

Support vector regression for warranty claim forecasting

Shaomin Wu¹, Artur Akbarov

*Cranfield University, School of Applied Sciences, Cranfield,
Bedfordshire MK43 0AL, United Kingdom*

Abstract

Forecasting the number of warranty claims is vitally important for manufacturers/warranty providers in preparing fiscal plans. In existing literature, a number of techniques such as log-linear Poisson models, Kalman filter, time series models, and artificial neural network models have been developed. Nevertheless, one might find two weaknesses existing in these approaches: (1) they do not consider the fact that warranty claims reported in the recent months might be more important in forecasting future warranty claims than those reported in the earlier months, and (2) they are developed based on repair rates (i.e, the total number of claims divided by the total number of products in service), which can cause information loss through such an arithmetic-mean operation.

To overcome the above two weaknesses, this paper introduces two different approaches to forecasting warranty claims: the first is a weighted support vector regression (SVR) model and the second is a weighted SVR-based time series model. These two approaches can be applied to two scenarios: when only claim rate data are available and when original claim data are available. Two case studies are conducted to validate the two modelling approaches. On the basis of model evaluation over six months ahead forecasting, the results show that the proposed models exhibit superior performance compared to that of multilayer perceptrons, radial basis function networks and ordinary support vector regression models.

Keywords: support vector regression, radial basis function network, warranty claims, neural networks, multilayer perceptron.

¹ Suggested Citation: Wu, S., Akbarov, A. Support vector regression for warranty claim forecasting (2011) *European Journal of Operational Research*, 213 (1), pp. 196-204.
Corresponding author: s.m.wu@kent.ac.uk

1 Introduction

A warranty is a contractual obligation incurred by a manufacturer (vendor or seller) in connection with the sale of a product. Nowadays, product warranty becomes increasingly more important in consumer and commercial transactions, and is widely used and serves many purposes [1–6]. The US Congress has enacted several acts (UCC, Magnusson Moss Act, Tread Act, etc.) over the last 100 years. The European Union passed legislation requiring a two-year warranty for all products sold in Europe [7].

Forecasting the number of warranty claims is vital for manufacturers/warranty suppliers in preparing fiscal plans. Starting with Kalbfleisch et al. [8], considerable research on forecasting warranty claims has been conducted (see [8–20], for example).

In existing literature, researchers have developed a number of forecasting techniques, including log-linear Poisson models, Kalman filter, time series models, and artificial neural network models such as multilayer perceptron (MLP) and radial basis function network (RBFN). Nevertheless, one might find two weaknesses in these approaches: (1) these modelling techniques do not consider the fact that the warranty claims reported in the recent months might be more important in forecasting future warranty claims than those reported in the earlier months, and (2) they are developed based on repair rates (i.e., the number of claims divided by the number of products in service), which can cause information loss through such an arithmetic-mean operation.

This paper develops two approaches to overcome the above two weaknesses by using weighted support vector regression models to forecast the number of warranty claims. The approaches developed can be used for the following two scenarios: when only claim rate data are available and when the original claim data are available. The performance of different modelling approaches has been compared on two industrial datasets.

In this paper, we will use the terms *product* and *item* interchangeably. The paper is structured as follows. Section 2 reviews prior work on warranty claim forecasting, and details the problems in warranty claim forecasting. Section 3 describes weighted support vector regression models. Section 4 applies the proposed models on two datasets: one is from an open publication by Wasserman and Sudjianto [19] and the other is warranty claims data collected from an electronics manufacturer. This section also compares the performance of the modelling techniques discussed in this paper. Section 5 draws conclusions and suggests further research work.

2 Problem description and prior work

This section first discusses the problem of warranty claim prediction and reviews existing approaches to tackle this problem.

2.1 Acronyms and notation

Table 1 and Table 2 show the acronyms and the notation used in this paper.

Table 1

Acronyms

ANN	artificial neural network
FBNR	failed but not reported
MIS	month in service
MLP	multilayer Perceptron
NHPP	non-homogeneous Poisson process
ORP	ordinary renewal process
RBFN	radial basis function networks
RBNF	reported but not failed
SVR	support vector regression
tSVR	weighted SVR-based time series model
wSSE	weighted sum squared error
wSVR	weighted support vector regression

2.2 Warranty claim data

We assume that claim data are aggregated on a monthly basis. These data can be expressed as shown in Table 3, where i ($i = 1, 2, \dots, m$) in the 1st column represents month of sale.

The question of interest is how to forecast warranty claims $x_{i,n_0+1}, \dots, x_{i,n_0+k}$ for $(n_0 + 1)$ -th, $(n_0 + 2)$ -th, $\dots, (n_0 + k)$ -th months given that claims data, $x_{i,1}, \dots, x_{i,n_0}$, until the n_0 -th month is available.

2.3 Estimating the number of warranty claims

Forecasting warranty claims, in general terms, is to forecast the expected number of claims within the warranty coverage. The choice of the appropriate model for forecasting the expected number of claims depends on a variety of factors such as: the nature of the product (repairable vs. non-repairable), warranty coverage scope (full vs. limited), the replacement policy (pro rata vs. free), and warranty dimension (one vs. two dimensional) [21].

According to [22], warranty claims might have the following two characteristics:

