

Dušan Marček – Milan Marček *

APPLICATION OF SUPPORT VECTOR MACHINES TO THE MODELLING OF INFLATION

In Support Vector Machines (SVM's), a non-linear model is estimated based on solving a Quadratic Programming (QP) problem. Based on work [1] we investigate the quantifying of econometric structural model parameters of inflation in Slovak economics. The theory of classical Phillips curve [8] is used to specify a structural model of inflation. We provide the fit of the models based on econometric approach for the inflation over the period 1993-2003 in the Slovak Republic, and use them as a tool to compare their approximation ability with those obtained using SVM's method. Some methodological contributions are made for SVM implementations to the causal economic modelling.

Keywords: SVM's, time series analysis, quadratic programming, econometric modelling

1. Introduction

Model specification and estimation are two major components in econometric modelling. They are often treated as two separate but closely related steps in econometric model building. In modelling economic quantities, probably the most important step consists of identifying the relevant influential factors.

This contribution considers the econometric modelling of inflation in the Slovak Republic. The main tools, techniques and concepts involved in econometric modelling of inflation are based on the Phillips concept [8]. According to the Phillips inflation theory the variable inflation is generated on a set of underlying assumptions. In any case, the analysed inflation rates are explained by the behaviour of another variable or a set of variables, in our case by the wages and the unemployment (independent variables).

In this paper the resulting SVM's are applied using an ε -insensitive loss function developed by V. Vapnik. We motivate the approach by seeking a function which approximates mapping from an input domain to the real numbers based on a small subset of training points. The paper is organised as follows. The next section will provide a quick overview of the concept of SVM's theory. Section 3 analyses the data, discusses the Engle-Granger estimator and SVM estimator, and presents the fitted inflation rate values by the classical regression methods and SVM's models in Slovak economics. A section of conclusions will close the paper.

2. Support Vector for Functional Approximation

This section presents quickly a relatively new type of learning machine – the SVM applied in the regression (functional approximation) problems. For details we refer to [11], [12]. 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 $[x_i, y_i] \in \mathcal{X} \times \mathcal{Y}$, $i = 1, 2, \dots, n$ 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 considers the regression functions of two forms [11]. The first one is

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

where α_i, α_i^* are positive real constants (Lagrange multipliers), b is a real constant, $\psi(\cdot, \cdot)$ is the kernel function. Admissible kernels have the following forms: $\psi(x_i, x_j) = x_i^T x_j$ (linear SVM) $\psi(x_i, x_j) = (x_i^T x_j + 1)^d$ (polynomial SVM of degree d), $\psi(x_i, x_j) = \exp(-\theta \|x_i - x_j\|_2^2)$ (radial basis SVM), where θ is a positive real constant and other (spline, b-spline, exponential RBF, etc.).

The second regression function is of the form

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

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 approximation function $f(x, w)$ is explicitly written as a function of the weights w that are subject of learning.

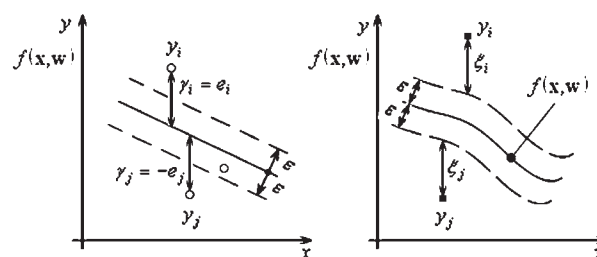


Fig. 1 The insensitive band for one dimensional linear (left), non-linear (right) function

* Dušan Marček¹, Milan Marček²

¹The Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 010 26 Žilina, E-mail: marcek@fria.utc.sk
Institute of Computer Science, The Silesian University Opava, Czech Republic, E-mail: dusan.marcek@fpf.slu.cz

²KIWA, Ltd., Priemyselná 2, 949 01 Nitra, E-mail: tech.kiwa@flynet.sk

The Support Vector 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.

