*Journal of Applied Statistics*, 2013
Vol. 40, No. 6, 1231–1245, http://dx.doi.org/10.1080/02664763.2013.785489

Taylor & Francis
Taylor & Francis Group

# A multi-index model for quantile regression with ordinal data

Hyokyoung Grace Hong[a]* and Jianhui Zhou[b]

*[a]Department of Statistics and Computer Information Systems, The City University of New York, One Bernard Baruch Way, Box 11-220, New York, NY 10010, USA; [b]Department of Statistics, University of Virginia, 107 Halsey Hall, P.O. Box 400135, Charlottesville, VA 22904, USA*

(*Received 1 October 2011; accepted 11 March 2013*)

In this paper, we propose a quantile approach to the multi-index semiparametric model for an ordinal response variable. Permitting non-parametric transformation of the response, the proposed method achieves a root-*n* rate of convergence and has attractive robustness properties. Further, the proposed model allows additional indices to model the remaining correlations between covariates and the residuals from the single-index, considerably reducing the error variance and thus leading to more efficient prediction intervals (PIs). The utility of the model is demonstrated by estimating PIs for functional status of the elderly based on data from the second longitudinal study of aging. It is shown that the proposed multi-index model provides significantly narrower PIs than competing models. Our approach can be applied to other areas in which the distribution of future observations must be predicted from ordinal response data.

**Keywords:** dimension reduction; health economics; multi-index model; ordinal response; quantile regression

## 1. Introduction

Aging population is becoming an increasingly urgent issue in many developed countries. In 2008, people aged 65 and over consisted of more than 13% of the total US population and are projected to amount to nearly 20% by 2030 [29]. The incredible gain in life expectancy, however, may not always come with healthy later life. Despite some controversy, declining trends in disability of the elderly were generally observed from the 1980s and 1990s [5,7]. Now these positive trends may be reversing since the new cohorts, mainly from baby boomers, exhibit worse health status and more disability [28]. The increased morbidity and disability in the rapidly growing elderly population would exert enormous strain on available human and financial resources. Therefore, an accurate assessment of the distribution of true health status of the elderly provides a crucial step toward gauging the economics of aging, including the ability of the elderly to remain in the work force and the health care expenditures as well as individual well-being [21].

*Corresponding author. Email: hyokyoung.hong@baruch.cuny.edu

Functional status (FS) is one of the most commonly used measures of health status [6,31] and has been extensively utilized to assess a senior's capacity to perform self-care and physical activities [3] and to predict survival and quality of life in the elderly [26,27]. Derived from self-reports on activities of daily living (ADL) and instrumental activities of daily living (IADL), FS is often classified into an ordinal scale according to the severity ratings of disability [2,13]. Even though numerous attempts have been made to estimate the effect of covariates on FS of the elderly [2,14,22], little work has been reported on predicting future FS via statistically valid prediction intervals (PIs). Furthermore, existing approaches to FS prediction are generally based on popular ordinal models such as ordered probit or ordered logit. As such, their prediction accuracy is likely to be seriously undermined when the assumption of a homoscedastic error distribution is violated.

In an effort to increase the flexibility and improve the predictive power of traditional approaches, Hong and He [9] developed the transformed ordinal regression quantile estimator, or TORQUE, a semiparametric ordinal model that includes the ordered logit and ordered probit models as special cases. Where these traditional models focus attention on estimating mean changes in the dependent variable, TORQUE produces estimates of conditional quantiles, thus providing a more complete picture of the covariate risk factors' effects. Moreover, the quantile-based TORQUE model was shown to be useful in building PIs, consistently producing narrower PIs than those based on parametric models for data with non-Gaussian error distributions. On the other hand, as a single-index model, the assumption that the residuals are uncorrelated to that single-index somewhat limits the model's applicability.

This work differs from the previous approaches as it incorporates a multi-index model in the quantile regression framework for the prediction of (ordinal) FS of the elderly. Moreover, we employ modern modeling techniques in the TORQUE setup in that the proposed model (1) accounts explicitly for remaining correlations between the covariates and the residuals from the single-index model, (2) utilizes the canonical correlation (CANCOR) method to decide the number of dimensions needed for the multi-index model, and (3) assumes that the link functions in the multi-index model are unknown and can be estimated via non-parametric methods.

Our new method significantly reduces the error variance, thus leading to more efficient PIs compared with the TORQUE model, and enjoys a root-$n$ rate of convergence. By estimating PIs for the FS of the elderly, we will present that the proposed multi-index model indeed provides significantly shorter PIs than the ordered probit or single-index TORQUE model. Since ordinal responses appear not only in aging research but also in other health-related sciences, social studies, and business and economics, the proposed method is potentially useful for a wide class of applications.

## 2. Statistical models

### 2.1 *Predecessor models*

Consider the following single-index model.

$$\Lambda(\tilde{Y}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon, \tag{1}$$

where $\Lambda$ is a monotone function and $\tilde{Y}$ is a jittered response variable. That is, $\tilde{Y}_i = Y_i + U_i$, where $Y_i$ is the ordinal response variable taking values in $\{1, 2, 3, \ldots\}$ and $U_i \sim \mathrm{Unif}[0, 1)$ are independent random samples. The vector $\mathbf{X} = (x_1, \ldots, x_p) \in \mathcal{R}^p$ is a $p$-dimensional predictor, $\boldsymbol{\beta}_0$ is a $p \times 1$ parameter vector, and $\epsilon$ represents a random error whose cumulative distribution $F$ is not specified.

Then, the $\tau$th conditional quantile of $\tilde{Y}$ given $\mathbf{X} = \mathbf{X}$ is written as

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

for some coefficients $\alpha_\tau \in \mathcal{R}$ and $\boldsymbol{\beta}_\tau \in \mathcal{R}^p$, where $\tau \in (0, 1)$, and $Q_\tau$ denotes the $\tau$th quantile.

Model (2) is the TORQUE model of Hong and He [9], which generalizes existing ordinal response models such as ordered probit and ordered logit. The TORQUE model has a couple of advantages compared with its counterparts. It is more robust against deviations from the assumptions on a specific error distribution and the parametric transformation of the response. Furthermore, the PI for the outcome variable using the TORQUE model gives more efficient prediction.

However, we note that Model (1) assumes that the random errors are independent of the single-index $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0$. When this assumption is violated, the model can be improved by introducing additional indices as in

$$Y = g(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1, \ldots, \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_k, \epsilon), \tag{3}$$

where $k$ is the number of indices, $\epsilon$ is a random error independent of $\mathbf{X}$, and $g$ is a unknown link function. Model (3) is called a multi-index model and many popular models are special cases of Equation (3). For example, for $k = 1$ if $g$ is a linear function, it is a linear regression model; if $g$ is a nonlinear function, it is a single-index model; if $Y$ is binary and $g$ is a certain parametric choice, it can produce the logit and probit models. The multi-index model has been applied in areas such as marketing and epidemiology (see [1,19,23]). The multi-index model assumes that all the relevant information provided by $\mathbf{X}$ for predicting $Y$ is contained in the $k$ linear combinations of $\mathbf{X}$. In the current literature on multi-index models, including Ichimura and Lee [10], Poirier [25], Picone and Butler [24], Wang *et al.* [30], among others, the number of indices usually needs to be specified in advance.

