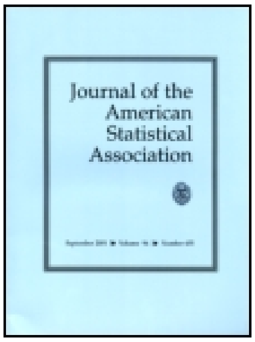

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://amstat.tandfonline.com/loi/uasa20

Prediction of Functional Status for the Elderly Based on a New Ordinal Regression Model

Hyokyoung Grace Hong & Xuming He

To cite this article: Hyokyoung Grace Hong & Xuming He (2010) Prediction of Functional Status for the Elderly Based on a New Ordinal Regression Model, Journal of the American Statistical Association, 105:491, 930-941, DOI: 10.1198/jasa.2010.ap08631

To link to this article: http://dx.doi.org/10.1198/jasa.2010.ap08631

Published online: 01 Jan 2012.

Prediction of Functional Status for the Elderly Based on a New Ordinal Regression Model

Hyokyoung Grace HONG and Xuming HE

The functional mobility of the elderly is a very important factor in aging research, and prognostic information is valuable in making clinical and health care policy decisions. We develop a predictive model for the functional status of the elderly based on data from the Second Longitudinal Study of Aging (LSOA II). The functional status is an ordinal response variable. The ordered probit model has been moderately successful in analyzing such data; however, its reliance on the normal distribution for its latent variable hinders its accuracy and potential. In this paper, we focus on the prediction of conditional quantiles of the functional status based on a more general transformation model. The proposed estimation procedure does not rely on any parametric specification of the conditional distribution functions, aiming to reduce model misspecification errors in the prediction. Cross-validation within the LSOA II data shows that our prediction intervals are more informative than those from the ordered probit model. Monte Carlo simulations also demonstrate the merits of our approach in the analysis of ordinal response variables.

KEY WORDS: Nonparametric transformation model; Ordinal data; Quantile regression; Second Longitudinal Study of Aging.

1. INTRODUCTION

One of the main goals in aging research is to improve health status for older individuals. Monitoring and predicting the health-related quality of life for the elderly is of particular importance because it helps public health policy makers and caretakers understand the special needs of the elderly and design appropriate interventions. Among many instruments to measure the health status of the elderly, the functional status is commonly used to represent a person's ability to perform self-care, self-maintenance, and physical activities (Bierman 2001). An assessment of functional status allows us to detect subtle yet measurable changes in individual health conditions. For this reason, we chose the functional status as a premise of analysis in this study.

It is good news that a number of panel surveys having relevance to aging research have become available over the last two decades. Specifically, the Second Longitudinal Study of Aging (LSOA II) provides many opportunities for statistical analysis of the characteristics of aging including the functional status. Functional status (FS) is conceptualized as the ability to carry out functional tasks, often measured by self-reports on activities of daily living (ADL) and instrumental activities of daily living (IADL). The ADLs refer to the basic activities of daily life such as bathing, dressing, getting in/out of chairs, and toileting. The IADLs include activities not necessary for fundamental functioning, but still useful in a community. Examples of IADLs are such as preparing meals, managing money, and performing light housework. In this study, FS is classified into ordered and integer values, ranging from 1 to 5, according to the severity ratings used in earlier reports on disability by Anderson et al. (1998) and Mor et al. (1994), among others. More specifically, the five levels of FS are: (1) independent without any ADL or IADL disability; (2) IADL disabled only; (3) moderately ADL disabled (1–2 ADLs impaired); (4) severely ADL disabled (≥ 3 ADLs impaired); and (5) deceased. An elderly person becomes increasingly less functional as the FS level increases. Our primary aim is to predict how FS changes over a two-year time span based on current health conditions.

By the nature of the clinical practice, as in the functional status, ordinal scales are common in the clinical research; see Dawson and Trapp (2004). The ordered probit and ordered logit models are frequently used for ordinal data analysis; see McCullagh and Nelder (1989). Inferences on the linear coefficients for the covariate are generally robust against the misspecification of the link function (Duan and Li 1987). However, the prediction of the probabilities or quantiles of the conditional distribution can be seriously distorted by a misspecified link. Semiparametric approaches to median regression for ordinal responses have been considered by Lee (1992) and Melenberg and van Soest (1996), but those methods do not provide root-n consistent estimates of the link functions.

In this paper, we use a semiparametric model with both flexibility and prediction accuracy in mind. The proposed model is a transformed quantile regression model based on the jittered response (with random noise added to smooth the response values) and includes the ordered probit and logit models as special cases. Our approach is related to data smoothing used by Machado and Santos Silva (2005) as well as transformations in quantile regression as discussed by Mu and He (2007). We allow a nonparametric monotone function as the transformation, but under iid errors and certain conditions on the predictor distribution, including what is commonly known as the *linearity assumption* in the dimension reduction literature, the proposed estimates attain the root-n rate of convergence. Furthermore, the proposed method has desirable robustness properties. Given the transformation, we may use quantile regression estimates for allowing different effects of the covariate in different regions of the conditional distribution, making our prediction robust against model misspecification.

Hyokyoung Grace Hong is Assistant Professor, Department of Statistics and CIS, The Baruch College of the City University of New York (E-mail: *hyokyoung.hong@baruch.cuny.edu*). Xuming He is Professor, Department of Statistics, University of Illinois at Urbana–Champaign, Champaign, IL 61820 (E-mail: *x-he@illinois.edu*). The research is partially supported by NSF grant DMS-0604229, NIH grant R01GM080503-01A1, U.S.A., National Natural Science Foundation of China grant 10828102, and a Changjiang Visiting Professorship at the Northeast Normal University, China. We thank the Editor, Associate Editor, and anonymous referees for their helpful comments and suggestions.

Journal of the American Statistical Association
September 2010, Vol. 105, No. 491, Applications and Case Studies
DOI: 10.1198/jasa.2010.ap08631

The rest of the paper is organized as follows. Section 2 describes the variables that we use from the LSOA II data. Section 3 presents the transformed ordinal regression quantile model and the estimation procedure that we shall use for constructing the prediction intervals of the functional status of the elderly. Some important properties of the estimates are discussed in Section 4, where the proposed method is also compared with the ordered probit model through Monte Carlo simulations. We also discuss the essential conditions required for those properties as well as some practical implications. We return to the LSOA II data in Section 5 and use the proposed methodology to predict the functional status of the elderly over a two-year time period. Our analysis shows that the proposed method leads to more informative prediction than the ordered probit model. We also discuss what we can learn from our analysis regarding some of the limitations in the LSOA designs for the prediction of functional status.

2. PREDICTORS FROM THE LSOA II DATA

The LSOA II is a publicly available dataset and a collaborative project of the National Center for Health Statistics (NCHS) and the National Institute on Aging (NIA). The subjects of LSOA II are a nationally representative sample comprised of 9447 noninstitutionalized civilian persons of 70 years of age or older at the time of the Second Supplement on Aging (SOA II) in the United States. The SOA II was conducted by the Centers for Disease Control and Prevention. Participants completed a baseline questionnaire in 1994–1996 and completed two followup questionnaires about two years apart in 1997–1998 and 1999–2000. The sampling weights are a product of four components, which take into account the complex multistage probability design (Skinner, Holt, and Smith 1989). The complete set of LSOA II data is available online from the LSOA website *http://www.cdc.gov/nchs/lsoa.htm*.