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

In regression, typically some error of approximation is used. They 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]. Left: quadratic 2-norm. Middle: absolute error 1-norm. Right: Vapnik's ε -insensitive loss function.

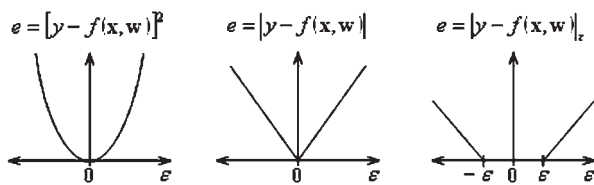


Fig. 2 Error (loss) functions

Formally, this results from solving the following Quadratic Programming problem

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

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

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

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

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]

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

subject to constraints

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

which leads to the solution of the QP problem:

$$\begin{aligned} \max_{\alpha_i, \alpha_i^*} - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* + \alpha_i)(\alpha_j^* + \alpha_j) \varphi(x_i^T x_j) - \\ - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*). \end{aligned} \quad (8)$$

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

$$f(x) = \sum_{i=1}^n (\alpha_i + \alpha_i^*) \psi(x_i, x_j) + b. \quad (9)$$

Finally, b is computed by exploiting the Karush-Kuhn-Trucker (KKT) conditions [3], i.e.

$$\begin{cases} b = y_k - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \psi(x_i, x_k) - \varepsilon & \text{for } \alpha_k \in (0, C) \\ b = y_k - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \psi(x_i, x_k) + \varepsilon & \text{for } \alpha_k^* \in (0, C) \end{cases} \quad (10)$$

3. Causal Models, Experimenting with Non-linear SV Regression

We demonstrate here the use of SV regression framework for dynamic modelling of economic time series where the time series or variable, say inflation, to be modelled can be explained by the behaviour of another variable or a set of variables. First, we present an econometric approach for modelling and investigating the relationship between the dependent variable of inflation measured by *CPI* (Consumption Price Index) and the two independent variables are the unemployment rate (*U*), and aggregate wages (*W*) in the Slovak Republic. Then, the SV regression is applied. Finally, the results are compared between a dynamic model based on econometric modelling and an SVR model.

To study the modelling problem of inflation quantitatively the quarterly data from 1993Q1 to 2003Q4 was collected concerning the consumption price index *CPI*, aggregate wages *W* and unemployment *U*. These variables are measured in logarithm, among others for the reason that the original data exhibit considerable inequalities of the variance over time, and the log transformation stabilises this behaviour. Fig. 3 illustrates the time plot of the *CPI* time series. This time series shows a slight decreasing trend without apparent periodic structure.

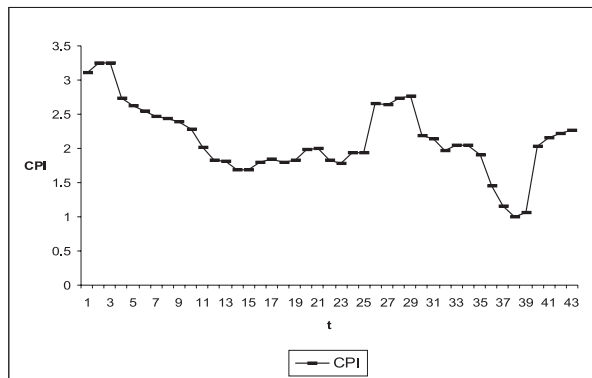


Fig. 3 Natural logarithm of quarterly inflation from January 1993 to December 2003

Experimenting with the linear transfer function models [1], the resulting reasonable model formulation was found

$$\hat{CPI}_t = 0.5941 - 0.0295W_{t-1} - 0.00359U_{t-1} + 0.84524CPI_{t-1} \quad (11)$$

