Partial quantile regression

Yadolah Dodge · Joe Whittaker

Abstract Partial least squares regression (PLSR) is a method of finding a reliable predictor of the response variable when there are more regressors than observations. It does so by eliciting a small number of components from the regressors that are inherently informative about the response. Quantile regression (QR) estimates the quantiles of the response distribution by regression functions of the covariates, and so gives a fuller description of the response than does the usual regression for the mean value of the response. We extend QR to partial quantile regression (PQR) when there are more regressors than observations. For each percentile the method provides a low dimensional approximation to the joint distribution of the covariates and response with a given coverage probability and which, under further linearity assumptions, estimates the corresponding quantile of the conditional distribution. The methodology parallels the procedure for PLSR using a quantile covariance that is appropriate for predicting a quantile rather than the usual covariance which is appropriate for predicting a mean value. The analysis suggests a new measure of risk associated with the quantile regressions. Examples are given that illustrate the methodology and the benefits accrued, based on simulated data and the analysis of spectrometer data.

Keywords Bilinear factor model · Longley data · Partial least squares · Quantile regression · Spectrometer data · Variance heterogeneity

Y. Dodge (✉)
University of Neuchatel, Neuchatel, Switzerland
e-mail: yadolah.dodge@unine.ch

J. Whittaker
Department of Mathematics and Statistics, Lancaster University, LA1 4YF Lancaster, UK
e-mail: joe.whittaker@lancaster.ac.uk

1 Introduction

Partial Least Squares Regression: Partial least squares regression (PLSR) is a procedure that concentrates on dimension reduction to make prediction and originates from work by Wold in the 1960 and 1970's. To paraphrase Martens and Naes (1989, p. 118–119) "PLS is a loose term for a family of …multivariate modelling methods derived from Herman Wold's basic concepts of iterative fitting of bilinear models in several blocks of variables. The chemometric version …was originally developed as a two block algorithm, consisting of a sequence of partial models fitted by least squares …motivated by …geometrical considerations (a sequence of orthogonal projections) rather than from a statistical optimization perspective."

While there is no universal agreement on the best description of the PLS procedure, if the variance matrix of the explanatory variables X is singular, it is acknowledged that PLS performs well when ordinary least squares fails. The singularity may be due either to collinearity or to so-called "fat" data when the size of the sample is less than the number of variables. PLS succeeds by projecting the data onto a lower dimensional subspace eliciting a relatively small number of components. In addition to its predictive role, plots of the PLS components and their loadings, are often informative, providing a visual description of the relationship between X and Y replicable in repeated experiments. PLS has many adherents, especially from chemometrics and food science; it is a technique that works well in practise, especially for the analysis of data incorporating spectrometer readings.

Martens and Naes (1989) give an extensive practical exposition. An excellent recent theoretical review of the statistical basis for PLS is given by Helland (2001). It is competitive with other similar statistical procedures, such as ridge regression (RR) and principal components regression (PCR); see, for instance, Frank and Friedman (1993) and more recently Hwang and Nettleton (2003). We note that the increasing deployment of electronic monitoring technology that makes "fat" data applications ever more common also makes PLS more relevant.

Quantile Regression: The seminal paper of Koenker and Bassett (1978) introduced the idea of regression quantiles that extend least absolute value, or L_1, regression with its robust qualities to the estimation of the quantiles of the residual distribution. The method is based on the quantile loss function, or check function, ρ, defined for a given $\tau \in (0, 1)$ by

$$\rho_\tau(y) = y\,(\tau - I_{y<0}(y)) \tag{1}$$

where I is the indicator function. It has the property that minimising the expectation $\mathsf{E}\rho_\tau(Y - \alpha)$ with respect to α gives the τth quantile, $F^{-1}(\tau)$, of Y. Quantile regression finds the estimate of β that minimises $\sum_j \rho_\tau(y_j - \beta^T x_j)$ from a sample of $j = 1, 2, \ldots, n$ observations on (x, y) where x is p-dimensional and y is scalar for a given τ. With $\tau = 0.5$ this becomes least absolute value regression estimating the median of Y for given x, see Dodge and Jureckova (2000) for further discussion.

The set of regression estimates evaluated may elucidate a variety of queries. Foremost is that plots of observed response values against median fitted values may be augmented by the quantiles at these points, indicating how either the response

distribution or the residual distribution changes with the level of the response. Additionally, plots of the elements of $\beta(\tau)$ against τ determine whether the covariates have a differing effect on the smaller values of the response variable compared to the larger.

Encouragingly the computational difficulties of large scale data analysis associated with the L_1 norm appear to be of a no greater order of magnitude than those associated with least squares and the L_2 norm, see Portnoy and Koenker (1997).

This paper develops partial quantile regression (PQR) which allows the computation of regression quantiles within a PLS type procedure. The principal interest is to enhance the fitted mean regression from PLSR with the additional information provided by analysing the quantile behaviour. PQR, as QR itself, allows the examination of possibly different regressions for each quantile, informative when the response distribution departs from the standard homoscedastic normal. Additionally, if different components are constructed for some quantiles it may lead to different interpretations of the way underlying components affect the response variable. The advantage that QR has over ordinary least squares (OLS) when the response variable is contaminated with outliers is also inherited by PQR.

The question of how to best modify the PLS algorithm to compute the quantiles and develop a procedure that separately estimates components for each quantile is addressed. In standard PLS a component is determined by the covariance of each of the individual explanatory variables with the response, using the ordinary covariance operator which predicts a mean value of the response. The extension to PQR developed here proposes a modification of the covariance, the quantile covariance, appropriate for predicting the τth quantile, and an associated measure of risk.

The algorithm for PQR is described in Sect. 2, and depends on the quantile covariance which is defined in the Appendix. Examples are given in Sect. 3. The first is a re-analysis of the well known Longley data to serve as a benchmark for PQR computations. It is followed by some experimental data obtained from a simulation from a bilinear factor model of interest to chemometricians, and simulations from a multiplicative regression model and from a switching regression model that exhibit variance heterogeneity and asymmetric tolerance intervals. Lastly, two examples of spectrometer readings, one based on samples of corn and one on samples of fish are analysed. In Sect. 4 the results of out of sample prediction are assessed and the paper ends with a short discussion in Sect. 5.

2 Methodology

2.1 The PLS algorithm

There exist many descriptions and even several versions of the PLS algorithm, for instance see Helland (1988), Martens and Naes (1989). These may differ in a variety of ways including: the way in which the same calculation is performed, for instance, from the output of a least squares fit or singular value decomposition, or by explicit calculation; the way in which variables are initialised and intermediate coefficients normalised; the way in which the successive steps are orthogonalised; additionally, they may differ according to whether they apply to a multivariate response. That there

is no universally agreed objective function for PLS is possibly the reason for a plethora
of approaches.

We describe the algorithms in population terms, that is, in terms of a p-dimensional
random vector X and a random scalar Y, rather than in terms of a sample. The advantage
is that we do not require the notation to handle the dimensions for the sample and
for the variables simultaneously. All the procedures may be described in terms of
expectations, so that sample, or empirical, versions are obtained by giving X and Y
the joint probability distribution that puts equal mass at the n sample observations
$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The expectations and covariances become finite sums
over the n observations, and to make this clear the expectation is written as E_n.

The linear least squares predictor of Y from X,

$$\mathsf{E}(Y|X) = \mathsf{E}Y + \mathsf{cov}(Y, X)\,\mathsf{var}(X)^{-1}[X - \mathsf{E}X], \qquad (2)$$

minimises the quadratic loss $\mathsf{E}(Y - \alpha - \beta^T X)^2$, see Whittaker (1990) or Christensen
(1991) for an exposition. It has its origins in the work of Doob who referred to it as a
linear expectation, see also Whittle (1983).

2.2 PLS: algorithm

The PLS algorithm is

0. Initialise: centre and scale the variables so that $\mathsf{E}(X_i) = 0$ and $\mathsf{var}(X_i) = 1$ for
   $i = 1, 2, \ldots, p$.
1. Repeat the steps:
   1.1 Compute a direction c from $c_i = \mathsf{cov}(Y, X_i)$ for $i = 1, 2, \ldots, p$, and nor-
       malise so that $c^T c = 1$.
   1.2 Form the 1-dimensional component $T = c^T X$ and the least squares predic-
       tors $\mathsf{E}(X_i|T)$. Save T.
   1.3 Adjust for T by replacing the elements of X by their residuals $X_i - \mathsf{E}(X_i|T)$.
2. Stop: according to a given criterion, retaining the components $(T_1, T_2, \ldots, T_k)$
   and form the final predictor

$$\mathsf{E}(Y|T_1, T_2, \ldots, T_k).$$

2.3 PLS: remarks

There are some obvious remarks to make. The PLSR procedure reduces the dimension
of the predictand space from p to k components which are, by construction, inherently
informative about Y. The direction based on $c_i = \mathsf{cov}(Y, X_i)$ may be justified by the
fact that it is the solution to maximising $\mathsf{cov}(Y, c^T X)$ subject to $c^T c$ constant. In
contrast to canonical correlation analysis, which employs the constraint $c^T \mathsf{var}(X)c$
constant, the computation of $\mathsf{var}(X)^{-1}$ is avoided.