Of the 9447 participants of the SOA II, 8300 complete records were available for the first followup study, due to missing information or dropout. Death was identified through interviews with family representatives. Out of those 8300 participants, 680 were missing at the time of the administration of the second followup, and 2169 deaths were reported, resulting in 5451 participants with complete information (including death) at the end of the second followup (Figure 1). Based on the data

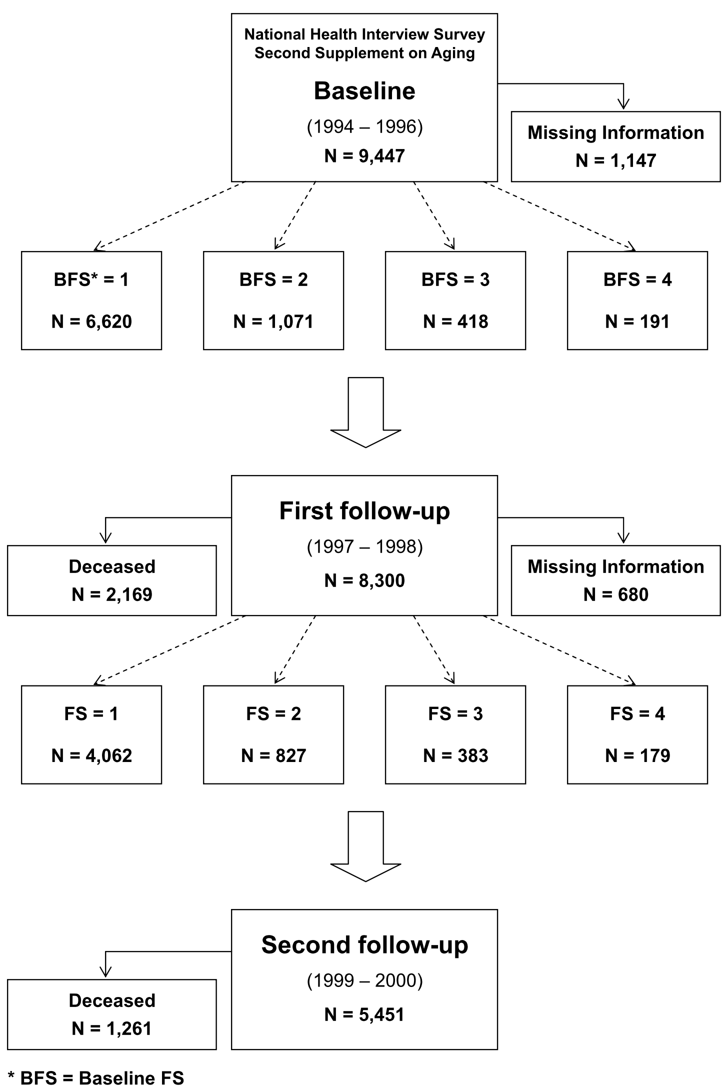

Figure 1. Flow diagram of participants in the Second Longitudinal Study of Aging.

Table 1. Summary of outcome, covariates, and descriptive measures of a dataset

Variables	Data sources	Explanation
Outcome		
Functional status (y)	First followup	Responses were classified as $1 =$ independent, $2 =$ IADL disabled only, $3 =$ moderately ADL disabled, $4 =$ severely ADL disabled, or $5 =$ death
Covariates		
Self-rated health (x_1)	Baseline	Coded into two levels ($0 =$ excellent/very good, $1 =$ good/fair/poor)
Diabetes (x_2)	Baseline	Classified as $0 =$ absent or $1 =$ present
Race (x_3)	Baseline	Classified as $1 =$ white or $-1 =$ non-white
Marital status (x_4)	Baseline	Classified as $1 =$ married or $-1 =$ not married (including respondents who were widowed, divorced, separated, or never married)
Age (x_5)	Baseline	Calculated from reported birth month and year
Education (x_6)	Baseline	Years ranged from 0 (never attended or kindergarten only) to 18 (indicating ≥ 6 years of college)
Sex (x_7)	Baseline	Classified as $1 =$ male or $-1 =$ female
Cancers (x_8)	Baseline	Classified as $0 =$ absent or $1 =$ present
CVD (x_9)	Baseline	Classified as $0 =$ absent or $1 =$ present
MSD (x_{10})	Baseline	Classified as $0 =$ absent or $1 =$ present
BMI (x_{11})	Baseline	Classified as $0 = \text{BMI} \geq 25$ or $1 = \text{BMI} < 25$
Smoking (x_{12})	Baseline	Classified as $0 =$ non-smoking or $1 =$ smoking
Condition (x_{13})	Baseline	The total number of self-reported chronic health conditions, ranging 0–11 (lower scores indicate fewer chronic health conditions) is recoded as $0 =$ number of condition ≤ 2 or $1 =$ number of condition ≥ 3
Lung disease (x_{14})	Baseline	Classified as $0 =$ absent or $1 =$ present

NOTE: Abbreviations: CVD, cardiovascular diseases, which include heart disease (heart attack), stroke, hypertension, and heart failure; MSD, musculoskeletal diseases, which include whether the respondent ever had a broken hip, osteoporosis, or arthritis; BMI, body mass index, which is calculated by dividing weight in kilograms by the square of height in meters; ADL, activities of daily living; IADL, instrumental activities of daily living.

from the first followup study (with 8300 subjects), we aim to develop a statistical model for predicting the functional status of the elderly over a two-year period. It is also possible to use the second followup data to predict functional status of a second two-year period, or to build a four-year predictive model using the longitudinal data from the entire study; see a cautionary note in the concluding section of the paper.

The sampling weights from the survey are used in the summary and in the estimation throughout our analysis (e.g., averages refer to weighted averages), but this point will not be repeated every time the results are discussed.

To select the predictors for functional status, we follow Anderson et al. (1998) and Lee et al. (2006). We consider two classes of hierarchical variables as predictors: sociodemographic variables and behavioral and biomedical variables.

The sociodemographic variables included in our analysis are age, marital status, race, gender, and education. Marital status was recorded as married or not married, the classification of not married including individuals who were widowed, never married, and seperated/divorced. Race was classified into two categories (white and nonwhite). Education is years of attainment ranging from 0 to 18. Age is calculated from birth month and year and ranging from 70 to 99.

We use a total of nine behavioral and biomedical variables: diabetes; cancers excluding minor skin cancer; chronic lung disease (bronchitis/emphysema); cardiovascular diseases (CVD) including heart disease, stroke, hypertension, or heart failure; musculoskeletal diseases (MSD) including whether the respondent ever had a broken hip, osteoporosis, or arthritis; self-rated health; number of chronic health conditions (diabetes, arthritis, heart disease, stroke, cancer, hypertension, asthma, etc.); smoking; and body mass index (BMI) (less than 25 or not). Table 1 describes the data sources and definition of the variables. It is worth noting, however, some variables in Table 1 could be further refined, and the variable selection procedure may help identify better predictors in future studies.

3. TRANSFORMED ORDINAL REGRESSION QUANTILE ESTIMATOR

3.1 Proposed Model

In this section, we develop a general methodology that will be used later for the analysis of the LSOA II data described in Section 2. Here, Y is an ordinal response variable taking values in $\{1, 2, 3, \ldots\}$, and $\mathbf{X} = (x_1, \ldots, x_p) \in \mathcal{R}^p$ is a p-dimensional predictor.

We conceive a random variable $\tilde{Y}$ by adding to Y, the ordinal response variable, U, a pseudo random variable from $\mathcal{U}[0, 1)$, that is, $\tilde{Y}_i = Y_i + U_i$, where $U_i \sim \mathcal{U}[0, 1)$ are independent draws. We shall refer to $\tilde{Y}_i$ as the jittered responses. Then we adopt the following model for the sample $(\mathbf{X}_i, Y_i)$:

$$\Lambda(\tilde{Y}_i) = \mathbf{X}_i^T \boldsymbol{\beta}_0 + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where Λ is a monotone function, and $\boldsymbol{\beta}_0$ is the vector of linear coefficients. Two basic conditions made in the paper are:

A0. The ϵ_i's are independent and identically distributed, whose distribution is unspecified.

A1. The conditional expectation $E(\mathbf{X}|\mathbf{X}^T\boldsymbol{\beta})$ exists and is linear in $\mathbf{X}^T\boldsymbol{\beta}$.

These two conditions make it easier to find a consistent estimator of $\boldsymbol{\beta}_0$, and Condition A0 leads to the root-n rate of convergence in our proposed estimate, but we shall also propose the quantile regression method that is more accommodating to a more general class of error distributions. Additional conditions and their implications will be discussed in Section 4.

Jittering enables us to specify a distributional relationship in the model, which is sufficient to identify the conditional quantiles of Y given $\mathbf{X}$. The use of a uniform distribution for U_i is for convenience and without loss of generality because a different choice for the jittering distribution simply leads to a different Λ function without changing the distribution of $\Lambda(\tilde{Y})$. Mathematically, we note that if $U_i = G(V_i)$ for any jittering distribution function G on $(0, 1)$, the corresponding link function in Model (1) becomes $\Lambda_G(y+v) = \Lambda(y+G(v))$ for $y \in \{1, 2, \ldots\}$ and $v \in (0, 1)$.

Location and scale normalizations are needed to make the parameters in Model (1) identifiable. The location normalization can be made by setting $\Lambda(\tilde{Y}_0) = 0$ for some prespecified $\tilde{Y}_0$. As in Chen (2002), we set the first coefficient of $\boldsymbol{\beta}_0$ to 1 for scale normalization. A common alternative is to assume that $\boldsymbol{\beta}_0$ is a unit vector. In either case, we must assume that $\boldsymbol{\beta}_0$ is a nonzero vector in the model.

