

# Quantile regression model coefficient CIs for linear expanding horn heteroscedastic cases using auxiliary regression and the empirical variance distribution

*John P. D. Martin*

*Sunday, February 7, 2016*

## Executive Summary

This paper investigates the application of the empirical variance distribution (evd) function (Martin (1-3)), to estimate the confidence interval bounds of the quantile regression model coefficient estimates (Koencker & Bassett (4)), for linear expanding horn heteroscedastic unweighted data cases. The heteroscedasticity is analysed and transformed by auxiliary regression to provide estimates of the slope variance and regression formula components of the intercept variance previously trialled for the homoscedastic case Martin (3).

Analysis of the coverage accuracy of the results is conducted by repeated sampling of several sample sizes to known regression models and error distributions. Importantly, the sampling also included different positions of the intercept in the known distributions, since the intercept variance changes in the heteroscedastic case depending on the position of the intercept within the observed data range. Table 1 displays the coverage performance of several quantile regression variance estimates of unweighted regression examples, for sample size 1000. The empirical variance distribution (evd) based coverage estimates are compared to default bootstrap quantile coverage calculated using 600 replicates (from the quantreg r package (5)).

**Table 1: Quantile regression confidence interval estimator coverage of sample size 1000 for (slope,intercept) respectively, based on 1000 repeated samples**

Regression $Y = \beta f(X) + x\varepsilon$	quantile	bootstrap coverage	bootstrap half CI	evd_aux reg coverage
Intercept at median(x)				
$y = 0.5 * x + x * rnorm(0, 1.5)$	0.1	(0.935,0.931)	(0.0089,1.0753)	(0.963,0.931)
“”	0.5	(0.935,0.937)	(0.0065,0.7866)	(0.956,0.921)
“”	0.9	(0.955,0.940)	(0.0089,1.0771)	(0.971,0.938)
Intercept at weighted median(x) using auxiliary regression model as weights				
$y = 0.5 * x + x * rnorm(0, 1.5)$	0.1	(0.935,0.948)	(0.0089,0.6851)	(0.963,0.958)
“”	0.5	(0.935,0.943)	(0.0065,0.5014)	(0.956,0.958)
“”	0.9	(0.955,0.954)	(0.0089,0.6928)	(0.971,0.965)
Intercept at median( $e^x$ )				
$y = 0.5 * e^x + x * rnorm(0, 1.5)$	0.1	(0.937,0.956)	(0.0111,8.0037)	(0.979,0.969)
“”	0.5	(0.955,0.951)	(0.0081,5.7855)	(0.977,0.950)
“”	0.9	(0.936,0.950)	(0.0111,8.0106)	(0.971,0.938)
Intercept at weighted median( $e^x$ ) using auxiliary regression model as weights				
$y = 0.5 * e^x + x * rnorm(0, 1.5)$	0.1	(0.937,0.962)	(0.0111,6.6743)	(0.979,0.971)
“”	0.5	(0.955,0.951)	(0.0081,4.7947)	(0.977,0.950)
“”	0.9	(0.936,0.946)	(0.0111,6.6838)	(0.971,0.956)

In Table 1, the regression degrees of freedom model adjustment  $\sqrt{n/(n-p-1)}$  for the evd based estimators was performed in the percentile scale.

It can be seen that for this modest sample size, the bootstrap estimator and the evd based confidence interval estimators exhibit nominal 95% coverage for univariate linear expanding horn heteroscedastic, independent error cases in slope estimates. The evd based intercept CI estimates have good coverage for intercepts in the middle of distributions.

For extreme quantiles in smaller samples the evd based slope CIs coverage performance is lower. For intercepts on the edge of distributions the evd based intercept CIs coverage performance are weaker. The good performance for evd based estimators is expected to be sensitive to how well the quantile regression is specified compared to the data. For Table 1 results, all the quantile regression fits are linear matching the specified distributions, however, if the expanding horn heteroscedasticity varies as  $x^2$  then the best quantile regression fitting is a second order polynomial model. This sensitivity to model specification for evd based estimators for heteroscedastic cases is part of the reason for the lower coverage with smaller sample sizes. The quantreg (5) default bootstrap estimator coverage is robust for smaller samples with linear or quadratic expanding horn heteroscedascity.

The first and second examples illustrate the same uniformly distributed x distribution with linear expanding horn heteroscedasticity for two different intercept values. The third and fourth examples illustrate the same exponential normal distribution (similar to the Engel food expenditure dataset behaviour) with expanding horn heteroscedasticity for two different intercept values. Examining the bootstrap intercept intervals for the four examples shown in Table 1, as well as other intercept values, the minimum intercept confidence interval length occurs near the weighted median value of the x distribution when weighted by the auxiliary regression model of the heteroscedasticity. This numerical bootstrap data on quantile regression intercept variance behaviour has been used to further develop the quantile regression evd estimator intercept confidence interval formula (presented in (3)) to work for the linear expanding horn heteroscedastic case.

## Introduction

Quantiles (6) are an order statistic of a distribution defined by the equivalent probability amount contained under the cumulative distribution function up to the (ordered) value of the quantile point.

That is, x is a k-th q-quantile for a variable X if

$$\Pr[X < x] \leq k/q \text{ or, equivalently, } \Pr[X \geq x] \geq (1 - k/q)$$

So the 25th percentile point is the 25/100 (k/q) 100-quantile point where 25% of the probability under the cumulative density function has occurred.

An equivalent calculation of quantile points has been demonstrated (4) using least absolute deviation (LAD) regression of the following quantile estimation function

$$\min_{b \in \mathbb{R}} \{\theta |x_t - b| + (1 - \theta) |x_t - b|\} \quad (1)$$

where  $\theta \equiv k/q$  and  $x_t$  are the sample/population elements of X. As the absolute value functions in equation 1, create a piecewise linear function shape (convex polytope) to the estimating function, linear programming techniques are required to solve the minimisation problem. As such, closed form expressions for the standard error of the quantile estimates are not available from this approach.

Another approach for estimated standard errors of the quantile estimation function solution is to concurrently calculate the standard errors of smoothed versions of the problem, Brown & Wang (7). Consistent with that approach, Martin (1) identified an analytic quadratic polynomial smoothing function, in the percentile scale, for the quantile estimating function of unweighted samples. This analytic function, only requires the sample size and selected quantile value, to calculate the sample quantile confidence interval (CI) bounds in the percentile scale.

Backtransforming to the original measurement scale, using the CI bounds in the quantile estimation function calculations on the sample distribution results in the empirical variance distribution (evd). As shown in (1), the evd is an asymmetric stepped sample CI, in contrast to smooth symmetric bootstrap sample CIs, but similar in morphology to the discrete cumulative density function (cdf).

In Martin (2), several evd based sample quantile CI estimators were shown to have nominal 95% performance for samples sizes 50-100-1000, except for extreme quantiles in the smallest samples. Some improvement in the sample quantile CI coverage was also shown to be possible for these extreme cases, via use of quantile regression extrapolated 0th,100th quantile sample bounds.

In Martin (3), the evd based sample quantile CI estimators (1,2) were trialled, assessed and adapted where required as quantile regression model coefficient CI estimators. Since the evd approximation to quantile estimating function, produces a smooth, differentiable approximation to the quantile estimating function (1), the evd based CI estimators trialled for homoscedastic iid cases were analogous to the ordinary least squares regression estimators where the median has replaced the mean estimate in the variance intercept formula. For regression with two explanatory variables,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (2)$$

the model slope(s) and intercept CI estimates are derived from empirical variance distribution (evd) approximations of the quantile regression residuals sample variance (distribution)  $s_{res}$  via the relationships

$$s_{\beta_1} = \frac{s_{res}}{\sqrt{var(x_1)(1 - r_{12qr})}} \quad (3)$$

$$s_{\beta_2} = \frac{s_{res}}{\sqrt{var(x_2)(1 - r_{12qr})}} \quad (4)$$

$$s_{\beta_0}(\theta) = \sqrt{s_{res}^2 + s_{\beta_1}^2 median(x_1)^2 + s_{\beta_2}^2 median(x_2)^2} \quad (5)$$

where the following covariance term had been trialled

$$r_{12qr} \sim \frac{cov(x_1, x_2)}{\sqrt{var(x_1)var(x_2)}} \quad (6)$$

In (3), equations (3), (4) & (5) were found to form a suitable basis to produce accurate evd based approximations to quantile regression model coefficient CIs for univariate homoscedastic iid cases, from quantile regression residuals.

**However, further research, for the current paper has revealed that the covariance term  $r_{12qr}$  in equation (6) is not accurate enough for use with bivariate quantile regression, it is sufficient for specific bivariate cases when  $r_{12qr} \rightarrow 0$  but more generalisation of the covariate adjustment is needed.**

Importantly, to keep the evd based variance estimation maximising the use of quantile regression calculations. The residuals distribution variance  $s_{res}$  is not calculated using sum of squares of errors of the residuals but is calculated by performing quantile regression calculations on the residual distribution using evd based estimates of  $(\theta_{LB}, \theta_{UB})$ , to then estimate the 2.5th & 97.5th bounds in the original measurement scale.

In this paper, the evd based quantile regression model coefficient CI estimators (3) have been successfully adapted to estimate CI intervals for linear expanding horn heteroscedastic, independent error cases. The results are then compared to the known population slope and intercept regression values as well as default quantreg (5) bootstrap estimates. To deal with the heteroscedasticity, auxiliary regression of the quantile regression residuals (analogous to generalised least squares regression) has proved essential to transform the

distribution before extracting useful and consistent evd quantile values for single covariate cases (and some limited bivariate cases).

It has been identified for linear expanding horn heteroscedastic cases, that the weighted median of the explanatory variable closely corresponds to the minimum CI of the dependent variable model estimate, for well specified models. The weights are obtained from auxiliary quantile regression fitting of the heteroscedasticity.

### Quantile regression model *slope* coefficient CI estimators for linear expanding horn heteroscedasticity using auxiliary regression and empirical variance distribution

A well specified quantile regression model is one where the density of data points beyond a given fitted quantile regression line is relatively independent of the explanatory variables when repeated sampling occurs from a population distribution. For example, with a linear explanatory variable dependence and linear expanding horn heteroscedasticity, linear quantile regression lines would be expected to occur for all quantile values, as shown in figure 1.