The general form of the predictor (2) is employed at two points in the algorithm, at
step 1.2 with a 1-dimensional predictor, and at step 2 with a k-dimensional predictor.

The initialisation step may start by setting $\mathsf{E}(Y) = 0$. Some versions of the algorithm may compute $\mathsf{E}(Y|T)$ at step 1.2 and the residual $Y - \mathsf{E}(Y|T)$ at step 1.3. However, this is strictly unnecessary as the subsequent iteration would compute

$$\mathsf{cov}(Y - \mathsf{E}(Y|T), X_i - \mathsf{E}(X_i|T))$$

but this covariance equals $c_i = \mathsf{cov}(Y, X_i - \mathsf{E}(X_i|T))$ due to the orthogonality of T and $X_i - \mathsf{E}(X_i|T)$. The final predictor $\mathsf{E}(Y|T_1, T_2, \ldots, T_k)$ may be computed as a sum because the resulting components are mutually orthogonal; in this case the whole procedure only requires the fitting of 1-dimensional random variables.

As the iterations continue the residual variances of the adjusted explanatory variables at step 1.3 decrease, and it is an important feature of PLS that these variables are not rescaled to the initial value of 1 before recomputing c_i at step 1.1. The covariance used to compute c_i is bilinear, and its relative size is directly proportional to the size of the residual variation in the explanatory variables. Explanatory variables with small residual variances are not well determined, and purely by chance may have a high correlation with Y. If the variables were to be rescaled, for instance, by replacing the covariance in step 1.1 by a correlation, the relative effect of the smaller ones would be magnified.

2.4 The PQR algorithm

We wish to retain as many of the good properties of PLS as possible, including good prediction, dimension reduction, the loadings plots, one dimensional computations, but, in addition, to provide new information with a description of how the quantiles of the response variable behave. Our strategy is to retain an identical algorithmic structure but to replace the expectation and covariance operators used to calculate $\mathsf{cov}(Y, X_i)$ and $\mathsf{E}(Y|T_1, T_2, \ldots, T_k)$ by the quantile covariance cov_τ at (11) and the quantile expectation E_τ defined at (10). The quantile expectation $\mathsf{E}_\tau(Y)$ is the τth-quantile of Y, and $\mathsf{cov}_\tau(Y, X)$ is a covariance type measure of the association between Y and X formulated in terms of predicting the τth-quantile of Y from X. The details are given in the Appendix.

2.5 Proposed PQR algorithm

Consider a p-dimensional random vector $X = (X_1, X_2, \ldots, X_p)$ and a random scalar Y. Fix the percentile τ. The PQR algorithm is

0. Initialise: centre and scale so that $\mathsf{E}(X_i) = 0$ and $\mathsf{var}(X_i) = 1$ for $i = 1, 2, \ldots, p$.
1. Repeat the steps:
    1.1 Compute a direction c from $c_i = \mathsf{cov}_\tau(Y, X_i)$ for $i = 1, 2, \ldots, p$, and normalise so that $c^T c = 1$.
    1.2 Form the component $T = c^T X$ and the least squares predictor $\mathsf{E}(X_i|T)$. Save T.
    1.3 Adjust for T by replacing the elements of X by their residuals $X_i - \mathsf{E}(X_i|T)$.

2. Stop: according to a given criterion, retaining the components $(T_1, T_2, \ldots, T_k)$ and form the final predictor

$$\mathsf{E}_\tau(Y|T_1, T_2, \ldots, T_k).$$

The algorithm for PQR follows the one for PLSR exactly with the redefinition of the expectation and covariance operators at steps 1.1 and 2. A consequence is that, though the retained components $(T_1, T_2, \ldots, T_k)$ remain mutually orthogonal, the quantile expectation $\mathsf{E}_\tau(Y|T_1, T_2, \ldots, T_k)$ cannot be formed as a sum, and a k-dimensional fit is required. Another difference is that it makes no sense to replace Y by its residual adjusted for the component T at step 1.3, as is optional in the PLS algorithm. The algorithm is repeated for a set $\mathcal{T}$ of values of τ; in practise the number of values is small.

2.6 The risk and number of components

It is now fairly widespread to use some form of cross validation (CV) to control model complexity, here the number of components; for instance, Martens and Martens (2000). The discussion of Hastie et al. (2001) makes clear that a CV type measure firstly, has both to assess variation from estimation in the training sample and from prediction on the test (or hold-out) sample; and secondly, is estimated empirically by resampling: repeatedly separating the data set into training and test samples, either systematically (e.g. traditional leave-one-out CV), or according using some form of randomisation. We suppose the training sample size n and the test sample size n^* are the same for all repetitions of the resampling scheme.

The natural criterion to assess overall model fit is the risk, defined as the expected loss for predicting Y from the k components. The risk for PQR is

$$Q_\tau(k) = \frac{1}{\tau(1-\tau)} \, \mathsf{E}\,\rho_\tau[Y - \mathsf{E}_\tau(Y|T_1, T_2, \ldots, T_k)], \qquad (3)$$

where the divisor in (3) attempts to standardise the risk for different values of τ, and is exact for a uniform distribution. The risk for PLSR is the usual root mean square error

$$Q(k) = \sqrt{\mathsf{E}[Y - \mathsf{E}(Y|T_1, T_2, \ldots, T_k)]^2}. \qquad (4)$$

To make the dependence on the covariates in the predictor explicit we write μ_τ^k for $\mathsf{E}_\tau(Y|T_1, T_2, \ldots, T_k)$ and note this is the linear combination $\mu_\tau^k = X' \beta_\tau^k = \mu(\beta_\tau^k, X)$ say, where the β are the implied regression coefficients. Using this and definition (3), the empirical PQR training risk is

$$\hat{Q}_\tau(k) = \frac{1}{\tau(1-\tau)} \, \mathsf{E}_n\,\rho_\tau[Y - \mu(\hat{\beta}_\tau^k, X)], \qquad (5)$$

based on a given training sample with coefficient estimates $\hat{\beta}_\tau^k$.

The test sample version gives X^* and Y^* equal probability mass at the n^* test sample observations. For a given training and test sample, the PQR test risk is

$$\frac{1}{\tau(1-\tau)} \; \mathsf{E}_{n^*} \, \rho_\tau [Y^* - \mu(\hat{\beta}_\tau^k, X^*)].$$

Finally, the CV measure for PQR is the average PQR test risk averaged over all repetitions, r, of training and test samples in the resampling scheme

$$CV_\tau(k) = \frac{1}{\tau(1-\tau)} \; \mathsf{E}_r \, \mathsf{E}_{n^*} \rho_\tau [Y^* - \mu(\hat{\beta}_\tau^k, X^*)]. \tag{6}$$

A similar argument expresses the CV measure for PLSR

$$CV(k) = \mathsf{E}_r \sqrt{} \, \mathsf{E}_{n^*} [Y^* - \mu(\hat{\beta}^k, X^*)]^2. \tag{7}$$

The decisions needed to implement this calculation include: the largest number of components considered; the sizes of the training and test samples, n and n^*; and whether the resampling is systematic or randomised and, if randomised, the number of repetitions r. There seem to be no universal rules, and we are fairly agnostic about the particular resampling scheme employed: one caveat is that an empirically determined minimal value of n is needed to obtain a sensible training sample estimate of $\hat{\beta}_\tau^k$, and so the total sample size available is one constraint; another constraint is the computing time taken to complete a single repetition.

We tend to favour a scheme that sets a minimal value of n, chooses n^* to be relatively large, and resamples randomly rather than systematically. This has the advantages of limiting the number of repetitions and so fits (r around 20 seeming adequate); and secondly, random resampling allows every configuration of test and training sample to be drawn, at least in principle. For instance, leave-one-out CV never considers predicting two observations, which if both aberrant, could give different assessments.

