Optimal hedging of prediction errors using prediction errors*

Yuji Yamada

Graduate School of Business Sciences, University of Tsukuba

3-29-1 Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan

E-mail: `yuji@gssm.otsuka.tsukuba.ac.jp`

Abstract

Wind power energy has been paid much attention recently for various reasons, and the production of electricity with wind energy has been increasing rapidly for a few decades. One of the most difficult issues for using wind power in practice is that the power output largely depends on the wind condition, and as a result, the future output may be volatile or uncertain. Therefore, the prediction of power output in the future is considered important and is key to electric power generating industries making the wind power electricity market work properly. However, the use of predictions may cause other problems due to "prediction errors."

In this work, we will propose a new type of weather derivatives based on the prediction errors for wind speeds, and estimate their hedge effect on wind power energy businesses. At first, we will investigate the correlation of prediction errors between the power output and the wind speed in a Japanese wind farm, which is a collection of wind turbines that generate electricity in the same location. Then we will develop a methodology that will optimally construct a wind derivative based on the prediction errors using nonparametric regressions. A simultaneous optimization technique of the loss and payoff functions for wind derivatives is demonstrated based on the empirical data.

Keywords: Wind power energy, Prediction errors, Weather derivatives, Minimum variance hedge, Non-parametric regression

1 Introduction

Predicting the future weather conditions is considered important in real businesses for many industries including electricity producers and suppliers, because their profit or loss is largely affected by the weather conditions. Under these circumstances, we may have a new risk when the prediction error exists. In this work, we will propose a new type of weather derivative (see, e.g., [6] for the introduction of weather derivatives) to effectively hedge the loss caused by prediction errors.

*The author would like to thank H. Fukuda, R. Tanikawa, and N. Hayashi from Earth Science Department, ITOCHU Techno-Solutions Corporation, for their helpful comments and discussions.

This work is motivated by a critical issue in wind power energy markets, which is explained as follows: In general, electricity companies must sell the output immediately because the electricity has to be consumed as soon as it is produced. Therefore, sales contracts need to be written in advance. However, in the case of electricity production using wind power energy, the power output largely depends on wind conditions, and as a result, tradable volume is uncertain.

What we can do is to predict the future power outputs and quote them in advance. But, this may cause another risk (or loss) associated with prediction errors of the power outputs. One of the objectives of this work is to hedge this type of risk using weather derivatives based on prediction errors on the wind condition. In contrast to the standard weather derivatives in which the underlying index is given by weather data only (such as temperature [1, 2, 3, 4, 8, 9, 14, 15]), the proposed weather derivative uses prediction data and the payoff depends on the difference between the actual data and the prediction data.

Here we consider the power output from a wind farm (WF), which is a collection of wind turbines that generate electricity in the same location. The power output is predicted using numerical weather prediction and the power generating properties for turbines, where a public weather forecasting company computes sophisticated values from Japan Meteorological Agency data. Because of this prediction mechanism, we have both the wind and power predictions data.

A possible sales contract of the power output using the prediction may be described as follows: The value of electricity generated by wind power is normally considered to be low due to the uncertainty of the tradable volume. Here we assume that the electricity price without prediction is estimated to be 3 yen per 1 kWh. On the other hand, the value of the electricity would be estimated to be higher, if the tradable volume were quoted in advance by prediction, but the seller has to guarantee the quoted volume or has to pay the penalty in case of shortages. Suppose that the value of electricity with prediction is given as 7 yen per 1 kWh and that the penalty of the shortage is 10 yen per 1 kWh. These assumptions are not so far from the current situation discussed in the prediction business [10]. In this case, the loss function caused by prediction errors is depicted in Fig. 1.1, which shows the relation between the prediction error for the power output $P - \hat{P}$ (the actual power output minus its prediction) and the loss caused by the prediction error. Note that, even if the prediction error is positive, we can also think of this situation as an opportunity loss to sell the output with a suitable price.

Based on the above discussions, we will first consider the following problems:

P1) Given the loss function and the payoff function of wind derivatives, find the optimal volume of wind derivative using a linear regression.

P2) Given the loss function, find the optimal payoff function of wind derivatives.

We will investigate the hedge effect of wind derivatives and show that using wind derivatives on prediction error of wind speed is highly effective to hedge the loss caused by prediction errors of power output.

Then we will consider a situation in which there already exists a standardized derivative contract with a certain payoff function, but there is some room for improvement on the loss function, e.g., for a WF owner. The problem can be thought of as a reverse problem of P2), which is given as follows:

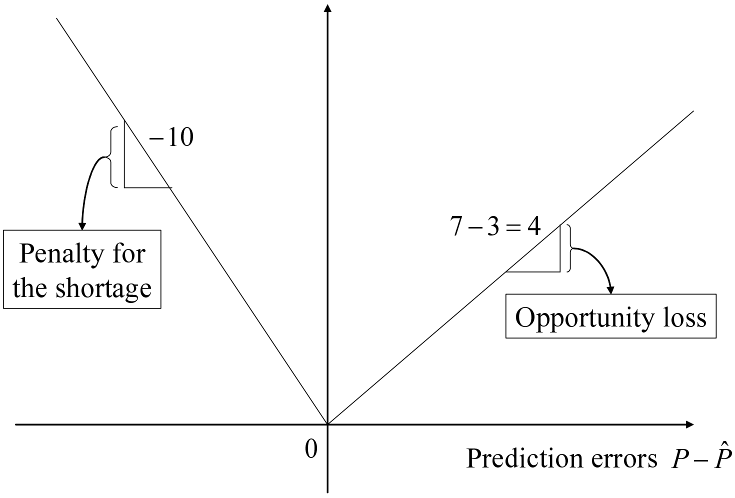

Figure 1.1: An example of loss function

P3) Given the payoff function of wind derivatives, find the optimal loss function against prediction errors of power output.

Finally, we will formulate a simultaneous optimization problem of payoff and loss functions as P4) below:

P4) Optimize the payoff function of wind derivatives and the loss function simultaneously.

The rest of this paper is organized as follows: In Section 2, we explain the definitions of loss and payoff functions, and formulate the first problem, P1), as the standard minimum variance hedging problem. After describing the motivation to introduce a non-parametric regression, we formulate the payoff function optimization problem based on the generalized additive model (GAM) in Section 3. It is shown that the loss function optimization problem may be solved using GAM as well in Section 4, where a simultaneous optimization problem is also formulated and an iterative algorithm is proposed. An empirical analysis and numerical experiments are performed in Section 5 to illustrate the hedge effect of the proposed wind derivatives. Finally, we explain the multi-period case in Section 6, and provide some concluding remarks in Section 7.

We use the following notation: For a sequence of observations of a variable, $x_n,\ n = 1, \ldots, N$, the sample mean and the sample variance are denoted by $\mathrm{Mean}(x_n)$ and $\mathrm{Var}(x_n)$, respectively. $\mathrm{Cov}(x_n,\ y_n)$ and $\mathrm{Corr}(x_n,\ y_n)$ represent the sample covariance and the sample correlation, respectively, where $y_n,\ n = 1, \ldots, N$ is a sequence of observations for another variable. The set of real number is denoted by $\Re$, and an $n \times m$ matrix with real entries is denoted by $A \in \Re^{n \times m}$.

2 Standard minimum variance hedging problem

At first, we will explain loss and payoff functions, and then formulate the first problem, P1), as the standard minimum variance hedging problem.

2.1 Loss and payoff functions

For simplicity, consider a wind power energy trade between two basic positions, a seller and a buyer, for the electricity output. Assume that the seller and the buyer carry out sales contracts based on the prediction of the power output. Let $n = 1, \ldots, N$ be the time index (say, hourly index) and define the following variables:

P_n: Total power output at time n

$\hat{P}_n$: Prediction of P_n (which is computed, e.g., 1 day in advance)

The buyer is willing to trade the power output by using the reference $\hat{P}_n$, and may require a penalty if the prediction error exceeds a certain level.

Let $\epsilon_{p,n}$ $(n = 1, \ldots, N)$ be the prediction error of the power output at time n, which causes a loss for the seller due to the penalty or opportunity loss to sell the output. Suppose that the loss associated with the prediction error of the power output is defined using a loss function as $\phi(\epsilon_{p,n})$. For instance, the loss function may be given as the one shown in Fig. 1.1 if the seller is a WF owner. Also, there is a case in which the prediction is sufficiently accurate or the prediction error is less than a certain (small) level. In this case, the seller can be thought of getting a bonus because of a higher price of power output with prediction, which results in a profit for the seller and makes the loss negative, i.e., $\phi(\epsilon_{p,n}) < 0$. We assume that

$$\mathrm{Mean}\left(\phi\left(\epsilon_{p,n}\right)\right) = 0 \tag{2.1}$$

so that the sum of profit/loss is zero on average.