(0.229) (0.3387) (0.1035)

$$R^2 = 0.7762; \quad F(3.40) = 45.089; \quad d = 1.49;$$

$$d_L = 1.38; \quad d_U = 1.67$$

The model specification of (11) is the lagged dependent variable model in which the dependent variables, lagged one period, appear as independent explanatory variables. This model is based on a distributed lag model [9]. One of the primary reasons for using this functional form is to determine the long-run response of dependent variable to change in each of the independent variables. The d statistic for model (11) is 1.49, falling in the inconclusive region, making the decision concerning first-order autoregression indeterminate. In this particular case, a decision has to be made whether or not to correct the autocorrelation. At this point, we will use this model and not correct for autocorrelation. In model (11), the regression coefficients β_1, β_2 have not the appropriate magnitude. In addition, they are statistically insignificant at five percent level. In all probability the lagged dependent variable CPI_{t-1} substitutes for the inclusion of other lagged independent variables (W_{t-1}, U_{t-1}). The inclusion of CPI_{t-1} and W_{t-1}, U_{t-1} is redundant and overspecified the model. A graph of the historical and

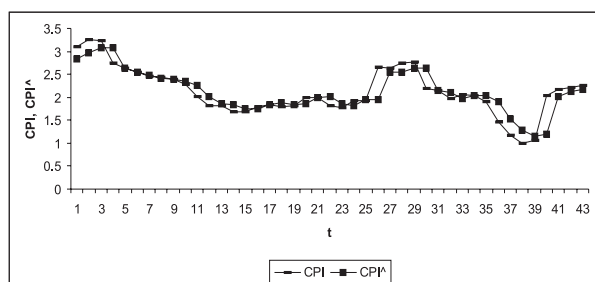


Fig. 4 Natural logarithm of actual and fitted inflation values (model (11))

the fitted values for inflation is presented in Fig. 4. The model follows the pattern of the current inflation very closely.

Finally another attempt was made supposing a more sophisticated dependence of current inflation on the previous observation performed with the help of SV regression. As stated in literature, we can not have a model in which the coefficients are statistically insignificant. We made an arbitrary decision which deleted "insignificant" explanatory variables. Then the equation

(11) becomes the first-order autoregressive process (Markov process) because the observation of CPI at time t depends only on the observation of CPI at time $t - 1$, i. e.

$$\hat{CPI}_t = \beta_0 + \beta_1 CPI_{t-1}, \quad (12)$$

where β_0, β_1 are unknown parameters. If CPI_t exhibits a curvilinear trend, one important approach to generating an appropriate model is to regress the CPI_t against time. In Tab. 1 the SVR results of inflation were also calculated using an alternative time series model expressed by the following SVR form

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

which is a time series model where $x_t = (1, 2, \dots, 43)$ is the vector of time sequence (regressor variable). We report the results in Fig. 5. Since this pattern of change is a common practice, we desire that our machine identify permanent changes and adjust the parameters to track the new process.

One crucial design choice is to decide on a kernel. Creating 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 insight into the application domain. The Fig. 5 shows SVM learning by using various kernels. In Fig. 5a we have a piecewise-linear approximating function, while in Fig. 5b and Fig. 5c we have a more complicated approximating function. Both functions agree with the training points, but they differ on the three y values, they assign to other x inputs. The functions in Fig. 5d and Fig. 5e apparently ignore some of the example points but are good for extrapolation. The true $f(x)$ is unknown, and without further knowledge, we have no way to prefer one of them, and so to resolve the design problem of choosing an appropriate kernel in our application. For example, the objective in pattern classification from sample data is to classify and predict successfully new data, while the objective in control applications is to approximate non-linear functions, or to make unknown systems follow the desired response.

Tab. 1 presents the results for finding the proper model by using the quantity R^2 (the coefficient of determination) on our application of the best approximation of the inflation rate. Since the estimation method for the SVR models is without any sound of statistical theory behind it, the values of the quantity R^2 are basically intuitive, there are no "statistically correct" coefficients of determination for SV regression models. As the calculating of R^2 uses the residuals of the goodness of fit of the least squares line to

the data, we will use the R^2 in the case of the SV regression line as well. As shown in Tab. 1 the “best” is 0.9999 for the time series models with the RBF kernel and quadratic loss functions. In the cases of causal models the best R^2 is 0.9711 with the exponential RBF kernel and ε -insensitive loss function (standard deviation $\sigma = 0.52$). The choice of σ was made in response to the data. In our case, the CPI , CPI_{t-1} time series have $\sigma = 0.52$. The radial basis function defines a spherical receptive field in and the variance σ^2 localises it.

