Support vector regression with asymmetric loss for optimal electric load scheduling

Jinran Wu

School of Mathematical Sciences, Queensland University of Technology, Australia

You-Gan Wang

*School of Mathematical Sciences, Queensland University of Technology, Australia

Yu-Chu Tian

School of Computer Science, Queensland University of Technology, Australia

Kevin Burrage

School of Mathematical Sciences, Queensland University of Technology, Australia

Department of Computer Science, University of Oxford, U.K.

and

Taoyun Cao

School of Statistics and Mathematics, Guangdong University of Finance and Economics, China.

May 4, 2020

*Corresponding author: you-gan.wang@qut.edu.au.

Support vector regression with asymmetric loss for optimal electric load scheduling

Abstract

In energy demand scheduling, 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 robustness. The electric load data from the state of New South Wales in Australia is used to show the superiority of our proposed framework. Our new framework results in a daily cost reduction ranging from 32.1% to 65.4% depending on the actual cost ratio of the two types of errors.

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

1 Introduction

Electric load scheduling is essential in energy system administration. More accurate prediction on the electric load can efficiently reduce the economic cost due to over-ordering or the under-ordering (Farzin et al., 2017). However, accurate prediction of the electric load is extremely challenging due to the non-linear, dynamic and complex system in demand (Yang et al., 2019). In case of under-predictions, the agency has to loan extra loads from other agencies at a much higher price. On the other hand, if more loads are ordered than needed, the agency will waste the remaining loads or save them in battery storage. Thus, the agency tends to make a biased load scheduling to reduce economic costs (Li and Chiang, 2016). However, current predictions with symmetric loss will result in more unnecessary economic costs although unbiased predictions can be obtained.

Two machine learning approaches with symmetric loss are popularly used in electric load scheduling, including artificial neural networks and support vector regressions (Kuster et al., 2017). The first approach works as a brain with a collection of neurons to learn the data pattern. This approach has many variations in load scheduling, such as, wavelet neural networks (Chen et al., 2009), RBF neural networks (Lu et al., 2016), boosted neural networks (Khwaja et al., 2017), and extreme learning machine (Wu et al., 2019). To handle these complex systems more efficiently, deep learning has recently been introduced. Kong et al. (2017) presented a long short-term memory (LSTM) recurrent neural network for short-term residual load forecasting. The second approach of support vector regression (SVR) combines L_2 regularization and an insensitive Laplace loss for the load scheduling; see basic SVR (Ceperic et al., 2013), weighted SVR (Elattar et al., 2010), on-line SVR (Vrablecová et al., 2018), and vector field-based SVR (Zhong et al., 2019). There are 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 (Vapnik, 2013) and exponential squared loss (Wang et al., 2013).

However, as recommended by Zellner (1986), symmetric loss is inappropriate for load scheduling. Thus, the forecasting model with asymmetric loss is recommended based on the real economic cost minimization for electric load scheduling (Croonenbroeck and Stadtmann, 2015). The influences from over-prediction and under-prediction bring different serious outcomes; thus, the type of predictions should be distinguished. Asymmetric loss considering different outcomes is more rational. So far, limited work on machine learning with asymmetric loss has been done. Linear-linear cost is one of the asymmetric loss functions, that applies two parameters to determine the the severity of the prediction type (over-predictions or under-predictions). It has been used in neural networks training to minimize costs in some specific applications. Crone (2002) introduced linear-linear cost in neural networks to predict optimum service levels in inventory management. Dress et al. (2018) developed a neural network with linear-linear cost for resale price forecasting to aid pricing decisions. A complex linear-linear cost was developed by Li and Chiang (2016) for tree regression training, where the prediction was divided into four types with different severities.

Different from the former asymmetric loss designs for real cost minimization, asymmetric support vector regression has been developed for more accurate regressions (Xu et al., 2009; Peng, 2010; Seok et al., 2010; Hwang, 2011; Stockman et al., 2012). Peng et al. (2014) modified twin support vector regression with two non-parallel bound functions for stable predictions. Huang et al. (2014) proposed an asymmetrical and quadratic loss function for SVR to accurately predict power usage. Xu et al. (2018) presented an asymmetric ν-twin

SVR based on the pinball loss function, which can enhance generalization ability by controlling the fitting error. Balasundaram and Meena (2019) 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 reviewed above, most current electric load scheduling systems are designed for error minimization with symmetric loss. Symmetric loss ignores the difference in penalties from over-prediction and under-prediction in load scheduling. A biased prediction with asymmetric loss is more appropriate in load scheduling. This means a cost-oriented asymmetric loss is demanded by considering the real penalties to the over-prediction and under-prediction. Furthermore, due to the solid theoretical foundation, a cost-oriented asymmetric SVR framework (AsySVR) is more promising for load scheduling. An asymmetric least absolute value regression (AsyLAV), as a special AsySVR framework, is presented to explore the relationship between asymmetric framework and quantile regression. This is motivated by the following three considerations.

a. The economic costs should be incorporated in electric load scheduling. A asymmetric loss considering different penalties for over-predictions and under-predictions is more appropriate in the forecasting system training for economic predictions.

b. Similar to insensitive Laplace loss, an insensitive linear-linear cost needs to be designed for the asymmetric SVR training in order to guarantee the robustness of the proposed asymmetric framework.

c. The effectiveness of the proposed asymmetric SVR framework can be employed in Australian electric load scheduling.

We incorporate a cost-oriented asymmetric loss into support vector regression, and develop a novel asymmetric SVR framework to reduce the economic costs on electric load scheduling. Additionally, in order to make the AsySVR framework robust, an insensitive linear-linear cost is designed, that ignores the smaller errors for the asymmetric SVR training. As a special AsySVR framework, an AsyLAV framework, is proposed by incorporating the linear-linear cost in LAV regression.

To apply our theoretical discussion, we focus on the state of New South Wales load scheduling project to minimize operational costs. Furthermore, the half-hourly electric load from February, 01, 2019 to March, 20, 2019 in NSW is used as experimental data, retrieved from the Australia Energy Market Operator (AEMO). As displayed in Fig. 1, the data set is divided into two groups: a training set (1344 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, the Australia's National Electricity Market (NEM) volume-weighted wholesale price (k_1) is around A\$82/MWh in NSW. Also, the NEW has real time balancing with the obligation before delivery. Thus, the price (k_2) is highly capped, at A\$$14,500/MWh$ (mid-2018). Apparently, there exists a bias in load scheduling considering the operational costs. Here 4 scenarios ($k_2 = 400$, 600, 800 and 1200) with $k_1 = 80$, are taken account in our NSW electric load scheduling. From an economic cost comparison, the asymmetric frameworks can significantly reduce the operational costs with the state of New South Wales electric load data.