We will consider a situation in which the seller with $\phi(\cdot)$ would like to compensate their loss on $\epsilon_{p,n}$ using a weather derivative on the prediction error of the wind speed. To this end, define the following variables:

W_n: Wind speed at time n

$\hat{W}_n$: Prediction of W_n (which is computed, e.g., 1 day in advance)

Let $\epsilon_{w,n}$ be the prediction error of the wind speed, and assume that the payoff of the wind derivative is defined by using a suitable payoff function of $\epsilon_{w,n}$ as $\psi(\epsilon_{w,n})$. Also, suppose that the weather derivative contract with a payoff function $\psi(\cdot)$ is carried out in advance without any cost and that $\psi(\epsilon_{w,n})$ satisfies the following condition:

$$\mathrm{Mean}\left(\psi\left(\epsilon_{w,n}\right)\right) = 0. \tag{2.2}$$

Note that condition (2.2) indicates that the physical probability measure provides a risk neutral probability measure, and that, in the case of simple forward contracts, $\psi(\epsilon_{w,n})$ may be given as a linear function, e.g.,

$$\psi\left(\epsilon_{w,n}\right) = \epsilon_{w,n}. \tag{2.3}$$

2.2 Minimum variance hedge

With the notation and definitions introduced in the previous subsection, the first optimization problem, P1), is formulated as follows:

Contract volume optimization problem:

$$\min_{\Delta \in \Re} \operatorname{Var}\left(\phi\left(\epsilon_{p,n}\right) + \Delta \psi\left(\epsilon_{w,n}\right)\right). \tag{2.4}$$

The contract volume optimization problem may be considered as the standard "minimum variance hedge," and the optimal volume Δ^* may be computed analytically as

$$\Delta^* = -\frac{\operatorname{Cov}\left(\phi\left(\epsilon_{p,n}\right),\ \psi\left(\epsilon_{w,n}\right)\right)}{\operatorname{Var}\left(\psi\left(\epsilon_{w,n}\right)\right)}. \tag{2.5}$$

To estimate the hedge effect, we define the variance reduction rate (VRR) as follows:

$$\mathrm{VRR} := \frac{\operatorname{Var}\left(\phi\left(\epsilon_{p,n}\right) + \Delta^* \psi\left(\epsilon_{w,n}\right)\right)}{\operatorname{Var}\left(\phi\left(\epsilon_{p,n}\right)\right)}. \tag{2.6}$$

Because the minimum variance can be computed as

$$\begin{aligned} & \operatorname{Var}\left(\phi\left(\epsilon_{p,n}\right) + \Delta^* \psi\left(\epsilon_{w,n}\right)\right) \\ = \quad & \operatorname{Var}\left(\phi\left(\epsilon_{p,n}\right)\right)\left(1 - \left[\operatorname{Corr}(\phi\left(\epsilon_{p,n}\right),\ \psi\left(\epsilon_{w,n}\right))\right]^2\right), \end{aligned} \tag{2.7}$$

we obtain

$$\mathrm{VRR} = 1 - \left[\operatorname{Corr}(\phi\left(\epsilon_{p,n}\right),\ \psi\left(\epsilon_{w,n}\right))\right]^2. \tag{2.8}$$

Note that VRR satisfies

$$0 \leq \mathrm{VRR} \leq 1 \tag{2.9}$$

and that a smaller VRR provides a better hedge effect in terms of minimum variance.

In the case of standard minimum variance hedge, the optimal volume is also found by solving a linear regression problem, where $\phi\left(\epsilon_{p,n}\right)$ is regressed with respect to $\psi\left(\epsilon_{w,n}\right)$, and the regression coefficient gives the optimal volume for fixed loss and payoff functions. On the other hand, we can expect to obtain a better hedge effect if we could optimize the payoff function of the weather derivative directly. This can be done by applying non-parametric regression techniques introduced in the next section, and we will find that using a non-parametric regression corresponds to optimizing the derivative contract directly by choosing a suitable payoff function.

3 Minimum variance hedging using non-parametric regression

In this section, we first introduce a non-parametric regression technique, and then formulate the second optimization problem, P2).

In the previous section, we showed that the contract volume optimization problem is formulated as standard minimum variance hedging and can be solved by applying linear regression. A similar idea may be employed to solve the payoff function optimization problem of P2) (or the loss function optimization problem of P3)) by introducing a non-parametric regression technique. Since we will apply a non-parametric regression to find a payoff function (or loss function) by assuming that a loss function (or payoff function) is fixed, it may be useful to specify which function is given explicitly. To this end, we use overlines as

$$\phi(\cdot) = \overline{\phi}(\cdot) \quad (\text{or } \psi(\cdot) = \overline{\psi}(\cdot))$$

to indicate that the loss function (or payoff function) is given.

3.1 Generalized additive models

The non-parametric regression technique introduced here is to find a (cubic) smoothing spline that minimizes the so-called penalized residual sum of squares (PRSS) among all regression spline functions with two continuous derivatives. Let y_n and x_n be dependent and independent variables, respectively, and express y_n as

$$y_n = h(x_n) + \epsilon_n, \quad \text{Mean}(\epsilon_n) = 0 \tag{3.1}$$

using a smooth function $h(\cdot)$ and residuals ϵ_n. Here the function $h(\cdot)$ is a (cubic) smoothing spline that minimizes the following PRSS,

$$\text{PRSS} = \sum_{n=1}^{N} (y_n - h(x_n))^2 + \lambda \int_{-\infty}^{\infty} \left(h''(x)\right)^2 \mathrm{d}x \tag{3.2}$$

among all functions $h(\cdot)$ with two continuous derivatives, where λ is a given parameter. In (3.2), the first term measures closeness to the data while the second term penalizes curvature in the function. Note that, if $\lambda = 0$ and $h(\cdot)$ is given by a polynomial function, the problem is reduced to the standard regression polynomial and is solved by the least squares method. It is shown that (3.2) has an explicit and unique minimizer and that a candidate of optimal λ may be found by using the so-called generalized cross validation criteria (See Appendix A). Note that regression splines can be extended to the multivariable case with additive sums of smoothing splines, known as generalized additive models (GAMs; see e.g., [7]). Also note that GAMs can be computed using free software "R (`http://cran.r-project.org/`)," and we will refer to the class of smoothing splines for non-parametric regression as GAMs in this paper. We will apply GAMs to solve P2)–P4) and estimate the hedge effect of wind derivatives.

Note that, instead of writing the problem as an unconstrained optimization problem, we can reformulate it as an optimization problem constrained on $h(\cdot)$ as follows:

$$\begin{aligned} \min_{h(\cdot)} \quad & \sum_{n=1}^{N} (y_n - h(x_n))^2 \\ \text{s.t.} \quad & \int_{-\infty}^{\infty} \left(h''(x)\right)^2 \mathrm{d}x \le \alpha \end{aligned} \tag{3.3}$$

where α is a given parameter. Based on the similar argument to that in Appendix A, we can verify that the objective function of problem (3.3) is quadratic subject to a convex constraint and that the minimization problem (3.3) is equivalent to the following problem,

$$\max_{\lambda > 0} \left\{ \min_{h(\cdot)} \left\{ \sum_{n=1}^{N} \{y_n - h(x_n)\}^2 + \lambda \left(\int \left\{h''(x)\right\}^2 \mathrm{d}x - \alpha \right) \right\} \right\}, \tag{3.4}$$

using a Lagrange multiplier $\lambda > 0$. Therefore, we see that fixing λ in (3.2) corresponds to fixing α in (3.3) and that the non-parametric regression problem using GAM may be recast as a minimization problem of the sample variance with a smooth constraint.