As we focus on the τth conditional quantile of $\tilde{Y}$ given $\mathbf{X} = \mathbf{x}$, we may extend Model (1) to

$$Q_\tau(\Lambda(\tilde{Y})|\mathbf{X}) = \alpha_\tau + \mathbf{X}^T\boldsymbol{\beta}_\tau \tag{2}$$

for some coefficients $\alpha_\tau \in \mathcal{R}$ and $\boldsymbol{\beta}_\tau \in \mathcal{R}^p$, where $\tau \in (0, 1)$, and Q_τ denotes the τth quantile. The upper tail corresponds to $\tau > 0.5$ and the lower tail corresponds to $\tau < 0.5$. It is easy to show that

$$Q_\tau(Y|\mathbf{X}) = \lfloor \Lambda^{-1}\big(Q_\tau(\Lambda(\tilde{Y})|\mathbf{X})\big)\rfloor, \tag{3}$$

where $\lfloor\cdot\rfloor$ denotes the integer part of any nonnegative number.

Under Model (1), the slope parameters $\boldsymbol{\beta}_\tau$ have to be the same for all τ, so the quantiles can also be written as

$$Q_\tau(\Lambda(\tilde{Y})|\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta}_0 + F_\epsilon^{-1}(\tau). \tag{4}$$

In this paper, two approaches to estimating the quantiles will be used. We refer to the τ-specific estimator of Model (2) as Method 1, and the common slope approach in (4) as Method 2 for estimating the conditional quantiles. It is clear that Method 2 relies more on the iid error assumption A0 in Model (1), but Method 1 is more robust against deviations of this assumption. We now describe the estimating procedures for the transformed ordinal quantile regression model, to be called the TORQUE model for short.

3.2 Method 1: TORQUE With Quantile Regression

This method consists of four steps.

Step 1. Start with an initial estimate $\mathbf{b}_n$ of $\boldsymbol{\beta}_0$, taken to be the least squares estimate of slope from regressing $\{\tilde{Y}_i\}$ on $\{\mathbf{X}_i\}$.

Step 2. Obtain the estimate of Λ at each $\tilde{y}$ as

$$\hat{\Lambda}_n(\tilde{y}) = \arg\max_{\Lambda\in M_\Gamma}\Bigg\{\Gamma_n(\tilde{y}, \Lambda, \mathbf{b}_n) = \sum_{i\neq j}\big(d_{i\tilde{y}} - d_{j\tilde{y}_0}\big) \times 1\{\mathbf{X}_i^T\mathbf{b}_n - \mathbf{X}_j^T\mathbf{b}_n \geq \Lambda\}\Bigg\}, \tag{5}$$

where M_Γ is a prechosen compact set in $\mathcal{R}$, $d_{i\tilde{y}} = 1\{\tilde{Y}_i \geq \tilde{y}\}$ and $d_{j\tilde{y}_0} = 1\{\tilde{Y}_j \geq \tilde{y}_0\}$ for some $\tilde{y}_0$ chosen by the user under the location normalization assumption of $\Lambda(\tilde{y}_0) = 0$.

Step 3. Calculate a regression quantile estimate of $(\alpha_\tau, \boldsymbol{\beta}_\tau)$ as

$$(\hat{\alpha}_\tau, \hat{\boldsymbol{\beta}}_{n,\tau}) = \arg\min_{\alpha\in\mathcal{R},\boldsymbol{\beta}\in\mathcal{R}^p}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i) - \alpha - \mathbf{X}_i^T\boldsymbol{\beta}), \tag{6}$$

where $\rho_\tau(r) = (\tau I(r > 0) + (1-\tau)I(r < 0))|r|$ is the quantile loss function of Koenker and Bassett (1978).

Step 4. The τth quantile of Y given $\mathbf{X}$ can then be estimated from (2) and (3) with $\alpha_\tau, \boldsymbol{\beta}_\tau$, and Λ substituted by their estimates.

The scale normalization on $\boldsymbol{\beta}_0$ is unnecessary for computing the quantile estimates based on (3). Step 2 uses the same estimation method as in Chen (2002), so we refer to Chen (2002) for more details on how the objective function $\Gamma_n(\tilde{y}, \Lambda, \mathbf{b}_n)$ in (5) can be evaluated with $O(n)$ computations instead of $O(n^2)$. The regression quantile estimate in Step 3 can be obtained by the R package *quantreg*.

3.3 Method 2: TORQUE With Quantiles of Residuals

This method differs from Method 1 only in the use of (4) instead of (2), where the parameter estimate is updated by

$$(\hat{\alpha}_n, \hat{\boldsymbol{\beta}}_n) = \arg\min_{\alpha\in\mathcal{R},\boldsymbol{\beta}\in\mathcal{R}^p}\sum_{i=1}^n |\hat{\Lambda}_n(\tilde{Y}_i) - \alpha - \mathbf{X}_i^T\boldsymbol{\beta}|, \tag{7}$$

and $\hat{\Lambda}_n$ is the same as in Method 1.

From the residuals $\epsilon_i^n = \hat{\Lambda}_n(\tilde{Y}_i) - \hat{\alpha}_n - \mathbf{X}_i^T\hat{\boldsymbol{\beta}}_n$, we estimate the τth quantile of $\Lambda(\tilde{Y})|\mathbf{X}$ using

$$\mathbf{X}^T\hat{\boldsymbol{\beta}}_n + \hat{F}_{\epsilon^n}^{-1}(\tau), \tag{8}$$

where $\hat{F}_{\epsilon^n}^{-1}(\tau)$ is the τth sample quantile of the residuals. The quantile of Y given $\mathbf{X}$ can be obtained in the same way as in Method 1.

The quantile estimates from Method 1 allow τ-specific coefficients, but we use a single transformation Λ in both Method 1 and Method 2. The use of a τ-specific transformation as in Mu and He (2007) would broaden the models but require much more intensive computations. In our analysis of the LSOA II data, we find that a single Λ function is adequate and is easier for interpretation.

3.4 Averaging Over Jittered Data

One noticeable feature of the TORQUE approach is the use of the jittered response $\tilde{Y}$. We observe $(\mathbf{X}_i, Y_i)$ but the values of U_i are independently drawn from the uniform distribution. If we had observed two samples from the same model, it would be sensible to use both. Using the same logic, we take advantage of multiple draws of U_i in reducing some variability of the estimates. More specifically, we follow an suggestion of Machado and Santos Silva (2005) of averaging the estimates over $\{\tilde{Y}_i^{(l)}, \mathbf{X}_i\}_{i=1}^n$, $\tilde{Y}_i^{(l)} = Y_i + U_i^{(l)}$, $l = 1, \ldots, m$, where $U_i^{(l)}$ are drawn independently from $\mathcal{U}[0, 1)$.

Let $\hat{\boldsymbol{\beta}}_n^{(l)}$ and $\hat{\Lambda}_n^{(l)}$ be the estimates of $\boldsymbol{\beta}_0$ and Λ based on the lth jittered sample. Then, the τth quantile of Y given $\mathbf{X}$ is estimated by the integer part of

$$\hat{Q}_\tau^m(\tilde{Y}_i|\mathbf{X}_i) = \frac{1}{m}\sum_{l=1}^m \hat{\Lambda}_n^{(l)^{-1}}\big(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_n^{(l)}\big).$$

If the parameters $\boldsymbol{\beta}_0$ and Λ are of interest, we take the averages

$$\hat{\boldsymbol{\beta}}_n^m = \frac{1}{m}\sum_{l=1}^m \hat{\boldsymbol{\beta}}_n^{(l)} \qquad \text{and} \qquad \hat{\Lambda}_n^m = \frac{1}{m}\sum_{l=1}^m \hat{\Lambda}_n^{(l)}.$$

Averaging the estimates from jittered data can reduce variability, but most of the reduction comes from the first 10 draws of U_i. In practice, we take m between 10 and 50 in our analysis.

4. PROPERTIES OF THE PROPOSED ESTIMATES

In this section, we discuss some of the basic properties of our proposed model and the estimates, and report some empirical performance measures from Monte Carlo studies. We also elaborate on the conditions required for the statistical properties presented in this paper, aiming to illuminate the potential as well as the limitations of the proposed method.