To this end, 3 main **contributions** of our work are as follows.

a. An insensitive linear-linear cost is modified for our asymmetric SVR training. This

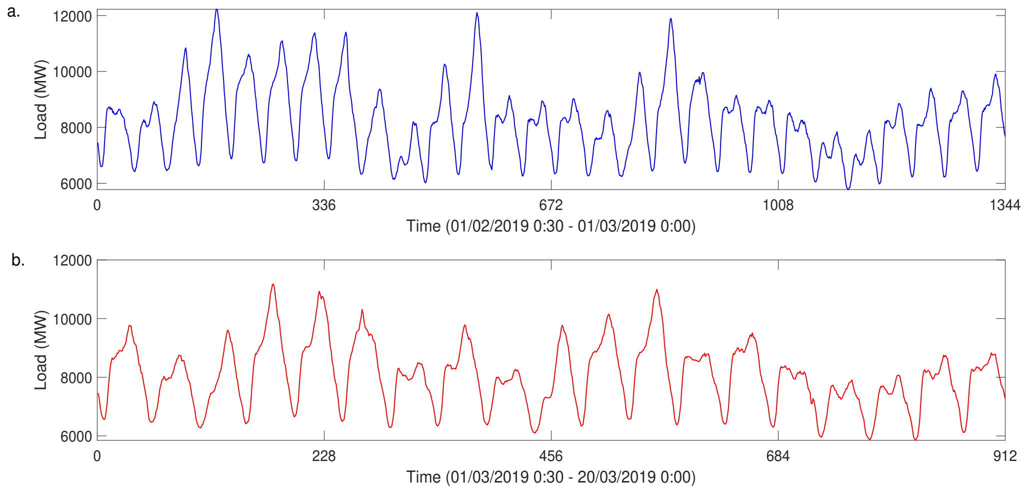

Figure 1: The electric load from NSW: a). the training data (30 days), and b). the test data (20 days).

insensitive cost can ensure the robustness of the forecasting system for electric load forecasting.

b. A cost-oriented asymmetric SVR framework with a modified linear-linear cost is proposed for load scheduling cost reduction, that can distinguish the type of predictions in energy operations.

c. In 4 electric load scheduling scenarios, the cost-oriented asymmetric loss is more practical and appropriate, that can significantly save mean daily costs from 32.1% to 65.4%.

The organization of this paper is as follows. Section 2 reviews the linear-linear cost, and presents our insensitive linear-linear cost; then, our proposed asymmetric framework is illustrated. Then, in Section 3, the performance of our proposed frameworks are evaluated in 4 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 support vector regression 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,

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

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 Granger (1969), the cost should represent the actual business objective, such as, the profit maximization. Thus, in our asymmetric loss designs, the linear-linear cost (LLC) is employed, which can directly quantify the economic costs on each error as (Granger, 1969),

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

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 $l(u|k_1, k_2) = \max(k_1 u^+, k_2 u^-)$. Also, we have, if $k_1 > 0$,

$l(u|k_1,k_2) = k_1 l(u|1,k)$, 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. 2, the parameter k shows the bias on two types of errors. Obviously, if k is larger than 1, under-prediction brings more serious cost than over-prediction. Otherwise, over-prediction is more costly.

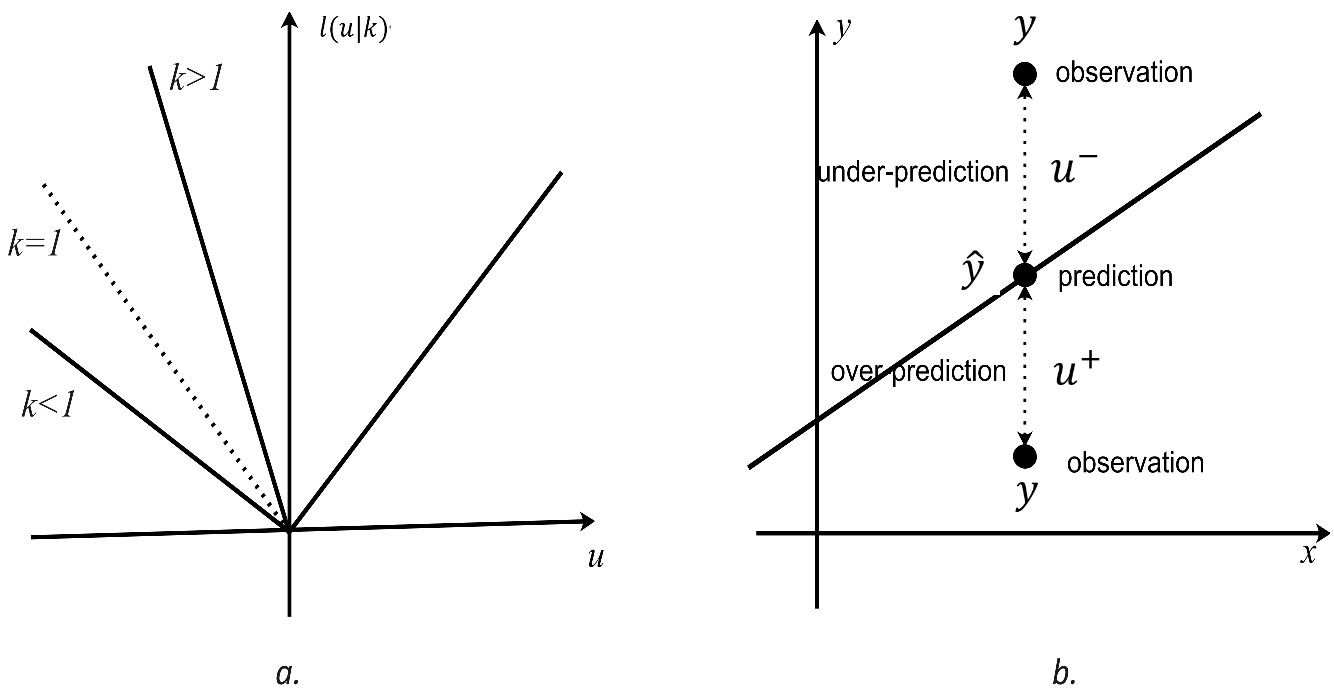

Figure 2: The linear-linear cost: (a). the proposed linear-linear cost, and (b). the illustration of prediction types.