Table 2

Notation

$F(t), R(t)$	$F(t) = 1 - R(t)$, where $F(t)$ is the cumulative distribution function and $R(t)$ is the reliability function of the products of interest
$f(t)$	$= \frac{dF(t)}{dt}$
$\lambda(t)$	failure intensity function
m	total number of sales months
m_j	$m_j = m$ if $j \leq n - m + 1$, and $m_j = n + 1 - j$ if $n - m + 1 < j \leq n$, that is, m_j is the number of observations in month j
N_i	number of products sold in the i th month
$M_{i,j}$	$M_{i,1} = N_i$, and $M_{i,j} = N_i - \sum_{i=1}^{j-1} x_{i,j}$ for $j > 1$, that is, $M_{i,j}$ is the remaining number of products in the market from sales batch N_i in month j
$\mu(t)$	$= F(t) + \int_0^t \mu(t-s)f(s)ds$, which is the renewal function
n	largest months of products in service
n_0	number of months used for training forecasting models
\tilde{N}	$= \sum_{i=1}^m N_i$, which is the total number of products sold
\tilde{N}_j	$= \sum_{i=1}^{m_j} N_i$, which is the total number of products aged j
$N_{ri}(t_1, t_2)$	expected number of warranty claims for the repairable products sold in the i -th month within time period (t_1, t_2) , for $t_2 > t_1$ and $t = 1, 2, \dots$
$N_{nri}(t_1, t_2)$	expected number of warranty claims for the non-repairable products sold in the i -th month within time period (t_1, t_2) , for $t_2 > t_1$ and $t = 1, 2, \dots$
p_1, p_2	positive integers
$p(t)$	probability that a claim is not executed on an item, given that the item fails at time t .
$q(t)$	probability that a claim is made on an item, given that the item does not fail at time t .
r_j	$= \sum_{i=1}^{m_j} x_{i,j}$, which is the total number of claims for products aged j
h_j	$= \frac{r_j}{N_j}$, which is the claim rate of the products aged j
$x_{i,j}$	number of warranty claims for products aged j and sold in the i -th month, where $i = 1, 2, \dots, m, j = 1, \dots, n$
\mathbf{h}_j	$= (h_{j-1}, \dots, h_{j-p_1})$
$\mathbf{z}_{i,j}$	$= (j, M_{i,j}, M_{i,j-1}, \dots, M_{i,j-p_2+1})$

- failed items might not be reported warranty, and
- warranty claims might not be due to product failures.

The above two factors are especially common in warranty claims of electronics products. These two factors are called FBNR (failed but not reported) and RBNF (reported but

Table 3
Warranty claims ($n \geq m$)

Sold month	Sold amount	Months in service (MIS)							
		1	2	...	$n - m + 1$	$n - m + 2$...	$n - 1$	n
1	N_1	x_{11}	x_{12}	...	$x_{1,n-m+1}$	$x_{1,n-m+2}$...	$x_{1,n-1}$	$x_{1,n}$
2	N_2	x_{21}	x_{22}	...	$x_{2,n-m+1}$	$x_{2,n-m+2}$...	$x_{2,n-1}$	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
$m - 1$	N_{m-1}	$x_{m-1,1}$	$x_{m-1,2}$...	$x_{m-1,n-m+1}$	$x_{m-1,n-m+2}$			
m	N_m	$x_{m,1}$	$x_{m,2}$...	$x_{m,n-m+1}$				

not failed) respectively. Both, the number of warranty claims due to FBNR and RBNF, account for quite a large proportion in the total number of claims [22]. The existence of these two factors makes it difficult to use the reliability function, $R(t)$, in forecasting warranty claims. The following demonstrates the estimation of the number of warranty claims for both repairable and non-repairable items.

2.3.1 Repairable items

We assume that a failed item will be returned to the manufacturer and the item can be restored to the same failure intensity as it had when it failed (i.e., minimal repair). Then within a given time interval (t_1, t_2) , the expected number of warranty claims for the products sold in the i -th month is given by

$$N_{ri}(t_1, t_2) = N_i \int_{t_1}^{t_2} (1 - p(v))\lambda(v)dv + N_i \int_{t_1}^{t_2} q(v)dv \quad (1)$$

In Eq. (1), the first term represents the number of claims made on a certain proportion of failed products dictated by the probability of raising a claim on a failed product $1 - p(t)$. This term takes into account only the number of failed products that were reported as warranty claims. Some product failures are not reported as warranty claims due to reasons such as obsolescence of the product, or new purchase. The second term represents the number of claims made on products that have not failed. That is, some warranty claims are raised due to other reasons than product failure, such as fraudulent claims, claims that can be resolved by giving appropriate instructions on how to operate the products.

2.3.2 Non-repairable items

We assume that a failed item will be returned to the manufacturer and a new identical item will be returned to the customer. Then within a given time interval (t_1, t_2) , the expected number of warranty claims for the non-repairable products sold in the i -th month is given by

$$N_{\text{wri}}(t_1, t_2) = N_i \left(\int_0^{t_2} (1 + \mu(t_2 - v))f(v)(1 - p(v))dv - \int_0^{t_1} (1 + \mu(t_1 - v))f(v)(1 - p(v))dv \right) + N_i \int_{t_1}^{t_2} q(v)dv \quad (2)$$

In Eq. (2), as in the case of Eq.(1), the first term represents the number of claims made on some failed products depending on $1 - p(t)$, and the second term represents the number of claims made on the products that have not failed.

2.3.3 Comments

A manufacturer might be able to estimate $f(t)$ based on historical data or expert opinion. However, estimating customer behaviour functions $p(t)$ and $q(t)$ might be difficult. This is especially true for electronics products such as computers, mobile phone, and so on. The users of these products might not be bothered to claim warranty, especially when the failed items have served for quite a long time and/or they are not expensive. In such scenarios, forecasting warranty claims based on $f(t)$ might not be appropriate.

2.4 Prior work on warranty claims forecasting

2.4.1 Existing approaches

There are a number of approaches developed to forecasting warranty claims. These approaches are briefly discussed below.

Lifetime distributions. Kleyner and Sandborn [9] present a warranty claim forecasting model based on a piecewise application of Weibull and exponential distributions. This model tries to capture the dynamic characteristic features of failure rates in both early failure period and the intrinsic failure period of the bathtub curve. Rai [10] presents a forecasting model incorporating calendar month seasonality, business days per month for authorised service centres, and sales ramp-up.

Stochastic processes. Kalbfleisch et al. [8] used a log-linear Poisson model to analyse and forecast warranty claims. In their work, they modelled warranty claims based on the date of warranty claim rather than the failure date, and therefore the reporting lag between occurrence of a claim and its entry to a database was taken into consideration. Dynamic linear models with leading indicators are also used in [14]. Kaminskiy [15] develop warranty claim forecasting models with three types of stochastic processes: the G-renewal process that is a generalised renewal process introduced by Kijima and Sumita [23], the ordinary renewal process (ORP), and the non-homogeneous Poisson process (NHPP). They found that GRP provides a higher accuracy compared to the ORP or the NHPP. Majeske [17] presents a NHPP-based technique that forecasts the total number of claims and the timing of claims during the vehicle lifetime. Fredette and Lawless [18] present forecasting methods for warranty claims using flexible non-homogeneous Poisson processes, where possible heterogeneity among the products is modelled using random effects.