4.1 Connection to the Ordered Probit Model

The use of the jittering in $\tilde{Y}$ may make Model (1) seemingly artificial, but it is simply a convenient device for model building. We now point out that the ordered probit and logit models, commonly used for the analysis of ordinal responses, are special cases of Model (1). To this end, we derive from Model (1) that

$$\begin{aligned} P(Y \geq j|\mathbf{X}) &= P(\tilde{Y} \geq j|\mathbf{X}) \\ &= P(\Lambda(\tilde{Y}) \geq \Lambda(j)|\mathbf{X}) \\ &= P(\epsilon \geq \Lambda(j) - \mathbf{X}^T\boldsymbol{\beta}|\mathbf{X}) \\ &= 1 - F_{\epsilon|\mathbf{X}}(\Lambda(j) - \mathbf{X}^T\boldsymbol{\beta}), \end{aligned}$$

where $F_{\epsilon|\mathbf{X}}$ is the conditional distribution of ϵ given $\mathbf{X}$. Especially, if F is symmetric about 0, then

$$P(Y \leq j|\mathbf{X}) = F_{\epsilon|\mathbf{X}}(\Lambda(j+1) - \mathbf{X}^T\boldsymbol{\beta}). \tag{9}$$

If $\Lambda(j+1) = \alpha_j$ and $F(\mathbf{x}) = \Phi(\mathbf{x})$, then it reduces to the ordered probit model. The same can be said about the ordered logit model. The derivation in this subsection does not rely on any technical assumptions made elsewhere in the paper.

4.2 Large-Sample Consistency

The consistency of the proposed estimates, under either Method 1 or Method 2, relies on consistency of the initial estimate $\mathbf{b}_n$ and the function estimate $\hat{\Lambda}_n$. Under appropriate conditions, we have the following Theorem 1, which implies that the proposed conditional quantile estimates at any $\tau \in (0, 1)$ is asymptotically consistent at the root-n rate. Additional technical conditions A2, A3, and B1–B5 are given in the Appendix in their mathematical forms, but a heuristic description of those conditions are listed here for easier understanding.

- A2. The $\mathbf{X}_i$'s are not concentrated in any proper subspace of $\mathcal{R}^p$.
- A3. $\alpha_\tau + \mathbf{X}^T\boldsymbol{\beta}_0$ is the unique τth conditional quantile of $\Lambda(Y)$ given $\mathbf{X}$, for some $\alpha_\tau \in \mathcal{R}$.
- B1. At least one significant predictor is of interval scale.
- B2. The predictor $\mathbf{X}$ has finite third moments.
- B3–B5. The function $\Lambda(\cdot)$ is well behaved, and the predictor distribution satisfies some mild regularity conditions.

Theorem 1. Under Conditions A0–A3 and B1–B5, we have

$$\sup_{\tilde{y}_a \leq \tilde{y} \leq \tilde{y}_b} |\hat{\Lambda}_n(\tilde{y}) - \Lambda(\tilde{y})| = O_p\big(n^{-1/2}\big) \tag{10}$$

and

$$\hat{\boldsymbol{\beta}}_{n,\tau} - \boldsymbol{\beta}_0 = O_p\big(n^{-1/2}\big), \tag{11}$$

where $\tilde{y}_a$ and $\tilde{y}_b$ are specified in Condition B3.

The results of Theorem 1, especially the parametric rate of convergence, depend critically on the conditions we have imposed. Conditions A0 and A1 described in Section 2 are arguably the most stringent. The iid error assumption is necessary to achieve the root-n rate of convergence when Λ is nonparametric in nature. In this paper, A1 is also used to ensure that the initial estimator $\mathbf{b}_n$ of $\boldsymbol{\beta}_0$ in Step 1 of the TORQUE method is consistent. Condition B1 is needed to get a consistent estimate of Λ, and the proposed method might not do very well in a purely factorial design.

Some further remarks on the conditions are in order. Condition A1 places a rigid assumption on the design distribution of $\mathbf{X}$, often called a *linearity condition* on the design. Strictly speaking, this condition amounts to requiring $\mathbf{X}$ to be elliptically symmetric. However, it has been well discussed in the dimension reduction literature that this is a convenient condition to use to find a consistent estimate of $\boldsymbol{\beta}_0$, and more importantly, many authors, including Hall and Li (1993) and Cook (1998), have noted that this condition may hold to a reasonable approximation in many regressions, especially when $\mathbf{X}$ is high dimensional. Furthermore, the basic conditions used in this paper can be relaxed to include certain heteroscedastic errors if Λ is parametric (e.g., Mu and He 2007), or if a lower rate of convergence is asked for. In such extensions, we need $\boldsymbol{\beta}$ to be τ-specific in A3, and the proposed TORQUE with quantile regression ("Method 1" in Section 3.2) follows this direction, making the results more robust against the deviations from the iid assumption.

4.3 Simulation Studies

To understand the finite sample performance of the proposed estimator, we use Monte Carlo studies to compare the proposed estimator with the ordered probit model (OPM). The simulation results show that our proposed approach does well in a variety of cases and outperforms the OPM when ϵ is not normally distributed.

Data are generated from Model (1) with four case studies specified below. Study 1 is chosen as a favorable case for OPM, because the OPM model assumption holds exactly in this case. Studies 2 and 3 are chosen to represent non-Gaussian error distributions. Study 4 is motivated by a robustness consideration, as it includes a binary predictor and heteroscedastic error in the model. Clearly, conditions A0 and A1 are violated in this study, but the scenario in Study 4 is quite realistic.

The sample size is fixed at $n = 1000$ for each dataset, and a total of 100 datasets are generated in each study. After $\tilde{Y}$ is generated, we obtain the ordinal counts $Y = 1, 2, 3, 4$ as the greatest integer function $Y = \lfloor \tilde{Y} \rfloor$ for $1 \le \tilde{Y} < 5$. In addition, we let $Y = 4$ when $\tilde{Y} \ge 5$ and $Y = 1$ when $\tilde{Y} < 1$.

Study 1. $2\tilde{y} = x_1 + x_2 + 5 + \epsilon$, where $x_1 \sim \mathrm{N}(0.5, 0.5)$, $x_2 \sim \mathrm{N}(0.5, 0.5)$, and $\epsilon \sim \mathrm{N}(0, 1)$ are independent random variables.

Study 2. $5\tilde{y} = x_1 + x_2 + \epsilon$, where $x_1 \sim \mathcal{U}(3, 8)$, $x_2 \sim \mathcal{U}(3, 8)$, and $\epsilon \sim \chi^2(3)$ are independent random variables.

Study 3. $7\log(\tilde{y}) = x_1 + x_2 + \epsilon$, where $x_1 \sim \mathcal{U}(0, 5)$, $x_2 \sim \mathcal{U}(0, 5)$, and $\epsilon \sim \mathrm{LN}(0, 0.75)$ are independent random variables.

Study 4. $7\log(\tilde{y}) = x_1 + 2x_2 + (1 + x_2)\epsilon$, where $x_1 \sim \mathrm{Bernoulli}(0.5)$, $x_2 \sim \mathcal{U}(0, 4)$, and $\epsilon \sim \chi^2(1)$ are independent random variables.

For each study, we report the following two performance measures for comparing TORQUE (Method 1) and OPM:

- MAE(y), the mean absolute error (or distance) between the predicted responses and the observed responses; The conditional median from the estimated model is taken as the predicted response, but MAE(y) is averaged over 500 testing datasets generated from the same model as the training datasets.
- MAE(p), the mean absolute error (or distance) between $\hat{\mathbf{p}}_i$ and $\mathbf{p}_i$ over i, where $\mathbf{p}_i = (p_{i1}, \ldots, p_{i4})$, with $p_{ij} = P(y_{ij} = j)$ based on the true model, and $\hat{\mathbf{p}}_i = (\hat{p}_{i4}, \ldots, \hat{p}_{ij})$, with $\hat{p}_{ij}$ obtained from the estimated model.

Note that MAE(y) focuses on the error of point prediction, but MAE(p) measures how well the model predicts the true probability distribution of Y given $\mathbf{X}$. The simulation results are summarized in Table 2.

From Table 2, we see small differences between TORQUE and OPM in terms of point prediction, but the differences in MAE(p) are more telling. In Study 1, where the OPM is exactly the right parametric model, MAE(p) of the OPM does indeed exhibit smaller values. When the OPM is a misspecified model (Studies 2–3), the TORQUE becomes more favorable. In Study 4 where Conditions A0 and A1 are not satisfied, TORQUE continues to outperform in all 100 datasets in our study. Overall, the gain from TORQUE is often substantial, making it more than worthwhile to pay the price in Study 1. When the two methods are similar in the accuracy of point prediction, a difference in MAE(p) indicates differential performance in other aspects. In our analysis of LSOA II data, we shall see that a major advantage of TORQUE is that it produces more informative prediction intervals.