The results shown in Tab.1 were obtained using ε -insensitive loss function ($\varepsilon = 0.2$), with different kernels and degrees of capacity $C = 10^5$. We used partly modified software developed by Steve. R. Gunn [4] to train the SV regression models. The use of

SV regression is a powerful tool to the solution of many economic problems. It can provide extremely accurate approximation of time series, the solution to the problem is global and unique. However, these approaches have several limitations. In general, as can be seen from equations (7), (8), the size of the matrix involved in the quadratic programming problem is directly proportional to the number of training data. For this reason there are many computing problems in which general quadratic programs become intractable in their memory and time requirements. To solve these problems many modified versions of SVM’s were introduced. For example the generalized version of the decomposition strategy is proposed by Osuna et al. [7], the so-called SVM^{light} proposed by Joachims, Thorsten [5] is an implementation of an SVM learner which addresses the problem of a large task, and finally, in [10]

Tab. 1 The SV regression results of different choice of the kernels on the training set (1993Q1 to 2003Q4). In last column the approximation performance is analysed. See text for details.

Fig. 5	MODEL	KERNEL	σ	DEGREE- d	LOSS FUNCTION	R_2
a	causal	exp. RBF	1		ε - insensitive	0.9711
b	causal	RBF	1		ε - insensitive	0.8525
c	causal	RBF	0.52		ε - insensitive	0.9011
d	causal	polynomial		2	ε - insensitive	0.7806
e	causal	polynomial		3	ε - insensitive	0.7860
f	time series	RBF	0.52		quadratic	0.9999