Artificial neural networks. Wasserman [11] and Wasserman and Sudjianto [19] use a neural network model, MLP (multi-layer perceptron), to forecast warranty claims. Rai and Singh [12] use another neural network model, RBFN (radial basis function network), to forecast warranty claims. Hrycej and Grabert [13] use MLP (multi-layer perceptron) to forecast warranty cost based on individual vehicle variables (i.e. age, monthly mileage rate, and road condition index) and the overall manufacturing quality fluctuation risk (i.e. different technical groups).

The Kalman filter and time series models. Singpurwalla and Wilson [14] consider using the Kalman filter to build forecasting models. Wasserman [11] develops linear regression models, first-order auto-regression time series models, and the Kalman filter models to forecast warranty claims. In his linear regression models, the number of months in service is used to forecast the number of repairs per 1000 items. Wasserman and Sudjianto [19] further compare the performance of linear regression models, time series models, the Kalman filter, the orthogonal series, and artificial neural networks in modelling and forecasting warranty claims. They conclude that the Kalman filter model offers a significant improvement over simple linear regression approach, but both the orthogonal series and the neural network models outperform the Kalman filter. In the same year, Chen et al. [20] also propose to model and forecast the number of warranty claims with the Kalman filter.

2.4.2 *Comments on the approaches reviewed above*

The existing approaches reviewed in the previous section have the following weaknesses.

- When fitting a lifetime probability distribution, one might assume that the warranty data represents the exact number of failed products (for example see [10]). However, such an assumption might not hold due to FBNR and/or RBNF phenomena, as discussed in Section 2.3.
- Stochastic process approaches might require rigid assumptions such as the claim rates following a specific form. For example the assumption of a power law process for NHPP[14,15,18,17]. Such an assumption might not hold in reality partly the fact that some users might not be bothered to execute warranty claims on product failures.
- The approaches developed on the basis of repair rates (or claims rates) [11,20,19,9] may cause information loss, as they are obtained as a ratio of warranty claims to the number of products in service (i.e. they integrate two observations into one).
- None of the above-reviewed modelling techniques has considered the fact that the warranty claims reported in the recent months might be more important for forecasting future claims than those reported in the earlier months. This might be due to the following fact: most of the existing research does not consider human factors (such as FBNR and RBNF) and assumes those items that have been claimed are failed ones and those ones that have not been claimed are non-failed ones[22].

To overcome the above weaknesses, we should develop a new approach that (1) employs original claim data (instead of claim rates/repair rates), which allows us to use all of the available information; (2) gives more weights to the recent warranty claims as the most recent data can be more useful for predicting future claims — due to the causes of FBNR and RBNF; and (3) considers a more flexible model structure (to use a nonparametric

approach rather than a parametric one).

To include all of these three requirements, we propose new warranty claims forecasting approaches:

a weighted support vector regression model with original claim data as model input and output variables, and an SVR-based time series model for the case where only claim rate data are available.

3 Weighted support vector regression

Support vector regression (SVR), developed in the context of statistical learning theory, is a novel nonparametric methodology for creating regression functions [24]. The aim of the SVR is to optimise generalisation performance. Artificial neural networks (ANN) have been criticised for their vulnerability to the over-fitting problem—which might result in a good fit when we train a model, but poor prediction ability when we test the model on a different test dataset. The SVR, in contrast, has an excellent generalisation performance, or prediction ability in the test dataset. Below we present a brief introduction to ϵ support vector regression, for a detailed overview of the SVR the reader is referred to [24,25]. For the applications of the SVR the reader is referred to [29,28,26,27].

3.1 Overview of ϵ -support vector regression

Suppose we are given training data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l) \subset \mathbb{R}^d \times \mathbb{R}$, where \mathbb{R}^d denotes the space of the independent variables and \mathbb{R} is the space of the dependent variable. For a linear function $g(\cdot)$:

$$g(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b_0, \quad \mathbf{w} \in \mathbb{R}^d, b_0 \in \mathbb{R}, \quad (3)$$

where $\langle \cdot, \cdot \rangle$ represents the dot product in \mathbb{R}^d , \mathbf{w} is the weights vector that needs to be estimated, and b_0 is the intercept. We aim to seek a function $g(\mathbf{x})$ that has a small norm of \mathbf{w} , or to ensure $g(\mathbf{x})$ as flat as possible. We can write this problem as a convex optimisation problem:

$$\begin{aligned} & \text{minimise} && \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{subject to} && \begin{cases} y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle - b_0 & \leq \epsilon \\ \langle \mathbf{w}, \mathbf{x}_j \rangle + b_0 - y_j & \leq \epsilon \end{cases} \end{aligned} \quad (4)$$

where $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle$. Such an optimisation problem implicitly assumes the existence of a solution for $g(\cdot)$. However, this is not always the case, and it is often desired to allow for some errors higher than ϵ . This problem is resolved by introducing slack variables, δ_j and δ_j^* (for detailed discussion, see [25]). Thus, the optimisation problem for finding suitable

$g(\cdot)$ can be stated as:

$$\begin{aligned} & \text{minimise} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^l (\delta_j + \delta_j^*) \\ & \text{subject to} && \begin{cases} y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle - b_0 & \leq \epsilon + \delta_j \\ \langle \mathbf{w}, \mathbf{x}_j \rangle + b_0 - y_j & \leq \epsilon + \delta_j^* \\ \delta_i, \delta_j^* & \geq 0 \end{cases} \end{aligned} \quad (5)$$

where l is the number of observations, δ_j and δ_j^* are error terms. The first term of the objective function in Eq. (5) will be minimised to ensure that the function $g(\cdot)$ is as flat as possible, and the second term, $C \sum_{i=1}^l (\delta_i + \delta_i^*)$, is known as the regularisation term and ensures that the optimisation problem posed by Eq. (5) is feasible. The parameter C determines the trade-off between the complexity of the model and errors. It is often determined by minimising some error function on a validation dataset or by the means of cross-validation.