Dimension reduction methods focus on the estimation of $k$ and then $\boldsymbol{\beta}_j$, $j = 1, 2, \ldots, \hat{k}$, in Model (3) without estimating the link function $g$. Many methods have been proposed to execute dimension reduction, including sliced inverse regression [17], sliced average variance estimation [4], principal Hessian directions [18], directional regression [15], contour regression [16], minimum average variance estimation [32], CANCOR method [8], and others.

### 2.2 *Proposed model*

In applications of Model (3), the required number of indices $k$ is generally unknown *a priori*. In this paper, we use the method of CANCOR to determine $k$. CANCOR is an appealing dimension reduction method due to its transparent interpretation of the estimates. It enables us to obtain the estimate of the dimension $\hat{k}$, and a set of the effective dimension reduction directions, $\boldsymbol{\beta}_j$, $j = 1, 2, \ldots, \hat{k}$, by finding the significant CANCORs between $\mathbf{X}$ and a set of B-spline basis functions of $Y$, $\pi(Y)$. The CANCORs are a sequence of maximized constrained correlations between $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_j$ and $\alpha_j^{\mathrm{T}}\pi(Y)$ subject to $\boldsymbol{\beta}_j$ and $\alpha_j$. The procedure can be carried out conveniently with existing functions in most statistical packages, such as SAS and R. The canonical direction estimates of $\mathbf{X}$ are the estimates of $\boldsymbol{\beta}_j, j = 1, 2, \ldots, \hat{k}$, where $\hat{k}$ is the number of significant CANCORs selected by the following sequential $\chi^2$ test. Denoting the estimated CANCORs by $\hat{\gamma}_i$ in a decreasing order, the $\chi^2$ test statistic is $-\{n - (p + H + m + 2)/2\}\sum_{i=s+1}^{p}\log(1 - \hat{\gamma}_i^2)$, where $n$ is the sample size, $p$ is the dimension of $\mathbf{X}$, $H$ and $m$ are the number of internal knots and the spline order, respectively, for generating the basis function, $\pi(Y)$. Under the null hypothesis of $\gamma_s^2 > \gamma_{s+1}^2 = 0$, i.e. there are $s$ significant CANCORs, the above test statistic has a $\chi^2$ distribution with $(p - s)(H + m - s - 1)$ degrees of freedom. The sequential $\chi^2$ test is performed using the above test statistic for $s = 0, 1, \ldots$, and the dimension $\hat{k}$ is selected as the smallest value of $s$ that makes the null hypothesis accepted. For detail on CANCOR and the sequential $\chi^2$ test, see Fung *et al.* [8].

Provided that the CANCOR method selects $k = 2$ in our application to the longitudinal study of aging (LSOA) II data, we explore a double-index transformed ordinal regression model

$$Y = g(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1, \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_2, \epsilon). \tag{4}$$

Among many possible structures for Model (4), we consider the following additive structure for the two indices as a simple tool to demonstrate the construction process for the double-indexed quantile model,

$$\Lambda_1(\tilde{Y}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1 + \epsilon_1 \tag{5}$$

and

$$\Lambda_2(\epsilon_1) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_2 + \epsilon_2, \tag{6}$$

for some unknown monotone transformations $\Lambda_1$ and $\Lambda_2$. The jittering in $\tilde{Y}$ is used as a convenient tool for converting discrete data into continuous data. Machado and Santos Silva [20] also applied this jittering technique to the discrete response data. See Koenker [11] for further discussion on jittering of discrete data.

We now present a set of conditions that can be used as basic building blocks for the root-$n$ rate of convergence in our proposed estimator.

(**C0**) $\epsilon_2$ are i.i.d. and independent of $\mathbf{X}$.
(**C1**) $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1$ and $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_2$ are independent.
(**C2**) Conditions **A1–A5** in Fung *et al.* [8].
(**C3**) Linearity condition of Li [17], i.e. for any given $b \in \mathcal{R}^p$, $E(\mathbf{X}^{\mathrm{T}}b \mid \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1, \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_2) = c_0 + \sum_{i=1}^{2} c_i\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_i$ for some constants $c_i$.

Conditions (**C0**) and (**C1**) together imply that $\epsilon_1$ is i.i.d. given $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1$, which leads to the root-$n$ rate of convergence in our proposed estimate. Conditions (**C2**) and (**C3**) are needed to ensure the root-$n$ rate of the initial estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ by CANCOR.

Combining Equations (5) and (6), we obtain

$$\Lambda_2(\Lambda_1(\tilde{Y}) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_2 + \epsilon_2. \tag{7}$$

Assume that the function $\Lambda_1(\cdot)$ and $\Lambda_2(\cdot)$ are strictly increasing on $[\tilde{y}_a, \tilde{y}_b]$ and $[\epsilon_{1,a}, \epsilon_{1,b}]$, the support of $\tilde{y}$ and $\epsilon_1$, respectively. The $\tau$th conditional quantile of $\tilde{Y}$ can then be expressed as

$$Q_\tau(\Lambda_2(\Lambda_1(\tilde{Y}) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1) \mid \mathbf{X}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_{2,\tau} \tag{8}$$

or

$$Q_\tau(\tilde{Y} \mid \mathbf{X}) = \Lambda_1^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1 + \Lambda_2^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_{2,\tau})). \tag{9}$$

The transition from Equation (8) to Equation (9) is justified by the monotone equivariance property of quantile regression, $Q_\tau(h(Y) \mid \mathbf{x}) = h(Q_\tau(Y \mid \mathbf{x}))$ for any monotone transformation $h$.

Finally, as shown in Hong and He [9] the conditional quantile of the ordinal response $Y$ is obtained from the jittered $Y$ by

$$Q_\tau(Y \mid \mathbf{X}) = \lfloor \Lambda_1^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1 + \Lambda_2^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_{2,\tau}))\rfloor, \tag{10}$$

where $\lfloor\cdot\rfloor$ denotes the greatest integer function.

### 2.3 *Estimation method*

Given the jittered observations $\{\tilde{Y}_i, \mathbf{X}_i\}_{i=1}^n$, the estimation of the proposed double-index transformed ordinal quantile regression model proceeds in the following steps.