5. APPLICATION TO THE LSOA II DATA

We now return to the LSOA II data described in Section 2 and use the proposed TORQUE approach for constructing the prediction intervals for functional status. We shall show that the TORQUE approach gives more informative predictions than does the OPM, and also discuss some of our findings.

5.1 Preliminary Analysis

Since the LSOA II was designed for multiple aims, we need to consider several issues as to how the data can be best used. A predictive model can be estimated using the data from the first followup or the second followup. Although the longitudinal aspect of the study is useful in a number of ways, a predictive model for the functional status over a two-year period may not be estimated from both followups by treating them as repeated measurements. The subjects in the first followup are a (weighted) random sample from a well-defined population, but the second followup is no longer a random subsample from the same population because only those who have survived for the first few years are included in the second followup. This difference is also noted by Dellapasqua, Colleoni, and Goldhirsch (2006), where it is reported that cancer survivors without other chronic diseases were significantly more likely to report poor FS than individuals without a history of cancer. The analysis presented in this paper will be based on data from the first followup, where we believe we have a random sample from the general population.

Figure 2 shows the change in the functional status from the baseline to the first followup about two years later. In this section, we use BFS for the baseline functional status and FS for the functional status two years later. For example, it is clear from Figure 2 that more than 60% of the elderly with BFS = 1 stay with FS = 1 two years later, and that the functional status does not always get worse over a two-year period.

The QQ plot of the residuals from the TORQUE model (Method 1) is shown in Figure 3, which clearly indicates that the residuals deviate quite substantially from a normal distribution, making the OPM hard to defend.

5.2 Model Estimation

Here, we consider the quartiles at $\tau = 0.25, 0.50$, and 0.75. Roughly 25% of the subjects in the survey died after the first

Table 2. Mean absolute error comparison of two methods in four studies

	Study 1		Study 2		Study 3		Study 4	
	TORQUE	OPM	TORQUE	OPM	TORQUE	OPM	TORQUE	OPM
MAE(y)	0.39	0.39	0.35	0.36	0.30	0.32	0.56	0.59
MAE(p)	0.08	0.04	0.06	0.18	0.05	0.28	0.18	0.29
	(5)	(95)	(100)	(0)	(100)	(0)	(100)	(0)

NOTE: Each number in parenthesis in the last row represents the number of datasets (out of 100) for which the method outperforms the competitor in terms of MAE(p).

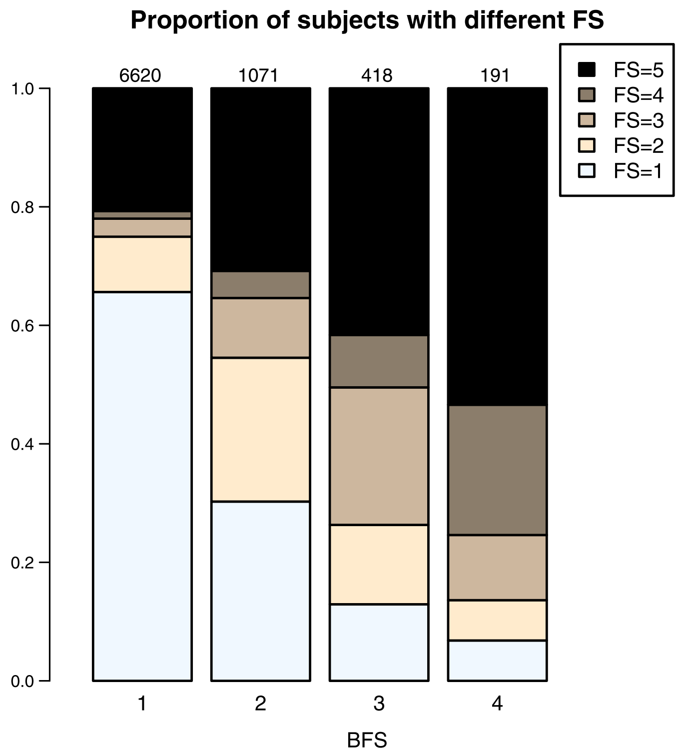

Figure 2. The vertical axis represents the proportion of subjects at different FS levels within each BFS subgroup. The values on top of the bars are the number of subjects within each subgroup. The online version of this figure is in color.

Table 3. 95% confidence interval for γ_k at each subgroup

BFS	γ_2	γ_3	γ_4
1	(0.29, 0.56)	(0.34, 0.64)	(0.38, 0.69)
2	(0.21, 0.70)	(0.68, 0.82)	(0.83, 0.93)
3	(0.00, 0.38)	(0.36, 0.67)	(0.81, 0.94)
4	(0.01, 0.53)	(0.26, 0.89)	(0.52, 0.94)

followup, so the prediction of the upper quantiles beyond the third quartile is not challenging.

In our analysis of the LSOA II data, the TORQUE models are fit to each subgroup of subjects with the same BFS values. To decide whether a single link function Λ fits all subgroups well, we show in Table 3 the bootstrap-based 95% confidence intervals for

$$\gamma_k = (\Lambda(k) - \Lambda(1))/(\Lambda(5) - \Lambda(1)), \qquad k = 2, 3, 4,$$

for each subgroup. The parameters γ_k describe the growth rates of Λ at different levels. There is significant statistical evidence that the link functions are different for subgroups with BFS $= 1$, 2, and 3 (or 4).

The estimated Λ functions are shown in Figure 4. The most interesting is the estimated Λ function for the subgroup with BSF $= 1$, because it is quite flat between 2 and 4 but has steep slopes before 2 and after 4. This Λ function indicates that there is not really much information in the TORQUE model that can distinguish the severities 2, 3, and 4 for functional status in

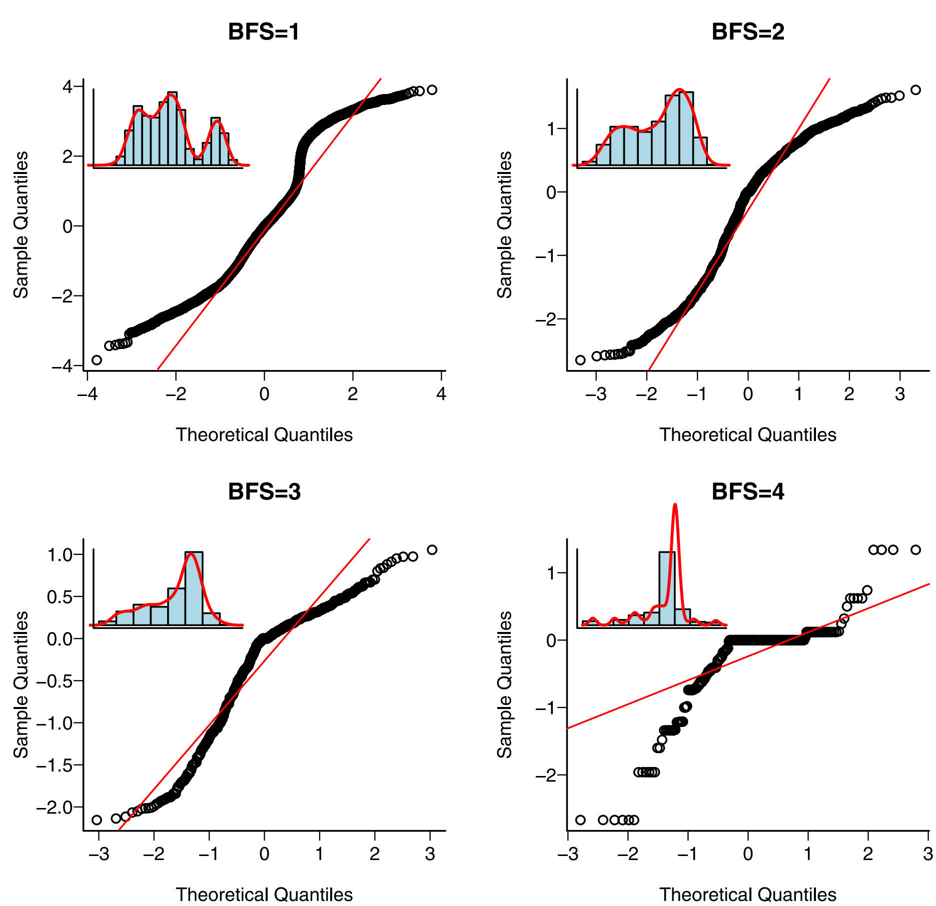

Figure 3. Q–Q plot for the Pearson residuals from the TORQUE model in the LSOA II data. The online version of this figure is in color.