3.2 Optimization of derivative contracts

It is in a position to formulate the the second optimization problem, i.e., the payoff function optimization problem, in the context of minimum variance hedge using non-parametric regression as follows:

Payoff function optimization problem:

$$\begin{aligned} \min_{\psi(\cdot)} \quad & \mathrm{Var}\left(\overline{\phi}\left(\epsilon_{p,n}\right) + \psi\left(\epsilon_{w,n}\right)\right) \\ \text{s.t.} \quad & \int_{-\infty}^{\infty} \left(\psi''(x)\right)^2 \mathrm{d}x \leq \alpha. \end{aligned} \tag{3.5}$$

The minimization problem (3.5) may be recast as (3.3) by taking $y_n = \overline{\phi}\left(\epsilon_{p,n}\right)$, $x_n = \epsilon_{w,n}$, and $h\left(\cdot\right) = -\psi\left(\cdot\right)$, and therefore, can be solved by applying GAM. Let $\psi^*(\cdot)$ be the optimal payoff function. Then VRR may be defined as

$$\mathrm{VRR} := \frac{\mathrm{Var}\left(\overline{\phi}\left(\epsilon_{p,n}\right) + \psi^*\left(\epsilon_{w,n}\right)\right)}{\mathrm{Var}\left(\overline{\phi}\left(\epsilon_{p,n}\right)\right)}. \tag{3.6}$$

Although it is possible to find the optimal payoff function by solving GAM once, it may be worthwhile to mention that we have a slight improvement by applying a linear regression after finding the optimal payoff function $\psi^*(\cdot)$ as

$$\min_{a \in \Re} \mathrm{Var}\left(\overline{\phi}\left(\epsilon_{p,n}\right) + a\psi^*\left(\epsilon_{w,n}\right)\right). \tag{3.7}$$

In this case, VRR may be given as

$$\mathrm{VRR} = \frac{\mathrm{Var}\left(\overline{\phi}\left(\epsilon_{p,n}\right) + a^*\psi^*\left(\epsilon_{w,n}\right)\right)}{\mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right)}. \tag{3.8}$$

or equivalently,

$$\mathrm{VRR} = 1 - \left[\mathrm{Corr}(\overline{\phi}\left(\epsilon_{p,n}\right),\ \psi^*\left(\epsilon_{w,n}\right))\right]^2. \tag{3.9}$$

where $a^* \in \Re$ is the regression coefficient to solve (3.7). Note that (3.9) is independent of a^*, or any scaling parameter to $\psi^*\left(\epsilon_{w,n}\right)$, and that it can be computed if $\psi^*\left(\cdot\right)$ is specified. Therefore, we use the right hand side of (3.9) as a proxy of VRR. It is readily confirmed that VRR in (3.6) is actually an upper bound of (3.9). However, as indicated in the end of Subsection 5.2, the gap between (3.6) and (3.9) is very small from our numerical experience.

4 Optimization with loss functions and simultaneous optimization

4.1 Optimal loss function

Next, we will consider a case in which a payoff function of wind derivative is given but we would like to find a loss function that is desirable for using the wind derivative, i.e., in a case where there already exists a standardized derivative contract with a certain payoff function, but there is some room for improvement on the loss function, e.g., for a WF owner. We assume that possible losses on $\epsilon_{p,n}$, $\phi(\epsilon_{p,n})$, has the same mean and variance, i.e., $\phi(\epsilon_{p,n})$ satisfies

$$\begin{aligned} \mathrm{Mean}\left(\phi\left(\epsilon_{p,n}\right)\right) &= 0, \qquad (4.1)\\ \mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) &= c. \qquad (4.2) \end{aligned}$$

We will compute an optimal loss function satisfying (4.2).

The loss function optimization problem is formulated as follows:

Loss function optimization problem:

$$\begin{aligned} \min_{\phi(\cdot)} \quad & \mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right) + \overline{\psi}\left(\epsilon_{w,n}\right)\right) \\ \text{s.t.} \quad & \int_{-\infty}^{\infty} \left(\phi''(x)\right)^2 \mathrm{d}x \le \alpha, \qquad (4.3) \\ & \mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) = c. \end{aligned}$$

Note that the constraint $\mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) = c$ is also quadratic if ϕ is given by a cubic natural spline function, and hence, the problem might be reformulated as an unconstrained optimization problem by introducing another Lagrangian term for the variance constraint. On the other hand, we can still apply GAM directly to solve the problem without the variance constraint (i.e., $\mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) = c$), similar to the payoff function optimization problem (3.5). Then we can scale the minimizing function so that it satisfies the variance constraint (4.2).

Let $\hat{\phi}(\cdot)$ be the optimizer of problem (4.3) without the variance constraint (i.e., $\mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) = c$), which can be computed by applying GAM. By scaling $\hat{\phi}(\cdot)$ to satisfy (4.2), we obtain the optimal loss function $\phi^*(\cdot)$ as follows:

$$\phi^*(\cdot) = \frac{c}{\mathrm{Var}\left(\hat{\phi}\left(\epsilon_{p,n}\right)\right)} \hat{\phi}(\cdot). \qquad (4.4)$$

Note that the optimal volume of wind derivative with the given payoff and loss functions, $\overline{\psi}(\cdot)$ and $\phi^*(\cdot)$, will be found by solving the standard minimum variance hedging problem as in Subsection 2.2, and VRR may be computed as

$$\mathrm{VRR} = 1 - \left[\mathrm{Corr}\left(\phi^*\left(\epsilon_{p,n}\right),\ \overline{\psi}\left(\epsilon_{w,n}\right)\right)\right]^2. \qquad (4.5)$$

4.2 Simultaneous optimization

It may be interesting to consider a simultaneous optimization of the payoff and loss functions, $\psi(\epsilon_{w,n})$ and $\phi(\epsilon_{p,n})$. Recall that VRR can be computed using the correlation between the payoff function and the loss function as

$$1 - \left[\mathrm{Corr}\left(\phi\left(\epsilon_{p,n}\right), \psi\left(\epsilon_{w,n}\right)\right)\right]^2.$$

Since the larger correlation the smaller VRR, the minimization of VRR boils down to the maximization of correlation between $\phi\left(\epsilon_{p,n}\right)$ and $\psi\left(\epsilon_{w,n}\right)$. Therefore, the simultaneous optimization of the payoff and the loss functions may be formulated as follows:

Simultaneous optimization problem:

$$\begin{aligned} \max_{\phi(\cdot),\psi(\cdot)} \quad & \mathrm{Corr}\left(\phi\left(\epsilon_{p,n}\right), \psi\left(\epsilon_{w,n}\right)\right) \\ \text{s.t.} \quad & \int_{-\infty}^{\infty} \left(\phi''(x)\right)^2 \mathrm{d}x \leq \alpha_\phi, \\ & \int_{-\infty}^{\infty} \left(\psi''(x)\right)^2 \mathrm{d}x \leq \alpha_\psi, \\ & \mathrm{Var}\left(\phi\left(\epsilon_{p,n}\right)\right) = c. \end{aligned} \tag{4.6}$$

The simultaneous optimization problem may be solved using an iterative algorithm by solving the payoff function optimization problem with $\phi(\cdot) = \overline{\phi}(\cdot)$ fixed, or the loss function optimization problem with $\psi(\cdot) = \overline{\psi}(\cdot)$ fixed, at each step. The following is the iterative algorithm:

Iterative algorithm:

1. Given $\phi(\cdot) = \overline{\phi}(\cdot)$, find $\psi(\cdot)$ to solve the payoff function optimization problem. Let $\psi^*(\cdot)$ be the optimal function, and let $\overline{\psi}(\cdot) = \psi^*(\cdot)$.
2. Given $\psi(\cdot) = \overline{\psi}(\cdot)$, find $\phi(\cdot)$ to solve the loss function optimization problem. Let $\phi^*(\cdot)$ be the optimal loss function and let $\overline{\phi}(\cdot) = \phi^*(\cdot)$.
3. Repeat Steps 2 and 3 until the objective function in (4.6) does not change.

Note that the optimal loss function obtained from the above iterative algorithm satisfies (4.2) and that we can consider additional constraints to take more realistic situations into account for the loss and payoff functions.

Remark 1 *The above iterative algorithm is formally in the class of so-called "Alternating Conditional Expectations (ACE) algorithm (see, e.g., Chapter 7 of [7])." The ACE algorithm seeks optimal transformations of $\theta(Y)$ and $f(X)$ for two random variables X and Y so that the squared error loss*

$$\mathbb{E}\left[\left(\theta(Y) - f(X)\right)^2\right]$$

is minimized. Since the zero functions trivially minimize the square error, ACE has a constraint so that $\theta(Y)$ has unit variance at each step, which is exactly the same as our variance constraint (4.2).

Note that the convergence of ACE algorithm is also discussed in [7], although we omit the details for brevity. Also note that, for solving the iterative algorithm, we may need to specify α_θ and α_ϕ. However, in stead of fixing these parameters a priori in the algorithm, an optimal selection of smoothing parameters for $\phi(\cdot)$ and $\psi(\cdot)$ may be applicable at each step by using GAMs (See Appendix A).