- *Step 1:* Estimation of the initial estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.
Obtain the initial estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ by applying CANCOR to $\{\tilde{Y}_i, \mathbf{X}_i\}$. By Fung *et al.* [8], the initial estimates $\hat{\boldsymbol{\beta}}_1^0$ and $\hat{\boldsymbol{\beta}}_2^0$ of CANCOR are root-$n$ consistent to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively.
- *Step 2:* Estimation of $\Lambda_1$ and $\boldsymbol{\beta}_1$.
*(a)* Obtain the estimate of $\Lambda_1$ at each given $\tilde{y}$ as

$$\hat{\Lambda}_1(\tilde{y}) = \arg \max_{\Lambda_1 \in M_{\Gamma_1}} \{\Gamma_1(\tilde{y}, \Lambda_1, \hat{\boldsymbol{\beta}}_1^0)\}, \tag{11}$$

where $\Gamma_1(\tilde{y}, \Lambda_1, \hat{\boldsymbol{\beta}}_1^0) = \sum_{i \neq j}(d_{i\tilde{y}} - d_{j\tilde{y}_0}) 1\{\mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_1^0 - \mathbf{X}_j^{\mathrm{T}} \hat{\boldsymbol{\beta}}_1^0 \geq \Lambda_1\}, M_{\Gamma_1}$ is a pre-specified compact set in $\mathcal{R}^1$, $d_{i\tilde{y}} = 1\{\tilde{Y}_i \geq \tilde{y}\}$, 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 $\Lambda_1(\tilde{y}_0) = 0$. For details on estimation of $\Lambda_1$, see Hong and He [9].
*(b)* Obtain $\hat{\boldsymbol{\beta}}_1$, the median absolute deviation estimate of $\boldsymbol{\beta}_1$, by regressing $\{\hat{\Lambda}_1(\tilde{Y}_i)\}$ on $\{\mathbf{X}_i\}$.
- *Step 3:* Estimation of $\Lambda_2$ and $\boldsymbol{\beta}_{2,\tau}$.
Calculate residuals, $e_{1,i} = \hat{\Lambda}_1(\tilde{Y}_i) - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_1$, for $i = 1, 2, \ldots, n$.
*(a)* Obtain the estimate of $\Lambda_2$ at each given $e_{1,i}$ as

$$\hat{\Lambda}_2(e_{1,i}) = \arg \max_{\Lambda_2 \in M_{\Gamma_2}} \{\Gamma_2(e_{1,i}, \Lambda_2, \hat{\boldsymbol{\beta}}_2^0)\}, \tag{12}$$

where $\Gamma_2(e_{1,i}, \Lambda_2, \hat{\boldsymbol{\beta}}_2^0) = \sum_{j \neq k}(d_{e,ij} - \tilde{d}_{e,k}) 1\{\mathbf{X}_j^{\mathrm{T}} \hat{\boldsymbol{\beta}}_2^0 - \mathbf{X}_k^{\mathrm{T}} \hat{\boldsymbol{\beta}}_2^0 \geq \Lambda_2\}, M_{\Gamma_2}$ is a pre-specified compact set in $\mathcal{R}^1$, $d_{e,ij} = 1\{e_{1,j} \geq e_{1,i}\}$, and $\tilde{d}_{e,k} = 1\{e_{1,k} \geq e_{1,0}\}$ for some $e_{1,0}$ which satisfies $\Lambda_2(e_{1,0}) = 0$.
*(b)* Estimate $\boldsymbol{\beta}_{2,\tau}$ in the quantile regression of Koenker and Bassett [12] as

$$\hat{\boldsymbol{\beta}}_{2,\tau} = \arg \min_{\boldsymbol{\beta}_2 \in \mathcal{R}^p} \sum_{i=1}^{n} \rho_\tau(\hat{\Lambda}_2(e_{1,i}) - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_2), \tag{13}$$

where $\rho_\tau(r) = \tau r - r1\{r < 0\}$ is the quantile loss function.

Finally, the $\tau$th quantile of $Y$ given $\mathbf{X}$ can be estimated by substituting ($\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_{2,\tau}$, $\Lambda_1$, $\Lambda_2$) in Equation (10) with their estimates ($\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_{2,\tau}, \hat{\Lambda}_1, \hat{\Lambda}_2$) obtained above.

Note that the estimated canonical covariate $\alpha_1^{\mathrm{T}} \pi(\tilde{Y})$ in CANCOR gives an initial estimate of $\Lambda_1(\tilde{Y})$ in the B-spline space. We update $\Lambda_1(\tilde{Y})$ in Step 2 based on the rank transformation method providing a monotone function estimate, which is essential for the jittered $\tilde{Y}$ and for identifying the conditional quantiles $Q_\tau(Y \mid \mathbf{X})$ in our method. The relationship between $\alpha_2^{\mathrm{T}} \pi(\tilde{Y})$ and $\Lambda_2(\epsilon)$ is however unknown due to the assumed model structure in Equation (7) and the internal optimization constraints in the CANCOR procedure. Both of the proposed additive Models (5) and (6) and the developed algorithm can be generalized to the model with $k$ ($k > 2$) indices. For the model generalization, we can recursively regress the residuals from the previous index on the current index. Accordingly, for the algorithm generalization, we should repeat Step 2 to estimate $\Lambda_i$ and $\boldsymbol{\beta}_i$ for $i = 1, \ldots, k - 1$, and estimate $\boldsymbol{\beta}_{k,\tau}$ as in Step 3.2. The conditional quantile of $Y$ can then be estimated as in Equation (10) using the estimates ($\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{\beta}}_{k,\tau}, \hat{\Lambda}_1, \ldots, \hat{\Lambda}_{k-1}, \hat{\Lambda}_k$).

### 2.4 *Consistency*

The consistency of the proposed estimator relies on the consistency of the initial estimate $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and the function estimate $\hat{\Lambda}_k$, $k = 1, 2$. Under appropriate conditions, we have the following Proposition 2.2, which implies that the proposed conditional quantile estimates at any value of $\tau$ in (0,1) are asymptotically consistent at the root-$n$ rate. We impose the following regularity conditions in addition to (**C0**)–(**C3**) to facilitate the proofs.

(**C4**) $\int f_1(-z_1)p_1(z_1 + \Lambda_1(\tilde{y}_0))p_1(z_1 + \Lambda_1(\tilde{y}))\,\mathrm{d}z_1$ is negative for each $\tilde{y} \in [\tilde{y}_a, \tilde{y}_b]$, and uniformly bounded away from zero.

(**C5**) $\int f_2(-z_2)p_2(z_2 + \Lambda_2(\epsilon_{1,0}))p_2(z_2 + \Lambda_2(\epsilon_1))\,\mathrm{d}z_2$ is negative for each $\epsilon_1 \in [\epsilon_{1,a}, \epsilon_{1,b}]$ and uniformly bounded away from zero.

(**C6**) Let $z_k = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_k$ for $k = 1, 2$. The conditional density of $z_k$ given $\mathbf{X} = t \in \mathcal{R}^p$ for any $t$, $p_k(s_k \mid t)$, and the density of $\epsilon_k, f_k(s_k)$ for $k = 1, 2$, are twice continuously differentiable in $s_k$, and the derivatives are uniformly bounded.

(**C7**) By scale normalization, we assume that the first element in $\boldsymbol{\beta}_k$ $(k = 1, 2)$ is 1, and the distribution of $x_1$ conditional on $\mathbf{X}$ has an everywhere positive density with respect to Lebesgue measure. Also, the support of $\mathbf{X}$ is not contained in any proper linear subspace of $\mathcal{R}^p$.

