Computational Statistics and Data Analysis 63 (2013) 50–62

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A quantile approach to the power transformed location–scale model

Hyokyoung Grace Hong

Department of Statistics and Computer Information Systems, Zicklin School of Business, Baruch College, The City University of New York, New York, NY 10010, United States

ARTICLE INFO

Article history:
Received 14 January 2012
Received in revised form 7 January 2013
Accepted 8 January 2013
Available online 4 February 2013

Keywords:
Location–scale model
Medical expenditure panel survey data
Power transformation
Quantile regression

ABSTRACT

The burgeoning growth of health care spending has become a major concern to policy makers, making the modeling of health care expenditure valuable in their decision-making processes. The challenges of health care expenditure analysis are two-fold: the exceptional skewness of its distribution as the top 5% of the population accounted for almost half of all spending and its heteroscedasticity. To address these concerns, the quantile regression model with power transformation has been employed, but at a price of the model complexity and analysis cost. In this article, we introduce a simpler quantile approach to the analysis of expenditure data by employing the location–scale model with an unknown link function to accommodate the heteroscedastic data with non-ignorable outliers. Specifically, in our approach a link function does not depend on quantiles; yet, it effectively fits the data as the slope coefficient depends on the quantiles. This parsimonious feature of our model helps us conduct a more intuitive and easily understood analysis for the whole distribution with fewer computational steps. Thus, it can be more widely applicable in practice. Additionally, simulation studies are conducted to investigate the model performance compared to other competing models. Analysis of the 2007 Medical Expenditure Panel Survey data using our model shows that aging and self-rated health tend to drive up costs. However, uninsured persons do not contribute to the high health cost. These findings suggest that careful monitoring of elderly's health status and a more aggressive preventive medicare system may contribute to slow down the explosion of medical costs.

1. Introduction

More money is spent on health care than ever in the United States. According to the recent report of World Health Statistics 2009, approximately 16% of US GDP was attributed to health care expenditure in 2007. The government, employers, and consumers alike are struggling to keep up with increasing health care costs. Therefore, many analysts have advocated controlling health care costs as a key step to economic stability.

One of the interesting features of health care spending in the US is its concentration of few individuals with extremely high spending. For instance, the 2007 Medical Expenditure Panel Survey (MEPS), available from http://www.meps.ahrq.gov/mepsweb/, suggests that the top 1% of the expenses accounted for 22% of total expenditures, the top 5% for almost 50% of all spending, but the bottom 50% of the population accounted for only 3% of total expenditures (Fig. 1). Naturally, individuals with high spending have been a focus of the analysis as they contribute disproportionately to total expenditure. In addition, some studies have reported that high expenses persist over time; individuals with high expenditures in one year are more likely to have high expenditures in subsequent years (Monheit, 2003). For these reasons,

E-mail address: hyokyoung.hong@baruch.cuny.edu.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2013.01.022

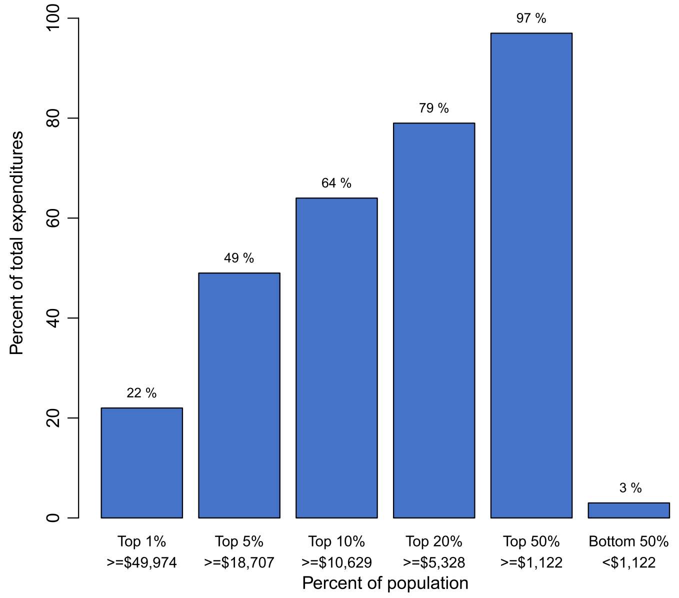

Fig. 1. Percent of total health care expenses incurred by different percentiles.

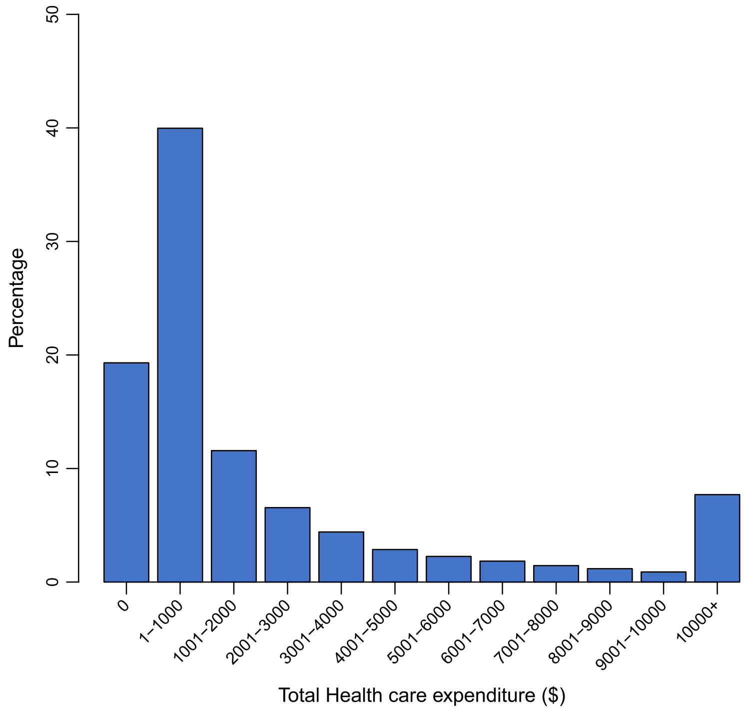

Fig. 2. Distribution of health care expenditures for the US population, 2007 MEPS.

an emphasis is placed on cost containment efforts for these individuals. There are some key challenges associated with analyzing health care expenditure data. They have a tendency to be skewed to the right due to a few cases requiring exceptionally expensive procedures (Fig. 2). Other than skewness, heteroscedasticity is also a feature of the expenditure data.

