Support vector regression with asymmetric loss for optimal electric load forecasting

Wu, Ryan, Wang, You-Gan, Tian, Yu-Chu, Burrage, Kevin and Cao, Taoyun

Support vector regression with asymmetric loss for optimal electric load forecasting

Jinran Wu [a,b], You-Gan Wang [*a,b], Yu-Chu Tian [c], Kevin Burrage [a,b,d] and Taoyun Cao [e]

[a]**School of Mathematical Sciences, Queensland University of Technology, Australia**

[b]**ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Australia**

[c]**School of Computer Science, Queensland University of Technology, Australia**

[d]**Department of Computer Science, University of Oxford, UK**

[e]**School of Statistics and Mathematics, Guangdong University of Finance and Economics, China**

Abstract

In energy demand forecasting, the objective function is often symmetric, implying that over-prediction errors and under-prediction errors have the same consequences. In practice, these two types of errors generally incur very different costs. To accommodate this, we propose a machine learning algorithm with a cost-oriented asymmetric loss function in the training procedure. Specifically, we develop a new support vector regression incorporating a linear-linear cost function and the insensitivity parameter for sufficient fitting. The electric load data from the state of New South Wales in Australia is used to show the superiority of our proposed framework. Compared with the basic support vector regression, our new asymmetric support vector regression framework for multi-step load forecasting results in a daily economic cost reduction ranging from 42.19% to 57.39%, depending on the actual cost ratio of the two types of errors.

Keywords: Asymmetric loss; Cost-orientation; Machine learning; Statistical modeling; Load forecasting.

[*]To whom all correspondence should be directed, *Email: you-gan.wang@qut.edu.au*

1 Introduction

Electric load forecasting is essential in energy system administration. Optimized objectives in machine learning for electric load forecasting are often symmetric, and the target is to predict the electric load as accurately as possible. However, the prediction of the electric load is extremely challenging due to the non-linear, dynamic, and complex system in demand [1]. As a result, the errors in real forecasting are hard to avoid. Moreover, in real applications, the penalties from under-predictions and over-predictions are very different based on the corresponding financial costs [2]. Apparently, predictions with symmetric loss will result in more unnecessary economic costs, although unbiased predictions can be obtained, and biased predictions considering different real costs are more reasonable.

However, symmetric loss is popularly used to obtain higher accurate load forecasting, such as with a neural network and its variations [3–5] and support vector regression (SVR) with its developments [6–8]. Even now, to handle these complex systems more efficiently for load forecasting, deep learning has been introduced, and its optimization objective is symmetric, such as mean square error for recurrent neural networks [9]. There is also a variety of different symmetric loss functions proposed in the literature, motivated by objectives in estimation efficiency, robustness to outliers, and computational convenience. Novel examples include ϵ-Laplace error [10] and exponential squared loss [11].

Interestingly, the concept of "asymmetric" loss has been introduced before in machine learning. However, in "asymmetric" machine learning, different distributions from positive errors and negative errors are considered to develop more accurate regressions [12–16]. [17] proposed an asymmetric and quadratic loss function for SVR to accurately predict power usage. [18] presented an asymmetric ν-twin SVR based on the pinball loss function, which can enhance generalization ability by controlling the fitting error. [19] incorporated asymmetric Huber and ϵ-insensitive Huber loss functions into SVR to avoid the interruption from asymmetric noise and outliers. To summarize, the motivation for the investigated asymmetric SVR frameworks is to improve the forecasting accuracy instead of minimizing the real cost.

As recommended by [20], penalties from over-predictions and under-predictions should be distinguished, so symmetric loss and "asymmetric" loss for highly accurate load forecasting are not rational in real applications. Thus, the forecasting model with real asymmetric loss is recommended based on the real economic cost minimization for electric load forecasting [21]. Asymmetric loss considering different outcomes is more rational. So far, there has been limited work on machine learning with asymmetric loss. Linear-linear cost is one of the asymmetric loss functions that is applied to two parameters to determine the severity of the prediction type (over-prediction or under-prediction). It has been used in neural networks training to minimize costs in some specific applications. [22] introduced linear-linear cost in neural networks to predict optimum service levels in inventory management. [23] developed a neural network with linear-linear cost for resale price forecasting to aid with pricing decisions. A complex linear-linear cost was developed by [2] for tree regression training, where the prediction was divided into four types with different severities.

As reviewed above, most current electric load forecasting systems are designed for error minimization with symmetric loss. Symmetric loss ignores the difference in penalties from over-prediction and under-prediction in load forecasting. A biased prediction with asymmetric loss is more appropriate in load forecasting. This means a cost-oriented asymmetric loss is demanded by considering the real penalties for over-prediction and under-prediction. Furthermore, due to the solid theoretical foundation of SVR, a cost-oriented asymmetric SVR framework (AsySVR) is more promising for load forecasting. Therefore, this is motivated by the following three considerations: (a) an economic cost is more rational as the objective function for electric load forecasting, where different economic penalties for over- and under-predictions as well as the generalization of the forecasting system should be considered; (b) a new design and corresponding training procedure for a cost-oriented AsySVR framework are required to show the performance of the proposed economic cost; and (c) the effectiveness of the proposed AsySVR framework should be validated by a load forecasting project.

Therefore, we incorporate a cost-oriented asymmetric loss into SVR, and develop a novel AsySVR framework to reduce the economic costs of electric load forecasting. Additionally, in order to overcome the over-fitting in AsySVR framework training, an insensitive linear-linear cost is designed that ignores the smaller errors to improve its generalization. We also show that the asymmetric least absolute value regression, as a special AsySVR framework, is equivalent to quantile regression. To apply our theoretical discussion, we focus on the state of New South Wales load forecasting project to minimize operational costs. To this end, the three main **contributions** of our work are as follows.

(a) An insensitive linear-linear cost (insensitive LLC) is modified based on the LLC, where an insensitive parameter ϵ is incorporated, to overcome the over-fitting for model training with good generalization. In addition, the proposed insensitive LLC considering different real economic penalties from over-predictions and under-predictions can minimize the cost of load forecasting.

(b) A cost-oriented AsySVR framework is developed for electric load forecasting, where the proposed insensitive LLC works as the objective function in the AsySVR training. Moreover, similar to the basic SVR training, the corresponding dual problem for our AsySVR framework is obtained for model training.

(c) In two multi-step electric load forecasting scenarios, the proposed AsySVR framework is more practical and appropriate, as it significantly reduces the daily costs from 42.19% to 57.39%, depending on the actual penalty ratio of two types of predictions.

Additionally, the nomenclature of the paper is listed in Table 1.

The organization of this paper is as follows. Section 2 reviews the LLC, and presents our insensitive LLC; then, our proposed asymmetric framework is illustrated. In Section 3, the performance of our proposed framework is evaluated in two scenarios with the state of New South Wales electric load data. Finally, Section 4 concludes the paper.

2 The proposed asymmetric SVR

To design our SVR with asymmetric loss framework (AsySVR), suppose there are training data $\{(x_i, y_i), i = 1, 2, \ldots, n\}$, and the function $f(\cdot)$ is formulated as

$$f(x) = \langle \omega, x \rangle + b \tag{1}$$

Table 1: Nomenclature

Notation	Description	Notation	Description
x_i	features of the i-th sample	y_i	response of the i-th sample
ω	normal vector	b	threshold
$\hat{y}$	predictions	y	observation
k_1	penalty for over-predictions	k_2	penalty for under-predictions
ϵ	insensitive parameter	ξ	slack variable for under-predictions
C	regularization parameter	L	Lagrange function
ξ^*	slack variable for over-predictions	η_i	Lagrange multiplier for ξ_i
$\langle\cdot,\cdot\rangle$	dot product in a Hilbert space	k	penalty ratio
η_i^*	Lagrange multiplier for ξ_i^*	ρ_τ	the loss function for τ-quantile regression
u	prediction errors	MWh	megawatt-hour
$A\$$	Australia dollar	MDC	mean daily cost
ϱ	proportion of over-predictions	κ	kernel function
α_i^*	Lagrange multiplier for $\langle\omega,x_i\rangle+b-y_i-\epsilon-\xi_i^*$	α_i	Lagrange multiplier for $y_i-\langle\omega,x_i\rangle-b-\epsilon-\xi_i$
τ	quantile of quantile regression	LLC	Linear-linear cost

with the normal vector ω, the threshold b, and the dot product $\langle\cdot,\cdot\rangle$ in a Hilbert space.

