MODELING AND FITTING QUANTILE DISTRIBUTIONS AND REGRESSIONS

Warren G. Gilchrist
Sheffield Hallam University. Sheffield. UK.

SYNOPTIC ABSTRACT

This paper explores the developing use of distributions defined by their quantile functions in both distributional and regression situations. It considers issues of both model construction and fitting. In particular it advocates the use of the ordered data in distributional estimation and seeks to emphasise that regression based on the use of quantiles is far broader than that usually encompassed by the phrase quantile regression. The emphasis is on parametric rather than non-parametric approaches.

Key Words and Phrases: distributions; lambda distributions; quantile functions; model construction, estimation, .parametric quantile regression.

1. INTRODUCTION.

Sir Francis Galton in the last half of the 19th century made two significant contributions to statistics. One concerned the handling of data, the other the parallel modelling of populations. On the data front he used the ordering of data as part of its analysis. For example to study the distribution of men's heights on a trip in Africa, he would get men from a tribe to stand in a line in height order, shortest on the left, tallest on the right, this gave a clear view of the distribution and led naturally to concepts of median, m, as a measure of position, and inter-quartile range, iqr, as a measure of spread. For populations a theoretical plot of x against the probability, p, of $X \leq x$ gave what Galton called the Ogive for the Normal distribution, Galton (1883) and Gilchrist (2005), and what Parzen termed the Quantile Function, QF, (Parzen (1979)). Thus in general Galton's model to describe distributions took the form $x = Q(p)$. Galton's term Ogive was later misused to refer to the CDF, $\Phi(x)$, of the Normal distribution. For the Normal Galton used the form $Q(p) = \lambda + \eta N(p)$, where N(p) is the quantile function for the standard Normal and is the function given in books of tables of statistics. The parameters λ and η have the meaning of position and scale parameters, for Galton the median, M, and the 'probable deviation' (IQR/2)

Galton's approach was almost forgotten and distributions for most of the next hundred years were defined and studied in terms of their distribution functions, F(x), and not their quantile functions. A return to a quantile formulation of distributional models began in the 1960s. A paper by Hastings et al (1947) suggested a formula for generating random variables with a useful and flexible shape. This was developed further by Tukey (1962) as a distribution and later called the Tukey-lamda distribution. The Lambda term arose from the use of subscripted lambdas as the parameters in the model. In this paper we wish to emphasise the differing nature and functions of the various parameters, so different symbols will be used for each parameter. In our notation the formula for the Tukey-lambda distribution is:

$$\lambda + (\eta/\alpha)[p^{\alpha} - (1 - p)^{\alpha}]$$

If the value p is replaced by random numbers from a uniform distribution on (0,1) then values are obtained from a flexible symmetric distribution with mean λ and scale parameter η. The parameter α controls the shape of the tails of the distribution and is thus referred to as a shape parameter. If a distribution is defined by the Cumulative Distribution Function (CDF), p = F(x), then it may be possible to invert the relation to give x = Q(p). But here there is no simple CDF. The function Q(p), the quantile function, provides an alternative way of defining a distribution and the Tukey-lambda distribution was the first to be defined solely on this basis. Thus we have a formal definition as

$$Q(p) = \lambda + (\eta/\alpha)[p^{\alpha} - (1 - p)^{\alpha}].$$

This distribution was later generalised to

$$Q(p) = \lambda + \eta[p^{\alpha} - (1 - p)^{\beta}],$$

called the Generalized Lambda Distribution, GLD, originally due to Ramberg and Schmeiser (1972). As wll be seen this has two shape parameters. For a detailed study of this distribution see Freimer, Kollia, Mudholka and Lin (1988) and the book by Karian and Dudewicz (2000), abbreviated in this paper to K and D. This book also covers the extension of the GLD to cover a wider range of distributional shapes and also extends it to a bivariate form. A book by Gilchrist (2000) also gives some extensions of the GLD to provide an alternative four-parameter version and a five-parameter version, see below. Once the quantile function is used to define distributions it is realised that many distributions have simple forms of quantile function. For example we have:

The Uniform Distribution on (0, 1) is $Q(p) = p$.

The Exponential Distribution is $Q(p) = -\eta \ln(1 - p)$.

The Pareto Distribution on $(1, \infty)$ is $Q(p) = 1/(1 - p)^{\beta}$.

The Logistic Distribution on $(-\infty, \infty)$ is $Q(p) = \lambda + \eta \ln[p/(1 - p)]$.

The aims of this paper are:

(a) to bring together some of the previous work on the Lambda family of distributions, extend it and study the properties and use of the consequent extended family of GLD Models. Section 3.

(b) To look at other distributions that can be defined in quantile function form. Section 4.

(c) To consider issues of model construction and identification for such distributional models. Section 5.

(d) To survey some of the approaches to the fitting of distributions in quantile function form. Section 6.

(e) To consider the implications of the quantile function approach to distributional modelling. Section 7.

(f) To critically review the current approach to quantile regression. Section 8.

(g) To extend the ideas developed for distributional modelling into the province of parametric quantile regression. Section 9.

(h) Our survey is brought to a close by some conclusions in Section 10.

Before beginning the main tasks however we need briefly to introduce the notations and basic results to be used.