Quantile regression, first introduced by Koenker and Bassett (1978), can be an effective way to model such skewed distributions with outliers as seen in medical expenditure data. By transforming the response variable, the linearity in the quantile regression setting can be better approximated. The power transformed quantile regression (PQR), employing the power transformed response variable originated from Box and Cox (1964), was proposed in this line and implemented by Powell (1991), Chamberlain (1994), Buchinsky (1995), Machado and Mata (2000), and Mu and He (2007). The basic assumption of the PQR approach is that the transformation parameter and the regression coefficient depend on quantiles and may change from quantile to quantile. From a modeling point of view, this may provide a great deal of flexibility to the model, but the increased complexity also causes some implementation issues in practice. In contrast, Hong and He (2010) considered the quantile approach with a single transformation. However, strong assumptions associated with *iid* error terms are needed for estimation consistency. In this paper, we introduce a simple power transformed location–scale model or the power transformed heterogeneous regression model (PHR) as an alternative approach to PQR. The location–scale model has been one of the most popular classes of models in the statistics and econometrics literature (refer to discussion Ruppert and Wand, 1994; Fan and Gijbels, 1996). In the location–scale model, the location part is of primary interest to the econometrician while the scale part is considered as a nuisance. Although PQR may work well for heteroscedastic data in general, the simpler structure of PHR can be preferable to PQR for the following reasons. (i) PHR employs a single transformation, letting only a scale factor adjust different conditional quantiles of the distribution. While the transformation parameter $\lambda(\tau)$ depends on the quantile for PQR, the common transformation parameter λ for all quantiles is sufficient for PHR. When different transformations are used for different quantiles, it becomes much harder to interpret or compare the transformed responses at the quantile of interest. In contrast, the single transformation brings a more straightforward interpretation of the slope parameter. (ii) Multiple transformations and quantile-dependent slopes parameters increase the chance of quantile crossing (Bassett and Koenker, 1982) for PQR. The crossing problem can be avoided by using PHR, which naturally satisfies monotonicity (He, 1997). (iii) PHR does not require a large sample size compared to PQR. The intensive simulation studies and the applications in our paper show that the performance of the proposed PHR model is on par with that of the PQR. Further, the PHR enjoys its simplicity in structure and computation; also it does not entail the *iid* error assumption, a large sample size, and multiple transformations. Thus, PHR allows for a single transformation, while ensuring the heteroscedasticity of the data can still be allowed. It is, therefore, a good alternative to the PQR approach.

The remainder of the article is organized as follows. In Section 2, we propose the model. In Section 3, the simulation studies are illustrated. In Section 4, the health expenditure 2007 MEPS data will be analyzed. Section 5 makes some concluding remarks. The main technical details of the approach are summarized in the Appendix.

2. Proposed model

We consider location–scale models with the transformed response:

$$\Lambda_\lambda(y_i) = x_i'\beta + (x_i'\gamma)u_i, \tag{1}$$

where the transformation $\Lambda_\lambda(y)$ is a monotone and continuous function of $R \to R$ for each λ in an interval $[a, b]$, y_i is the response for the *ith* subject, x_i is its covariate, and $u_i \sim G_u$ is independent of x_i. In Model (1) we let the data determine the exact transformation entailing a linear relationship between the dependent variable and the covariates. The common example of this type of transformation is the Box–Cox transformation, and the idea is to find a parameter λ such that

$$\Lambda_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0. \end{cases}$$

This transformation requires that the values of y must be positive in order for the function to be defined everywhere.

The quantiles of Model (1) have the following form:

$$Q_\tau(\Lambda_\lambda(y_i|x_i)) = x_i'\beta + (x_i'\gamma)G_{u_i}^{-1}(\tau), \tag{2}$$

where $G_{u_i}^{-1}(0.5)$ is assumed to be zero. We refer to Model (2) as the power transformed location–scale model or power transformed heteroscedastic regression model (PHR). Using the well-known equivariance property of order statistics under a monotone transformation, the conditional quantile of y_i given x_i can be easily derived instead of $\Lambda_\lambda(y_i)$ as

$$Q_\tau(y_i|x_i) = \Lambda_\lambda^{-1}(x_i'\beta + x_i'\gamma G_{u_i}^{-1}(\tau)) = [\lambda x_i'(\beta + \gamma G_{u_i}^{-1}(\tau)) + 1]^{1/\lambda}. \tag{3}$$

As an alternative to PHR, we can consider the following models.

- The power transformed quantile regression model (PQR):

$$Q_\tau(\Lambda_{\lambda(\tau)}(y_i|x_i)) = x_i'\beta(\tau). \tag{4}$$

This is equivalent to $\Lambda_{\lambda(\tau)}(y_i|x_i) = x_i'\beta(\tau) + u_i(\tau)$, where u_i is the error term, whose τth quantile is zero conditional on x_i. Thus the conditional quantile of y_i given x_i becomes

$$Q_\tau(y_i|x_i) = \Lambda_{\lambda(\tau)}^{-1}(x_i'\beta(\tau)) = [\lambda(\tau)x_i'\beta(\tau) + 1]^{1/\lambda(\tau)}. \tag{5}$$

- The power transformed location regression model (PLR):

$$Q_\tau(\Lambda_\lambda(y_i|x_i)) = x_i'\beta + G_{u_i}^{-1}(\tau). \tag{6}$$

The conditional quantile for y_i at x_i is then given by

$$Q_\tau(y_i|x_i) = \Lambda_\lambda^{-1}(x_i'\beta + G_{u_i}^{-1}(\tau)) = [\lambda(x_i'\beta + G_{u_i}^{-1}(\tau)) + 1]^{1/\lambda}.$$

Model (4) is variant of Model (1), which considers the constant transformation parameter λ regardless of τ. This model was proposed by Buchinsky (1995), and Mu and He (2007) among others. In Model (4), both the transformation parameter $\lambda(\tau)$ and the regression coefficient $\beta(\tau)$ depend on τ and may change from quantile to quantile. Model (6), however, assumes *iid* errors, and thus a single transformation suffices and the same slope coefficients can be used for different quantiles. Therefore, the model complexity for PHR can be considered between PQR and PLR with PQR being the more complex model.

2.1. Estimation of λ, β, and γ

The joint estimation of λ, β, and γ is not easy, but for a given λ, Model (2) reduces to a simplified location–scale model.

Step 1 (*Estimating* λ *and* β). Our approach to estimating λ is similar to Powell (1991) and Chamberlain (1994). The sample estimator for λ can be obtained as a solution to

$$\hat{\lambda} = \arg\min_{\lambda} \left(\sum_{i=1}^{n} \left| \Lambda_\lambda(y_i) - x_i'\hat{b}(\lambda) \right| \right), \tag{7}$$

where

$$\hat{b}(\lambda) = \arg\min_{b} \sum_{i=1}^{n} |\Lambda_\lambda(y_i) - x_i' b|$$

is a median regression estimate of β for given λ. Clearly, $\hat{\lambda}$ can be computed using a grid search.

Step 2 (*Estimating* γ). In this step, we substitute $\hat{\lambda}$ and $\hat{\beta} = \hat{b}(\lambda)$ into Model (2) and estimate γ. We define the residuals $r_i = \Lambda_{\hat{\lambda}}(y_i) - x_i'\hat{\beta}$. Regress $|r_i|'s$ on $x_i's$ to obtain the median regression coefficient $\hat{\gamma}$ and the fitted values $s_i = x_i'\hat{\gamma}$. Note that this median regression does not guarantee the fitted function s to be nonnegative. In this case, some ad hoc approaches such as replacing s_i with $\max\{s_i, \min_j\{s_j > 0\}\}$ would be used to exclude the negative s. A detailed discussion on the treatment of such cases can be found in He (1997).

