t-1}, \dots$. If the successive observations are highly dependent, we should use in model past values of the time series variable and (or) current and past values of the error terms $\{\varepsilon_t\}$. There are available techniques which are designed to exploit this dependency and which will generally produce superior forecasts. Many of these techniques are based on developments in time series analysis recently presented by Box and Jenkins [1].

In Support Vector Machines (SVM's), a non-linear model is estimated based on solving a Quadratic Programming (QP) problem. In the next section of this article, we briefly describe the framework of SVM's methods and support vector (SV) regressions within which our empirical investigation is conducted. Section 3 provides a fit of the SV regression model using the Matlab program, discusses the circumstances under which SV regression outputs are conditioned and corresponding interpretation of SV regression results is also considered. A section of conclusions will close the paper.

1 Support Vector Machine for Functional Approximation

This section presents quickly a relatively new type of learning machine – the SVM's applied in the regression (functional approximation) problems. For details we refer to [2, 4, 6, 7, 9, 10, 11]. The general regression learning task is set as follows. The learning machine is given n training data, from which it attempts to learn the input-output relationship $y = f(x)$, where $\{x_i, y_i \in \mathfrak{R}^n \times \mathfrak{R}, i = 1, 2, \dots, n\}$ consists of n pairs $\{y_i, x_i\}_{i=1}^n$. The x_i denotes the i th input and y_i is the i th output. The SVM's considers regression functions of two forms [4, 7]. The first one is

$$f(\mathbf{x}) = \sum_{\substack{i=1 \\ j=1}}^n (\alpha_i - \alpha_i^*) \psi(\mathbf{x}_i, \mathbf{x}_j) + b \quad (1)$$

where α_i, α_i^* are positive real constants (Lagrange multipliers), b is a real constant (see formulas (3)-(5)), $\psi(\cdot, \cdot)$ is the kernel function. Admissible kernels have the following forms: $\psi(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$

(linear SVM's) $\psi(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^d$ (polynomial SVM's of degree d),
 $\psi(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\theta \|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$ (radial basis SVM's), where θ is a positive real constant and other
(spline, b-spline, exponential RBF, etc.).

The second approximation function is of the form [5]

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^n w_i \varphi_i(\mathbf{x}) + b \quad (2)$$

where $\varphi(\cdot)$ is a non-linear function (kernel) which maps the input space into a high dimensional feature space. In contrast to Eq. (1), the regression function $f(\mathbf{x}, \mathbf{w})$ is explicitly written as a function of the weights \mathbf{w} that are subject of learning.

The SV regression approach is based on defining a loss function that ignores errors that are within a certain distance of the true value. This type of function is referred to as an ε -insensitive loss function (see Fig. 1 and Fig. 2).

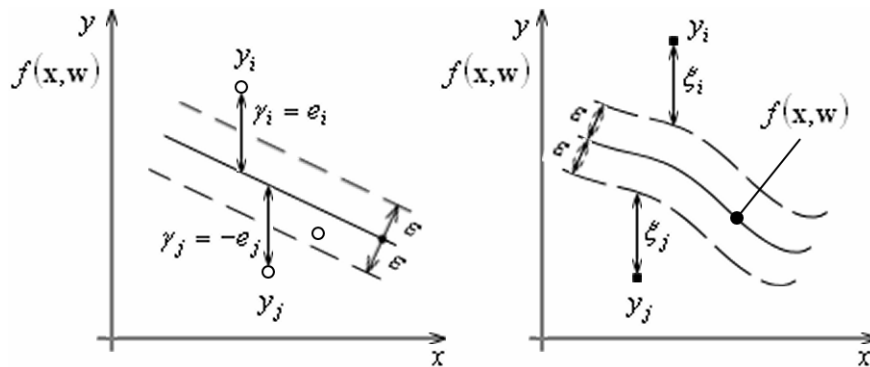


Fig. 1: The insensitive band for one dimensional linear (left), non-linear (right) function

Fig. 1 shows an example of an one dimensional function with an ε -insensitive band. The variables ζ, ζ^* measure the cost of the errors on the training points. These are zero for all points inside the ε -insensitive band, and only the points outside the ε -tube are penalised by the so called Vapnik's ε -insensitive loss function.

In regression, there are different error (loss) functions in use and that each one results from a different final model. Fig. 2 shows the typical shapes of three loss functions [2, 4]. Left: quadratic 2-norm. Middle: absolute error 1-norm. Right: Vapnik's ε -insensitive loss function.