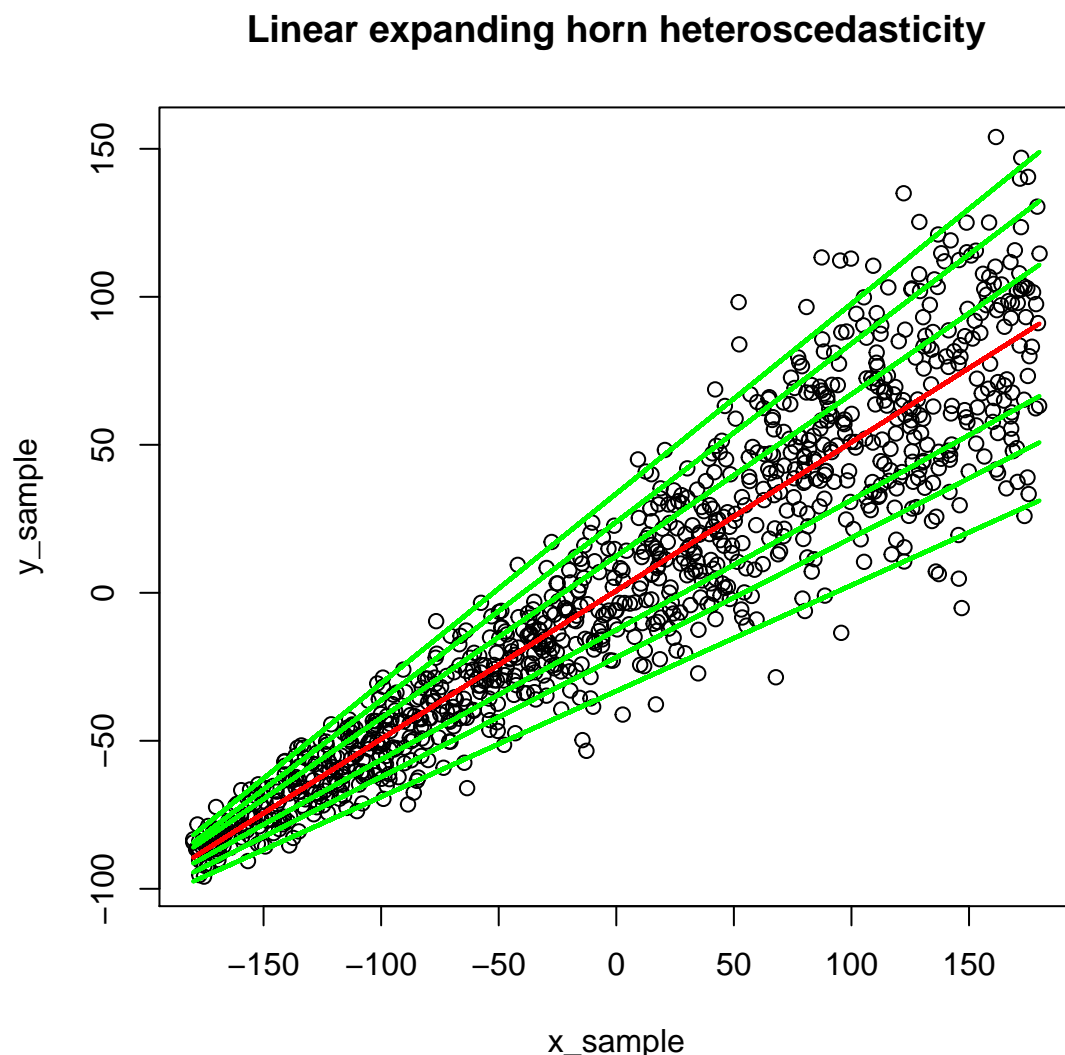


Figure 1: Linear heteroscedasticity with small homoscedastic component

As shown in Figure 2, the quantile regression model residuals exhibit heteroscedasticity. Initially, the evd based model slope coefficient CI estimator for homoscedastic case (3) was trialled explicitly on the heteroscedastically scattered residual distribution but undercoverage of the evd based model slope coefficient CI estimates was clearly evident.

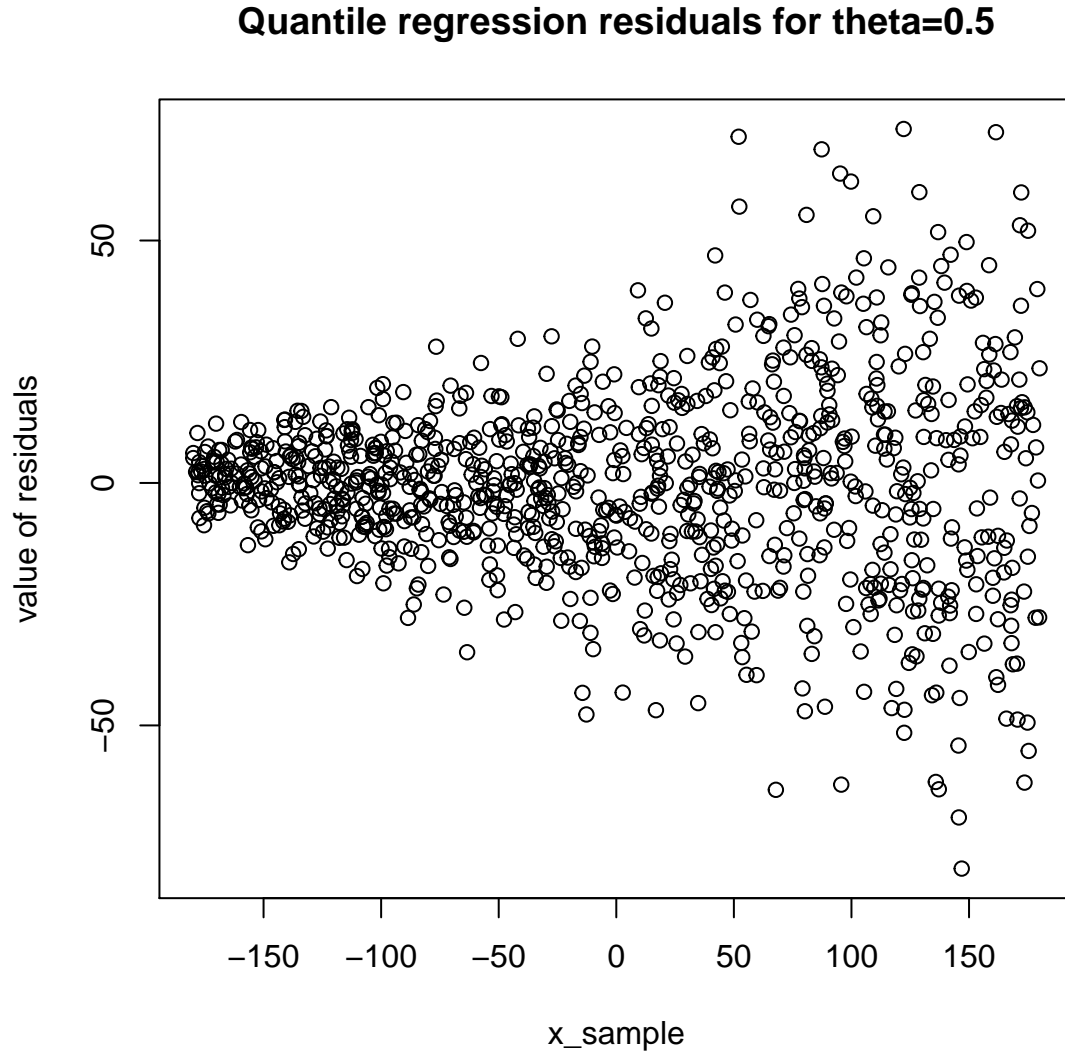


Figure 2: Original residuals distribution

Following the approach of auxiliary regression, the heteroscedascity in the residuals can be modelled using quantile regression applied to the absolute values of the residuals, as shown in Figure 3.

### Auxiliary Quantile regression of absolute values of residuals

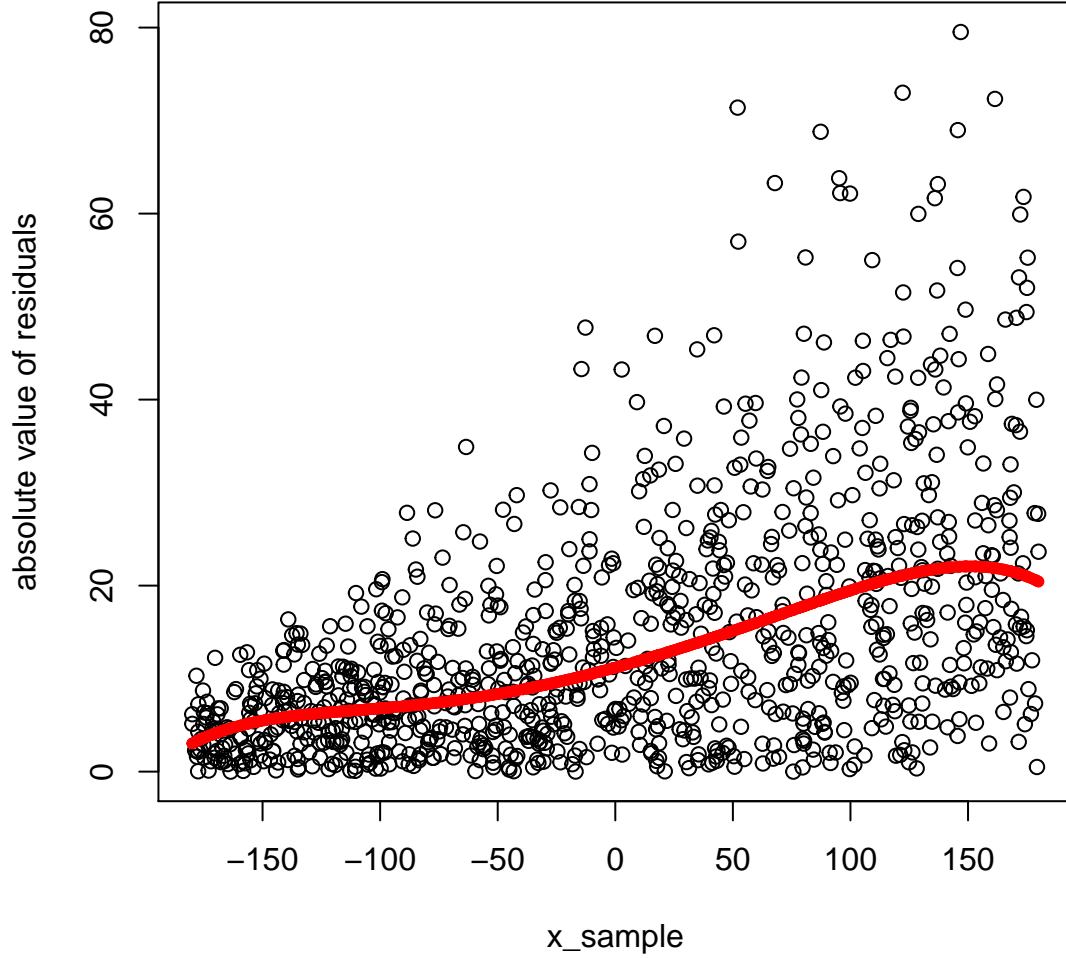


Figure 3: Auxiliary regression of heteroscedasticity in residuals distribution

The auxiliary regression model can be used explicitly to adjust the residuals distribution as shown in Figure 4. As part of the adjustment procedure, the area of the homoscedastic transformed residuals distribution was explicitly adjusted to be equal to the area of the original heteroscedastic residuals distribution for the 2.5th & 97.5 quantile bounds.