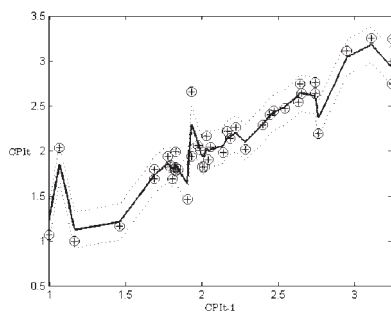


Fig. 5a

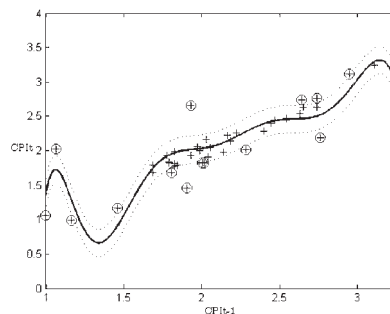


Fig. 5b

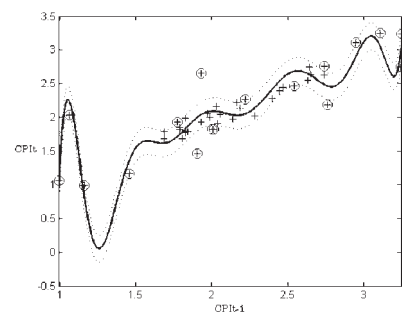


Fig. 5c

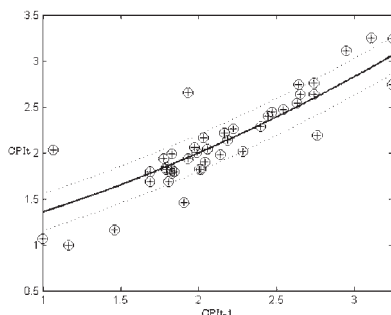


Fig. 5d

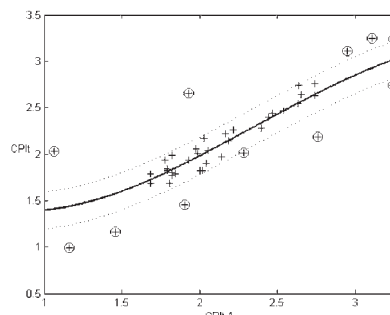


Fig. 5e

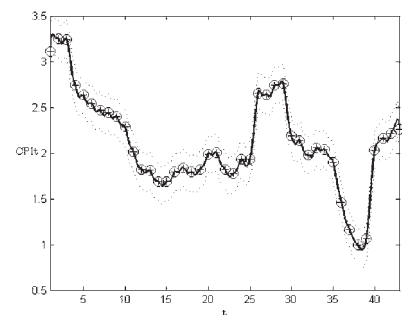


Fig. 5f

Fig. 5 Training results for different kernels, loss functions and of the SV regression (see Tab. 1). The original functions (plus points), the estimated functions (full line), the tube (dotted lines) are shown. Fig. a, b, d, e, f correspond to a good choice of the parameters, Fig. c corresponds to a bad choice.

a modified version of SVM's so-called least squares SVM's (LS-SVM's) is introduced for classification and non-linear function estimation.

4. Conclusion

In this paper, we have examined the SVM approach to study linear and non-linear models on a time series of inflation in the Slovak Republic. For the sake of approximation abilities we evaluated eight models. Two models are based on causal multiple regression in time series analysis, and six models are based on the Support Vector Machines methodology. Using the disposable data a very appropriate econometric model is the regression (11) in

which the lagged dependent variable CPI_{t-1} can substitute for the inclusion of other lagged independent variables (W_{t-1} , U_{t-1}). The benchmarking was performed between traditional statistical approaches and SVMs in regression approximation tasks. The SVM approach was illustrated on the regression function of (12), which was developed by statistical tools. This problem was readily solved by a SV regression with excellent approximation performance as it is visually clear from Fig. 5. Finally, the paper made some methodological contribution for SVM implementations to the causal economic modelling.

Acknowledgement

This work was supported by the grants VEGA 1/2628/05, VEGA 1/0499/03 and GAČR 402/05/2768.

References

- [1] ADAMČA, J., MARČEK, M., PANČÍKOVÁ, L.: *Some results in econometric modelling and forecasting of the inflation in Slovak economics*. Journal of Economics, 52, 2004, No. 9, (1080–1093)
- [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 NY, 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] JOACHIMS, THORSTEN: *Making large-Scale SVM Learning Practical*. In Advances in Kernel Methods - Support Vector Learning, chapter 11. MIT, 1999
- [6] JOHANSEN, S.: *Statistical analysis of cointegrating vectors*. Journals of Economic Dynamics and Control, 12, 1988, (231–254)
- [7] OSUNA, R., FREUND, GIROSI, F.: *An improved training algorithm for support vector machines*. In J. Principe, L. Gile, N. Morgan, and E. Wilson, editors, Neural Networks for Signal Processing VII - Proceedings of the 1997 IEEE Workshop, New York. IEEE, (276–285)
- [8] PHILLIPS, A. W.: *The Relation between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861–1957*, Economica, November 1958
- [9] PINDYCK, R., RUBINFELD, D.: *Economic Models and Economic Forecasts*. 3d ed., Mc Graw-Hill Publishing Company, New York, 1991
- [10] SUYKENS, J., A., K., VAN GESTEL, T., DE BRABANTER, J., DE MOOR, B., VANDEWALLE, J.: *Least Squares Support Vector Machines*. World Scientific Pub. Co., Singapore, 2002
- [11] VAN GESTEL, T., SUYKENS J., BART DE MOOR, DIRK-EMMA BAESTAENS: *Volatility tube support vector machines*. Neural Network World 1-2/2000, (287-297)
- [12] VAPNIK, V.: *Statistical learning theory*. John-Wiley, New York, 1990
- [13] VAPNIK, V.: *The nature of statistical learning theory*. Springer Verlag, New York, 1995.