(**C8**) There exists a constant $C > 0$ such that

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

(**C9**) $\boldsymbol{\beta}_1$ is the unique minimizer of $E[\rho_\tau(\Lambda_1(\tilde{Y}) - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \rho_\tau(\Lambda_1(\tilde{Y}))]$, and $\boldsymbol{\beta}_2$ is the unique minimizer of $E[\rho_\tau(\Lambda_2(\epsilon_1) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}) - \rho_\tau(\Lambda_2(\epsilon_1))]$.

For explanations of conditions (**C4**)–(**C9**), we refer to Hong and He [9].

Lemma 2.1 *Assume conditions (**C2**) and (**C3**), the initial estimates $\hat{\beta}_1^0$ and $\hat{\beta}_2^0$ are root-n consistent to $\beta_1$ and $\beta_2$ in direction.*

*Proof of Lemma 2.1* This proof is ascribed to Theorem 1 of Fung *et al.* [8]. ■

Proposition 2.2 *Under conditions (**C0**)–(**C9**),*

$$\sup_{\tilde{y}_a \le \tilde{y} \le \tilde{y}_b} |\hat{\Lambda}_1(\tilde{y}) - \Lambda_1(\tilde{y})| = O_p(n^{-1/2}),$$

$$\sup_{\epsilon_{1,a} \le \epsilon_1 \le \epsilon_{1,b}} |\hat{\Lambda}_2(\epsilon_1) - \Lambda_2(\epsilon_1)| = O_p(n^{-1/2}),$$

$$\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 = O_p(n^{-1/2}),$$

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

*Proof of Proposition 2.2* Lemma 2.1 implies that $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent estimators of $\beta_0$ and $\beta_1$, respectively. Conditions (**C0**) and (**C1**) imply that $\epsilon_1$ is i.i.d. given $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_1$ and $\epsilon_2$ is i.i.d. Thus, the conclusion of the proposition holds by sequentially applying Theorem 1 of Hong and He [9] to $(\Lambda_1, \beta_1)$ and $(\Lambda_2, \beta_{2,\tau})$. ■

The conditions (**C0**) and (**C1**) we imposed on Proposition 2.2 are needed to achieve the root-$n$ rate of convergence when $\Lambda$ is non-parametric. These conditions can be relaxed to accommodate heteroscedastic errors if $\Lambda$ is parametric. Also note that our proposed $\tau$-specific quantile estimator does not heavily rely on the i.i.d. error assumption, and thus is robust against the class of more general error distributions.

Similar to the proof of Proposition 2.2, given the consistency of the initial estimators for the general multi-index model (3) by Fung *et al.* [8], the root-$n$ consistency for $\hat{\beta}_i$, $i = 1, 2, \ldots, k$, by the proposed method can be established for the general multi-index model with $k > 2$, by iteratively applying Theorem 1 of Hong and He [9].

## 3. Monte Carlo studies

We carry out simulation studies to investigate the performance of the proposed multi-index Torque model (to be denoted by M-TORQUE). We consider two criteria for evaluating the performance. The first criterion is the mean coverage probability (denoted by $\bar{C}$), that is, the average coverage probability of the estimated $(\tau_2 - \tau_1) \cdot 100\%$ PI, which is estimated by $(\hat{Q}_{\tau_1}(Y \mid \mathbf{X}), \hat{Q}_{\tau_2}(Y \mid \mathbf{X}))$, where $\hat{Q}_\tau(Y \mid \mathbf{X})$ is computed as in Equation (10) by substituting the estimates $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_{2,\tau}, \hat{\Lambda}_1, \hat{\Lambda}_2)$.

The second criterion is the mean length of PI, denoted by $\bar{L} = \sum_{i=1}^{n} L_i/n$, where $n$ is the sample size. Here the length of a PI is simply $L = \hat{Q}_{\tau_2}(Y \mid \mathbf{X}) - \hat{Q}_{\tau_1}(Y \mid \mathbf{X})$. Thus if the response takes values $y = 1, 2, 3, 4, 5$, the possible values of $L$ would be 0, 1, 2, 3, 4. An effective method is expected to have a smaller $\bar{L}$ while maintaining the targeted coverage probability, i.e. $(\tau_2 - \tau_1) \cdot 100\%$. We considered $\tau_1 = 0.25$ and $\tau_2 = 0.75$ for a 50% PI, and $\tau_1 = 0.1$ and $\tau_2 = 0.9$ for an 80% PI.

We consider the double-index designs and compare the performance of M-TORQUE with the TORQUE for the single-index model of Hong and He [9] and the ordinal probit regression model (OPM).

For each example, the sample size is fixed at $n = 400$ for each data set, and a total of 100 data sets are generated in each study. After $\tilde{Y}$ is generated, we obtain the ordinal counts $Y = 1, 2, 3, 4, 5$ by the greatest integer function $Y = \lfloor \tilde{Y} \rfloor$ for $1 \leq \tilde{Y} < 6$, with $Y = 5$ when $\tilde{Y} \geq 5$ and $Y = 1$ when $\tilde{Y} < 1$.

*Example 1* $\sqrt{2\tilde{y}} = x_1 + x_2 + \epsilon_1, \ln(10\epsilon_1) = x_1 + 2x_2 + \epsilon_2, x_1 \sim \text{Unif}(0.5, 1), x_2 \sim \text{Unif}(0.5, 1)$, and $\epsilon_2 \sim t(1)$.

*Example 2* $\tilde{y}^2 = 10x_1 + x_2 + \epsilon_1$, $\ln(\epsilon_1) = x_1 x_2 + \epsilon_2$, $x_1 \sim \text{Ber}(0.5)$, $x_2 \sim \text{Unif}(0, 1)$, and $\epsilon_2 \sim t(1)$.

The simulation results are reported in Table 1. Due to the discrete nature of the response variables, the coverage probability tends to be higher or lower than the targeted probability.

Table 1. Results for Examples 1 and 2. The numbers reported are the mean of $\bar{C}$ and $\bar{L}$ over 100 generated data sets.

| | | 50% PI | | 80% PI | |
|---|---|---|---|---|---|
| Example | Method | $\bar{C}$ | $\bar{L}$ | $\bar{C}$ | $\bar{L}$ |
| Example 1 | OPM | 0.88 | 3.29 | 1.00 | 4.00 |
| | TORQUE | 0.84 | 3.22 | 1.00 | 4.00 |
| | M-TORQUE | 0.70 | 2.62 | 0.75 | 2.85 |
| Example 2 | OPM | 0.91 | 1.44 | 0.97 | 3.82 |
| | TORQUE | 0.79 | 0.15 | 0.95 | 3.51 |
| | M-TORQUE | 0.64 | 0.14 | 0.79 | 2.50 |