2.1 The linear-linear cost

Let $u=\hat{y}-y$ be the prediction error representing over-prediction if $u>0$ or under-prediction if $u<0$. As noted by [24], the cost should represent the actual business objective, such as, the profit maximization. Thus, in our asymmetric loss designs, the LLC is employed with two different penalties k_1 and k_2 for over-predictions and under-predictions, respectively, which can directly quantify the economic costs of each error as [24],

$$l\left(u|k_1,k_2\right)=\begin{cases} k_1u^+, & u>0,\\ k_2u^-, & u\leqslant 0\end{cases} \tag{2}$$

with $u^+=\max\{u,0\}$ and $u^-=\max\{-u,0\}$. Corresponding to the over-prediction and under-prediction, two positive parameters, k_1 and k_2, determine the economic severity of a given error type. Note that $l(u|k_1,k_2)=\max(k_1u^+,k_2u^-)$. Also, we have $k_1>0$, $l(u|k_1,k_2)=k_1l(u|1,k)$, where $k=k_2/k_1$. Therefore, without loss of generality, we can assume $k_1=1$ and let $k=k_2/k_1$ for convenience as

$$l\left(u|k\right)=\begin{cases} u^+, & u>0,\\ ku^-, & u\leqslant 0.\end{cases} \tag{3}$$

Notice two penalty parameters, k_1 and k_2, are given based on economic cost calculations. Thus, as visualized in Fig. 1, the parameter k shows the bias on two types of errors. Obviously, if k is larger than 1, under-prediction brings a higher cost than over-prediction. Otherwise, over-prediction is more costly.

2.2 The linear-linear cost with an insensitive parameter

To overcome over-fitting of the model with LLC, according to good performance of the insensitive Laplace loss by [10], an insensitive LLC is developed with an insensitive parameter ϵ, that can ignore the smaller

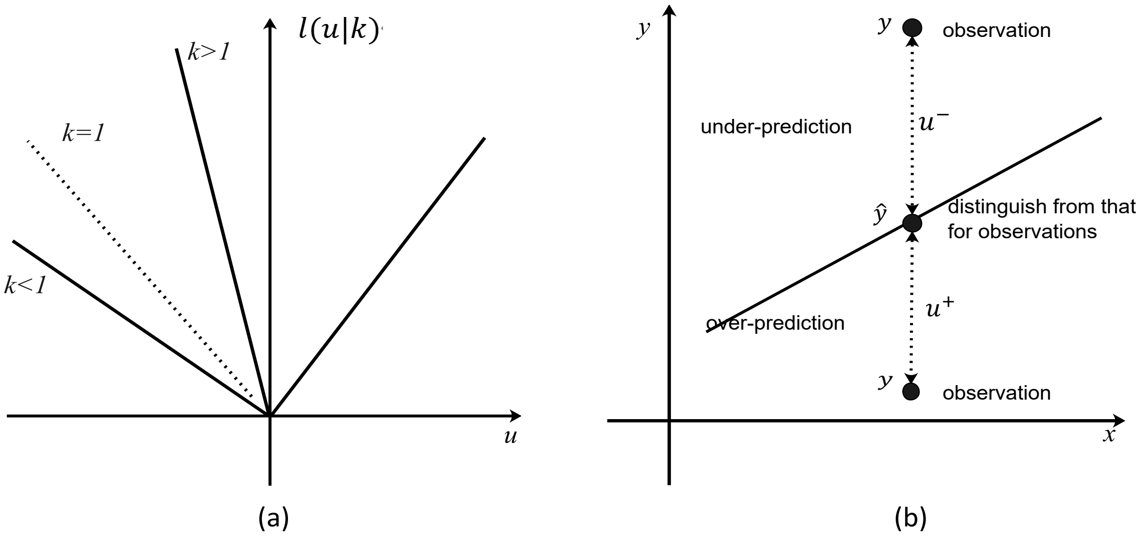


Figure 1: The linear-linear cost: (a) the proposed linear-linear cost, and (b) the illustration of prediction types.

errors ($|u| \leqslant \epsilon$). The proposed insensitive LLC is formulated as

$$l\left(u|k\right)=\begin{cases} u^{+}-\epsilon, & u>\epsilon, \\ 0, & |u|\leqslant\epsilon, \\ k\left(u^{-}-\epsilon\right), & u<-\epsilon. \end{cases} \tag{4}$$

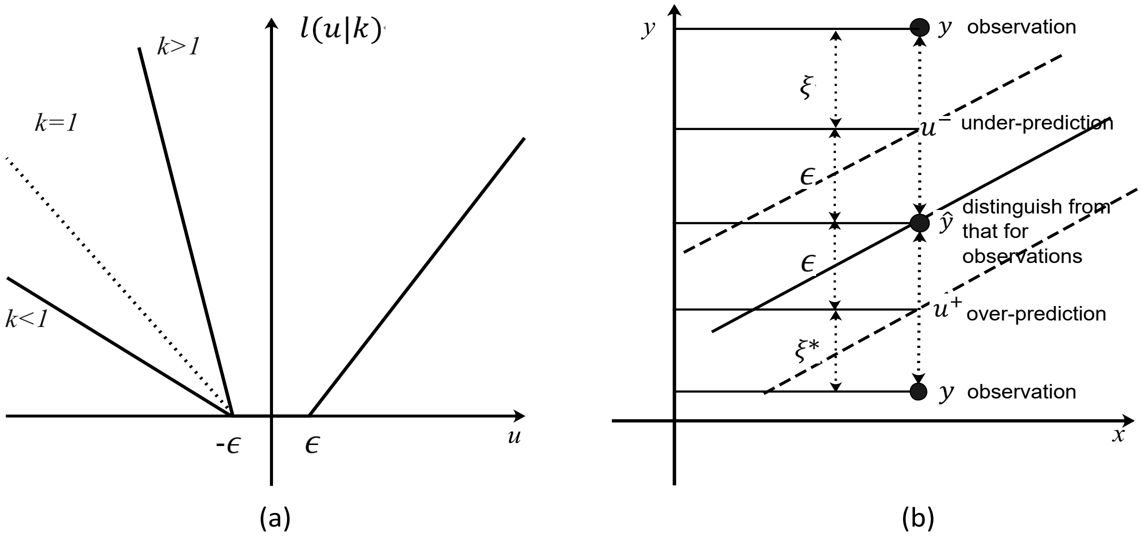


Figure 2: The proposed insensitive linear-linear cost: (a) the proposed insensitive linear-linear cost, (b) the illustration of prediction types.

Different from Fig. 1, two soft margins are introduced in our insensitive LLC. As shown in Fig. 2, only the larger prediction errors located out of margins are counted to calculate the cost. The slack variables for the under-prediction and over-prediction are defined as $\xi = u^- - \epsilon$ and $\xi^* = u^+ - \epsilon$, respectively. More specifically, the proposed insensitive LLC can guarantee the model's fitting, especially in complex networks.

2.3 The new support vector regression

Now considering the optimized loss in ϵ-SVR [25], we incorporate the insensitive LLC, Eq. (4), in the SVR structure to develop an AsySVR. Moreover, the corresponding convex optimization problem (primal objective function) with slack variables ξ_i and ξ_i^* for the cost-oriented asymmetric framework is formulated as

$$\begin{aligned} &\min_{\omega,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\left(\|\omega\|^2\right) + C\sum_{i=1}^{n}(k\xi_i + \xi_i^*) \\ &\text{s.t.} \quad \begin{cases} \langle \omega, x_i \rangle + b - y_i \leqslant \epsilon + \xi_i, \\ y_i - \langle \omega, x_i \rangle - b \leqslant \epsilon + \xi_i^*, \\ \xi_i \geqslant 0,\ \xi_i^* \geqslant 0,\ i = 1, 2, \ldots, n \end{cases} \end{aligned} \tag{5}$$

with a regularization parameter C. Here, the optimized objective is designed with cost-oriented penalties. In our case, we consider a simple condition, under which over-prediction ($u > \epsilon$) will be punished with k_1 for each unit loss, while under-prediction ($u < -\epsilon$) will be punished with k_2 for each unit loss. The weight k is calculated by two penalties from a real scenario. Apparently, according to different real practices, the optimized objective can be generalized to L_2 regularization with a specific piecewise function. It should be noted that conventional asymmetric cost functions are for accurate regressions to address noises or outliers in the dataset.