$$model_{residualadjust} = \frac{(model_{residual})}{(auxiliary_{modelvalue})} * \frac{\int_{.025}^{.975} model_{residual} dx}{\int_{.025}^{.975} \frac{(model_{residual})}{(auxiliary_{modelvalue})} dx} \quad (7)$$

### Homoscedastically transformed residuals distribution using auxiliary regression

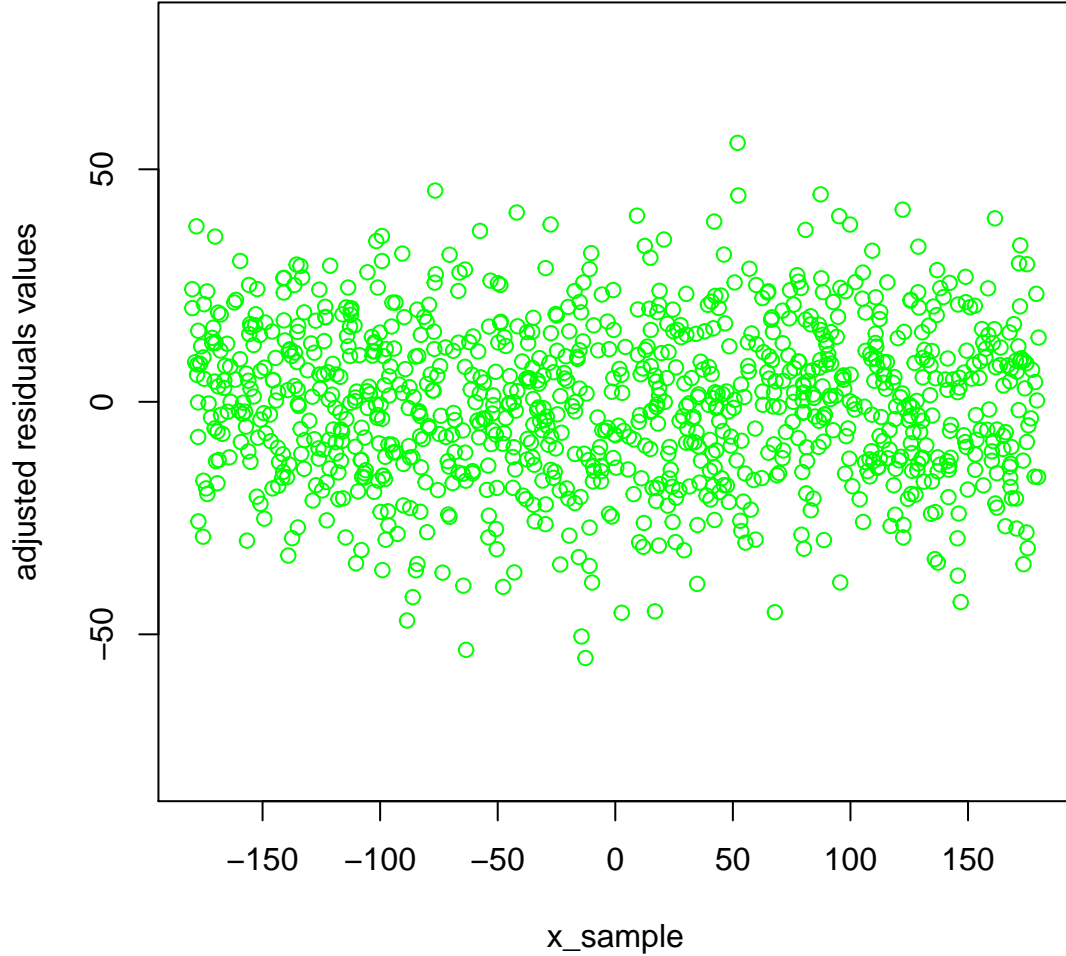


Figure 4: Adjusted residuals for evd based model slope CI estimation

In the examples, to be shown, the evd based model slope coefficients CI estimators based on the auxiliary regression adjusted residuals distribution give good coverage performance.

$$s_{\beta_1} = \frac{s_{\text{residualadjust(maxhalf evdCI)}}}{\text{var}(x)} \quad (8)$$

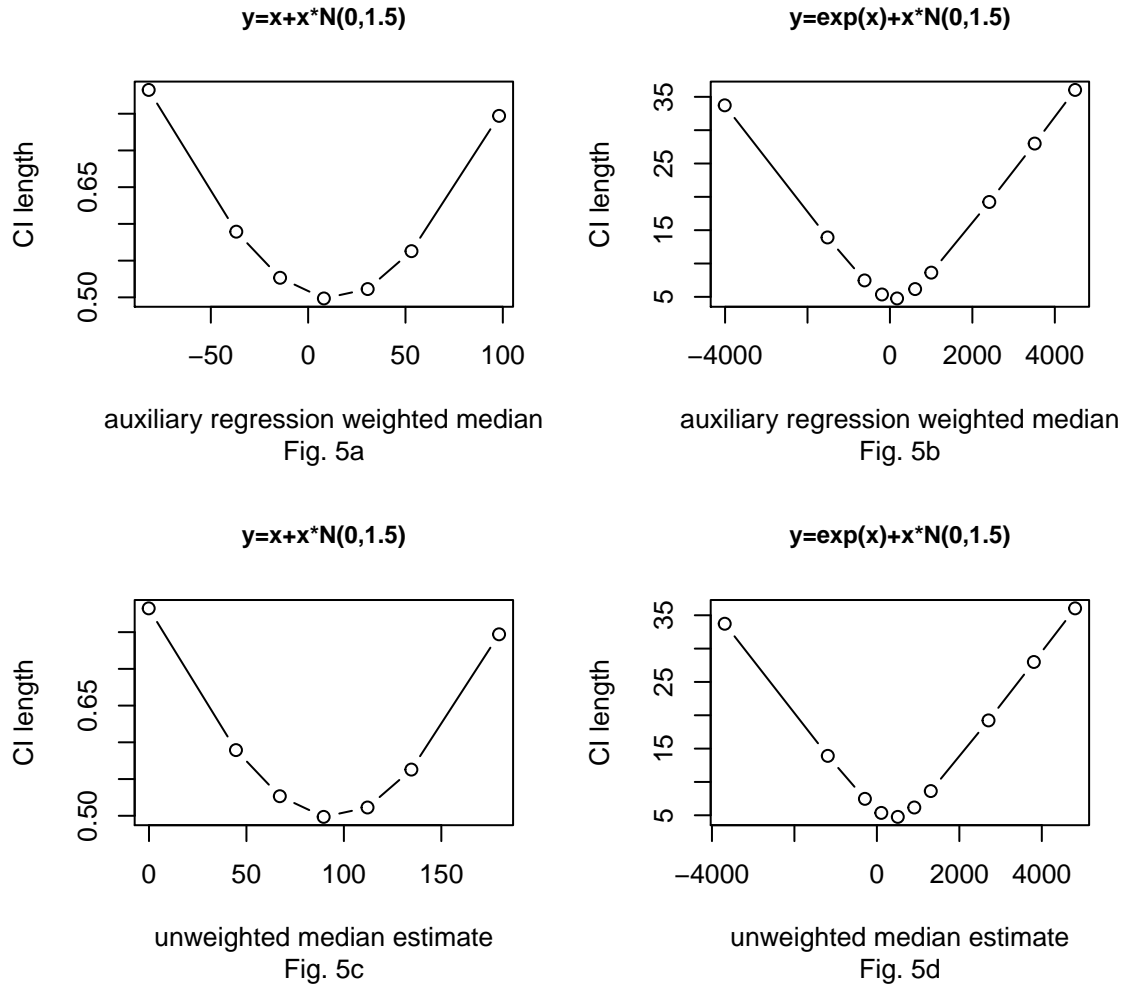
It has been found that using  $\theta = 0.5$  for auxiliary quantile regression modelling of the absolute residuals is satisfactory for linear expanding horn heteroscedasticity, regardless of the quantile value investigated in the quantile regression. Also, consistent with the evd based model slope coefficients CI estimators in (3), it is necessary to take the maximum half CI of the asymmetric evd CIs of the (transformed) residuals distribution as the estimated symmetric confidence interval estimate of the slope coefficient.

## Quantile regression model *intercept* coefficient CI estimators for linear expanding horn heteroscedasticity using auxiliary regression and empirical variance distribution

In investigating possible evd based model intercept coefficient CI estimators, it is instructive to graph the default bootstrap model intercept coefficient CI length as the intercept position within a fixed scatterplot pattern. As shown in figure 5, for two examples, the minimum in the bootstrap model intercept coefficient CI length for linear heteroscedasticity is close to the weighted median of the explanatory variable, where the weights are obtained by auxiliary regression of the residuals.

As expected, the intercept CI is much smaller if the intercept is centrally located in the measured data. However, it is apparent that the unweighted median is no longer the minimum CI length in the presence of linear heteroscedasticity.

Figure 5: bootstrap intercept CI length as a function of auxiliary model weighted median or unweighted sample



Using this insight, the evd based regression intercept estimator (5) applied to linear expanding horn heteroscedasticity was trialed sourcing,

- (i) the first term in the sqrt expression, from the (auxiliary regression) weighted quantile regression of the



- original residuals distribution, and
- (ii) the  $\text{median}(x_i)$  terms in the sqrt expression were converted to (auxiliary regression) weighted medians.