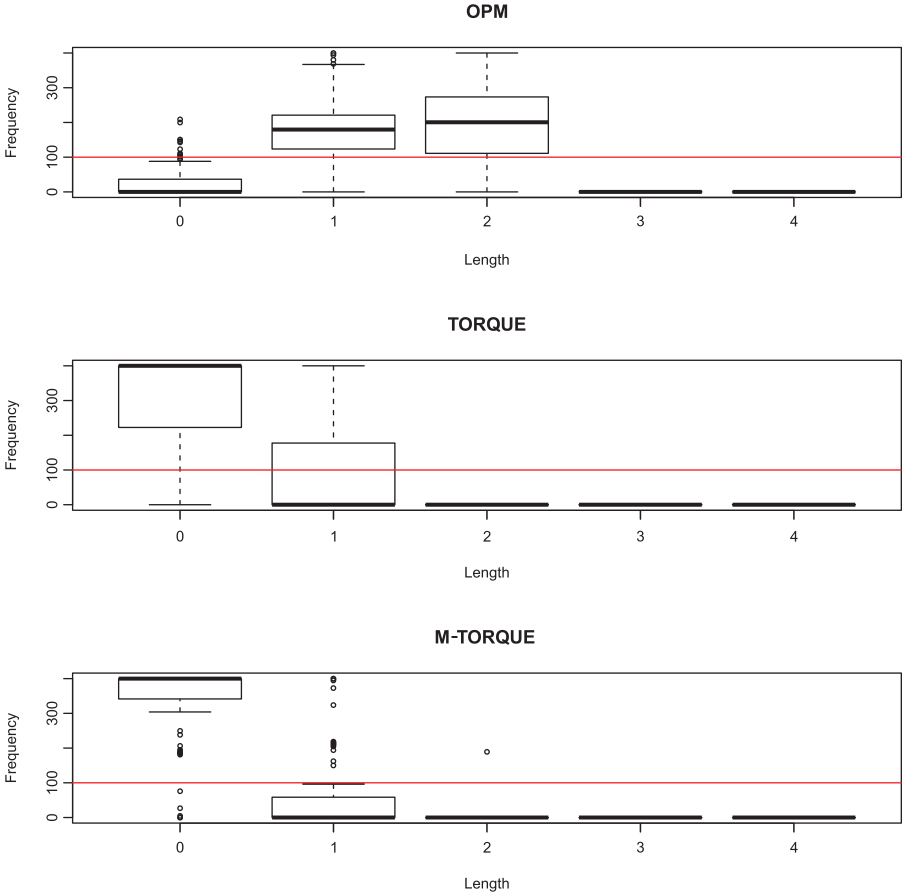

Figure 1. Example 2, boxplot of the length of the PI ($L$) for 100 generated data sets.

Example 1 is a double-index model with additive structure, which is considered in our paper. The result shows that the proposed M-TORQUE is successful to reduce the mean length, $\bar{L}$, compared with the single-index models OPM and TORQUE. For 80% PI, the mean coverage for both OPM and TORQUE is almost 100%, indicating the PI is too wide. However, M-TORQUE gives the coverage probability which is close to the targeted coverage probability while maintaining the shortest $\bar{L}$.

Example 2 is also a double-index model, but it takes a more complicated structure than the additive one considered in this paper, since $\Lambda_2(\epsilon_1)$ is a non-additive function of $x_1$ and $x_2$. Although the design we considered in Example 2 is a more general form than Model (7), the performance of M-TORQUE is quite efficient compared with OPM, whose $\bar{L}$ is more than 10 times larger than that of M-TORQUE for the 50% PI. For the 50% PI, TORQUE also performs well; however, as shown in Figure 1, M-TORQUE has the highest percentage of cases with $L = 0$ or $L = 1$. And the merits of M-TORQUE are even better illustrated when we consider the 80% PI, where the coverage probability of M-TORQUE is very close to the targeted probability.

Overall, M-TORQUE reduces the length of the PI while it achieves closer coverage probabilities to the targeted ones compared with other competitors by reducing the error variance when the true model is the double-index model.

## 4. Application to the functional status of the elderly

### 4.1 *Data description: the second longitudinal study of aging*

The data are from the second LSOA II study. The LSOA II is a collaborative project of the National Center for Health Statistics and the National Institute on Aging, which represents the national elderly population in the USA. In the baseline survey in 1994–1996 there were 9447 nationally representative, non-institutionalized United States civilian persons of age 70 years or older. Participants completed a baseline questionnaire in 1994–1996 (Wave 1) and completed two follow-up questionnaires about two years apart in 1997–1998 (Wave 2) and 1999–2000 (Wave 3). The complete set of LSOA II data is available on the LSOA website http://www.cdc.gov/nchs/lsoa.htm. Our response variable $Y$ is the FS of the elderly from Wave 2, and we use 14 covariates from Wave 1. The FS is defined in terms of ADL and IADL. The ADL is a measure of simple functions, such as bathing, dressing, eating, getting in/out of bed or chairs, and toileting. On the other hand, the IADL requires more complexity and interaction with the external environment, such as preparing meals, managing money, performing light housework, use of telephone, use of transportation, and taking medications. In this paper, we formulate the FS as follows: FS $=1$ for independent without disability; FS $=2$ for IADL disabled only; FS $=3$ for moderately ADL disabled (1–2 ADLs impaired); FS $=4$ for severely ADL disabled ($\geq$ 3 ADLs impaired); and FS $=5$ for deceased.

Initially, 9447 people participated in the survey of Wave 1. After removing 1147 subjects for missing information or drop-out, 8300 subjects were available. Among them, 6620 participants had independent FS at the baseline in the survey 1994–1996 of Wave 1. Even though other participants with different baseline functional status (BFS) would also be important, we focus on those 6620 subjects with BFS $=1$ who were healthy at the survey of Wave 1, and represent a majority of participants. A brief summary of the FS distributions at each wave is shown in Figure 2. The 14 covariates used in our analysis are described in Table 2.

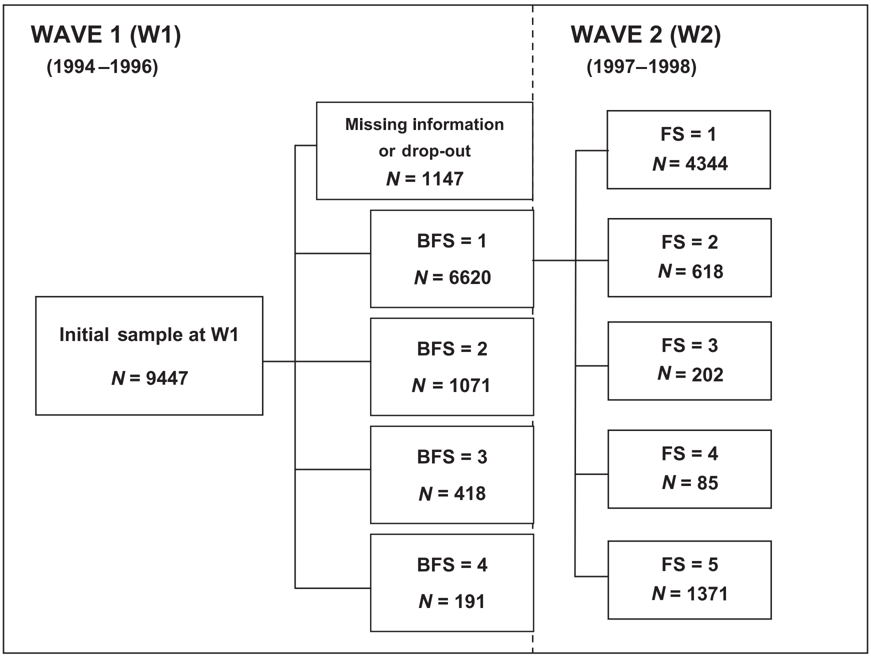

Figure 2. Flowchart of participants at Wave 1 and Wave 2 in LSOA II study. BFS, baseline functional status; FS, functional status.