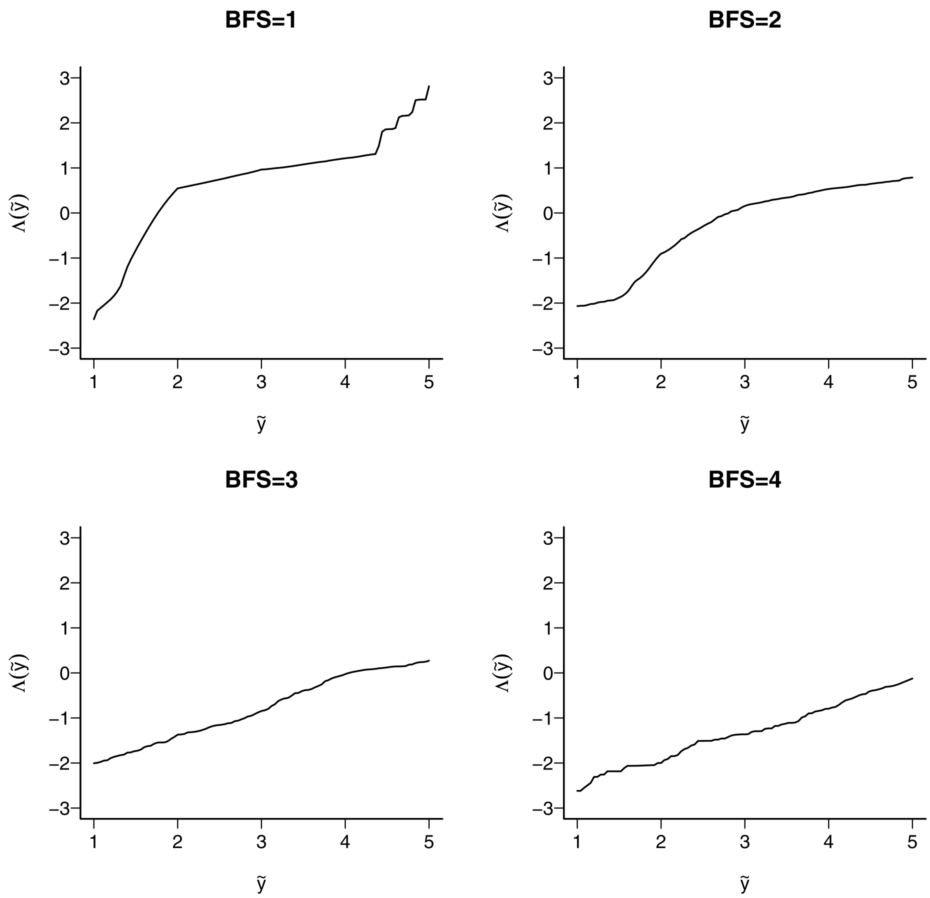

Figure 4. Estimation results for Λ of the group at each BFS.

the prediction, and the predictors are most helpful in separating the two extremes (1 and 5) from the others. The estimated Λ functions for other subgroups are closer to linear functions, so the resulting fits are approximately linear models with non-Gaussian errors.

For the most interesting subgroup with BFS $= 1$, Table 4 gives the scale-normalized estimates of $\boldsymbol{\beta}_0$ from TORQUE (Method 1) for the quartiles as well as those from the OPM. We note from the OPM results that on average self-rated health (x_1), marital status (x_4), age (x_5), education (x_6), and MSD (x_{10}) are important factors in predicting FS. However, the coefficients from TORQUE suggest that their relative contributions may vary slightly in two tails of the FS distribution. For example, relative to self-rated health, age and education have greater contributions at the lower quartiles, but MSD has a greater contribution at the higher quartiles. The different impacts of those predictors at different quartiles might not be statistically significant, so our analysis is exploratory rather than confirmatory. Detailed results for the other subgroups are not provided in the paper but are available from either author upon request.

5.3 Evaluation of Predictive Performance

To allow a fair assessment of the predictions based on TORQUE and the OPM, we used a random split of the data to form an estimation sample and a validation sample. To highlight the main point, we only discuss the results for the largest subgroup of 6620 subjects with BFS $= 1$.

We start with the investigation of the coverage and length of the 50% prediction intervals of the response variable, Y, based on Methods 1 and 2 of TORQUE, as well as the OPM for the comparison purpose. We use the interval between $\hat{Q}_{0.25}(Y|\mathbf{X})$ and $\hat{Q}_{0.75}(Y|\mathbf{X})$ as the 50% prediction interval under all methods.

Table 4. $\boldsymbol{\beta}$ estimation for BFS $= 1$

	OPM			TORQUE		
Variables	$\hat{\boldsymbol{\beta}}$	SE	t-value	$\hat{\boldsymbol{\beta}}_{n,\tau=0.25}$	$\hat{\boldsymbol{\beta}}_{n,\tau=0.50}$	$\hat{\boldsymbol{\beta}}_{n,\tau=0.75}$
x_1	1.00		8.61*	1.00	1.00	1.00
x_2	0.43	0.06	2.19	−0.40	0.37	0.69
x_3	−0.18	0.03	−1.90	−0.15	−0.13	−0.26
x_4	−0.22	0.02	−3.67	−0.15	−0.10	−0.18
x_5	0.13	0.00	12.47	0.30	0.17	0.12
x_6	−0.12	0.00	−7.09	−0.20	−0.17	−0.11
x_7	0.17	0.02	2.87	0.15	0.10	0.23
x_8	0.85	0.09	2.81	0.85	1.30	1.23
x_9	0.20	0.04	1.59	−0.10	0.07	0.18
x_{10}	−0.45	0.03	−3.62	−0.30	−0.33	−0.57
x_{11}	0.24	0.03	2.13	0.00	0.23	0.28
x_{12}	0.50	0.05	2.75	0.45	0.60	0.54
x_{13}	0.17	0.04	1.05	1.15	0.43	−0.14
x_{14}	0.31	0.06	1.40	1.30	0.60	0.26

NOTE: *The first coefficient of $\boldsymbol{\beta}$ is set to 1 in all cases. The t-value for this coefficient under OPM is obtained prior to scaling.

Table 5. Frequency of the prediction intervals of different lengths when BFS = 1

Model		L: 0	1	2	3	4	Total
OPM	Est	903 (23%)	1325 (33%)	461 (12%)	201 (5%)	1082 (27%)	3972
	Val	636 (24%)	870 (33%)	308 (12%)	125 (5%)	709 (27%)	2648
M1	Est	611 (15%)	1033 (26%)	1848 (47%)	360 (9%)	120 (3%)	3972
	Val	434 (16%)	677 (26%)	1226 (46%)	225 (8%)	86 (3%)	2648
M2	Est	302 (8%)	1692 (43%)	1944 (49%)	33 (1%)	1 (0%)	3972
	Val	218 (8%)	1136 (43%)	1264 (48%)	29 (1%)	1 (0%)	2648

NOTE: Abbreviations: M1, Method 1; M2; Method 2, Est, Estimation Data; Val, Validation Data. Without the rounding-off errors, the percentages provided in each entry should add up to 100 in each row.

Table 5 reports the frequency of the nominal 50% prediction intervals of different length (L). The length L ranges from 0 to 4 because FS ranges from 1 to 5. From Table 5, we see that the length 4 appears much less frequently under the TORQUE model than under the OPM. Since $L = 4$ means that the prediction interval covers all possible values of FS, we say that the OPM is less informative in prediction. If we compute the coverage probabilities of those prediction intervals, they are all around 80%, well exceeding the nominal level, due to the discreteness of FS.

5.4 Mortality Rates

Lee et al. (2006) developed a prognostic index for predicting mortality risk in community-dwelling older adults employing data from 11,701 of the study participants in the United States. It was based on a one-page questionnaire with 12 questions to predict the mortality risk of the elderly (see Table 6). The index weighed different mortality risk factors according to a simple point system by means of the predictors such as age, comorbidities, and (baseline) functional difficulties. The weights were determined by the odds ratio estimates from the logit model. This method is currently a standard in medical and public health literature.