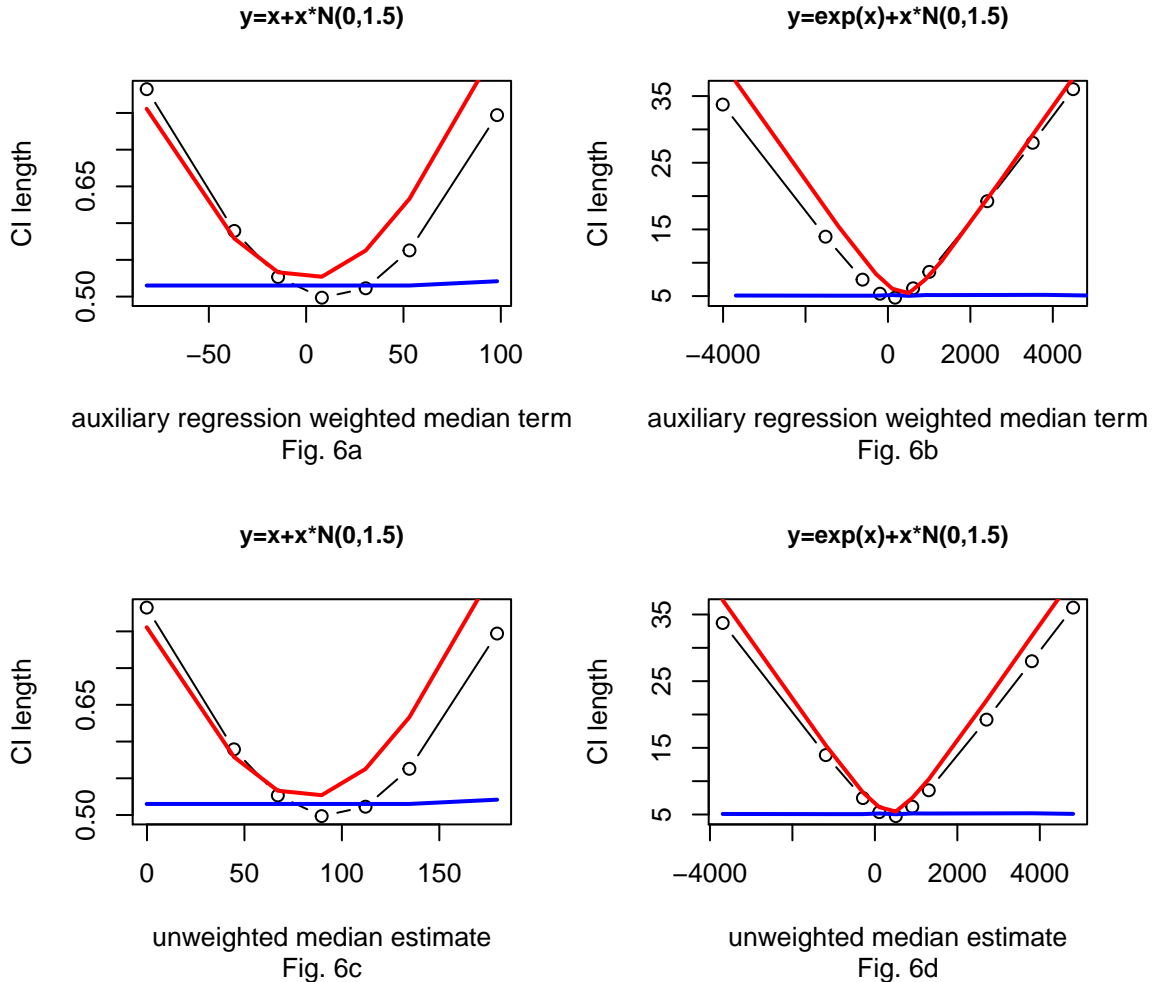
$$s_{\beta_0}(\theta) \sim \sqrt{s(\text{median}(\text{residuals}, w_{auxreg}))^2 + s_{\beta_1}^2 (0 - \text{median}(x_1, w_{auxreg}))^2} \quad (9)$$

On assessing (9), better coverage performance was occurring if  $s_{\beta_i}$  used in the intercept CI was reduced from the evd based max half CI, to mean half CI estimates as was found optimal for sample quantile CIs (1). These findings result in the following evd based model coefficient intercept CI estimator

$$s_{\beta_0}(\theta) = \sqrt{s(\text{median}(\text{residuals}, w_{auxreg}))^2 + s_{\beta_1}^2 \text{meanhalfCI}(0 - \text{median}(x_1, w_{auxreg}))^2} \quad (10)$$

As shown in figure 6, this estimator has good performance compared to the default bootstrap estimates for linear expanding horn hetercedasticity. Namely,  $s(\text{median}(\text{residuals}, w_{auxreg}))$  term closely corresponds to the minimum intercept CI length and  $s_{\beta_1} \text{meanhalfCI}$  term provides the best match to the curvature to the bootstrap intercept CI length as the intercept position is varied within a fixed scatterplot shape.

Figure 6: bootstrap and evd based intercept CI length as a function of auxiliary model weighted median or unweighted sample  
black – bootstrap, red – evd symm, blue – evd CI median(residuals, w\_auxreg)



In practice, this evd based model CI method to approximate quantile regression model coefficient CIs for linear expanding horn heteroscedastic id cases contains the following steps.

- (a) Firstly, the quantile estimating function is used to perform the quantile regression modelling and derive the residuals distribution,
- (b) in the presence of linear heteroscedasticity in the residuals, auxiliary regression of the absolute residuals is conducted using quantile regression for  $\theta = 0.5$  to obtain a set of transformation weights,
- (c) **for slope CI estimates**; for the model slope coefficient CI estimates, the residuals are transformed to homoscedastic form using the transformation weights and normalisation adjustment of the original and transformed cdf totals (7)
- (d) next, using only the sample size and the given quantile value  $\theta \in (0, 1)$ , the evd variance estimator approach provides the confidence interval bound values  $(\theta_{LB}, \theta_{UB})$  for the homoscedastically transformed quantile regression residuals distribution,
- (e) the quantile estimating function is then used on the residuals distribution to backtransform the evd based model slope coefficient(quantile) confidence interval bound values  $(\theta_{LB}, \theta_{UB})$  to the original measurement scale. For the slope CI estimates, the **maximum half CI** of the backtransformed results are used,
- (f) **for intercept (and point) CI estimates**; the **mean half CI** of the backtransformed evd slope coefficient CI estimates are obtained and will be used in equation (10),
- (g) the weighted median(s) of the explanatory variable(s) are calculated,
- (h) the variance of the weighted median of the original heteroscedastic residuals distribution is calculated using weighted quantile regression on the residuals distribution to backtransform the evd based model slope coefficient(quantile) confidence interval bound values  $(\theta_{LB}, \theta_{UB})$  to the original measurement scale,
- (i) the regression formula equation (10), is used based on (f-h) to calculate evd based approximation estimates of the quantile regression model intercept coefficient CIs. For other point estimates besides the intercept, replace 0 in the RHS of (10) with x value of interest.

In the implemented code, the regression degrees of freedom adjustment  $\sqrt{\frac{n}{(n-p-1)}}$  required for model coefficient CIs has been applied in the percentile frame  $(\theta_{LB}, \theta_{UB})$ . This is in contrast, to (3) where two versions of regression degrees of freedom adjustment were assessed.

### estimating model slope(s) and intercept CIs

The output of the evd based estimators of the residuals variance/CI were then used to estimate the model slope(s) and intercepts using equations (8) & (10). For bivariate cases, the approximation () was also included in the calculations. The coverage performance was then assessed and compared to the known population regression model and default quantreg bootstrap estimates using 600 replicates as shown in the next section.

## Assessing coverage performance for different homoscedastic datasets

In the following figures, the coverage performance of evd based model coefficients CI estimators (based on 1000 resamples) for three linear expanding horn heteroscedastic quantile regression cases.

Three samples size  $n=50,100,1000$  were trialled and four intercept positions, (i) left edge, (ii) auxiliary regression weighted median, (iii) unweighted median and (iv) right edge.

Given the scatterplot (heteroscedasticity) does not change shape under intercept shifts, the same model slope coefficient CI estimates are expected.

$$\text{model (i)} \quad y = \frac{(runif(n, -180, 180) + x_{shift})}{2} + (3 + \frac{(runif(n, -180, 180) + x_{shift})}{20} - \frac{x_{shift}}{20}) rnorm(n, 0, 1.5)$$

A simple linear relationship between the dependent variable (Y) and explanatory variable ( $runif(n, -180, 180)$ ) with linear expanding horn heteroscedasticity (see figure 1), where the quantile regression for different quantiles results in a set of diverging regression lines.

$$\text{model (ii)} \quad y = \frac{(exp(rnorm(n, 7.5, .5)) + x_{shift})}{2} + (3 + \frac{(exp(rnorm(n, 7.5, .5)) + x_{shift})}{20} - \frac{x_{shift}}{20}) rnorm(n, 0, 1.5)$$

Also a linear relationship between the dependent variable (Y) and an exponential normal explanatory variable  $exp(rnorm(n, 7.5, .5))$  with linear expanding horn heteroscedasticity. The behaviour is a synthetic simulation of the Engel food expenditure dataset. Again, the quantile regression for different quantiles results in a set of diverging regression lines. This example is to more thoroughly test coverage performance for Engel data scenario with a known population distribution.

model (iii)

$$y = \frac{(runif(n, -180, 180) + x_{shift})}{2} + (runif(n, -10, 10) + z_{shift}) + (3 + \frac{(runif(n, -180, 180) + x_{shift})}{20} - \frac{x_{shift}}{20}) rnorm(n, 0, 1.5)$$

To further test the evd based model intercept coefficient CI performance. A bivariate example where the covariance between the two explanatory variables is expected to be small and hence coverage performance may be considered less sensitive to the current inaccuracy in  $r_{12qr}$  estimator.

In the figures below, the coverage is calculated for 15 quantile points (0.025, 0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, 0.95, 0.975). The black points and lines indicate the default quantreg bootstrap estimates ( $R=600$ ).

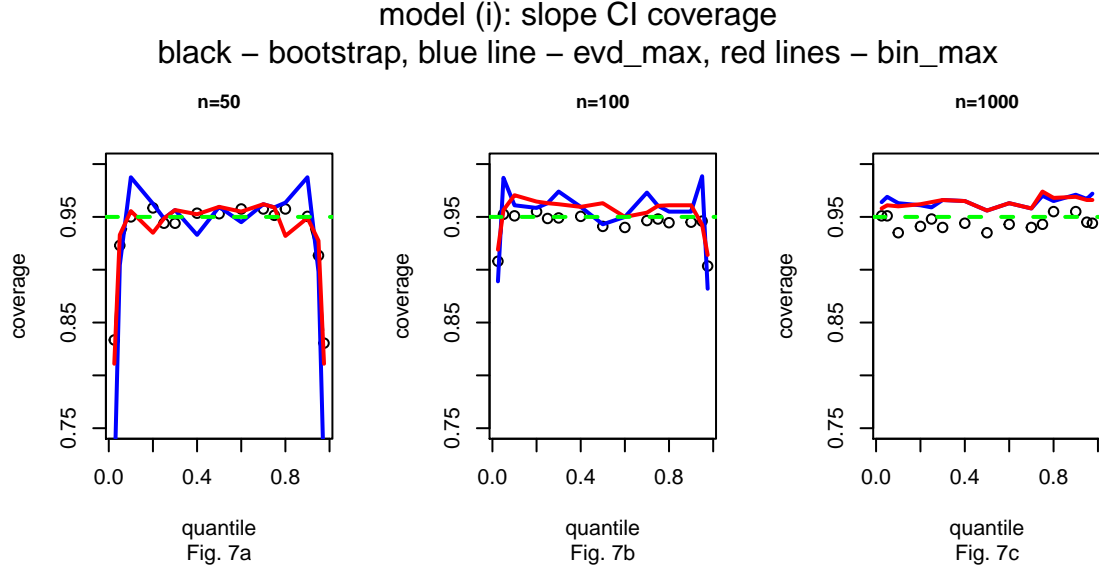
For graphs, containing the model slope coefficient CI coverage, the blue lines indicate the `evd_max` CI estimators, and the red lines indicate the `bin_max` CI estimators described in (3), which includes degrees of freedom correction in percentile scale. This overlap of estimators and subfigures allows a visual comparison of the effects of sample size, and estimator type.