2.7 Interpretation of quantile regressions

PLSR approximates the joint distribution of $(X_1, \ldots, X_p, Y)$ with p large by that of $(T_j, \; j = 1, \ldots, k, \; Y)$ with k small. PQR approximates the joint distribution of $(X_1, \ldots, X_p, Y)$ by that of $(T_{\tau j}, \; j = 1, \ldots, k; \; Y)$ for each $\tau \in \mathcal{T}$, and so provides a set of approximations. These approximations are accurate for describing the conditional distribution of Y given X, under further modelling assumptions.

Consider the PQR approximation, and take the simplest case in which $k = 1$ and where the loadings corresponding to different percentiles τ are the same. Whatever the percentile τ the PQR summary is (T_1, Y) where T_1 is one dimensional.

Suppose the data are generated by the linear additive error model $Y = \alpha + \gamma T_1 + E$ where $E \perp\!\!\!\perp T_1$. We use the symbol $\cdot \perp\!\!\!\perp \cdot$ to denote independence. The quantile of E is

$e_\tau = \mathsf{E}_\tau(Y - \alpha - \gamma T_1)$, whatever the distribution of T_1. The regression quantile of Y is interpretable as the quantile of E because of the conditional interpretation of the model: for given $T_1 = t_1$ the quantile of Y is $\alpha + \gamma t_1 + e_\tau$. Plotted against t_1 the quantile regressions are parallel lines.

However, if $E \perp\!\!\!\perp T_1$ is relaxed and, for instance, the error variance grows with T_1, a heteroscedastic regression model, the regression quantile of Y is $\alpha_\tau + \gamma_\tau t_1$ and the regression quantiles plotted against t_1 fan out. In this case the regression line $(t_1, y - \alpha_\tau - \gamma_\tau t_1)$ can no longer be interpreted as a regression quantile of an error E. But with the additional modelling assumption that the quantiles of the conditional distribution of Y given X are linear in T_1, the regression line becomes a statement about the conditional distribution of Y given T_1.

In general the conditional quantile differs from the linear regression quantile. In this case the quantile regression line $(t_1, \alpha_\tau + \gamma_\tau t_1)$ is interpretable as a **coverage property** in the joint distribution of (T_1, Y), explicitly: the proportion of points in (T_1, Y) that fall below the line is τ. The example in the Appendix illustrates this point, and the quantile regression serves as a good approximation to the quantiles of the conditional distribution.

We now turn to the interpretation of PQR components when components at different values of τ are different linear functions of X. Suppose the two percentiles for which the components are different are τ_1 and τ_2, again suppose $k = 1$ and drop the $j = 1$ suffix. The approximating distributions are (T_{τ_1}, Y) and (T_{τ_2}, Y), and the regression quantile of Y, at τ_1 say, is $\alpha_{\tau_1} + \gamma_{\tau_1} t_{\tau_1}$. The coverage property interpretation is that in the **joint** distribution (T_{τ_1}, Y) the proportion of points that lie below the line $(t_{\tau_1}, \alpha_{\tau_1} + \gamma_{\tau_1} t_{\tau_1})$ is τ_1.

Making the further modelling assumption that the conditional distribution of Y given T_{τ_1} has linear quantile regressions, gives a stronger **conditional** interpretation to the regressions.

A further example in the Appendix of a "triangular" distribution shows regression quantiles that differ for different τ.

The coverage property interpretation extends to arbitrary values of k and τ. The practical implication of different components for different τ is that different linear combinations of X are best for predicting different quantiles of Y.

3 Examples

The four initial examples are chosen to exemplify the PQR technique. The first is the small and well known Longley data set; the second is a simulation from a bilinear latent variable model; the third is a multiplicative regression in which the variance increases with mean; and the fourth is an instance of a non-linear switching model. We then give two applications for data of interest to chemometrics.

The calculations for PQR require computation of quantile regression estimates and were performed using the R software available from http://lib.stat.cmu.edu/R/CRAN, and downloading the library for quantile regression, provided by Koenker, http://www.econ.uiuc.edu/~roger/research/home.html.

Table 1 Longley example: ordinary and implied regression coefficients

Coefficients	a	b_1	b_2	b_3	b_4	b_5	b_6
Fitting all variables							
OLS	0	0.0463	−1.01	−0.538	−0.205	−0.101	2.48
0.25	−0.0546	0.302	−1.72	−0.625	−0.234	0.165	2.74
0.50	−0.0197	−0.0227	−1.48	−0.597	−0.231	−0.136	3.09
0.75	0.0425	−0.0401	−0.243	−0.452	−0.177	0.0577	1.58
Fitting 2 components							
PLS	0	0.242	0.264	−0.092	0.129	0.228	0.236
0.25	−0.202	0.443	0.561	0.114	0.131	0.0240	−0.147
0.50	0.0422	0.212	0.134	0.00110	−0.0331	0.259	0.362
0.75	0.146	0.211	0.221	0.0624	0.0775	0.224	0.236

3.1 Example 1: Longley's data

We analyse the data discussed by Longley (1967) in order to provide a simple benchmark for the computations required for PQR. It is a well-known example of collinear regression that consists of seven macroeconomic variables, observed yearly from 1947 to 1962. It serves as a good example for a dimension reduction procedure such as PLSR, because all explanatory variables appear informative, but when all are included in the regression the results are unstable.

The explanatory variables are GNP price deflator, GNP, unemployment, armed forces size, adult population, and year, and the number of people employed is taken as the response. Here $n = 16$ and $p = 6$. All the explanatory and the response variables are standardised to have mean zero and variance 1. This implies that coefficients from single variable regressions may be interpreted as correlation coefficients and, in other regressions, have the same order of magnitude as a correlation.

The upper half of Table 1 displays the regression coefficients for the 6 explanatory variables, and the intercepts, obtained from a standard multiple regression (OLS) and three standard quantile regressions with $\tau = 0.25, 0.5, 0.75$. The pattern is similar for all four regressions: several of the values are well outside $(-1, 1)$ with, for instance, large values of b_6 in part compensated for by large negative values of b_2 reflecting the near collinearity of this data set.

The tables and plots obtained from a PLSR analysis, similar to the analyses of Martens and Naes (1989), and parallel results from a PQR using the percentiles $\tau = 0.25, 0.5, 0.75$, are presented here.

The estimated least squares risk (4) and quantile risk (3) from fitting an increasing number of components is plotted in Fig. 1a.

For both PLSR and PQR the risk is substantially reduced by including one component and there may be an argument for including a second, or even a third, in the predictor.

The standard deviations of the components are plotted in Fig. 1b. The standard deviations of PLS components 2 and 3 are larger than their PQR counterparts, but otherwise the pattern is similar for all regressions. The values for components 4, 5 and
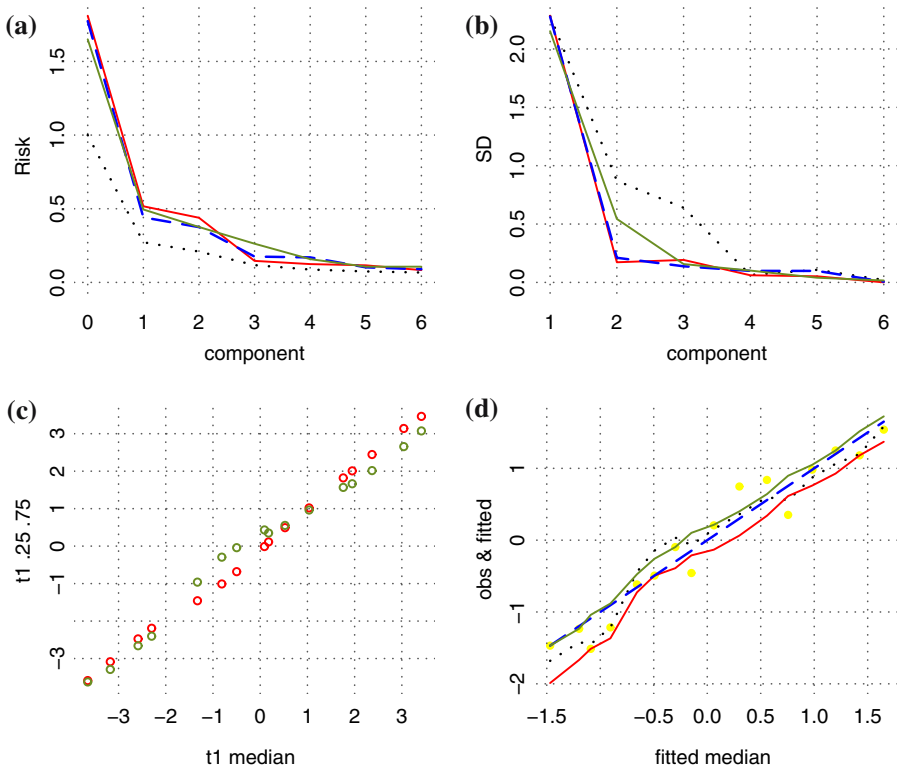
Fig. 1 PLS and PQR analysis for the Longley data: **a** estimated risk after fitting upto 6 components; **b** standard deviations of the first 6 components; **c** PQR component plot for upper and lower quartiles plotted against the median component ($k = 1$); **d** observed and fitted values based on $k = 2$ components plotted against the fitted median. Colour code: observations yellow points; PLSR dotted black line, PQR $\tau = 0.25$ red line or points, PQR $\tau = 0.5$ dashed thick blue line, PQR $\tau = 0.75$ green line or points

6 are near zero and reflect the collinearity in the data set. Further analyses here are based on choosing two components for each of the four regressions.

