Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

A Corrected Plug-In Method for Quantile Interval Construction through a Transformed Regression

Z. L. Yang and Y. K. Tse

School of Economics and Social Sciences, Singapore Management University
90 Stamford Road, Singapore 178903. `zlyang@smu.edu.sg`; `yktse@smu.ed.sg`

October 2006

We propose a corrected plug-in method for constructing confidence intervals of the conditional quantiles of an original response variable through a transformed regression with heteroscedastic errors. The interval is easy to compute. Factors affecting the magnitude of the correction are examined analytically through the special case of Box-Cox regression. Monte Carlo simulations show that the new method works well in general, and is superior over the commonly used delta method and the quantile regression method. An empirical application is presented.

KEY WORDS: Analytical correction; Transformation; Heteroscedasticity; Finite sample performance; Living standards in South Africa.

1. INTRODUCTION

In this paper we develop a general method to calculate confidence limits for conditional quantiles of an original response variable in a transformed regression model with heteroscedasticity. The method is built upon normal inference theories for linear models, with adjustments to account for the estimation of the transformation and weighting parameters, thus is expected to perform well in small samples.

We consider general transformations beyond the well-known Box-Cox transformation. This extension is important in two aspects. First, the Box-Cox power transformation works only for nonnegative observations. As economic data are not always positive, a more general transformation family that allows for both positive and negative observations is desirable. Second, even if the data observations are nonnegative, the Box-Cox transformation is unable to transform the data to exact normality unless the transformation parameter is zero.

Our model allows for heteroscedasticity with the variance specified as a general function of a set of weighting variables. When both transformation and weighting parameters are known, the standard inference methods lead directly to a confidence interval for the transformed quantile through a non-central t statistic, which upon an inverse transformation gives a confidence interval for the conditional quantile itself. When the transformation and weighting parameters are unknown, a *plug-in* type of confidence interval (i.e., with plug-in for both the transformation and the weighting parameters) for

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

this conditional quantile can be obtained by evaluating the lower and upper confidence limits at the estimated transformation and weighting parameters.

We show that the above plug-in method leads to liberal confidence intervals for the conditional quantiles even in large samples. We develop a simple and general method of correcting the underlying statistic to give a *corrected plug-in* confidence interval for the conditional quantile. This correction method works for any monotonic transformation and a fairly general weighting function. We give special attention to the Box-Cox transformation model with homoscedastic errors to provide analytical details that help identify factors affecting the magnitude of the correction.

There are general parametric or non-parametric methods available in the literature for the type of problems we consider in this paper. Two closely related ones are the *delta method* and the likelihood ratio (LR) test method. As the LR test method is too computationally involved, we choose the delta method as the benchmark for our comparison within the same model. Another method, based on a different model, the *quantile regression*, has become popular in economics research. Thus it would be interesting to give a comparison between our method and the quantile regression method.

Our extensive Monte Carlo results reveal the following interesting regularities: (i) our method performs very well in general, (ii) it is quite robust against mild departures from normality, (iii) it clearly outperforms the delta method in the cases of small samples, upper extreme quantiles, high nonlinearity, or out-of-sample forecast, and (iv) it outperforms the method based on the Box-Cox quantile regression. Of particular importance is that when the main interest is in extrapolation (out-of-sample forecasting), our method provides satisfactory results, whereas the delta method may lead to unreliable inferences with deteriorating performance as the extrapolation point moves away from the design region even for large samples. The simulation results also show that the correction effect can be enormous, indicating the routine use of the plug-in method may be inappropriate – a fact that has not received much attention in the literature of the Box-Cox transformation.

Unlike the interval estimation for a fixed parameter or a parametric function as we consider in this paper, Collins (1991) examined the use of the plug-in technique for the prediction intervals of a future response Y_0 in a homogeneous Box-Cox regression and showed using simulation that the uncertainty about the transformation parameter is relatively unimportant. This is in line with the remark by Carroll and Ruppert (1991, p.297) that the uncertainty in predicting a future response is mostly due to its variability about its mean, not to the uncertainty about the parameters. In other words, a crucial factor for the predictability of Y_0 is the variability of Y_0, not the variability of the parameter estimates. Thus, estimating the transformation and weighting parameters

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

does not affect the asymptotic performance of the prediction interval. In this paper, however, we are considering the confidence interval of the conditional quantile, and our results show that the variability of the parameter estimates play a very important role.

The plan of this paper is as follows. Section 2 provides the corrected plug-in confidence interval for a regression quantile with a general monotonic transformation and homoscedastic errors. An analytical expression for the correction factor is given for the case of the Box-Cox transformation model. In Section 3 we extend the results to a transformation model with heteroscedastic errors. Section 4 investigates the small-sample performance of the quantile limits using Monte Carlo method, and compares it with alternative methods based on the same model (namely, the delta and plug-in methods) or on a different model (namely, the quantile regression method). Section 5 provides an empirical application of our method and Section 6 concludes the paper.

2. TRANSFORMATION MODEL WITH HOMOSCEDASTIC ERRORS

Let $\mathbf{Y}$ be an $n \times 1$ vector of response observations, $h(\mathbf{Y}, \lambda)$ a vector of transformed observations, and $\mathbf{X}$ an $n \times k$ matrix the columns of which contain the values of the explanatory variables $X_1, X_2, \ldots, X_k$. The usual Box-Cox regression model (Box and Cox, 1964) is

$$h(\mathbf{Y}, \lambda) = \mathbf{X}\beta + \sigma\mathbf{e}, \tag{1}$$

where β is a $k \times 1$ vector of parameters, σ is the error standard deviation, $\mathbf{e}$ is an $n \times 1$ vector of independent and identically distributed normal variates with zero mean and unit variance, and $h(\cdot, \lambda)$ is a general monotonic increasing function. Given the model assumptions, the maximum likelihood estimators (MLE) of β, σ^2 and λ are given by

$$\hat{\lambda} = \arg\min_{\ell} \dot{J}^{-1}(\ell)\|\mathbf{M}h(\mathbf{Y}, \ell)\|, \tag{2}$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'h(\mathbf{Y}, \hat{\lambda}), \tag{3}$$

$$\hat{\sigma}^2 = \frac{1}{n}\|\mathbf{M}h(\mathbf{Y}, \hat{\lambda})\|^2, \tag{4}$$

where $\|\cdot\|$ is the Euclidian norm, $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\dot{J}(\ell)$ is the geometric mean of $\{h_y(Y_i, \ell) = \partial h(Y_i, \ell)/\partial Y_i, i = 1, \cdots, n\}$. Note that $\hat{\beta}$ and $\hat{\sigma}^2$ are analytic functions of $\hat{\lambda}$, and we shall denote these quantities as $\hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}^2(\hat{\lambda})$, respectively. When λ is known, the MLE of β and σ are $\hat{\beta}(\lambda)$ and $\hat{\sigma}(\lambda)$, respectively. Equivalently, $\hat{\lambda}$ solves

$$-n\frac{h'(\mathbf{Y}, \lambda)\mathbf{M}h_\lambda(\mathbf{Y}, \lambda)}{h'(\mathbf{Y}, \lambda)\mathbf{M}h(\mathbf{Y}, \lambda)} + \sum_{i=1}^{n}\frac{h_{y\lambda}(Y_i, \lambda)}{h_y(Y_i, \lambda)} = 0, \tag{5}$$

where $h_\lambda(\mathbf{Y}, \lambda) = \{h_\lambda(Y_i, \lambda)\}_{n\times 1}$ with $h_\lambda(Y_i, \lambda) = \frac{\partial}{\partial\lambda}h(Y_i, \lambda)$, and $h_{y\lambda}(Y_i, \lambda) = \frac{\partial^2}{\partial Y_i \partial\lambda}h(Y_i, \lambda)$.

2.1 The Plug-in Quantile Limits

Consider the problem of estimating the τ-quantile of Y_0 conditional upon a given observation x_0. We denote the conditional quantile as y_τ (the conditional information x_0 is suppressed). As the transformation is monotonic, we have $h(y_\tau, \lambda) = x_0'\beta + \sigma z_\tau$, where z_τ is the τ-quantile of the standard normal variate. A natural estimate of $h(y_\tau, \lambda)$ is $x_0'\hat{\beta}(\lambda) + \hat{\sigma}^*(\lambda) z_\tau$, where $\hat{\sigma}^{*2}(\lambda) = \hat{\sigma}^2(\lambda)[n/(n-k)]$ is an unbiased estimator of σ^2. As $\mathrm{Var}[x_0'\hat{\beta}(\lambda)] = \sigma^2 \kappa_{n0}^{-2}$, where $\kappa_{n0}^{-2} = x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0$, it is natural to consider the following pivotal quantity for inference about y_τ:

$$T_\tau(\lambda) = \frac{x_0'\hat{\beta}(\lambda) + \hat{\sigma}^*(\lambda) z_\tau - h(y_\tau, \lambda)}{\kappa_{n0}^{-1}\hat{\sigma}^*(\lambda)}. \tag{6}$$

Note that $h(y_\tau, \lambda) = x_0'\beta + \sigma z_\tau$, and $T_\tau(\lambda)$ can be rewritten as

$$\frac{\kappa_{n0}[x_0'\hat{\beta}(\lambda) - x_0'\beta]/\sigma - \kappa_{n0} z_\tau}{\hat{\sigma}^*(\lambda)/\sigma} + \kappa_{n0} z_\tau, \tag{7}$$

from which we can see that $T_\tau(\lambda) \sim t_{n-k}(-\kappa_{n0} z_\tau) + \kappa_{n0} z_\tau$, where $t_\nu(\delta)$ denotes a noncentral t random variable with ν degrees of freedom and noncentrality parameter δ. Thus, we obtain the mean μ_T and variance σ_T^2 of $T_\tau(\lambda)$ as

$$\mu_T = \kappa_{n0} z_\tau \left[1 - \left(\frac{n-k}{2} \right)^{1/2} \frac{\Gamma((n-k-1)/2)}{\Gamma((n-k)/2)} \right], \tag{8}$$

$$\sigma_T^2 = \frac{n-k}{n-k-2}(1 + \kappa_{n0}^2 z_\tau^2) - \frac{n-k}{2}\kappa_{n0}^2 z_\tau^2 \left[\frac{\Gamma((n-k-1)/2)}{\Gamma((n-k)/2)} \right]^2. \tag{9}$$

Note that when $\tau = 0.5$, $z_\tau = 0$, $\mu_T = 0$, $\sigma_T = (n-k)/(n-k-2)$, and $T_\tau(\lambda)$ follows a central t distribution.

Denote the α-quantile of $t_\nu(\delta)$ by $t_\nu^\alpha(\delta)$. The following inequality: $t_{n-k}^{\alpha/2}(-\kappa_{n0} z_\tau) + \kappa_{n0} z_\tau \le T_\tau(\lambda) \le t_{n-k}^{1-\alpha/2}(-\kappa_{n0} z_\tau) + \kappa_{n0} z_\tau$, holds with probability $1-\alpha$, from which we obtain an exact $100(1-\alpha)\%$ confidence interval (CI) for $h(y_\tau, \lambda)$ as

$$\left\{ x_0'\hat{\beta}(\lambda) - t_{n-k}^{1-\alpha/2}(-\kappa_{n0} z_\tau)\frac{\hat{\sigma}^*(\lambda)}{\kappa_{n0}}, \quad x_0'\hat{\beta}(\lambda) - t_{n-k}^{\alpha/2}(-\kappa_{n0} z_\tau)\frac{\hat{\sigma}^*(\lambda)}{\kappa_{n0}} \right\}. \tag{10}$$

Applying inverse transformation to the lower limit $L(\lambda)$ and the upper limit $U(\lambda)$ in (10) gives the following *exact* $100(1-\alpha)\%$ CI for y_τ:

$$\left\{ h^{-1}[L(\lambda), \lambda], \ h^{-1}[U(\lambda), \lambda] \right\}. \tag{11}$$

We shall call the confidence interval of y_τ the quantile limits (QL). In practical situations, λ is often unknown and has to be estimated. In this case, a popular approach is to 'plug'

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

the MLE $\hat{\lambda}$ into (11) for the unknown λ (Collins, 1991; Hahn and Meeker, 1991), so that the CI for y_τ becomes

$$\left\{h^{-1}[L(\hat{\lambda}),\hat{\lambda}],\ h^{-1}[U(\hat{\lambda}),\hat{\lambda}]\right\}. \tag{12}$$

The interval (12), referred to as the *plug-in quantile limits* (PQL), is sometimes used without accounting for the extra variability introduced by $\hat{\lambda}$. In the next sub-section we shall show that such an extra variability cannot be ignored, and shall discuss the procedure to correct for the PQL to account for the estimation of λ.

2.2 Asymptotics of the Plug-in Method

Let $\psi = (\beta', \sigma, \lambda)'$, $v^2(\psi)$ be the asymptotic variance of $\sqrt{n}(\hat{\lambda}-\lambda)$, and $T_\tau(\hat{\lambda})$ be the statistic obtained by replacing λ in $T_\tau(\lambda)$ by its MLE $\hat{\lambda}$. Thus, the validity of PQL depends on whether $T_\tau(\hat{\lambda})$ agrees with $T_\tau(\lambda)$. In what follows, we assume that a quantity bounded in probability has a finite expectation.

Theorem 1. *Assume the following are true: i)* $\hat{\lambda} \xrightarrow{p} \lambda$ *and* $\sqrt{n}(\hat{\lambda}-\lambda)/v(\psi) \xrightarrow{D} N(0,1)$*; ii)* $\mathbf{X}'h_\lambda(\mathbf{Y},\lambda)/n$*,* $\mathbf{X}'h_{\lambda\lambda}(\mathbf{Y},\lambda)/n$*, and* $h'(\mathbf{Y},\lambda)\mathbf{M}h_\lambda(\mathbf{Y},\lambda)/n$ *converge in probability; and iii)* $\mathbf{X}'\mathbf{X}/n$ *converges to a positive definite matrix. Then we have,*

$$T_\tau(\hat{\lambda}) = T_\tau(\lambda) + \frac{\sqrt{n}(\hat{\lambda}-\lambda)}{v(\psi)}c(\psi) + o_p(1). \tag{13}$$

Furthermore, $T_\tau(\lambda)$ *and* $\hat{\lambda}$ *are asymptotically independent, so that*

$$\mathrm{E}[T_\tau(\hat{\lambda})] = \mathrm{E}[T_\tau(\lambda)] + o(1), \tag{14}$$

$$\mathrm{Var}[T_\tau(\hat{\lambda})] = \mathrm{Var}[T_\tau(\lambda)] + c^2(\psi) + o(1), \tag{15}$$

where

$$c(\psi) = \lim_{n\to\infty} \frac{x_0'\mathrm{E}[\hat{\beta}_\lambda(\lambda)] + \frac{1}{2\sigma}z_\tau \mathrm{E}[\hat{\sigma}^*_\lambda(\lambda)] - h_\lambda(y_\tau,\lambda)}{\sqrt{n}\sigma/[\kappa_{n0}\tau(\psi)]}. \tag{16}$$

The proof of Theorem 1 is given in Appendix A. The constant $c(\psi)$ quantifies the effect of estimating λ on the pivotal quantity $T_\tau(\lambda)$ and on the PQL, hence we refer to it as the *correction factor*. As $c(\psi)$ is of order $O(1)$, $T_\tau(\hat{\lambda})$ and $T_\tau(\lambda)$ differ by a term of order $O_p(1)$. In particular, they have the same limiting mean, but the limiting variance of $T_\tau(\hat{\lambda})$ is larger than that of $T_\tau(\lambda)$. This indicates that the QL without accounting for the estimation of λ is too tight. The value of $c(\psi)$ generally depends on ψ, as well as z_τ and x_0. Its detailed structure will be examined in Section 2.4.