Eq. (5) can be restated in terms of the Lagrangian multipliers, which allows the extension of the linear $g(\cdot)$ to nonlinear functions and makes it is easier to construct algorithms for solving the optimisation problem. After constructing the Lagrangian and taking the derivatives it is possible to arrive at the following reformulation of Eq. (5):

$$\begin{aligned} & \text{maximise} && \begin{cases} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{cases} \\ & \text{subject to} && \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (6)$$

where α and α^* are the Lagrangian multipliers of the first two constraints in Eq. (5) respectively. For all data points that lie inside the ϵ -tube, if $|g(\mathbf{x}_i) - y_i| < \epsilon$, α_i and α_i^* vanish. That is, we do not need all \mathbf{x}_i to describe \mathbf{w} . The data points whose α_i, α_i^* do not vanish are called support vectors. As a result of the Lagrangian expansion, we also have:

$$g(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b. \quad (7)$$

This is known as the support vector expansion. This expansion shows that the complexity of a function's representation by support vectors is independent of the dimensionality of the input space, and depends only on the number of support vectors. Only for $|g(\mathbf{x}_i) - y_i| \geq \epsilon$ the Lagrange multipliers may be nonzero.

The complexity of the SVR model depends on the number of support vectors and the regularisation constant C . High values of C increase the cost of errors in the training set and force a more accurate model for $g(\cdot)$. Such a model may not generalise well and lead to overfitting. Conversely, smaller values of C can lead to much simpler models allowing for more errors and result in underfitting.

For nonlinear functions, SVR can be applied in the so-called feature space, where $g(\cdot)$ becomes nonlinear. This can be done by mapping: $\Phi : \mathbb{R}^d \rightarrow F^d$, where F^d is some d

dimensional feature space. As can be noted from Eq. (7) the SVR algorithm needs only dot products, hence it is sufficient to determine $k(\mathbf{x}_i, \mathbf{x}) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle$. Thus, we have:

$$g(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}) + b, \quad (8)$$

where $k(\mathbf{x}_i, \mathbf{x})$ is known as the Kernel function.

Below we consider two different approaches to building forecasting models: the first one is a weighted support vector regression (SVR) model, and the second one is a weighted SVR-based time series model.

3.2 Time series model

The time series models are built based on warranty claim rate, h . The model can be expressed as follows:

$$h_j = g_1(h_{j-1}, h_{j-2}, \dots, h_{j-p_1}) + e_j. \quad (9)$$

where p_1 is the order and e_j is the error. Eq. (9) can also be re-written as

$$h_j = g_1(\mathbf{h}_j) + e_j.$$

A similar model has been considered by Wasserman and Sudjianto [19] for data from automotive industry, where in addition to h_{j-1} the claim rate for the previous year car model was also used as an independent variable.

We can use support vector regression modified by Cao and Tay[30], which is called tSVR (or weighted SVR-based time series model) hereafter, to build the above time series model.

The tSVR aims to find a function $g_1(\cdot)$ which has at most ϵ_1 deviation from the actually observed values of h :

$$g_1(\mathbf{h}_j) = \mathbf{w}_1^T \phi_1(\mathbf{h}_j) + b_1 \quad (10)$$

where b_1 is the intercept, \mathbf{w}_1 is the weights vector and ϕ_1 is a mapping to some feature space. During the estimation process, the explicit mapping to ϕ_1 is generally dealt away with by making use of kernels $k(\mathbf{h}_i, \mathbf{h}_j) = \phi_1(\mathbf{h}_i) \phi_1(\mathbf{h}_j)$ (see [25] for more details). The resulting optimisation problem can be stated as follows:

$$\begin{aligned} & \text{minimise} \quad \frac{1}{2} \|\mathbf{w}_1\|^2 + \sum_{j=1}^{n_0} c_{1,j} (\zeta_j + \zeta_j^*) \\ & \text{subject to} \quad \begin{cases} h_j - \mathbf{w}_1^T \phi_1(\mathbf{h}_j) - b_1 \leq \epsilon_1 + \zeta_j \\ \mathbf{w}_1^T \phi_1(\mathbf{h}_j) + b_1 - h_j \leq \epsilon_1 + \zeta_j^* \\ \zeta_j, \zeta_j^* \geq 0 \end{cases} \end{aligned} \quad (11)$$

where n_0 is the number of observations, h_j is the actual warranty claim rate at time j . The second term in the optimisation function represents the regularisation term. Here, the values of $c_{1,j}$ control the trade off between the complexity of the model and the

errors ζ_j and ζ_j^* . The variable values of $c_{1,j}$ as opposed to constant C allow to give different importance for errors depending on their temporal distance. For example, the errors on most recent observations can be given more importance than the errors on earlier observations. The specification of $c_{1,j}$ for all values of j depends on the specific data set and can be determined through cross-validation.

Although model (9) and model (11) have considered the fact that recent claims should have more weights than the older ones, they still use claim rates as dependent and independent variables. Below we introduce a new modelling approach that employs original warranty claims as model's dependent variable.

Model (11) can deal with the situation when only claim rates are available, but original claims cannot be obtained.

3.3 Regression models

The regression models are built as following.

$$x_{i,j} = g_2(j, M_{i,j}, M_{i,j-1}, \dots, M_{i,j-p_2+1}) + e_{i,j}, \quad (12)$$

where $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, m$. Eq. (12) can also be rewritten as

$$x_{i,j} = g_2(\mathbf{z}_{i,j}) + e_{i,j}.$$

A weighted SVR (wSVR) can be expressed as follows.

$$g_2(\mathbf{z}_{i,j}) = \mathbf{w}_2^T \phi_2(\mathbf{z}_{i,j}) + b_2 \quad (13)$$

where b_2 is the intercept, \mathbf{w}_2 is the weights vector, and ϕ_2 is a mapping to some feature space.