Table 2. Variables used in the analysis.

| Variable | Description |
|---|---|
| SRH | 0 if excellent/very good; 1 if good/fair/poor |
| Diabetes | 0 if absent; 1 if present |
| Race | 1 if white; $-1$ if non-white |
| Marital status | 1 if married; $-1$ if not married[a] |
| Age | Years |
| Education | Years ranged from 0 to 18 |
| Sex | 1 if male; $-1$ if female |
| Cancers | 0 if absent; 1 if present |
| CVD | 0 if absent; 1 if present |
| MSD | 0 if absent; 1 if present |
| BMI | 0 if BMI $\geq$ 25; 1 if BMI $<$ 25 |
| Smoking | 0 if non-smoking; 1 if smoking |
| Condition | 0 if the total number of self-reported chronic health conditions $\leq$2;<br>1 if number of conditions $\geq$3 |
| Lung disease | 0 if absent; 1 if present |

Notes: SRH, self-rated health; CVD, cardio vascular diseases; MSD, muscular skeletal diseases; BMI, body mass index.

[a] Not married includes respondents who were widowed, divorced, separated, or never married.

### 4.2 *Results*

In this section, we apply the proposed prediction method in Section 3 to the LSOA II data as we construct the PI for the FS of the elderly over a two-year period (Wave 1 to Wave 2 in Figure 2). The sampling weights from the survey are used in the weighted CANCOR and all other estimation procedures in this section. The prediction results from the proposed method are also compared with the results from the other existing methods.

We focus our attention on the quantile levels of $\tau = 0.25$ and $\tau = 0.75$, and generate the 50% PI accordingly. Since roughly 25% of the subjects in the survey died after the first follow-up, the prediction of the upper quantiles beyond the third quartile is not useful. To validate our model performance, we divided the data randomly to form an estimation sample ($n = 3972$) and a validation sample ($n = 2648$). The estimation sample was utilized to estimate coefficients in the model, and subsequently, the FS in the validation sample was predicted using those coefficient estimates obtained from the estimation sample. For the 6620 subjects with BFS = 1, CANCOR selects two indices, suggesting that two indices are needed to thoroughly convey the predictive information contained in the original 14 covariates of the LSOA II data.

The plots of the estimated monotone transformations $\Lambda_1$ and $\Lambda_2$ in the proposed model are shown in Figures 3 and 4. The plot of estimated $\Lambda_1$ against FS on $x$ axis shows a steep increase in slope between FS = 1 and FS = 2. However, in the interval [FS = 2, FS = 5] the slope becomes flatter. This behavior in $\Lambda_1$ function suggests a lack of information in the prediction to distinguish the severities among FS being 2, 3, 4, and 5, and the predictors are most helpful in separating independent FS (FS = 1) from the poorer FS ($2 \leq \text{FS} \leq 5$). The plot of estimated $\Lambda_2$ depicts that residuals $\epsilon_1$ from Equation (12) are transformed to have a s-shaped curve, which separates the lower and higher residuals.

Table 3 shows the estimated coefficients based on OPM, TORQUE, and M-TORQUE at different quantiles. The first coefficient (SRH) is set to 1 for identification. The direction $\hat{\boldsymbol{\beta}}_2$ of M-TORUQE indicates the second most informative index since the estimated direction $\hat{\boldsymbol{\beta}}_2$ has the highest correlation with the transformed residuals, among all linear combinations that are uncorrelated to the first estimated combination. Most notably, cancer has a highly significant value at the second index, implying that cancer is the most important factor (after adjusting for the most important

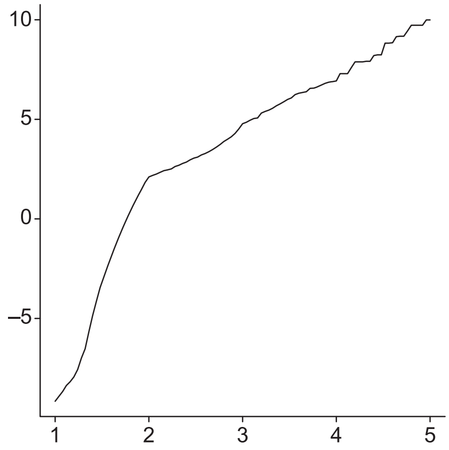

Figure 3. Plot of estimated $\Lambda_1(\tilde{Y})$.

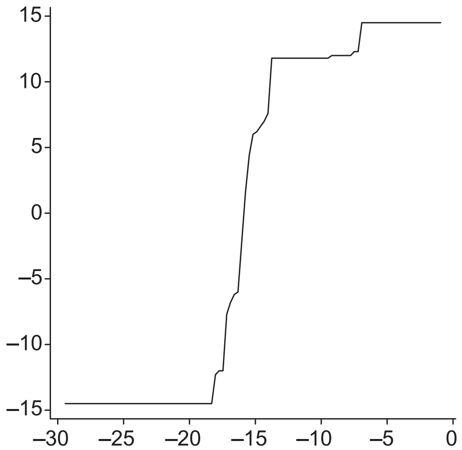

Figure 4. Plot of estimated $\Lambda_2(\epsilon_1)$.

index of covariates) in predicting FS. Thus this additional information contributes to the accurate prediction of FS of the elderly. Furthermore, smoking is not significant in the single-index, but in the second index of M-TORQUE smoking shows consistently higher coefficients than other covariates throughout all quantiles.

As shown in Table 4, M-TORQUE did not give an outstanding performance in predicting median FS compared with the other competing models, since other models were on par with M-TORQUE at predicting the ordinal FS.

However, Table 5, which reports the frequency of the number of subjects in each PI length, clearly shows the benefits of M-TORQUE. First, we define the length $L$ of the 50% PI of $Y_i$ as $|Q_{0.75}(Y_i) - Q_{0.25}(Y_i)|$. The possible values of $L$ are 0, 1, 2, 3, and 4 in our application since the FS is an integer value ranging from 1 to 5. As shown in Table 5, more than half of the subjects (58%) in the estimation data have $L = 0$, i.e. $Q_{0.25}(Y_i) = Q_{0.75}(Y_i)$, and only 2% have $L = 4$ when M-TORQUE is used. The reduced error variance from the double-index model helps to estimate the upper and lower quartiles more efficiently, thus resulting in the most informative (shortest) PI for the FS among three models.