We can adopt the same approach (i.e., the logit modeling) to predict the two-year mortality rates from the LSOA II data and ask how the results compare with the proposed TORQUE estimates. Note that TORQUE is applied to the five-level ordinal variable FS, but we can easily compute the probability that FS = 5 (death) based on the estimated TORQUE model. Using the same set of predictors discussed in Section 2, we assess the performance of three mortality estimates based on the following measure:

$$\mathrm{MAE}(\tilde{p}) = \sum_{i}^{N} |\hat{p}_i - I(\mathrm{FS}_i = 5)|/N,$$

where $\hat{p}_i$ is an estimator of $P(\mathrm{FS}_i = 5)$, $I(\mathrm{FS}_i = 5)$ is the observed indicator of death for the ith subject, and N is the total number of subjects in the estimation or validation data (after adjusting for the sampling weights in the LSOA II data). This measure of $\mathrm{MAE}(\tilde{p})$ is a variant of $\mathrm{MAE}(p)$ in Section 4.3, but specialized to binary responses. For the real data, we have no knowledge about the true probability of Y; therefore, we cannot evaluate $\mathrm{MAE}(p)$ directly. Instead, $\mathrm{MAE}(\tilde{p})$ measures how well our model fits the observed response $[I(\mathrm{FS}_i = 5)]$ on the training and, perhaps more sensibly, on the validation data. Table 7 shows that TORQUE outperforms both the logit and the probit models applied to the binary response for predicting mortality rates with 20% or more reductions in $\mathrm{MAE}(\tilde{p})$.

Lee et al. (2006)'s prognostic index, constructed based on the odds ratio of a logistic regression model, is a useful and handy tool in the prediction of mortality. However, its simplistic structure hinders its prediction accuracy. The proposed TORQUE

Table 6. Prognostic index to predict four-year mortality risk in community-dwelling elders

Patient characteristic	Points
Age (years)	
60–64	1
65–69	2
70–74	3
75–79	4
80–84	5
≥85	7
Male sex	2
Diabetes	1
Cancer (not including minor skin cancer)	2
Chronic lung disease (limits activities or individual requires aided oxygen)	2
Heart failure	2
Body mass index < 25 kg/m^2	1
Current smoker	2
Functional difficulties caused by health or memory problems	
Bathing	2
Managing money or finances	2
Walking several blocks	2
Pulling or pushing large objects (e.g., living room chair)	1

Point total	Total: Predicted four-year mortality risk (%)
0 to 5	<4
6 to 9	15
10 to 13	42
≥14	64

NOTE: Adapted from Lee et al. Development and validation of a prognostic index for four-year mortality in older adults (Lee et al. 2006, p. 805).

Table 7. Performance measure MAE($\tilde{p}$) for the prediction of mortality when BFS = 1

Method	TORQUE		Probit		Logit		OPM	
Data sets	Est	Val	Est	Val	Est	Val	Est	Val
MAE($\tilde{p}$)	0.24	0.25	0.31	0.32	0.31	0.32	0.31	0.31

approach has been shown to improve on the logit and probit models based on the same set of predictors. To see whether this improvement was due to the use of 5-level ordinal scales, we also included the results from the OPM based on the same ordinal scales. It is clear from the comparison that the improvement was due to the flexibility of the TORQUE model.

5.5 Challenges in the Analysis

Our proposed TORQUE model is able to provide a useful predictive model for the functional status of the elderly based on the LSOA II data, but a number of challenges remain for further studies.

In our analysis of the LSOA II data, we have restricted attention to an additive index of the predictors, and no interaction terms have been used. Variable selection and model building for the TORQUE model remain an open topic for research.

Missing information occurred for more than 10% of the participants in the first followup. We have implicitly assumed in our analysis that they were missing at random. A more careful study of the participants with partially observed data would be helpful. If the functional status variable is not missing, we may model the binary predictors and the continuous predictors by log-linear models and then multivariate normal (conditional on the binary predictors) to perform multiple imputation as described in Schafer (1997).

The LSOA II data, and future LSOA studies of similar nature, allow us to update the prediction of the functional status every two years. The TORQUE model can be fitted to the first and second followups simultaneously to predict FS in two consecutive two-year windows. If a number of followups were available, an appropriate Markovian model might prove to be more useful to model the transition between the levels of FS. A Markovian model would allow better handling of missing data and eliminate the need to consider subgroups. The main challenge is finding good specifications of the conditional transition probabilities.

Another potentially useful modification is to account for the varying time lags between two consecutive measurements. In reality, the time lags between baseline and the first followup are not constant. The exact times of measurements per subject are not available in the published LSOA II data, but the time stamps can easily be made available in any future studies of this type. When the time stamps are available, the TORQUE model can include time differences between two measurements as a predictor similar to the approach of Wei and He (2006).

6. CONCLUSION

In this article we have provided a new approach for analyzing the functional status for the elderly from the LSOA II data where the functional status is treated naturally as an ordinal response. The proposed model generalizes the ordered probit or logit model, and features jittering, a nonparametric link function, and semiparametric quantile estimation. However, it uses a linear index of the predictors to control the model complexity. We demonstrated through simulation experiments that the proposed method works well for data fitting and prediction in a variety of settings, and a comparison with the ordered probit model showed that the method led to more informative prediction of the functional status, as well as the mortality rates from the LSOA II data.

The large-sample properties of the proposed method given in this article rely on two basic conditions, the iid error structure and the linearity condition of the design $\mathbf{X}$ in Model (1). We do not expect these conditions to hold exactly in applications. We suggested TORQUE with quantile regression as a robust procedure against heteroscedasticity. The linearity condition is assumed to enable us to use a simple initial estimate of $\boldsymbol{\beta}_0$. We have some empirical evidence, consistent with the findings in the dimension reduction literature, that violations of the linearity condition on $\mathbf{X}$ are often quite benign in how they affect the proposed method. Concerned users of the TORQUE model may seek other preliminary estimators of $\boldsymbol{\beta}_0$ that are consistent at given quantile levels without the linearity condition used in this paper.

There are at least two ways that our work can help clinicians and policy makers for the care of the elderly. First, TORQUE enables clinicians and policy makers to construct a better prognostic index for the mortality rates than do conventional methods like the logit model. Second, the prediction intervals given by TORQUE for the levels of the functional status are more informative than any point prediction. For example, the ability to say with sufficient confidence that an individual with given predictors will have a functional status at least as good (or poor) as its current level in two years can mean a lot to a clinician who needs to communicate with caretakers of the elderly. The prediction intervals obtained from TORQUE are much less sensitive to model misspecification than the ordered probit or logit models.

Finally, we note that the analysis of the LSOA II data in this paper uses existing studies in the aging research literature for selecting predictors. Generalization to multi-index models and the issue of variable selection within the semiparametric models require future research.

APPENDIX A: TECHNICAL CONDITIONS

The consistency result given in Section 4.2 is established under a series of conditions on the model. We refer to Section 2 for the basic conditions A0 and A1. The rest of the conditions are listed in this section.

A2. There exists a constant $C > 0$ such that

$$\inf_{\|\boldsymbol{\gamma}\|=1} \frac{1}{n}\sum_{i=1}^{n} |\mathbf{X}_i^T \boldsymbol{\gamma}| > C \quad \text{for all } n \text{ almost surely.}$$

A3. $\boldsymbol{\beta}_0$ is the unique minimizer of $E[\rho_\tau(\Lambda(\tilde{Y}_i) - \mathbf{X}_i^T\boldsymbol{\beta}) - \rho_\tau(\Lambda(\tilde{Y}_i))]$.

B1. By scale normalization, we assume that the first coefficient of $\boldsymbol{\beta}_0$ is 1, and the distribution of x_1 conditional on $\mathbf{X}_{-1} = (x_2, \ldots, x_p)$ has an everywhere positive density with respect to the Lebesgue measure. Also, the support of $\mathbf{X}$ is not contained in any proper linear subspace of $\mathcal{R}^p$.

B2. $\mathbf{X}$ has finite third moments.

B3. The function $\Lambda(\cdot)$ is strictly increasing on $[\tilde{y}_a, \tilde{y}_b]$, the support of $\tilde{Y}$, and $\Lambda(\tilde{y}_0) = 0$ by location normalization. Furthermore, there exists a positive number ϵ^* and a compact interval M_Λ such that $[\Lambda(\tilde{y}_a - \epsilon^*), \Lambda(\tilde{y}_b + \epsilon^*)] \subset M_\Lambda$.

B4. The conditional density of $\mathbf{Z} = \mathbf{X}^T\boldsymbol{\beta}$ given $\mathbf{X}_{-1} = \mathbf{t} \in \mathcal{R}^{p-1}$ for any $\mathbf{t}$ and the density of ϵ, $p_t(s|\mathbf{t})$ and $f(s)$, are twice continuously differentiable in s, and the derivatives are uniformly bounded.

B5. $V(\tilde{y}) = -\int f(-z)p(z + \Lambda(\tilde{y}_0))p(z + \Lambda(\tilde{y}))\,dz$ is negative for each $\tilde{y} \in [\tilde{y}_a, \tilde{y}_b]$, and uniformly bounded away from zero.