2.3 The Corrected Plug-in Quantile Limits

We now derive the corrected plug-in quantile limits. From the results of Theorem 1, it is clear that the following adjusted pivotal quantity

$$T_\tau^*(\hat{\lambda}) = \frac{T_\tau(\hat{\lambda}) - C_m(\psi)}{C_s(\psi)}, \tag{17}$$

with $C_m(\psi) = \mu_T(1 - C_s(\psi))$ and $C_s(\psi) = (1 + c^2(\psi)/\sigma_T^2)^{\frac{1}{2}}$, has the same asymptotic mean and variance as $T_\tau(\lambda)$. This leads immediately to the following *corrected plug-in quantile limits* (CPQL):

$$\left\{h^{-1}[L^*(\hat{\lambda}), \hat{\lambda}], \;\; h^{-1}[U^*(\hat{\lambda}), \hat{\lambda}]\right\}, \tag{18}$$

with the two adjusted end points before transformation given by

$$L^*(\hat{\lambda}) \;=\; x_0'\hat{\beta}(\hat{\lambda}) + \widehat{C_m^*(\psi)} - t_{n-k}^{1-\alpha/2}(-\kappa_{n0}z_\tau)\widehat{C_s(\psi)}\frac{\hat{\sigma}^*(\hat{\lambda})}{\kappa_{n0}}, \tag{19}$$

$$U^*(\hat{\lambda}) \;=\; x_0'\hat{\beta}(\hat{\lambda}) + \widehat{C_m^*(\psi)} - t_{n-k}^{\alpha/2}(-\kappa_{n0}z_\tau)\widehat{C_s(\psi)}\frac{\hat{\sigma}^*(\hat{\lambda})}{\kappa_{n0}}, \tag{20}$$

and $\widehat{C_m^*(\psi)} = \hat{\sigma}^*(\hat{\lambda})[1 - \widehat{C_s(\psi)}][z_\tau + \mu_T/\kappa_{n0}]$. What remains now is a method to estimate $c(\psi)$ so that the estimates $\widehat{C_s(\psi)}$ and $\widehat{C_m^*(\psi)}$ can be obtained. From the way $c(\psi)$ is defined, it is natural to introduce the following estimate:

$$\widehat{c(\psi)} = \frac{x_0'\hat{\beta}_\lambda(\hat{\lambda}) + z_\tau\hat{\sigma}_\lambda^*(\hat{\lambda}) - h_\lambda(\hat{y}_\tau, \hat{\lambda})}{\hat{\sigma}^*(\hat{\lambda})/[\kappa_{n0}J^{\lambda\lambda}(\hat{\psi})^{1/2}]}, \tag{21}$$

where $J^{\lambda\lambda}(\psi)$ is the last diagonal element of $J^{-1}(\psi)$. Note that $\widehat{c(\psi)}$ is a consistent estimate of $c(\psi)$, thus, the CPQL given in (18) has correct coverage asymptotically.

The results above apply to any monotonic transformation. This is important as in practice one may need to use different transformations for different types of data. For example, when data are nonnegative it is popular to use the Box-Cox power transformation:

$$h(y, \lambda) = \begin{cases} (y^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0, \\ \log y, & \text{if } \lambda = 0. \end{cases} \tag{22}$$

This transformation, however, has a well-known truncation problem. To circumvent this problem, the dual power transformation (Yang, 2006) may be used, which takes the form

$$h(y, \lambda) = \begin{cases} (y^\lambda - y^{-\lambda})/2\lambda, & \lambda > 0, \\ \log y, & \lambda = 0. \end{cases} \tag{23}$$

For data that can be both positive and negative, suitable transformations are available in Manly (1976), John and Draper(1980), Bickel and Doksum (1981), Burbidge *et al.* (1988) and Yeo and Johnson (2000). For proportion and percentage data, suitable transformations are available in Atkinson (1985).

2.4 The Box-Cox Power Transformation

As discussed earlier, it is important to examine the detailed analytical structure of $c(\psi)$ so that factors affecting the magnitude of the correction can be identified. In this section, we consider the Box-Cox transformation for this purpose. We denote the Hadamard operation (elementwise multiplication) of two vectors by $\odot$. As a convention, functions applied to a vector, b say, are carried out elementwise, e.g., $\log b = \{\log b_i\}_{n\times 1}$. For the Box-Cox power transformation, we have the following result.

Theorem 2. *Assume the conditions of Theorem 1 hold. Let $h(\cdot,\lambda)$ be the Box-Cox power transformation, and $\eta = \{\eta_i\} = \mathbf{X}'\beta$. We have, for large n, when $\lambda \neq 0$,*

$$c(\psi) = \frac{x_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[(1_n+\lambda\eta)\odot\phi+\frac{1}{2}\lambda\sigma\theta]-(1+\lambda\eta_0)\phi_0+\lambda^2 a(z_\tau)}{\lambda\sigma k_{n0}^{-1}\left(\| \mathbf{M}(\theta^{-1}\odot\phi+\frac{1}{2}\theta)\|^2+2\|\phi-\bar{\phi}\|^2+\frac{3}{2}\|\theta\|^2\right)^{1/2}} + O_p(n^{-1/2})+O_p(\theta_M^3), \tag{24}$$

and when $\lambda = 0$,

$$c(\psi) = \frac{x_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\eta^2+\sigma^2 1_n)-\eta_0^2+2a(z_\tau)}{k_{n0}^{-1}\left(\|\mathbf{M}\eta^2\|^2+8\sigma^2\|\eta-\bar{\eta}\|^2+6n\sigma^4\right)^{1/2}}+O_p(n^{-1/2}), \tag{25}$$

where $\phi = \log(1+\lambda\eta)$, $\theta = \{\theta_i\} = \lambda\sigma(1_n+\lambda\eta)^{-1}$, $\theta_M = \max\{|\theta_i|\}$, $\bar{\eta} = 1_n'\eta/n$, $\bar{\phi} = 1_n'\phi/n$, $1_n = \{1\}_{n\times 1}$, $\eta_0 = x_0'\beta$, $\theta_0 = \lambda\sigma/(1+\lambda\eta_0)$, $\phi_0 = \log(1+\lambda\eta_0)$, and

$$a(z_\tau) = \begin{cases} \sigma z_\tau\left(\frac{1}{n-k}\sum_{i=1}^n m_{ii}(\phi_i-\frac{3}{2}\theta_i^2)-\phi_0-\frac{1}{2}\theta_0 z_\tau+\frac{1}{2}\theta_0^2 z_\tau^2\right)\lambda^{-1}, & \text{if } \lambda\neq 0,\\ \sigma z_\tau\left(\frac{1}{n-k}\sum_{i=1}^n m_{ii}\eta_i-\eta_0-\frac{1}{2}\sigma z_\tau\right), & \text{if } \lambda = 0,\end{cases} \tag{26}$$

with m_{ii} being the ith *diagonal element of* $\mathbf{M}$.

The proof of Theorem 2 is in Appendix A. Note that when $\tau = 0.5$, $a(z_\tau) = 0$ and Theorem 2 above reduces to Theorem 2 in Yang (2002). Equation (24) implies that the truncation effect of the Box-Cox transformation is small when θ_M is small, which is true when either σ is small, or λ is small, or η are large. From the theorem, we see that $c(\psi)$ depends on

(a) the distance between x_0 and the center of the design proportionally,
(b) the value of z_τ through the term $a(z_\tau)$,
(c) the model structure through the $\mathbf{M}$ term,
(d) the mean spread through the term $\|\phi-\bar{\phi}\|^2$ or $\|\eta-\bar{\eta}\|^2$, and
(e) the residual standard deviation σ.

Among these factors, the distance from x_0 to the center of the design measured by $\kappa_{n0} = \{x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0\}^{-\frac{1}{2}}$ is most important. The further x_0 is away from the center of the design, the larger is the correction. In this case, routine use of the QL without accounting for the transformation estimation may be highly misleading. This is in contrast to Carroll and Ruppert (1981), who claimed that the cost of estimating the transformation in making inference about the median is generally not severe.

Of secondary importance are the model structure and the mean spread. More structured models (i.e., models with more continuous type of covariates), and models with a large variation in their means may be associated with a smaller effect in estimating the transformation. Finally, when θ_M is small the result of Theorem 2 provides a simple estimate of $c^2(\psi)$ by replacing ψ by its MLE $\hat{\psi}$.

3. TRANSFORMATION MODEL WITH HETEROSCEDASTIC ERRORS

Although the purposes of applying data transformation technique are to induce normality, linear model structure and constancy of error variance (Box and Cox, 1964), these three goals may not be achieved simultaneously. We now consider an extension of model (1) in which the errors are heteroscedastic, i.e,

$$h(Y_i, \lambda) = x_i'\beta + \sigma\omega(v_i, \gamma)e_i, \qquad i = 1, \cdots, n, \tag{27}$$

where $\omega(v_i, \gamma) \equiv \omega_i(\gamma)$ is the weighting function, v_i is a vector of observations on a set of variables, called the weighting variables, and γ is a vector of weighting parameters. A special weighting function is $\omega^2(v_i, \gamma) = \exp(v_i'\gamma)$ (Harvey, 1976). The weighting variables may include (some of) the regressors and/or other variables. Define $\Omega(\gamma) =$ diag $\{\omega_1^2(\gamma), \cdots, \omega_n^2(\gamma)\}$. If the weights are given, equation (27) can be converted into the form of equation (1) by premultiplying ω_i^{-1} on each side of the equation. Let $\Omega^{-\frac{1}{2}}(\gamma)$ $=$ diag $\{\omega_1^{-1}(\gamma), \cdots, \omega_n^{-1}(\gamma)\}$. Thus, the results given in Section 2 apply after replacing h by $\Omega^{-\frac{1}{2}}(\gamma)h$, $\mathbf{X}$ by $\Omega^{-\frac{1}{2}}(\gamma)\mathbf{X}$, x_0 by $\omega_0^{-1}(\gamma)x_0$, and $h(y_\tau, \lambda)$ by $\omega_0^{-1}(\gamma)h(y_\tau, \lambda)$.

When the weights are unknown, the maximum likelihood method can be used to estimate jointly the parameters. The log-likelihood function is, ignoring the constant,

$$\ell(\beta, \sigma^2, \gamma, \lambda) = -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^n \log\omega_i(\gamma) - \frac{1}{2\sigma^2}\sum_{i=1}^n \left[\frac{h(y_i, \lambda) - x_i'\beta}{\omega_i(\gamma)}\right]^2 + \sum_{i=1}^n \log h_y(y_i, \lambda). \tag{28}$$

For given γ and λ, the MLEs of β and σ are

$$\hat{\beta}(\gamma, \lambda) = [\mathbf{X}'\Omega^{-1}(\gamma)\mathbf{X}]^{-1}\mathbf{X}'\Omega^{-1}(\gamma)h(\mathbf{Y}, \lambda), \tag{29}$$

$$\hat{\sigma}^2(\gamma, \lambda) = \frac{1}{n}[h(\mathbf{Y}, \lambda) - \mathbf{X}\hat{\beta}(\gamma, \lambda)]'\Omega^{-1}(\gamma)[h(\mathbf{Y}, \lambda) - \mathbf{X}\hat{\beta}(\gamma, \lambda)]. \tag{30}$$

Substituting $\hat{\beta}(\gamma,\lambda)$ and $\hat{\sigma}(\gamma,\lambda)$ into the log-likelihood function gives the following concentrated likelihood function of γ and λ, ignoring the constant,

$$\ell_c(\gamma,\lambda) = \ell[\hat{\beta}(\gamma,\lambda),\hat{\sigma}^2(\gamma,\lambda),\gamma,\lambda] = -\frac{n}{2}\log\sum_{i=1}^{n}\left[\dot{\omega}(\gamma)s_i(\gamma,\lambda)/\dot{J}(\lambda)\right]^2, \tag{31}$$

where $s_i(\gamma,\lambda) = [h(y_i,\lambda) - x_i'\hat{\beta}(\gamma,\lambda)]/\omega_i(\gamma)$, and $\dot{\omega}(\gamma)$ and $\dot{J}(\lambda)$ are the geometric means of $\omega_i(\gamma)$ and $h_y(Y_i,\lambda)$, respectively. The full MLEs can thus be written as

$$(\hat{\gamma}',\hat{\lambda})' = \arg\min_{(\gamma,\lambda)}\left[\dot{\omega}(\gamma)/\dot{J}(\lambda)\right]^2\sum_{i=1}^{n}s_i^2(\gamma,\lambda), \tag{32}$$

$$\hat{\beta}(\hat{\gamma},\hat{\lambda}) = [\mathbf{X}'\Omega^{-1}(\hat{\gamma})\mathbf{X}]^{-1}\mathbf{X}'\Omega^{-1}(\hat{\gamma})h(\mathbf{Y},\hat{\lambda}), \tag{33}$$

$$\hat{\sigma}^2(\hat{\gamma},\hat{\lambda}) = \frac{1}{n}h'(\mathbf{Y},\hat{\lambda})\mathbf{M}'(\hat{\gamma})\Omega^{-1}(\hat{\gamma})\mathbf{M}(\hat{\gamma})h(\mathbf{Y},\lambda), \tag{34}$$

where $\mathbf{M}(\gamma) = I_n - \mathbf{X}[\mathbf{X}'\Omega^{-1}(\gamma)\mathbf{X}]^{-1}\mathbf{X}'\Omega^{-1}(\gamma)$.

Similar to the case of homoscedastic errors, to estimate the τ-quantile y_τ of the response Y_0 at values x_0 and v_0 of the regressors and weighting variables, respectively, we start with the pivotal quantity for the case of known λ and γ, i.e.,

$$T_\tau(\gamma,\lambda) = \frac{x_0'\hat{\beta}(\gamma,\lambda) + \hat{\sigma}^*(\gamma,\lambda)\omega(v_0,\gamma)z_\tau - h(y_\tau,\lambda)}{\kappa_{n0}^{-1}(\gamma)\hat{\sigma}^*(\gamma,\lambda)}, \tag{35}$$

where $\kappa_{n0}^{-2}(\gamma) = x_0'[\mathbf{X}'\Omega^{-1}(\gamma)\mathbf{X}]^{-1}x_0$, and $\hat{\sigma}^{*2}(\gamma,\lambda) = \hat{\sigma}^2(\gamma,\lambda)[n/(n-k)]$ is an unbiased estimator of σ^2 for given γ and λ. Now $T_\tau(\gamma,\lambda) \sim t_{n-k}[-\delta(\gamma)] + \delta(\gamma)$, where $\delta(\gamma) = \kappa_{n0}(\gamma)\omega_0(\gamma)z_\tau$. Thus, when γ and λ are known the confidence limits for y_τ can be constructed in the same way as in equations (10) and (11).

When both γ and λ are unknown and are replaced by their MLEs, the pivotal quantity becomes

$$T_\tau(\hat{\gamma},\hat{\lambda}) = \frac{x_0'\hat{\beta}(\hat{\gamma},\hat{\lambda}) + \hat{\sigma}^*(\hat{\gamma},\hat{\lambda})\omega(v_o,\hat{\gamma})z_\tau - h(y_\tau,\hat{\lambda})}{\kappa_{n0}^{-1}(\hat{\gamma})\hat{\sigma}^*(\hat{\gamma},\hat{\lambda})}. \tag{36}$$

The PQL can be constructed in the same way as in equation (12). The issue now is the adjustment of the PQL to account for the estimation of the weighting parameter as well as the transformation parameter. The following theorem provides a convenient way to perform the adjustment. We now denote $\psi = (\beta',\sigma^2,\gamma',\lambda)'$.