Next, in order to reduce the complexity of the primal objective optimization, a dual problem is recommended where dimensionality depends only on the number of support vectors. Thus, a Lagrange function is constructed from the primal objective function and the corresponding constraints with a set of variables as

$$\begin{aligned} L := &\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(k\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ &+ \sum_{i=1}^{n}\alpha_i\left(y_i - \langle \omega, x_i \rangle - b - \epsilon - \xi_i\right) \\ &+ \sum_{i=1}^{n}\alpha_i^*\left(\langle \omega, x_i \rangle + b - y_i - \epsilon - \xi_i^*\right) \end{aligned} \tag{6}$$

where $\eta_i, \eta_i^*, \alpha_i$, and α_i^* are Lagrange multipliers.

Then, the saddle point condition can be calculated by the zero of the partial derivatives of L with respect to the primal variables $(\omega, b, \xi_i, \xi_i^*)$ as

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0, \tag{7}$$

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\, x_i = 0, \tag{8}$$

$$\frac{\partial L}{\partial \xi_i} = kC - \alpha_i - \eta_i = 0, \tag{9}$$

and

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0. \tag{10}$$

After that, the corresponding dual problem can be obtained as

$$
\begin{aligned}
\max_{\alpha,\alpha^*} \; & \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) \\
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i^T x_j \\
\text{s.t.} \; & \begin{cases} \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0, \\ 0 \leqslant \alpha_i \leqslant kC, \\ 0 \leqslant \alpha_i^* \leqslant C, \end{cases}
\end{aligned}
\tag{11}
$$

which is required to meet the Karush–Kuhn–Tucker (KKT) condition [26]

$$
\begin{cases}
\alpha_i \left(y_i - \langle \omega, x_i \rangle - b - \epsilon - \xi_i\right) = 0, \\
\alpha_i^* \left(\langle \omega, x_i \rangle + b - y_i - \epsilon - \xi_i^*\right) = 0, \\
\alpha_i \alpha_i^* = 0, \; \xi_i \xi_i^* = 0, \\
(kC - \alpha_i)\, \xi_i = 0, \; (C - \alpha_i^*)\, \xi_i^* = 0.
\end{cases}
\tag{12}
$$

Finally, substitute Eq. (8) for Eq. (1), and our proposed asymmetric SVR framework with the Lagrange multipliers α_i and α_i^* can be estimated as

$$
f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{13}
$$

where b can be estimated by the KKT condition. Here, similar to the basic SVR, our corresponding dual problem can be solved by quadratic programming [27] or sequential minimal optimization [28]. It should be mentioned that the proposed framework for training large sample sets can process within the allowed computing resources. Our computation requirement is as same as the original ϵ-SVR, thus the scalability is unchanged, and any efficient training procedures in SVR can be adopted to our framework. For example, sequential minimal optimization can solve the proposed dual problem for training large sample sets following the search strategies by [29].

Now, the procedure of our proposed AsySVR framework can be given as follows:

Step 1 Calculate the penalty ratio k using the different penalties k_1 and k_2 from a real scenario.

Step 2 Obtain the corresponding specific dual problem Eq. (11) by substituting the penalty ratio k.

Step 3 Input training data.

Step 4 Solve the dual problem to achieve α_i, α_i^* and b with two proper hyper-parameters, C and ϵ.

Step 5 Predict the test data using Eq. (13) .

For a clearer illustration, the flowchart of our proposed AsySVR framework is displayed in Fig. 3.

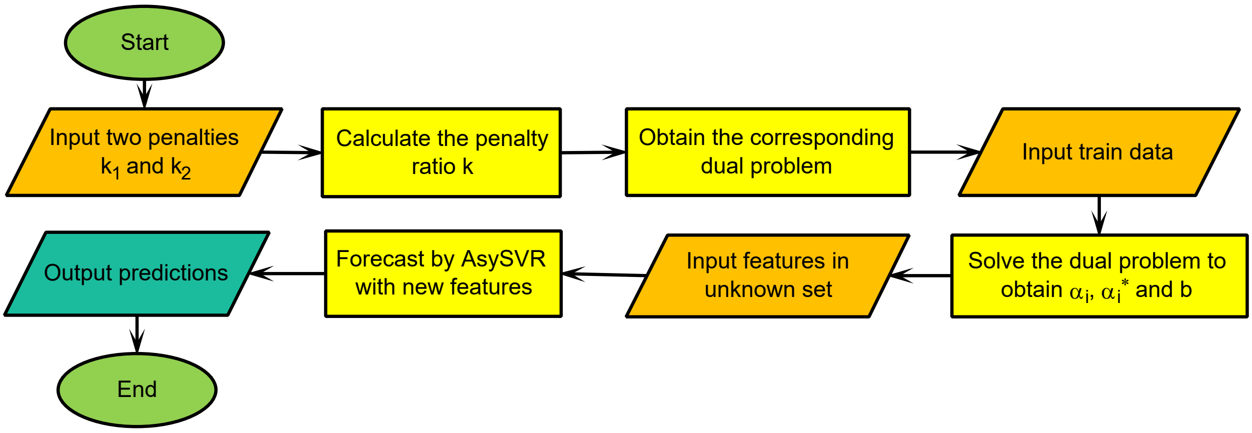


Figure 3: The flowchart of the proposed AsySVR framework

2.4 Connection to quantile regression

In the special case of the asymmetric SVR framework with $\epsilon = 0$ and $C \longrightarrow +\infty$, the optimized objective function (Eq. (5)) can become equivalent to

$$\mathcal{L} = \sum_{i=1}^{n} l(u_i|k). \tag{14}$$

Note that $l(u_i|k)$ is equivalent to (up to a constant $\tau = k/(1+k)$)

$$\rho_\tau = \begin{cases} \tau u^-, & u < 0, \\ (1-\tau)u^+, & u \geqslant 0, \end{cases} \tag{15}$$

which is the optimized objective function for τ-quantile regression, also known as the least absolute value regression under asymmetric loss [30–32]. The case where $k = 1$ or $\tau = 0.5$ corresponds to median regression. Thus, this optimization problem can be solved via quantile regression [33].

3 The case study

In this section, our proposed cost-oriented asymmetric framework, AsySVR, is evaluated by a multi-step New South Wales (NSW) electric load forecasting project.

3.1 Data

The half-hourly electric load from February, 01, 2019 to March, 20, 2019 in NSW is used as experimental data, which was retrieved from the Australia Energy Market Operator (AEMO). As displayed in Fig. 4, the dataset is divided into two groups: a training set ($1,344$ data points) from 01/02/2019 0:30 - 01/03/2019 0:00, and a test set (912 data points) from 01/03/2019 0:30 - 20/03/2019 0:00.

Moreover, according to the report from the World Nuclear Association in 2019, Australia's National Electricity Market (NEM) volume-weighted wholesale price (k_1) was around A\$$82/MWh$ in NSW (here, the Australian dollar is denoted as A\$). Also, the NEW had real-time balancing with the obligation before delivery. Thus, the price (k_2) is highly capped at A\$$14,500/MWh$ (mid-2018) (here, the megawatt-hour is