5 Empirical analysis and numerical experiment

In this section, we demonstrate the solutions P1)–P4) and estimate their hedge effect using empirical data for the power output, wind speed, and their predictions. Here we consider the power output from a wind farm (WF) located in Japan, where the power output from the WF is predicted based on the numerical weather prediction and the power generating properties for turbines. The numerical weather prediction consists of the following two steps:

- Japan Meteorological Agency announces the hourly data of regional spectral models for the next 51 hours twice a day (9am and 9pm).
- Using them as initial and boundary values, a public weather forecasting company computes more sophisticated values for the next day's hourly data by 12pm.

5.1 Preliminary

5.1.1 Data description

In this paper, we use the prediction data obtained from the Local Circulation Assessment and Prediction System (LOCALS) developed by the ITOCHU Techno-Solutions Corporation for the wind speed and the power output of a wind farm in Japan [5]. The data set is given as follows:[1]

Data specifications:
Realized and predicted values of total power output for the WF, and those of wind speed for the observation tower in the WF.

Data period:
2002–2003 (1 year), hourly data, everyday

Total number of data:
8,000 for each variable excluding missing values

Let $n = 1, \ldots, N$ be the time index (where $N \simeq 8,000$), and assume that the actual power output and the wind speed at time n are, respectively, denoted by P_n and W_n. Also, let $\hat{P}_n$ and $\hat{W}_n$ be the predictions of the corresponding power output and the wind speed obtained from LOCALs, which are computed by noon one day before the actual data is observed. Fig. 5.1 shows a scatter diagram for the wind speed W_n and the power output P_n, where the power output P_n is normalized so that its maximum equals 100. From Fig. 5.1, we can see that:

- The generator starts providing the power output when the wind speed exceeds around 2 [m/s].
- The power output increases with the wind speed between 5–15 [m/s].

Also note that, because each electricity generator is controlled so that the maximum output does not exceed a certain value, the total output is also bounded as shown in Fig. 5.1.

[1] All the data used in this paper were provided by ITOCHU Techno-Solutions Corporation.

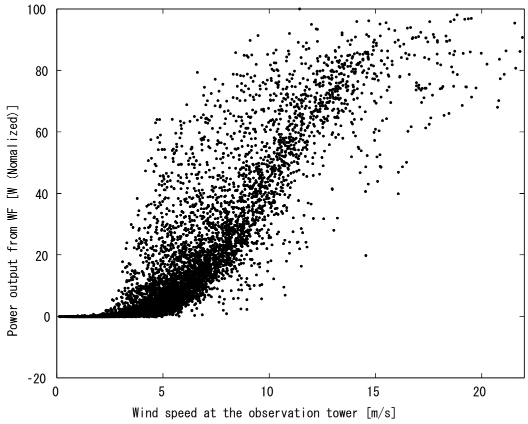

Figure 5.1: Wind speed W_n [m/s] vs. Power output P_n [W]

5.1.2 Prediction error of the wind speed

Fig. 5.2 shows a partial residual plot for

$$W_n = a_w \hat{W}_n + b_w + \epsilon_{w,n}, \quad n = 0, \ldots, N, \ \mathrm{Mean}\,(\epsilon_{w,n}) = 0 \tag{5.1}$$

i.e., the scatter diagram of $\left(\hat{W}_n,\ W_n - b_w\right)$, where a_w and b_w are a regression coefficient and intercept, respectively, and $\epsilon_{w,n}$ is a residual satisfying $\mathrm{Mean}\,(\epsilon_{p,n}) = 0$. The partial regression line is depicted using a solid straight line shown in Fig. 5.2. In this case, the sample variance of residuals is found to be

$$\mathrm{Var}\,(\epsilon_{w,n}) \simeq 5.12. \tag{5.2}$$

On the other hand, the regression spline $f(\cdot)$ to fit the same data of Fig. 5.2 is shown as a solid line in Fig. 5.3, where $f(\cdot)$ satisfies

$$W_n = f\big(\hat{W}_n\big) + \epsilon_{w,n}. \tag{5.3}$$

using GAM. In this case, the sample variance of the residuals is

$$\mathrm{Var}\,(\epsilon_{w,n}) \simeq 4.95 \tag{5.4}$$

Noting that the sample variance of the measured values is computed as "11.0," we can say that the variance of the wind speed is reduced by 50% (from "11.0" to "5.12") using the predicted value and the linear regression, and it is improved a little using GAM, i.e., from "5.12" to "4.95." In this section, we define the prediction error of the wind speed as the one given by GAMs, i.e., $\epsilon_{w,n}$ in (5.3).

5.1.3 Prediction error of the power output

Similarly, we can draw a partial residual plot for the power output P_n with respect to the predicted value $\hat{P}_n$ as shown in Fig. 5.4, where the solid line is obtained from a linear regression for partial

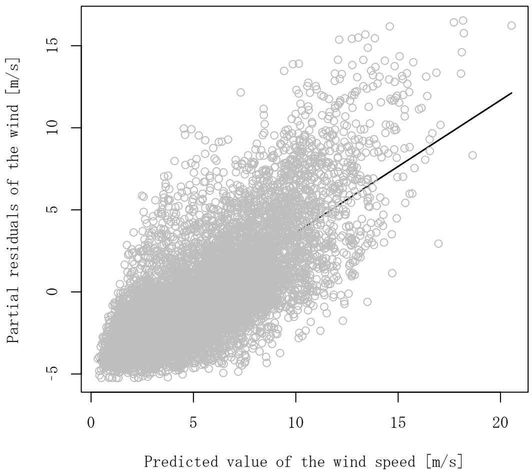

Figure 5.2: Predicted vs. Measured values for the wind speed

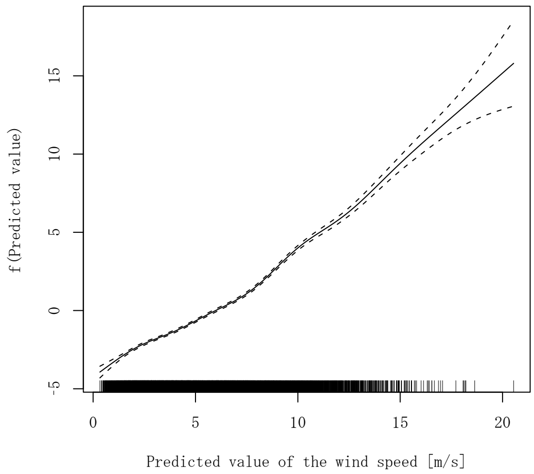

Figure 5.3: Spline regression function for the wind speed

residuals. In this case, the sample variance of the residuals is found to be "249." The solid line in Fig. 5.5 refers to the regression spline function $g(\cdot)$ satisfying

$$P_n = g\big(\hat{P}_n\big) + \epsilon_{p,n}, \quad n = 0, \dots, N \tag{5.5}$$

using GAM. Note that the sample variance of residuals in this case is given as "239," whereas the sample variance of the measured value of the power output is "504." Similar to the wind speed case, we can say that the variance of the wind speed is reduced to less than half (from "504" to "249") using the predicted value and the linear regression, and it is improved a little using GAM, i.e., "249" to "239."

Although we should be able to define the prediction error of the power output using the residual in (5.5), it might be worthwhile to mention that there is another way to define the prediction error of the power output. As stated in the beginning of this section, the power output is predicted using numerical weather prediction, and therefore, we can define a regression model such that the power output P_n is a dependent variable and the wind speed prediction $\hat{W}_n$ is an independent variable, i.e.,

$$P_n = h\big(\hat{W}_n\big) + \epsilon_{p,n}, \tag{5.6}$$

where $h(\cdot)$ is a regression spline that minimizes PRSS.