Theorem 3. *Under the specification of the model in equation (27), we assume further that the following are true: i)* $\hat{\lambda} \xrightarrow{p} \lambda$ *and* $\hat{\gamma} \xrightarrow{p} \gamma$*, ii)* $\sqrt{n}[(\hat{\gamma}-\gamma)',(\hat{\lambda}-\lambda)]' \xrightarrow{D} N(0,\Sigma)$*, iii)* $\mathbf{X}'\Omega^{-1}(\gamma)h_\lambda(\mathbf{Y},\lambda)/n$*,* $\mathbf{X}'\Omega^{-1}(\gamma)h_{\lambda\lambda}(\mathbf{Y},\lambda)/n$*, and* $h'(\mathbf{Y},\lambda)\mathbf{M}'(\gamma)\Omega^{-1}(\gamma)\mathbf{M}(\gamma)h_\lambda(\mathbf{Y},\lambda)/n$ *converge in probability, and iv)* $\mathbf{X}'\Omega^{-1}(\gamma)\mathbf{X}/n$ *converges to a positive definite matrix. Then, we have,*

$$T_\tau(\hat{\gamma},\hat{\lambda}) = T_\tau(\gamma,\lambda) + b_1'\sqrt{n}(\hat{\gamma}-\gamma) + b_2\sqrt{n}(\hat{\lambda}-\lambda) + o_p(1), \tag{37}$$

where

$$b_1 = \lim_{n\to\infty} \frac{x_0' \mathrm{E}[\hat{\beta}_\gamma(\gamma,\lambda)] + z_\tau \omega_0(\gamma)\mathrm{E}[\hat{\sigma}^*_\gamma(\gamma,\lambda)] + \hat{\sigma}^*(\gamma,\lambda)\omega_{0\gamma}(\gamma) z_\tau}{\sqrt{n}\sigma\kappa_{n0}^{-1}(\gamma)}, \tag{38}$$

$$b_2 = \lim_{n\to\infty} \frac{x_0' \mathrm{E}[\hat{\beta}_\lambda(\gamma,\lambda)] + z_\tau \omega_0(\gamma)\mathrm{E}[\hat{\sigma}^*_\lambda(\gamma,\lambda)] - h_\lambda(y_\tau,\lambda)}{\sqrt{n}\sigma\kappa_{n0}^{-1}(\gamma)}. \tag{39}$$

Furthermore, $T_\tau(\gamma,\lambda)$ is asymptotically independent of $\hat{\gamma}$ and $\hat{\lambda}$, so that

$$\mathrm{E}[T_\tau(\hat{\gamma},\hat{\lambda})] = \mathrm{E}[T_\tau(\gamma,\lambda)] + o(1), \tag{40}$$

$$\mathrm{Var}[T_\tau(\hat{\gamma},\hat{\lambda})] = \mathrm{Var}[T_\tau(\gamma,\lambda)] + c^2(\psi) + o(1) \tag{41}$$

where $c^2(\psi) = b'\Sigma b$, and $b' = (b_1', b_2)$.

The proof of Theorem 3 is given in Appendix A. Again, Theorem 3 shows that estimating the weighting and transformation parameters causes the variance of the pivotal quantity to be inflated. Thus, the pivotal quantity has to be corrected in order for the subsequent inferences to be valid. A consistent estimator of $c(\psi)$ is given by

$$\widehat{c(\psi)} = (\hat{b}'\hat{\Sigma}\hat{b})^{1/2}, \tag{42}$$

with $\hat{\Sigma}/n = J^{22}(\hat{\psi})$, the $(\gamma',\lambda)'$ diagonal block of $J^{-1}(\psi)$ evaluated at $\hat{\psi}$, and

$$\hat{b}_1 = \frac{x_0'\hat{\beta}_\gamma(\hat{\gamma},\hat{\lambda}) + z_\tau\omega_0(\hat{\gamma})\hat{\sigma}^*_\gamma(\hat{\gamma},\hat{\lambda}) + \hat{\sigma}^*(\hat{\gamma},\hat{\lambda})\omega_{0\gamma}(\hat{\gamma})z_\tau}{\sqrt{n}\hat{\sigma}^*(\hat{\gamma},\hat{\lambda})\kappa_{n0}^{-1}(\hat{\gamma})}, \tag{43}$$

$$\hat{b}_2 = \frac{x_0'\hat{\beta}_\lambda(\hat{\gamma},\hat{\lambda}) + z_\tau\omega_0(\hat{\gamma})\hat{\sigma}^*_\lambda(\hat{\gamma},\hat{\lambda}) - h_\lambda(\hat{y}_\tau,\hat{\lambda})}{\sqrt{n}\hat{\sigma}^*(\hat{\gamma},\hat{\lambda})\kappa_{n0}^{-1}(\hat{\gamma})}. \tag{44}$$

Using Theorem 3, the pivotal quantity $T_\tau(\hat{\gamma},\hat{\lambda})$ can be corrected as

$$T^*_\tau(\hat{\gamma},\hat{\lambda}) = \frac{T_\tau(\hat{\gamma},\hat{\lambda}) - C_m(\psi)}{C_s(\psi)}, \tag{45}$$

where $C_m(\psi) = \mu_T(\gamma)(1 - C_s(\psi))$ and $C_s(\psi) = \sqrt{1 + c^2(\psi)/\sigma_T^2(\gamma)}$, with $\mu_T(\gamma)$ and $\sigma_T^2(\gamma)$ being the mean and variance of $T_\tau(\gamma,\lambda)$, which have similar expressions as μ_T and σ_T^2 in equations (8) and (9) with a different noncentrality parameter. The implementation of the CPQL requires several partial derivatives. All the partial derivatives have analytical expressions except $\hat{\beta}_\gamma(\gamma,\lambda)$ and $\hat{\sigma}^*_\gamma(\gamma,\lambda)$ which may be calculated numerically. The CPQL takes the form

$$\left\{h^{-1}[L^*(\hat{\gamma},\hat{\lambda}),\hat{\lambda}],\ h^{-1}[U^*(\hat{\gamma},\hat{\lambda}),\hat{\lambda}]\right\}, \tag{46}$$

with the two adjusted end points before transformation being

$$L^*(\hat{\gamma}, \hat{\lambda}) = x_0'\hat{\beta}(\hat{\gamma}, \hat{\lambda}) + \widehat{C_m^*(\psi)} - t_{n-k}^{1-\alpha/2}[-\delta(\hat{\gamma})]\widehat{C_s(\psi)}\frac{\hat{\sigma}^*(\hat{\gamma}, \hat{\lambda})}{\kappa_{n0}(\hat{\gamma})}, \tag{47}$$

$$U^*(\hat{\gamma}, \hat{\lambda}) = x_0'\hat{\beta}(\hat{\gamma}, \hat{\lambda}) + \widehat{C_m^*(\psi)} - t_{n-k}^{\alpha/2}[-\delta(\hat{\gamma})]\widehat{C_s(\psi)}\frac{\hat{\sigma}^*(\hat{\gamma}, \hat{\lambda})}{\kappa_{n0}(\hat{\gamma})}, \tag{48}$$

and $\widehat{C_m^*(\psi)} = \hat{\sigma}^*(\hat{\gamma}, \hat{\lambda})[1 - \widehat{C_s(\psi)}][\omega_0(\hat{\gamma})z_\tau + \mu_T(\hat{\gamma})/\kappa_{n0}(\hat{\gamma})]$.

4. MONTE CARLO RESULTS

The results given in Sections 2 and 3 provide asymptotically correct confidence limits for the regression quantiles, namely the CPQL. In this section we examine the small-sample performance of the CPQL through Monte Carlo simulations. In particular, we (i) compare the CPQL with the delta-method quantile limits (DQL) and the PQL; (ii) investigate the robustness of the CPQL against certain departures from normality of the error distribution, and (iii) compare the CPQL with the quantile regression quantile limits (QRQL), i.e., the quantile limits obtained from the quantile regression method with the Box-Cox transformation applied to the response. Appendix B gives some technical details for the DQL, and Appendix C describes briefly the QRQL.

We consider cases of homoscedastic and heteroscedastic errors. For homoscedastic errors we consider the following model as the data generation process (DGP):

$$\text{Model 1:} \quad h(Y_i, \lambda) = \beta_0 + \beta_1 X_i + \sigma e_i, \qquad i = 1, \cdots, n. \tag{49}$$

For heteroscedastic errors we use the following DGP:

$$\text{Model 2:} \quad h(Y_i, \lambda) = \beta_0 + \beta_1 X_i + \sigma \exp(\gamma X_i) e_i, \qquad i = 1, \cdots, n. \tag{50}$$

In each of the above models, X_i are fixed values of the regressor that are uniformly spaced in $[0, X_m]$, where X_m is the maximum value of X_i in the setup. For Model 1, we consider both the Box-Cox and dual-power transformations. For Model 2, to simplify the numerical computations, only the Box-Cox transformation is considered.

The Monte Carlo experiment is described as follows. For a given set of parameter values, we generate n standard normal random numbers (e_i). Using these random numbers we calculate the values of Y_i based on the assumed DGP and estimate the model parameters. We then compute the 95% quantile limits. This process is repeated 10,000 times (except the case of QRQL where 1,000 is used due to the computational demand). The proportion of the quantile limits that cover the true quantile provides a Monte Carlo estimate of the coverage probability of the quantile limits.

4.1. Small Sample Properties of CPQL

In this subsection, we focus on the small sample properties of the CPQL and its comparison with the DQL and the PQL as these methods are based on the same model. We conduct Monte Carlo experiments based on the DGPs specified by Models 1 and 2, using the following parameter configurations: $\beta_0 = 5$, $\beta_1 = 1$, $\sigma = \{0.1, 0.5, 1.0, 2.0\}$, $\gamma = \{0.01, 0.1, 0.5\}$, $\lambda = \{0, 0.01, 0.1, 0.5\}$, $\tau = \{0.01, 0.05, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99\}$, $n = \{15, 60, 200, 500\}$, and $X_m = \{10, 50, 100\}$. We shall report some selected results based on Model 1. The complete results are available from the authors on request.

First, Figure 1 plots the values of $c(\psi)$ (calculated based on the results of Theorem 2) versus x_0 when the DGP is Model 1 and the Box-Cox transformation (BCT) is used. The values of X used in this computation are uniformly spaced in [0, 50]. From the results we see that the required variance correction factor $c(\psi)^2$ is generally quite large, especially when x_0 is far from the in-sample values. This suggests that it is important to make corrections to the plug-in method. It is also interesting to note that the value of c_0 is quite robust with respect to the sample size n, and changes only very little with respect to τ.

Figures 2 and 3 plot the empirical relative frequency of coverage of the true quantiles at the x_0 value using the three methods, based on, respectively, the Box-Cox transformation and dual-power transformation (DPT). It is clear from the plots that the CPQL has the best performance in all cases, with its empirical coverage probabilities generally very close to the nominal level even when sample size is 15. In contrast, the PQL has the poorest performance. The empirical coverage probabilities do not get nearer to the nominal value of 95% when the sample size increases, and quickly go down when x_0 moves away from the design region. The performance of the DQL depends on several factors. In what follows we summarize some regularities that seem to have emerged from the experiment. These regularities are in line with the general observations drawn from Theorem 2. Our discussion will focus on the following aspects: (i) the effect of the nonlinearity parameter λ, (ii) the effect of the sample size, (iii) performance in out-of-sample forecast, and (iv) the effect of heteroscedasticity.

Degree of nonlinearity. The transformation is more nonlinear for smaller λ. The Monte Carlo results show that λ has a rather significant effect on the performance of the DQL, especially when τ is large. Generally speaking, when the degree of nonlinearity is high, the empirical coverage of the DQL can be far below the nominal level even when x_0 is within the design region [0, 50]. For example, from the plot corresponding to $n = 15$ and $\tau = 0.99$, we see that the coverage probability for the nominal 95% CIs of the 0.99-quantile ranges roughly from 0.825 to 0.85 for $x_0 \in [0, 50]$. Increasing λ from

0.01 to 0.1 (i.e., reducing the nonlinearity) improves the coverage of the DQL to between 0.86 and 0.90 (results not reported for brevity), but it is still far below the nominal level of 0.95. In contrast, the empirical coverage probabilities of the CPQL method are quite close to the nominal 0.95 value in all the cases considered.

Sample size. One of the striking phenomenon observed from the simulation results is that the CPQL performs reasonably well even when the sample size is only 15. When the sample size increases, the coverage of CPQL quickly converges to its nominal level. The empirical coverage of the DQL improves as the sample size increases. However, in many situations it can still be significantly below its nominal level even when the sample size is as large as 500 (see the last six plots of Figure 3). As the PQL is an incorrect QL, its performance does not necessarily improve when the sample size is increased.

Out-of-sample forecast. Out-of-sample forecast or extrapolation beyond the data range is an important topic in duration and event-time analyses. In this case, the largest difference between CPQL and DQL is observed. The plots in Figures 2 and 3 show that as x_0 moves away from the design region [0, 50], the empirical coverage probability stays very close to the nominal level for the CPQL in all cases. However, it drops significantly for the DQL in the cases of high degrees of nonlinearity even when the sample size is as large as 500, and simply breaks down for the PQL. The results clearly demonstrate the superior performance of the corrected plug-in method, the unsatisfactory performance of the delta method, and the poor performance of the plug-in method.

Heteroscedasticity. We also performed simulation using the DGP specified by Model 2 where a simple heteroscedastic structure is imposed. The results (not reported for brevity) show that estimating the weighting parameter does not have significant effect on the performance of the CPQL method. In contrast, the PQL method may perform very poorly. Once again, the performance of the delta method may be quite poor if λ is small or x_0 is outside the design region. In all cases, the performance of the delta method is dominated by the CPQL method.

Another undesirable property of the delta method is that it may end up with a negative lower limit for the supposedly positive y_τ. This phenomenon, however, cannot happen for the CPQL method under the dual-power transformation or other transformations that are free of the truncation problem. For the case of the Box-Cox transformation, the quantity $(1+\lambda L^*)$ or $(1+\lambda U^*)$ could end up being negative (depending on whether λ is positive or negative) so that $h^{-1}(L^*, \lambda)$ or $h^{-1}(U^*, \lambda)$ does not exist. When this happens, either the lower limit of the CPQL should be set to zero or the upper limit of the CPQL should be set to infinity. However, the chance of this happening is negligible if the truncation probability is small.

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

4.2. Robustness of the CPQL

As mentioned at the beginning of Section 3, the purposes of applying the data transformation technique are to induce normality, simplicity (of model structure) and homoscedasticity. However, it is generally recognized that with a single transformation these three goals may not be achieved simultaneously, in particular the normality and homoscedasticity in the context of economics studies. Section 3 relaxes the homoscedasticity assumption by adding a general heteroscedastic structure. Model structure is an issue of concern only in the modeling stage, thus what is left is the normality. We now relax the normality assumption and examine the effect of the non-normality of e_i on the coverage probability of the CPQL through Monte Carlo simulations. We argue that although a transformation may not be able to induce exact normality, it is able to bring the data to "near normality". Therefore, it is reasonable to investigate the performance of the CPQL only under a "mild" departure from normality. We assume that e_i are independent and identically distributed (iid) with zero mean and unit variance, and consider the cases that they are drawn from a normal-mixture, a t-distribution with ν degrees of freedom t_ν, and a normal-gamma mixture. Thus, departure from normality is due to changes in tail areas, or an increase in kurtosis or skewness or both.