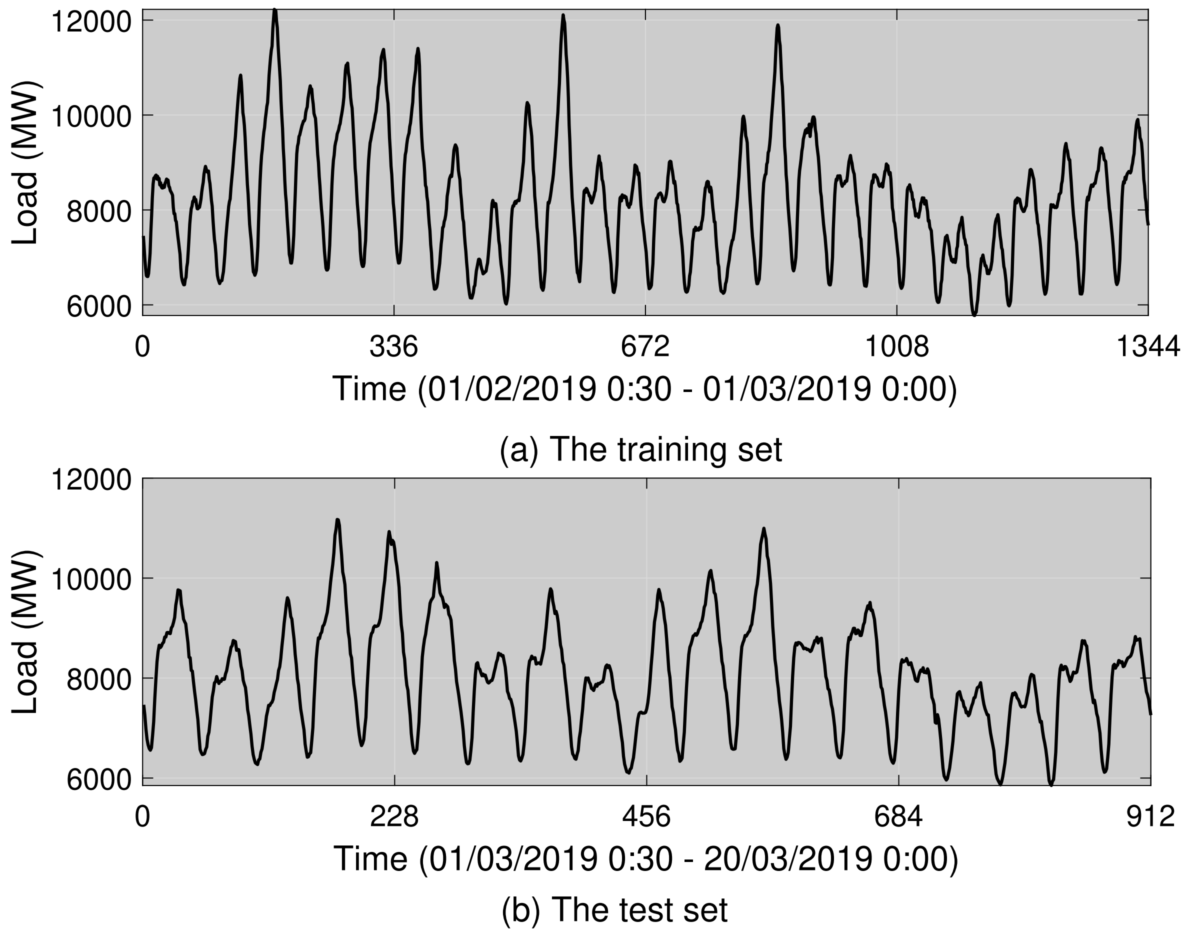


Figure 4: The electric load from NSW: (a) the training data (30 days), and (b) the test data (20 days).

denoted as MWh).

In the Australian electricity market, the electricity retailer buys some electric loads at a spot price and on-sale to the end-use customer. Due to the volatility of the spot price, the retailer and the power generator often sign a hedge contract to decrease financial risks by locking in a fixed price for the electric load in the short term. Therefore, in our study, the whole demand side of NSW is regarded as a retailer. Assume the actual order exactly follows the prediction. Then, in the case of under-predictions, the electricity retailer has to buy extra loads from NEM at a very high price. On the other hand, over-prediction will lead to the disposal of the unused order. As a result, from a economic perspective, the economic penalties for over-predictions and under-predictions are different in the assumed NSW load forecasting project.

3.2 Evaluation criterion

Different from the conventional evaluation criterion, the mean daily cost (MDC, unit: A$) is used to show the effectiveness of our AsySVR framework in two scenarios as

$$\mathrm{MDC} = \frac{48}{n}\sum_{i=1}^{n}\left(k_1 \cdot |u_i| \cdot I(u_i \geqslant 0) + k_2 \cdot |u_i| \cdot I(u_i < 0)\right) \tag{16}$$

where $u_i = \hat{y}_i - y_i$ represents the prediction error, and I is the indicator function. Thus, a smaller MDC value means a more economical daily optimal load prediction.

Furthermore, the proportion ϱ is employed to show the proportion of over-predictions (or the bias of predictions) as

$$\varrho = \frac{1}{n}\sum_{i=1}^{n} I(u_i > 0). \tag{17}$$

From Eq. (17), we see that if ϱ is smaller than 0.5, the forecasting tends toward under-prediction. Otherwise, it tends toward over-prediction.

3.3 Experimental settings

In the multi-step load forecasting project, the proposed AsySVR is targeted at predicting the one-, three-, and five-step ahead loads. As for the input, due to the high correlation with our responses [34,35], the half-hourly loads of the previous day (48 data points) are taken as the main input. In addition, we follow the idea of the intra-day cycles modeling for the temperature factor from [36] and introduce the harmonic functions $\sin(2\pi t/T)$, and $\cos(2\pi t/T)$ (with time order t, and the cycle length $T = 48$ since the half-hourly data is employed in our study), to describe the daily seasonality as the additional inputs. Furthermore, the linear function $\kappa(x_i, x_j) = x_i \cdot x_j'$ is chosen as the kernel, and the proposed AsySVR framework is trained via quadratic programming in our study. Here, to avoid the impact from the scale of the electric load, the data is normalized in our experiments. All experiments were implemented in R 3.6.2 on HPC with ncpus 8 and 32GB memory.

Moreover, the proposed asymmetric framework, AsySVR, is validated in two scenarios as: (1) Scenario 1: $k_1 = 80$, $k_2 = 600$; and (2) Scenario 2: $k_1 = 80$, $k_2 = 900$ (here, the unit is A$/MWh).

3.4 The hyper-parameter selection for AsySVR

As shown in Eq. 5, there are two hyper-parameters (the insensitive parameter ϵ and the regularization parameter C). To tune these hyper-parameters, the 5-fold cross-validation is used for our proposed AsySVR framework. Furthermore, the alternative values for C are set as: 0.1, 1, 10, 100, and 1,000, while these for ϵ are set as: 0, 0.0001, 0.001, 0.01, 0.1, 0.2, and 0.4. These alternative values can be evaluated using cross-validation, in the training set, and the corresponding MDC in the two scenarios are displayed in Fig. 5

According to the results from cross-validation, the optimal hyper-parameters obtained are shown in Table 2, and the whole results are provided in the Supplementary Material.

Table 2: The optimal results for the hyper-parameters by cross-validation: (C, ϵ)

	1-step ahead	3-step ahead	5-step ahead
Scenario 1	$(1, 0.001)$	$(1000, 0)$	$(1, 0.0001)$
Scenario 2	$(10, 0.0001)$	$(1000, 0.0001)$	$(1, 0.001)$

3.5 Experimental results

In the two investigated scenarios, with the optimal hyper-parameter in Table 2, the proposed AsySVR is trained in the total training set for one-, three-, and five-step ahead load forecasting, respectively. Then, the experimental results of the proposed AsySVR framework are plotted in Fig. 6.

As illustrated in Fig. 6, three interesting points can be achieved using our proposed AsySVR framework. The most obvious finding is that the predictions in both scenarios are biased. For example, at $k_2 = 600$, the mean of residual distributions for all one-, three-, and five-step ahead predictions are significantly larger than 0, as shown in Fig. 6a. Another one is that the accuracy of load prediction decreases with the forecasting step increasing. From Fig. 6b, we can see that compared with the predictions for one-step ahead, the five-step ahead forecasting results are far from the block line (which is denoted as the most ideal performance). The last point is that more biases are introduced when the penalty k_2 increases. From Table 3, we can see that in one-step ahead forecasting, the ϱ proportion by AsySVR is 0.90 for Scenario 1, while it is 0.93 for Scenario 2. To explore the underlying mechanism, the proposed AsySVR framework considers the different penalties for over-predictions and under-predictions in model training; as a result, with k_2 increasing, the prediction by the AsySVR tends toward over-prediction, which can significantly reduce the economic cost for load forecasting.

In short, our proposed AsySVR framework with insensitive LLC can effectively reduce the economic cost for load forecasting in a real energy operation.