For graphs, containing the model intercept coefficient CI coverage, the blue lines indicate the `evd_max` CI estimators, and the red lines indicate the `bin_max` CI estimators. In subfigures (a-c), (d-f) & (g-i) the intercept is on the left edge, median( $x, w\_auxreg$ ) & right edge of the data. This overlap of estimators and subfigures allows a visual comparison of the effects of sample size, heteroscedasticity magnitude and estimator type.

Figures 7 & 8, display the coverage performance of bootstrap, `bin_max` & `evd_max` for the slope and intercept respectively of model (i) as a function of the sample size.

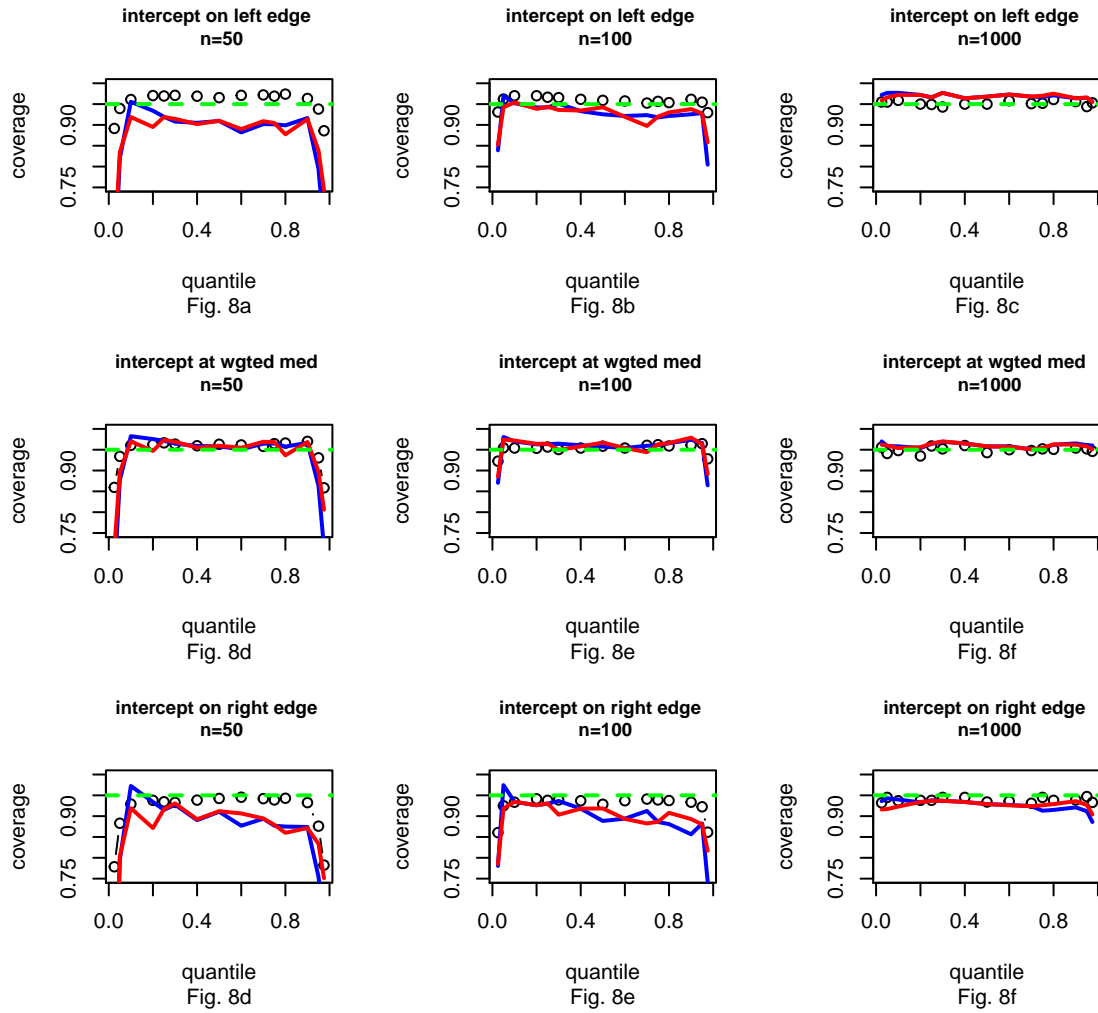
Figures 9 & 10, display the coverage performance of bootstrap, `bin_max` & `evd_max` for the slope and intercept respectively of model (ii) as a function of the sample size.

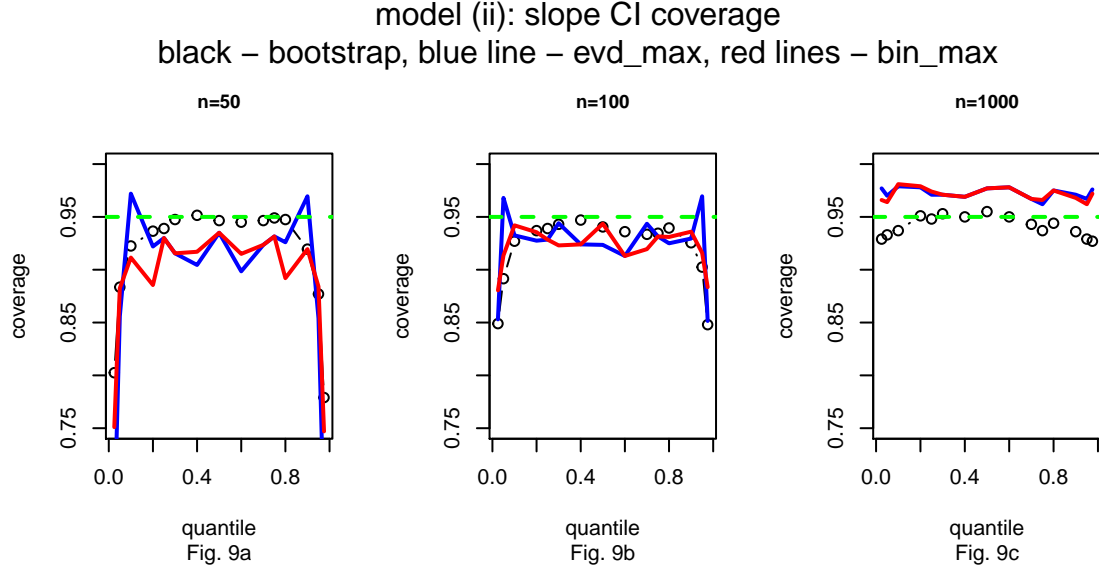
Figures 11-13, display the coverage performance of bootstrap, `bin_max` & `evd_max` for the slope and intercept respectively of model (iii) as a function of the positions of the intercepts of the two explanatory variables.



In figures 7 & 8, for  $n=1000$ , nominal 95% coverage is observed for bootstrap, evd\_max, bin\_max estimators for all quantiles (0.025-0.975) in slope CIs and  $> 96\%$  for the evd based CIs of the intercept. For smaller samples, bin\_max and evd\_max based CIs estimates exhibit poorer coverage for extreme quantiles. Good intercept coverage performance is observed for the intercept CI estimates regardless of sample size if the intercept of the explanatory variable is positioned at the (auxiliary regression) weighted median. This result indicates the first term in equation (10) is performing excellently.

model (i): intercept CI coverage  
black – bootstrap, blue line – evd\_max, red lines – bin\_max

