

Retraction

Retracted: Sublinear Expectation Nonlinear Regression for the Financial Risk Measurement and Management

Discrete Dynamics in Nature and Society

Received 14 November 2013; Accepted 14 November 2013

Copyright © 2013 Discrete Dynamics in Nature and Society. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper titled “Sublinear Expectation Nonlinear Regression for the Financial Risk Measurement and Management” [1], published in *Discrete Dynamics in Nature and Society*, has been retracted as it was submitted for publication by Yunquan Song without the knowledge and approval of the co-author Lu Lin.

References

- [1] Y. Song and L. Lin, “Sublinear expectation nonlinear regression for the financial risk measurement and management,” *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 398750, 10 pages, 2013.

Research Article

Sublinear Expectation Nonlinear Regression for the Financial Risk Measurement and Management

Yunquan Song^{1,2} and Lu Lin¹

¹ School of Mathematics, Shandong University, Jinan 250100, China

² College of Science, China University of Petroleum, Qingdao 266580, China

Correspondence should be addressed to Lu Lin; linlu@sdu.edu.cn

Received 24 March 2013; Accepted 30 May 2013

Academic Editor: Ivan Ivanov

Copyright © 2013 Y. Song and L. Lin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Financial risk is objective in modern financial activity. Management and measurement of the financial risks have become key abilities for financial institutions in competition and also make the major content in finance engineering and modern financial theories. It is important and necessary to model and forecast financial risk. We know that nonlinear expectation, including sublinear expectation as its special case, is a new and original framework of probability theory and has potential applications in some scientific fields, specially in finance risk measure and management. Under the nonlinear expectation framework, however, the related statistical models and statistical inferences have not yet been well established. In this paper, a sublinear expectation nonlinear regression is defined, and its identifiability is obtained. Several parameter estimations and model predictions are suggested, and the asymptotic normality of the estimation and the mini-max property of the prediction are obtained. Finally, simulation study and real data analysis are carried out to illustrate the new model and methods. In this paper, the notions and methodological developments are nonclassical and original, and the proposed modeling and inference methods establish the foundations for nonlinear expectation statistics.

1. Introduction

Finance is the core of economy, and financial safety is directly related to economic safety. Financial risk management is a huge field with diverse and evolving components, as evidenced by both its historical development and current best practice. One such component—probably the key component—is risk measurement. The 2007-2008 financial crisis and its long-lasting aftermath make people more aware that it is a very urgent and necessary thing to model and forecast financial risk.

It is well known that among all the assumption conditions imposed to the classical statistical models, the most vital one is of course that the models under study have a certain probability distribution that may or may not be known. The classical linear expectation and determinant statistics are built on such a distribution certainty or model certainty. The distribution certainty, however, is not always the case in practice, such as in risk measure and superhedging in finance (see, e.g., El Karoui et al. [1], Artzner et al. [2] Chen and Epstein [3],

Follmer and Schied [4]). Without the distribution certainty, the resulting expectation is of nonlinearity, usually. The earlier work on nonlinear expectation may ascend to Huber [5] in the sense of robust statistics or ascent to Walley [6] in the sense of imprecise probabilities. In recent decades, the theory and methodology of nonlinear expectation have been well developed and received much attention in some application fields such as finance risk measure and control. A typical example of the nonlinear expectation, called *g*-expectation (small *g*), was introduced in Peng [7] in the framework of backward stochastic differential equations. As a further development, *G*-expectation (big *g*) and its related versions are proposed by Peng [8]. Under the nonlinear expectation framework, the most common distribution is the so-called *G*-normal distribution, which was first introduced in Peng [8]. Furthermore, as the theoretical basis of the nonlinear expectation, the law of large numbers as well as the central limit theorem is also established by Peng [9, 10]. Also, from different points of view, many authors studied the nonlinear expectation, its application, and the related issues; see, for

example, Briand et al. [11], Coquet et al. [12], Denis and Martini [13], Denis et al. [14], Gao [15], Li and Peng [16], Rosazza [17], Soner et al. [18], and Xu and Zhang [19]. Other references include Chen and Peng [20], Peng [21–24], Soner et al. [25–27], and Song [28], among many others.

Contrary to the fast development of the nonlinear expectation in probability theory, little attention has been paid to the related statistical models and statistical inferences to the best of our knowledge. Although the earlier work of Huber [5] refers initially to the upper and the lower expectations, a special nonlinear expectation, the main aspects focus on robust statistics, and the underlying true model is supposed implicitly to have a certain distribution. Gross error model, for example, contains a certain true distribution in the contaminated distribution set, and on such a distribution set, the upper and the lower expectations can be defined; see, for example, Strassen [29] and Huber [5]. In classical statistical frameworks, heteroscedastic model may be the closest one to the model uncertainty aforementioned, but it only has the variance uncertainty, and the corresponding inference methods do not involve any notion of the nonlinear expectation. In the nonparametric framework, the model structure is not given, and in Bayesian framework, the model parameter is random. But the two statistical frameworks are essentially different from the model uncertainty aforementioned, and the corresponding methods are completely unrelated to any nonlinear expectation. In time series models, although the data depend on observation time, the strict stationarity or the weak stationarity is required to guarantee the certainty of statistical inferences. In one word, under the classical statistics frameworks, including parameter models, nonparametric models, Bayes models, and time series models, the defined expectations are of linearity. Without this linearity, it is essentially difficult or impossible by the classical methods to achieve the classical certain conclusions, such as estimation consistency, asymptotic normality of the estimation.

Under the model-uncertainty frameworks, the classical statistics methods are no longer available, usually. The classical maximum likelihood, for example, is nonexistent or can not be uniquely determined due to without a certain likelihood function. Also the classical least squares is invalid because it is required that the data are derived from a certain distribution, such as normal distribution. Moreover, the classical statistical models such as linear regression model, may not be well-defined as their identifiability depends on the mean-certainty; without the mean-certainty, the regression notion has to be redefined so that the new one is of identifiability. Thus, to achieve the target of statistics inference, it is necessary to develop new statistical frameworks and new statistical methods.

Lin et al. [30] establish a framework of sublinear expectation regression for the model with distribution uncertainty. Based on a sublinear expectation space, a sublinear expectation linear regression is defined, and its identifiability is achieved. We often meet nonlinear model in the study of finance risk measure and management. It is a kind of simple and approximate methods that we deal with nonlinear model with the theory of linear model. The approximate often brings many problems. Moreover, it is not consistent with the fact of

the conclusion. As a result, since the actual model is nonlinear model, we should use the method of nonlinear science to deal with the actual model. Motivated by Lin et al. [30], we propose a sublinear expectation nonlinear regression in this paper. And we achieve its identifiability. Our model is always available for both cases of the variance uncertainty and the mean-variance uncertainty. Unlike the classical regression, the new model tends to use a large value to predict response variable and obtains the mini-max prediction risk. It implies that our method is a robust strategy and therefore has potential applications in finance risk measure and management. New parameter estimation methods are suggested, and the resulting estimators are asymptotically normal distributed for the case of high-frequency data. It is worth mentioning that under the model-uncertainty framework, the certainty-statistical inferences are established in this paper, including the parameter certainty, the prediction certainty, and the distribution certainty of the parameter estimation. The notions and methodologies developed here are nonclassical and original, and the theoretical framework establishes the foundations for nonlinear expectation statistics.