Table 3. OPM, TORQUE, and M-TORQUE coefficients for predictors at $\tau = 0.25, 0.5, 0.75$.

| | | TORQUE | | | M-TORQUE | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | $j = 1$ | | $j = 2$ | |
| | OPM | $\tau = 0.25$ | $\tau = 0.5$ | $\tau = 0.75$ | $\tau = 0.5$ | $\tau = 0.25$ | $\tau = 0.5$ | $\tau = 0.75$ |
| SRH | $1.00_{(0.17*)}$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Diabetes | $0.56_{(0.30)}$ | 0.84 | 0.48 | 0.52 | 0.73 | −0.67 | 2.77 | 0.71 |
| Race | $-0.23_{(0.14)}$ | −0.19 | −0.15 | −0.25 | −0.15 | 0.12 | 1.11 | −0.32 |
| Married | $-0.25_{(0.09*)}$ | −0.30 | −0.27 | −0.17 | −0.19 | 2.66 | 1.01 | −0.17 |
| Education | $0.15_{(0.02*)}$ | 0.26 | 0.19 | 0.11 | 0.17 | 0.74 | −0.15 | −0.02 |
| Age | $-0.14_{(0.02*)}$ | −0.17 | −0.14 | −0.11 | −0.12 | −0.55 | 0.08 | 0.06 |
| Sex | $0.15_{(0.09)}$ | 0.16 | 0.15 | 0.18 | 0.15 | −1.02 | 0.83 | 0.11 |
| Cancer | $1.16_{(0.45*)}$ | 1.88 | 1.07 | 1.27 | 0.95 | 2.09 | 3.43 | $> 3 \times 10^2$ |
| CVD | $0.44_{(0.19*)}$ | 0.29 | 0.36 | 0.38 | 0.62 | −1.25 | 1.95 | 2.05 |
| MSD | $-0.63_{(0.18*)}$ | −0.25 | −0.29 | −0.67 | −0.20 | 0.37 | −1.00 | −0.62 |
| BMI | $0.36_{(0.17*)}$ | 0.41 | 0.38 | 0.40 | 0.38 | 0.37 | 2.55 | 0.63 |
| Smoke | $0.50_{(0.27)}$ | 0.31 | 0.52 | 0.42 | 0.83 | 3.74 | 3.01 | 2.47 |
| Condition | $0.03_{(0.24)}$ | 0.28 | 0.15 | −0.29 | −0.10 | 1.35 | −3.02 | −0.50 |
| Lung disease | $0.49_{(0.33)}$ | 0.91 | 0.50 | 0.26 | 0.58 | −0.81 | −0.53 | −0.62 |

Notes: For OPM the numbers in parentheses are standard errors. A $|t|$-value greater than 2 is marked with a ‘*’.
The slope of the first coefficient (SRH) for the three models is set to 1 for identification.
For M-TORQUE, $j$ denotes the first and second indices in our double-index model.

Table 4. Mean absolute error comparison of three models.

| | Model | | |
|---|---|---|---|
| | OPM | TORQUE | M-TORQUE |
| Estimation | 1.00 | 0.99 | 0.99 |
| Validation | 1.02 | 1.01 | 1.01 |

Table 5. Frequencies (percentage in parentheses) of PI lengths $L = 0, 1, 2, 3, 4$.

| | $L$ | | | | |
|---|---|---|---|---|---|
| Method | 0 | 1 | 2 | 3 | 4 |
| *OPM* | | | | | |
| Estimation$_{(n_1=3972)}$ | $903_{(23\%)}$ | $1325_{(33\%)}$ | $461_{(12\%)}$ | $201_{(5\%)}$ | $1082_{(27\%)}$ |
| Validation$_{(n_2=2648)}$ | $636_{(24\%)}$ | $870_{(33\%)}$ | $308_{(12\%)}$ | $125_{(5\%)}$ | $709_{(27\%)}$ |
| *TORQUE* | | | | | |
| Estimation$_{(n_1=3972)}$ | $630_{(16\%)}$ | $812_{(20\%)}$ | $1896_{(48\%)}$ | $520_{(13\%)}$ | $114_{(3\%)}$ |
| Validation$_{(n_2=2648)}$ | $449_{(17\%)}$ | $538_{(20\%)}$ | $1264_{(48\%)}$ | $308_{(12\%)}$ | $89_{(3\%)}$ |
| *M-TORQUE* | | | | | |
| Estimation$_{(n_1=3972)}$ | $2293_{(58\%)}$ | $1374_{(35\%)}$ | $178_{(4\%)}$ | $42_{(1\%)}$ | $85_{(2\%)}$ |
| Validation$_{(n_2=2648)}$ | $1625_{(61\%)}$ | $829_{(31\%)}$ | $105_{(4\%)}$ | $36_{(1\%)}$ | $53_{(2\%)}$ |

It is important to note that a shorter PI length inevitably means a lower coverage probability, and thus the coverage percentage for M-TORQUE will necessarily be lower than those of TORQUE and OPM. Therefore, we wish to investigate the agreement of the observed M-TORQUE coverage rate with the targeted coverage probability of 50%. Table 6 shows that indeed OPM outperforms

Table 6. Coverage probabilities for 50% PIs by PI length $L$.

| | OPM | | TORQUE | | M-TORQUE | |
|---|---|---|---|---|---|---|
| $L$ | Estimation | Validation | Estimation | Validation | Estimation | Validation |
| 0 | 0.83 | 0.81 | 0.84 | 0.82 | 0.75 | 0.73 |
| 1 | 0.78 | 0.78 | 0.81 | 0.79 | 0.67 | 0.68 |
| 2 | 0.77 | 0.75 | 0.76 | 0.75 | 0.71 | 0.52 |
| 3 | 0.80 | 0.75 | 0.66 | 0.64 | 0.70 | 0.66 |
| 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\bar{C}$ | 0.85 | 0.84 | 0.76 | 0.77 | 0.73 | 0.71 |
| $\bar{L}$ | 1.77 | 1.73 | 1.56 | 1.54 | 0.54 | 0.50 |

Note: $\bar{C}$ is the weighted mean coverage probability; $\bar{L}$ is the weighted mean PI length.

M-TORQUE in terms of coverage. However, as long as actual coverage is 50% or greater, the precise value is not of great interest with an ordinal response variable. Coverage probabilities of PIs for ordinal responses tend to exceed the nominal rate, and therefore a more meaningful evaluation of competing methods is given by comparing interval lengths (provided the desired coverage probability is met). From Tables 5 and 6 it is clear that M-TORQUE has the lowest mean PI length, with no sacrifice in coverage performance.

The shorter PI is particularly important in predicting the FS of the elderly as we have only five levels of functional outcome. For example, if the FS of an old individual was predicted to be ‘healthy’ but with a PI of length $L = 4$, then his/her FS could range from ‘healthy’ to ‘dead’, which is not useful in practice.

Despite the increased complexity, the new analysis provides a safeguard against the potential effects of the remaining correlation between covariates and the residuals in the single-index TORQUE model. These findings indicate that the proposed double-index transformed ordinal quantile regression model significantly improves the prediction of the FS for the elderly in the LSOA II data set.