The implied regression coefficients in the bottom half of Table 1 based on fitting just two components are rather different from the OLS and QR estimates in the upper half of the table. Their overall magnitude is now plausible. Since the variables are standardised to mean 0, the intercepts are plausible with $(-0.202, 0.042, 0.146)$ estimating the quartiles of conditional distribution of the response at the mean value of the explanatory variables. The regression equations for the PLSR and the PQR with $\tau = 0.5$ and $\tau = 0.75$ are broadly similar. However, the regression equation for the lower quartile with $\tau = 0.25$ puts less weight on variables 5 and 6 and more on variables 1 and 2.

The loadings, the values of the coefficients c_1 and c_2, are displayed in Table 2. The scatterplot of the two loadings (not shown) indicate that four of the 6 explanatory variables cluster together, and a pairs plot (not shown) indicates that it is these variables that account for the collinearity. The vector c_1 is broadly the same for all regressions but the second displays some somewhat complex differences.

Table 2 Longley example: variable loadings of first two components

Variable	1	2	3	4	5	6
c_1						
PLS	0.472	0.478	0.244	0.222	0.467	0.472
0.25	0.478	0.471	0.269	0.186	0.457	0.483
0.50	0.485	0.465	0.245	0.284	0.465	0.438
0.75	0.377	0.402	0.240	0.554	0.425	0.388
c_2						
PLS	0.108	0.199	−0.962	0.122	0.0509	0.0805
0.25	0.294	0.499	−0.064	0.0443	−0.396	−0.708
0.50	0.426	0.284	0.0251	−0.0326	0.509	0.69
0.75	0.469	0.484	0.0472	−0.122	0.478	0.548

Figure 1c is the PQR component plot with the components for the upper and lower quartiles for $k = 1$ plotted against the median component. It shows very little difference between the three components.

Figure 1d plots the observed values of Y and the fitted quantiles against the median fitted value $\mathsf{E}_{0.5}(Y|T_1, T_2)$. These PQ regressions respect the monotonicity property: that at a given x the fitted values with τ_1 exceed those with τ_2 whenever $\tau_1 > \tau_2$.

3.2 Example 2: bilinear factor model data

We sample from a bilinear factor model, described in the PLS literature by Martens and Naes (1989)

$$X = AZ + E \quad \text{and} \quad Y = BZ + F,$$

where X and Y are linear combinations of a latent vector Z with independent standard normal errors E and F. The data used below was generated with Z consisting of independent standard normal variables. The A and B matrices are partitioned so that $X = (X_1, X_2)$ where X_1 depends on (Z_1, Z_3), X_2 depends on (Z_2, Z_3), but Y depends on (Z_1, Z_2). The correlation induced by Z_3 leads to two components extracted. The sample size was $n = 120$ and the number of explanatory variables $p = 32$.

Figure 2 summarises the results. In panel (a) the risks for each quantile indicate that $k = 2$ components are needed (confirmed by cross validation).

Figure 2b plots observed and fitted values based on $k = 2$ components against the fitted median. The variation in the median is matched by the upper and lower fitted quantiles, and this is more clearly displayed by Loess smoothing in Fig. 2d, using the default settings of the R function. The smoothed quantiles are approximately parallel, which reproduces a feature of the generating model.

The residuals, $Y - \mathsf{E}_{0.5}(Y|T_1, T_2)$, together with the (smoothed) median adjusted quantiles $\mathsf{E}_\tau(Y|T_1, T_2) - \mathsf{E}_{0.5}(Y|T_1, T_2)$ for $\tau = 0.25, 0.75$, are plotted against the rank order of the residual in Fig. 2c. Approximately 1/4, 1/2 and 3/4 of the residuals fall
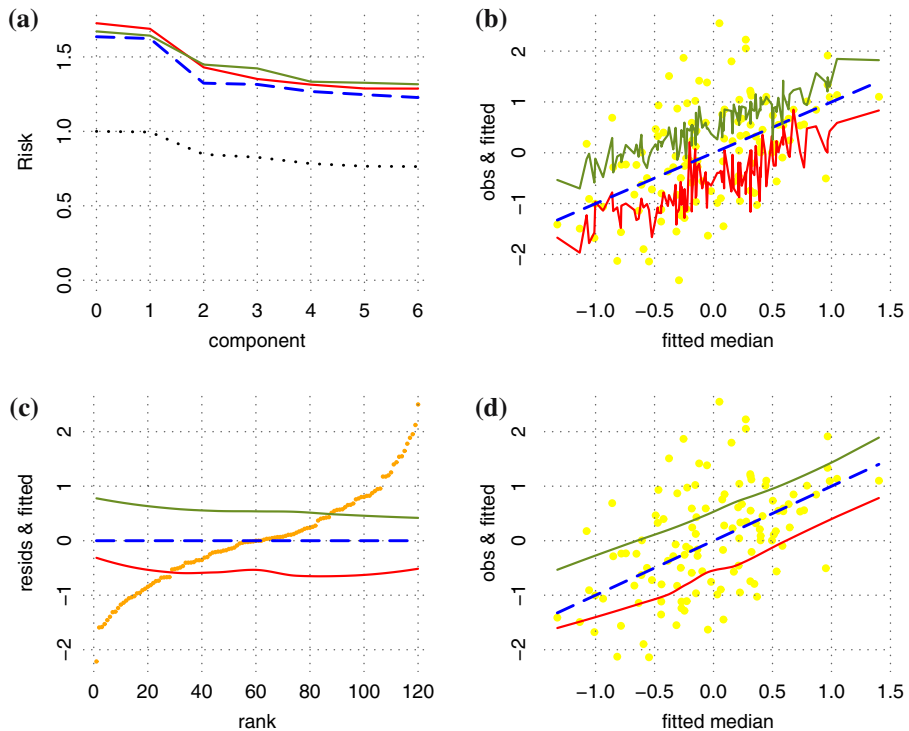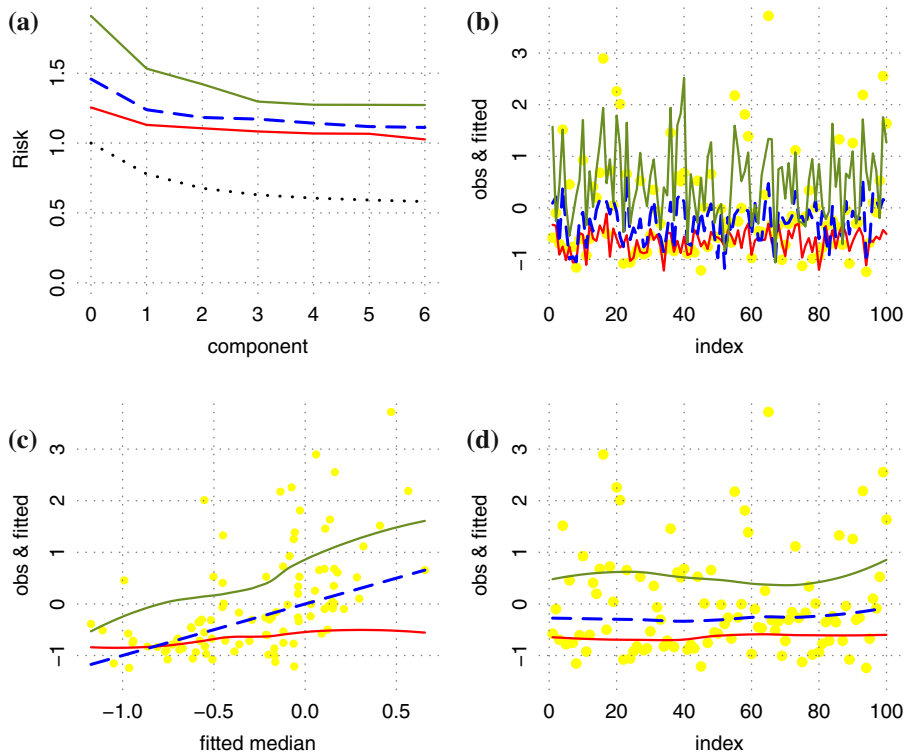
Fig. 2 Bilinear factor data, $n = 120$, $p = 32$: **a** estimated risk after fitting upto 6 components; **b** observed and fitted values based on $k = 2$ components plotted against the fitted median; **c** rank plot of ordered residuals with associated smoothed median adjusted fitted values; **d** observed and smoothed fitted values based on $k = 2$ components plotted against the fitted median. Colour code: observations yellow points; residuals orange points; PLSR dotted black line, PQR $\tau = 0.25$ red line, PQR $\tau = 0.50$ dashed thick blue line, PQR $\tau = 0.75$ green line

below the lower quantile, the median and the upper quantile, respectively, as expected under this generating model.