Fig. 5.6 shows the relation between the predicted values of the wind speed and the measured values for the power output, where the solid line in Fig. 5.7 is the regression spline $h(\cdot)$. In this case, the sample variance of the residuals is computed as

$$\mathrm{Var}\left(\epsilon_{p,n}\right) \simeq 254 \tag{5.7}$$

which is, in fact, higher than the one given by (5.5). However, it will turn out that using the prediction error in (5.6) provides not only a better hedge effect but also a smaller variance of the hedged loss

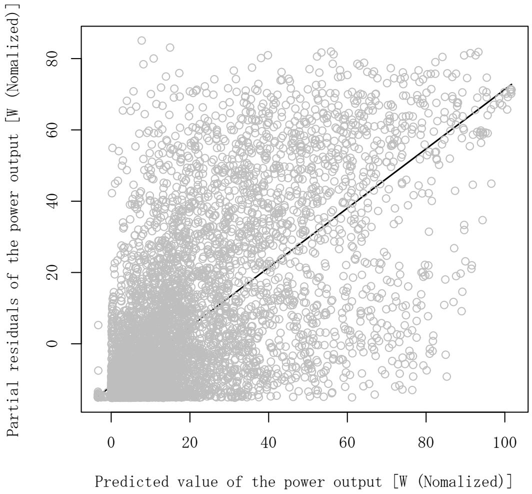

Figure 5.4: Predicted vs. Measured values for the power output

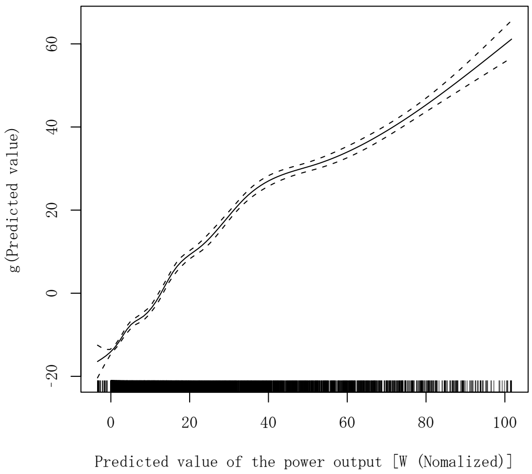

Figure 5.5: Spline regression function for the power output

when combining with the optimal wind derivative. Therefore, we will use the residual $\epsilon_{p,n}$ in (5.6) to define the prediction error of the power output. An empirical analysis using the prediction error defined by the residual in (5.5) may be found in [13].

5.2 Construction of wind derivatives and their hedge effect

Next, we will construct wind derivatives and demonstrate their hedge effect on wind power energy businesses.

5.2.1 Linear function's case

We first solve the minimum variance hedging problem for the simplest case where the loss and the payoff functions are both linear. Let

$$\phi\left(\epsilon_{p,n}\right) = \epsilon_{p,n}, \quad \psi\left(\epsilon_{w,n}\right) = \epsilon_{w,n} \tag{5.8}$$

without loss of generality. In this case, the problem is reduced to solving a linear regression for the following regression function:

$$\epsilon_{p,n} = a_w \epsilon_{w,n} + \eta_n, \tag{5.9}$$

where η_n is a residual. Since the linear regression computes a_w that minimizes variance of $\eta_n = \epsilon_{p,n} - a_w \epsilon_{w,n}$, the regression coefficient provides the optimal volume as

$$\Delta^* = -a_w \tag{5.10}$$

in the problem (2.4) under condition (5.8), where

$$a_w = \frac{\mathrm{Cov}\left(\epsilon_{p,n},\ \epsilon_{w,n}\right)}{\mathrm{Var}\left(\epsilon_{w,n}\right)}. \tag{5.11}$$

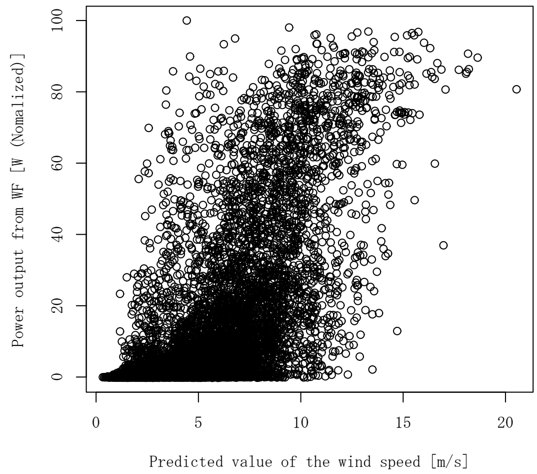

Figure 5.6: Predicted value of the wind speed vs. Measured values for the power output

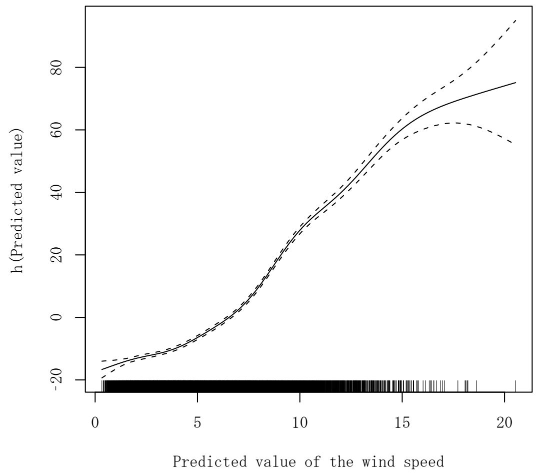

Figure 5.7: Spline regression function for the power output using the wind speed prediction

Fig. 5.8 shows a scatter plot of $\epsilon_{w,n}$ vs. $\epsilon_{p,n}$ with a linear regression line. The sample correlation is computed as

$$\mathrm{Corr}\left(\epsilon_{p,n},\ \epsilon_{w,n}\right) \simeq 0.76. \tag{5.12}$$

and VRR as

$$\mathrm{VRR} = 1 - \mathrm{Corr}(\epsilon_{p,n},\ \epsilon_{w,n})^2 \simeq 0.43. \tag{5.13}$$

We see that the prediction errors of the wind speed and the power output, $\epsilon_{w,n}$ and $\epsilon_{p,n}$, are highly correlated and that the sample variance is reduced to 43% from the original one using the wind derivative in the case where the loss and the payoff functions are both linear.

Now, we apply GAMs to compute an optimal payoff function. The solid line in Fig. 5.9 shows the optimal payoff curve obtained by solving the optimization problem (3.5) when $\phi(\cdot)$ is linear. In this case, the VRR is computed as

$$\mathrm{VRR} = \frac{\mathrm{Var}\left(\epsilon_{p,n} + \psi^*\left(\epsilon_{w,n}\right)\right)}{\mathrm{Var}\left(\epsilon_{p,n}\right)} \simeq 0.407. \tag{5.14}$$

where $\psi^*(\cdot)$ is the optimal payoff function. Moreover, the variance of the hedged loss $\epsilon_{p,n} + \psi^*\left(\epsilon_{w,n}\right)$ is computed as

$$\mathrm{Var}\left(\epsilon_{p,n} + \psi^*\left(\epsilon_{w,n}\right)\right) \simeq 103. \tag{5.15}$$

The above variance is actually lower than that of the hedged loss using (5.5) with the optimal wind derivative, which is computed as "119." Therefore, we see that, even though the variance of the original loss might be larger, it can be reduced more effectively by combining it with the wind derivative if we define the prediction error by (5.6) instead of (5.5).

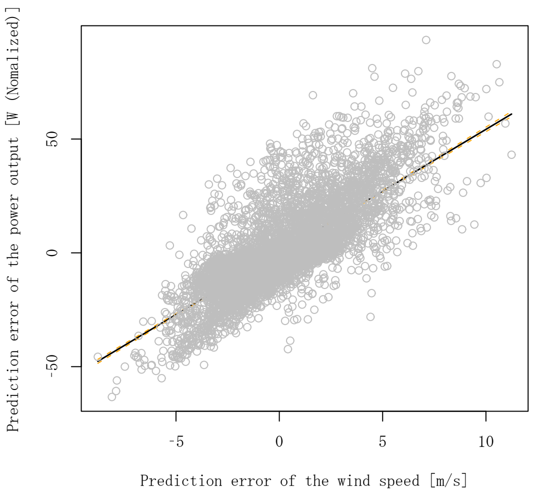

Figure 5.8: Wind speed prediction error ($\epsilon_{w,n}$) vs. Power output prediction error ($\epsilon_{p,n}$)