The wSVR aims to find a function $g_2(\cdot)$ which has at most ϵ_2 deviation from the actually observed values of $x_{i,j}$, and at the same time is as flat as possible. Allowing for some errors higher than ϵ_2 , $\xi_{i,j}$ and $\xi_{i,j}^*$, the optimisation problem for finding suitable $g_2(\cdot)$ can be stated as:

$$\begin{aligned} & \text{minimise} \quad \frac{1}{2} \|\mathbf{w}_2\|^2 + \sum_{j=1}^{n_0} \left(c_{2,j} \sum_{i=1}^{m_j} (\xi_{ij} + \xi_{ij}^*) \right) \\ & \text{subject to} \quad \begin{cases} x_{i,j} - \mathbf{w}_2^T \phi_2(\mathbf{z}_{i,j}) - b_2 \leq \epsilon_2 + \xi_{ij} \\ \mathbf{w}_2^T \phi_2(\mathbf{z}_{i,j}) + b_2 - x_{i,j} \leq \epsilon_2 + \xi_{ij}^* \\ \xi_{ij}, \xi_{ij}^* \geq 0 \end{cases} \end{aligned} \quad (14)$$

As in the previous model, the values of $c_{2,j}$ control the trade-off between the complexity of the model and errors. Here, however, the errors, ξ_{ij} and ξ_{ij}^* , on the even-aged products received in different months are assigned the same importance; for example, for a given

MIS (month in service), the weights are the same, but the weights are different over different MIS.

Model (14) can deal with the situation when the original claims are available.

3.4 Selection of $c_{1,j}$ and $c_{2,j}$

We borrow the approach introduced by Cao and Tay [30] to search for the optimal values of $c_{1,j}$ and $c_{2,j}$:

$$c_j = a_0 \frac{2}{1 + \exp(a_1 - 2a_1 \frac{j}{n_0})} \quad (15)$$

where $c_j = c_{1,j}$ for the tSVR, $c_j = c_{2,j}$ for the wSVR, a_0 and a_1 are unknown parameters that need to be estimated empirically. c_j is an increasing function with respect to j . Ascending values of c_j allow more errors for distant data points.

4 Case studies

This section aims to compare the performance of several forecasting methods commonly used in practice. The comparison is made based on two different datasets, one from the previous study by [19], Table 4, and the other from an electronics manufacturer, Table 5.

4.1 The datasets

The first dataset is shown in Table 4, we refer to this dataset as WS dataset. The time in service column in the table is given in terms of months. The prior and current year model columns represent repair rate per 1000 cars. Wasserman and Sudjianto [19] use this dataset to test a number of algorithms: time series model, Kalman filter and MLP. These models are defined for current year model in terms of time in service and prior year model.

The second dataset is supplied by an electronics manufacturer, whose name will not be disclosed here for the confidentiality reason. The products are a type of internet networking equipment. The original dataset has the same format as shown in Table 3 and includes three variables: monthly sales amounts, monthly warranty claims, and months in service (MIS). In order to keep the original data confidential, here we present a table of the claim rate per MIS of the original data. The claim rate data is presented in Table 5, where the MIS columns show the number of MIS and the h columns show claim rate per 10000 products. The warranty claims data of this dataset has the following properties:

- The warranty claims are aggregated on a monthly basis.
- The warranty policy of the product is a long-term warranty policy, and it expires when the technological obsolescence has been announced.

Table 4

Warranty claims data (adopted from [19]). Time in service is given in terms of months, and prior and current year data are given in terms of repair rate per 1000 cars.

Time in service	Prior model year	Current model year	Time in service	Prior model year	Current model year	Time in service	Prior model year	Current model year
1	0.55	1.27	13	78.2	53.3	25	167	172
2	4.68	1.58	14	84.0	61.8	26	173	186
3	9.88	1.58	15	84.7	75.9	27	179	199
4	12.8	7.71	16	89.0	90.8	28	185	211
5	15.5	11.4	17	97.8	102	29	187	228
6	20.0	18.7	18	110	105	30	189	249
7	25.9	26.3	19	125	113	31	196	271
8	33.1	35.4	20	139	125	32	205	290
9	37.5	42.0	20	139	125	33	211	300
10	45.4	46.4	22	158	139	34	213	328
11	56.4	48.2	23	163	159	35	215	375
12	68.9	48.8	24	166	165	36	221	430

- The products are repairable. Once a claim is received, the manufacturer will immediately send out an identical product to the customer as a replacement. The received (or claimed) product will be tested and/or repaired. The sent-out product can be at any age: it can be new, or can be old. Due to this age uncertainty, it is difficult to estimate the number of claims/failures within a specific time period with conventionally used statistical methods such as the renewal process or NHPP.

Table 5

Warranty claims data from an electronics manufacturer. MIS is given in terms of months, and h is given in terms of warranty claim rate per 10000 items.

MIS	h	MIS	h	MIS	h	MIS	h	MIS	h
1	3.89	8	21.08	15	20.20	22	17.84	29	14.43
2	9.42	9	22.69	16	19.98	23	17.46	30	19.12
3	14.23	10	20.23	17	20.05	24	14.70	31	17.92
4	17.41	11	23.07	18	16.68	25	14.68	32	23.39
5	20.71	12	21.94	19	17.36	26	18.29	33	17.28
6	21.87	13	19.79	20	16.56	27	18.51	34	41.91
7	22.61	14	18.25	21	18.58	28	19.71		

4.2 Alternative modelling methods and performance measure

We apply the following five modelling methods: MLP, RBFN, SVR, tSVR, and wSVR on the above two datasets. We code our computational programs with Java and borrow some

functions from two data mining packages, Weka² and LIBSVM³. Each modelling method has its own set of hyper-parameters that need to be specified by users. Below we briefly describe what hyper-parameters have been used for each of the forecasting methods.

4.2.1 Modelling methods used for comparison

Radial basis function network (RBFN). The RBFN used here has the following three hyper-parameters: (1) the number of Gaussian radial basis functions (GRBF) for the K-means algorithm to generate, (2) the minimum of standard deviation (MSD) for the GRBFs, and (3) the ridge value (R) for linear regression that is used for combining the outputs of GRBFs.

Multilayer perceptron (MLP). The MLP used in this experiment is a network with a single hidden layer that uses the backpropagation algorithm to find its optimal parameters. The hyper-parameters of this network are: (1) the number of hidden nodes in the hidden layer (NHN), which controls the complexity of the network, (2) the learning rate (LR), for determining the step size for the optimisation problem, and (3) the momentum (M), which is used to control convergence.

Support vector regression (SVR). The SVR has the following three main hyper-parameters: (1) the constant C which determines the trade-off between the complexity of the model and tolerance of errors, (2) ϵ —which specifies the margins for negligible errors, and (3) γ —the parameter of the RBF kernel for the case of SVR with RBF kernel.