Figures 4-6 present selected plots of the coverage probabilities of the CPQL versus x_0 under Model 2 with $X_i \in [0, 10]$ and $\sigma = \gamma = \lambda = 0.1$. The complete results are available from the authors on request. In Figure 4, $e_i \sim (1-p)\phi(x) + \frac{p}{1.5}\phi(\frac{x}{1.5})$ where ϕ denotes the standard normal density function with p chosen to be 0.1, 0.2 and 0.3, representing 10%, 20% and 30% contaminations, respectively. In Figure 5, $e_i \sim t_\nu$ with $\nu = 20, 12$ and 8. Finally in Figure 6, $e_i \sim (1-p)\phi(x) + p\ g(x; a, b)$ where $g(x; a, b)$ denotes the gamma density with the scale parameter a chosen to be 1 and the shape parameter b chosen to be 2 or 1, and the mixing probability p being 0.1, 0.2 and 0.3.

To see how much the error distribution deviates from normality, we calculate the skewness and kurtosis (which assume values of 0 and 3 for the case of exact normality). For the case of normal-mixture, the skewness is zero, and the kurtosis is 3.33, 3.48 and 3.52, respectively for $p = 0.1$, 0.2 and 0.3. For the case of the t errors, the skewness is again zero, and the kurtosis is 3.375, 3.75 and 4.5, respectively, for $\nu = 20$, 12 and 8. For the normal-gamma(1,2) mixture, the skewness and kurtosis are (0.14, 0.28, 0.42) and (3.3, 3.6, 3.9), corresponding to $p = (0.1, 0.2, 0.3)$. Finally for the normal-gamma(1,1) mixture, the skewness and kurtosis are (0.2, 0.4, 0.6) and (3.6, 4.2, 4.8), corresponding to $p = (0.1, 0.2, 0.3)$.

From the plots in Figures 4-6, the following general observations emerge: (i) the CPQL is in general quite robust against mild departures from normality, (ii) the CPQLs for middle quantiles (0.25, 0.5 and 0.75) are more robust against non-normality than those for the extreme quantiles, and iii) in the cases that the distribution of e_i is skewed to the right (the case of normal-gamma mixture), both the middle and lower quantiles are quite robust against non-normality; only the upper extreme quantiles (0.9, 0.95 and 0.99) are sensitive to non-normality. However, if the contamination rate is kept within 20% for the normal-gamma(1,2) mixture, and 10% for the normal-gamma(1,1) mixture, the CPQL still provides a reasonable coverage percentage for the upper quantiles. Note that the results corresponding to $\tau = \{0.05, 0.25, 0.95\}$ are available from the authors.

4.3. A Comparison of CPQL with QRQL

Quantile regression introduced by Koenker and Bassett (1978) has become a popular method in economics research (e.g., Chamberlain, 1994; Buchinsky, 1994, 1998; Machado and Mata, 2005; Maitra and Vahid, 2006). Applying the Box-Cox transformation to the response (Powell, 1991) gives additional flexibility to the already flexible quantile regression (Buchinsky, 1995; Machado and Mata, 2000; Tian and Wei, 2002). However, little is known about the finite sample performance of the quantile regression quantile limits (QRQL). Also, as pointed out by Kocherginsky *et al.* (2005), calculation of the confidence intervals using the quantile regression method may require a large sample. Thus, it would be interesting to compare our method with the quantile regression method in quantile limits construction. To be consistent in the functional forms of the models involved so that a fair comparison can be made, we compare the CPQL based on the Box-Cox heteroscedastic regression in (27) with the QRQL based on the Box-Cox quantile regression. A brief description of the construction of the QRQL based on the Box-Cox quantile regression is given in Appendix C.

As the simulation for QRQL is computationally very demanding, we restrict the Monte Carlo samples for the estimation of the coverage probability to 1,000. The number of runs for CPQL is still 10,000. Two Monte Carlo experiments are conducted based on the DGP specified by Model 2, one with iid standard normal errors, and the other with the iid normal-gamma(1,2) mixture. As we are comparing quantile limits constructed from two different models, it is necessary to compare their coverage probability as well as their average interval length. The results are summarized in Tables 1 and 2. From the results we see that the CPQL clearly outperforms the QRQL in terms of both the coverage probability and average length of the confidence intervals. In particular, the CPQL not only has a higher coverage which is generally very close to the nominal level of 0.95, but also has a shorter length than QRQL on average. The coverage of QRQL can be significantly below the nominal level, in particular for $x_0 = 0$.

5. AN EMPIRICAL APPLICATION

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

To illustrate the application of our methodology, we consider an empirical example using the data on living standards in South Africa 1993-98, recently analyzed by Maitra and Vahid (2006) using quantile regression. These authors take into account the effect of sample attrition or missing data – certain families dropped out of the study in 1998. To keep our discussions within the scope of this paper, we do not address the issue of sample attrition, and use only the 1993 data with 1354 observations.

We use the identical set of variables as in Maitra and Vahid (2006), namely, the per capita household expenditure (the response variable); the age of the household head (`AGEHD`); the age squared (`AGEHD2`); a dummy to indicate whether the household head is a female (`FHH`); the highest level of education attained by the household head, which is accounted for by including three dummies: `HDEDUC1` (primary), `HDEDUC2` (secondary) and `HDEDUC3` (more than secondary), all referred to the category of no education; the total number of children in the household (`TOTCHILD`, individuals aged 0-17); the total number of working aged adults (`TOTADULT`, males aged 18-64 and females aged 18-59); the total number of elderly in the household (`TOTELDER`, individuals above the working age); the race dummy `BLACK`; the location dummies `RURAL` (to account for rural residence) and `NATAL` (to account for the residence in the province of NATAL). As in Maitra and Vahid (2006), we also include the `ATTRITE` dummy in the model, which indicates whether a household leaves the sample in the latter survey in 1998. Furthermore, the interactions of `ATTRITE` with other regressors, denoted by the prefix A-, are also considered.

The reason for repeating this study is clear: to see if our results are comparable to those obtained by Maitra and Vahid (2006) using quantile regression. Note that Maitra and Vahid fixed the functional form for the expenditure distribution to be log-linear, i.e., they run a linear quantile regression on the log per capita household expenditure. In contrast, we fit a Box-Cox heteroscedastic regression with the response (per capita household expenditure) subject to an unknown Box-Cox power transformation.

Major questions to be answered in this empirical application include: (i) are the estimated relationships consistent with those obtained from the quantile regression? (ii) is the Box-Cox heteroscedastic regression a viable approach to examining the living standards in South Africa in the sense that, like the quantile regression, it allows one to examine the relationship between the explanatory variables and dependent variables at different points on the expenditure distribution, and to reflect the changes of such a relationship as one moves along the expenditure distribution? (iii) is the CPQL indeed more reliable than DQL, PQL and QRQL?

The model estimates are summarized in Table 3, which should be compared with the results contained in Table 4 of Maitra and Vahid (2006). From Table 3, we find that (a) the coefficients of the major family characteristics: `FHH`, `HEEDUC1`, `HDEDUC2`, HDEDUC3, BLACK, NATAL, RURAL, TOTCHILD and TOTADULT, are all highly significant and have the same sign as those based on quantile regression; (b) ATTRITE and its interactions are collectively significant at the 1% level of significance; (c) the transformation parameter is significantly different from 0, suggesting that the log-linear functional form may be inappropriate, and (d) there is indeed significant evidence for the existence of heteroscedasticity. Overall, the estimated relationship is in general consistent with that based on the quantile regression with the log per capita household expenditure as the dependent variable.

The total number of children is one of the most important household characteristics that affect the family living standards. In Figure 6 we plot the quantiles and the quantile limits at various quantile levels against TOTCHILD. Although the number of families with TOTCHILD larger than 8 is quite small, we consider values of TOTCHILD ranging from 0 to 15 to examine the performance of various quantile limits in out-of-sample forecast. For clarity, the first half of Figure 6 (nine plots) compares CPQL with DQP and PQL, and the second half compares CPQL with QRQL. From the first nine plots we see that the CPQL, PQL and DQL can differ substantially, especially at the lower quantiles and the upper extreme quantiles, with PQL being the most narrow, followed by DQL and CPQL. This suggests that the practical application of PQL and DQL can be quite misleading. From the second nine plots we see that the point estimates of the quantiles in the original scale based on the Box-Cox heteroscedastic regression (solid line) and the quantile regression (dotted line) do not differ much, which suggests that, like the quantile regression, Box-Cox heteroscedastic regression also allows one to examine the relationship between the explanatory variables and dependent variables at different points on the expenditure distribution, and to reflect the changes of such a relationship as one moves along the expenditure distribution. However, the two methods may produce quite different quantile limits as seen from the plots in the second half of Figure 6. Based on the simulation results given in Tables 1 and 2, CPQL may be more reliable than QRQL.

6. CONCLUSIONS AND DISCUSSIONS

Transformation is a popular technique for analyzing skewed data found in many economic applications. When the transformation is known, the usual inference theories can be applied and simple inverse transformations lead to inferences on the original response. However, the transformation parameter is often unknown and has to be estimated from the data. A common practice in this case is to use the so-called 'plug-in' method, i.e., plugging the estimated unknown parameters into the confidence-interval formula. This practice ignores the effect of estimating the transformation and/or the weighting, and is shown to be asymptotically invalid.

We modify the plug-in method to give an asymptotically correct confidence interval. The gains in introducing the corrected plug-in method are significant as the Monte Carlo results show that it performs very well even in small samples and dominates uniformly the delta method. A Monte Carlo comparison with the confidence interval obtained through a different model, the Box-Cox quantile regression, shows that the corrected plug-in method is superior in terms of both the coverage probability and the average length of the interval. An empirical application on living standards in South Africa shows that the Box-Cox heteroscedastic regression provides comparable results to those based on quantile regression.

There are important implications for the corrected plug-in method. For example, it can be applied to a model where both the response and regressors are transformed. Generally speaking, the methodology can be applied to any situation within the likelihood inference framework where normal inference theory is available when certain parameters (such as the transformation parameters) are known, but not available when these parameters are unknown and have to be estimated from the same set of data. See Yang *et al.* (2006) for more discussions.

It may be desirable to give a more thorough investigation on the finite sample performance of QRQL, in particular when the construction of QRQL is based on a different method of covariance estimation such as the resampling method of Tian and Wei (2002). However, such a study is beyond the scope of this paper, and thus is left for a future study.

We end the paper by offering some intuitive explanations on why our method performs better than the delta method. The performance of the large-sample normal-theory method can be improved by considering a proper transformation. For example, if θ is a positive quantity (e.g., a quantile duration or event time), then it is better to assume $\log\hat{\theta}$ (rather than $\hat{\theta}$) follow an asymptotic normal distribution; if θ is restricted to be between 0 and 1 (e.g., a survivor function), confidence interval based on $\log[\hat{\theta}/(1-\hat{\theta})]$ usually performs better than that based on $\hat{\theta}$ (Hahn and Meeker, 1991, p.239; Agresti, 1990, p.54). A quantile must be positive if the original observations are. Delta method works directly with the quantile, whereas our method works first on the transformed quantile and then retransform. Thus, it is expected that our method works better in finite samples. More importantly, as mentioned at the beginning of the paper, our method is able to take the advantage of the exact finite sample distribution of the underlying (λ-known) pivotal quantity, thus is expected to perform well in small samples.

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

APPENDIX A: PROOF OF THE THEOREMS

Proof of Theorem 1. First-order Taylor expansions of $\hat\beta(\hat\lambda)$, $h(y_\tau, \hat\lambda)$ and $\hat\sigma^*(\hat\lambda)$ give

$$\begin{aligned}\hat\beta(\hat\lambda) &= \hat\beta(\lambda) + \hat\beta_\lambda(\lambda)(\hat\lambda - \lambda) + O_p(n^{-1}),\\ h(y_\tau, \hat\lambda) &= h(y_\tau, \lambda) + h_\lambda(y_\tau, \lambda)(\hat\lambda - \lambda) + O_p(n^{-1}),\\ \hat\sigma^*(\hat\lambda) &= \hat\sigma^*(\lambda) + \hat\sigma^*_\lambda(\lambda)(\hat\lambda - \lambda) + O_p(n^{-1}).\end{aligned}$$

Combining the above, we have

$$\begin{aligned}& x_0'\hat\beta(\hat\lambda) + \hat\sigma^*(\hat\lambda) z_\tau - h(y_\tau, \hat\lambda)\\ = \;& x_0'\hat\beta(\lambda) + \hat\sigma^*(\lambda) z_\tau - h(y_\tau, \lambda) + (\hat\lambda - \lambda)[x_0'\hat\beta_\lambda(\lambda) + \hat\sigma^*_\lambda(\lambda) z_\tau - h_\lambda(y_\tau, \lambda)] + O_p(n^{-1}).\end{aligned}$$

As $1/\hat\sigma^*(\hat\lambda) = 1/\hat\sigma^*(\lambda) + O_p(n^{-1/2}) = 1/\sigma + O_p(n^{-1/2})$, the first part of Theorem 1 (equation (13)) is obtained. Note that $T_\tau(\lambda)$ depends only on the constrained (upon λ) MLE of β and σ, and $\hat\lambda$ is the unconstrained MLE of λ. Thus, by Lemma 1 of Yang *et al.* (2006), $T_\tau(\lambda)$ and $\hat\lambda$ are asymptotically independent. The rest of the theorem is then straightforward.

Proof of Theorem 2. The key to the proof of this theorem is the derivation of explicit expressions for $v^2(\psi)$ defined in Theorem 1. From the asymptotic theories of likelihood inference, $v^2(\psi)$ involves elements of the expected information matrix. After simplification, it can be shown that $v^2(\psi)$ depends on the quantities $\mathrm{E}[h_\lambda(Y_i, \lambda)]$, $\mathrm{Var}[h_\lambda(Y_i, \lambda)]$, $\mathrm{E}[e_i h_\lambda(Y_i, \lambda)]$, and $\mathrm{E}[e_i h_{\lambda\lambda}(Y_i, \lambda)]$. The expressions for $h_\lambda(Y_i, \lambda)$ and $h_{\lambda\lambda}(Y_i, \lambda)$ are given in Appendix B. Thus, when $\lambda = 0$, direct substitution of $\log Y_i = x_i'\beta + \sigma e_i$ into the above leads to explicit expressions for $v^2(\psi)$. For $\lambda \neq 0$, it is necessary to approximate $\log Y_i$. A Taylor series expansion provides the following approximation

$$\lambda \log Y_i = \phi_i + \theta_i e_i - \frac{1}{2}\theta_i^2 e_i^2 + O_p(\theta_i^3),$$

which gives, after some algebra, the following approximate expression for $v^2(\psi)$,

$$v^2(\psi) = \begin{cases} \frac{n\lambda^2}{\|M(\theta^{-1}\odot\phi + \frac{1}{2}\theta)\|^2 + 2\|\phi - \bar\phi\|^2 + \frac{3}{2}\|\theta\|^2} + O_p(n^{-1/2}) + O_p(\theta_M^3), & \text{if } \lambda \neq 0,\\ \frac{4n\sigma^2}{\|M\eta^2\|^2 + 8\sigma^2\|\eta - \bar\eta\|^2 + 6n\sigma^4} + O_p(n^{-1/2}), & \text{if } \lambda = 0.\end{cases}$$

The other terms in $c(\psi)$ can be evaluated in the same manner, completing the proof.