## 5. Conclusion

Data with ordinal responses, such as FS in aging studies, are commonly used in many fields. We introduced a flexible multi-index model for transformed ordinal quantile regression, which generalizes the ordered probit or logit model, and incorporates jittering, a non-parametric link function, semiparametric quantile estimation, and dimension reduction. The application of the proposed method to the LSOA II data showed that the prediction of FS for the elderly can be meaningfully improved, compared with that based on single-index models in Hong and He [9]. Specifically, PIs for FS estimated by the proposed method achieve a shorter length while keeping almost the same coverage as the ordered probit and TORQUE models.

The models proposed here have wide applications to a variety of fields. The benefit of more efficient prediction is considerable, since a shorter PI can substantially reduce time and cost through more optimal use of existing data. For example, suppose a company is conducting a survey of consumer intention to buy certain goods by an ordinal scale: strongly disagree, disagree, neutral, agree, and strongly agree. If we can efficiently predict a range containing the true outcome (here purchasing or not purchasing goods) for a person with specified characteristics, the company can formulate a more effective marketing plan.

Undoubtedly, the gain of using multiple indices depends on how a multi-index model is structured. In this paper, we adopted an additive structure for the two indices for the sake of simplicity although it is certainly not the only way to construct the multi-index model. Prediction of the FS based on more complicated structures of the multi-index model is a subject for future study.

## Acknowledgement

The authors thank Xuming He for his helpful comments on an earlier draft of this manuscript.

## References

[1] U. Amato, A. Antoniadis, and I. De Feis, *Dimension reduction in functional regression with applications*, Comput. Stat. Data Anal. 50 (2006), pp. 2422–2446.

[2] R.T. Anderson, M.K. James, M.E. Miller, A.S. Worley, and C.F.J. Longino, *The timing of change: Patterns in transitions in functional status among elderly persons*, J. Gerontol. Ser. B 53 (1998), pp. 17–27.

[3] A. Bierman, *Functional status the sixth vital sign*, J. Gen. Intern. Med. 16 (2001), pp. 785–786.

[4] R.D. Cook and S. Weisberg, *Discussion of sliced inverse regression for dimension reduction*, J. Am. Stat. Assoc. 86 (1991), pp. 328–332.

[5] E.M. Crimmins, *Trends in the health of the elderly*, Annu. Rev. Public Health 25 (2004), pp. 79–98.

[6] M. Extermann, J. Overcash, and G. Lyman, *Comorbidity and FS are independent in older cancer patients*, J. Clin. Oncol. 16 (1998), pp. 1582–1587.

[7] V.A. Freedman, L.G. Martin, and R.F. Schoeni, *Recent trends in disability and functioning among older adults in the United State: A systematic review*, J. Am. Med. Assoc. 288 (2002), pp. 3137–3146.

[8] W.K. Fung, X. He, L. Liu, and P. Shi, *Dimension reduction based on canonical correlation*, Stat. Sin. 12 (2002), pp. 1093–1113.

[9] G.H. Hong and X. He, *Prediction of functional status for the elderly based on a new ordinal regression model*, J. Am. Stat. Assoc. 105 (2010), pp. 930–941.

[10] H. Ichimura and L. Lee, *Semiparametric least squares estimation of multiple index models: Single equation estimation*, in *Parametric and Semiparametric Methods in Econometrics and Statistics*, W.A. Barnett, J.L. Powell, and G. Tauchen, eds., Cambridge University Press, New York, 1991, pp. 3–49.

[11] R. Koenker, *Quantile Regression*, Cambridge University Press, New York, 2005.

[12] R. Koenker and G. Bassett Jr., *Regression quantiles*, Econometrica 46 (1978), pp. 33–50.

[13] R. Koenker and Q. Zhao, *L-estimation for linear heteroscedastic models*, J. Nonparametric Stat. 3 (1994), pp. 223–235.

[14] Y. Lee, *The predictive value of self assessed general, physical, and mental health on functional decline and mortality in older adults*, Epidemiol. Community Healths 54 (2000), pp. 123–129.

[15] B. Li and S. Wang, *On directional regression for dimension reduction*, J. Am. Stat. Assoc. 102 (2007), pp. 997–1008.

[16] B. Li, H. Zha, and F. Chiaromonte, *Contour regression: A general approach to dimension reduction*, Ann. Stat. 33 (2005), pp. 1580–1616.

[17] K.C. Li, *Sliced inverse regression for dimension reduction (with discussion)*, J. Am. Stat. Assoc. 86 (1991), pp. 316–342.

[18] K.C. Li, *On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma*, J. Am. Stat. Assoc. 87 (1992), pp. 1025–1039.

[19] L. Li and H. Li, *Dimension reduction methods for microarrays with application to censored survival data*, Bioinformatics 20 (2004), pp. 3406–3412.

[20] J.A.F. Machado and J.M.C. Santos Silva, *Quantiles for counts*, J. Am. Stat. Assoc. 100 (2005), pp. 1226–1237.

[21] L.G. Martin, R.F. Schoeni, and P.M. Andreski, *Trends in health of older adults in the United States: Past, present, future*, Demography 47 (2010), pp. S17–S40.

[22] L. McGuire, E. Ford, and U. Ajan, *The impact of cognitive functioning on mortality and the development of functional disability in older adults with diabetes: The second longitudinal study on aging*, BMC Geriatr. 6 (2006), p. 8.

[23] P. Naik, M. Wedel, and W. Kamakura, *Multi-index binary response analysis of large databases*, J. Bus. Econ. Stat. 28 (2010), pp. 67–81.

[24] G. Picone and J. Butler, *Semiparametric estimation of multiple equation models*, Econ. Theory 16 (2000), pp. 551–575.

[25] D. Poirier, *Partial observability in bivariate probit models*, J. Econ. 12 (1980), pp. 209–217.

[26] D. Reuben, L. Rubenstein, and S. Hirsch, *Value of FS as a predictor of mortality: Results of a prospective study*, Am. J. Med. 93 (1992), pp. 663–669.

[27] K. Rockwood, K. Stadnyk, and C. MacKnight, *A brief clinical instrument to classify frailty in elderly people*, Lancet 353 (1999), pp. 205–206.

[28] T.E. Seeman, S.S. Merkin, E.M. Crimmins, and A.S. Karlamangla, *Disability trends among older american: National health and nutrition examination surveys, 1988–1994 and 1999–2004*, Am. J. Public Health 100 (2010), pp. 100–107.

[29] E.F. Thomson, B. Yu, A. Nuru-Jeter, J. Guralnik, and M. Minkler, *Basic ADL disability and functional limitation rates among older Americans from 2000–2005: The end of the decline?* J. Gerontol. A: Biol. Sci. Med. Sci. 64 (2009), pp. 1333–1336.

[30] J.L. Wang, L. Xue, L. Zhu, and Y. Chong, *Estimation for a partial-linear single-index model*, Ann. Stat. 38 (2010), pp. 246–274.

[31] A.W. Wu, K.A. Cagney, and S.P.D. John, *Health status assessment completing the clinical database*, J. Gen. Intern. Med. 12 (1997), pp. 254–255.

[32] Y. Xia, H. Tong, W.K. Li, and L. Zhu, *An adaptive estimation of dimension reduction space (with discussion)*, J. R. Stat. Soc. Ser. B 64 (2002), pp. 363–410.