Weighted SVR-based time series model (tSVR). In the case of tSVR, instead of C , as used in the above SVR, we use an adaptive $C(\cdot)$ (as shown in Eq. (15)). Thus, in addition to optimising the values of a_0 and a_1 in Eq. (15), we also need to optimise the values of ϵ and γ .

Weighted support vector regression. In the case of wSVR, it has the same hyper-parameters as those of tSVR.

To search for the optimal hyper-parameters of the above models, we use genetic algorithms (GA), which are adaptive search and optimisation algorithms that are based on the principles of natural evolution theory. The reader is referred to [31] for an introduction to GA.

The following GA parameter settings are used for estimating the hyper-parameters of the forecasting methods. The crossover is implemented using a single point crossover, where for a pair of chromosomes a gene is selected randomly and swapped including all of the subsequent genes between the two chromosomes. The crossover rate is set to 35% of the population size. The mutation rate is set to 8.3%. The population size is kept fixed throughout the consecutive evolutions. 90% of the original population is selected for the next generation with the addition of duplicate chromosomes. A typical population size used for the experiments below is 75 with 50 population evolutions. The GA is implemented based on Java open source component JGAP - Java Genetic Algorithms Package³.

² <http://www.cs.waikato.ac.nz/ml/weka/>, accessed on 01 October 2010

³ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, accessed on 01 October 2010

4.2.2 Performance measure

To evaluate model performance, we conduct six months ahead forecasting in this paper. In order to compare the performance of different forecasting methods on the data, we use the following metrics to measure the performance of a model:

MSE: For the time series model for WS dataset, we use the following mean squared error (MSE) as a performance metric.

$$\text{MSE} = \frac{1}{6} \sum_{j=n_0+1}^{n_0+6} (h_j - \hat{h}_j)^2 \quad (16)$$

where \hat{h}_j is a claim rate estimated by a model. For example, if $p_1 = 1$ in Eq. (9), then \hat{h}_j , for $j = n_0 + k$ and $k = 1, 2, \dots, 6$, can be obtained with the following rule:

$$\hat{h}_{n_0+k} = \begin{cases} g_1(h_{n_0}) & \text{for } k = 1 \\ g_1(\hat{h}_{n_0+k-1}) & \text{for } 2 \leq k \leq 6 \end{cases} \quad (17)$$

wSSE: For the regression models, we use the following weighted sum squared error (wSSE) to measure model performance:

$$\text{wSSE} = \sum_{j=n_0+1}^{n_0+6} \sum_{i=1}^{m_j} \frac{(x_{i,j} - \hat{x}_{i,j})^2}{M_{i,j}} \quad (18)$$

where $\hat{x}_{i,j}$ is the number of claims estimated by a model. For example, if $p_2 = 1$ in Eq. (12), then $\hat{x}_{i,j}$, for $j = n_0 + k$ and $k = 1, 2, \dots, 6$, can be obtained with the following rule:

$$\hat{x}_{i,n_0+k} = \begin{cases} g_2(j, M_{i,n_0+k}) & \text{for } k = 1 \\ g_2(j, \hat{M}_{i,n_0+k}) & \text{for } 2 \leq k \leq 6, \end{cases} \quad (19)$$

where $\hat{M}_{i,n_0+k} = \hat{M}_{i,n_0+k-1} - \hat{x}_{i,n_0+k-1}$ and $\hat{M}_{i,n_0+1} = M_{i,n_0} - x_{i,n_0}$.

For the dataset shown in Table 4, following [19], we use the data of the first 30 months as the training dataset and the data from 31st to 36th months as the validation dataset. The hyperparameters of a given forecasting method are selected from the model with the smallest MSE over the validation dataset based on six months ahead forecasting using Eq. (16). The results of estimated MSEs are presented in Table 6.

For the dataset shown in Table 5, we divide the dataset into three sets: the training dataset, the validation dataset, and the test dataset. The training dataset contains the data including the first 22 MIS, the validation dataset includes claims with MIS between 23 and 28, and test dataset contain claims with MIS between 29 and 34. The hyperparameters of the forecasting methods are optimised to yield the smallest MSE for the models with claim rates as dependent variables (or the time series models) and the smallest wSSE for the models with original claims as dependent variables (or the regression models) over the validation dataset. After the optimal hyperparameters have been selected, we merge

the training dataset and the validation dataset to create a new training dataset. We then train a new model with the optimal hyperparameters on the new training dataset. The predictions are made based on Eq. (17) and Eq. (19) accordingly. In order to compare the time series models against the regression models, we convert the predicted rate values, \hat{h}_j , into the original data format, $\hat{x}_{i,j}$, and evaluate the prediction performance on the test dataset based on Eq. (18). This is due to fact that in practise, we are mainly interested in predicting the actual warranty claims data rather than the rates. Therefore, evaluation of the prediction performance based on wSSE for both time series and regression models seems to be appropriate. This is not possible to implement on the WS dataset as the original warranty claims data is not publicly available.

Since the GA might reach a local minimum, we repeat the evaluation experiment for each model 30 times. The results of these experiments are then compared using the t-test. The prediction performance (i.e. the values of MSE and wSSE) of different forecasting models (i.e. MLP, RBFN, SVR, tSVR and wSVR) are shown in Table 7, and Table 9.

4.3 Results on the WS dataset

Table 6 shows the values of MSE for each forecasting model. The best result by Wasserman and Sudjianto [19] was obtained by using a two-node neural network with error on the validation set being $MSE = 78.18$. For SVR and tSVR, the MSE for the validation set is lower than 78.18. It can be seen that between SVR and tSVR, tSVR performs slightly better than SVR.

Table 6

Mean squared error over training and validation dataset for WS dataset.