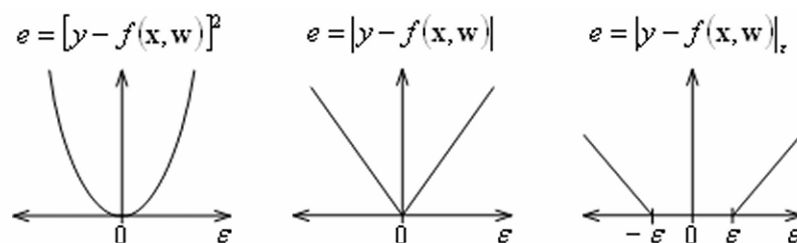


Fig. 2: Error (loss) functions

Formally this results from solving the following Quadratic Programming problem

$$\min_{\mathbf{w}, b, \xi, \xi^*} R(\mathbf{w}, \xi, \xi^*) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

where C is the value of capacity degrees

$$\text{subject to } \begin{cases} y_i - \mathbf{w}^T \varphi(\mathbf{x}) - b \leq \varepsilon + \xi_i & i = 1, 2, \dots, n, \\ \mathbf{w}^T \varphi(\mathbf{x}) + b - y_i \leq \varepsilon + \xi_i^* & i = 1, 2, \dots, n, \\ \xi_i, \xi_i^* \geq 0 & i = 1, 2, \dots, n. \end{cases} \quad (4)$$

To solve (3), (4) one constructs the Lagrangian

$$\begin{aligned} L_p(\mathbf{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) \\ = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \varphi(\mathbf{x}_i) + b) \\ - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b) - \sum_{i=1}^n (\beta_i \xi_i + \beta_i^* \xi_i^*) \end{aligned} \quad (5)$$

by introducing Lagrange multipliers $\alpha_i, \alpha_i^* \geq 0, \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n$. The solution is given by the saddle point of the Lagrangian [3]

$$\max_{\alpha_i, \alpha_i^*, \beta_i, \beta_i^*} \min_{\mathbf{w}, b, \xi_i, \xi_i^*} L_p(\mathbf{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) \quad (6)$$

subject to constrains

$$\begin{cases} \frac{\partial L_p}{\partial \mathbf{w}} = 0 & \rightarrow \mathbf{w} = \sum_{i=1}^n (\alpha_i^* - \alpha_i) \varphi(\mathbf{x}_i), \\ \frac{\partial L_p}{\partial b} = 0 & \rightarrow \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \\ \frac{\partial L_p}{\partial \xi_i} = 0, \frac{\partial L_p}{\partial \beta_i} = 0 & \rightarrow 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n, \\ \frac{\partial L_p}{\partial \xi_i^*} = 0, \frac{\partial L_p}{\partial \beta_i^*} = 0 & \rightarrow 0 \leq \alpha_i^* \leq C, \quad i = 1, \dots, n, \end{cases} \quad (7)$$

which leads to the solution of the QP problem:

$$\max_{\alpha_i, \alpha_i^*} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \varphi(\mathbf{x}_i^T \mathbf{x}_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad (8)$$

subject to (7).

After computing Lagrange multipliers α_i and α_i^* , one obtains of the form of (1), i.e.

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \psi(\mathbf{x}_i, \mathbf{x}_j) + b. \quad (9)$$

By substituting the first equality constraint of (7) in (2), one obtains the regression hyperplane as

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (10)$$

Finally, b is computed by exploiting the Karush-Kuhn-Trucker (KKT) conditions [3]. These conditions state that at the optimal solution the product between the dual variables and constraints has to vanish. For the SVM's, we obtain from, $\beta_i, \xi_i = 0$ and $\beta_i^*, \xi_i^* = 0$, $i = 1, 2, \dots, n$ that $(C - \alpha_i)\xi_i = 0$ and $(C - \alpha_i^*)\xi_i^* = 0$, respectively. Hence, $\xi_i = 0$ ($\xi_i^* = 0$) if and only if $C = \alpha_i$ ($C = \alpha_i^*$). From this and KKT conditions, we obtain

$$\begin{cases} b = y_k - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \psi(\mathbf{x}_i, \mathbf{x}_k) - \varepsilon & \text{for } \alpha_k \in (0, C), \\ b = y_k - \sum_{i=1}^n (\alpha_i - \alpha_i^*) \psi(\mathbf{x}_i, \mathbf{x}_k) + \varepsilon & \text{for } \alpha_k^* \in (0, C). \end{cases} \quad (11)$$

When solving the QP-problem with an interior point method [7, 8], it is more convenient to exploit primal-dual properties and to obtain b as a by-product of the optimization algorithm. Also note that for all samples inside the ε -band, α_i and α_i^* are zero. Therefore, we have a *sparse* expansion (1) of \mathbf{w} in terms of \mathbf{x}_i . The samples that come with non-vanishing coefficients are called Support Vectors.