Under this model the conditional distribution of Y given X is normal with homoscedastic error terms, and it is easy to verify that the theoretical loadings in PQR are the same for each percentile τ. This is empirically verified in the plots of the PQR loadings (not shown here). In this example the implied regression coefficients of Y on X have a characteristic structure dependent upon A and B, which is reproduced empirically in an index plot (also not shown). Interestingly this reveals that PLS estimates have smaller sampling variation than PQ estimates because sampling under a Gaussian model makes LS most efficient.

3.3 Example 3: heteroscedastic regression data

A multiplicative model is generated using the log-normal distribution. A linear predictor, η, is formed as an average of the elements of X which are generated independently

Fig. 3 Heteroscedastic regression data, $n = 100$, $p = 60$: **a** estimated risk after fitting upto 6 components; **b** index plot for observed and fitted quartiles; **c** observed and (smoothed) fitted quartiles based on $k = 2$ components plotted against the fitted median; **d** index plot for observed and (smoothed) fitted quartiles. Colour code: observations yellow points; PLSR dotted black line, PQR $\tau = 0.25$ red line or points, PQR $\tau = 0.50$ dashed thick blue line, PQR $\tau = 0.75$ green line or points

standard normal, and Y is sampled by specifying $\log(Y)$ is normal with mean η and variance 1. The model is characterised by an increasing variance dependent on the mean. A sample of $n = 100$ observations based on $p = 60$ explanatory variables was generated, and the results are displayed in Fig. 3.

The declining risk from fitting upto 6 components is displayed in Fig. 3a. An index plot of the observed and fitted quantiles is displayed in Fig. 3b. The smoothed version of this panel is displayed in Fig. 3d below and the lack of symmetry in the response distribution becomes evident.

In Fig. 3c the observations and the (smoothed) fitted quantile regressions are plotted against the fitted median, and show (i) that the variation in the median is quite substantial; (ii) the fitted quantile regressions separate with increasing median and do not form parallel curves, and (iii) here the lower quartile and the median actually intersect, in part a function of the Loess smoothing.

Clearly, given the generating model, a more efficient analysis would be to transform the response by taking logs. The PQR components plot (not shown) display substantial
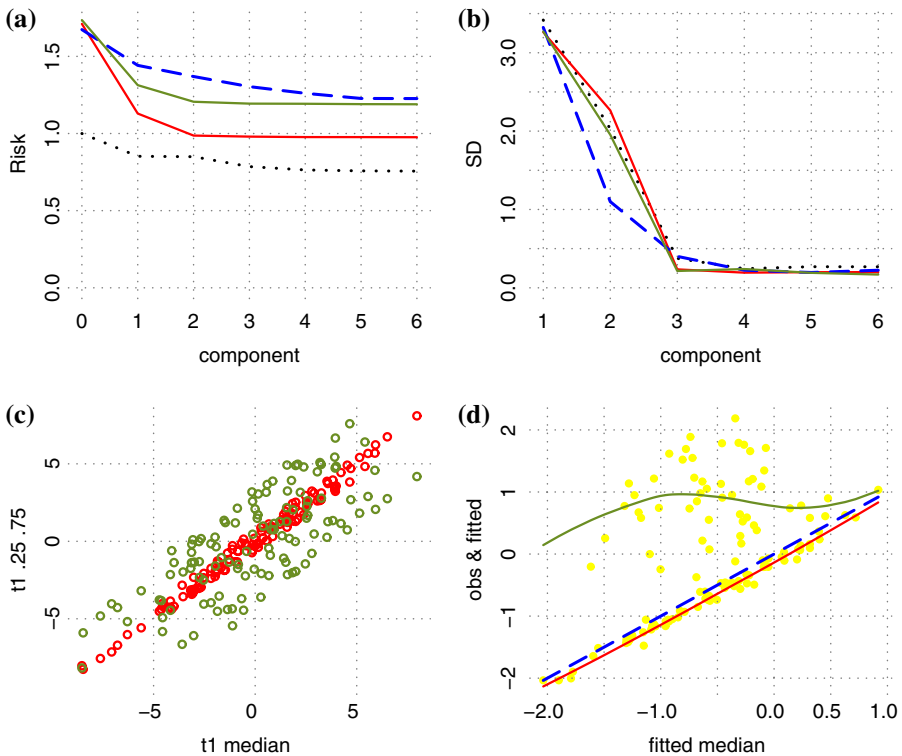
Fig. 4 Switching model data, $n = 120$, $p = 20$: **a** estimated risk after fitting upto 6 components; **b** standard deviations of the first 6 components; **c** PQR component plot for upper and lower quartiles plotted against the median component ($k = 1$); **d** observed and smoothed fitted values based on $k = 2$ components plotted against the fitted median. Colour code: observations yellow points; PLSR dotted black line, PQR $\tau = 0.25$ red line or points, PQR $\tau = 0.50$ dashed thick blue line, PQR $\tau = 0.75$ green line or points

scatter among the components indicating different components are produced at the three percentiles.

3.4 Example 4: data from a switching regression model

Underlying latent variables have distributions $M \sim$ Uniform, $Z \sim$ N and, to induce a switch, $S \sim$ Bernoulli on $\{-1, 1\}$. The covariates are partitioned into $X = (X_1, X_2)$, and are related to the latent variables by $X_1 | M \sim N(m1, I)$, $X_2 | Z \sim N(z1, I)$, and $Y | S, M \sim N(sm + z, 1)$. The non linear model generated shows that scatter plots of elements of X_1 with Y are somewhat triangular with the lower quantile decreasing with X but the upper quartile increasing.

A sample of $n = 120$ observations based on $p = 20$ explanatory variables was generated, and the results are displayed in Fig. 4.

The risks displayed in Fig. 4a indicate two components are needed. In Fig. 4c the PQR components plot shows that the lower quantile and the median components are

almost identical, but that the upper quantile component substantially differs. The effect of different components is displayed in the final panel of Fig. 4d where the (smoothed) upper quantile regression does not form a parallel curve.

The PQR loadings plot (not shown) of the loadings for the first components against those for the second indicate that the plot corresponding to the upper quantile differ from the other two.

3.5 Example 5: NIR spectra of corn samples

NIR (near-infra-red) spectrometer data on $n = 80$ samples of corn is available as a Matlab file from http://software.eigenvector.com/Data. The file contains the absorbances of the samples at each of $p = 700$ wavelengths equally spaced in the range 1,100–2,498 nm, together with the measured percentage contents of moisture, oil, protein and starch. Here we give the PQR for protein content on the wavelengths for spectrometer M5.

The decline in risk plotted in Fig. 5a suggests that $k = 5$ components are needed for PLSR and for PQR, irrespective of the quantile. This is verified by cross validation. The first component accounts for the vast majority of variation in the explanatory variables, and the standard deviations of the remaining components are roughly in ratio of 1:25.

There are three additional plots in Fig. 5 to the previous examples. Panels (c) and (d) show PQR loadings plot with p points representing the covariates for $\tau = 0.5$; and $\tau = 0.75$ respectively. The plots for $\tau = 0.25$ and for PLSR (both not shown) are very similar to (c) with $\tau = 0.5$ and consists of a single cluster at a fairly constant value of first loading. In contrast the plot for $\tau = 0.75$ indicates two clear clusters. The effect of this difference makes itself apparent in the index plot of the implied regression coefficients of Fig. 5e, where there is a noticeable difference for wavelengths between 225 and 375, for the upper quantile.

Features such as these clusters or different responses at different percentiles may well arise from skew or heterogeneous distributions. One might further argue that this is scientifically interesting, as such distributions may be manifested by some hidden but scientifically meaningful feature of the data.