Proof of Theorem 3. First-order Taylor expansion of the numerator of $T_\tau(\hat\gamma, \hat\lambda)$ gives

$$\begin{aligned}& x_0'\hat\beta(\hat\gamma, \hat\lambda) + \hat\sigma^*(\hat\gamma, \hat\lambda)\omega_0(\hat\gamma) z_\tau - h(y_\tau, \hat\lambda)\\ = \;& x_0'\hat\beta(\gamma, \lambda) + \hat\sigma^*(\gamma, \lambda)\omega_0(\gamma) z_\tau - h(y_\tau, \lambda)\\ & + [x_0'\hat\beta_\lambda(\gamma, \lambda) + \hat\sigma^*_\lambda(\gamma, \lambda)\omega_0(\gamma) z_\tau - h_\lambda(y_\tau, \lambda)](\hat\lambda - \lambda)\\ & + [x_0'\hat\beta_\gamma(\gamma, \lambda) + \hat\sigma^*_\gamma(\gamma, \lambda)\omega_0(\gamma) z_\tau + \hat\sigma^*(\gamma, \lambda)\omega_{0\gamma}(\gamma) z_\tau](\hat\gamma - \gamma) + O_p(n^{-1}).\end{aligned}$$

As $1/\hat{\sigma}^*(\hat{\gamma},\hat{\lambda}) = 1/\hat{\sigma}^*(\gamma,\lambda) + O_p(n^{-1/2}) = 1/\sigma + O_p(n^{-1/2})$, the above leads to the first part of the theorem. The asymptotic independence between $T_\tau(\gamma,\lambda)$ and $\{\hat{\gamma},\hat{\lambda}\}$ is a consequence of Lemma 1 of Yang *et al.* (2006). The rest of the proof is straightforward.

APPENDIX B: DELTA METHOD QUANTILE LIMITS

Carroll and Ruppert (1991) recommended using the delta method or the likelihood ratio test method to construct confidence intervals for y_τ. Implementation of the delta method is straightforward. Implementation of the likelihood ratio test method requires constrained maximization. To apply the delta method, let $\hat{\psi}$ be the MLE of ψ where ψ may or may not contain γ. Write $\hat{\beta}(\hat{\lambda})$ or $\hat{\beta}(\hat{\gamma},\hat{\lambda})$ as $\hat{\beta}$, and $\hat{\sigma}(\hat{\lambda})$ or $\hat{\sigma}(\hat{\gamma},\hat{\lambda})$ as $\hat{\sigma}$. Suppose that the distribution of $\hat{\psi}$ is asymptotically normal with mean ψ and covariance matrix $I^{-1}(\psi)$. Then the distribution of the MLE of y_τ, $\hat{y}_\tau = h^{-1}[(x_0'\hat{\beta} + \hat{\sigma}\omega_0(\hat{\gamma})z_\tau), \hat{\lambda}] \equiv g(\hat{\psi})$, is asymptotically normal with mean $g(\psi)$ and variance $g_\psi'(\psi)I^{-1}(\psi)g_\psi(\psi)$. The variance can be consistently estimated by $g_\psi'(\hat{\psi})J^{-1}(\hat{\psi})g_\psi(\hat{\psi})$ with $J(\hat{\psi})$ being the observed information matrix evaluated at $\hat{\psi}$. Thus, a $100(1-\alpha)\%$ large-sample confidence interval for y_τ is given by

$$\hat{y}_\tau \pm z_{\alpha/2}\sqrt{g_\psi'(\hat{\psi})J^{-1}(\hat{\psi})g_\psi(\hat{\psi})}.$$

Implementation of the delta-method QL requires the quantile function and its partial derivatives, as well as the observed information matrix. The details are provided below.

The quantile function and partial derivatives. The quantile function for the general transformation model with heteroscedastic errors is defined as

$$g(\psi) = h^{-1}[x_0'\beta + \sigma\omega_0(\gamma)z_\tau, \lambda].$$

For the Box-Cox transformation model with heteroscedastic errors, $g(\psi) = (1 + \lambda\mu_0)^{1/\lambda}$, where $\mu_0 = x_0'\beta + \sigma\omega_0(\gamma)z_\tau$. Hence, the partial derivatives of $g(\psi)$ are

$$\begin{aligned}
g_\beta(\psi) &= (1+\lambda\mu_0)^{(1-\lambda)/\lambda}x_0 \\
g_{\sigma^2}(\psi) &= \frac{1}{2\sigma}(1+\lambda\mu_0)^{(1-\lambda)/\lambda}\omega_0(\gamma)z_\tau \\
g_\gamma(\psi) &= (1+\lambda\mu_0)^{(1-\lambda)/\lambda}\sigma z_\tau\omega_{0\gamma}(\gamma) \\
g_\lambda(\psi) &= g(\psi)\left[\mu_0\lambda^{-1}(1+\lambda\mu_0)^{-1} - \lambda^{-2}\log(1+\lambda\mu_0)\right].
\end{aligned}$$

Setting $\omega_0(\gamma) = 1$ in the above expressions and removing the $g_\gamma(\psi)$ element gives the partial derivatives of the quantile function for the Box-Cox transformation model with homoscedastic errors.

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

For the dual-power transformation model with heteroscedastic errors, we have $g(\psi) = [\lambda\mu_0 + (1+\lambda^2\mu_0^2)^{1/2}]^{1/\lambda}$, and the partial derivatives are

$$\begin{aligned}
g_\beta(\psi) &= g(\psi)x_0(1+\lambda^2\mu_0^2)^{-1/2}\\
g_{\sigma^2}(\psi) &= \frac{1}{2\sigma}g(\psi)\omega_0(\gamma)z_\tau(1+\lambda^2\mu_0^2)^{-1/2}\\
g_\gamma(\psi) &= g(\psi)\sigma\omega_{0\gamma}(\gamma)z_\tau(1+\lambda^2\mu_0^2)^{-1/2}\\
g_\lambda(\psi) &= g(\psi)\left[\mu_0\lambda^{-1}(1+\lambda^2\mu_0^2)^{-1/2} - \lambda^{-2}\log(\lambda\mu_0 + (1+\lambda^2\mu_0^2)^{1/2})\right].
\end{aligned}$$

Again, setting $\omega_0(\gamma) = 1$ and removing the $g_\gamma(\psi)$ gives the partial derivatives of the quantile function in the dual-power transformation model with homoscedastic errors.

The observed information matrix. For models with general transformation and weighting functions, the elements of the observed information matrix are given by:

$$\begin{aligned}
J_{\beta\beta} &= \frac{1}{\sigma^2}\sum_{i=1}^n \frac{x_i x_i'}{\omega_i^2(\gamma)}\\
J_{\sigma^2\sigma^2} &= \frac{1}{\sigma^6}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]^2}{\omega_i^2(\gamma)} - \frac{n}{2\sigma^4}\\
J_{\gamma\gamma} &= \sum_{i=1}^n\left(\frac{\omega_{i\gamma\gamma'}(\gamma)}{\omega_i(\gamma)} - \frac{\omega_{i\gamma}(\gamma)\omega_{i\gamma}'(\gamma)}{\omega_i^2(\gamma)}\right) - \frac{1}{\sigma^2}\sum_{i=1}^n[h(Y_i,\lambda)-x_i'\beta]^2\left(\frac{\omega_{i\gamma\gamma'}(\gamma)}{\omega_i^3(\gamma)} - \frac{3\omega_{i\gamma}(\gamma)\omega_{i\gamma}'(\gamma)}{\omega_i^4(\gamma)}\right)\\
J_{\lambda\lambda} &= \frac{1}{\sigma^2}\sum_{i=1}^n \frac{h_\lambda^2(Y_i,\lambda) + [h(Y_i,\lambda)-x_i'\beta]h_{\lambda\lambda}(Y_i,\lambda)}{\omega_i^2(\gamma)} - \sum_{i=1}^n\left(\frac{h_{y\lambda\lambda}(Y_i,\lambda)}{h_y(Y_i,\lambda)} - \frac{h_{y\lambda}^2(Y_i,\lambda)}{h_y^2(Y_i,\lambda)}\right)\\
J_{\beta\sigma^2} &= \frac{1}{\sigma^4}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]x_i}{\omega_i^2(\gamma)}\\
J_{\beta\gamma} &= \frac{2}{\sigma^2}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]x_i\omega_{i\gamma}(\gamma)}{\omega_i^3(\gamma)}\\
J_{\beta\lambda} &= \frac{1}{\sigma^2}\sum_{i=1}^n \frac{h_\lambda(Y_i,\lambda)x_i}{\omega_i^2(\gamma)}\\
J_{\sigma^2\gamma} &= \frac{1}{\sigma^4}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]^2\omega_{i\gamma}(\gamma)}{\omega_i^3(\gamma)}\\
J_{\sigma^2\lambda} &= -\frac{1}{\sigma^4}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]h_\lambda(Y_i,\lambda)}{\omega_i^2(\gamma)}\\
J_{\gamma\lambda} &= -\frac{2}{\sigma^2}\sum_{i=1}^n \frac{[h(Y_i,\lambda)-x_i'\beta]h_\lambda(Y_i,\lambda)\omega_{i\gamma}(\gamma)}{\omega_i^3(\gamma)}.
\end{aligned}$$

Now, for the Box-Cox transformation, we have $h_y(y,\lambda) = y^{\lambda-1}$, $h_{y\lambda}(y,\lambda) = y^{\lambda-1}\log y$, $h_{y\lambda\lambda}(y,\lambda) = y^{\lambda-1}(\log y)^2$, and (for $y > 0$)

$$h_\lambda(y,\lambda) = \begin{cases} \frac{1}{\lambda}[1+\lambda h(y,\lambda)]\log y - \frac{1}{\lambda}h(y,\lambda), & \lambda \neq 0,\\ \frac{1}{2}(\log y)^2, & \lambda = 0, \end{cases}$$

$$h_{\lambda\lambda}(y,\lambda) = \begin{cases} h_\lambda(y,\lambda)(\log y - \frac{1}{\lambda}) + \frac{1}{\lambda^2}[h(y,\lambda) - \log y], & \lambda \neq 0, \\ \frac{1}{3}(\log y)^3, & \lambda = 0. \end{cases}$$

For the dual-power transformation, we have

$$\begin{aligned} h_\lambda(y,\lambda) &= \frac{1}{2\lambda}(y^\lambda + y^{-\lambda})\log y - \frac{1}{\lambda}h(y,\lambda) \\ h_{\lambda\lambda}(y,\lambda) &= h(y,\lambda)(\frac{2}{\lambda^2} + (\log y)^2) - \frac{1}{\lambda^2}(y^\lambda + y^{-\lambda})\log y \\ h_y(y,\lambda) &= \frac{1}{2}(y^{\lambda-1} + y^{-\lambda-1}), \\ h_{y\lambda}(y,\lambda) &= \frac{1}{2}(y^{\lambda-1} - y^{-\lambda-1})\log y \\ h_{y\lambda\lambda}(y,\lambda) &= \frac{1}{2}(y^{\lambda-1} + y^{-\lambda-1})(\log y)^2. \end{aligned}$$

Setting $\omega_i(\gamma) \equiv 1$ gives the information matrix for models with homoscedastic errors. Furthermore, for the Box-Cox transformation model with homoscedastic errors, the observed information matrix reduces to

$$\begin{aligned} J_{\beta\beta} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} \\ J_{\sigma^2\sigma^2} &= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\ J_{\lambda\lambda} &= \frac{1}{\sigma^2}[h'_\lambda(\mathbf{Y},\lambda)h_\lambda(\mathbf{Y},\lambda) + (h(\mathbf{Y},\lambda) - \mathbf{X}\beta)'h_{\lambda\lambda}(\mathbf{Y},\lambda)] \\ J_{\beta\sigma^2} &= \frac{1}{\sigma^4}\mathbf{X}'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\ J_{\beta\lambda} &= -\frac{1}{\sigma^2}\mathbf{X}'h_\lambda(\mathbf{Y},\lambda) \\ J_{\sigma^2\lambda} &= -\frac{1}{\sigma^4}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'h_\lambda(\mathbf{Y},\lambda). \end{aligned}$$

For the dual-power transformation model with homoscedastic errors, the information matrix takes identical forms as the above except that the quantity $4\sum_{i=1}^n(\log Y_i)^2/(Y_i^\lambda + Y_i^{-\lambda})^2$ has to be subtracted from $J_{\lambda\lambda}$.

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

APPENDIX C: QUANTILE LIMITS BASED ON THE BOX-COX QUANTILE REGRESSION

Box-Cox quantile regression assumes $h(y_{i\tau}, \lambda(\tau)) = x_i'\beta(\tau)$, where $y_{i\tau}$ is the τ-quantile of the response y_i, x_i is the ith row of $\mathbf{X}$ and h is the Box-Cox power transformation. Equivalently, the model can be written as $y_{i\tau} = (1+\lambda(\tau)x_i'\beta(\tau))^{1/\lambda(\tau)}$. The τ-dependent parameters $\beta(\tau)$ and $\lambda(\tau)$ can be estimated jointly by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - (1+\lambda(\tau)x_i'\beta(\tau)))^{1/\lambda(\tau)}$$

where ρ is the check function taking the form $\rho_\tau(u) = \tau|u|\mathbf{1}_{u\geq 0} + (1-\tau)|u|\mathbf{1}_{u<0}$ with $\mathbf{1}$ denoting the indicator function.

As the Box-Cox quantile regression is a special case of the general nonlinear quantile regression, one can simply use the `nlrq` procedure in the `quantreg` package in `R` to perform the model estimation. See Koenker (2005, Appendix A) for a vignette on quantile regression in `R`, in particular nonlinear quantile regression. However, unlike its linear counter part (the `rq` procedure), the `nlrq` procedure does not directly provide a covariance estimate, which is necessary for constructing the confidence interval for y_τ, the τ-quantile of y_0 at x_0. Following is an outline for the covariance estimation.

Let $\psi(\tau) = (\beta(\tau)', \lambda(\tau))'$, and $\hat{\psi}(\tau)$ be the estimate of $\psi(\tau)$ defined above. Let $\dot{y}_{i\tau}$ be the derivative of $y_{i\tau}$ with respect to ψ. Under regularity conditions (e.g., Koenker, 2005, Sec. 4.4), $\hat{\psi}(\tau)$ is a consistent estimate of $\psi(\tau)$, and is asymptotically normal, i.e.,

$$\sqrt{n}(\hat{\psi}(\tau) - \psi(\tau)) \xrightarrow{D} N(0, \tau(1-\tau)D_1^{-1}D_0D_1)$$

where $D_0 = \lim_{n\to\infty}\sum \dot{y}_{i\tau}\dot{y}_{i\tau}'$, and $D_1 = \lim_{n\to\infty}\sum f_i(y_{i\tau})\dot{y}_{i\tau}\dot{y}_{i\tau}'$, and f_i denotes the density function of the error corresponding to the ith observation evaluated at $y_{i\tau}$. As the explicit expression for $\dot{y}_{i\tau}$ is available (see Appendix B), an estimate of $\dot{y}_{i\tau}$ can be obtained by simply replacing $\psi(\tau)$ by $\hat{\psi}(\tau)$, which gives a consistent estimate of D_0. What is left is a consistent estimate for $f_i(y_{i\tau})$. We adopt, as did Machado and Mata (2000), the sandwich estimator suggested by Hendricks and Koenker (1992) for linear quantile regression, i.e., replacing the linear function $x_i'\hat{\beta}(\tau)$ everywhere by $(1+\hat{\lambda}(\tau)x_i'\hat{\beta}(\tau))^{1/\hat{\lambda}(\tau)}$. See Koenker (2005, Sec. 3.4.2) for a description of the linear case. Finally, the confidence interval for $y_\tau = (1+\lambda(\tau)x_0'\beta(\tau))^{1/\lambda(\tau)}$ can be obtained through an application of the delta method. Alternatively, one may use the resampling method of Tian and Wei (2002) to estimate the covariance, which is computationally much more demanding, in particular, in conducting Monte Carlo simulations, than the sandwich method.