3 Causal Models, Experimenting with Non-linear SV Regression

In previous section, we quickly introduced some of the main ideas of curve fitting based on solution quadratic problem. In this section we will discuss the problem of selecting the appropriate functional form of the SV regression model. The average nominal wages W_t can be described by the following regression equation

$$W_t = b + a W_{t-4} + \varepsilon_t \quad (12)$$

where a, b are the parameters, ε_t is the disturbance term. We demonstrate here the use of the SV regression framework for estimating the model given by Eq. (12). If W_t exhibits a curvilinear trend, one important approach for generating an appropriate functional non-linear form of the model is to use the SV regression in which the W_t is regressed either against W_{t-4} or the time by the form

$$\hat{W}_t = \sum_{i=1}^n w_i \varphi_i(\mathbf{x}_t) + b \quad (13)$$

where $\mathbf{x}_t = (W_{t-1}, W_{t-4}, \dots)$ is the vector of time sequence (regressor variable), or

$$\hat{W}_t = \sum_{i=1}^n w_i \varphi_i(\mathbf{x}_t) + b \quad (14)$$

where $\mathbf{x}_t = (1, 2, \dots, 60)$ is the vector of time sequence (regressor variable). Our next step is the evaluation of the goodness of last three regression equations to the data insite the estimation period expressed by the Mean Square Error MSE_A , and the forecast summary statistics the MSE_E for each of the models out of the estimation period.

One crucial design choice in constructing an SV machine is to decide on a kernel. The choosing of good kernels often requires lateral thinking: many measures of similarity between inputs have been developed in different contexts, and understanding which of them can provide good kernels depends on the insight into the application's domains. Tab 1 and corresponding Fig. 3 shows SVM's learning of the historical period illustrating the actual and the fitted values by using various kernels.

Tab. 1 presents the results for finding the proper model by using the quantities MSE. As shown in Tab. 1, the model that generates the “best” forecasts is the model with $MSE_E = 0.0012$ (Fig. 3a).

Tab. 1 SV regression results of three different choice of the kernels and the results of the dynamic model on the training set (1993Q1 to 2003Q4). In two last columns the fit to the data and forecasting performance respectively are analysed.

Fig.	MODEL	KERNEL	σ	DEGREE- d	LOSS FUNCTION	MSE_A	MSE_E
3a	causal (13)	RBF	1.6		ε - insensitive	0.0016	0.0012
3b	causal (13)	RBF	0.52		ε - insensitive	0.0012	0.0318
3c	causal (13)	Exp. RBF	1.0		ε - insensitive	0.000	0.0038
3d	Time S. (14)	RBF	1.0		ε - insensitive	0.000	none
	causal (12)					0.0026	0.0026

The results shown in Tab.1 were obtained using degrees of capacity $C = 10^4$. The insensitivity zone ε and the degrees of capacity are most relevant coefficients. An increase in ε means a reduction in requirements for the accuracy of the fit to the data. To learn the SV regression machine we used partly modified software developed by Steve R. Gunn [4].

The use of an SV machine is a powerful tool to the solution many economic problems. It can provide extremely accurate approximating functions for time series models, the solution to the problem is global and unique.