Forecasting method	Hyper-parameters	MSE on training dataset	MSE on validation set
RBFN	GRBF=6, MSD=2.3768, R=0.0111	20.82	85.68
MLP	NHN=10, LR=0.7351, M=0.6972	223.25	77.17
SVR	C=904.109, ϵ =6.2478, γ =0.03838	16.61	74.11
tSVR	a_0 =752.818, a_1 =9.3178, ϵ =5.9305, γ =0.0215	14.74	71.62

4.4 Results on the industrial dataset

The experiments for the claim data shown in Table 5 were conducted using the following search bounds for the hyperparameters of the forecasting methods. For MLP, we have, $NHN = \{1,2,3,4\}$, $LR \in [0.2, 0.8]$, and M is set to a fixed value as the number of iterations is sufficiently large, 500, $M = 0.2$. For RBFN, we have, $GRBF = \{1,2,3,4\}$, $R \in [0, 1]$ and $MSD = 0.1$. For SVR, we have $C \in [1, 1000]$, $\epsilon \in [0, 1]$, and γ is set to a fixed value of one over the number of attributes. For the wSVR method, we have $b \in [1, 1000]$, $a \in [0, 10]$ and ϵ and γ are set to the same values as in the case of SVR. Here, we present the results for models with $p_1 = 1$ and $p_2 = 1$.

4.4.1 Time series models

Table 7 shows the means of wSSE values (and their standard deviations in the brackets) on validation datasets and test datasets. T-test for sample means with unequal variances are also carried out for comparing the performance of different models, as shown in Table 8. We can find that, on validation datasets, MLP outperforms both RBF and SVR, and tSVR outperforms all of the other three approaches. tSVR also outperforms all of the other three approaches on test datasets.

Table 7

Means (standard deviations) of the performance of the four time series models on the industrial dataset.

Dataset	MLP	RBF	SVR	tSVR
validation dataset	0.9264(8.23E-08)	0.9532(2.04E-31)	0.9286(4.34E-07)	0.9178(8.28E-07)
test dataset	0.9871(6.4E-05)	0.9508(0.00)	0.9545(6.9E-06)	0.9487(1.75E-07)

Table 8

Comparison of the four time series models on the industrial dataset: t-values.

Dataset	MLP vs.	MLP vs.	MLP vs.	RBF vs.	RBF vs.	SVR vs.
	RBF	SVR	tSVR	SVR	tSVR	tSVR
validation dataset	-511.61	-16.48	49.46	204.73	213.13	52.54
test dataset	24.87	21.18	26.22	-7.84	26.58	12.14

4.4.2 Regression models

Table 9 shows the means of wSSE values (and their standard deviations in the brackets) on validation datasets and test datasets and Table 10 shows the results of T-tests carried out for comparing the performance of the four regression models. We can see that wSVR outperforms all of the other methods on both validation and test datasets.

Table 9

Means (standard deviations) of the outcomes of the four regression models on the industrial dataset.

Dataset	MLP	RBF	SVR	wSVR
validation dataset	0.5874(2.37E-05)	0.5898(2.37E-05)	0.5726(1.18E-6)	0.5678(7.57E-07)
test dataset	0.7858(0.004)	0.5417(2.22E-12)	0.5215(1.21E-05)	0.5186(3.02E-06)

Table 10

Comparison of the four regression models on the industrial dataset: t-values.

Dataset	MLP vs.	MLP vs.	MLP vs.	RBF vs.	RBF vs.	SVR vs.
	RBF	SVR	wSVR	SVR	wSVR	wSVR
validation dataset	-2.69	16.30	21.78	86.89	138.83	18.96
test dataset	21.01	22.72	22.30	31.79	72.86	4.09

4.4.3 Comparison of the prediction performance between the forecasting models

It is interesting and important to compare the performance of the time series models and the regression models, based on which we can select the optimally performed model. We plot box-plots of the distributions of the wSSE from both the time series models and the regression models on the industrial dataset in Figure 1. Then, on both validation datasets (see Figure 1(a)) and test datasets (see Figure 1(b)), we can see the prediction performance of the regression models outperform that of the times series models. This implies that the prediction performance of the models with original warranty claims as model's dependent variables outperform that of those models with claim rates/repair rates as model's dependent variables.

From these figures, we can also see that the MLP regression models perform similarly to the other models on the validation datasets, but have poor prediction ability on test datasets.

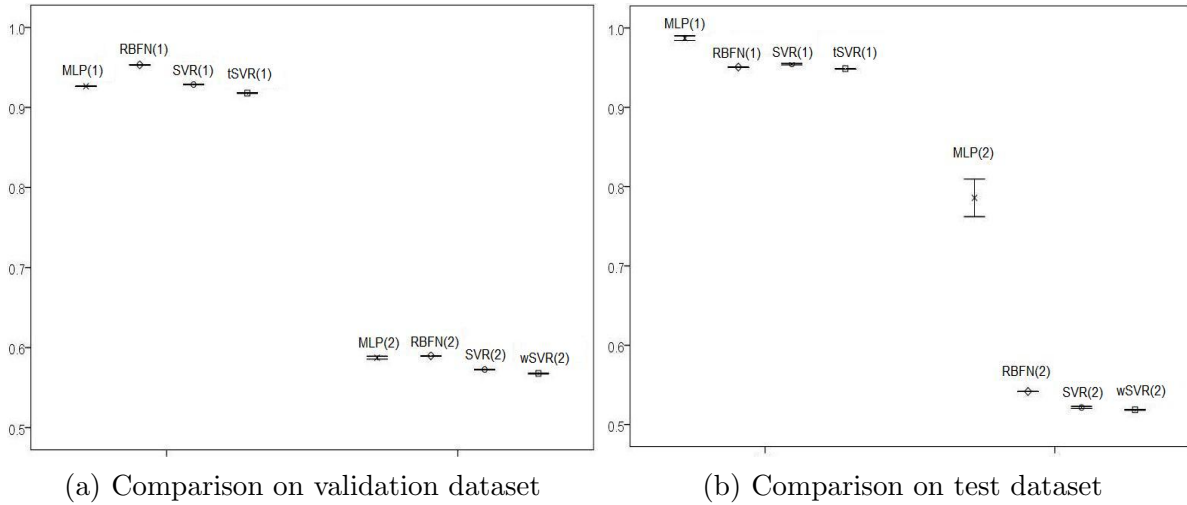


Fig. 1. Comparison between forecasting models 1 and 2. (The labels ended with (1) are time series models, and those labels ended with (2) are regression models.)