ACKNOWLEGEMENTS

The authors are grateful to the Editor, Torben Andersen, an Associate Editor and two anonymous referees for their constructive comments that have led to significant improvements in the paper. We are also grateful to Roger Koenker for his help on the use of the `QUANTREG` package, and to Lu Tian and Lee-Jen Wei for sharing with us their working paper.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis.* John Wiley & Sons

Atkinson, A. C. (1985). *Plots, Transformation and Regression.* Oxford University Press.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, **76**, 296–311.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation (with discussion). *Journal of the Royal Statistical Society B*, **26**, 211–252.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963–1987: application of quantile regression. *Econometrica*, **62**, 405–458.

Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987. *Journal of Econometrics*, **65**, 109-154.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, **13**, 1–30.

Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, **83**, 123–127.

Carroll, R. J. and Ruppert, D. (1981). On prediction and power transformation family. *Biometrika*, **68**, 609–615.

Carroll, R. J. and Ruppert, D. (1991). Prediction and tolerance intervals with transformation and/or weighting. *Technometrics*, **33**, 197–210.

Chamberlain, G. (1994). Quantile regression, censoring and the structure of wages. In *Advances in Econometrics*, Sims C. (eds), Cambridge University Press, 171-209.

Collins, S. (1991). Prediction Techniques for Box-Cox Regression Models. *Journal of Business and Economic Statistics*, **9**, 267–277.

Hahn, G. J. and Meeker, W.Q. (1991). *Statistical Intervals, A Guide for Practitioners.* John Wiley & Sons.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, **44**, 461–465.

Hendricks, W. and Koenker, R. (1992). Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. *Journal of the American Statistical Association*, **87**, 58-68.

John, J. A. and Draper, N. A. (1980). An alternative family of transformations. *Applied Statistics*, **29**, 190–197.

Kocherginsky, M., He, X. and Mu, Y. (2005). Practical Confidence Intervals for Regression Quantiles. *Journal of Computational and Graphical Statistics*, **14**, 41–55.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Machado, J. A. F. and Mata, J. (2000). Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics*, **15**, 253–274.

Machado, J. A. F. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, **20**, 445–465.

Manly, B. F. (1976). Exponential data transformation. *The Statistician*, **25**, 37–42.

Maitra, P. and Vahid, F. (2006). The effect of household characteristics on living standards in South Africa 1993 – 98: a quantile regression analysis with sample attrition. *Journal of Applied Econometrics*, forthcoming.

Powell, J. (1991). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium on Economic Theory and Econometrics*, Barnett W., Powell J., Tauchen G. (eds), Cambridge University Press: New York; 357-384.

Tian, L. and Wei, L. J. (2002). Heteroscedastic Quantile Regression for Parametrically Transformed Response. Manuscript, Department of Biostatistics, Harvard University.

Yang, Z. (2002). Median estimation through a regression transformation. *Canadian Journal of Statistics*, **30**, 235–242.

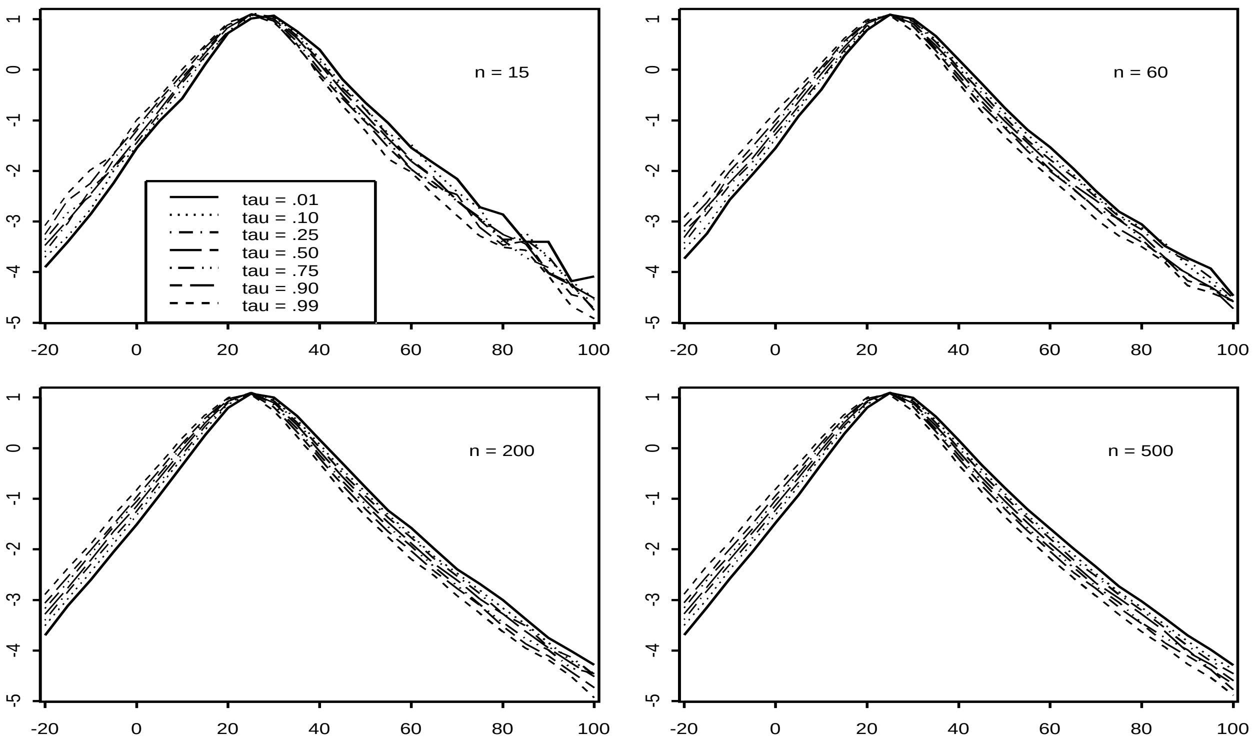
Yang, Z. (2006). A modified family of power transformations. *Economics Letters*, **92**, 14–19.

Yang, Z., Tse, Y. K. and Bai, Z. D. (2006). Statistics with estimated parameters. *Statistica Sinica*, forthcoming.

Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformation to improve normality or symmetry. *Biometrika*, **87**, 954–959.

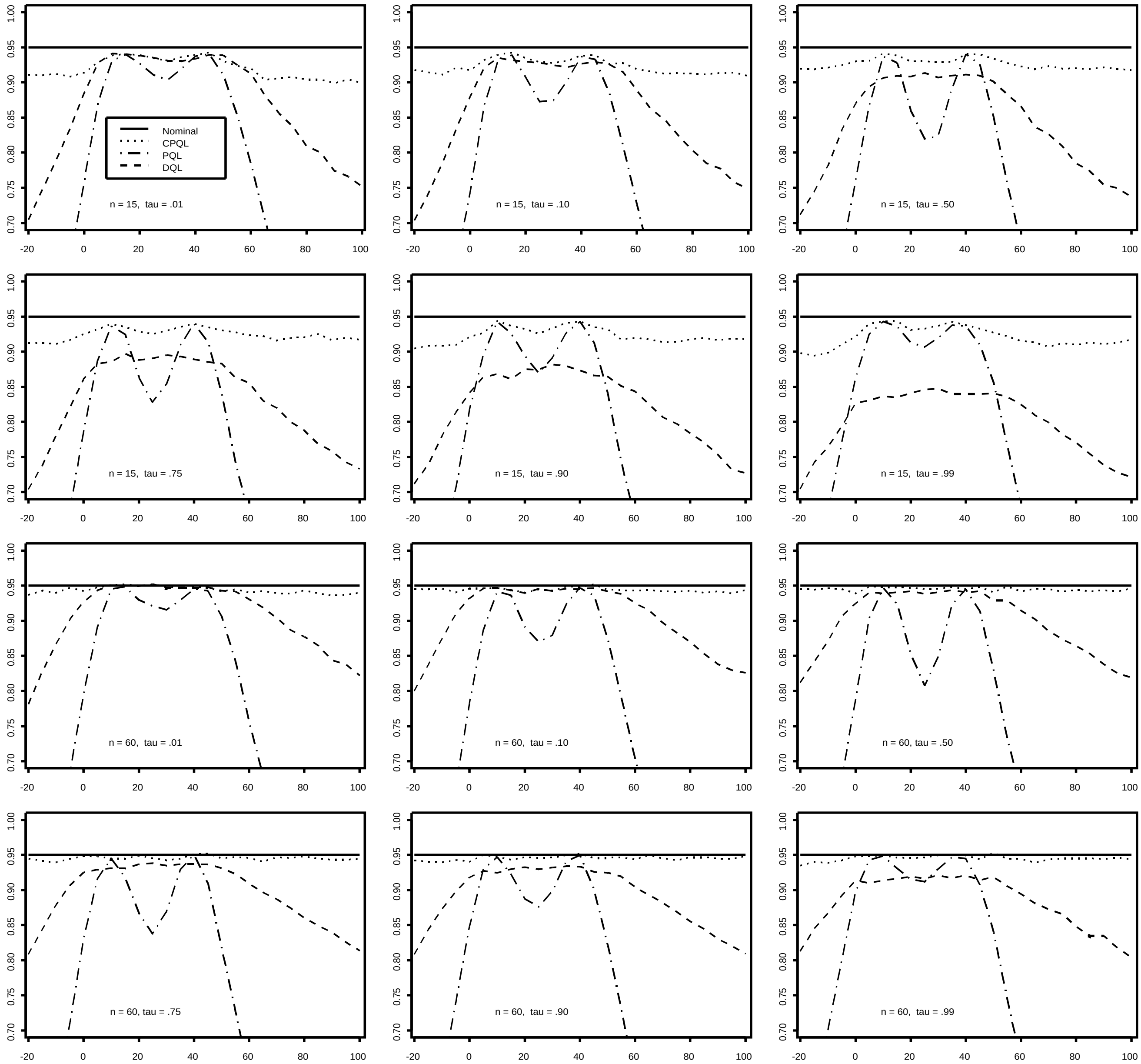
Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 1: Plots of $c(\psi)$ versus x_0: Model 1 with BCT, $X_i \in [0, 50]$, $\lambda = 0.01$, $\sigma = 1$



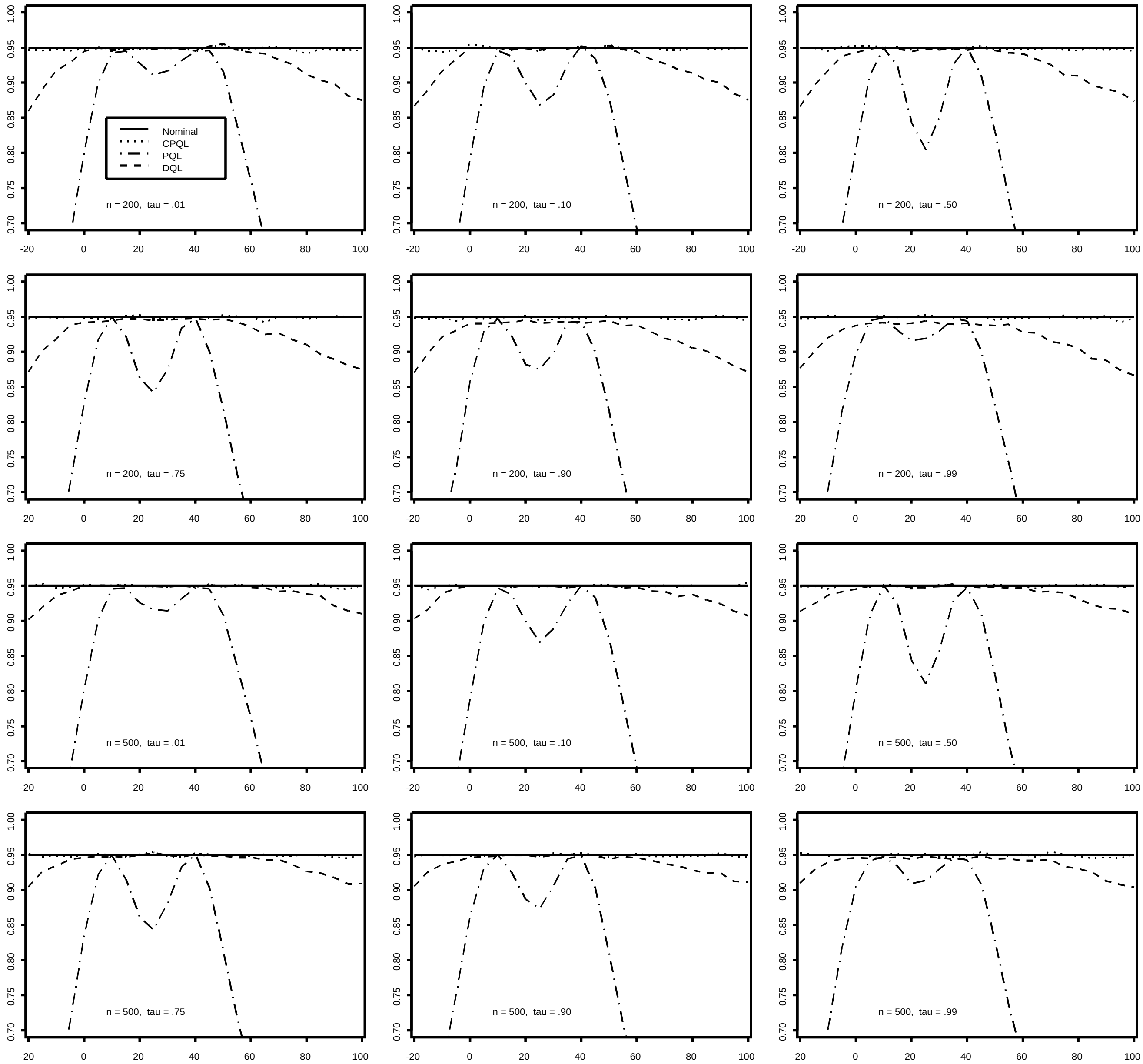
Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 2: Coverage Probability vs x_0: Model 1 with BCT, $X_i \in [0, 50]$, $\lambda = .01$, $\sigma = 1$



Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 2: Cont'd



Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 3: Coverage Probability vs x_0: Model 1 with DPT, $X_i \in [0, 50]$, $\lambda = .1$, $\sigma = 2$