The remainder of the paper is organized in the following way. In Section 2, a sublinear expectation nonlinear regression model is built, and its identifiability is obtained. The estimation and prediction methods are suggested in Section 3, and the asymptotic normality of the estimators and the mini-max property of the predictions are established as well in this section. Simulation study and real data analysis are carried out in Section 4 to illustrate the new model and methodology. The proofs of the theorems and the definition of the sublinear expectation space are postponed to appendices.

2. Sublinear Expectation Nonlinear Regression

In this section we establish a framework of sublinear expectation nonlinear regression, including modeling, estimation, prediction, and the asymptotic properties.

2.1. Model. We consider the following nonlinear regression model:

$$Y = f(\mathbf{x}, \beta) + \varepsilon, \quad (1)$$

where Y is a scalar response variable, $\mathbf{x} = (X_1, \dots, X_p)'$ is the associated p -dimensional covariate having a certain distribution $F_{\mathbf{x}}(x)$, and $\beta = (\beta_1, \dots, \beta_p)'$ is a p -dimensional vector of unknown parameters. We assume that the function $f(\cdot)$ is a known function, and it is twice continuously differentiable. Furthermore, it is supposed that the error ε is independent of \mathbf{x} . We need the independency condition only for simplicity. The idea and methodology developed in the following can be extended to the dependent case, but the notations and algorithm are relatively complex. It is worth pointing out that the essential difference from the classical nonlinear regression model is that here the error ε has distribution uncertainty, which is defined in the following way.

Let Ω be a given set, and let \mathcal{H} be a linear space of real valued functions defined on Ω . Furthermore, let \mathbb{E} denote

a sublinear expectation: $\mathcal{H} \rightarrow \mathcal{R}$, satisfying monotonicity, constant preserving, subadditivity, and positive homogeneity; for the details of the definitions; see appendices. The triple $(\Omega, \mathcal{H}, \mathbb{E})$ is then called a sublinear expectation space. In this paper, we assume that ε is defined on a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$. It can be seen from the definition that the probability distribution of ε is uncertain. For regression analysis, we suppose that \mathcal{H} contains linear and quadratic functions, and although the sublinear expectation \mathbb{E} is supposed to be existent, its exact form may be unknown. Thus, a remarkable point of view is that since regression analysis depends mainly on expectation, we here only define a sublinear expectation space, instead of the well-accepted probability space.

As was known by Peng [10, 31], the sublinear expectation of a function $g(\varepsilon) \in \mathcal{H}$ can be expressed as a supremum of linear expectations.

Lemma 1. *There exists a family of linear expectations $\{E_f : f \in \mathcal{F}\}$ defined on \mathcal{H} such that*

$$\mathbb{E}[g(\varepsilon)] = \sup_{f \in \mathcal{F}} E_f[g(\varepsilon)] \quad \text{for } g \in \mathcal{H}, \quad (2)$$

and there exists a $f_g \in \mathcal{F}$ such that

$$\mathbb{E}[g(\varepsilon)] = E_{f_g}[g(\varepsilon)]. \quad (3)$$

Write $\bar{u} = \mathbb{E}[\varepsilon]$, $\underline{u} = -\mathbb{E}[-\varepsilon]$, $\bar{\sigma}^2 = \mathbb{E}[\varepsilon^2]$, $\sigma^2 = -\mathbb{E}[-\varepsilon^2]$. Then, the intervals $[\underline{u}, \bar{u}]$ and $[\sigma^2, \bar{\sigma}^2]$ characterize the mean-uncertainty and the variance uncertainty of ε , respectively.

When \mathbf{x} is a random variable, for regression modeling, it is required to define the sublinear conditional expectation $\mathbb{E}[Y | \mathbf{x}]$. Lin et al. [30] gave the definition of the sublinear conditional expectation $\mathbb{E}[Y | \mathbf{x}]$. For example, by the representation theorem given previously, the previous $\mathbb{E}[Y | \mathbf{x}]$ can be defined as

$$\mathbb{E}[Y | \mathbf{x}] = \sup_{f_{Y|\mathbf{x}} \in \mathcal{F}_{Y|\mathbf{x}}} E_{f_{Y|\mathbf{x}}}[Y | \mathbf{x}], \quad (4)$$

where $\{E_{f_{Y|\mathbf{x}}} : f_{Y|\mathbf{x}} \in \mathcal{F}_{Y|\mathbf{x}}\}$ is a family of linear conditional expectations. With this definition, the properties of monotonicity, constant preserving, subadditivity, and positive homogeneity given in appendices still hold.

2.2. G-Normal Regression. We first consider the case when the error ε is supposed to be G-normally distributed; namely,

$$\varepsilon \sim \mathcal{N} = N(\{0\} \times [\underline{\sigma}^2, \bar{\sigma}^2]). \quad (5)$$

Under this situation, ε has a certain zero mean, but its variance is uncertain, a special distribution uncertainty. As was defined by Peng [8], ε is called G-normally distributed if it is defined on a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$ and satisfies that for each $a, b \geq 0$, $a\varepsilon + b\bar{\varepsilon} \stackrel{d}{=} \sqrt{a^2 + b^2}\varepsilon$, where $\bar{\varepsilon}$ is an independent copy of ε and “ $\stackrel{d}{=}$ ” stands for equal in distribution. For the definition and the representation of G-normal distribution, see Peng [8]. It follows from the cash

translatability of the sublinear conditional expectation that for regression model (1), if ε is G-normally distributed as in (5), then

$$\mathbb{E}[Y | \mathbf{x}] = f(\mathbf{x}, \beta). \quad (6)$$

The above relationship (6) could be thought of as a G-normal expectation nonlinear regression because \mathbb{E} is the G-normal expectation, a special sublinear expectation.

Remark 2. (1) Note that \mathbf{x} has an identical distribution. Then, if ε is G-normally distributed as in (5), the G-expectation of Y is identifiable in the sense that $\mathbb{E}(Y | \mathbf{x})$ can be uniquely determined by $f(\mathbf{x}, \beta)$ as in (6). (2) Here we emphasize the use of G-normal regression because a quadratic loss function will be employed in the following to construct a quasi-maximum likelihood estimation; for the details, see the next section. In fact the notion proposed here can be directly extended to general mean-certainty sublinear expectation regressions. Specifically, we only assume that ε has the mean certainty, instead of G-normal distribution. Under this situation, model (6) could be regarded as a mean-certainty sublinear expectation regression.

2.3. Sublinear Expectation Regression. Now we investigate the model in which the error ε has both the mean uncertainty and the variance uncertainty. By the cash translatability of the sublinear conditional expectation, we have

$$\mathbb{E}[Y | \mathbf{x}] = f(\mathbf{x}, \beta) + \bar{u}. \quad (7)$$

This model could be thought of as a sublinear expectation nonlinear regression because \mathbb{E} is a sublinear expectation.