3.6 Comparative analysis

To show the efficiency of our proposed AsySVR framework, five popular frameworks (least square regression (LS) [37], least absolute value regression (LAV) [38], SVR [6,39], and multilayer perception (MLP) [40,41]) with quantile regression (QUANTILE) are investigated in the two scenarios. The parameter settings for SVR are set as the recommendation in [42] as the regularization parameter 1 and

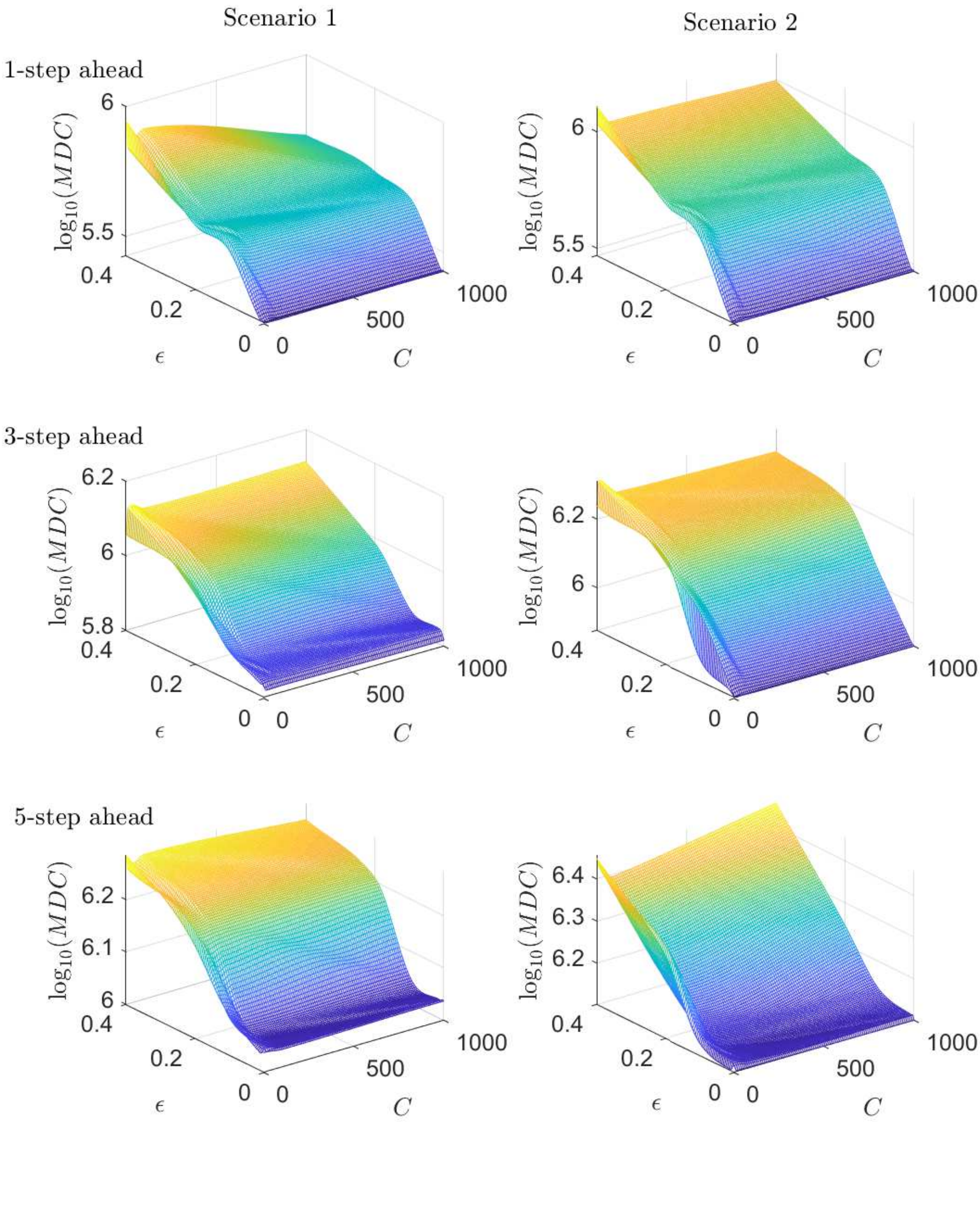
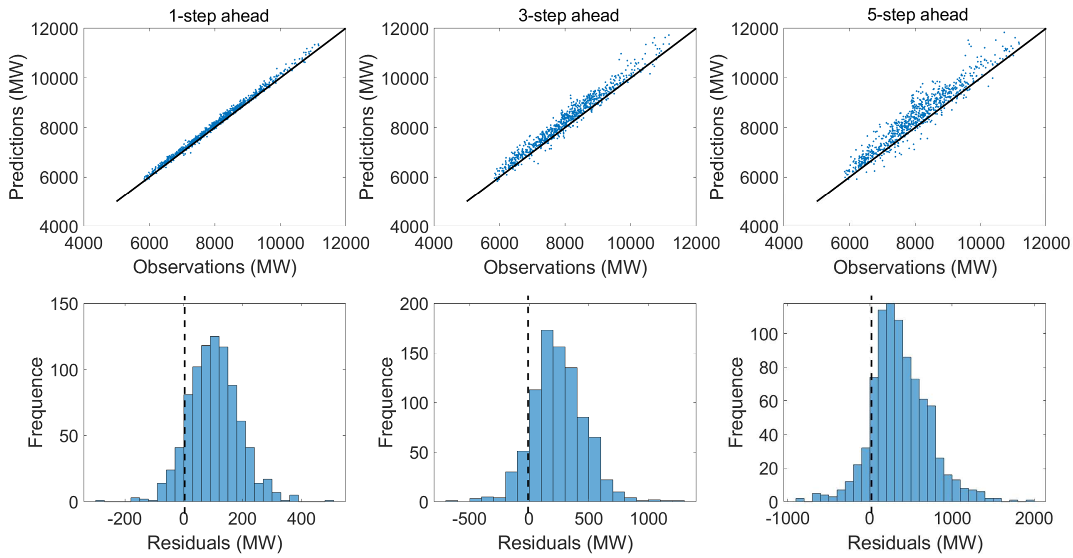
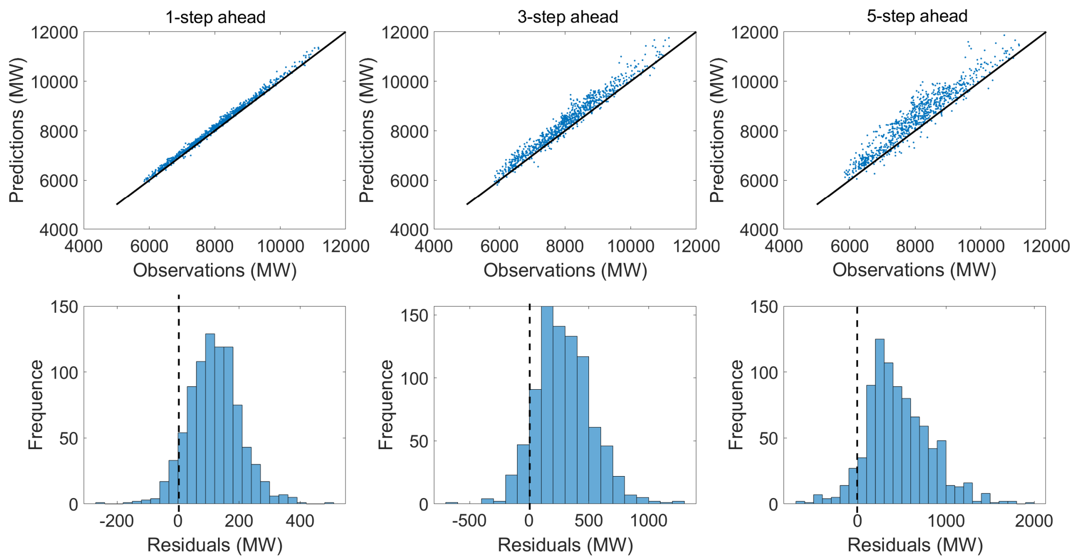


Figure 5: The multi-step MDC by 5-fold cross-validation in the two scenarios.



(a) Scenario 1



(b) Scenario 2

Figure 6: The predictions and corresponding residuals in two scenarios: the black line is denoted as $y_i = y_i$.

the insensitive parameter 0.1, while the structure of MLP is set as 50-20-1 with the logistic activation function by the R Package "neuralnet" [43]. Here, the quantile regression is implemented by using R Package "quantreg" [44]. According to the discussion in Sect. 2.4, the quantiles τ for the two scenarios are calculated as 88.24% and 91.84%, respectively.