Step 3 (*Estimating* $Q_\tau(y_i|x_i)$). Choose $\hat{\eta}_\tau$ to minimize $\Sigma_i \rho_\tau(r_i - \eta s_i)$ over η, where $\rho_\tau(t) = t\,\{\tau - I(t < 0)\}$ is the "check function" (Koenker and Bassett, 1978), for $I(A)$ denoting the indicator function of the event A (it takes the value one if A is true, and is zero otherwise).

Finally, the τth quantile of y_i has the form of

$$[\hat{\lambda} x_i'(\hat{\beta} + \hat{\gamma}\hat{\eta}_\tau) + 1]^{1/\hat{\lambda}}. \tag{8}$$

2.2. Consistency

We shall now prove the consistency of λ, β, and γ under the following assumptions. We assume the following conditions.

A1. u has a median of 0.
A2. $x_i'\gamma > 0$ for all $x \in D_o$, where D_o is the compact domain. $i = 1, \ldots, n$.
A3. There exists a constant $C > 0$ such that $\inf_{\|\delta\|=1} n^{-1} \sum_{i=1}^{n} |x_i'\delta| > C$ for all n almost surely.
A4. β_0 is the unique minimizer of $E[\rho_\tau(\Lambda_\lambda(y_i) - x_i'\beta) - \rho_\tau(\Lambda_\lambda(y_i))]$.

Theorem 2.1. *Consider the location model from* (6)

$$\Lambda_\lambda(y) = x_i'\beta_0 + u_i.$$

Under conditions A1–A4 *and a given consistent estimator* $\hat{\lambda}$ *for* λ,

$$\hat{\lambda} = \arg\min_{\lambda} \left(\sum_{i=1}^{n} \left| \Lambda_\lambda(y_i) - x_i'\hat{b}(\lambda) \right| \right), \tag{9}$$

where

$$\hat{b}(\lambda) = \arg\min_{b} \sum_{i=1}^{n} |\Lambda_\lambda(y_i) - x_i' b|$$

is a median regression estimate of β_0 *for given* λ.

The consistent estimator of β_0 *is*

$$\hat{\beta}_\tau = \arg\min_{\beta} \sum \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\beta).$$

The additional conditions are stated as B1–B4.

B1. The random vector x_i is independently distributed across i.
B2. The parameter vector $\theta = (\beta, \lambda)$ is an element of a compact parameter space Θ.

B3. For all i, the transformation $\Lambda_\lambda(x_i'\beta)$ is well defined for all $\theta \in \Theta$ and is continuous in Θ.

B4. For $\theta \in \Theta$, there is some random variable $L(x_i)$ such that $|\Lambda_\lambda(x_i'\beta)| \leq L(x_i)$ with $E[L(x_i)]^r \leq \bar{L} < \infty$, for some $\bar{L}$ and $r < 1$. Furthermore, the function

$$\frac{1}{n}\sum_{i=1}^{n}(\Lambda_\lambda(x_i'\beta) - \Lambda_{\lambda^*}(x_i'\beta^*))^2$$

is continuous in $\theta \in \Theta$ for every $\theta^* \in \Theta$ and uniformly for n.

The conditions imposed here are quite similar to those imposed by Powell (1991) and Oberhofer (1982), except to the extent that the conditions are stated in terms of the Box–Cox transformation parameter λ rather than the general transformation of y.

Theorem 2.2. *Consider Model (2). Under conditions* B1–B4 *and the conditions of Theorem 2.1, as n goes to infinity,* $\hat{\lambda} \to \lambda$, $\hat{\beta} \to \beta$, $\hat{\gamma} \to \gamma$, *and the* τ*th quantile is consistent for the true conditional quantile function* $\Lambda_\lambda^{-1}(x_i'\beta + x_i'\gamma G_{u_i}^{-1}(\tau))$, *where* $G_u^{-1}(\tau)$ *is the* τ*th quantile of the error distribution of* u.

3. Simulation studies

Simulation studies are conducted to compare the numerical performance of the proposed estimator versus two competing models, PQR and PLR. Throughout each simulation, bivariate observations (x_i, y_i) with sample sizes of $n = 120$ and $n = 600$ are generated according to the following model:

Study 1: $y^{0.5} = 10 + 2x + u, u \sim N(0, 1)$,
Study 2: $\log(10y) = 0.3x + u, u \sim N(0, 1)$,
Study 3: $y^{0.5} = 10 + 2x + (x/5)u, u \sim N(0, 1)$,
Study 4: $10\log(10y) = 3x + (x/6)u, u \sim U(-0.5, 0.5)$,
Study 5: $y^{0.5} = 10 + 2x + 0.1(3 + \log(x))u, u \sim N(0, 1)$,
Study 6: $y^{0.5} = 10 + 2x + 0.1(x^2 + x + 1)u, u \sim N(0, 1)$,

where the independent variable x takes the values $x_i = i/5$, $i = 1, \ldots, n$. We repeat each scenario 400 times.

The performance of the three models, PQR, PLR, and PHR are compared using the coverage probabilities obtained for each technique. Let $C_{\mathcal{D}}(\alpha_0, \alpha_1)$ denote the simulated (empirical) coverage corresponding to the $(\alpha_1 - \alpha_0)$% nominal prediction interval (PI) obtained within the domain of $\mathcal{D}$. In Table 1, the simulation results for the coverage probability of the 90% PI including the mean squared error (MSE) are presented. For Studies 1–6, all three models have satisfactory overall coverage probabilities in the domain of $S = S_1 \cup S_2 \cup S_3$, where $S_1 = \{x_i,\ 1 \leq i \leq n/3\}, S_2 = \{x_i, n/3 + 1 \leq i \leq 2n/3\}$, and $S_3 = \{x_i, 2n/3 + 1 \leq i \leq n\}$. Thus, the subdomains of S, i.e., S_1, S_2, and S_3, are three mutually exclusive sets with the same number of data values.

Studies with the heteroscedastic errors (i.e., Studies 3–4) and Study 6 are designed to mimic the distribution of the medical expenditure having greater variability at older ages; S_1 would be representing the young–old and S_3 for the oldest–old group.

The results are reported in Tables 1–3 and include the average coverage probabilities of the simulated estimates and $1000\times$ mean square error $(\text{mean}(\hat{\omega} - \omega_0)^2)$, where $\hat{\omega}$ is the observed coverage probabilities and ω_0 is the targeted coverage probability. Coverage probabilities are investigated on the whole (S) and on the subdomains (S_1, S_2, and S_3) for each study. A number of observations can be made from Table 1. (i) Three models, PQR, PLR, and PHR, have satisfactory overall coverages, but greater variability in coverage probabilities at the subdomains are observed. (ii) As expected PLR performs well when the error distribution is homoscedastic compared to studies with the heteroscedastic error. When the scale term is not equal to 1 (Studies 3–6), PLR's coverage probabilities within the subdomains tend to be over/under-estimated. Figs. 3 and 4 shows the data generated from Study 4 with different sample sizes, along with the lower and upper quantiles, $Q_{0.05}(y|x)$ and $Q_{0.95}(y|x)$. The 90% prediction interval (PI), $(Q_{0.05}(y|x), Q_{0.95}(y|x))$ for PLR is shown as the short dotted lines. At small x values (S_1), PLR's PI covers the most samples, but at very high x values (S_3) the coverage probabilities are far beyond the targeted coverage probabilities of 0.90. Apparently, PLR fails to account for non-constant variability and tends to overestimate the coverage probability at S_1 and underestimate it at S_2 compared to PQR and PHR. These latter models have more flexibility in fitting the quantiles that can be adjusted for the heteroscedastic data characteristics. (iii) PQR performs reasonably well, but for most studies the coverage probabilities of PHR are closer to 0.90 at each subdomain than PQR when the sample sizes are small ($n = 120$).