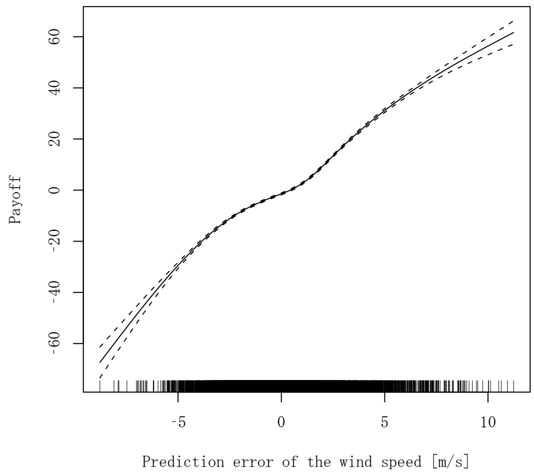

Figure 5.9: Optimal payoff function on $\epsilon_{w,n}$

5.2.2 Piecewise linear function's case

Next, we will consider the case in which the loss function $\phi(\cdot) = \overline{\phi}(\cdot)$ is given as shown in Fig. 1.1 with zero mean constraint (2.1), i.e.,

$$\overline{\phi}\left(\epsilon_{p,n}\right) := 4\left|\epsilon_{p,n}\right|^{+} + 10\left|\epsilon_{p,n}\right|^{-} - \mu \tag{5.16}$$

where

$$\mu := \mathrm{Mean}\left(4\left|\epsilon_{p,n}\right|^{+} + 10\left|\epsilon_{p,n}\right|^{-}\right).$$

and $|\cdot|^{+}$ and $|\cdot|^{-}$ are defined as

$$|x|^{+} := \max\left(x,\ 0\right), \quad |x|^{-} := \min\left(x,\ 0\right)$$

for $x \in \Re$. The solid line in Fig. 5.10 shows the optimal payoff function to solve the problem (3.5). In this case, VRR in (3.6) is computed as

$$\mathrm{VRR} = 0.5461946\cdots \tag{5.17}$$

whereas the right hand side of (3.9) is found to be

$$1 - \left[\mathrm{Corr}(\phi\left(\epsilon_{p,n}\right),\ \psi^{*}\left(\epsilon_{w,n}\right))\right]^{2} = 0.5461927\cdots. \tag{5.18}$$

From this example, we see that VRR can be approximated by (3.9) with high accuracy.

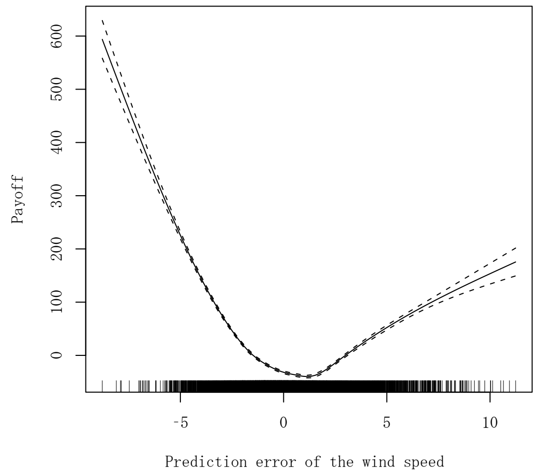

Figure 5.10: Optimal payoff function on the wind speed prediction error $\epsilon_{w,n}$

5.3 Optimal loss function and simultaneous optimization

In this subsection, we first provide an illustrative example of solving P3) to compute an optimal loss function, and then solve the simultaneous optimization problem of P4).

Since the linear correlation between $\epsilon_{p,n}$ and $\epsilon_{w,n}$ is high in this example, it would be more interesting to consider the case where a payoff function is non-linear with respect to $\epsilon_{w,n}$. Therefore, we assume that there already exists a derivative contract with the payoff being proportional to the size of the wind speed prediction error, $|\epsilon_{w,n}|$. Noting that $\psi(\epsilon_{w,n})$ satisfies (2.2), such a payoff function may be given as

$$\psi(\epsilon_{w,n}) = \overline{\psi}(\epsilon_{w,n}) := |\epsilon_{w,n}| - \mathrm{Mean}(|\epsilon_{w,n}|), \tag{5.19}$$

Fig. 5.11 shows the payoff function with respect to $\epsilon_{w,n}$ given in (5.19).

Now we will solve P3) with the given payoff function in (5.19). Assume that the sample variance of the loss, $\phi(\epsilon_{p,n})$, satisfies

$$\mathrm{Var}(\phi(\epsilon_{p,n})) = \mathrm{Var}(\epsilon_{p,n}) \tag{5.20}$$

and we solve the problem (4.3) with the assumption that the optimal loss function satisfies the above variance constraint. The solid line in Fig. 5.12 shows the optimal loss function, which is obtained by applying GAM and scaling the minimizing function as in (4.4). In this case, VRR is found to be

$$\mathrm{VRR} \simeq 0.56. \tag{5.21}$$

Next, we demonstrate the simultaneous optimization of P4). Here we also introduce a nonlinearity using the absolute value of $\epsilon_{w,n}$. Assume that the payoff of the wind derivative is a function of $|\epsilon_{w,n}|$, and consider a maximization problem of

$$\mathrm{Corr}(\phi(\epsilon_{p,n}), \psi(|\epsilon_{w,n}|)). \tag{5.22}$$

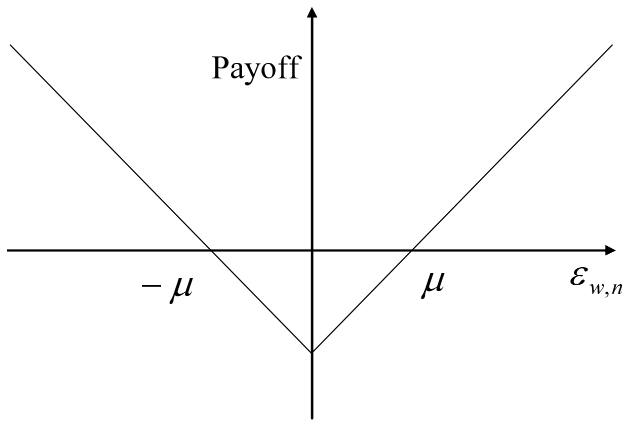

Figure 5.11: Given payoff function with respect to the wind speed prediction error $\epsilon_{w,n}$

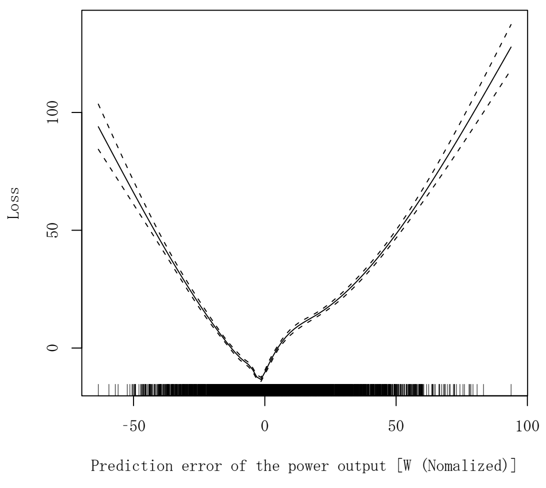

Figure 5.12: Optimal loss function on the power output prediction error $\epsilon_{p,n}$

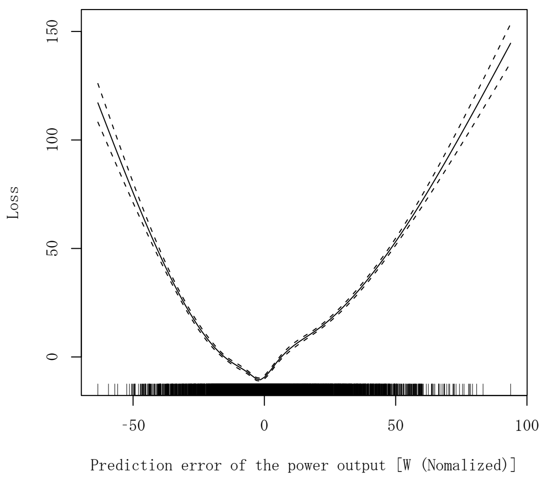

Figure 5.13: Optimal loss function after the 8th iteration

We apply the iterative algorithm for a fixed loss function $\phi(\cdot)$ or a fixed payoff function $\psi(\cdot)$ at each step to maximize (5.22). Assume that the payoff function is initially set to the one given in (5.19) and we solve the loss function optimization problem. The initial loss function in this case is given by the one shown in Fig. 5.12. We repeatedly apply Steps 1 and 2 in the iterative algorithm until the objective function does not change or the relative change of the values of the objective function is less than a sufficiently small number. In this example, we obtained

$$\mathrm{VRR} = 0.53, \tag{5.23}$$

after the 8th iteration. Fig. 5.13 shows the optimal loss function after the 8th iteration, where the loss function is scaled to satisfy the variance constraint (5.20). We see that the loss function became smoother compared to the one given in Fig. 5.12.