Remark 3. (1) If $\underline{u} < \bar{u}$, then, given \mathbf{x} , the sublinear expectation of Y has a shift \bar{u} , and more precisely, the sublinear expectation of Y has the framework of (7). (2) In the face of the mean uncertainty, we still can uniquely determine the parameter vector β and then use the mean-shift framework $f(\mathbf{x}, \beta) + \bar{u}$ instead of $f(\mathbf{x}, \beta)$ to predict the response variable Y . Such a framework reflects the robust feature of sublinear expectation nonlinear regression. If Y is a measure of the risk of a financial product, then the sublinear expectation regression tends to use a relatively large value to predict the risk and moreover, and the increment of the risk measure is just the sublinear expectation \bar{u} of the error ε .

3. Estimation and Prediction

It is supposed in this section that the dimension p of β is fixed. Let $\{(Y_i, \mathbf{x}_i) : i = 1, \dots, N\}$ be a sample from model (1), satisfying

$$Y_i = f(\mathbf{x}_i, \beta) + \varepsilon_i, \quad i = 1, \dots, N. \quad (8)$$

Unlike the classical ones, here Y_1, \dots, Y_N may have distribution uncertainty because of the distribution uncertainty of $\varepsilon_1, \dots, \varepsilon_N$. Then the corresponding estimation method should be different from the classical ones that apply only to linear expectation regression models.

We now introduce a mini-max method to construct the estimator of β .

3.1. The Case of the Mean Certainty. We first consider the case of ε having the mean certainty. Because Y has the sublinear expectation $f(\mathbf{x}, \beta)$ given \mathbf{x} , theoretically, we should choose β , so that it can minimize the sublinear expectation square loss:

$$\mathbb{E} \left[(Y - f(\mathbf{x}, \beta))^2 \right]. \quad (9)$$

We can easily verify that the previous sublinear expectation square loss is a convex function of β . Thus the optimization problem has a unique global optimal solution. The previous is in fact a sublinear expectation least squares.

Remark 4. It is worth mentioning that under G-normal distribution, we have that if φ is a convex function, then

$$\mathbb{E} [\varphi(\varepsilon)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(\bar{\sigma}u) \exp \left\{ -\frac{u^2}{2} \right\} du, \quad (10)$$

and if φ is a concave function, then

$$\varphi(\varepsilon) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(\underline{\sigma}u) \exp \left\{ -\frac{u^2}{2} \right\} du. \quad (11)$$

For the details see Peng [8]. These imply that under the convex function and concave function spaces, the G-normal has density functions $(1/\sqrt{2\pi\bar{\sigma}}) \exp\{-u^2/2\bar{\sigma}^2\}$ and $(1/\sqrt{2\pi\underline{\sigma}}) \exp\{-u^2/2\underline{\sigma}^2\}$, respectively. Therefore, the previous sublinear expectation least squares could be thought of as a quasi maximum likelihood.

To implement the estimation procedure, we need the following assumption.

(CI) There exists an index decomposition: I_j , $i = 1, \dots, m$, such that when $(ij) \in I_j$, $\varepsilon_{i1}, \dots, \varepsilon_{im}$ are independent and have an identical distribution.

We suppose from now on that the numbers of elements in I_j , $i = 1, \dots, m$, are equal; that is, $n_1 = n_2 = \dots = n_m = n$, without loss of generality. Because it is assumed that $\varepsilon_{i1}, \dots, \varepsilon_{in}$ are identically distributed, the independence of $\varepsilon_{i1}, \dots, \varepsilon_{in}$ in the condition (CI) is the same as in the linear expectation framework, instead of the independence in the nonlinear expectation. Here we need the independence only for simplicity. Without the independence, for example, $\varepsilon_{i1}, \dots, \varepsilon_{in}$ are weakly dependent, and the conclusions given in the following still hold; for weakly dependent processes and the properties of the estimation, see, for example, Rosenblatt [32, 33], Kolmogorov and Rozanov [34], Bradley and Bryc [35], and Lu and Lin [36]. Furthermore, a common decomposition is built according to the observation time order; more precisely, $\varepsilon_1, \dots, \varepsilon_N$ are reindexed as $\varepsilon_{ij} = \varepsilon_{(i-1)n+j}$, $i = 1, \dots, m$, $j = 1, \dots, n$, and then the index sets I_j 's are defined as $I_j = \{(ij) : j = 1, \dots, n\}$. It is known that in a small time interval, the characteristic of the data could be regarded as to be changeless exactly or approximately. With this point of view, the condition (CI) is relatively mild. Also we can decompose the index sets according to the values of Y in a descending order, for example. Moreover, we will further weaken (CI) and suggest a data-driven decomposition after Theorem 5 given in the following.

Denote by F_i the common distribution function of ε_{ij} , $(ij) \in I_i$. By the representation theorem of sublinear expectation given in (2), the sublinear expectation loss (9) can be written as $\max_{1 \leq i \leq m} E_{F_i} [(Y - f(\mathbf{x}, \beta))^2]$, and therefore its empirical version is

$$\max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2. \quad (12)$$

By minimizing the previous empirical square loss, we obtain the mini-max estimator of β as

$$\hat{\beta}_G = \arg \min_{\beta \in \mathcal{B}} \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2. \quad (13)$$

It can be easily verified that $\max_{1 \leq i \leq m} (1/n) \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2$ is a convex function of β . Thus the resulting estimator $\hat{\beta}_G$ is a unique global optimal solution in the previous optimization problem. Since there is in general no explicit formula for the sublinear expectation nonlinear estimator $\hat{\beta}$, the minimization of (13) must usually be carried out by some iterative method. There are two general types of iteration methods: the Newton-Raphson iteration and the Gauss-Newton iteration. Denote $\sigma_i^2 = E(\varepsilon_{ij}^2)$ for $(ij) \in I_i$ and $\sigma_{i_*}^2 = \max_{1 \leq i \leq m} \sigma_i^2$, and for simplicity, assume that

$$\sigma_{i_*}^2 > \sigma_i^2 \quad \text{for all } i \neq i_*. \quad (14)$$

The previous mini-max estimator is asymptotic normally distributed.

In order to give the following theorem, we need the following conditions:

- (I) the parameter space \mathcal{B} is compact (closed and bounded), and β_0 is its interior point;
- (II) the following inequality holds: $(1/n) \sum_{j=1}^n f(\mathbf{x}_{ij}, \beta_1) f(\mathbf{x}_{ij}, \beta_2)$ converges uniformly in $\beta_1, \beta_2 \in \mathcal{B}$;
- (III) $\lim_{n \rightarrow \infty} (1/n) \sum_{j=1}^n [f(\mathbf{x}_{ij}, \beta_0) - f(\mathbf{x}_{ij}, \beta)]^2 \neq 0$ if $\beta \neq \beta_0$;
- (IV) $\lim_{n \rightarrow \infty} (1/n) \sum_{j=1}^n (\partial f / \partial \beta) |_{\beta_0} (\partial f / \partial \beta') |_{\beta_0} (\equiv C)$ exist and is nonsingular;
- (V) $(1/n) \sum_{j=1}^n (\partial f / \partial \beta) (\partial f / \partial \beta')$ converges uniformly in β in an open neighborhood of β_0 ;
- (VI) $(1/n) \sum_{j=1}^n [\partial^2 f / \partial \beta_i \partial \beta_j]^2$ converges uniformly in β in an open neighborhood of β_0 .