Table 2 shows that with the larger sample size ($n = 600$) the performance of PQR is improved compared to the smaller sample size. Particularly, when the scale term is non-linear (Study 6), the performance of PQR is best among the other models. However, the results from Study 5 indicate that a nonlinear scale term can actually be favorable to PHR. The data plots of Study 5 and Study 6 (not shown here) suggest a greater degree of non-constant error variances for Study 6 than Study 5.

On another note, in the medical cost application it is often of interest to estimate high quantiles. For example, the 0.95th quantile of the total medical spending ($Q_{0.95}(y|x)$) consists of a considerable percentage of the total medical costs in the

Table 1
Coverage of Monte Carlo: $C_{\mathcal{D}} = C_{\mathcal{D}}(\alpha_0, \alpha_1)$ with $(\alpha_0, \alpha_1) = (0.05, 0.95)$ with $n = 120$. In simulation, 400 replications have been carried out. The coverage probabilities are investigated at the whole domain, $S = \{S_1, S_2, S_3\}$ as well as $S_1 = \{x_i, i = 1, \dots, 40\}$, $S_2 = \{x_i, i = 41, \dots, 80\}$, and $S_3 = \{x_i, i = 81, \dots, 120\}$. The numbers in parentheses are 1000 × MSE.

$C_{\mathcal{D}}$	Model	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6
C_S	PQR	0.905	0.902	0.903	0.902	0.904	0.908
		(0.124)	(0.091)	(0.010)	(0.087)	(0.121)	(0.131)
	PLR	0.900	0.900	0.900	0.900	0.900	0.900
		(0)	(0)	(0)	(0)	(0)	(0)
	PHR	0.898	0.898	0.885	0.886	0.901	0.838
		(0.006)	(0.121)	(0.944)	(0.979)	(0.100)	(0.552)
C_{S_1}	PQR	0.915	0.934	0.935	0.941	0.927	0.933
		(0.848)	(1.496)	(1.434)	(1.832)	(1.242)	(1.489)
	PLR	0.902	0.946	1.000	0.999	0.954	1.000
		(1.481)	(3.226)	(10)	(9.846)	(3.698)	(9.967)
	PHR	0.900	0.898	0.847	0.856	0.900	0.737
		(2.453)	(3.184)	(10.490)	(10.832)	(3.071)	(39.303)
C_{S_2}	PQR	0.888	0.865	0.868	0.852	0.871	0.920
		(1.212)	(1.696)	(1.618)	(2.757)	(1.679)	(1.643)
	PLR	0.902	0.906	0.970	0.933	0.892	0.961
		(1.468)	(1.464)	(5.654)	(1.871)	(1.625)	(5.170)
	PHR	0.901	0.903	0.902	0.911	0.890	0.905
		(1.532)	(1.465)	(1.406)	(1.268)	(1.598)	(1.965)
C_{S_3}	PQR	0.911	0.908	0.905	0.912	0.913	0.871
		(0.748)	(0.467)	(0.364)	(0.301)	(0.614)	(1.682)
	PLR	0.896	0.848	0.730	0.768	0.854	0.739
		(1.268)	(4.521)	(29.592)	(18.215)	(3.473)	(27.356)
	PHR	0.894	0.895	0.907	0.889	0.914	0.874
		(2.475)	(1.976)	(0.667)	(1.010)	(1.632)	(2.184)

Note: $C_{\mathcal{D}}(\alpha_0, \alpha_1)$ denotes the coverage probability for prediction intervals for a response variable between α_0- and α_1-th quantile at the domain $\mathcal{D}$.
The results for PLR at the whole C_S are mostly 0.9 throughout all 400 repetitions. Thus the zero MSE are reported.
PHR seems to have a less satisfactory coverage at the entire domain compared to the competing models, but the coverage is more close to the targeted coverage at the subdomains for the most studies.

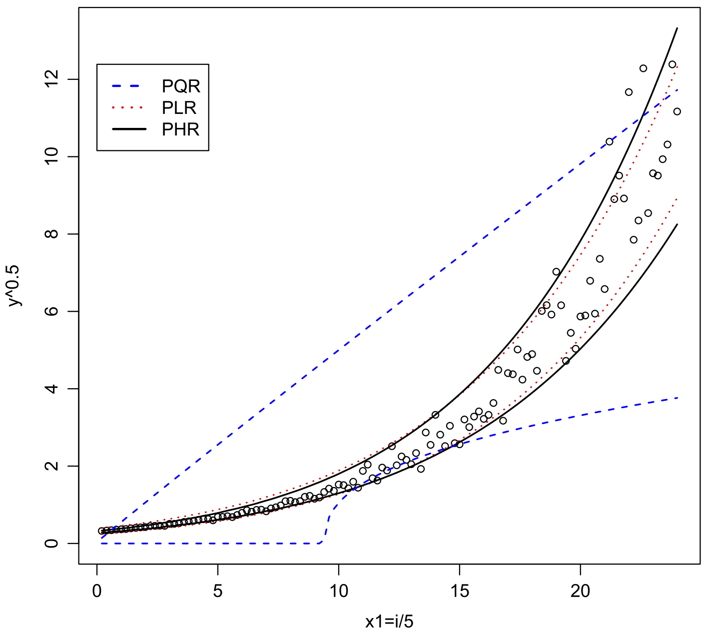

Fig. 3. Plot of curves from the three models at lower and upper quantiles using Study 4 with $n = 120$. Note that the inverse transformation function in Models (2), (4) and (6), $\Lambda^{-1}_{\lambda(\tau)}(t) = (\lambda(\tau)t + 1)^{1/\lambda(\tau)}$ (if PHR or PLR, $\tau = 0.5$) is not well defined if $\lambda(\tau)t + 1 < 0$. Therefore, we set $\Lambda^{-1}_{\lambda(\tau)}$ to be zero if $\lambda(\tau)t + 1 < 0$. This event happens to PQR frequently at the lower quantile.

population. Failure to accurately predict this population whose medical spending costs are extremely high can cause large economic losses. This highly skewed distribution motivates us to investigate coverage probability beyond $Q_{0.95}(y|x)$.