2.2 The linear-linear cost with an insensitive parameter

To enhance the robustness of the model with LLC, according to good performance of the insensitive Laplace loss by Vapnik (2013), an insensitive linear-linear cost (insensitive LLC) is developed with an insensitive parameter ϵ, that can ignore the smaller errors ($|u| \leqslant \epsilon$). The proposed insensitive LLC is formulated as,

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

Different from Fig. 2, two soft margins are introduced in our insensitive LLC. As shown in Fig. 3, only the larger prediction errors located out of margins are counted to calculate the cost. Defined as: $\xi = u^{-} - \epsilon$ and $\xi^{*} = u^{+} - \epsilon$. More specifically, the proposed insensitive LLC can guarantee the model's robustness, especially in complex networks.

2.3 The new support vector regression

Now considering the optimized loss in ϵ-SVR (Smola and Schölkopf, 2004), we incorporate the insensitive LLC, Eq. (4), in the SVR structure to develop an asymmetric SVR. Moreover, the corresponding convex optimization problem (primal objective function) with

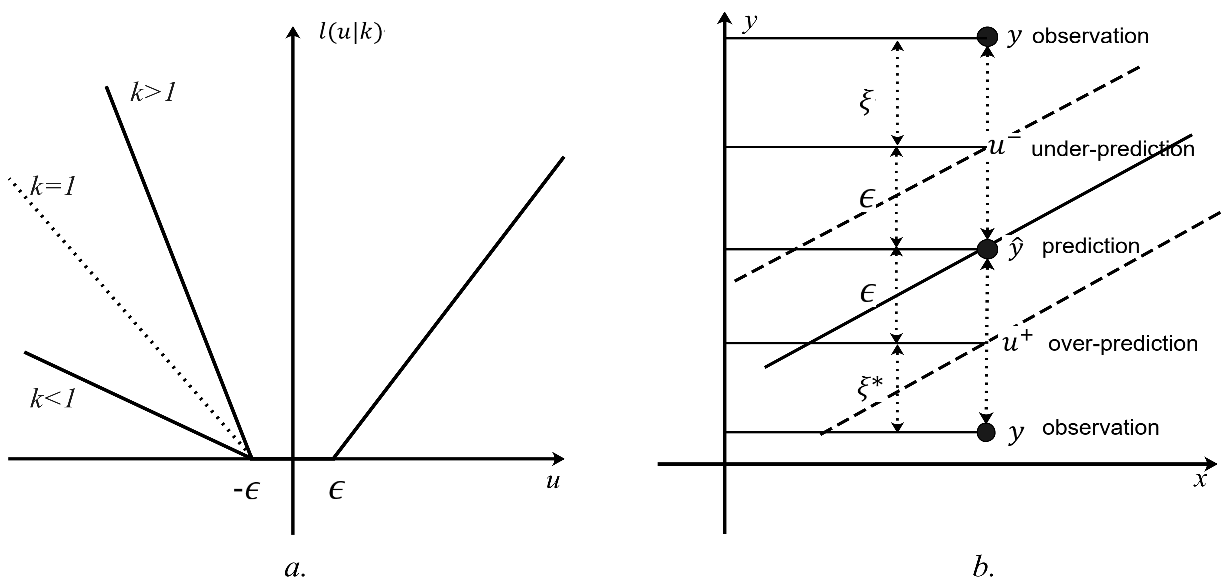

Figure 3: The proposed insensitive linear-linear cost: (a). the proposed insensitive linear-linear cost, (b). the illustration of prediction types.

slack variables ξ_i and ξ_i^* for the cost-oriented asymmetric framework is formulated as,

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

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 noticed that conventional asymmetric cost functions are for accurate regressions to address noises or outliers in dataset.

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

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

with the Lagrangian L 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

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

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

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

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

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 (Hanson and Mond, 1987),

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

Finally, substitute Eq. (8) in 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 KKT condition.

2.4 Connection to quantile regression

In the special case when $\epsilon = 0$ and $C \longrightarrow +\infty$, the asymmetric SVR framework, the loss becomes 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 loss function for τ-quantile regression, also known as the least absolute value regression under asymmetric loss (AsyLAV) (Christoffersen and Diebold, 1997; Wang et al., 2009; Fu and Wang, 2012). The case $k = 1$ or $\tau = 0.5$ corresponds to median regression.

Thus, this optimization problem can be obtained via quantile regression (Koenker and Hallock, 2001).

3 The case study

In the section, two cost-oriented asymmetric frameworks, asymmetric LAV regression (AsyLAV), and asymmetric SVR (AsySVR), are evaluated on NSW electric load scheduling data. 4 symmetric frameworks, least square regression (LS), least absolute value regression (LAV), support vector regression (SVR), and multilayer perception (MLP), are chosen as the benchmark models to show the superiority of our proposed asymmetric frameworks.

3.1 Evaluation criterion

Different from the conventional evaluation criterion, the mean daily cost (MDC, unit: AUS$) is used to show the effectiveness of our asymmetric frameworks in 4 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 $u_i = \hat{y}_i - y_i$ representing the prediction error and I is the indicator function. Apparently, a smaller MDC value means a better daily optimal load scheduling.

Furthermore, the proportion α is employed to show the proportion of over-prediction,

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

From Eq. (17), if α is smaller than 0.5, the forecasting tends to under-prediction. Otherwise, the over-prediction is preferable.

3.2 Experimental settings

In the one-step ahead load scheduling project, the target is the next half-hourly load. As for the input, due to the high correlation with next step load (Kebriaei et al., 2011; Wang and Wu, 2017), the half-hourly load of the previous day (48 data points) is taken as the main input. Moreover, the harmonic functions, $\sin(2\pi t/T)$, and $\cos(2\pi t/T)$ (with time order t, and the cycle length $T = 48$), describing the daily seasonality, is taken as the additional input (Stolwijk et al., 1999).