Theorem 5. For the mean-certainty model, if the condition (CI) holds, and conditions (I), (II), and (III) hold, as $n \rightarrow \infty$ and with probability tending to 1, the estimator defined in (13) satisfies (a)

$$\hat{\beta}_G \rightarrow \beta_0. \quad (15)$$

(b) In addition to the previous condition, if conditions (IV), (V), and (VI) hold, as $n \rightarrow \infty$,

$$\sqrt{n} (\hat{\beta}_G - \beta) \xrightarrow{d} N(0, \sigma_{i_*}^2 C^{-1}), \quad (16)$$

where \xrightarrow{d} stands for convergence in distribution, and $N(0, \sigma_i^2 C^{-1})$ is a classical normal distribution.

This theorem establishes the theoretical foundation for further statistical inferences such as constructing confidence interval and test statistic. We can see that the condition (CI) can be replaced by the following relatively weak condition:

(CI') $\varepsilon_{i_1}, \dots, \varepsilon_{i_n}$ are independent and have an identical distribution.

This condition only involves the errors with indexes in I_i . However, recognizing the fact that the number n of data in each small time slice I_i should be relative large, the conditions (CI) and (CI') only apply to high-frequency data. Moreover, by the two conditions, it is implicitly assumed that the index compositions I_i , $i = 1, \dots, m$, or I_{i^*} are known completely. Under some situations, however, it is difficult or impossible to get such exact compositions in advance. Thus, the data-driven decompositions are desired in practice. Now we briefly discuss this issue. By the condition (CI'), the proof of Theorem 5 and (3), the mini-max estimator in (13) can be approximately recasted as

$$\hat{\beta}_G = \arg \min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^n [Y_{i_j} - f(\mathbf{x}_{i_j}, \beta)]^2. \quad (17)$$

Thus, a simple approach is to identify I_{i^*} or its subset. Let $I_i^0 = \{(ij) : j = 1, \dots, n^0\}$, $i = 1, \dots, m^0$ be the initial compositions according to the observation time order, for example, where $n^0 > p$. Note that under the model with the mean certainty, the common LS estimator $\hat{\beta}_{LS}$ of β is consistent. We then arrange $\sum_{j=1}^{n^0} [Y_{ij} - f(\mathbf{x}_{ij}, \hat{\beta}_{LS})]^2$, $i = 1, \dots, m^0$, in the descending order as

$$\begin{aligned} \sum_{j=1}^{n^0} [Y_{i_1j} - f(\mathbf{x}_{i_1j}, \hat{\beta}_{LS})]^2 &\geq \sum_{j=1}^{n^0} [Y_{i_2j} - f(\mathbf{x}_{i_2j}, \hat{\beta}_{LS})]^2 \\ &\geq \dots \geq \sum_{j=1}^{n^0} [Y_{i_mj} - f(\mathbf{x}_{i_mj}, \hat{\beta}_{LS})]^2. \end{aligned} \quad (18)$$

When p is relatively small, the index set $I_{i_1}^0 = \{(i_1j) : j = 1, \dots, n^0\}$ can be chosen as an initial choice of I_{i^*} or a subset of I_{i^*} . We can use the data in $I_{i_1}^0$, together with approximate formula (17) to build the estimator. Since the data size in $I_{i_1}^0$ may be small, it is necessary to enlarge the initial choice $I_{i_1}^0$. To this end, we consider the following hypothesis testing:

$$H_0 : \sigma_2^2 = v_1^2 \iff H_1 : \sigma_2^2 < v_1^2, \quad (19)$$

where σ_2^2 is the supposed variance of ε_{i_2j} for $(i_2j) \in I_{i_2}^0 = \{(i_2j) : j = 1, \dots, n^0\}$ and $v_1^2 = (1/(n^0 - p)) \sum_{j=1}^{n^0} [Y_{i_1j} - f(\mathbf{x}_{i_1j}, \hat{\beta}_{LS})]^2$. The classical methods can be used to test the hypothesis H_0 . If H_0 is not rejected, then $I_{i_1}^0 \cup I_{i_2}^0$ could be chosen as an enlarged choice of I_{i^*} . The procedure is repeated

until the remainder variances are significantly smaller than v_1^2 .

After the estimator $\hat{\beta}_G$ being obtained, a natural prediction of Y is

$$\hat{Y} = f(\mathbf{x}, \hat{\beta}_G). \quad (20)$$

If the model uncertainty is ignored and the common least squares (LS) method is used to construct the estimator $\hat{\beta}_{LS}$ of β , then the LS based prediction is

$$\hat{Y} = f(\mathbf{x}, \hat{\beta}_{LS}). \quad (21)$$

Comparing the two estimators by maximum prediction risk and average prediction risk, we have the following conclusion.

Theorem 6. *Under the condition of the mean certainty, whether the variance uncertainty exists or not, the following relationships always hold:*

$$\begin{aligned} \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_G)]^2 &\leq \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_{LS})]^2, \\ \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_G)]^2 &\geq \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_{LS})]^2. \end{aligned} \quad (22)$$

Remark 7. The theorem indicates that the sublinear expectation nonlinear regression is a robust strategy that can reduce the maximum prediction risk. Thus, it can be expected that such a regression could be useful for measuring and controlling financial risk.

3.2. The Case of the Mean-Variance Uncertainty. We now consider the case of ε having both the mean-uncertainty and the variance uncertainty. In this case Y has the sublinear expectation $f(\mathbf{x}, \beta) + \bar{u}$ given \mathbf{x} . Theoretically, we should choose β , so that it can minimize the sublinear expectation square loss:

$$\mathbb{E} [(Y - f(\mathbf{x}, \beta) - \bar{u})^2]. \quad (23)$$

However, we cannot directly complete the estimation procedure as \bar{u} is unknown usually. We then design a profile estimation procedure as follows. Let $\hat{\beta}$ be an initial estimator of β , which may be the estimator obtained in case of the mean certainty or by common least squares. We then estimate \bar{u} by

$$\hat{\bar{u}} = \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \hat{\beta})] \quad (24)$$

and finally estimate β by

$$\tilde{\beta}_G = \arg \min_{\beta \in \mathcal{B}} \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta) - \hat{\bar{u}}]^2. \quad (25)$$

Denote $u_i = E[\varepsilon_{ij}]$, $\sigma_i^2 = E(\varepsilon_{ij} - u_i)^2$, $v_i^2 = \sigma_i^2 + (\bar{u} - u_i)^2$, and $\sigma_{k^*}^2 = \max\{v_i^2 : i = 1, \dots, m\}$, and for simplicity, assume

$\sigma_{k_*}^2 > v_i^2$ for all $i \neq k_*$. By the same argument as in Theorem 5, we can prove that the estimator $\tilde{\beta}_G$ is asymptotically normal distributed. The following theorem presents the details.