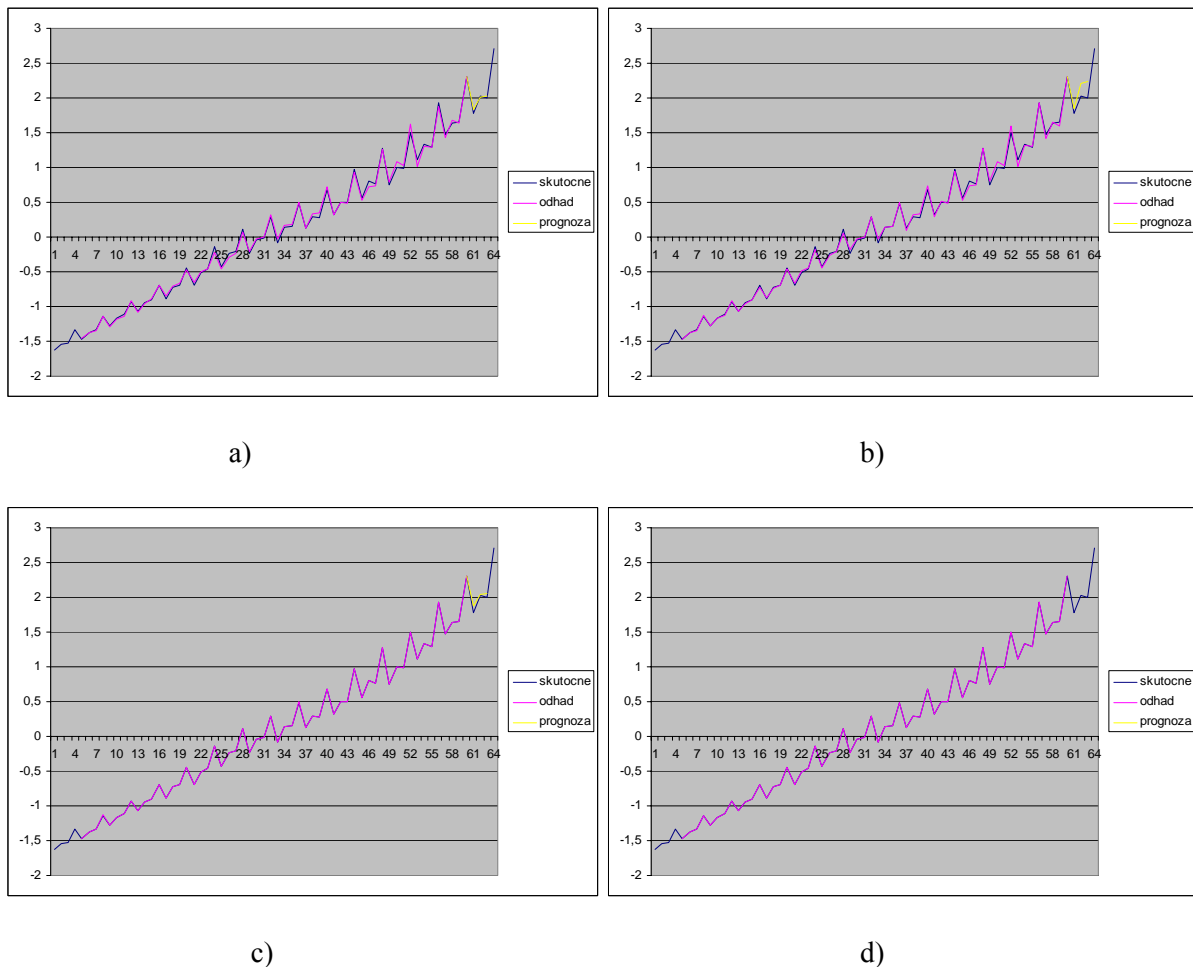


Fig. 3: Training results for different kernels, loss functions and σ of the SV regression (see Tab. 1).

4 Conclusion

In this paper, we have examined the SVM's approach to study linear and non-linear models on the time series of wages in the Slovak Republic. For the sake of calculating the measure of the goodness of fit of the regression model to the data we evaluated five models. One model was based on causal regression and for models on the Support Vector Machines methodology. The benchmarking was performed between traditional statistical techniques and SVM's method in regression tasks. The SVM's approach was illustrated on the conventional regression function. As it visually is clear from Fig. 3, this problem was readily solved by a SV regression with excellent fit of the SV regression models to the data.

Acknowledgement: This work was supported by Slovak grant foundation under the grant No. 1/2628/05 and from the Grant Agency of the Czech Republic under the grant No. 402/05/2768. I thank B. Kočvara for computational support.

References:

- [1] BOX, G.,E., JENKINS, G., M.: *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, CA 1976.
- [2] CRISTIANINI, N., SHAWE-TAYLOR, J.: *An introduction to support vector machines*. Cambridge University Press, 2000
- [3] FLETCHER, R.: *Practical methods of optimization*. John Wiley and Sons, Chichester and New York, 1987
- [4] GUNN, S., R.: *Support Vector Machines for Classification and Regression*. Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.
- [5] KEEMAN, V.: *Learning and Soft Computing – Support Vector Machines, Neural Networks and Fuzzy logic Models*. Massachusetts Institute of Technology, 2001.
- [6] SUYKENS, J., A., K.: *Least squares support vector machines for classification and nonlinear modelling*, *Neural Network World* 1-2, (2000): 29-47.
- [7] VAN GESTEL, T., SUYKENS J., BART DE MOOR, DIRK-EMMA BAESTAENS: *Volatility tube support vector machines*. *Neural Network World*, (2000), No.1-2: 287-297.
- [8] VANDERBEI, R., J., *LOQO: An interior point code for quadratic programming*. TRSOR-94-15, *Statistics and Operation Research*, Princeton Univ., NY, 1994.
- [9] VAPNIK, V.: *Statistical learning theory*, John-Wiley, New-York, 1998.
- [10] VAPNIK, V.: *The nature of statistical learning theory*, Springer Verlag, New-York, 1995.
- [11] VAPNIK, V.: *The support vector method of function*. In: *Nonlinear Modeling: Advanced Black-Box Techniques*, Suykens, J.A.K., Vondewalle, J. (Eds.), Kluwer Academic Publishers, Boston, 1998: 55-85.