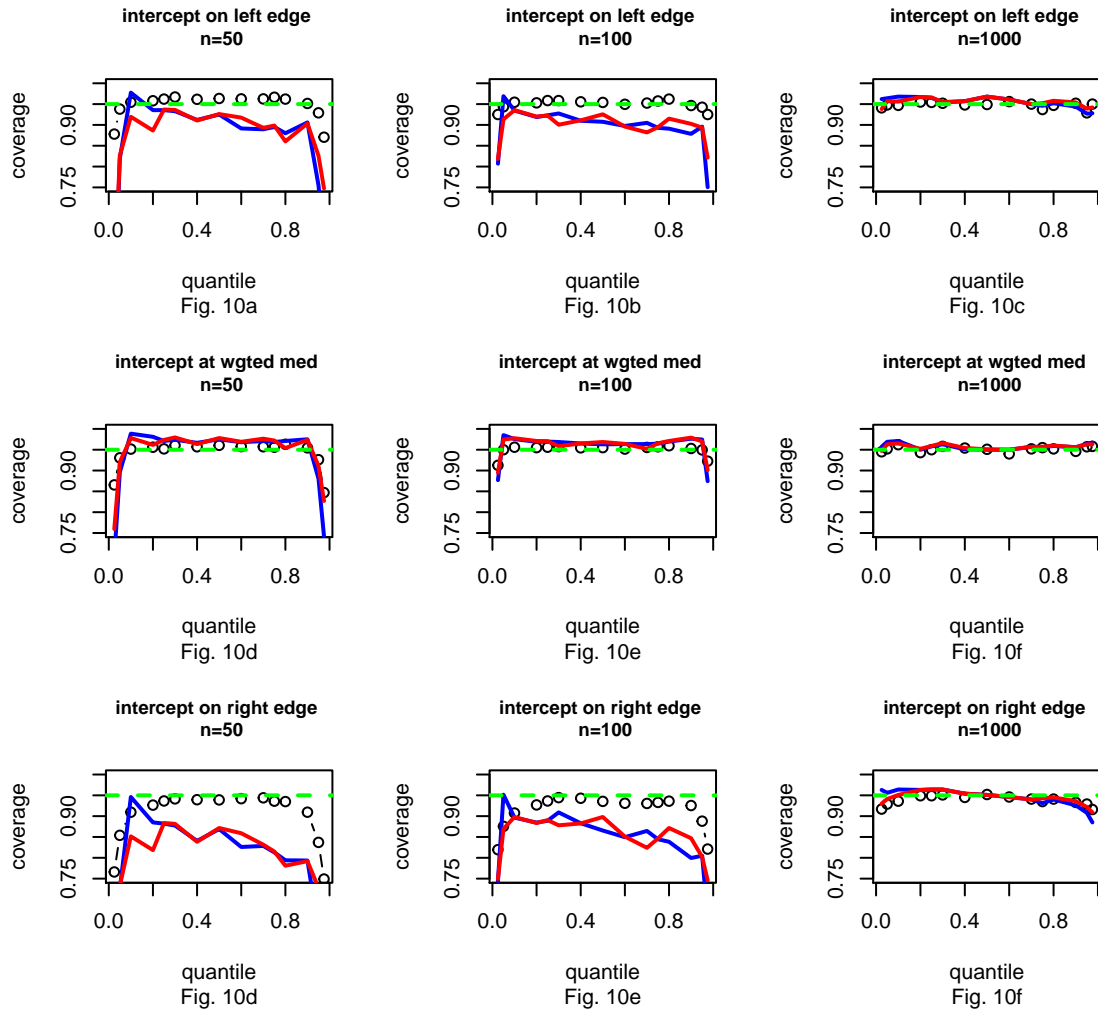


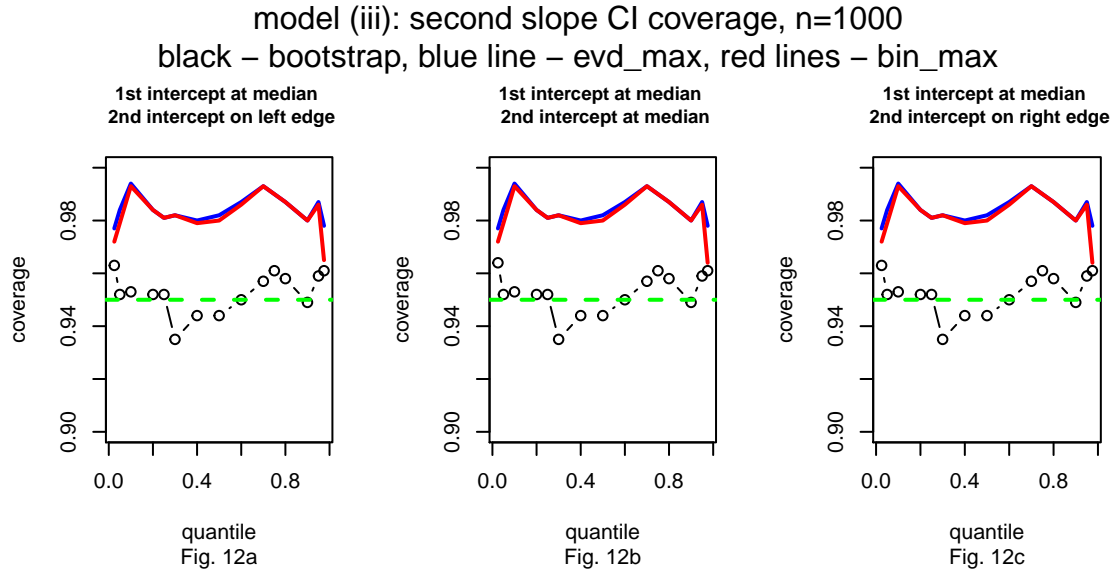
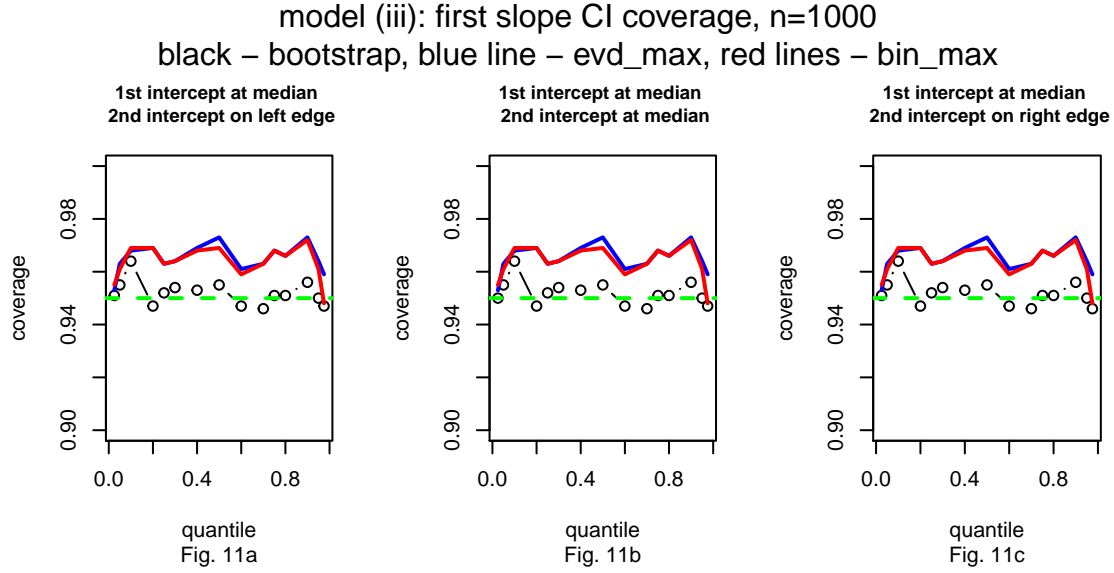


In figures 9 & 10, for  $n=1000$ , nominal 95% coverage is observed for bootstrap, evd\_max, bin\_max estimators for all quantiles (0.025-0.975) in slope CIs and  $> 96\%$  for the evd based CIs of the intercept. For smaller samples, bin\_max and evd\_max based CIs estimates exhibit poorer coverage for extreme quantiles. Comparison of figure 8 & 10 indicates weaker performance in figures 10a & 10c. Part of this difference may be due to the unequal spread of  $x$  values above & below, fitted quantile regression lines in model (ii) compared to the (random) uniform model (i). This unevenness in the data scatter is more apparent for small samples and so raises the question of what is the lowest MSE quantile regression model that could be specified with small samples. For small samples, the default bootstrap estimator is quite robust to model specification in comparison.

Looking at figures 10d-10f, the first term in equation (10) is again performing excellently with evd\_max and bin\_max intercept CI estimates.

model (ii): intercept CI coverage  
black – bootstrap, blue line – evd\_max, red lines – bin\_max



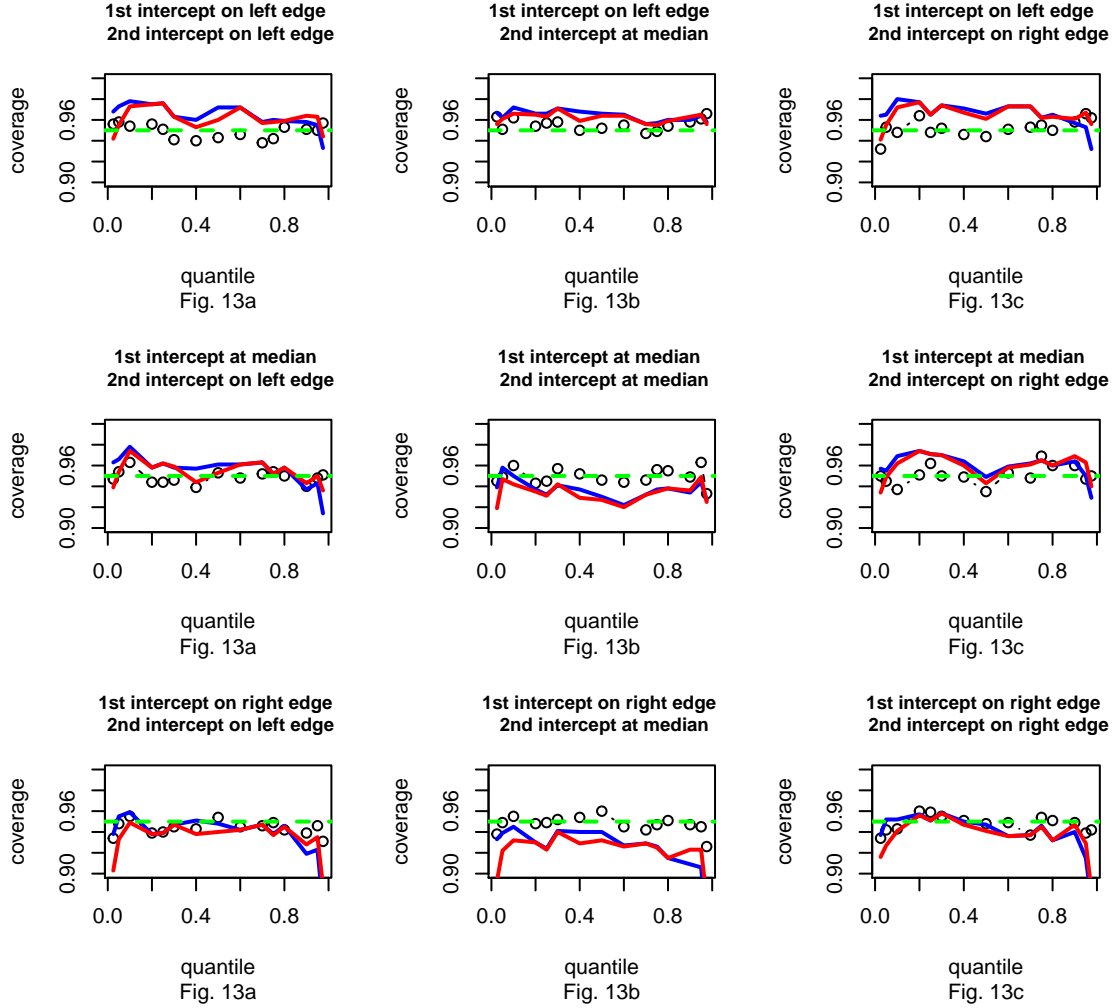


In figure 11, for  $n=1000$ , the slope CI coverage of the first variable which is causing the heteroscedasticity are at  $>96\%$  coverage performance. In figure 12, the slope CI coverage of the second variable which does not contribute to the heteroscedasticity exhibits higher overcoverage. Part of this poorer performance will be related to the inaccuracy in the covariance correction term .

Looking at the intercept CI coverage performance in figure 13, as the intercept positions are varied, the performance can exhibit slight undercoverage to slight overcoverage, in this specific bivariate case.



model (iii): intercept CI coverage, n=1000  
black – bootstrap, blue line – evd\_max, red lines – bin\_max



## Comparing coverage performance for a small sample “real world” example

To round off the CI performance comparison, appendix A, contains a comparison of the evd based model coefficient CI estimator to the quantreg default bootstrap and rank inversion CI estimates for the Engel food expenditure dataset [8]. Providing this example also allows, the r markdown version of this paper [9], to conveniently contain a concise version of the linear expanding horn heteroscedasticity evd based model coefficient CI estimator algorithm rather than the lengthy repeated sampling version.

For the Engel dataset, the linear expanding horn heteroscedasticity evd based model **slope** coefficient CI estimator produces improved results over the homoscedastic estimator (3). For model intercept coefficient CI estimates, the improvement is only small though it must be kept in mind that the Engel data intercept position is an extrapolated value and quantile crossing is observed in the linear quantile regression model fits (both conditions not considered in this paper).

## Conclusions

For linear expanding horn heteroscedastic, independent error, unweighted cases, auxiliary regression modelling of the heteroscedasticity provides improved evd based approximations to quantile regression model coefficient CIs. Good coverage performance for model coefficient CIs is observed, for moderate ( $n=1000$ ) to large samples.

In the presence of heteroscedasticity, the evd based results have shown that the minimum CI occurs at approximately  $\text{median}(x, w_{\text{auxreg}})$ . Further, the evd based value of the CI for this minimum point estimate is robust to sample size in univariate case.

The evd based estimates are sensitive to the accuracy of the specification of the quantile regression model, the current covariance term correction needs improvement and a jackknife version of the evd based approach could be considered.