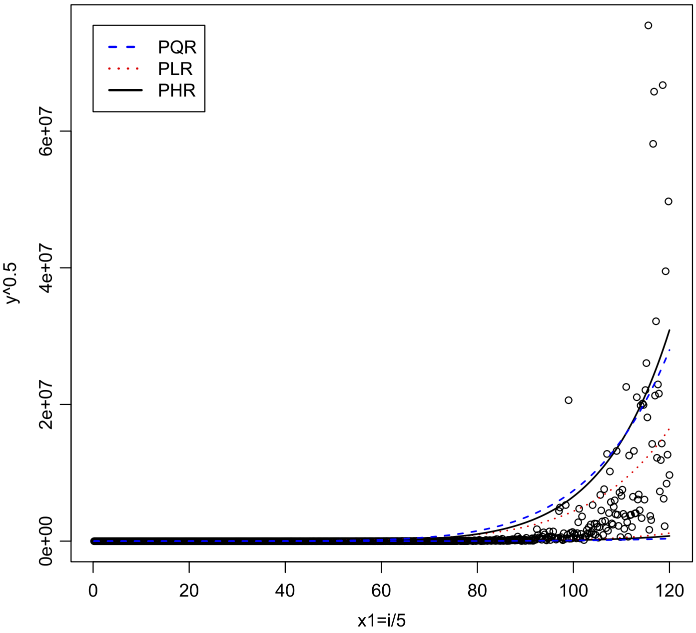

Fig. 4. Plot of curves from the three models at lower and upper quantiles using Study 4 with $n = 600$. Note that the estimated quantile for PHR (solid lines) and PQR (long dashed lines) were quite close particularly at the lower x values so that they appear to be a single line in this plot.

Table 2
Coverage of Monte Carlo: $C_{\mathcal{D}} = C_{\mathcal{D}}(\alpha_0, \alpha_1)$ with $(\alpha_0, \alpha_1) = (0.05, 0.95)$ with $n = 600$. In simulation, 400 replications have been carried out. The coverage probabilities are investigated at the whole domain, $S = \{S_1, S_2, S_3\}$ as well as $S_1 = \{x_i, i = 1, \dots, 200\}$, $S_2 = \{x_i, i = 201, \dots, 400\}$, and $S_3 = \{x_i, i = 401, \dots, 600\}$. The numbers in parentheses are $1000 \times$ MSE.

$C_{\mathcal{D}}$	Model	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6
C_S	PQR	0.901	0.901	0.900	0.900	0.901	0.901
		(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
	PLR	0.900	0.900	0.900	0.900	0.900	0.900
		(0)	(0)	(0)	(0)	(0)	(0)
	PHR	0.901	0.899	0.893	0.898	0.901	0.888
		(0.002)	(0.055)	(0.229)	(0.106)	(0.007)	(0.234)
C_{S_1}	PQR	0.924	0.926	0.917	0.927	0.933	0.904
		(0.708)	(0.691)	(0.434)	(0.726)	(1.189)	(0.111)
	PLR	0.899	1.000	1.000	1.000	0.946	0.997
		(0.302)	(9.982)	(10.000)	(10.000)	(2.283)	(9.510)
	PHR	0.900	0.922	0.877	0.912	0.897	0.859
		(0.420)	(2.051)	(2.527)	(2.109)	(0.560)	(2.643)
C_{S_2}	PQR	0.859	0.855	0.881	0.858	0.841	0.898
		(2.173)	(2.013)	(0.816)	(1.786)	(3.847)	(0.253)
	PLR	0.900	0.866	0.983	0.890	0.891	0.928
		(0.292)	(1.214)	(6.946)	(0.219)	(0.377)	(1.482)
	PHR	0.902	0.860	0.901	0.877	0.889	0.882
		(0.279)	(1.649)	(0.336)	(0.657)	(0.428)	(0.862)
C_{S_3}	PQR	0.921	0.920	0.903	0.916	0.929	0.900
		(0.611)	(0.435)	(0.193)	(0.281)	(0.994)	(0.106)
	PLR	0.901	0.834	0.717	0.810	0.864	0.774
		(0.275)	(4.482)	(33.486)	(8.302)	(1.613)	(16.669)
	PHR	0.902	0.915	0.902	0.905	0.915	0.922
		(0.401)	(0.645)	(0.157)	(0.362)	(0.534)	(0.766)

Note: $C_{\mathcal{D}}(\alpha_0, \alpha_1)$ denote the simulated (empirical) coverage corresponding to the $(\alpha_1 - \alpha_0)$% nominal prediction interval (PI) obtained within the domain of $\mathcal{D}$.
The results for PLR at $C_{\mathcal{S}}$ are mostly 0.9 throughout all 400 repetitions. Thus the zero MSE are reported. Also, PQR consistently estimates the overall coverage with the larger sample size compared to Table 1.
The performance of PHR in terms of estimating the coverage probabilities at the subdomains are better than the competing models for the most studies except Study 6.

Table 3 shows the coverage probability beyond $Q_{0.95}(y|x)$. If the coverage is higher (or lower) than 5%, it indicates that $Q_{0.95}(y|x)$ is underestimated (or overestimated). Again, the three models give the successful coverage probabilities close

Table 3
Coverage of Monte Carlo: $C_{\mathcal{D}}(\alpha_1, 1)$ with $\alpha_1 = 0.95$. In simulation, the sample sizes of $n = 120$ and $n = 600$ are used, and 400 replications have been carried out. The numbers in parentheses are $1000 \times$ MSE.

n	Model	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6
120	PQR	0.068	0.083	0.090	0.092	0.077	0.070
		(0.629)	(1.393)	(1.757)	(1.900)	(0.976)	(0.111)
	PLR	0.053	0.085	0.134	0.145	0.074	0.132
		(0.789)	(2.176)	(7.553)	(9.125)	(1.434)	(9.510)
	PHR	0.057	0.060	0.053	0.075	0.049	0.070
		(1.095)	(0.992)	(1.098)	(1.026)	(0.701)	(2.643)
600	PQR	0.060	0.076	0.068	0.077	0.062	0.058
		(0.181)	(0.688)	(0.436)	(0.732)	(0.241)	(0.118)
	PLR	0.051	0.150	0.142	0.150	0.690	0.112
		(0.168)	(10.000)	(8.476)	(10.000)	(0.527)	(4.198)
	PHR	0.052	0.076	0.051	0.074	0.045	0.041
		(0.215)	(1.061)	(0.305)	(0.777)	(0.174)	(0.336)

Note: $C_{\mathcal{D}}(\alpha_0, \alpha_1)$ denote the simulated (empirical) coverage corresponding to the $(\alpha_1 - \alpha_0)$% nominal prediction interval (PI) obtained within the domain of $\mathcal{D}$.

to 5% on the whole S. However, our following investigation reveals that at different subsets the coverage probabilities are not consistently 5% and tend to be under or overestimated. For example, at S_3, a particularly interesting subset, only PHR is able to consistently cover 5%. With heteroscedastic errors, PLR tends to underestimate $Q_{0.95}(y|x)$. Also, PQR requires a large sample size for good performance due to its model complexity; however, PHR and PLR are less influenced by the sample size.