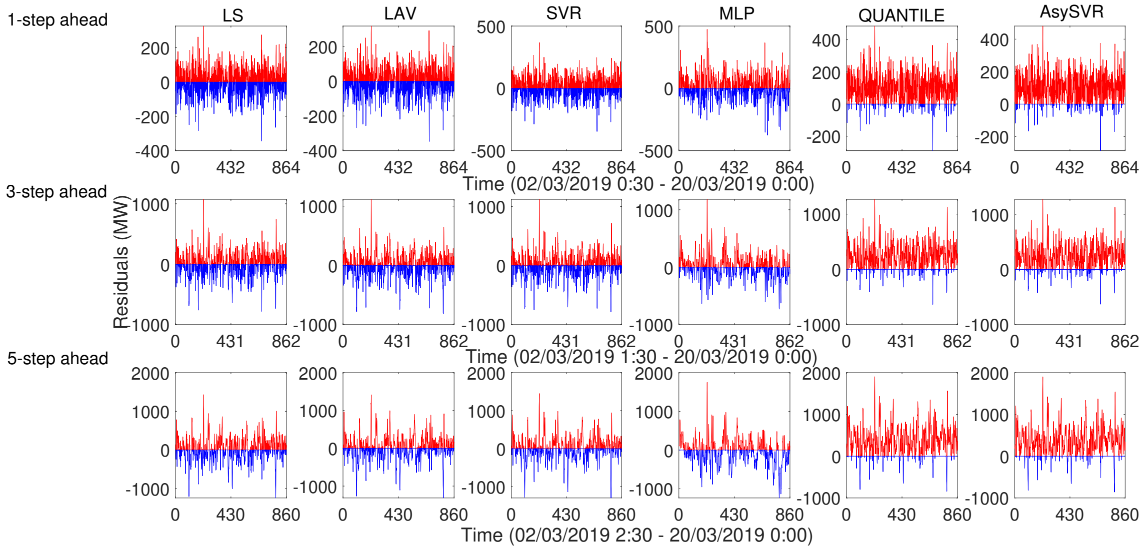
After the implementation for all benchmark frameworks, the residuals and economic costs by four symmetric frameworks and two asymmetric frameworks (QUANTILE and AsySVR) in two scenarios are shown in Fig. 7 and Fig. 8, respectively. Here, it should be mentioned that the optimized objective for the investigated "asymmetric" SVR frameworks is to obtain a highly accurate prediction, and it is very different from our insensitive LLC considering economic costs; however, our proposed cost-oriented AsySVR framework is more effective from an economic perspective. Therefore, their "asymmetric" SVR frameworks are not considered in our comparative analysis.

As shown in Fig. 7 and 8, the bias brought by the proposed AsySVR framework decreases the accuracy of forecasting. Meanwhile, the economic costs for under-predictions and over-predictions are balanced by the bias. This is because different penalties are considered in our asymmetric framework to balance the daily cost for the over-prediction and under-prediction. Apparently, in Scenario 1 ($k_1 = 80$, and $k_2 = 600$), compared to symmetric frameworks, more over-predictions are offered by our proposed asymmetric framework. From the daily cost comparison, the symmetric frameworks balance their loss with more costs for the under-predictions, while the AsySVR framework can give biased predictions with a good economic-cost balance.

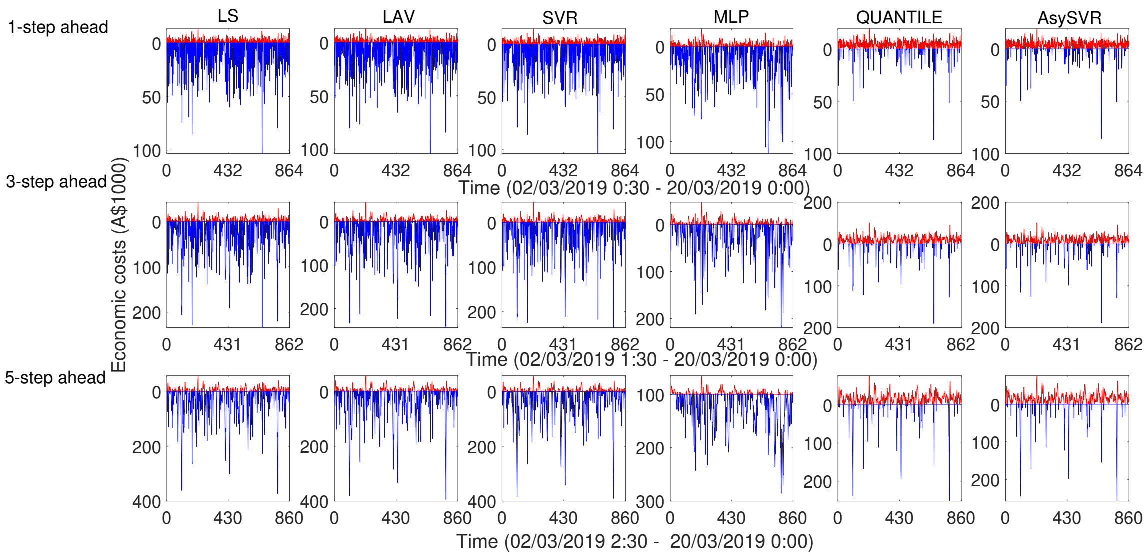
In addition, from the two scenarios, it can be seen that when the penalty k_2 increases, our proposed AsySVR framework is more effective with economic costs. Especially as shown in Fig. 8, it almost forecasts values by AsySVR that are over-predicted to pay less cost on under-predictions.

The multi-step forecasting performance (MDC and ϱ) and the execution time for four symmetric frameworks and two asymmetric frameworks in two scenarios are reported in Table 3. Obviously, according to the MDC index, our proposed AsySVR framework is the most superior in reducing the daily economic costs in electric load forecasting. One of the most obvious cases is Scenario 2 for five-step ahead forecasting, where the MDCs for LS, LAV, SVR, MLP, QUANTILE, and AsyVR, are A\$2,597,415.73, A\$2,592,080.52, A\$2,595,644.69, A\$3,275,893.63, A\$1,187,334.86, and A\$1,185,369.98, respectively. Furthermore, in the case where both ϱ proportions are 0.93 for QUANTILE and AsySVR, those for the symmetric framework ranges from 0.47 to 0.52. This shows our proposed framework can balance the daily costs by adding more biases in predictions. Furthermore, compared with the QUANTILE, the proposed AsySVR framework also obtains good improvements in our study. As shown in Table 3, compared with QUANTILE, $A\$1,014.94 - A\$4,171.78$ of savings in the daily economic cost can be achieved using our AsySVR framework. Particularly, when the step increases, residuals for the load forecasting are larger, and more biases are brought by our AsySVR framework, and the daily economic cost reduces more significantly.

Moreover, our proposed AsySVR framework is also very computationally efficient in the 1,344 historical load data training for the multi-step forecasting. For instance, in Scenario 2, for three-step-ahead forecasting, although the computational cost of the AsySVR (12.63s) is higher than those of LS (0.01s), LAV (0.03s), SVR (0.93s), and QUANTILE (10.66s), our proposed framework is more efficient than MLP (53.57s).