Theorem 8. *For mean-variance uncertainty, if the condition C1 holds, and conditions (I), (II), and (III) hold, as $n \rightarrow \infty$ and with probability tending to 1, the estimator defined in (25) satisfies (a)*

$$\tilde{\beta}_G \rightarrow \beta_0. \quad (26)$$

(b) *In addition to the previous condition, if conditions (IV), (V), and (VI) hold, as $n \rightarrow \infty$,*

$$\sqrt{n}(\tilde{\beta}_G - \beta) \xrightarrow{d} N(0, \sigma_{k_*}^2 C^{-1}), \quad (27)$$

where \xrightarrow{d} stands for convergence in distribution and $N(0, \sigma_{k_*}^2 C^{-1})$ is a classical normal distribution.

For proof of the theorem, see appendices. This theorem establishes a foundation for further statistical inference and data analysis. Here we also need to check the condition C1. From the estimation procedure given previously, we see that it is asymptotically equivalent to determine two index sets, with which the mean of the error and $(1/n) \sum_{j=1}^n [Y_{k_*j} - f(\mathbf{x}_{k_*j}, \beta) - \hat{u}]^2$ achieve the maximum values \bar{u} and $v_{k_*}^2$, respectively. The approaches are similar to that used in case of the mean-certainty and thus the details are omitted here.

With the estimator, a natural prediction of Y is

$$\tilde{Y} = f(\mathbf{x}, \tilde{\beta}_G) + \bar{u}. \quad (28)$$

Similar to the properties in Theorem 6, the prediction \tilde{Y} can obtain mini-max prediction risk.

Theorem 9. *Whether the mean uncertainty and the variance uncertainty exist or not, the following relationship always holds:*

$$\begin{aligned} & \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \tilde{\beta}_G) - \hat{u}]^2 \\ & \leq \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_{LS}) - \hat{u}]^2. \end{aligned} \quad (29)$$

It shows that our proposal is a robust strategy and is therefore useful for measuring and controlling financial risk. Meanwhile, the simulation study given in Section 4 will verify that when model has the mean-variance uncertainty, the average prediction error of the new method is usually smaller than that of the LS method; namely,

$$\begin{aligned} & \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \tilde{\beta}_G) - \bar{u}]^2 \\ & < \frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}, \hat{\beta}_{LS}) - \bar{u}]^2. \end{aligned} \quad (30)$$

It is because the prediction bias of $f(\mathbf{x}, \hat{\beta}_{LS})$ is between \underline{u} and \bar{u} , which is not ignorable special for the case of $\underline{u}\bar{u} > 0$.

TABLE 1: Simulation results of estimation and prediction for Experiment 1 with $m = 10$ and $n = 10$.

MSE	β_1	β_2	β_3	MPE	APE
$\tilde{\beta}_G$	0.0089	0.0327	0.0385	7.0248	4.6245
$\hat{\beta}_{LS}$	0.0043	0.0068	0.0052	8.3124	3.6279

4. Simulation Study and Real Data Analysis

4.1. Simulation Study. In this section we present several simulation examples to compare the finite-sample performances of the sublinear expectation nonlinear regression proposed in this paper. To get comprehensive comparisons, we use mean square error (MSE), maximum prediction error (MPE), and average prediction error (APE) to assess the different methods. From the simulations given, we will get the following findings: (1) the new methods can significantly reduce the MPE under all the situations; (2) when the model has the mean certainty, the advantages of the new methods over the LS methods are not very obvious; (3) for the model with the mean uncertainty, the prediction of the LS methods does not work and even breaks down, but the new methods can get valid prediction because the impact of the mean uncertainty on the new methods can be successfully eliminated by the use of the sublinear expectation of the error.

Experiment 1. We first consider the following simple nonlinear model:

$$Y = f(\mathbf{x}, \beta) + \varepsilon, \quad (31)$$

where $f(\mathbf{x}, \beta) = (\beta_3 / (\beta_1 + \beta_2))(\exp\{-\beta_1 X_1\} - \exp\{\beta_2 X_2\} + \exp\{\beta_3 X_3\})$. In the simulation procedure, the regression coefficients are chosen as $\beta_k = 1$, $k = 1, 2, 3$, and the observation values of X_k are independent and identically distributed from $N(10, 2)$, $k = 1, 2, 3$. We choose $\varepsilon \sim N(\{0\} \times [0, 3])$, a G-normal distribution with certain zero mean. In this case, the model has the mean certainty. The following way is used to generate the data of G-normal distribution, approximately. Generate variance values σ_i^2 , $i = 1, \dots, m$, from the uniform distribution $U[0, 3]$, and then generate the values ε_{ij} , $j = 1, \dots, n$, of ε from the common normal distribution $N(0, \sigma_i^2)$.

For $m = 10$ and $n = 10$, the simulation results are reported in Table 1, in which MSE, MPE, and APE denote the mean squared error, maximum prediction error, and average prediction, respectively; for the definitions of MPE and APE see Theorem 6. It is clear that the MSE and APE of common LS estimation $\hat{\beta}_{LS}$ are significantly smaller than those of the G-normal estimation $\tilde{\beta}_G$. Such a result is not surprising because, under the mean-certainty model, the common LS estimation $\hat{\beta}_{LS}$ is consistent, but the construction of the new estimation $\tilde{\beta}_G$ only uses the data in a small time interval (the number of the data used to construct the estimator is only 10). On the other hand, the MPE by the new one $\tilde{\beta}_G$ is significantly small than that by the LS estimator $\hat{\beta}_{LS}$, which implies that the new method can reduce the maximum prediction risk and therefore is a robust strategy.

TABLE 2: Simulation results of estimation and prediction for Experiment 2 with $m = 10$ and $n = 10$.

MSE	β_1	β_2	β_3	MPE	APE
$\hat{\beta}_G$	0.2485	0.3697	0.3289	16.0952	7.3548
$\hat{\beta}_{LS}$	0.1324	0.2819	0.2305	39.4257	22.6957

TABLE 3: Simulation results of estimation and prediction for Experiment 3.

MSE	β_1	β_2	β_3	MPE	APE
$\hat{\beta}_G$	0.2374	0.4738	0.4326	13.2590	7.8835
$\hat{\beta}_{LS}$	0.1962	0.2653	0.2739	39.0364	22.4357

The simulation results in Table I indicate that when model has the mean certainty, the advantages of the new methods over the common LS are not rather obvious. Moreover, the new methods even have the disadvantage of instability. In the following, we will see that when model has the mean-uncertainty, our new methods have rather clear advantages over the LS based methods.