## References

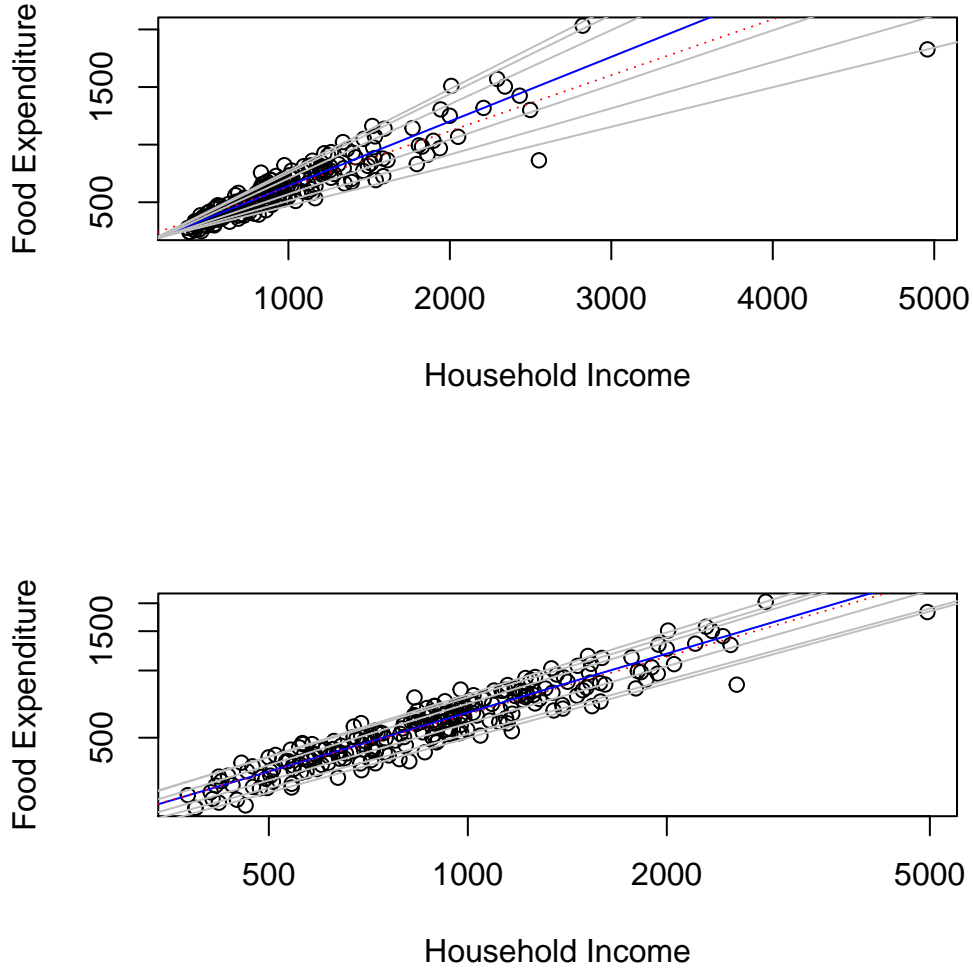
1. Martin J.P.D., 2015, <http://dx.doi.org/10.6084/m9.figshare.1566828>
2. Martin J.P.D., 2015, <http://dx.doi.org/10.6084/m9.figshare.1591019>
3. Martin J.P.D., 2015, <http://dx.doi.org/10.6084/m9.figshare.2055882>
4. Koencker, R. W. & Bassett G., *Econometrica*, 1978, vol. 46, issue 1, pages 33-50
5. Koencker, R. W., Portnoy S. et al, <https://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
6. <https://en.wikipedia.org/wiki/Quantile>
7. Brown, B. M. and Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* 92 149-158. MR2158616
8. <https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf>
9. [https://github.com/johnpdmartin/sampling-investigations/blob/master/quantile\\_regression\\_model\\_coefficient\\_CIs\\_for\\_some\\_heteroscedastic\\_cases\\_using\\_aux\\_reg\\_and\\_evd.Rmd](https://github.com/johnpdmartin/sampling-investigations/blob/master/quantile_regression_model_coefficient_CIs_for_some_heteroscedastic_cases_using_aux_reg_and_evd.Rmd)

## Appendix A: Comparison of model coefficient CIs for Engel dataset

The Engel dataset is a small sample example of the heteroscedastic relationship between food expenditure and household income and is included in the quantreg package as an quantile regression example.

In log linear form, by taking the logarithms of both food expenditure and household income prior to quantile regression analysis, the error distribution is almost homoscedastic in nature, as shown in Figure A1-b.

## Linear Quantile regression modelling of Engel dataset



**Figure A1: Engel dataset on food expenditure as a function of household income**

Figure A2, shows the model coefficients CI estimates of the intercept and slope, using

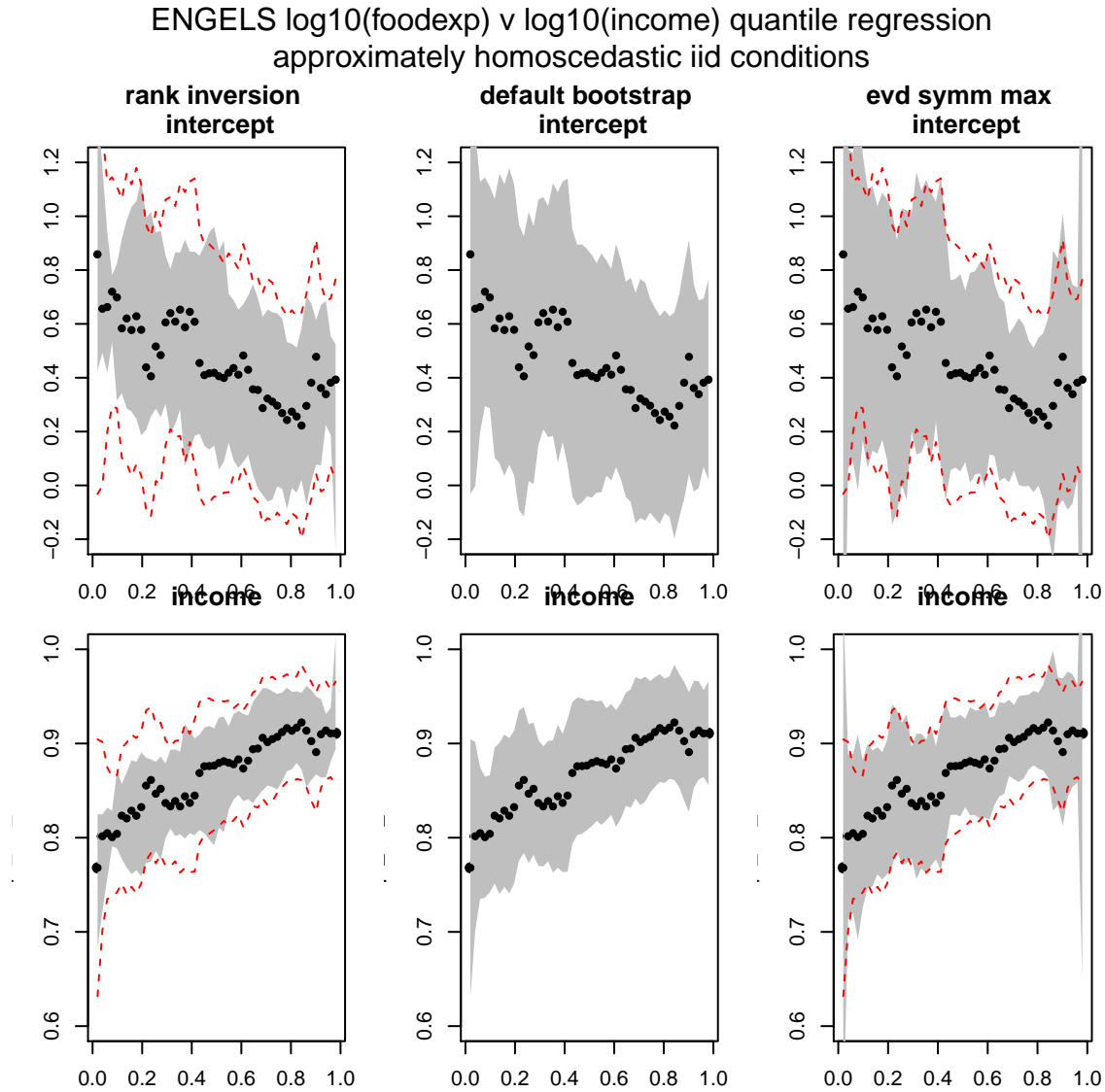
1. rank inversion
2. quantreg package default bootstrap (delete-d-group jackknife)
3. evd\_symm\_max with the degrees of freedom calculated in the percentile scale

where the outline of the bootstrap CIs is given in all the sub-graphs.

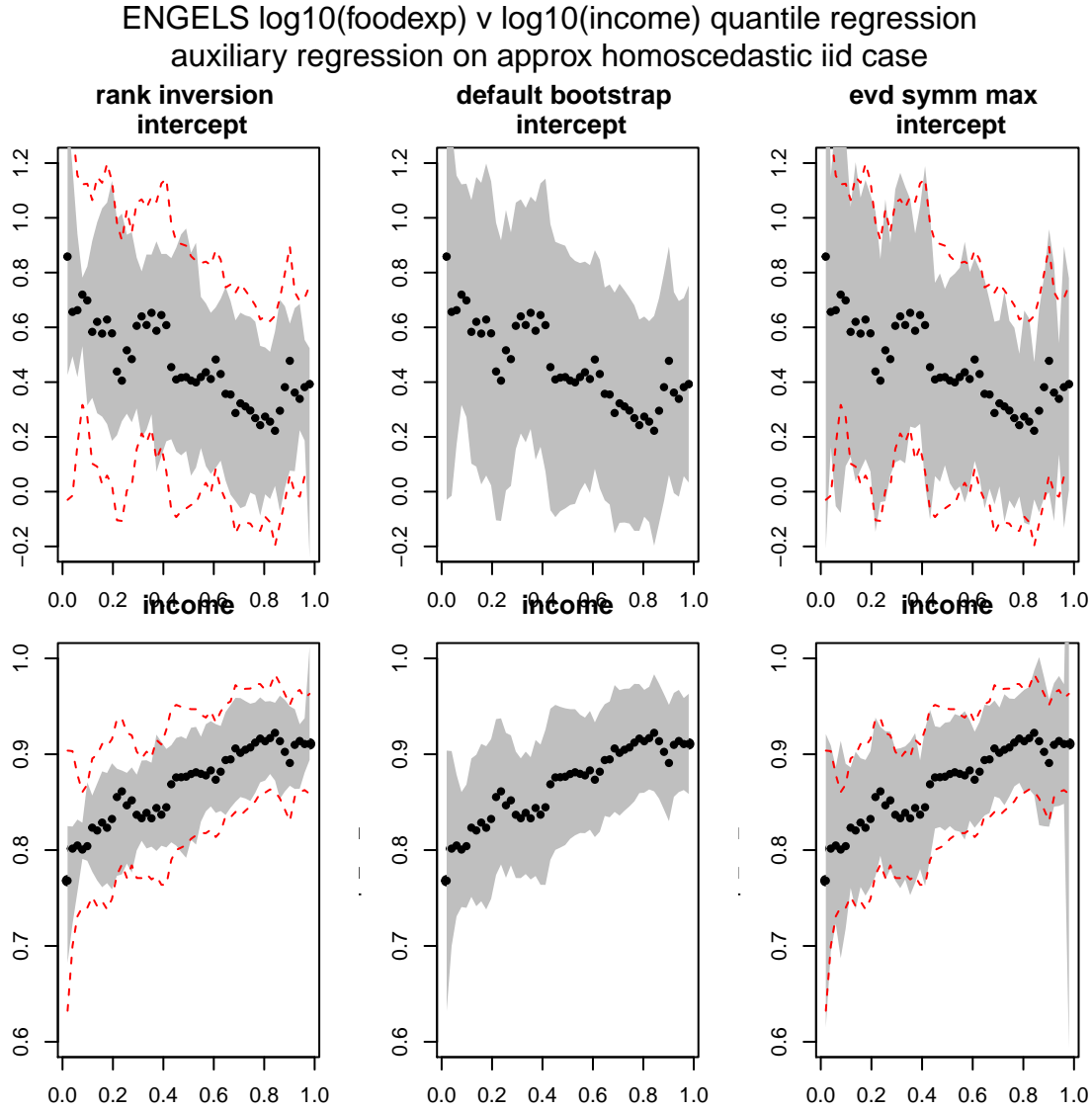
The rank inversion method was replaced as the quantreg default quantile regression CI estimator by the bootstrap method. As seen in figure A2, for this quasi-iid homoscedastic case, the bootstrap and evd\_symm\_max estimates are similar for many quantiles except for minor differences for regions near 0.4-0.5 and 0.6-0.8. The difference occur in both the slope and intercept CI estimates suggesting a breakdown of the iid conditions for the evd\_symm\_max method. The older rank inversion method exhibits smaller CIs compared to the other two methods for quasi-iid homoscedastic conditions.