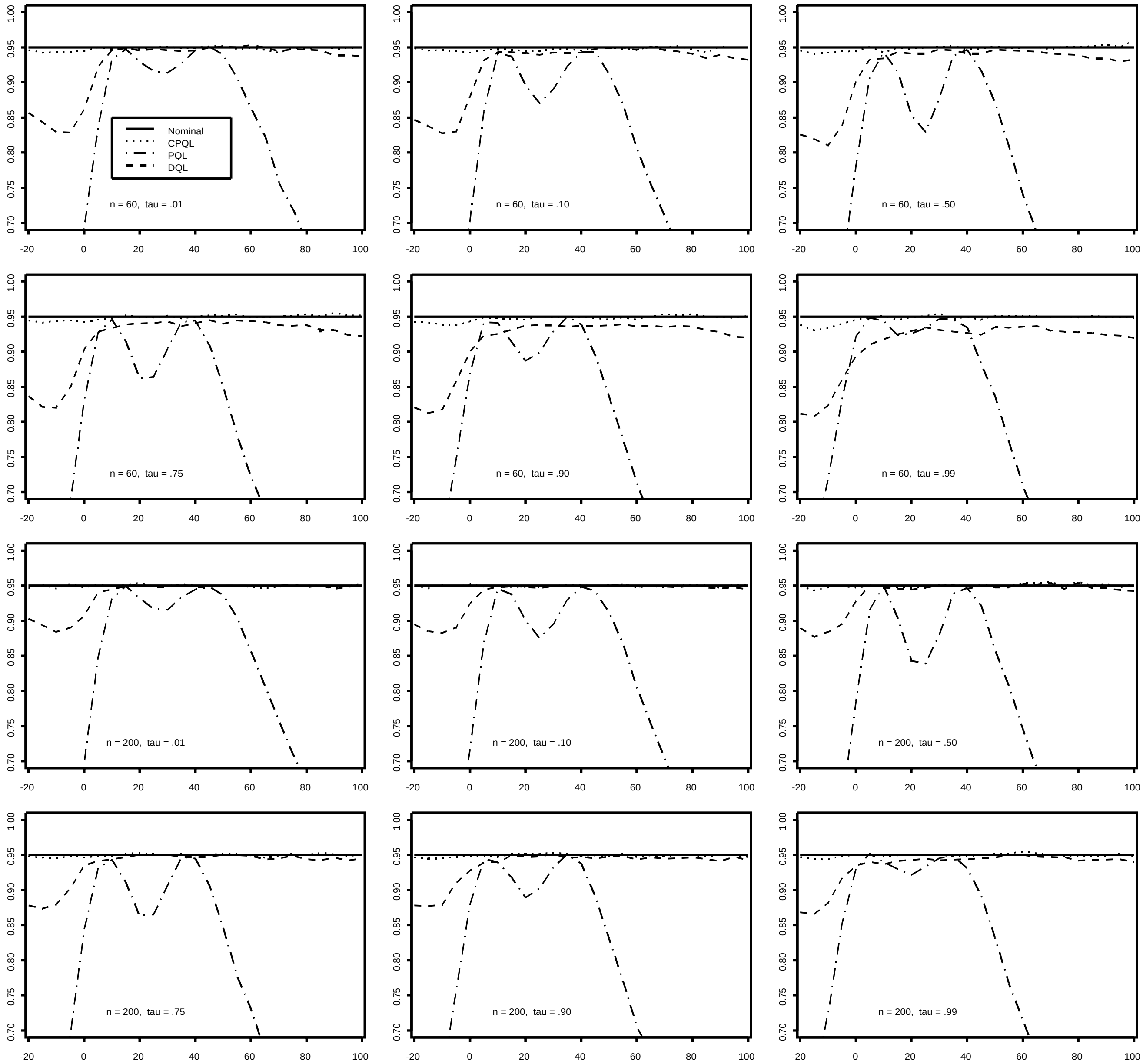


Figure 3: Continued with $\lambda = .1$ (first two rows), and $\lambda = .01$ (last two rows)

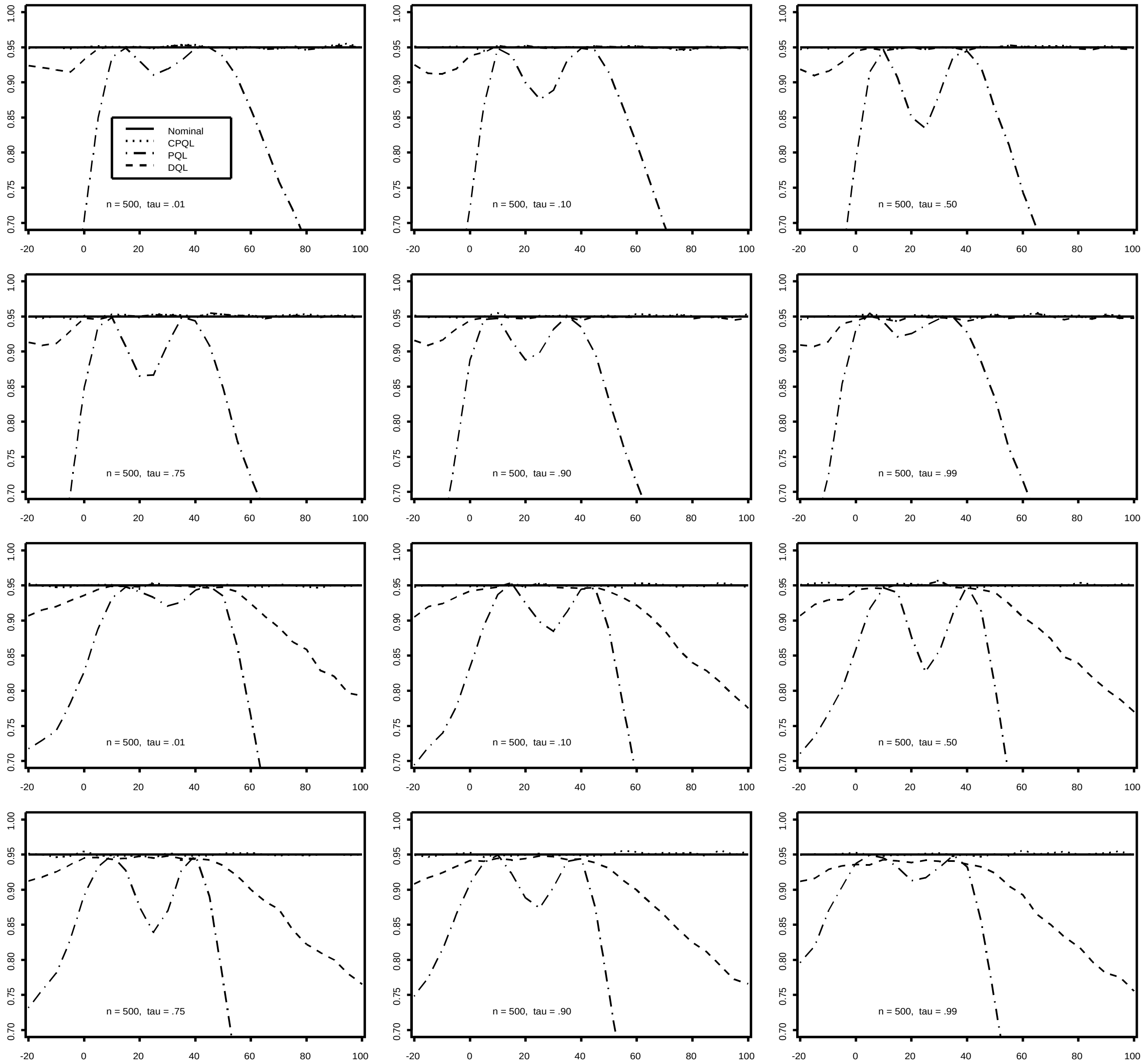


Figure 4: Coverage Probability of CPQL: Model 2 with BCT and Normal-Mixture Error

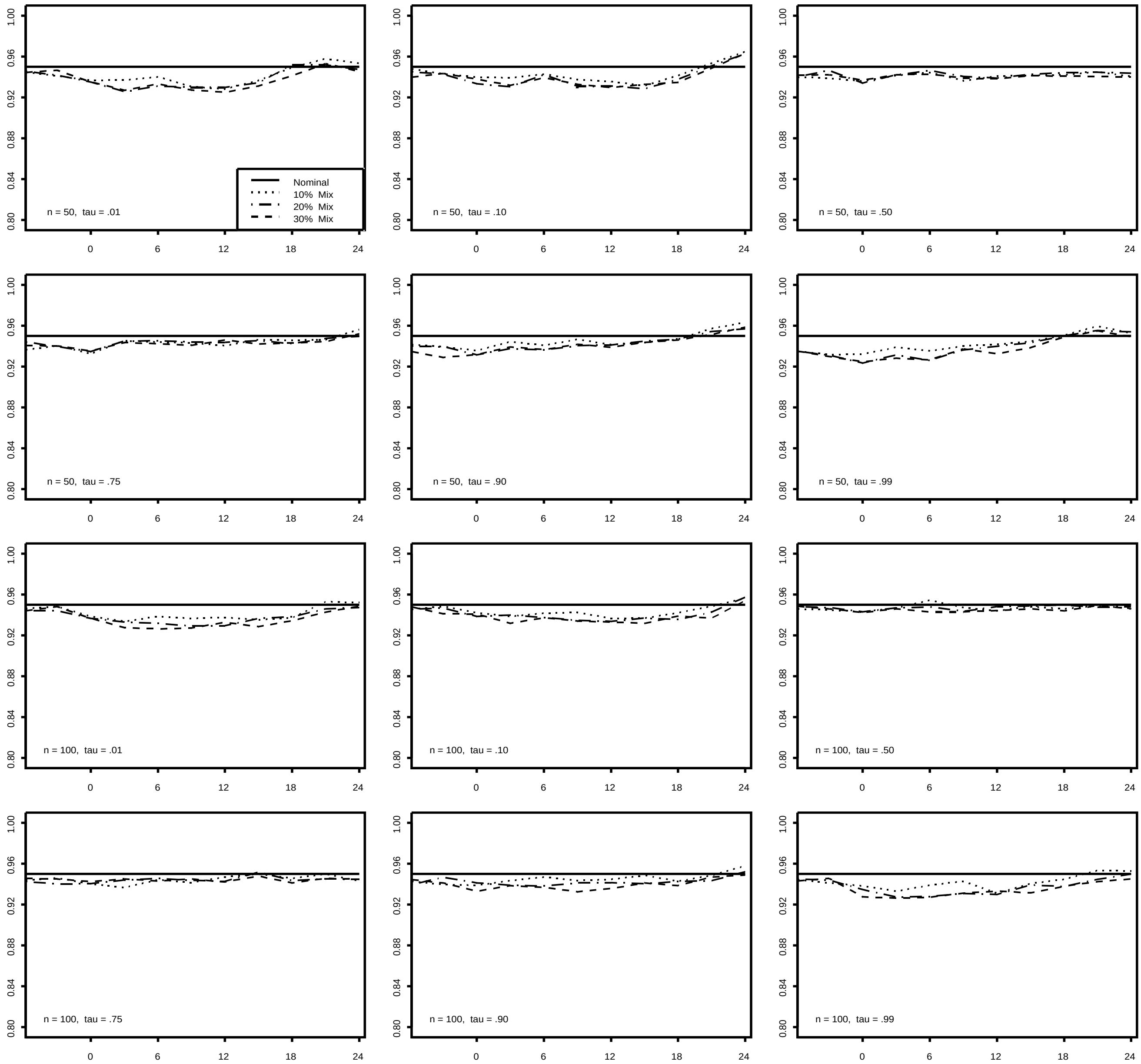


Figure 5: Coverage Probability of CPQL: Model 2 with BCT and t_ν Errors

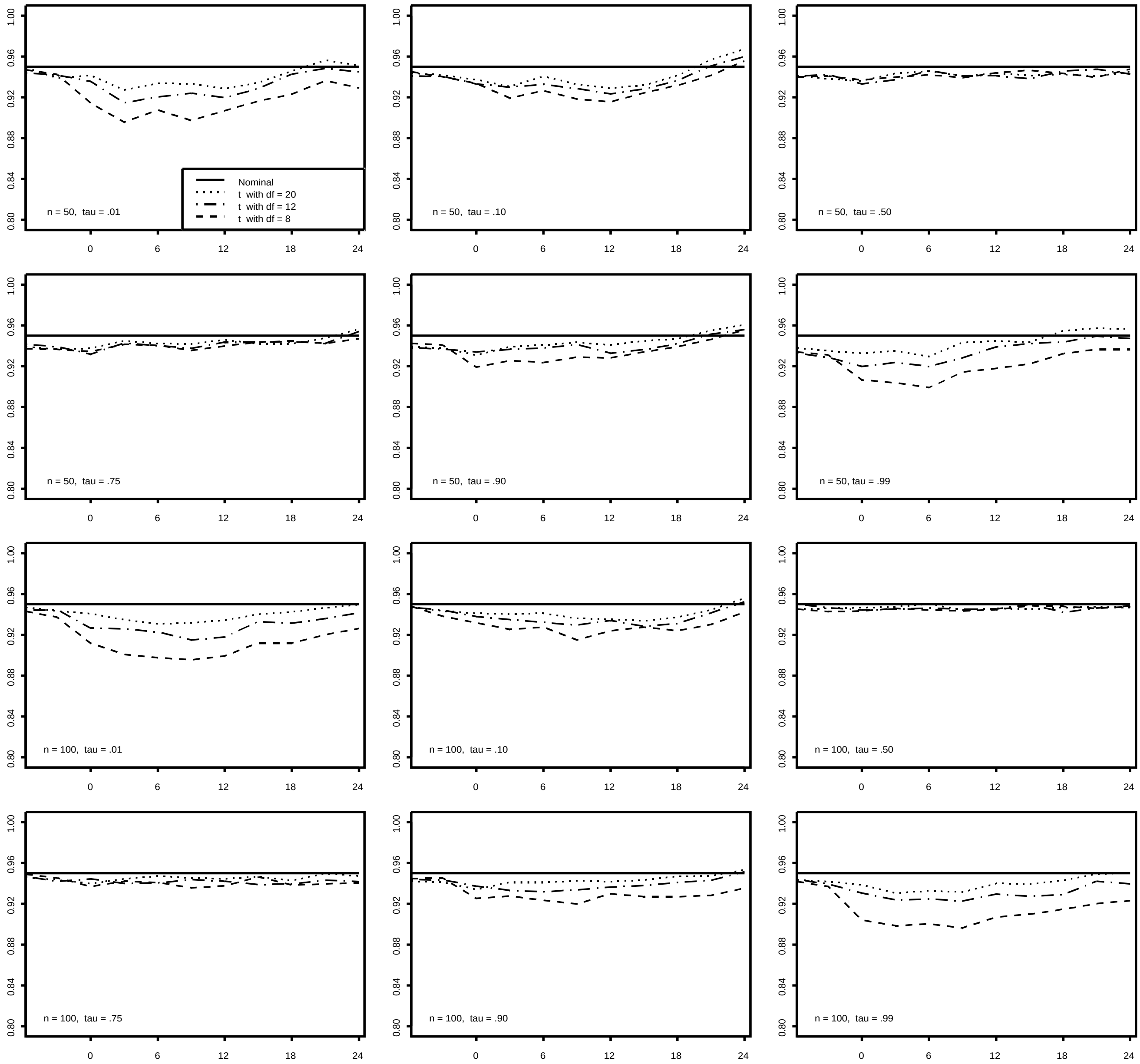
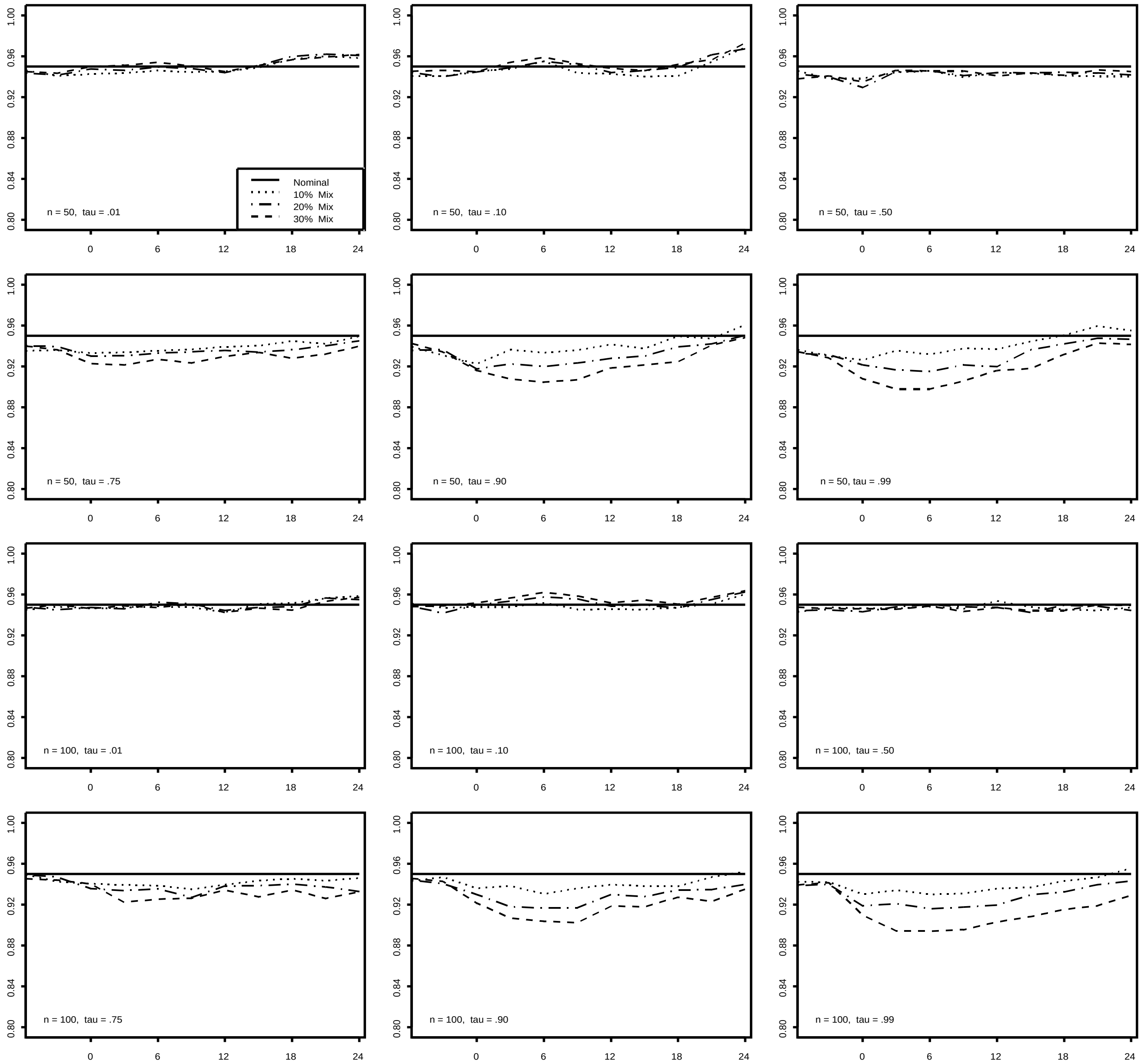