Condition A1 enables us to use the results of Li and Duan (1989) to verify that the initial estimate $\mathbf{b}_n$ is consistent (in direction). Conditions B1–B5 are sufficient to verify the uniform consistency of $\hat{\Lambda}_n$ using the results of Chen (2002). Conditions A2 and A3 ensure that the quantile estimates $\hat{\boldsymbol{\beta}}_{n,\tau}$ are consistent. For an explanation of Conditions B1–B5, we refer to Chen (2002).

APPENDIX B: SKETCH OF PROOF FOR THEOREM 1

By Li and Duan (1989, theorem 5.1), the initial estimate $\mathbf{b}_n$ converges at the root-n rate to $\boldsymbol{\beta}_0$ in direction due to Condition A1. By Chen (2002, theorem 1, p. 1687), we have a Bahadur representation of $n^{1/2}(\hat{\Lambda}_n(\tilde{y}) - \Lambda(\tilde{y}))$, which implies (10). It remains to show (11). First, we show consistency.

Let $\epsilon_0 = \min\{\tau, 1-\tau\}/2$. By the definition of ρ_τ,

$$\rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i) - \mathbf{X}_i^T\boldsymbol{\beta}) \ge (1-\tau)|\hat{\Lambda}(\tilde{Y}_i) - \mathbf{X}_i^T\boldsymbol{\beta}| \ge \epsilon_0|\mathbf{X}_i^T\boldsymbol{\beta}|,$$

when $|\mathbf{X}^T\boldsymbol{\beta}|$ is sufficiently large. Let $\boldsymbol{\gamma} = \boldsymbol{\beta}/\|\boldsymbol{\beta}\|$. Then

$$\begin{aligned}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i) - \mathbf{X}_i^T\boldsymbol{\beta}) - \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i)) &\ge \epsilon_0 \sum_{i=1}^n |\mathbf{X}_i^T\boldsymbol{\beta}| - \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i)) \\ &= \|\boldsymbol{\beta}\|\epsilon_0 \sum_{i=1}^n |\mathbf{X}_i^T\boldsymbol{\gamma}| - \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i)).\end{aligned}$$

By the assumption A2, if $\|\boldsymbol{\beta}\| > \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}_i))C/\epsilon_0$, then the above quantity will be positive, which means $\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}) > \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}))$. However, $\hat{\boldsymbol{\beta}}_{n,\tau}$ is chosen to minimize the objective function, thus

$$\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}) < \sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y})).$$

This conversely implies $\|\hat{\boldsymbol{\beta}}_{n,\tau}\| < M$, for some positive constant M.

Suppose that $\|\hat{\boldsymbol{\beta}}_{n,\tau} - \boldsymbol{\beta}_0\| \ge \epsilon_0$, along a subsequent of n, still denoted by n for simplicity. If $\hat{\boldsymbol{\beta}}_{n,\tau}$ does not converge to $\boldsymbol{\beta}_0$, due to the boundedness of $\hat{\boldsymbol{\beta}}_{n,\tau}$, there exists a further subsequence, still called $\hat{\boldsymbol{\beta}}_{n,\tau}$, such that $\hat{\boldsymbol{\beta}}_{n,\tau} \to \boldsymbol{\beta}_1 \ne \boldsymbol{\beta}_0$. By the continuity of $E\rho_\tau(\cdot)$ and the uniqueness of $\boldsymbol{\beta}_0$, we have

$$E\big[\rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_1) - \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0)\big] > \eta_0 > 0. \quad \text{(B.1)}$$

By the law of large numbers, $n^{-1}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0) \to E\rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0)$, and

$$n^{-1}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}) \to E\rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_1).$$

Then we obtain from (B.1), for sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0) < \frac{1}{n}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}) - \frac{\eta_0}{2}. \quad \text{(B.2)}$$

By (10), we have, for sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0) < \frac{1}{n}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0) + \frac{\eta_0}{4} \quad \text{(B.3)}$$

and

$$\frac{1}{n}\sum_{i=1}^n \rho_\tau(\Lambda(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}) - \frac{\eta_0}{4} < \frac{1}{n}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}). \quad \text{(B.4)}$$

Combining (B.2)–(B.4) implies that if $\hat{\boldsymbol{\beta}}_{n,\tau}$ does not converge to $\boldsymbol{\beta}_0$, then

$$\frac{1}{n}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\boldsymbol{\beta}_0) < \frac{1}{n}\sum_{i=1}^n \rho_\tau(\hat{\Lambda}_n(\tilde{Y}) - \mathbf{X}^T\hat{\boldsymbol{\beta}}_{n,\tau}).$$

This contradicts the definition of $\hat{\boldsymbol{\beta}}_{n,\tau}$. Therefore $\hat{\boldsymbol{\beta}}_{n,\tau}$ is consistent. Now making use of (10) and following He and Shao (1996), we arrive at (11). In fact, a Bahadur representation of $\hat{\boldsymbol{\beta}}_{n,\tau}$ can also be obtained, but the influence function involved in the representation is complicated in form and its asymptotic variance is difficult to estimate. We forgo the details in this paper.

[Received November 2008. Revised November 2009.]

REFERENCES

Anderson, R. T., James, M. K., Miller, M. E., Worley, A. S., and Longino Jr., C. F. (1998), "The Timing of Change: Patterns in Transitions in Functional Status Among Elderly Persons," *Journals of Gerontology, Ser. B*, 53, 17–27. [930,932]

Bierman, A. (2001), "Functional Status: The Sixth Vital Sign," *Journal of General Internal Medicine*, 16 (11), 785–786. [930]

Chen, S. (2002), "Rank Estimation of Transformation Models," *Econometrica*, 70, 1683–1697. [933,940]

Cook, R. (1998), *Regression Graphics: Ideas for Studying Regressions Through Graphics*, New York: Wiley-Interscience. [934]

Dawson, B., and Trapp, R. (2004), *Basic & Clinical Biostatistics*, New York: McGraw-Hill. [930]

Dellapasqua, S., Colleoni, M., and Goldhirsch, A. (2006), "Focusing on Physical Function Limitations in Elderly Women Surviving Cancer: Any Opportunity for Improvement?" *Nature Clinical Practice Oncology*, 3 (11), 598–599. [935]

Duan, N., and Li, K. C. (1987), "Distribution-Free and Link-Free Estimation for the Sample Selection Model," *Journal of Econometrics*, 35, 25–35. [930]

Hall, P., and Li, K. (1993), "On Almost Linearity of Low Dimensional Projections From High Dimensional Data," *The Annals of Statistics*, 21 (2), 867–889. [934]

He, X., and Shao, Q. (1996), "A General Bahadur Representation of M-Estimators and Its Application to Linear Regression With Nonstochatic Designs," *The Annals of Statistics*, 24 (6), 2608–2630. [940]

Koenker, R. W., and Bassett Jr., G. (1978), "Regression Quantiles," *Econometrica*, 46 (1), 33–50. [933]

Lee, M. (1992), "Median Regression for Ordered Discrete Response," *Journal of Econometrics*, 51 (1), 59–77. [930]

Lee, S. J., Lindquist, K., Segal, M. R., and Covinsky, K. E. (2006), "Development and Validation of a Prognostic Index for 4-Year Mortality in Older Adults," *The Journal of the American Medical Association*, 295, 801–808. [932,938]

Li, K. C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics*, 17 (3), 1009–1052. [940]

Machado, J. A. F., and Santos Silva, J. M. C. (2005), "Quantiles for Counts," *Journal of the American Statistical Association*, 100, 1226–1237. [930,934]

McCullagh, P., and Nelder, J. (1989), *Generalized Linear Models*, London: Chapman & Hall. [930]

Melenberg, B., and van Soest, A. (1996), "Measuring the Costs of Children: Parametric and Semiparametric Estimators," *Statistica Neerlandica*, 50, 171–192. [930]

Mor, V., Wilcox, V., Rakowski, W., and Hiris, J. (1994), "Functional Transitions Among the Elderly: Patterns, Predictors, and Related Hospital Use," *American Journal of Public Health*, 84 (8), 1274–1280. [930]

Mu, Y., and He, X. (2007), "Power Transformation Toward a Linear Regression Quantile," *Journal of the American Statistical Association*, 102, 269–279. [930,933,934]

Schafer, J. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman & Hall. [939]

Skinner, C., Holt, D., and Smith, T. (1989), *Analysis of Complex Surveys*, New York: Wiley. [931]

Wei, Y., and He, X. (2006), "Conditional Growth Charts" (with discussion), *The Annals of Statistics*, 34 (6), 2069–2097. [939]