Two asymmetric frameworks, the least-absolute-value regression with asymmetric loss (AsyLAV), and AsySVR, are validated in 4 scenarios as: 1). $k_1 = 80$, and $k_2 = 400$; 2). $k_1 = 80$, and $k_2 = 600$; 3). $k_1 = 80$, and $k_2 = 800$; and 4). $k_1 = 80$, and $k_2 = 1200$.

For the parameter setting in the AsySVR framework, the regularization coefficient C and the insensitive parameter ϵ are set as $C = 10$ and $\epsilon = 0.001$ (Scenario 1 and 2), $C = 1$ and $\epsilon = 0.001$ (Scenario 3), and $C = 100$ and $\epsilon = 0.001$ (Scenario 4), by the 5-folder cross-validation shown in Appx. A. Furthermore, the linear function is chosen as the kernel in our study.

3.3 Experimental results

The partial experimental results ($02/02/2019\ 0:30 - 04/03/2019\ 0:00$) of two asymmetric frameworks, AsyLAV and AsySVR, are shown in Fig. 4 and Fig. 5, respectively.

As illustrated in Fig. 4, all predictions by AsyLAV are larger than that by LAV in the 4 scenarios. For example, at $k_2 = 1200$, the AsyLAV curve (red curve) is significantly beyond the LAV curve (blue curve). Meanwhile, compared with LAV, the prediction with

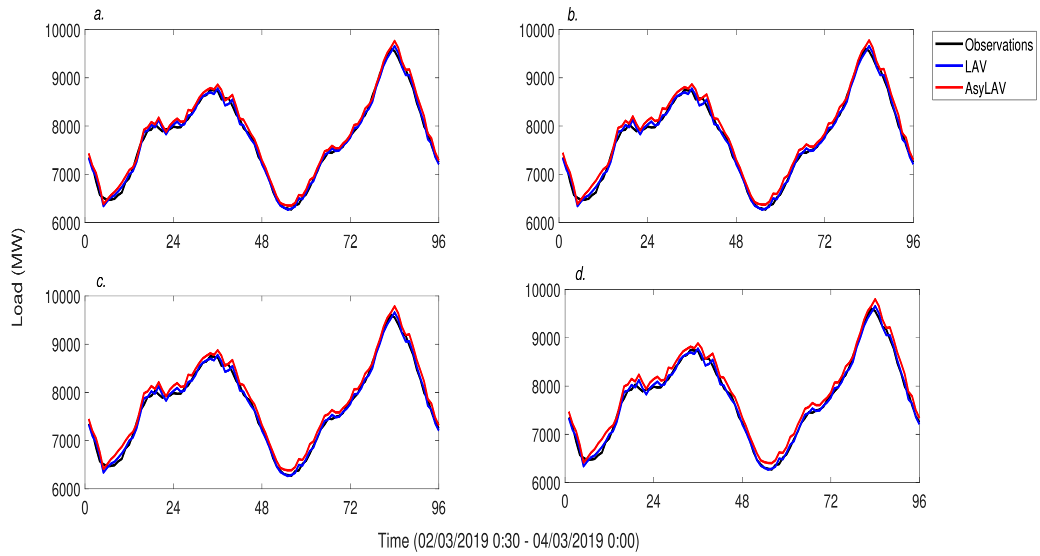

Figure 4: The predictions using LAV and AsyLAV in 4 scenarios: (a). $k_2 = 400$, (b). $k_2 = 600$, (c). $k_2 = 800$, and (d). $k_2 = 1200$.

bias is given by our AsyLAV. From Tab. 1, the α proportion is 0.522 by LAV, while that is 0.941 provided by AsyLAV, at $k_2 = 1200$. In addition, the AsyLAV can efficiently decrease the economic cost. One of the significant scenarios is Scenario 4 where AsyLAV (MDC: $315,220$ AUD) can save around twice the costs as LAV (MDC: $956,133$ AUD) for daily load scheduling.

From Fig. 5, compared with the basic SVR, similar findings can be obtained by AsySVR. Furthermore, as illustrated in Tab. 1, AsySVR is superior to AsyLAV in cost saving. For example, in Scenario 1, the MDC by AsySVR is $A\$238,963$, while the MDC by AsyLAV is $A\$240,084$. This can demonstrate the insensitive LLC is more efficient than LLC.

In short, the asymmetric loss (LLC or insensitive LLC) brings more biases in the prediction, that can significantly decrease daily costs in the load scheduling. In particular, the

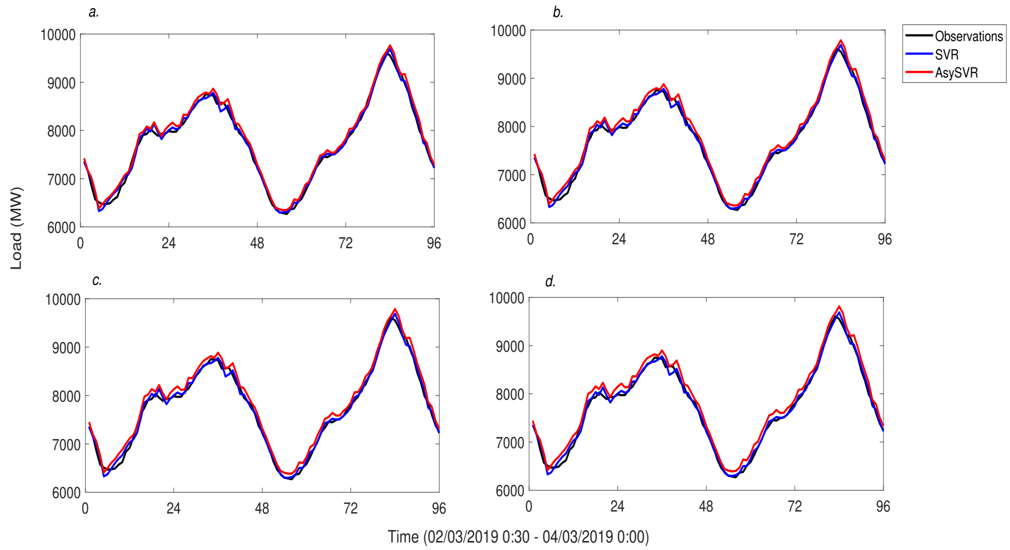

Figure 5: The predictions using SVR and AsySVR in 4 scenarios: (a). $k_2 = 400$, (b). $k_2 = 600$, (c). $k_2 = 800$, and (d).$k_2 = 1200$.

proposed insensitive LLC is superior in the daily cost minimization.