4. Application to the health care expenditures

The Medical Expenditure Panel Survey (MEPS), sponsored by Agency for Healthcare Research and Quality (AHRQ), is a longitudinal survey conducted to generate nationally representative estimates of health care use, health care expenditures, sources of payment, health insurance coverage and health status for the U.S. civilian noninstitutionalized population. The MEPS is unique in its ability to link data on individuals and households (including demographics, health status, health conditions, health insurance, employment, and income) to detailed information on their use of and expenses for health care. Medical expenditure quantifies the total annual medical spending (including insurance spending and annual out-of-pocket spending), measured in dollars. Here we focus on the MEPS of the calendar year of 2007. The MEPS in Year 2007 provides information on 30,964 subjects among a nationally representative sample of the civilian non-institutionalized population. The expenditure data from the MEPS exhibit a marked positive skewness with a few high expenditure respondents and many low and zero expenditure respondents. As a consequence of the departure from a normal distribution, the MEPS data are often analyzed after transformation.

Zero medical expenditure is recorded for about 17% of people, and therefore it is not meaningful to estimate the lower quantiles. Note that Model (2) is not well defined if y is not strictly positive. Therefore, we make y strictly positive by adding one to y. Since the proposed sampling method does not produce equal probability samples, in the estimation procedure for parameters we account for the sampling weights.

Table 4 presents the characteristics of the MEPS data for the people in the top 5% spending and for the entire population. There are disparities in the percentages of each covariate between the two groups. The analysis shows that people with high health care expenses are likely to be older, insurance holders, and in a self-rated fair or poor physical health status. However, note that the percentages in Table 4 are marginal values. To identify the socioeconomic factors associated with medical spending and estimate their high conditional quantiles, we consider medical expenditure as a function of socioeconomic factors, including age, sex, marital status, ethnicity/race, self-rated health (SRH), education, poverty status, and health insurance type.

Having enough sample data for each covariate in the highest spending group is helpful to obtain more consistent results. Thus, all variables except age, which is continuous, are categorized into two levels: male vs. female, married vs. non-married (single, widowed, divorced/separated), black vs. non-black, good SRH (excellent/very good/good) vs. poor SRH (poor/fair), $<$ high school education vs. $\geq$ high school education, higher income (middle income/high income) vs. lower income (negative or poor/ near poor/low income), and insured (private/public) vs. uninsured. In A vs. B, the category on B is entered as a reference in the regression model. We limit our interest to adults as most selected predictors are inapplicable to children. Among a total of 21,782 adults, 18,197 adults are used to avoid missing values. Then we divide the sample into two parts, a training set ($n = 10{,}000$) and a testing set ($n = 8{,}197$) for corroborating the model performance.

First, in order to see whether the single transformation is appropriate for our MEPS data, we estimate $\lambda(\tau)$ using Model (4) on the set of $\tau \in \{0.50, 0.55, \ldots, 0.90, 0.95\}$. Fig. 5 suggests that it would not be necessary to have different transformation parameters at different quantiles, and the PHR model would work desirably. Therefore, we use a fixed $\hat{\lambda} = \hat{\lambda}(\tau = 0.5) = 0.3$ for all quantiles. Table 5 shows the parameter estimates from PHR.

Table 4
Characteristics of the top 5% spending and the entire sample.

	Top 5%	Entire sample
MEPS sample size(n)	901	18 197
US weighted sample size	9 848 610	199 077 260
Mean age (years)	60.1	47.1
Female (%)	56.1	52.4
Race/ethnicity (%)		
Hispanic	7.7	13.3
Black	11.6	11.1
Asian	3.3	4.3
Other	2.0	2.0
White	75.4	69.3
Marital status (%)		
Married	57.9	55.7
Single	11.1	23.8
Widowed	14.8	6.5
Divorced/separted	16.2	14.1
Education (%)		
Less than high school	19.0	15.3
High school graduate	47.3	49.2
Some college	33.7	35.6
Health insurance type (%)		
Private	68.5	70.0
Public	30.0	15.2
Uninsured	1.5	14.8
Income (%)		
Poor	12.6	10.4
Near poor	5.5	4.0
Low income	15.9	12.9
Middle income	26.7	30.8
High income	39.3	41.9
Self-rated health (SRH) (%)		
Excellent	7.6	25.5
Very good	18.7	33.9
Good	32.4	27.9
Fair	23.7	9.6
Poor	17.6	3.1

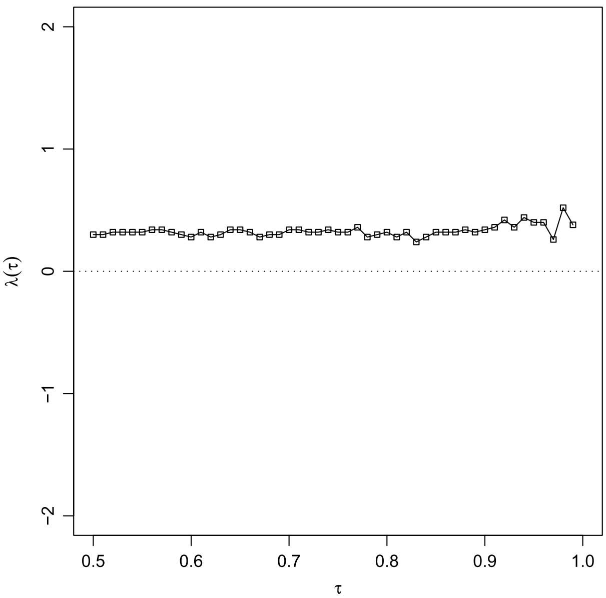

Fig. 5. Estimated transformation $\lambda(\tau)$ vs τ.

Analysis of the results shows that (i) the small p-values for age, health status, poverty, and uninsured predictors of γ coefficients in Table 5 indicate that heteroscedasticity is still present in the residuals. Therefore, it could potentially undermine the accuracy of estimating the quantiles if we ignored the heteroscedasticity. (ii) Older age, female, poor

Table 5
PHR parameter estimates of the location and scale parts using the 2007 MEPS data.

	Variables	Estimates	Pr(> |t|)
Location-part	(Intercept)	4.76	0.000
	Age	0.11	0.000
	Male	−1.53	0.000
	Married	0.01	0.881
	Black	−0.91	0.000
	Poor SRH	3.48	0.000
	< high school education	−0.81	0.000
	Lower income	−0.30	0.024
	Uninsured	−3.72	0.000
Scale-part	(Intercept)	1.85	0.000
	Age	0.01	0.000
	Male	−0.07	0.367
	Married	0.02	0.765
	Black	0.09	0.344
	Poor SRH	0.80	0.000
	< high school education	0.08	0.467
	Lower income	0.25	0.013
	Uninsured	−0.26	0.008

Abbreviation; SRH, self rated health.
p-values of the median regression parameter are reported.