Figure 6: Coverage Probability of CPQL: Model 2 with BCT and Normal-Gamma(1,2) Mixture



Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 6: Continued with Normal-Gamma(1,1) Mixture

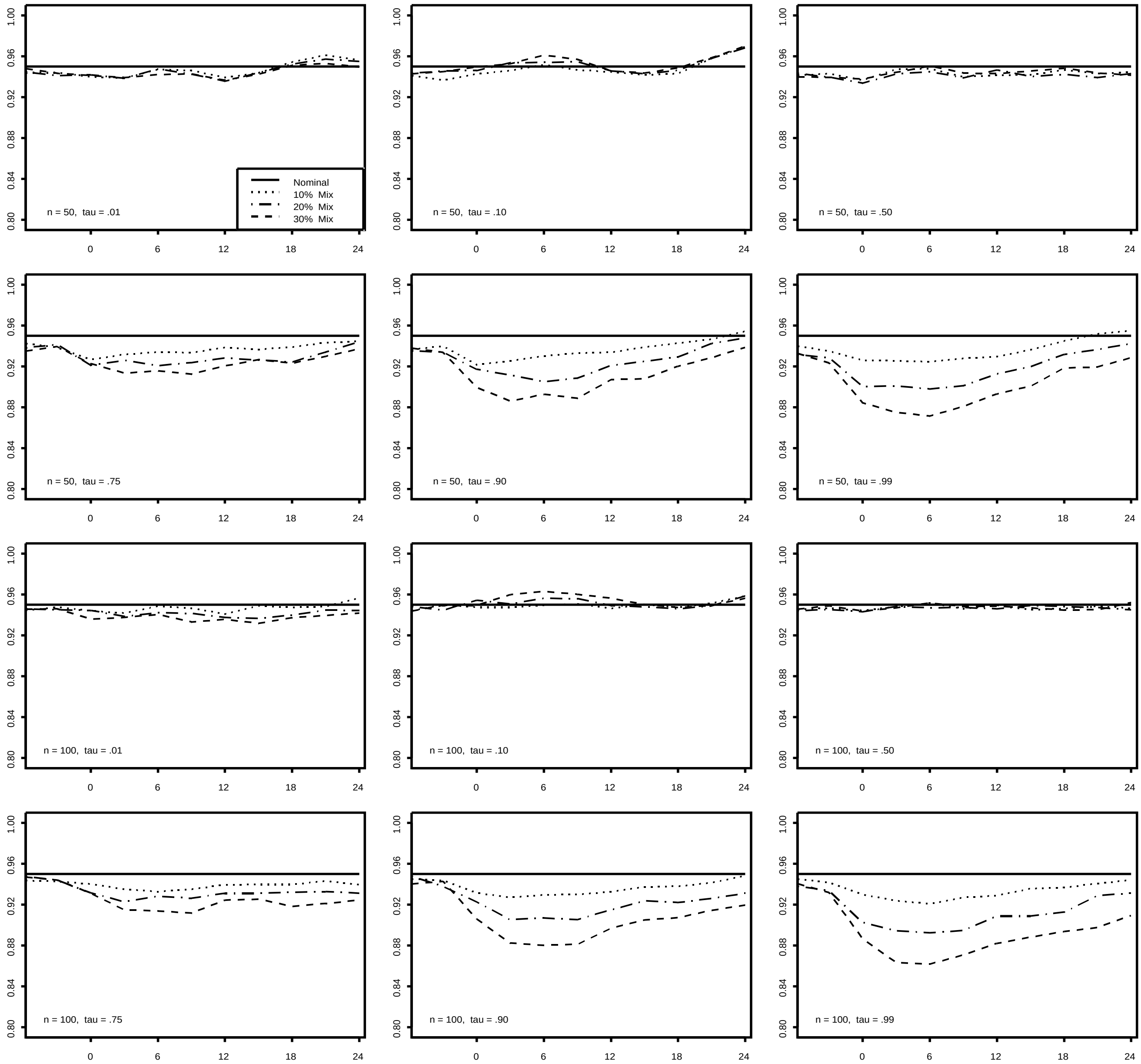


Table 1. Coverage Probability and Ratio (R) of the Average Length of QRQL to that of CPQL: Model 2, Box-Cox Transformation with iid normal errors.

p	x_0	$n=200$			$n=500$			$n=1000$		
		CPQL	QRQL	R	CPQL	QRQL	R	CPQL	QRQL	R
0.10	0	0.948	0.854	5.88	0.947	0.851	4.28	0.950	0.878	4.16
	3	0.950	0.894	2.42	0.948	0.924	1.75	0.946	0.954	1.71
	6	0.947	0.820	1.95	0.948	0.919	1.95	0.954	0.955	2.12
	9	0.950	0.804	1.68	0.946	0.884	1.43	0.953	0.951	1.54
	12	0.944	0.762	2.82	0.945	0.895	2.00	0.954	0.934	2.14
	15	0.944	0.794	2.89	0.950	0.898	2.42	0.947	0.931	2.44
0.25	0	0.946	0.864	3.77	0.944	0.844	3.42	0.953	0.878	3.36
	3	0.943	0.949	1.51	0.947	0.944	1.51	0.952	0.967	1.46
	6	0.950	0.902	1.81	0.947	0.914	1.83	0.950	0.944	1.83
	9	0.949	0.834	1.34	0.951	0.909	1.37	0.951	0.921	1.39
	12	0.948	0.855	1.87	0.947	0.915	1.89	0.947	0.922	1.87
	15	0.947	0.854	2.06	0.948	0.930	2.09	0.949	0.931	2.08
0.50	0	0.944	0.839	3.50	0.950	0.849	3.27	0.953	0.855	3.20
	3	0.949	0.950	1.46	0.951	0.963	1.49	0.952	0.945	1.42
	6	0.951	0.890	1.76	0.947	0.921	1.80	0.946	0.933	1.81
	9	0.943	0.841	1.33	0.951	0.913	1.38	0.946	0.936	1.38
	12	0.952	0.854	1.85	0.946	0.928	1.85	0.946	0.934	1.85
	15	0.945	0.866	2.09	0.950	0.895	2.04	0.951	0.933	2.03
0.75	0	0.942	0.844	3.89	0.948	0.873	3.65	0.954	0.859	3.45
	3	0.949	0.939	1.56	0.949	0.948	1.50	0.949	0.968	1.47
	6	0.948	0.879	1.82	0.949	0.929	1.86	0.954	0.945	1.87
	9	0.946	0.860	1.37	0.947	0.917	1.39	0.946	0.925	1.38
	12	0.948	0.867	1.92	0.952	0.906	1.88	0.947	0.928	1.92
	15	0.948	0.862	2.11	0.946	0.926	2.16	0.948	0.914	2.14
0.90	0	0.947	0.863	5.36	0.951	0.868	4.58	0.949	0.871	4.29
	3	0.950	0.907	1.95	0.952	0.955	1.82	0.950	0.962	1.70
	6	0.952	0.867	2.00	0.948	0.948	2.17	0.950	0.965	2.17
	9	0.955	0.798	1.38	0.948	0.929	1.54	0.946	0.952	1.57
	12	0.944	0.801	2.08	0.948	0.911	2.27	0.953	0.926	2.25
	15	0.949	0.831	2.69	0.953	0.925	2.60	0.951	0.928	2.56

Note: all results are based on $(\beta_0, \beta_1, \sigma, \gamma, \lambda) = (5, 1, 0.1, 0.1, 0.1)$, and $X_i \in [0, 10]$

Table 2. Coverage Probability and Ratio (R) of the Average Length of QRQL to that of CPQL: Model 2, Box-Cox Transformation with iid normal-gamma(1,2) mixtures.

p	x_0	$n=200$ CPQL	QRQL	R	$n=500$ CPQL	QRQL	R	$n=1000$ CPQL	QRQL	R
		$e_i \sim 0.85N(0,1)+0.15\text{Gamma}(1,2)$								
0.25	0	0.949	0.852	3.38	0.954	0.838	3.12	0.952	0.848	3.12
	3	0.952	0.921	1.41	0.956	0.907	1.37	0.955	0.911	1.33
	6	0.956	0.894	1.67	0.954	0.879	1.68	0.957	0.891	1.70
	9	0.956	0.882	1.25	0.957	0.915	1.27	0.956	0.934	1.28
	12	0.949	0.886	1.71	0.955	0.900	1.71	0.955	0.900	1.71
	15	0.949	0.903	1.93	0.953	0.887	1.88	0.953	0.871	1.88
0.50	0	0.946	0.849	3.41	0.949	0.848	3.17	0.952	0.894	3.25
	3	0.949	0.954	1.47	0.949	0.951	1.46	0.947	0.947	1.43
	6	0.949	0.928	1.81	0.952	0.946	1.81	0.948	0.926	1.83
	9	0.947	0.875	1.36	0.948	0.910	1.36	0.947	0.937	1.39
	12	0.943	0.871	1.82	0.950	0.906	1.81	0.948	0.928	1.87
	15	0.947	0.894	2.05	0.950	0.919	2.05	0.953	0.928	2.03
0.75	0	0.937	0.844	3.96	0.943	0.863	3.62	0.946	0.888	5.04
	3	0.937	0.942	2.01	0.934	0.948	2.49	0.941	0.954	2.03
	6	0.938	0.926	1.89	0.933	0.922	2.20	0.935	0.947	2.04
	9	0.929	0.861	1.41	0.932	0.922	1.45	0.932	0.929	1.44
	12	0.942	0.866	2.09	0.941	0.913	1.96	0.940	0.934	2.04
	15	0.942	0.895	2.28	0.943	0.906	2.28	0.938	0.921	2.30
		$e_i \sim t_{12}$								
0.25	0	0.943	0.876	10.03	0.945	0.863	6.46	0.950	0.842	9.30
	3	0.944	0.924	5.29	0.943	0.908	3.96	0.948	0.881	2.74
	6	0.947	0.862	2.11	0.944	0.837	1.87	0.947	0.822	1.84
	9	0.939	0.865	1.30	0.949	0.932	1.42	0.945	0.943	1.42
	12	0.943	0.861	2.20	0.948	0.895	2.14	0.940	0.902	2.32
	15	0.944	0.858	2.63	0.944	0.891	2.35	0.945	0.884	2.78
0.50	0	0.946	0.866	3.39	0.948	0.894	3.13	0.949	0.883	3.12
	3	0.952	0.949	1.35	0.948	0.956	1.33	0.948	0.953	1.31
	6	0.948	0.920	1.65	0.946	0.944	1.69	0.949	0.937	1.71
	9	0.948	0.853	1.26	0.946	0.919	1.30	0.953	0.935	1.31
	12	0.951	0.877	1.71	0.948	0.919	1.74	0.951	0.945	1.70
	15	0.948	0.880	1.94	0.951	0.912	1.92	0.950	0.931	1.92
0.75	0	0.943	0.860	5.68	0.949	0.856	3.76	0.942	0.873	3.42
	3	0.946	0.941	4.03	0.944	0.950	1.61	0.948	0.956	2.47
	6	0.947	0.891	1.75	0.947	0.921	2.05	0.940	0.937	1.92
	9	0.946	0.857	1.33	0.940	0.917	1.35	0.947	0.933	1.35
	12	0.945	0.854	1.90	0.945	0.913	1.88	0.947	0.925	1.84
	15	0.943	0.869	2.09	0.943	0.917	2.29	0.948	0.908	2.18

Note: all results are based on $(\beta_0, \beta_1, \sigma, \gamma, \lambda) = (5, 1, 0.1, 0.1, 0.1)$, and $X_i \in [0, 10]$

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Table 3. Box-Cox Heteroscedastic Regression of Living Standards in South Africa - 1993 Data

Variable	Estimate	Std. Error	Variable	Estimate	Std. Error
CONSTANT	8.71851***	0.6696	ATTRITE	1.38437*	0.7307
AGEHD	0.00040	0.0095	A-AGEHD	-0.03606	0.0267
AGEHD2	-0.00001	0.0001	A-AGEHD2	0.00025	0.0003
FHH	-0.18628***	0.0522	A-FHH	-0.21057	0.1283
HDEDUC1	0.17760***	0.0567	A-HDEDUC1	-0.18016	0.1554
HDEDUC2	0.43306***	0.0814	A-HDEDUC2	-0.54390***	0.2002
HDEDUC3	1.32563***	0.2130	A-HDEDUC3	-0.49665	0.3341
BLACK	-1.56720***	0.2022	A-BLACK	0.25899	0.2239
NATAL	-0.78227***	0.1052	A-NATAL	0.03956	0.1589
RURAL	-0.46405***	0.0719	A-RURAL	-0.22907	0.1637
TOTCHILD	-0.11958***	0.0149	A-TOTCHILD	0.01052	0.0284
TOTADULT	-0.07465***	0.0140	A-TOTADULT	-0.00515	0.0322
TOTELDER	0.00746	0.0474	A-TOTELDER	-0.28129**	0.1354
σ	0.95480***	0.2710	Likelihood Ratio Test for		
λ	0.06840***	0.0196	ATTRITE effect = 26.4332***		
γ_AGE	-0.00240	0.0018			
γ_TOTCHLD	-0.02100**	0.0090	*** Significant at 1%		
γ_TOTADULT	-0.03190***	0.0105	** Significant at 5%		
γ_TOTELDER	0.04730	0.0447	* Significant at 10%		

Appeared in: Journal of Business and Economic Statistics, 2007, 25, 356-376.

Figure 7: Plots of Confidence Intervals based on Household Expenditure Data