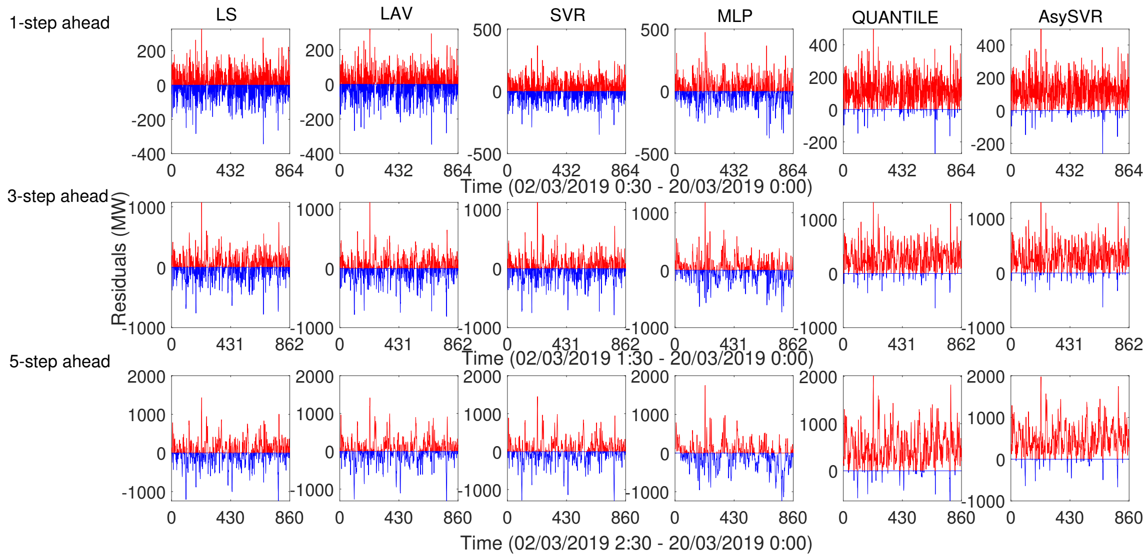


(a) Residuals

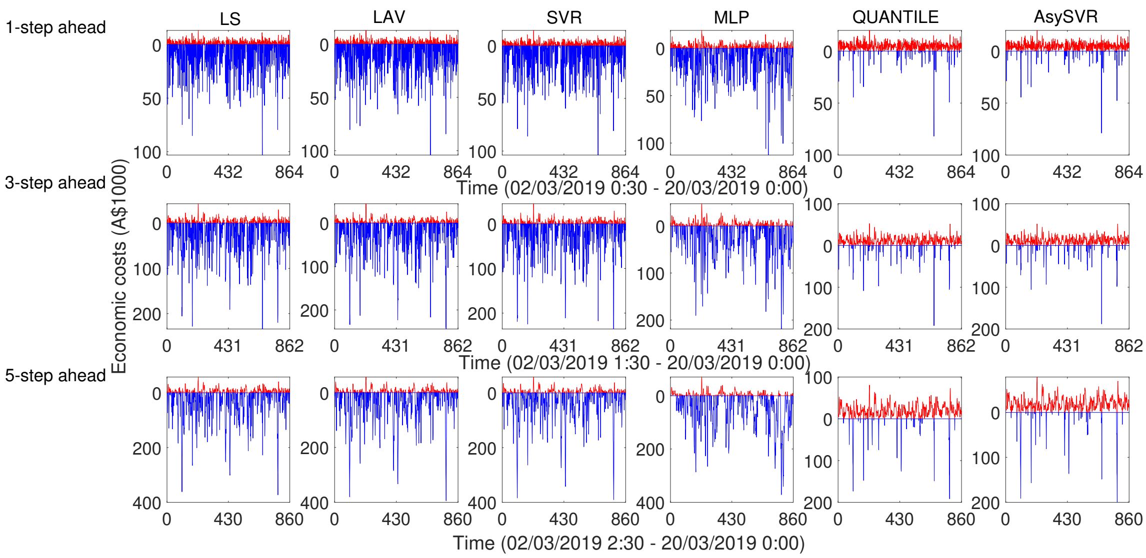


(b) Economic costs

Figure 7: The residuals and corresponding economic costs in Scenario 1. Red color represents over-predictions while blue color represents under-predictions



(a) Residuals



(b) Economic costs

Figure 8: The residuals and corresponding economic costs in Scenario 2. Red color represents over-predictions while blue color represents under-predictions.

Table 3: The MDC, ϱ proportion, and execution time for multi-step load forecasting: (MDC: A$, Time: s)

Panel A: Scenario 1 ($k_1 = 80$ and $k_2 = 600$)									
	1-step ahead			3-step ahead			5-step ahead		
	MDC	ϱ	Time	MDC	ϱ	Time	MDC	ϱ	Time
LS	501,762.58	0.52	0.01	1,229,119.94	0.52	0.01	1,814,764.51	0.52	0.01
LAV	508,926.23	0.52	0.05	1,242,653.32	0.52	0.03	1,806,520.82	0.51	0.03
SVR	490,778.12	0.56	0.34	1,234,895.85	0.52	0.93	1,809,207.12	0.51	1.08
MLP	601,167.73	0.51	38.42	1,210,044.20	0.50	53.57	2,252,860.23	0.47	113.77
QUANTILE	270,722.45	0.89	10.71	686,994.43	0.89	10.66	1,046,921.01	0.90	11.51
AsySVR	267,883.97	0.90	18.16	682,822.22	0.89	12.21	1,045,906.07	0.90	12.60
Panel B: Scenario 2 ($k_1 = 80$ and $k_2 = 900$)									
	1-step ahead			3-step ahead			5-step ahead		
	MDC	ϱ	Time	MDC	ϱ	Time	MDC	ϱ	Time
LS	721,003.86	0.52	0.01	1,762,172.70	0.52	0.01	2,597,415.73	0.52	0.01
LAV	732,502.42	0.52	0.05	1,787,169.87	0.52	0.03	2,592,080.52	0.51	0.03
SVR	698,736.06	0.56	0.34	1,773,109.21	0.52	0.93	2,595,644.69	0.51	1.08
MLP	861,442.16	0.51	38.42	1,739,095.62	0.50	53.57	3,275,893.63	0.47	113.77
QUANTILE	299,656.34	0.93	11.02	768,859.99	0.90	10.66	1,187,334.86	0.93	11.63
AsySVR	297,699.83	0.93	13.62	765,841.96	0.91	12.63	1,185,369.98	0.93	13.52

Additionally, another interesting finding is that the performance of MLP is inferior among all investigated frameworks. The first reason is that there is a strong linear relationship among the investigated load series according to the performance from forecasting benchmark frameworks with symmetric loss, and the MLP is designed to extract the nonlinear relationship. The second is that different from the conventional error index for the accuracy, the new index based on the mean daily cost (MDC) is presented to measure the daily economic cost. Since there are two different economic penalties on the over-prediction and the under-prediction, the forecasting model with the asymmetric loss is superior to the MLP.

To summarize, the proposed AsySVR framework can reduce the daily economic costs by introducing more biases in forecasting results with low computation costs. Meanwhile, under the proposed cost oriented insensitive LLC, the proposed AsySVR framework can obtain the lowest daily costs among all considered frameworks. The source code of our AsySVR framework is available at: https://github.com/wujrtudou/AsySVR_cost_minimisation.git

4 Conclusion

In this study, a cost-oriented asymmetric framework, AsySVR, was developed for electric load forecasting. Based on different penalties for over-predictions and under-predictions, an insensitive LLC was proposed to train the asymmetric SVR framework for real economic cost minimization. In particular, compared with a conventional SVR framework, in our NSW load forecasting project, the proposed asymmetric framework achieved excellent daily economic cost reductions with A\$222,894.15 (one-step ahead), A\$552,073.63 (three-step ahead), and A\$763,301,05 (five-step ahead) in Scenario 1, and A\$401,036.23 (one-step ahead), A\$1,007,267.25 (three-step ahead), and A\$1,410,274.71 (five-step ahead) in Scenario 2. With the step increases, the residuals increased, and the daily economic cost reductions were more significant. Moreover an insensitive LLC was designed for our AsySVR framework, where an insensitive parameter ϵ is used to address the over-fitting. From an economic cost comparison, it can be seen that the AsySVR framework with the insensitive LLC was the most superior for economic cost reduction in both scenarios. Additionally, we also show that the proposed AsySVR framework was essentially the same as the quantile regression. In brief, our proposed AsySVR framework is an efficient and promising tool for electric load forecasting.

The electricity demand data inevitably often contains Gaussian noises and heavy-tailed non-Gaussian noises. The distribution of the noises does not affect the validity of our framework but the extent of improvement does depend on the distribution. More economic benefit may be obtained when the data follows heavy tailed distributions implying more serious errors in predictions.The distribution of the noise affects the parameter estimation, and careful modeling of the noise distribution can improve the parameter estimation with limited impact on the overall loss function. More theoretic work can be carried out in this direction. Our key contribution in the proposed framework is to minimize the economic loss instead of the conventional statistical objective function. In addition, the proposed framework is a regularized forecasting approach, where the regularizer is the squared norm of the estimating function in some reproducing kernel Hilbert space. In the linear case, according to Theorem 1 of [45], the asymptotic results hold. Much more theoretic work can be done from a statistical perspective following the work by [46], for example. Fast convergence in computation when the training size is large is another big data issue worth research attention.

There are still some limitations in our work. The first limitation is that our electric load forecasting is based on time series modeling, ignoring some environmental factors, such as temperature and humidity. Furthermore, in our AsySVR framework, the cross-validation was used to select the regularization parameter C and the insensitive parameter ϵ, which selects the parameters from the alternative values with large computational costs for the AsySVR training. Additionally, the penalties for over-predictions and under-predictions should be obtained (or forecasted) to improve the effectiveness of a symmetric SVR framework.

In the future, there are many research directions. For electric load forecasting, more economical and environmental factors can be incorporated to establish a highly accurate forecasting system, promoting the economic cost reduction. Moreover, our proposed asymmetric frameworks can be extended to other economic cost minimizations in operational management, such as the renewable energy bidding. In addition, some advanced approaches can be developed for the parameter selection in our AsySVR framework. Furthermore, our proposed insensitive LLC can be extended to other state-of-the-art methods, such as recurrent neural network architecture, for big data forecasting. Finally, since the load series is temporal data, some incremental learning methods [47] can be designed for our proposed AsySVR framework to speed up the computation for large-scale data. In addition, parallel computing [48] can be implemented for our new framework in big data training.