3.4 Comparative analysis

To show the efficiency of our two asymmetric frameworks, 4 popular symmetric frameworks, least square regression (LS) (Papalexopoulos and Hesterberg, 1990), least absolute value regression (LAV) (Soliman et al., 1997), support vector regression (SVR) (Mohandes, 2002; Ceperic et al., 2013), and multilayer perception (MLP) (Peng et al., 1992; Hippert et al., 2001), are investigated in the 4 scenarios. The residuals and daily costs by 4 symmetric frameworks and 2 asymmetric frameworks in Scenario 1, Scenario 2, Scenario 3, and Scenario 4, are shown in Fig. 6, Fig. 7, Fig. 8, and Fig. 9, respectively.

As shown in Fig. 6, the bias brought by asymmetric frameworks (AsyLAV and AsySVR)

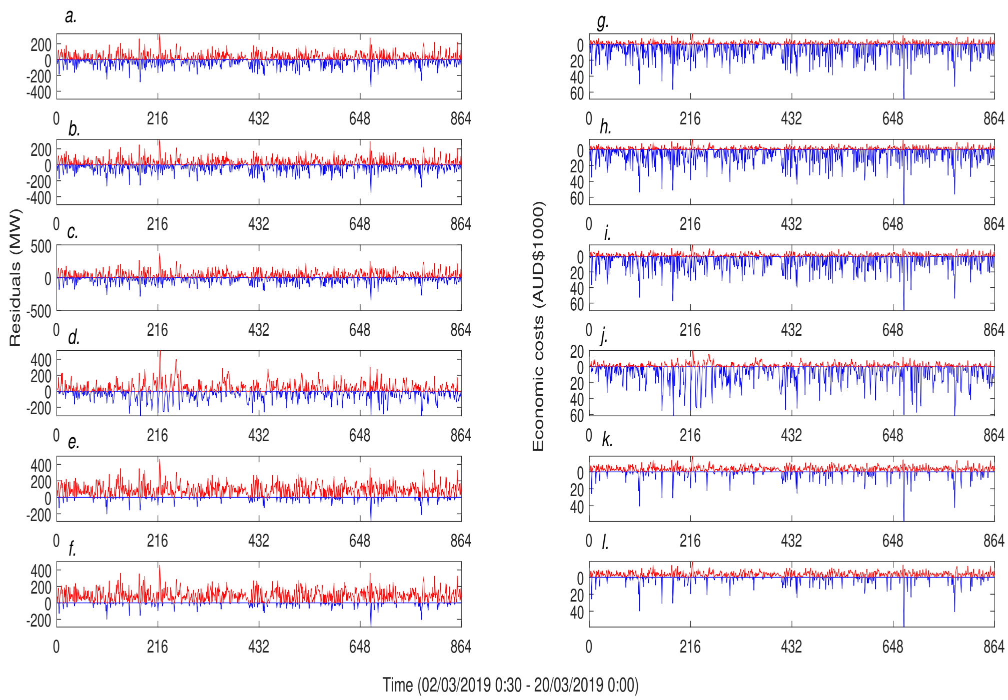

Figure 6: The residuals and corresponding daily costs in Scenario 1: LS (a & g), LAV (b & h), SVR(c & i), MLP (d & k), AsyLAV (e & k), and AsySVR(f & l). Red color represents over-prediction error while blue color represents under-prediction.

decreases the accuracy of forecasting. Meanwhile, the daily costs on 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 the under-estimation. Apparently, in Scenario 1 ($k_1 = 80$, and $k_2 = 400$), compared to symmetric frameworks, more over-predictions are offered by our proposed asymmetric frameworks. From the daily cost comparison, the symmetric frameworks balance their loss with more costs on the under-predictions, while the asymmetric frameworks can give biased predictions in good daily-cost balance.

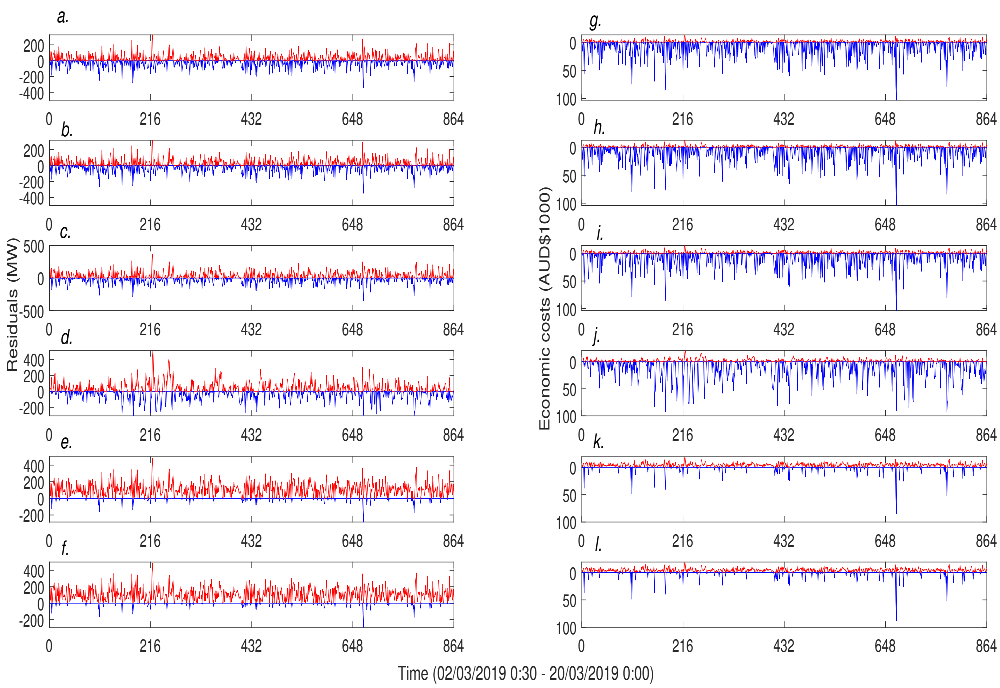

Figure 7: The residuals and corresponding daily costs in Scenario 2: LS (a & g), LAV (b & h), SVR(c & i), MLP (d & k), AsyLAV (e & k), and AsySVR (f & l). Red color represents over-prediction error while blue color represents under-prediction.