6 Extension to multi-period case

In the previous sections, we have implicitly assumed that the same contract volume and/or payoff functions may be used at each time period in a day, i.e., Δ is independent of time period in a day. On the other hand, if we carry out a wind derivatives contract based on the multiple time periods in a day (e.g., the prediction errors for 1pm, 2pm, and 3pm), we might need to consider a different contract volume for each time period. Since these volume may also depend on the correlation between prediction errors of different time periods, another new problem formulation in a multi-period framework is required. With this consideration in mind, here we consider the multi-period case where the wind derivative contract is carried out based on prediction errors of multiple time periods.

We will introduce a special notation to define the multi-period case as follows, where n denotes a daily time index and i an hourly time index satisfying $i = 0, 1, \ldots, 23$.

$P_n^{(i)}$: The average power output between i o'clock and $i+1$ o'clock at day n

$\hat{P}_n^{(i)}$: Prediction of $P_n^{(i)}$, which is computed by 12 o'clock at day $n-1$.

$W_n^{(i)}$: The average wind speed between i o'clock and $i+1$ o'clock at day n

$\hat{W}_n^{(i)}$: Prediction of $W_n^{(i)}$, which is computed by 12 o'clock at day $n-1$.

6.1 Contract volume optimization problem of wind derivatives

Let $\epsilon_{p,n}^{(i)}$ and $\epsilon_{w,n}^{(i)}$ be prediction errors of the power output and the wind speed (with day and time indices, n and i), respectively. Here the prediction errors may be computed similar to the previous sections. For instance, $\epsilon_{w,n}^{(i)}$ is obtained by regressing $W_n^{(i)}$ with respect to $\hat{W}_n^{(i)}$, e.g.,

$$W_n^{(i)} = a_w^{(i)} \hat{W}_n^{(i)} + b_w^{(i)} + \epsilon_{w,n}^{(i)} \tag{6.1}$$

where $a_w^{(i)}$ and $b_w^{(i)}$ are regression coefficients.

Also, let $\phi^{(i)}(\cdot)$ and $\psi^{(i)}(\cdot)$ be loss and payoff functions which define the loss on the prediction error $\epsilon_{p,n}^{(i)}$ of the power output and the payoff of wind derivative on $\epsilon_{w,n}^{(i)}$, respectively. Then, we can formulate the contract volume optimization problem for the multi-period case as follows, where s is a start time and u is an end time of the contract satisfying $0 \le s \le u \le 23$:

Contract volume optimization problem (multi-period case):

$$\min_{\Delta_s,\dots,\Delta_u} \ \mathrm{Var}\left[\sum_{i=s}^{u} \phi_i\left(\epsilon_{p,n}^{(i)}\right) + \sum_{i=s}^{u} \Delta_i \psi_i\left(\epsilon_{w,n}^{(i)}\right)\right] \tag{6.2}$$

Note that the contract volume optimization problem may be solved by applying the linear multiple regression to find the optimal volume

$$\Delta_i^* \in \Re, \ i \in \{s, \dots, u\}.$$

Similarly, we can formulate the payoff function optimization problem by applying GAM.

6.2 Autocorrelation of the errors and illustrative example

At first, we examine the daily auto-correlation of wind speed prediction errors. Let us consider the wind speed in the period of $i = 12, \ 13, \ 14, \ 15$. In these periods, we can suppose that the electricity consumption would be maximum in a day. Given the same data set as in Section 5, we compute the prediction errors of the wind speed for $i = 12, \ 13, \ 14, \ 15$ using the linear regression as follows,

$$W_n^{(i)} = a_w^{(i)} \hat{W}_n^{(i)} + b_w^{(i)} + \epsilon_{w,n}^{(i)}, \quad i = 12, \ 13, \ 14, \ 15. \tag{6.3}$$

where $a_w^{(i)}$ and $b_w^{(i)}$ are regression coefficients and are given as

$$a_w^{(12)} = 0.872, \quad a_w^{(13)} = 0.881, \quad a_w^{(14)} = 0.880, \quad a_w^{(15)} = 0.844$$
$$b_w^{(12)} = 0.794, \quad b_w^{(13)} = 0.771, \quad b_w^{(14)} = 0.677, \quad b_w^{(15)} = 0.793.$$

Figs. 6.1–6.4 show the (daily) autocorrelation functions (ACFs) for $\epsilon_{w,n}^{(12)}$, $\epsilon_{w,n}^{(13)}$, $\epsilon_{w,n}^{(14)}$ and $\epsilon_{w,n}^{(15)}$, where the dashed lines denote the 95% confidence intervals. From these figures, we see that the autocorrelations of the prediction errors for wind speed are very small and within the confidence intervals (except for lags 8 of $\epsilon_{w,n}^{(13)}$ and 1 of $\epsilon_{w,n}^{(14)}$).

We solve the contract volume optimization problem for the simplest case in which the loss and the payoff functions are given as

$$\phi\left(\epsilon_{p,n}^{(i)}\right) = \epsilon_{p,n}^{(i)}, \quad \phi\left(\epsilon_{w,n}^{(i)}\right) = \epsilon_{w,n}^{(i)}, \quad i = 12, \ 13, \ 14, \ 15.$$

In this case, the total loss is defined by the sum of $\epsilon_{p,n}^{(i)}$ and the problem can be rewritten as follows:

$$\min_{\Delta_{12},\dots,\Delta_{15}\in\Re} \ \mathrm{Var}\left[L_n + \Delta_{12}\epsilon_{w,n}^{(12)} + \Delta_{13}\epsilon_{w,n}^{(13)} + \Delta_{14}\epsilon_{w,n}^{(14)} + \Delta_{15}\epsilon_{w,n}^{(15)}\right], \quad L_n := \sum_{i=12}^{15} \epsilon_{p,n}^{(i)} \tag{6.4}$$

Here we computed the prediction errors of the power outputs, $\epsilon_{p,n}^{(i)}$, by using the linear regression,

$$P_n^{(i)} = a_p^{(i)} \hat{W}_n^{(i)} + b_p^{(i)} + \epsilon_{p,n}^{(i)}, \quad i = 12, \ 13, \ 14, \ 15. \tag{6.5}$$

similar to the numerical experiments in Section 5. After solving the problem (6.4), we obtained

$$\mathrm{VRR} = 0.414. \tag{6.6}$$

Note that we can extend the above results to address more sophisticated loss and payoff functions, although we omit the details for brevity.