5 Conclusions

This paper has introduced two new approaches to forecasting warranty claims. We have also compared the performance of five modelling techniques, namely, multilayer perceptrons (MLP), radial basis function networks (RBFN), support vector regressions (SVR), weighted SVR-based time series models (tSVR), and weighted SVR regression models (wSVR), on two warranty claim datasets. The main findings of this paper, based on the two warranty claim data collected from two different industries, are

- the prediction performance of the weighted SVR-based time series models outperform that of MLP, RBFN and SVR;
- the prediction performance of the weighted SVR regression models outperform that of MLP, RBFN and SVR;

- the prediction performance of the models with original warranty claims as model's dependent variables outperforms that of the models with claim rates/repair rates as model's dependent variables.
- the main reason why the weighted methods have better prediction accuracy is that the most recent data is given more weights than earlier data.

This paper has only considered the problem of six-month ahead forecasting for warranty claims. From a perspective of warranty suppliers/manufacturers, long-term forecasting (such as 24 months, 36 months, even longer) of warranty claims is vitally important in their fiscal plans. Our future research will focus on long-term forecasting for warranty claims.

Acknowledgement

We are grateful to the reviewers for their helpful comments that lead improvement on the paper. This research is supported by Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (EPSRC Grant reference: EP/G039674/1).

References

- [1] J. Bai, H. Pham, Cost analysis on renewable full-service warranties for multi-component systems, *European Journal of Operational Research* 168 (2006) 492–508.
- [2] S. Wu, H. Li, Warranty cost analysis for products with a dormant state, *European Journal of Operational Research* 182 (3) (2007) 1285–1293.
- [3] S. Wu, M. Xie, Warranty cost analysis for nonrepairable services products, *International Journal of Systems Science* 39 (3) (2008) 279–288.
- [4] B. Kim, S. Park, Optimal pricing, EOL (end of life) warranty, and spare parts manufacturing strategy amid product transition, *European Journal of Operational Research* 188 (3) (2008) 723–745.
- [5] W. M. Yeo, X. M. Yuan, Optimal warranty policies for systems with imperfect repair, *European Journal of Operational Research* 199 (1) (2009) 187–197.
- [6] S. Wu, P. Longhurst, Optimising age-replacement and extended non-renewing warranty policies in lifecycle costing, *International Journal of Production Economics* 130 (2) 262–267.
- [7] D. N. P. Murthy, I. Djameludin, New product warranty: A literature review, *International Journal of Production Economics* 79 (3) (2002) 231–260.
- [8] J. D. Kalbfleisch, J. F. Lawless, J. A. Robinson, Methods for the analysis and prediction of warranty claims, *Technometrics* 33 (3) (1991) 273–285.
- [9] A. Kleyner, P. Sandborn, A warranty forecasting model based on piecewise statistical distributions and stochastic simulation, *Reliability Engineering and System Safety* 88 (3) (2005) 207–214.

- [10] B. Rai, Warranty spend forecasting for subsystem failures influenced by calendar month seasonality, *IEEE Transactions on Reliability* 48 (4) (2009) 649–657.
- [11] G. S. Wasserman, An application of dynamic linear models for predicting warranty claims, *Computers and Industrial Engineering* 22 (1) (1992) 37–47.
- [12] B. Rai, N. Singh, Forecasting warranty performance in the presence of the 'maturing data' phenomenon, *International Journal of Systems Science* 36 (7) (2005) 381–394.
- [13] T. Hrycej, M. Grabert, Warranty cost forecast based on car failure data, in: *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2007, Celebrating 20 years of neural networks, Orlando, Florida, USA, August 12-17, 2007*, 2007, pp. 108–113.
- [14] N. D. Singpurwalla, S. Wilson, Warranty problem: its statistical and game theoretic aspects, *SIAM Review* 35 (1) (1993) 17–42.
- [15] M. P. Kaminskiy, V. V. Krivtsov, G-renewal process as a model for statistical warranty claim prediction, *Proceedings of the Annual Reliability and Maintainability Symposium* (2000) 276–280.
- [16] D. Stephens, M. Crowder, Bayesian analysis of discrete time warranty data, *Journal of the Royal Statistical Society Series C* 53 (1) (2004) 195–217.
- [17] K. Majeske, A non-homogeneous poisson process predictive model for automobile warranty claims, *Reliability Engineering and System Safety* 92 (2) (2007) 243–251.
- [18] M. Fredette, J. F. Lawless, Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims, *Technometrics* 49 (1) (2007) 66–80.
- [19] G. S. Wasserman, A. Sudjianto, A comparison of three strategies for forecasting warranty claims, *IIE Transactions (Institute of Industrial Engineers)* 28 (12) (1996) 967–977.
- [20] J. Chen, N. J. Lynn, N. D. Singpurwalla, Forecasting warranty claims, chapter 31, *Product Warranty Handbook*, Marcel Dekker, 1996, pp. 803–816.
- [21] W. Blischke, D. Murphy, *Warranty cost analysis*, Marcel Dekker, New York, 1992.
- [22] S. Wu, Warranty claim analysis considering human factors, *Reliability Engineering and System Safety* 96 (2011) 131–138.
- [23] M. Kijima, N. Sumita, A useful generalization of renewal theory: Counting process governed by non-negative markovian increments, *Journal of Applied Probability* 23 (1986) 71–88.
- [24] V. Vapnik, *The nature of statistical learning theory*, Springer, New York, 1995.
- [25] A. Smola, B. Scholkopf, A tutorial on support vector regression, *Statistics and Computing* 14 (3) (2004) 199–222.
- [26] N. K. Ahmed, A. F. Atiya, N. E. Gayar, H. El-Shishiny, An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews* 29 (5-6) (2010) 594–621.
- [27] P. Lingras, C. Butz, Rough support vector regression, *European Journal of Operational Research* 206 (2) (2010) 445–455.
- [28] R. Carbonneau, K. Laframboise, R. Vahidov, Application of machine learning techniques for supply chain demand forecasting, *European Journal of Operational Research* 184 (3) (2008) 1140–1154.

- [29] T. Trafalis, R. Gilbert, Robust classification and regression using support vector machines, *European Journal of Operational Research* 173 (3) (2006) 893–909.
- [30] L. Cao, F. Tay, Support vector machine with adaptive parameters in financial time series forecasting, *IEEE Transactions on Neural Networks* 14 (6) (2003) 1506–1518.
- [31] M. Vose, *The Simple Genetic Algorithm: Foundations and Theory*, MIT, 1999.