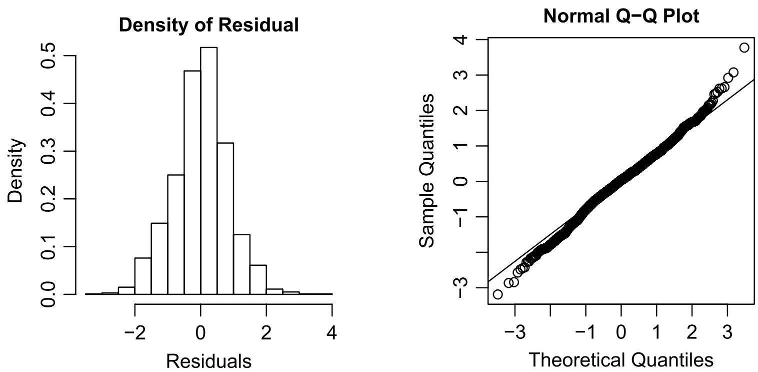

Fig. 6. The histogram of residuals after the Box–Cox transformation.

SRH, $\geq$ high school education, high income, and holding insurance are positively associated with medical expenditure. Importantly, the predictors of poor SRH and holding insurance greatly impact the high quantiles of medical expenditure. (iii) Black individuals and uninsured people from lower income families received significantly disparate care at high expenditure levels.

There are numerous benefits with better prediction of high quantiles of medical expenditure. It could alleviate the risk of catastrophic out-of-pocket medical care cost by focusing on patients with those high risk characteristics. Moreover, it can help us develop improved access to quality care among minorities with critical health issues. In this regard, we investigate the performance of the coverage probability at high quantiles, $\{y : y > \hat{Q}_{0.95}(y|x)\}$ with three models.

In practice, the traditional Box–Cox transformation toward the normality assumption is widely used in many fields when there is no prior information about an appropriate transformation. Although the density plot in Fig. 6 shows that the Box–Cox transformation considerably corrects a skewness in the MEPS data, longer right tails in both the density plot and the quantile–quantile (Q–Q) plot of residuals still question whether the normality assumption is appropriate.

Therefore, we introduce the new model based on the traditional Box–Cox transformation toward the normality assumption, and refer to it as the power transformed normal regression (PNR). Then, its quantile can be computed as

$$Q_\tau(y_i|x_i) = [1 + \lambda(x_i'\zeta + \sigma\Phi^{-1}(\tau))]^{1/\lambda},$$

where Φ is the cumulative distribution function for the Gaussian distribution with mean $x_i'\zeta$ and standard deviation σ. This PNR approach is compared with the previously introduced PQR and PLR as well as our proposed PHR model.

In Table 6 we estimate the coverage probability beyond the estimated 0.95th quantile of the expenditure for the four models. The deviations from the 0.05 level can be interpreted as under/over-estimated quantiles. Basically, the idea of Table 6 is that if the model is valid, it should give a consistent coverage probability at each level of the variables (e.g. male or female). In this application with the MEPS data set, PNR and PLR fail to give a consistent probability, especially over health status, income, and insured status, indicating that the data is not homoscedastic. As expected given the large sample size, we

Table 6
The coverage probabilities beyond the 0.95th quantile by the four models: PQR, PLR, PHR, and PNR.

Variables	Estimation				Validation			
	PQR	PLR	PHR	PNR	PQR	PLR	PHR	PNR
Sex								
Female	0.050	0.052	0.051	0.049	0.045	0.045	0.047	0.045
Male	0.050	0.049	0.055	0.035	0.053	0.051	0.057	0.041
Marital status								
Married	0.050	0.049	0.051	0.038	0.052	0.050	0.053	0.041
Widowed, divorced, never married	0.051	0.052	0.055	0.046	0.045	0.045	0.051	0.045
Race/ethnicity								
Black	0.050	0.063	0.058	0.056	0.057	0.076	0.063	0.069
Non-black	0.051	0.049	0.052	0.039	0.048	0.044	0.051	0.040
SRH								
Poor/fair health status	0.051	0.087	0.042	0.046	0.048	0.084	0.037	0.048
Good/excellent health status	0.050	0.045	0.055	0.041	0.049	0.043	0.054	0.042
Education level								
$<$ high school education	0.052	0.066	0.048	0.056	0.038	0.052	0.040	0.048
$\geq$ high school education	0.050	0.048	0.054	0.039	0.051	0.047	0.054	0.042
Income level								
Low income ($1 \leq$ poverty level ≤ 3)	0.050	0.070	0.056	0.058	0.050	0.059	0.055	0.060
High income ($4 \leq$ poverty level ≤ 5)	0.051	0.043	0.052	0.035	0.049	0.044	0.051	0.037
Insured status								
Uninsured	0.052	0.039	0.051	0.056	0.053	0.040	0.061	0.072
Private/ public insurance holder	0.050	0.053	0.053	0.039	0.048	0.049	0.051	0.038

observe that PQR is slightly better than PHR. However, when it comes to estimating the probability at different quantiles, say upper 0.99th quantile instead of 0.95th quantile, PQR needs to find different transformations and slope coefficients according to the transformed response. Since only the scale factors are adjusted for the different quantile for PHR, Table 6 is sufficient to explain the whole distribution of health care spending.

5. Conclusion

An observation that a relatively small group of individuals accounts for a large fraction of spending in the healthcare system motivates us to investigate the detailed causes for higher spending. The findings of our study reveal that aging tends to drive up costs and poor self-rated health is a good indicator for the higher costs. However, uninsured persons do not contribute to the high health cost. This result would suggest that careful monitoring of elderly's health status is important and a more aggressive preventive medicare system can reduce the high medical costs. Methodologically, estimating high quantiles in the medical expenditure data that exhibit a highly skewed and strongly heteroscedastic distribution is one of the key challenges in analyzing the Medical Expenditure Panel Survey (MEPS) data. In order to deal with these features of the MEPS data, we have proposed a quantile approach to the power transformed location–scale model. When the errors are not homoscedastic but the variability across the different range of quantiles is not evident, the power transformed location–scale model (PHR) can be a useful alternative to the power transformed quantile regression model (PQR).

The MEPS data suggests that the transformation parameters are quite stable over a set of quantiles, and thus the location–scale model using a single transformation performs well. The simulation studies demonstrate that our method can effectively estimate high quantiles when a data set presents either homoscedasticity or heteroscedasticity even with relatively small sample sizes, whereas the performance of PQR depends more on the sample size. In the aspect of data analysis, unlike PQR, PHR provides more intuitive and easier interpretations when the multiple quantiles are compared simultaneously since the different transformations on response variables are not required for different quantiles. Furthermore, an embarrassing phenomenon of quantile crossing may occur for PQR since each conditional quantile function is independently estimated. This event does not satisfy our theoretical and empirical assumption that the lower quantile level should not cross the higher quantile level. However, the proposed PHR model has the property of non-crossing quantiles. As a result, PHR can be more useful in practice if there is no clear benefit to employing the more complex-structured PQR model.