3.6 Example 6: dry weight determination of cod fillets

This data, Andersen and Rinnan (2002), gives measurements of the distribution of water within fresh cod, as an example of heterogeneity within biological materials. Five cod were caught in Aresund, Denmark, in August 2000. One fillet from each fish was used for the analysis. The fillets were divided in squares of $1.5\,\mathrm{cm}^2$ retaining information on the anatomical position of the squares. The NMR measurements were performed on a pulsed NMR analyzer, and further details are given in the paper. Only even echoes were recorded, which gives 512 echoes measured for each sample. All measurements were taken at the same temperature.

The water content was determined on the same samples as used for the NMR measurements, after the NMR relaxations were measured. The fish samples were kept
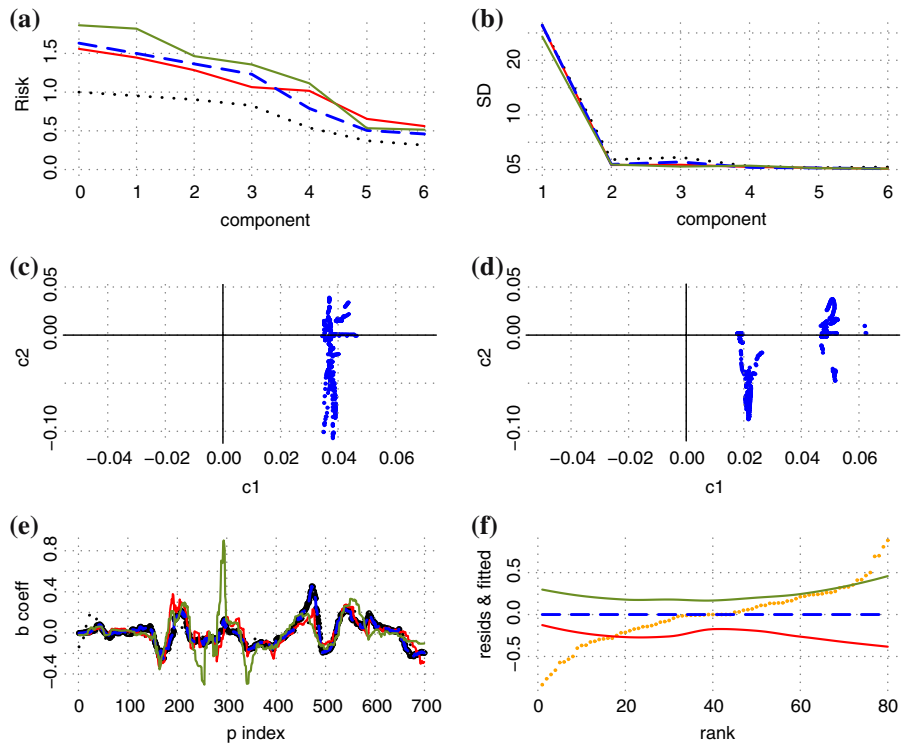
Fig. 5 Corn sample data, $n = 80$, $p = 700$: **a** estimated risk after fitting upto 6 components; **b** standard deviations of the first 6 components; **c** PQR loadings plot for $k = 1$ against $k = 2$ with $\tau = 0.5$; **d** PQR loadings plot for $k = 1$ against $k = 2$ with $\tau = 0.75$; **e** index plot of implied regression coefficients based on $k = 5$ components; **f** rank plot of ordered residuals with (smoothed) median adjusted fitted values. Colour code: residuals orange points, covariates blue points; PLSR dotted black line, PQR $\tau = 0.25$ red line, PQR $\tau = 0.5$ dashed thick blue line, PQR $\tau = 0.75$ green line

in the small glass tubes, dried, and weighed before and after drying. The data consists of the $p = 512$ NMR decays and the water content on the $n = 254$ samples from the five fillets.

The principal objective is to predict water content directly from NMR measurements.

The decline in risk plotted in Fig. 6a suggests that $k = 3$ components are needed for PQR, irrespective of the quantile. The picture is not so clear for PLSR which exhibits a rather smooth decline in risk. Cross validation supported $k = 3$ components for PQR, and, in fact, for PLSR.

The explanation for the lack of determination in the PLSR risk can be seen Fig. 6c where two outliers are apparent. While the fitted PQ regressions ignore the magnitude of the residuals the PLS fitted values, joined by dots, attempt to approximate these outlying points by introducing extra components.

The PQR component plot in Fig. 6b indicates a slight difference in orientation between the component for the lower quartile and for the median (the latter and the
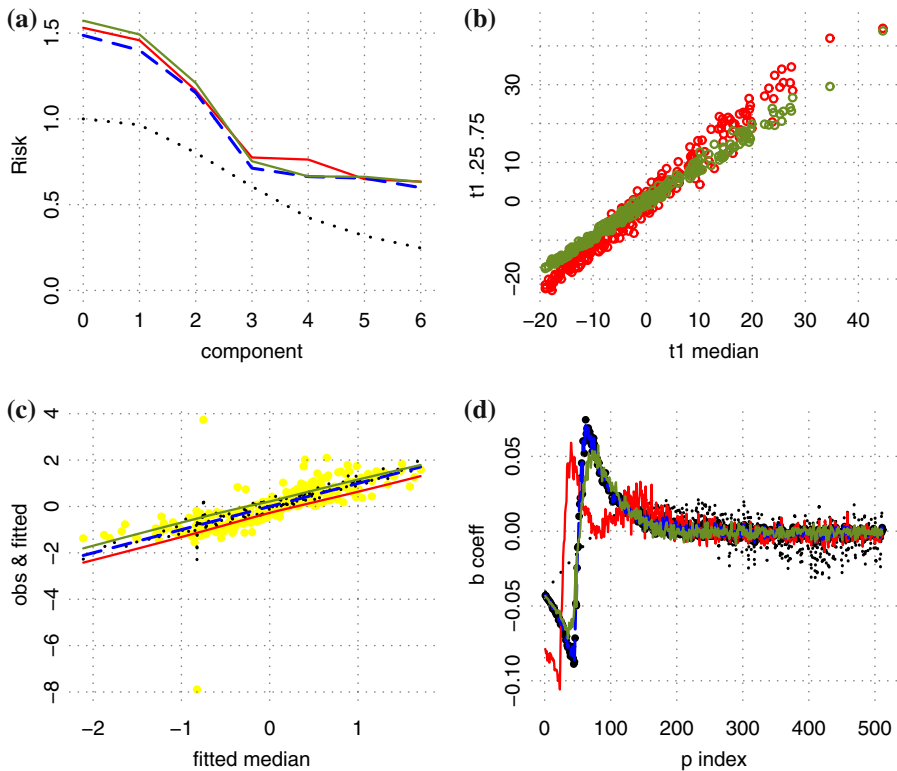
Fig. 6 Cod fillet data, $n = 254$, $p = 512$: **a** estimated risk after fitting upto 6 components; **b** PQR component plot for upper and lower quartiles plotted against the median component ($k = 1$); **c** observed and smoothed fitted values based on $k = 3$ components plotted against the fitted median; **d** index plot of implied regression coefficients based on $k = 3$ components. Colour code: observations yellow points; PLSR dotted black line, PQR $\tau = 0.25$ red line or points, PQR $\tau = 0.50$ dashed thick blue line, PQR $\tau = 0.75$ green line or points

upper quartile component are approximately the same). The index plot of the implied regression coefficients in Fig. 6d show there is a difference at the lower quantile. While the pattern is similar, there is a shift to the lower wavelengths in predicting the lower quantile.

4 Prediction results

We briefly summarise the empirical results for out of sample prediction on the examples discussed in the previous section. The re-sampling method used is the same for the assessment of both simulated data and published data.

At each percentile τ, quantile regression estimates $E_\tau(Y|x)$, a linear function of the original predictands, x, that may be used to predict the quantile of Y at a new value $x = x^o$ say. An assessment of the accuracy of our method may be made by evaluating the frequency with which a new value y^o lies in the inter-quantile

Table 3 Out of sample prediction: numbers falling in the 4 inter-quartile intervals for predicting a hold out sample of size 20, and two Pearson's chi-squared statistics, based on 20 re-samples

Data	Numbers in interval				Chi-squares	
	0–0.25	0.25–0.50	0.50–0.75	0.75–1	3df	60df
(a)	133	72	83	112	23.1	92.4
(b)	115	101	80	104	6.4	82.4
(c)	116	64	107	113	17.7	72.4
(d)	112	90	98	100	2.5	52.8

Data: (a) heteroscedastic regression data, $n = 60$, $p = 8$; (b) heteroscedastic regression data, $n = 60$, $p = 20$; (c) bilinear factor data, $n = 60$, $p = 80$; (d) corn sample data, $n = 60$, $p = 700$

intervals $\{E_{\tau_1}(Y|x^o) < y^o \leq E_{\tau_2}(Y|x^o)\}$, for chosen values of (τ_1, τ_2). In principle this frequency should be $\tau_2 - \tau_1$.