Comparing figure A2 & A3 shows only a slight improvement due to the auxiliary regression adjustment to the

evd based estimators compared to the delete-d-group jackknife CI estimates. For this dataset the intercept is a extrapolated x value.



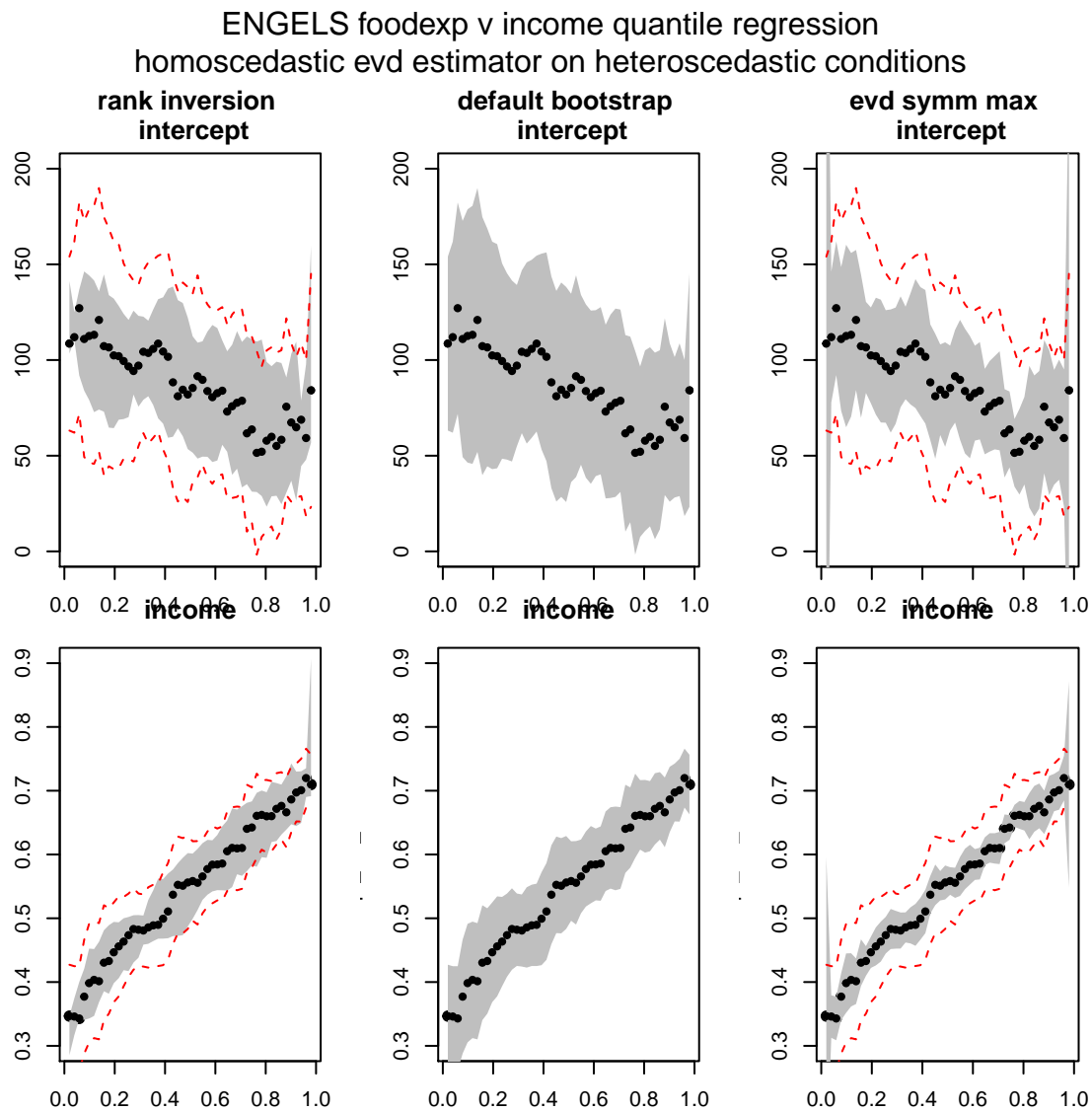
*Figure A2: Estimated intercept and slope model coefficient CIs for log-log transformed Engel dataset, homoscedastic model*



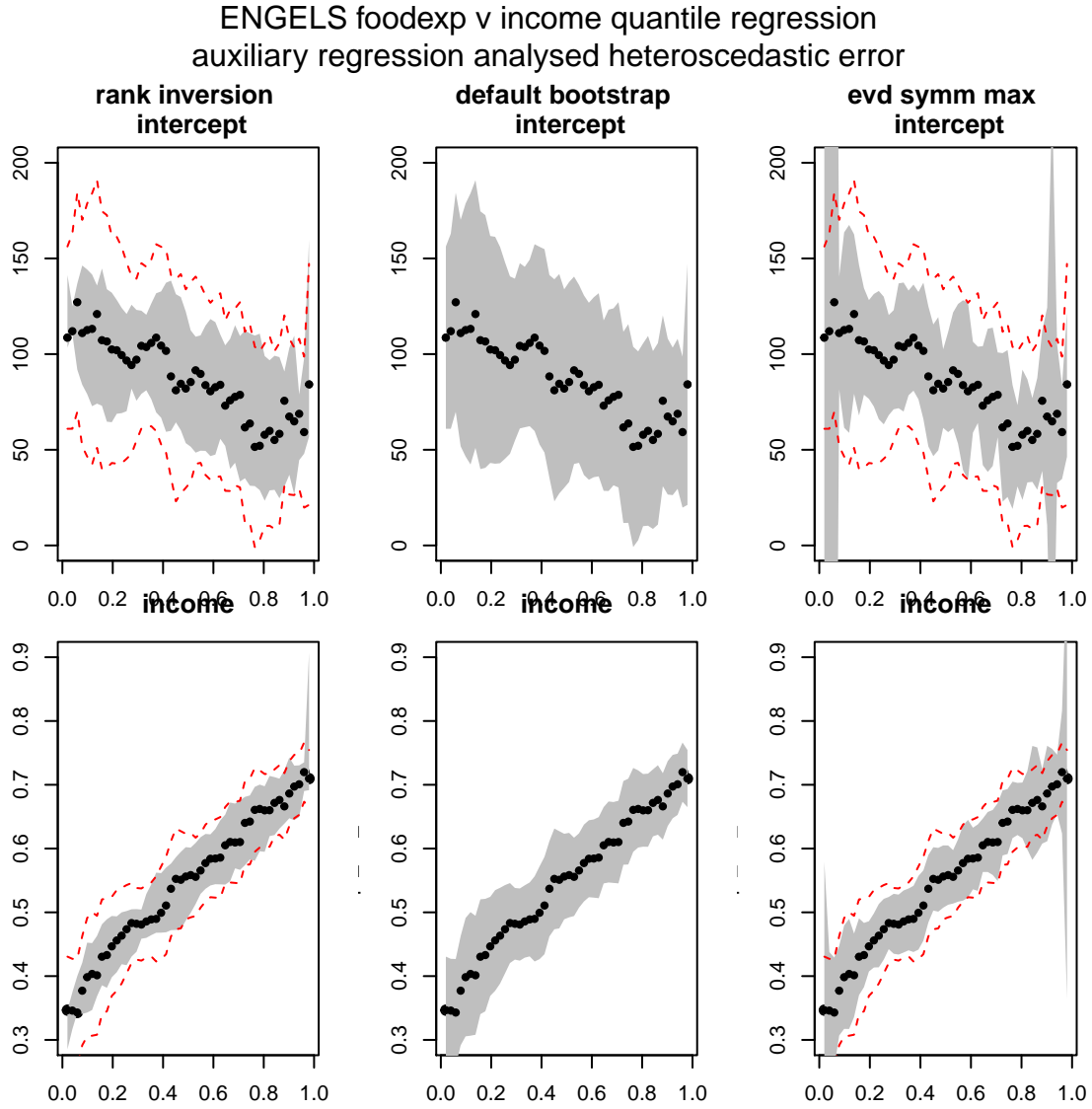
**Figure A3: Estimated intercept and slope model coefficient CIs for log-log transformed Engel dataset using heteroscedastic correction**

Figure A4 & A5 gives the comparative performance of the three CI estimators for the original heteroscedastic scattered dataset. It can be seen going from A4 to A5, that the auxiliary regression adjustment to the evd based estimators does usefully improve the coverage of the slope evd\_symm\_max CIs to a closer match with the delete-d-group jackknife CI estimates.

In comparison, the intercept CIs are only slightly increased. However, it needs to be kept in mind that for the Engel dataset, the intercept is a extrapolated x value and quantile crossing is observed in the quantile regression model fit.



*Figure A4: Estimated model coefficient CIs for non-transformed Engel dataset using evd based homoscedastic iid estimators*



*Figure A5: Estimated model coefficient CIs for non-transformed Engel dataset using evd based linear expanding horn heteroscedasticity estimators*