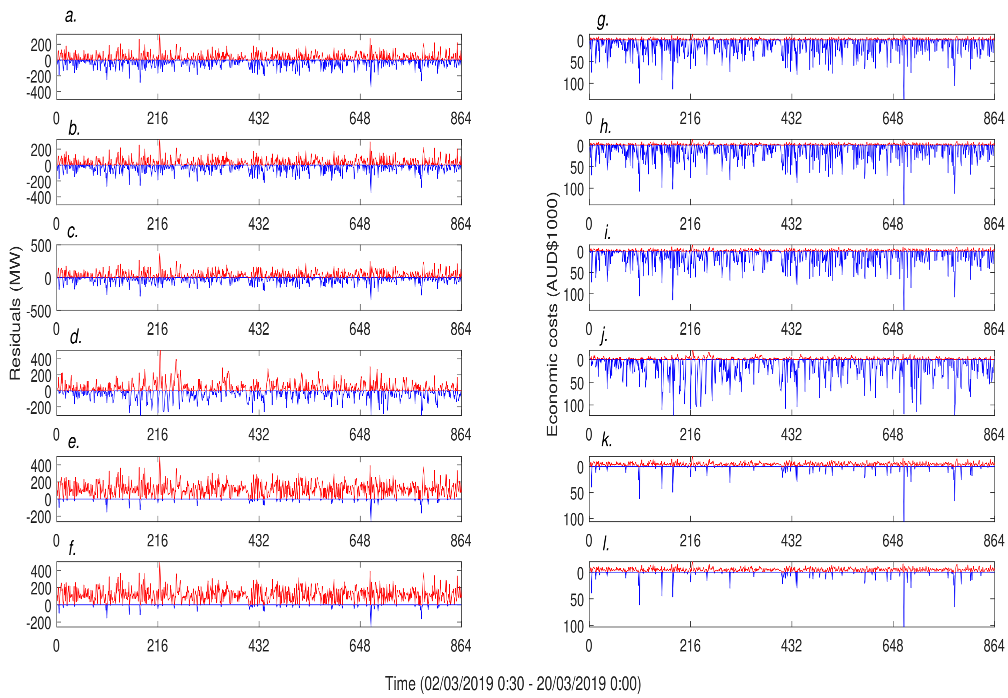

Figure 8: The residuals and corresponding daily costs in Scenario 3: LS (a & g), LAV (b & h), SVR(c & i), MLP (d & k), AsyLAV (e & k), and AsySVR(f & l). Red color represents over-prediction error while blue color represents under-prediction.

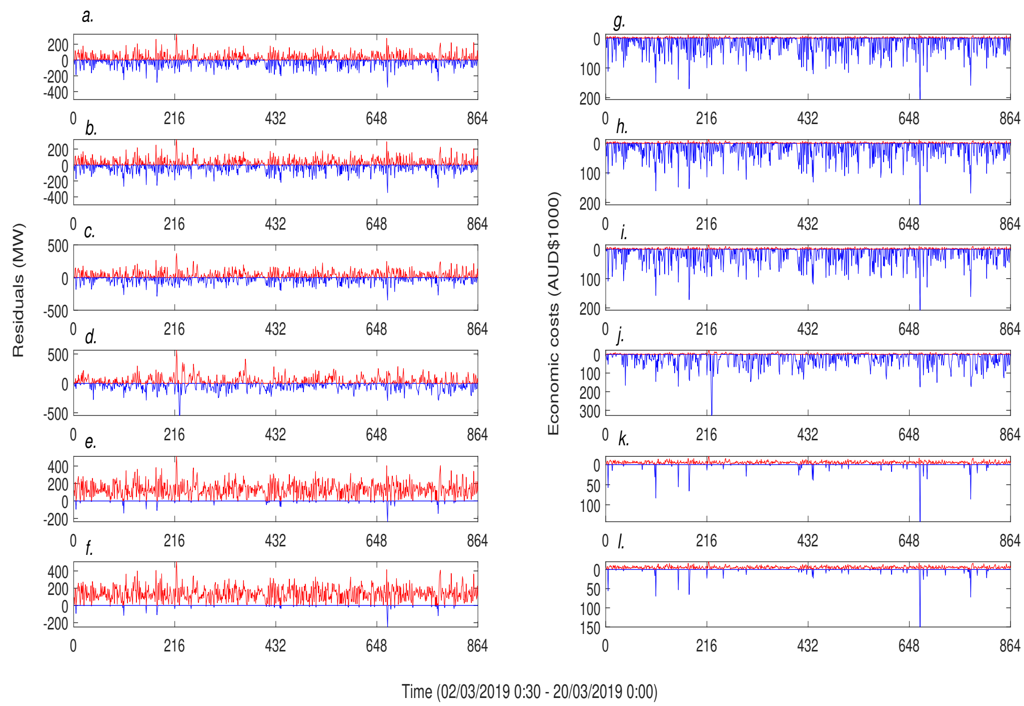

Figure 9: The residuals and corresponding daily costs in Scenario 4: LS (a & g), LAV (b & h), SVR(c & i), MLP (d & k), AsyLAV (e & k), and AsySVR (f & l). Red color represents over-prediction error while blue color represents under-prediction.

In addition, from 4 scenarios, when the penalty k_2 increases, our proposed asymmetric frameworks in daily costs are more efficient. Especially in Fig. 8, almost forecasting values by asymmetric frameworks are over-predicted to pay less costs on under-predictions.

Table 1: The MDC and α proportion for daily load scheduling: Relative performance of symmetric frameworks and asymmetric frameworks. (MDC: AUD)

	Symmetric frameworks				Asymmetric frameworks	
	LS	LAV	SVR	MLP	AsyLAV	AsySVR
Scenario 1: $k_1 = 80$, $k_2 = 400$						
MDC	355,602	359,891	352,140	465,857	240,084	238,963
α	0.520	0.522	0.562	0.507	0.843	0.837
Scenario 2: $k_1 = 80$, $k_2 = 600$						
MDC	501,763	508,951	490,778	656,635	270,851	267,383
α	0.520	0.522	0.562	0.507	0.895	0.896
Scenario 3: $k_1 = 80$, $k_2 = 800$						
MDC	647,923	658,012	629,417	847,413	289,023	288,339
α	0.520	0.522	0.562	0.507	0.921	0.922
Scenario 4: $k_1 = 80$, $k_2 = 1200$						
MDC	940,245	956,133	906,694	1,228,970	315,220	313,863
α	0.520	0.522	0.562	0.507	0.941	0.950

The prediction performance for 4 symmetric frameworks and two asymmetric frameworks in 4 scenarios are recorded in Tab. 1. Obviously, based on the MDC, our proposed asymmetric frameworks, AsyLAV and AsySVR, can significantly reduce the daily costs in electric load scheduling. One of the most obvious cases is Scenario 4, where the MDCs for LS, LAV, SVR, MLP, AsyLAV, and AsyVR, are $940,245$, $956,133$, $906,694$, $1,228,970$, $315,220$, and $313,863$ AUD, respectively. Furthermore, the α proportion for the asymmetric frameworks range from 0.837 to 0.950, while those for the symmetric frameworks ranges from 0.507 to 0.562, in 4 scenarios. This shows our proposed framework can balance the daily costs by adding more biases in predictions.