We use a simple design to evaluate this assertion. The intervals are defined by $(0, 0.25)$, $(0.25, 0.5)$, $(0.5, 0.75)$ and $(0.75, 1)$, making four equal probability intervals in all. The predictions are made on a hold out sample consisting of a random 25% subset of the original data, with 75% of the data used for estimating the regression functions. In fact, for each of the data sets assessed, the hold out sample is of size 20 and the estimation is based on $n = 60$ observations. The observed numbers from the hold out sample falling in these intervals should be uniformly distributed in the intervals and the goodness of fit is measured by Pearson's chi-squared statistic.

Each data set is a single sample of size 80 taken from: (a) the heteroscedastic regression model with $p = 8$, (b) the heteroscedastic regression model with $p = 20$, (c) the bilinear factor model with $p = 80$, and (d) the corn sample data with $p = 700$.

There are two χ^2 statistics reported in Table 3. Each resample is classified into 4 inter quartile categories giving an observed Pearson χ^2 statistic on 3 df, which are then summed over the 20 resamples to give an overall χ^2 statistic on 60 df. The reported Pearson χ^2 statistic on 3 df in Table 3 is obtained from the overall number of the 400 resamples falling into the 4 inter quartile intervals. The upper 95% quantile of the χ_3^2 distribution is 7.82, and that of χ_{60}^2 is 79.1.

In each of the four resampling experiments rather too many of the 400 outcomes fell below the lowest quartile. There are also somewhat too many in the highest interval, but this is not so marked. There are consequently too few in the two central inter-quartile intervals, but this is not uniform. Interestingly the worst behaving data set (a) is the one with the fewest covariates.

5 Discussion

Summary: We have extended PLSR to PQR so describing the quantiles of the response variable in terms of regression functions when there are comparatively fewer observations than explanatory variables. The methodology parallels the procedure for PLS, using a quantile covariance that is appropriate for predicting a quantile rather than the usual covariance which is appropriate for predicting a mean value, and the analysis

suggests a new measure of risk associated with the quantile regressions. For each percentile the method provides a low dimensional approximation to the joint distribution of the covariates and response with a given coverage probability and which, under further linearity assumptions, estimates the corresponding quantile of the conditional distribution. Examples are given that illustrate the methodology based on simulated data and on the analysis of spectrometer data.

Theory: While we have proposed the quantile extension to PLS, and shown by example the interest, we have not established theoretical properties of the estimates. In particular the verisimilitude of statements related to quantile coverage properties such as $P(\{(X, Y) : Y \leq \gamma_1 T_1 + \cdots + \gamma_k T_k\}) = \tau$, when sample estimates are substituted for population quantities remain to be demonstrated. Estimation variability is important in small samples and needs to be taken into account. For instance, the out-of-sample predictions for 25% of the data from the quantile regressions estimated from 75% of the data, while in the right direction, are significantly different from the predicted number in the inter-quartile intervals. There does seem to be some pattern in deviations from predicted which suggest certain areas to research. We would point out that PLS itself suffers from a similar lack of provable finite sample properties, and while there is some asymptotic (in n) theory available it is not especially relevant. Further work, both numerical and analytical, is needed.

A feature of the small sample sizes typical in chemometrics is the variability of the fitted quantile regressions evident, say, in Fig. 2b, for $\tau = 0.25, 0.75$, though it is no larger than the variability of the fitted PLS regression. There are no QR algorithms, as yet, that simultaneously estimate quantiles for a set of percentiles τ, see the discussion to Portnoy and Koenker (1997).

Quantiles are only well defined when the response variable is one dimensional so that one complication of the PLS algorithm is avoided.

Relationship to robust estimation: In this paper we have focussed on estimating the quantile of the response variable. A rather different aim might be to develop a similar methodology to robustify the PLSR procedure. The setting $\tau = 0.5$ gives least absolute value regression for Y and lessens the influence of outliers in the response variable on the fitted model. Similar protection occurs for other values of τ. However, if extreme values of the covariates are a concern then a better approach would be to robustly estimate the joint variance matrix of these variables directly, see Hubert and Branden (2003) and the references therein.

Remarks on the quantile covariance: The quantile covariance cov_τ is used at two points in the PQR algorithm: to compute $\mathsf{cov}_\tau(Y, X_i)$ for the component loadings and to calculate the predictor $E_\tau(Y|T_1, T_2, \ldots, T_k)$ at the final stage. The overall objective is to best estimate the quantile of Y, corresponding to the percentile τ, using ancillary information from the covariates. The example in the Appendix makes clear that cov_τ estimates the right quantity, and also shows that it may differ at different τ. Consequently the PQR algorithm combines the covariates with proportional weights calculated at the relevant quantile. The final predictor is linear in the components so these weights directly feed through to the estimated quantile $E_\tau(Y|T_1, T_2, \ldots, T_k)$ in much the same way as the standard $\mathsf{cov}(Y, X_i)$ feeds through to the final PLS predictor.

Several sets of components: By comparison with PLSR an additional complication for PQR is that there may be more than one set of components to consider, one for each τ. Not only may there be different components at different τ but perhaps different numbers of components as well. If so this provides evidence that the distribution of (X, Y) is more complex than perhaps initially envisaged. While a single set of components for the estimation of all quantiles is desirable in principle, it should be an empirical finding of the analysis rather than a predicated outcome.

Additional inference: There are some features of PQR analysis that require additional investigation, especially those concerned with inference. The loadings plot with spectrometry data indicate that certain parts of the spectrum may play a small role in modifying the response, and testing that subsets of the loadings are zero in the fitted components is of interest. Tests for the equality of some or all loadings at different values of τ should govern how many sets of components need to be retained. The work of Koenker and Machado (1999) may be useful here.

Appendix: quantile covariance

The quadratic covariance used to predict the mean values for given covariates, is extended to a quantile covariance appropriate for predicting quantiles. We consider the case where X is p-dimensional, though the application to PQR only requires $p = 1$ for the calculation of the loadings, and $p = k$ for the calculation of the final predictor.

Linear least squares prediction, see Whittle (1983), Whittaker (1990), or Christensen (1991), formulated as an optimisation problem leads to certain identities for expectations and covariances. Suppose that X is p-dimensional random vector and Y is a random scalar, with a joint distribution in which the relevant moments exist, then the identities

$$\mathsf{E}(Y) = \arg\inf_{\alpha} \mathsf{E}(Y - \alpha)^2, \tag{8}$$

$$\mathsf{cov}(Y, X)^T = \arg_{\beta}\inf_{\alpha, \beta} \mathsf{E}(Y - \alpha - \beta^T \mathsf{var}(X)^{-1}[X - E(X)])^2, \tag{9}$$

are well known. The predictor (2) above is the solution to the optimisation problem implied by (9). As such the optimising values of α and β obtained are summaries of the joint distribution of (X, Y) related to centering and covariation, and are not measures on the conditional distribution of Y given X.

The quantile expectation and quantile covariance are defined by replacing the quadratic loss function on the right of (8) and (9) by the quantile loss function at (1) giving

$$\mathsf{E}_{\tau}(Y) = \arg\inf_{\alpha} \mathsf{E}\, \rho_{\tau}(Y - \alpha), \tag{10}$$

$$\mathsf{cov}_{\tau}(Y, X)^T = \arg_{\beta}\inf_{\alpha, \beta} \mathsf{E}\, \rho_{\tau}(Y - \alpha - \beta^T \mathsf{var}(X)^{-1}[X - E(X)]). \tag{11}$$

The expectations on the right of (10) and (11) are standard. Retaining the multiplication $\mathsf{var}(X)^{-1}$ in (11) maintains the right linearity properties for the covariance.

The linearity properties of these operators are summarised in

$$\mathsf{E}_\tau(\alpha + \beta Y) = \alpha + \beta \, \mathsf{E}_\tau(Y),$$
$$\mathsf{cov}_\tau(\alpha + \beta Y, a + BX) = \beta \, \mathsf{cov}_\tau(Y, X)B^T,$$

where the constants α, $\beta \geq 0$, a and B are of the appropriate dimension, and follow from the linearity property of the quantile loss function

$$\rho_\tau(\beta y) = \beta \rho_\tau(y) \quad \text{for} \quad \beta \geq 0.$$

Several standard properties of the quadratic expectation and covariance do not hold for the quantile covariance. In particular, the expectation $\mathsf{E}_\tau(Y_1 + Y_2)$ is not additive, and the covariance cov_τ is not symmetric. In the quadratic case the values of α that minimises the implied objective function in (8) and in (9) are the same, but this is not the case with the quantile objective functions, and the linear quantile predictor of Y from X is

$$\mathsf{E}_\tau(Y|X) = \widehat{\alpha} + \mathsf{cov}_\tau(Y, X) \, \mathsf{var}(X)^{-1}[X - E(X)], \tag{12}$$

where $\widehat{\alpha}$ is the minimiser in (11), rather than

$$\mathsf{E}_\tau(Y) + \mathsf{cov}_\tau(Y, X) \, \mathsf{var}(X)^{-1}[X - E(X)].$$

This point may be argued by deriving the normal equations for (11).