Experiment 2. We reconsider the nonlinear model

$$Y = f(\mathbf{x}, \beta) + \varepsilon, \quad (32)$$

where $f(\mathbf{x}, \beta) = (\beta_3/(\beta_1 + \beta_2))(\exp\{-\beta_1 X_1\} - \exp\{\beta_2 X_2\} + \exp\{\beta_3 X_3\})$, which is the same in form as in Experiment 1. However, here the model has the mean-variance uncertainty as $\varepsilon \sim N([3, 5] \times [0, 3])$. The other experiment conditions are designed as $X_k \sim N(0, 1)$, $k = 1, 2, 3$, $m = 10$, and $n = 20$. The values of ε are generated by the following way. First, the values u_i of the mean and the values σ_i^2 of the variance are generated from the uniform distributions $U[3, 5]$ and $U[0, 4]$, respectively, and then the values ε_{ij} , $j = 1, \dots, n$ of ε are generated from the common normal distribution $N(u_i, \sigma_i^2)$ for $i = 1, \dots, m$. The simulation results are listed in Table 2. For the MSE of the parameter estimation, the results are similar to those in Experiment 1; that is, the MES of the LS estimation is smaller than that of the new estimation because the new method only uses the data in a small subinterval. However, when the mean uncertainty and variance uncertainty appear in the model, both the MPE and the APE of the new one are significantly smaller than those of the LS estimator. Particularly, the prediction by the LS seems to be totally invalid. It indicates that ignoring the model uncertainty will lead to a serious prediction risk.

4.2. Real Data Analysis

Experiment 3. In economics, the Cobb-Douglas functional form of production functions is widely used to represent the relationship of an output to inputs. It was proposed by Knut Wicksell (1851–1926) and tested against statistical evidence by Charles Cobb and Paul Douglas in 1900–1928. We consider the Cobb-Douglas production function with an additive error

$$Y = \beta_1 L^{\beta_2} K^{\beta_3} + \varepsilon_t, \quad (33)$$

where Y is total production (the monetary value of all goods produced in a year), L is labor input, K is capital input, β_1 is total factor productivity, β_2 and β_3 are the output elasticities of labor and capital, respectively. These values are constants determined by available technology. We assume that the model has the mean-variance uncertainty as $\varepsilon \sim N([0, 1] \times [0, 1])$.

Here, the statistical data comes from China Statistical Yearbook (2003), total production Y is the gross domestic product (GDP), labor input L is employment, and capital input K is fixed asset investment. The simulation results are listed in Table 3. When the mean uncertainty and variance-uncertainty appear in the model, both the MPE and the APE of the new one are significantly smaller than those of the LS estimator. Particularly, the prediction by the LS seems to be totally invalid. It indicates that ignoring the model-uncertainty will lead to a serious prediction risk.

The simulation results in Table 3 indicate that when model has the mean certainty, the advantages of the new methods over the common LS are not rather obvious. Moreover, the new methods even have the disadvantage of instability. In the following, we will see that when model has the mean-uncertainty, our new methods have rather clear advantages over the LS based methods.

Appendices

A. Definition of Sublinear Expectation

Let Ω be a given set, and let \mathcal{H} be a linear space of real valued functions defined on Ω . Suppose that $\mathbb{E} : \mathcal{H} \rightarrow \mathbb{R}$ satisfies the following properties: for all $U, V \in \mathcal{H}$,

- (i) monotonicity: if $U \geq V$, then $\mathbb{E}[U] \geq \mathbb{E}[V]$;
- (ii) constant preservation: $\mathbb{E}[c] = c$ for any constant c ;
- (iii) subadditivity: $\mathbb{E}[U + V] \leq \mathbb{E}[U] + \mathbb{E}[V]$;
- (iv) positive homogeneity: $\mathbb{E}[\lambda U] = \lambda \mathbb{E}[U]$ for each $\lambda \geq 0$.

Then $(\Omega, \mathcal{H}, \mathbb{E})$ is called a sublinear expectation space.

It can verified that (iii) and (iv) together imply

- (v) convexity: $\mathbb{E}[\alpha U + (1 - \alpha)V] \leq \alpha \mathbb{E}[U] + (1 - \alpha)\mathbb{E}[V]$ for $\alpha \in [0, 1]$.

Furthermore, (ii) and (iii) together lead to

- (vi) cash translatability: $\mathbb{E}[U + c] = \mathbb{E}[U] + c$ for any constant c .

B. Proofs

Proof of Theorem 5. It follows from (CI) that

$$\frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2 = \sigma_i^2 + o_p\left(\frac{1}{n}\right). \quad (B.1)$$

Consequently,

$$\max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2 = \sigma_{i_*}^2 + o_p\left(\frac{1}{n}\right). \quad (B.2)$$

Then, when n is large enough,

$$\begin{aligned} & \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)]^2 \\ &= \frac{1}{n} \sum_{j=1}^n [Y_{i_*, j} - f(\mathbf{x}_{i_*, j}, \beta)]^2 + o_p\left(\frac{1}{n}\right). \end{aligned} \quad (\text{B.3})$$

It implies that when n is large enough, $\widehat{\beta}_G$ is actually the common LS estimator of β obtained by the data with index in the small time interval I_{i_*} . By the asymptotic normality of the LS estimation under linear expectation framework, we can get the asymptotic normality of $\widehat{\beta}_G$. That is, first, let us consider the part (a) of the Theorem 5 as follows:

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n [Y_{i_*, j} - f(\mathbf{x}_{i_*, j}, \beta)]^2 \\ &= \frac{1}{n} \sum_{j=1}^n \varepsilon_t^2 + \frac{2}{n} \sum_{j=1}^n [f(\mathbf{x}_{i_*, j}, \beta_0) - f(\mathbf{x}_{i_*, j}, \beta)] \varepsilon_t \\ &+ \frac{1}{n} \sum_{j=1}^n [f(\mathbf{x}_{i_*, j}, \beta_0) - f(\mathbf{x}_{i_*, j}, \beta)]^2 \\ &= A_1 + A_2 + A_3. \end{aligned} \quad (\text{B.4})$$

First, $P \lim A_1 = \sigma_{i_*}^2$ by a law of large numbers. Secondly, for fixed β_0 and β , $P \lim A_1 = 0$ follows from the convergence of $(1/n) \sum_{j=1}^n [f(\mathbf{x}_{i_*, j}, \beta_0) - f(\mathbf{x}_{i_*, j}, \beta)]^2$ by Chebyshev's inequality

$$\begin{aligned} & P \left[\left\{ n^{-1} \sum_{j=1}^n [f(\mathbf{x}_{i_*, j}, \beta_0) - f(\mathbf{x}_{i_*, j}, \beta)] \varepsilon_i \right\}^2 > \varepsilon^2 \right] \\ &< \frac{\sigma_{i_*}^2}{\varepsilon^2 n^2} \sum_{j=1}^n [f(\mathbf{x}_{i_*, j}, \beta_0) - f(\mathbf{x}_{i_*, j}, \beta)]^2. \end{aligned} \quad (\text{B.5})$$

By the condition (II), we know that the uniform convergence of A_2 follows from the uniform convergence of the right-hand side of (B.5). Having thus disposed of A_1 and A_2 , we need only to prove that $P \lim A_3$ is uniquely minimized at β_0 . By the condition (II), we know that $\lim_{n \rightarrow \infty} (1/n) \sum_{j=1}^n [f(\mathbf{x}_{ij}, \beta_0) - f(\mathbf{x}_{ij}, \beta)]^2 \neq 0$ if $\beta \neq \beta_0$. Thus, we get the result that as $n \rightarrow \infty$ and with probability tending to 1, $\widehat{\beta}_G \rightarrow \beta_0$.