Furthermore, compared with the basic LLC, our proposed insensitive LLC also obtains good improvements in our study. As shown in Tab. 1, $A\$684 - A\$3,468$ of savings in the daily cost can be achieved using our insensitive LLC.

Additionally, the quantile regression is also implemented in the daily load scheduling project. The quantile parameter setting τ and experimental results are detailed in Appx. B. Compared with the special asymmetric framework, AsyLAV, the quantile regression obtains similar two indexes for daily scheduling.

To summarize, 3 points can be concluded. The first is that the asymmetric frameworks can reduce the daily costs by introducing more biases in the scheduling. Secondly, the asymmetric LAV framework is the same as the quantile regression in essence. Finally, our proposed insensitive LLC is more efficient than LLC in load scheduling, and the proposed AsySVR framework obtains the lowest daily costs among all considered frameworks.

4 Conclusion

In the paper, a cost-oriented asymmetric frameworks, AsySVR, is developed for electric load scheduling. Based on different penalties to over-predictions and under-predictions, an insensitive linear-linear cost is proposed to train the asymmetric SVR framework for real daily cost minimization. Particularly, compared with conventional SVR framework, in our NSW load scheduling project, the proposed asymmetric framework can achieve excellent daily cost reductions as: $113,177$ AUD (Scenario 1), $223,389$ AUD (Scenario 2), $341,078$ AUD (Scenario 3), and $592,831$ AUD (Scenario 4). Additionally, an insensitive LLC is designed for our AsySVR framework, where an insensitive parameter ϵ is used to improve the robustness. From economic cost comparison, the AsySVR framework with the insensitive LLC is the most superior for daily cost reduction in all scenarios. Additionally, we also show that the AsyLAV framework is the same as the quantile regression in essence. Furthermore, our proposed AsySVR framework is an efficient and promising tool for electric load scheduling.

However, there are still some limitations in our work. The first limitation is that our electric load scheduling is based on time series modeling, ignoring some environmental factors, such as, temperature, humidity. Furthermore, in our AsySVR framework, the cross-validation is used to select the regularization parameter C and the insensitive parameter ϵ, that selects the parameters from the alternative values with large computational costs for the AsySVR training.

In future, there are many research directions. For electric load scheduling, more economical and environmental factors can be incorporated to establish a high-accurate forecasting system, promoting the economic cost reduction. Moreover, our proposed asymmetric frameworks can be extended to other economic cost minimization 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 employed in other state-of-the-art methods, such as recurrent neural network architecture. Finally, the generalized cost-oriented asymmetric loss can be explored in other specific practices.

Acknowledgements

This work is in part supported by Australian Research Council (ARC) Discovery Project (DP160104292), the ARC Center of Excellence for Mathematical and Statistical Frontiers, Guangdong Basic and Applied Basic Research Foundation 2020A1515011580.

Appendices

A The hyper-parameter selection for AsySVR

The 5-folder cross-validation is used to select the regularization parameter C and the insensitive parameter ϵ for our proposed AsySVR framework. Furthermore, the alternative values for C are set as: 0.01, 0.1, 1, 10, and 100, while these for ϵ are set as: 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 MDC in 4 scenarios are displayed in Fig. 10. Based on the results from cross-validation, the parameter are set as: Scenario 1 (C : 10, ϵ : 0.001), Scenario 2 (C : 10, ϵ : 0.001), Scenario 3 (C : 1, ϵ : 0.001), and Scenario 4 (C : 100, ϵ : 0.001).

B Quantile regression for the load scheduling

In the section, the quantile regression is developed in the daily load scheduling. According to the discussion in Sect. 2.4, the quantiles τ for the 4 scenarios are calculated as: 83.3%, 88.2%, 90.9%, and 93.8%, respectively. Two indices, MDC and the α proportion, by using

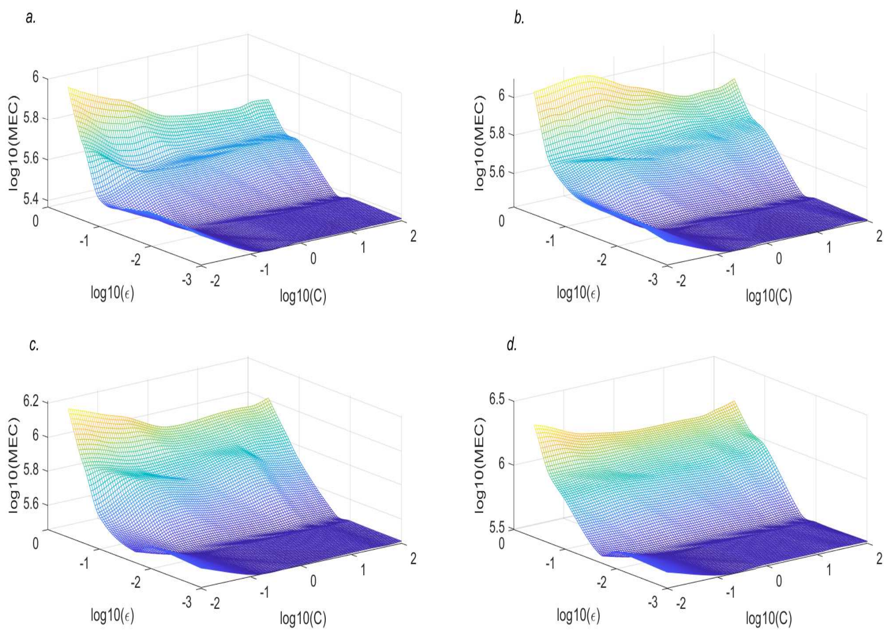

Figure 10: The MDC with different parameter settings in 4 scenarios: (a). $k_2 = 400$, (b). $k_2 = 600$, (c). $k_2 = 800$, and (d). $k_2 = 1200$.

the quantile regression, are shown in Tab. 2. As illustrated before, it can be observed that the α value is close to the quantile value in each scenario. Furthermore, the performance of the quantile regression is consistent with the AsyLAV framework.