Take the simple case where $p = 1$, $\mathsf{E}(X) = 0$ and $\mathsf{var}(X) = 1$, the normal equations for α and β are

$$\tau = \mathsf{E}I_{Y-\alpha-\beta X \leq 0} \quad \text{and} \quad \tau \, \mathsf{E}X = \mathsf{E}XI_{Y-\alpha-\beta X \leq 0}, \tag{13}$$

so that separate estimation of α and β is not in general possible.

Figure 7 illustrates the notion of quantile covariance.

In the upper left panel the joint distribution of (X, Y) is a uniform distribution on a polygon with the indicator of the region below an arbitrary line $y = \alpha + \beta x$ shaded, where here $\beta < 0$. The point of the Figure is to emphasize that the quantile covariance is a measure in the joint distribution of (X, Y) and not in the conditional distribution of Y given $X = x$. Symmetry considerations indicate that choosing α and β to minimise (11) such that probability of the shaded region is τ, leads to the covariance $\mathsf{cov}_\tau(Y, X)$ taking the value equal to the slope of the polygon, whatever the value of τ.

In the upper right panel is a "triangular" distribution composed of two polygonal regions having equal probabilities, and a uniform distribution within both regions. The quantile regression lines show at $\tau = 1/3$ that $\mathsf{cov}_\tau(Y, X) = 0$, while at $\tau = 2/3$ that $\mathsf{cov}_\tau(Y, X) > 0$. In fact $\mathsf{cov}_\tau(Y, X) = 0$ for all $\tau < 1/2$ and $\mathsf{cov}_\tau(Y, X) > 0$ for all $\tau > 1/2$ in this example.
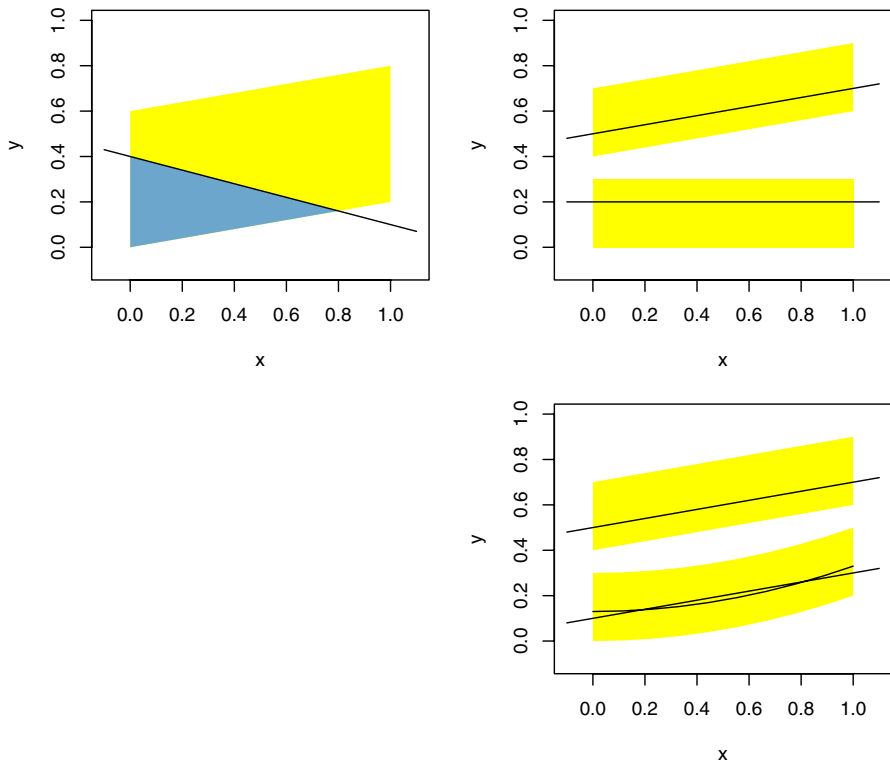
Fig. 7 Contour plots of three joint distributions of (X, Y). Upper left panel: a polygonal uniform distribution with the indicator of the region below an arbitrary line $y = \alpha + \beta x$. Upper right panel: a uniform distribution over two polygonal regions with equal probability, together with the different quantile regression lines at $\tau = 1/3$ and $\tau = 2/3$. Lower right panel: a uniform distribution over two regions with equal probability, together with the conditional quantile function and linear quantile regression at τ ($< 1/2$), and another at $\tau = 2/3$

More generally, when $p = 1$ and X is standardised, $\text{cov}_\tau (Y, X)$ is interpretable as the slope of the line minimising (11) in the joint distribution of (X, Y) for a given τ. The example also may be used to indicate why it would be wrong to base a measure of covariance on minimising (11) without the intercept parameter as

$$\mathsf{E}\, \rho_\tau (Y - \beta^T \text{var}(X)^{-1}[X - E(X)])$$

would lead to a line with a given coverage probability in the joint distribution, but with no obvious connection to the quantile of Y.

The third example shows that the linear quantile predictor may well differ from the conditional quantile function. In the lower right panel the lower polygon of the "triangular" distribution is replaced by a quadratically shaped region, again with uniform probability over this region. For $\tau < 1/2$ the conditional quantile is a quadratic function of X, while the linear quantile predictor is necessarily linear in X. The probability, in the joint distribution, of falling below the linear quantile regression is τ

Finally we make two points: (i) we note that it is possible to show that $Y \perp\!\!\!\perp X$ if and only if $\text{cov}_\tau(Y, X) = 0$ for all τ; and (ii) the notation of linear expectation in linear least squares and linear quantile prediction is useful, but does not lead to a more general calculus of either least squares or of quantile expectation.

Computation

For a given distribution p_{XY}, with $p = 1$, computation of $\text{cov}_\tau(Y, X)$ requires the analytic solution to (13). For samples from p_{XY} the solution needed is to $\tau = \mathsf{E}_n I_{Y-\alpha-\beta X \leq 0}$ and $\tau \mathsf{E}_n X = \mathsf{E}_n X I_{Y-\alpha-\beta X \leq 0}$, which in coordinate notation become

$$\tau = n^{-1} \sum_{i=1}^{n} I_{y_i-\alpha-\beta x_i \leq 0} \quad \text{and} \quad \tau \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i I_{y_i-\alpha-\beta x_i \leq 0}.$$

These, and the more general optimisation problem (11) with E_n replacing E, are solved numerically.

References

Andersen C, Rinnan A (2002) Distribution of water in fresh cod. Lebensm. Wiss. U.Technol 35:687–696

Christensen R (1991) Linear models for multivariate, time series, and spatial data. Springer, New York

Dodge Y, Jureckova J (2000) Adaptive regression. Springer, New York

Frank I, Friedman J (1993) A statistical view of some chemometrics regression tools. Technometrics 35:109–135

Hastie T, Tibshirani R, Friedman J (2001) Elements of statistical learning. Springer, New York

Helland I (1988) On the structure of partial least squares regression. Commun Stat Simulation Comput 17:581–607

Helland IS (2001) Some theoretical aspects of partial least squares regression. Chem Intell Laboratory Syst 58:97–107

Hubert M, Vanden Branden K (2003) Robust methods for partial least squares regression. J Chemometrics 17:537–549

Hwang J, Nettleton D (2003) Principal components regression with data-chosen components. Technometrics 45:70–79

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33–50

Koenker R, Machado J (1999) Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc 94:1296–1310

Longley J (1967) An appraisal of least-squares programs from the point of view of the user. J Am Stat Assoc 62:819–841

Martens H, Martens M (2000) Multivariate analysis of quality. Wiley, New York

Martens H, Naes T (1989) Multivariate calibration. Wiley, Chichester

Portnoy S, Koenker R (1997) The gaussian hare and the laplacian tortoise: computability of squared-error vs. absolute-error estimators, with discussion. Stat Sci 12:279–300

Whittaker J (1990) Graphical models in applied multivariate statistics. Wiley, Chichester

Whittle P (1983) Prediction and regulation by linear least-square methods. University of Minnesota Press, Minneapolis