To note, our study has a few limitations. First, we assume the linear estimation of the scale parameter for the sake of simplicity. Further investigation is required when a true scale parameter has a nonlinear form. Also, one of the interesting features of the MEPS data is that medical expenditure is censored at zero. Therefore, taking into account the non-negative values with a point mass at zero can potentially lead to a more precise analysis. In this article, we have not considered the medical expenditure distribution in this way and are currently researching to accommodate these aspects of the model.

Acknowledgment

The author thanks Xuming He for helpful comments on and discussion around earlier versions of this draft.

Appendix. Technical proofs

Lemma A.1. *Consider Model (4). Under conditions* B1–B4*,*

$$(\hat{\beta}(\tau), \hat{\lambda}(\tau)) = \arg\min_{(\beta,\lambda)} \sum_{i=1}^{n} \rho_\tau(\Lambda_{\lambda(\tau)}(y_i) - x_i'\beta(\tau)), \quad \textit{for } \tau = 0.5$$

is consistent.

The proof of Lemma A.1. A rigorous derivation can be found in Powell (1991) and Oberhofer (1982). Buchinsky (1995) also develops the theory of the Box–Cox quantile regression for the case of discrete regressors where the estimation of $(\hat{\beta}, \hat{\lambda})$ can be achieved by minimum distance methods. □

The proof of Theorem 2.1. First we show that $\hat{\beta}_\tau$ is bounded. Let $\epsilon_0 = \min\{\tau, 1-\tau\}/2$ for $\tau \in (0, 1)$. Then, the following inequality holds.

$$\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\beta) - \sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i)) \geq \epsilon_0 \sum_{i=1}^{n} |x_i'\beta| - \sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i)).$$

Denote $\delta = \beta/\|\beta\|$. For a constant C in the assumption C3, if $\|\beta\| > \sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i))C/\epsilon_0$, then $\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\beta) > \sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i))$. However, by definition $\hat{\beta}_\tau$ is chosen to minimize the objective function, thus $\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\hat{\beta}_\tau) < \sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i))$ holds. This implies that $\|\hat{\beta}_\tau\| < M$, for some positive constant M.

Now suppose that $\hat{\beta}_\tau$ does not converge to β_0, i.e., $\|\hat{\beta}_\tau - \beta_0\| \geq \epsilon_0$, along a subsequence of n, still denoted by n for simplicity. Due to the boundedness of $\hat{\beta}_\tau$, there exists a further subsequences, still called $\hat{\beta}_\tau$, such that $\hat{\beta}_\tau \to \beta^* \neq \beta_0$. By the continuity of $E\rho_\tau(\cdot)$ and the uniqueness of β_0, there exists k_0 s.t.

$$E[\rho_\tau(\Lambda_\lambda(y) - x'\beta^*) - \rho_\tau(\Lambda_\lambda(y) - x'\beta_0)] > k_0 > 0. \tag{A.1}$$

By the law of large numbers, $n^{-1}\sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(y_i) - x_i'\hat{\beta}_\tau) \to E[\rho_\tau(\Lambda_\lambda(y) - x'\beta^*)]$. For sufficiently large n, (A.1) leads to the following inequality,

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(y_i) - x_i'\beta_0) < \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(y_i) - x_i'\hat{\beta}_\tau) - \frac{k_0}{2}. \tag{A.2}$$

Since $\hat{\lambda}$ is consistent by Lemma A.1, we have, for sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\beta_0) < \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(y_i) - x_i'\beta_0) + \frac{k_0}{4} \tag{A.3}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_\lambda(y_i) - x_i'\hat{\beta}_\tau) - \frac{k_0}{4} < \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\hat{\beta}_\tau). \tag{A.4}$$

Putting (A.2)–(A.4) all together,

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\beta_0) < \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(\Lambda_{\hat{\lambda}}(y_i) - x_i'\hat{\beta}_\tau).$$

This contradicts the definition of $\hat{\beta}_\tau$ that minimizes the objective function, meaning $\hat{\beta}_\tau$ converges to β_0. Therefore, $\hat{\beta}_\tau$ is consistent. □

The proof of Theorem 2.2. The proof of consistency of $\hat{\lambda}$ follows immediately from Lemma A.1. Also, given $\hat{\lambda} \to \lambda$ as n goes to infinity, particularly for $\tau = 0.5$, Theorem 2.1 implies $\hat{\beta}$ in Model (2) is the consistent estimator of β. Now we remain to prove the consistency of $\hat{\gamma}$. Let $\hat{u} = \Lambda_{\hat{\lambda}}(y) - x_i'\hat{\beta}$ and $u = \Lambda_\lambda(y) - x'\beta$. By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(u_i) \to E\rho_\tau(u). \tag{A.5}$$

For sufficiently large n, since both $\hat{\lambda}$ and $\hat{\beta}$ are consistent, by continuity

$$\frac{1}{n}\sum_{i=1}^{n} (\rho_\tau(\hat{u}_i) - \rho_\tau(u_i)) \to 0. \tag{A.6}$$

From (A.5)–(A.6), we obtain $n^{-1}\sum_{i=1}^{n}\rho_\tau(\hat{u}_i) - E\rho_\tau(u) \rightarrow 0$. By making use of similar arguments that we used for Theorem 2.1, $\hat{\gamma}$ which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{u}_i - x_i'\gamma|$$

is consistent. Therefore, the τth quantile is consistent for the true conditional quantile function $\Lambda_\lambda^{-1}(x_i'\beta + x_i'\gamma G_{u_i}^{-1}(\tau))$. □

References

Bassett Jr., G.W., Koenker, R., 1982. An empirical quantile function for linear models with iid errors. Journal of the American Statistical Association 77, 407–415.
Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society, Series B 26, 211–252.
Buchinsky, M., 1995. Estimating the asymptotic covariance matrix for quantile regression models a monte carlo study. Journal of Econometrics 68, 303–338.
Chamberlain, G., 1994. Quantile regression censoring and the structure of wages. In: Sims, C. (Ed.), Advances in Econometrics. Elsevier, New York, NY.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications, first ed. In: CRC Monographs on Statistics & Applied Probability, Chapman & Hall.
He, X., 1997. Quantile curves without crossing. The American Statistician 51, 186–192.
Hong, G.H., He, X., 2010. Prediction of functional status for the elderly based on a new ordinal regression model. Journal of the American Statistical Association 105, 930–941.
Koenker, R., Bassett Jr., G.W., 1978. Regression quantiles. Econometrica 46, 33–50.
Machado, J., Mata, J., 2000. Box–Cox quantile regression and the distribution of firm sizes. Journal of Applied Econometrics 15, 253–274.
Monheit, A.C., 2003. Persistence in health expenditures in the short run: prevalence and consequences. Medical Care 41, 53–64.
Mu, Y., He, X., 2007. Power transformation toward a linear regression quantile. Journal of the American Statistical Association 102, 269–279.
Oberhofer, W., 1982. The consistency of nonlinear regression minimizing the l_1 norm. Annals of Statistics 10, 316–319.
Powell, J.L., 1991. Estimation of monotonic regression models under quantile restrictions. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium on Economic Theory and Econometrics. Cambridge University Press, New York, NY.
Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. Annals of Statistics 22, 1346–1370.