Table 2: The MDC (in AUS$) and α proportion for daily load scheduling by quantile regression.

Scenario	τ	MDC	α
1	83.3 %	242,167	0.845
2	88.2 %	274,169	0.883
3	90.9 %	290,238	0.916
4	93.8 %	315,828	0.950

References

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

Ceperic, E., V. Ceperic, and A. Baric (2013). A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems 28*(4), 4356–4364.

Chen, Y., P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, P. B. Friedland, and S. J. Rourke (2009). Short-term load forecasting: Similar day-based wavelet neural networks. *IEEE Transactions on Power Systems 25*(1), 322–330.

Christoffersen, P. F. and F. X. Diebold (1997). Optimal prediction under asymmetric loss. *Econometric theory 13*(6), 808–817.

Crone, S. F. (2002). 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, pp. 2374–2380. IEEE.

Croonenbroeck, C. and G. Stadtmann (2015). Minimizing asymmetric loss in medium-term wind power forecasting. *Renewable Energy 81*, 197–208.

Dress, K., S. Lessmann, and H.-J. von Mettenheim (2018). Residual value forecasting using asymmetric cost functions. *International Journal of Forecasting 34*(4), 551–565.

Elattar, E. E., J. Goulermas, and Q. H. Wu (2010). 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.

Farzin, H., R. Ghorani, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie (2017). A market mechanism to quantify emergency energy transactions value in a multi-microgrid system. *IEEE Transactions on Sustainable Energy 10*(1), 426–437.

Fu, L. and Y.-G. Wang (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis 56*(8), 2526–2538.

Granger, C. W. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society 20*(2), 199–207.

Hanson, M. and B. Mond (1987). Necessary and sufficient conditions in constrained optimization. *Mathematical programming 37*(1), 51–58.

Hippert, H. S., C. E. Pedreira, and R. C. Souza (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems 16*(1), 44–55.

Huang, X., L. Shi, K. Pelckmans, and J. A. Suykens (2014). Asymmetric ν-tube support vector regression. *Computational Statistics & Data Analysis 77*, 371–382.

Hwang, C. (2011). Asymmetric least squares regression estimation using weighted least squares support vector machine. *Journal of the Korean Data & Information Science Society 22*, 999–1005.

Kebriaei, H., B. N. Araabi, and A. Rahimi-Kian (2011). Short-term load forecasting with a new nonsymmetric penalty function. *IEEE Transactions on Power Systems 26*(4), 1817–1825.

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

Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of economic perspectives 15*(4), 143–156.

Kong, W., Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang (2017). Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid 10*(1), 841–851.

Kuster, C., Y. Rezgui, and M. Mourshed (2017). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society 35*, 257–270.

Li, G. and H.-D. Chiang (2016). Toward cost-oriented forecasting of wind power generation. *IEEE Transactions on Smart Grid 9*(4), 2508–2517.

Lu, Y., T. Zhang, Z. Zeng, and J. Loo (2016). An improved rbf neural network for short-term load forecast in smart grids. In *2016 IEEE International Conference on Communication Systems (ICCS)*, pp. 1–6. IEEE.

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

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

Peng, T., N. Hubele, and G. Karady (1992). Advancement in the application of neural networks for short-term load forecasting. *IEEE Transactions on Power Systems 7*(1), 250–257.

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

Peng, X., D. Xu, and J. Shen (2014). A twin projection support vector machine for data regression. *Neurocomputing 138*, 131–141.

Seok, K. H., D. Cho, C. Hwang, and J. Shim (2010). Support vector quantile regression using asymmetric e-insensitive loss function. In *2010 2nd International Conference on Education Technology and Computer*, Volume 1, pp. V1–438. IEEE.

Smola, A. J. and B. Schölkopf (2004). A tutorial on support vector regression. *Statistics and computing 14*(3), 199–222.

Soliman, S., S. Persaud, K. El-Nagar, and M. El-Hawary (1997). 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.

Stockman, M., R. S. El Ramli, M. Awad, and R. Jabr (2012). 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)*, pp. 651–656. IEEE.

Stolwijk, A., H. Straatman, and G. Zielhuis (1999). Studying seasonality by using sine and cosine functions in regression analysis. *Journal of Epidemiology & Community Health 53*(4), 235–238.

Vapnik, V. (2013). *The nature of statistical learning theory.* Springer science & business media.

Vrablecová, P., A. B. Ezzeddine, V. Rozinajová, S. Šárik, and A. K. Sangaiah (2018). Smart grid load forecasting using online support vector regression. *Computers & Electrical Engineering 65*, 102–117.

Wang, X., Y. Jiang, M. Huang, and H. Zhang (2013). Robust variable selection with exponential squared loss. *Journal of the American Statistical Association 108*(502), 632–643.

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

Wang, Y.-G., Q. Shao, and M. Zhu (2009). Quantile regression without the curse of unsmoothness. *Computational statistics & data analysis 53*(10), 3696–3705.

Wu, J., Z. Cui, Y. Chen, D. Kong, and Y.-G. Wang (2019). A new hybrid model to predict the electrical load in five states of australia. *Energy 166*, 598–609.

Xu, S.-Q., Q.-Y. Luo, G.-H. Xu, and L. Zhang (2009). Asymmetrical interval regression using extended ε-svm with robust algorithm. *Fuzzy Sets and systems 160*(7), 988–1002.

Xu, Y., X. Li, X. Pan, and Z. Yang (2018). Asymmetric ν-twin support vector regression. *Neural Computing and Applications 30*(12), 3799–3814.

Yang, Y., J. Che, C. Deng, and L. Li (2019). Sequential grid approach based support vector regression for short-term electric load forecasting. *Applied energy 238*, 1010–1021.

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

Zhong, H., J. Wang, H. Jia, Y. Mu, and S. Lv (2019). Vector field-based support vector regression for building energy consumption prediction. *Applied Energy 242*, 403–414.