(b) With ease of presentation, we denote $S_n(\beta) = \sum_{j=1}^n [Y_{i_*, j} - f(\mathbf{x}_{i_*, j}, \beta)]^2$. Because $f(\cdot)$ is twice continuously differentiable with respect to β , the asymptotic normality of the estimator $\widehat{\beta}_G$ can be derived from the following Taylor expansion:

$$\left. \frac{\partial S_n(\beta)}{\partial \beta} \right|_{\widehat{\beta}_G} = \left. \frac{\partial S_n(\beta)}{\partial \beta} \right|_{\beta_0} + \left. \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'} \right|_{\beta^*} (\widehat{\beta}_G - \beta_0), \quad (\text{B.6})$$

where $\partial^2 S_n(\beta) / \partial \beta \partial \beta'$ is a $p \times p$ matrix of second-order derivatives and β^* lies between $\widehat{\beta}_G$ and β_0 . Since the left-hand

side of (B.6) is zero (because $\widehat{\beta}_G$ minimizes $S_n(\beta)$), from (B.6) we obtain

$$\sqrt{n}(\widehat{\beta}_G - \beta_0) = - \left[\left. \frac{1}{n} \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'} \right|_{\beta^*} \right]^{-1} \left. \frac{1}{\sqrt{n}} \frac{\partial S_n(\beta)}{\partial \beta} \right|_{\beta_0}. \quad (\text{B.7})$$

Thus, we are done if we can show that (i) the limit distribution of $(1/\sqrt{n})(\partial S_n(\beta) / \partial \beta) |_{\beta_0}$ is normal and (ii) $(1/n)(\partial^2 S_n(\beta) / \partial \beta \partial \beta') |_{\beta^*}$ converges in probability to a nonsingular matrix. We will consider these two statements in turn.

The proof of statement (i) is straightforward. Differentiating $S_n(\beta) = \sum_{j=1}^n [Y_{i_*, j} - f(\mathbf{x}_{i_*, j}, \beta)]^2$ with respect to β , we obtain

$$\frac{\partial S_n(\beta)}{\partial \beta} = -2 \sum_{j=1}^n [Y_{i_*, j} - f(\mathbf{x}_{i_*, j}, \beta)] \frac{\partial f}{\partial \beta}. \quad (\text{B.8})$$

Evaluating (B.8) at β_0 and dividing it by \sqrt{n} , we have

$$\left. \frac{1}{\sqrt{n}} \frac{\partial S_n(\beta)}{\partial \beta} \right|_{\beta_0} = - \frac{2}{\sqrt{n}} \sum_{j=1}^n \varepsilon_i \left. \frac{\partial f}{\partial \beta} \right|_{\beta_0}. \quad (\text{B.9})$$

By the condition (IV) and the Lindberg-Feller central limit theorem, we can get

$$\left. \frac{1}{\sqrt{n}} \frac{\partial S_n(\beta)}{\partial \beta} \right|_{\beta_0} \rightarrow N(0, 4\sigma_{i_*}^2 C). \quad (\text{B.10})$$

Proving (ii) poses a more difficult problem. Differentiating (B.8) again with respect to β and dividing by n yields

$$\begin{aligned} \frac{1}{n} \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'} &= \frac{2}{n} \sum_{j=1}^n \frac{\partial f}{\partial \beta} \frac{\partial f}{\partial \beta'} \\ &- \frac{2}{n} \sum_{j=1}^n [f(\mathbf{x}, \beta_0) - f(\mathbf{x}, \beta)] \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'} \\ &- \frac{2}{n} \sum_{j=1}^n \varepsilon_i \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'}. \end{aligned} \quad (\text{B.11})$$

We must show that each of the three terms in the right-hand side of (B.11) converges almost surely to a nonstochastic function uniformly in β . By the conditions (V) and (VI), we can get

$$P \lim \left. \frac{1}{n} \frac{\partial^2 S_n(\beta)}{\partial \beta \partial \beta'} \right|_{\beta^*} = 2C. \quad (\text{B.12})$$

Finally, from (B.7), (B.10), and (B.12), we obtain

$$\sqrt{n}(\widehat{\beta}_G - \beta_0) \xrightarrow{d} N(0, \sigma_{i_*}^2 C^{-1}). \quad (\text{B.13})$$

Thus we prove the conclusion of theorem. \square

Proof of Theorem 6. The definitions of the two estimations lead directly to the conclusions of the theorem. \square

Proof of Theorem 8. From the proof of Theorem 5, we see that $\widehat{\beta}_G$ is actually the common LS estimator of β obtained by data

$(Y_{i,j}, \mathbf{x}_{i,j}), j = 1, \dots, n$. Thus $\widehat{\beta}_G = \beta + o_p(1/n)$, where β is the true regression coefficient given by (7) in the mean-certainty model. Moreover, by the same argument as used in the proof of Theorem 5, we have

$$\max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)] = \max_{1 \leq i \leq m} E_{F_i}[\varepsilon] + o_p\left(\frac{1}{n}\right). \quad (\text{B.14})$$

The previous discussion ensures that

$$\begin{aligned} \widehat{u} &= \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta)] + o_p\left(\frac{1}{n}\right) \\ &= \max_{1 \leq i \leq m} E_{F_i}[\varepsilon] + o_p\left(\frac{1}{n}\right) = \bar{u} + o_p\left(\frac{1}{n}\right), \end{aligned} \quad (\text{B.15})$$

where F_i is the distribution of data in I_i . Consequently,

$$\begin{aligned} \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta) - \widehat{u}]^2 \\ = \max_{1 \leq i \leq m} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta) - \bar{u}]^2 + o_p\left(\frac{1}{n}\right). \end{aligned} \quad (\text{B.16})$$

On the other hand,

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta) - \bar{u}]^2 \\ = \frac{1}{n} \sum_{j=1}^n [\varepsilon_{ij} - u_i - (\bar{u} - u_i)]^2 \\ = \sigma_i^2 + (\bar{u} - u_i)^2 + o_p\left(\frac{1}{n}\right). \end{aligned} \quad (\text{B.17})$$

Then,

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n [Y_{ij} - f(\mathbf{x}_{ij}, \beta) - \bar{u}]^2 \\ = \frac{1}{n} \sum_{j=1}^n [Y_{k,j} - f(\mathbf{x}_{k,j}, \beta) - \bar{u}]^2 + o_p\left(\frac{1}{n}\right), \end{aligned} \quad (\text{B.18})$$

where $\mathbf{x}_{k,j}$ is the covariate matrix with index in I_{k^*} . By the similar steps of of Theorem 5, we can prove the conclusion of theorem. \square

Proof of Theorem 9. The proof of the theorem follows directly from the definitions of the two estimators. \square

Acknowledgments

This research was supported by NNSF Project (11171188, 11221061, and 11231005) of China, NSF and SRRF Projects (ZR2010AZ001 and BS2011SF006) of Shandong Province of China, and K C Wong-HKBU Fellowship Programme for Mainland China Scholars 2010-11.