CRediT authorship contribution statement

Jinran Wu: Conceptualization, Investigation, Methodology, Validation, Formal analysis, Visualization, Software, Writing-original draft. **You-Gan Wang:** Supervision, Funding acquisition, Project administration, Methodology, Writing-review & editing. **Yu-Chu Tian:** Supervision. **Kevin Burrage:** Writing-review & editing. **Taoyun Cao:** Writing-review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the three reviewers for their constructive comments and suggestions, which have led to a much improved paper. Computational (and/or data visualization) resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia. This work is in part supported by Australian Research Council (ARC) Discovery Project (DP160104292), the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), under grant number CE140100049, Guangdong Basic and Applied Basic Research Foundation (2020A1515011580) and Provincial key platforms and major scientific research projects of Guangdong universities (2018GKTSCX010).

References

[1] Youlong Yang, Jinxing Che, Chengzhi Deng, and Li Li. Sequential grid approach based support vector regression for short-term electric load forecasting. *Applied Energy*, 238:1010–1021, 2019.

[2] Guo Li and Hsiao-Dong Chiang. Toward cost-oriented forecasting of wind power generation. *IEEE Transactions on Smart Grid*, 9(4):2508–2517, 2016.

[3] Ying Chen, Peter B Luh, Che Guan, Yige Zhao, Laurent D Michel, Matthew A Coolbeth, Peter B Friedland, and Stephen J Rourke. Short-term load forecasting: Similar day-based wavelet neural networks. *IEEE Transactions on Power Systems*, 25(1):322–330, 2009.

[4] AS Khwaja, X Zhang, A Anpalagan, and B Venkatesh. Boosted neural networks for improved short-term electric load forecasting. *Electric Power Systems Research*, 143:431–437, 2017.

[5] Jinran Wu, Zhesen Cui, Yanyan Chen, Demeng Kong, and You-Gan Wang. A new hybrid model to predict the electrical load in five states of australia. *Energy*, 166:598–609, 2019.

[6] Ervin Ceperic, Vladimir Ceperic, and Adrijan Baric. A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems*, 28(4):4356–4364, 2013.

[7] Ehab E Elattar, John Goulermas, and Q Henry Wu. Electric load forecasting based on locally weighted support vector regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(4):438–447, 2010.

[8] Hai Zhong, Jiajun Wang, Hongjie Jia, Yunfei Mu, and Shilei Lv. Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*, 242:403–414, 2019.

[9] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2017.

[10] Vladimir Vapnik. *The nature of statistical learning theory.* Springer science & business media, 2013.

[11] Xueqin Wang, Yunlu Jiang, Mian Huang, and Heping Zhang. Robust variable selection with exponential squared loss. *Journal of the American Statistical Association*, 108(502):632–643, 2013.

[12] Shao-Qing Xu, Qiang-Yi Luo, Guang-Hui Xu, and Lei Zhang. Asymmetrical interval regression using extended ε-svm with robust algorithm. *Fuzzy Sets and Systems*, 160(7):988–1002, 2009.

[13] Xinjun Peng. Tsvr: an efficient twin support vector machine for regression. *Neural Networks*, 23(3):365–372, 2010.

[14] Kyung Ha Seok, Daehyeon Cho, Changha Hwang, and Jooyong Shim. Support vector quantile regression using asymmetric e-insensitive loss function. In *2010 2nd International Conference on Education Technology and Computer*, volume 1, pages V1–438. IEEE, 2010.

[15] Changha Hwang. Asymmetric least squares regression estimation using weighted least squares support vector machine. *Journal of the Korean Data & Information Science Society*, 22:999–1005, 2011.

[16] Mel Stockman, Randa S El Ramli, Mariette Awad, and Rabih Jabr. An asymmetrical and quadratic support vector regression loss function for beirut short term load forecast. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 651–656. IEEE, 2012.

[17] Xiaolin Huang, Lei Shi, Kristiaan Pelckmans, and Johan AK Suykens. Asymmetric ν-tube support vector regression. *Computational Statistics & Data Analysis*, 77:371–382, 2014.

[18] Yitian Xu, Xiaoyan Li, Xianli Pan, and Zhiji Yang. Asymmetric ν-twin support vector regression. *Neural Computing and Applications*, 30(12):3799–3814, 2018.

[19] S Balasundaram and Yogendra Meena. Robust support vector regression in primal with asymmetric huber loss. *Neural Processing Letters*, 49(3):1399–1431, 2019.

[20] Arnold Zellner. Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394):446–451, 1986.

[21] Carsten Croonenbroeck and Georg Stadtmann. Minimizing asymmetric loss in medium-term wind power forecasting. *Renewable Energy*, 81:197–208, 2015.

[22] Sven F Crone. Training artificial neural networks for time series prediction using asymmetric cost functions. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.*, volume 5, pages 2374–2380. IEEE, 2002.

[23] Korbinian Dress, Stefan Lessmann, and Hans-Jörg von Mettenheim. Residual value forecasting using asymmetric cost functions. *International Journal of Forecasting*, 34(4):551–565, 2018.

[24] Clive WJ Granger. Prediction with a generalized cost of error function. *Journal of the Operational Research Society*, 20(2):199–207, 1969.

[25] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004.

[26] MA Hanson and B Mond. Necessary and sufficient conditions in constrained optimization. *Mathematical Programming*, 37(1):51–58, 1987.

[27] Edgar Osuna, Robert Freund, and Federico Girosi. An improved training algorithm for support vector machines. In *Neural networks for signal processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pages 276–285. IEEE, 1997.

[28] John Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.

[29] Léon Bottou and Chih-Jen Lin. Support vector machine solvers. *Large Scale Kernel Machines*, 3(1):301–320, 2007.

[30] Peter F Christoffersen and Francis X Diebold. Optimal prediction under asymmetric loss. *Econometric Theory*, 13(6):808–817, 1997.

[31] You-Gan Wang, Quanxi Shao, and Min Zhu. Quantile regression without the curse of unsmoothness. *Computational Statistics & Data Analysis*, 53(10):3696–3705, 2009.

[32] Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.

[33] Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156, 2001.

[34] Hamed Kebriaei, Babak N Araabi, and Ashkan Rahimi-Kian. Short-term load forecasting with a new nonsymmetric penalty function. *IEEE Transactions on Power Systems*, 26(4):1817–1825, 2011.

[35] Yamin Wang and Lei Wu. Improving economic values of day-ahead load forecasts to real-time power system operations. *IET Generation, Transmission & Distribution*, 11(17):4238–4247, 2017.

[36] Luciana Magnano and JW Boland. Generation of synthetic sequences of electricity demand: Application in south australia. *Energy*, 32(11):2230–2243, 2007.

[37] Alex D Papalexopoulos and Timothy C Hesterberg. A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, 5(4):1535–1547, 1990.

[38] SA Soliman, S Persaud, K El-Nagar, and ME El-Hawary. Application of least absolute value parameter estimation based on linear programming to short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 19(3):209–216, 1997.

[39] Mohamed Mohandes. Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*, 26(4):335–345, 2002.

[40] TM Peng, NF Hubele, and GG Karady. Advancement in the application of neural networks for short-term load forecasting. *IEEE Transactions on Power Systems*, 7(1):250–257, 1992.

[41] Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, 2001.

[42] Evgenia Dimitriadou, Kurt Hornik, Friedrich Leisch, David Meyer, and Andreas Weingessel. Misc functions of the department of statistics (e1071), tu wien. *R Package*, 1:5–24, 2008.

[43] Frauke Günther and Stefan Fritsch. neuralnet: Training of neural networks. *The R Journal*, 2(1):30–38, 2010.

[44] Roger Koenker, Stephen Portnoy, Pin Tian Ng, Achim Zeileis, Philip Grosjean, and Brian D Ripley. Package 'quantreg', 2012.

[45] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[46] Gilles Blanchard, Olivier Bousquet, Pascal Massart, et al. Statistical performance of support vector machines. *The Annals of Statistics*, 36(2):489–531, 2008.

[47] Youlong Yang, Jinxing Che, Yanying Li, Yanjun Zhao, and Suling Zhu. An incremental electric load forecasting model based on support vector regression. *Energy*, 113:796–808, 2016.

[48] Shirin Tavara. Parallel computing of support vector machines: a survey. *ACM Computing Surveys (CSUR)*, 51(6):1–38, 2019.