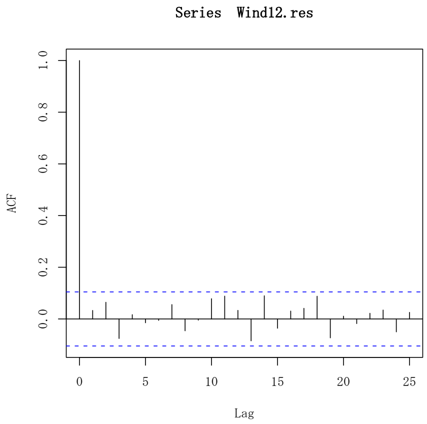

Figure 6.1: ACF for $\epsilon_{w,n}^{(12)}$

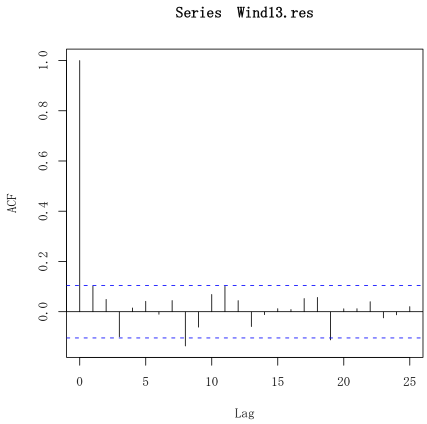

Figure 6.2: ACF $\epsilon_{w,n}^{(13)}$

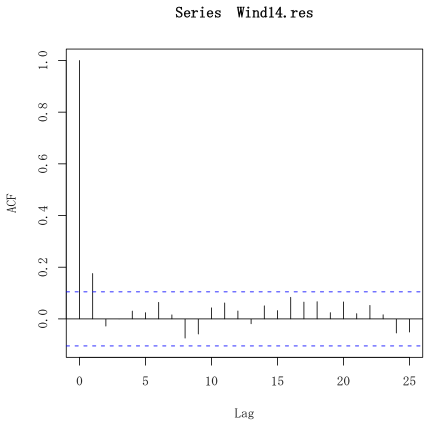

Figure 6.3: ACF for $\epsilon_{w,n}^{(14)}$

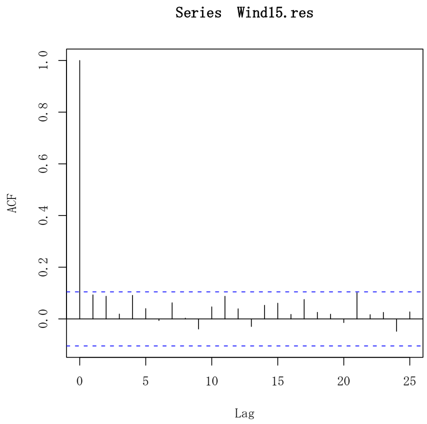

Figure 6.4: ACF for $\epsilon_{w,n}^{(15)}$

7 Concluding remarks

In this work, we have proposed a new type of weather derivatives based on the prediction errors for wind speeds and estimated their hedge effect on wind power energy businesses. At first, we explained some properties of the loss for a WF caused by prediction errors of the power output, and characterized it using a loss function on the error. We then formulated four types of optimization problems: 1) Contract volume optimization problem, 2) Payoff function optimization problem, 3) Loss function optimization problem, and 4) Simultaneous optimization problem. It was shown that the contract volume optimization problem may be reduced to the standard minimum variance hedge and is solved by applying linear regression. The idea of standard minimum variance hedging was generalized to the payoff function optimization problem by introducing a non-parametric regression technique based on smooth splines (or GAMs). We also showed that the loss function optimization problem may be solved by applying GAMs, and a simultaneous optimization technique of the loss and payoff functions for wind derivatives was demonstrated by applying GAMs iteratively. An empirical analysis and numerical experiments were performed to illustrate the hedge effect of the proposed wind derivatives.

The main contribution of this paper is summarized as follows:

- The paper is the first to provide a type of weather derivative contracts based on the prediction errors, which might be applicable for other situations (or businesses) and/or other indices such as temperature, rain falls, and so on.
- The paper provides an application of non-parametric regression techniques in the context of minimum variance hedge using smooth functions, which can be thought of a generalization of the standard minimum variance hedge based on linear regression.

Although we assumed that the payoff functions are just smooth, the approximation of these functions using the standard payoff functions for puts or calls may be required in practice when the standardized derivative contracts are only available. Also, the convergence of the iterative algorithm for simultaneous optimization is an important issue. These are interesting topics to be discussed further in the future work.

References

[1] D.C. Brody, J. Syroka, and M. Zervos, "Dynamical pricing of weather derivatives," Quantitative Finance, vol. 2, pp. 189–198, 2002.

[2] M. Cao and J. Wei, "Weather Derivatives Valuation and Market Price of Weather Risk," Journal of Futures Market, vol. 24, issue 11, pp. 1065–1089, 2004.

[3] M. Davis, "Pricing weather derivatives by marginal value," Quantitative Finance, vol. 1, pp. 305–308, 2001.

[4] Y. Emoto and T. Misawa, "Marginal value approach to pricing temperature options and its empirical example on daily temperatures in Nagoya," Discussion Papers in Economics No. 450, Nagoya City University , 2006.

[5] S. Enomoto, N. Inomata, T. Yamada, H. Chiba, R. Tanikawa, T. Oota, and H. Fukuda, "Prediction of power output from wind farm using local meteorological analysis," Proc. of European Wind Energy Conference, pp. 749–752, 2001.

[6] H. Geman, *Insurance and Weather Derivatives*, Risk Books, 1999.

[7] T. Hastie and R. Tibshirani, *Generalized Additive Models*, Chapman & Hall, 1990.

[8] T. Kariya (2003), "Weather Risk Swap Valuation," Working Paper, Institute of Economic Research, Kyoto University, Japan, 2003.

[9] E. Platen and J. West, "Fair Pricing of Weather Derivatives," Asia-Pacific Financial Markets, **11**(1), pp. 23–53, 2004.

[10] T. Takano, "Natural Energy Power and Energy Storing Technology," The transactions of the Institute of Electrical Engineers of Japan (B), vol. 126, no. 9, pp. 857–860, 2006 (in Japanese).

[11] K. Takezawa, *Introduction to Nonparametric Regression*, John Wiley, 2006.

[12] S.N. Wood, *Generalized Additive Models: An Introduction with R*, Chapman & Hall, 2006.

[13] Y. Yamada, "Optimal design of wind derivatives based on prediction errors," JAFEE Journal, vol. 7, pp. 152–181, 2008 (in Japanese).

[14] Y. Yamada, "Valuation and Hedging of Weather Derivatives on Monthly Average Temperature," Journal of Risk, vol. 10, no. 1, pp. 101–125, 2007.

[15] Y. Yamada, M. Iida, and H. Tsubaki, "Pricing of Weather Derivatives Based on Trend Prediction and Their Hedge Effect on Business Risks," Proceeding of the Institute of Statistical Mathematics, vol. 54, no. 1, pp. 57–78, 2006 (in Japanese).

Appendix

A Solution method and selection of smoothing parameters

It is known that the smoothing spline function that minimizes PRSS is given by a cubic natural spline of the following form [11, 12]:

$$h(x) = c_0 + c_1 x + \frac{1}{12} \sum_{n=1}^{N} w_n \left| x - x_n \right|^3, \tag{A.1}$$

where c_0, c_1, and $w_n,\ n = 1, \ldots, N$ are parameters to be found by minimizing PRSS for given $\lambda > 0$.

Let

$$\boldsymbol{y} := [y_1,\ y_2, \ldots, y_N]^\top \in \Re^N, \quad \boldsymbol{Q} := \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \in \Re^{N\times 2},$$

$$\boldsymbol{R} := \begin{bmatrix} 0 & \frac{|x_1-x_2|^3}{12} & \frac{|x_1-x_3|^3}{12} & \cdots & \frac{|x_1-x_N|^3}{12} \\ \frac{|x_2-x_1|^3}{12} & 0 & \frac{|x_2-x_3|^3}{12} & \cdots & \frac{|x_2-x_N|^3}{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{|x_N-x_1|^3}{12} & \frac{|x_N-x_2|^3}{12} & \frac{|x_N-x_3|^3}{12} & \cdots & 0 \end{bmatrix} \in \Re^{N\times N}.$$

Then PRSS in (3.2) may be recast as follows:

$$\mathrm{PRSS} = (\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{c} - \boldsymbol{R}\boldsymbol{w})^\top (\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{c} - \boldsymbol{R}\boldsymbol{w}) + \lambda \int_{-\infty}^{\infty} \left(h''(x)\right)^2 \mathrm{d}x, \tag{A.2}$$

where

$$\boldsymbol{c} := [c_0,\ c_1]^\top \in \Re^2, \quad \boldsymbol{w} := [w_1, \ldots, w_N]^\top \in \Re^N.$$

We see that the first term of the right hand side of equation (A.2) is quadratic with respect to $\boldsymbol{c} \in \Re^2$ and $\boldsymbol{w} \in \Re^N$.

Moreover, as shown in [11], the second term (related to the smoothing condition) may also be represented as a quadratic function, i.e.,

$$\lambda \int_{-\infty}^{\infty} \left(h''(x)\right)^2 \mathrm{d}x = \lambda \boldsymbol{w}^\top \boldsymbol{R}\boldsymbol{w}.$$

As a result, PRSS is given by

$$\mathrm{PRSS} = (\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{c} - \boldsymbol{R}\boldsymbol{w})^\top (\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{c} - \boldsymbol{R}\boldsymbol{w}) + \lambda \boldsymbol{w}^\top R\boldsymbol{w}. \tag{A.3}$$

Therefore, for any given $\lambda > 0$, the minimization of PRSS may be solved as a convex quadratic optimization problem.

For choosing the smoothing parameter λ, the cross validation criteria may be constructed by leaving points $(x_n,\ y_n)$ out one at a time and estimating the smooth at x_n based on the remaining $N-1$ points as

$$CV(\lambda) = \frac{1}{N}\sum_{n=1}^{N} \left(y_n - \hat{h}_\lambda^{-n}(x_n)\right)^2 \tag{A.4}$$

where $\hat{h}_\lambda^{-n}(x_n)$ indicates the fit at x_n, computed by leaving out the nth data point as shown in [7]. We can use $CV(\lambda)$ for searching the minimizing λ and set it as a candidate of optimal λ in the sense of cross validation. Note that in the algorithm implemented in "R," the so-called generalized cross validation criteria is used for computing an optimal λ more efficiently.