References

- [1] N. E. Karoui, S. Peng, and M. C. Quenez, "Backward stochastic differential equations in finance," *Mathematical Finance*, vol. 7, no. 1, pp. 1–71, 1997.
- [2] Ph. Artzner, F. Delbaen, J. M. Eber, and D. Heath, "Thinking coherently," *RISK*, vol. 10, pp. 86–71, 1997.
- [3] Z. Chen and L. Epstein, "Ambiguity, risk, and asset returns, in continuous time," *Econometrica*, vol. 70, no. 4, pp. 1403–1443, 2002.
- [4] H. Follmer and A. Schied, *Statistic Finance, An Introduction in Discrete Time*, Walter de Gruyter, Berlin, Germany, 2nd edition, 2004.
- [5] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, NY, USA, 1981.
- [6] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall/CRC, New York, NY, USA, 1991.
- [7] S. Peng, "Backward SDE and related g-expectations," in *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak, Eds., vol. 364 of *Pitman Research Notes in Mathematics Series*, pp. 141–159, 1997.
- [8] S. Peng, "G-expectation, G-brownian motion and related stochastic calculus of itô's type," in *The Abel Symposium 2005, Abel Symposia 2*, Benth et al., Ed., pp. 541–567, Springer, New York, NY, USA, 2006.
- [9] S. Peng, "Law of large numbers and central limit theorem under nonlinear expectations," In press.
- [10] S. Peng, "Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation," *Stochastic Processes and their Applications*, vol. 118, no. 12, pp. 2223–2253, 2008.
- [11] Ph. Briand, F. Coquet, Y. Hu, J. Mémin, and S. Peng, "A converse comparison theorem for BSDEs and related properties of g-expectation," *Electronic Communications in Probability*, vol. 5, pp. 101–117, 2000.
- [12] F. Coquet, Y. Hu, J. Mémin, and S. Peng, "Filtration-consistent nonlinear expectations and related g-expectations," *Probability Theory and Related Fields*, vol. 123, no. 1, pp. 1–27, 2002.
- [13] L. Denis and C. Martini, "A theoretical framework for the pricing of contingent claims in the presence of model uncertainty," *The Annals of Applied Probability*, vol. 16, no. 2, pp. 827–852, 2006.
- [14] L. Denis, M. Hu, and S. Peng, "Function spaces and capacity related to a sublinear expectation: application to G-brownian motion paths," *Potential Analysis*, vol. 34, no. 2, pp. 139–161, 2011.
- [15] F. Gao, "Pathwise properties and homeomorphic flows for stochastic differential equations driven by G-Brownian motion," *Stochastic Processes and their Applications*, vol. 119, no. 10, pp. 3356–3382, 2009.
- [16] X. Li and S. Peng, "Stopping times and related Itô's calculus with G-Brownian motion," *Stochastic Processes and their Applications*, vol. 121, no. 7, pp. 1492–1508, 2011.
- [17] G. E. Rosazza, "Risk measures via g-expectations," *Insurance*, vol. 39, no. 1, pp. 19–34, 2006.
- [18] H. M. Soner, N. Touzi, and J. Zhang, "Martingale representation theorem for the G-expectation," *Stochastic Processes and their Applications*, vol. 121, no. 2, pp. 265–287, 2011.
- [19] J. Xu and B. Zhang, "Martingale characterization of G-Brownian motion," *Stochastic Processes and their Applications*, vol. 119, no. 1, pp. 232–248, 2009.
- [20] Z. Chen and S. Peng, "A general downcrossing inequality for g-martingales," *Statistics and Probability Letters*, vol. 46, no. 2, pp. 169–175, 2000.

- [21] S. Peng, “Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob-Meyer’s type,” *Probability Theory and Related Fields*, vol. 113, no. 4, pp. 473–499, 1999.
- [22] S. Peng, “Filtration consistent nonlinear expectations and evaluations of contingent claims,” *Acta Mathematicae Applicatae Sinica English Series*, vol. 20, no. 2, pp. 191–214, 2004.
- [23] S. Peng, “Nonlinear expectations and nonlinear Markov chains,” *Chinese Annals of Mathematics B*, vol. 26, no. 2, pp. 159–184, 2005.
- [24] S. Peng, “Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations,” *Science in China A*, vol. 52, no. 7, pp. 1391–1411, 2009.
- [25] H. M. Soner, N. Touzi, and J. Zhang, “Quasi-sure stochastic analysis through aggregation,” *Electronic Journal of Probability*, vol. 16, pp. 1844–1879, 2011.
- [26] H. M. Soner, N. Touzi, and J. Zhang, “Wellposedness of second order backward SDEs,” *Probability Theory and Related Fields*, vol. 153, pp. 149–190, 2012.
- [27] M. Soner, N. Touzi, and J. Zhang, “Dual formulation of second order target problems,” *Annals of Applied Probability*, vol. 23, pp. 308–347, 2013.
- [28] Y. Song, “Uniqueness of the representation for G-martingales with finite variation,” *Electronic Journal of Probability*, vol. 17, article 24, 2012.
- [29] V. Strassen, “Messfehler und information,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 2, pp. 267–284, 1964.
- [30] L. Lin, Y. Shi, X. Wang, and S. Yang, “Sublinear expectation regression,” *Annals of Statistics*. In press.
- [31] S. Peng, “Nonlinear expectations and stochastic calculus under uncertainty,” In press.
- [32] M. Rosenblatt, “A central limit theorem and a strong mixing condition,” *Proceedings of the National Academy of Sciences*, vol. 42, pp. 43–47, 1956.
- [33] M. Rosenblatt, “Density estimates and Markov sequences,” in *Nonparametric Techniques in Statistical Inference*, M. Puri, Ed., pp. 199–210, Cambridge University Press, London, UK, 1970.
- [34] A. N. Kolmogorov and U. A. Rozanov, “On the strong mixing conditions of a stationary Gaussian process,” *Theory of Probability and Its Applications*, vol. 5, no. 2, pp. 204–208, 1960.
- [35] R. C. Bradley and W. Bryc, “Multilinear forms and measures of dependence between random variables,” *Journal of Multivariate Analysis*, vol. 16, no. 3, pp. 335–367, 1985.
- [36] C. R. Lu and Z. Y. Lin, *Limit Theories for Mixing Dependent Variables*, Science Press, Beijing, China, 1997.