2. NOTATION AND QUANTILE FUNCTION PROPERTIES.

We introduced the quantile function form for a distribution as the inverse of the CDF, p = F(x). Often we will be dealing with distributions that have no simple explicit CDF. We therefore need to note that the formal definition of the QF is:

$$Q(p) = \inf\{x: \Pr(X \leq x) \geq p\}$$

The use of quantile functions to define statistical models moves emphasis away from classical measures of structure based on moments. However moments have very simple definitions in terms of the quantile function, thus the rth moment about the origin is

$$\mu'_r = \int_0^1 [Q(p)]^r \, dp.$$

More natural quantities for describing aspects of distributional shape in quantile form are the median, M=Q(0.5), the quartiles LQ=Q(1/4), UQ=Q(3/4), and the deciles, LD=Q(1/10), UD=Q(9/10). These lead to

(a) measures of spread given by the inter-quartile and inter-decile ranges, IQR =UQ-LQ, IDR=UD-LD.
(b) Measures of skewness are given by Galton's Skewness Coefficient GQ = [UQ+LQ-2M]/IQR and the Tail Ratio TR=[M-LD]/[UD-M].
(c) There are a variety of measures of tail shape. One simple one, that we will use later, is Moor's Kurtosis defined by

$$K = [(Q(7/8)-Q(5/8))+(Q(3/8)-Q(1/8))]/IQR$$

A further measure of tail shape is the ratio T = IQR/IDR

These measures, except the last three, were originally due to Galton. The books by Gilchrist and K & D both discuss measures of shape but use slightly different sets as shown in Table 1. To study the shape of such distributions further we note that the derivative of the quantile function, the quantile density function, q(p), relates to the derivative of the CDF, the pdf, f(x), by the relation f(Q(p))q(p) = 1, since

$(dp/dx)(dx/dp) = 1$. So, in an obvious notation, $f_p(p) = 1/q(p)$, where $f_p(p)$ can be called the p-pdf or, by Parzen (1969), the density quantile function.

3. THE LAMBDA FAMILY OF DISTRIBUTIONS.

3.1. Introduction

By far the most used distributions in quantile form have been the Lambda distributions, see the survey in Karian and Dudewicz (2000). The aim of this section is to explore this family, building up by the number of parameters involved, and looking in particular at the interpretation of these parameters in describing distributional shape.

3.2. The Three Parameter Tukey-lambda Distribution.

Consider first the form of the simple Tukey-lambda distribution.

$$Q(p) = \lambda + (\eta/\alpha)[p^{\alpha} -(1 - p)^{\alpha}].$$

There are three parameters: λ is the median, as is seen by putting $p = 0.5$, so this acts as a position parameter, η is a scale parameter as is seen by observing that if we set the median as zero any change in η simply multiplies the quantile accordingly.

To consider the third parameter, first note that the 'standard form' of the model, (setting $\lambda = 0$ and $\eta = 1$) can be written as $(1/\alpha)[p^{\alpha}] + (1/\alpha)[-(1 - p)^{\alpha}]$. A general feature of a quantile model is that if $R(p)$ is a quantile function on $(0, c)$ then $-R(1-p)$ represents the reflected distribution on $(-c, 0)$, where c is some constant, the reflection taking place in the $x = 0$ axis. Thus we can see that for positive α the model represent a distribution $R(p) = (1/\alpha)[p^{\alpha}]$, the Power distribution , on $(0, 1/\alpha)$, and $(1/\alpha)[-(1 - p)^{\alpha}]$ the Reflected Power distribution, on $(-1/\alpha, 0)$. The Tukey-lambda distribution thus combines, adds, the two, with the reflected distribution dominating for small p, i.e. the extreme left tail, and the Power dominating for p approaching one, i.e. the extreme right tail. The parameter α thus controls the shape of the power distribution, and is referred to as the shape parameter. For the power distribution the p-pdf is $f(p)=p^{1-\alpha}$. It will be evident that for $\alpha < 1$, this gives an increasing shape, and for $\alpha > 1$ it shows a decreasing curve. For $\alpha = 1$ the distribution is uniform. Combining the

Power and its reflection thus gives a U shaped curve for $\alpha < 1$, a uniform for $\alpha = 1$ and a unimodal distribution for $\alpha > 1$. The latter includes the logistic, approximate Normal and a heavy tailed form, according to the value of α. The multiplier $(1/\alpha)$ is used to ensure a proper increasing quantile function for the case where α is negative. In this case the term $(1/\alpha)[-(1 - p)^{\alpha}]$ is the quantile function for a Generalised Pareto distribution on $(-1/\alpha, \infty)$. The term $(1/\alpha)[p^{\alpha}]$ is the Reflected Pareto, and the Tukey-lambda thus has Pareto shaped tails in both extremes. Figure 1 shows how the two tail quantile functions are combined to give the Tukey-lambda distribution. In the limiting case of $\alpha = 0$ the distributions becomes the Logistic distribution:

$$Q(p) = \lambda + \eta \ln[p/(1 - p)].$$

Some of the main properties of the symmetric Tukey-lambda distribution are shown in Table 2.

3.3. Four Parameter Distributions.

We now consider generalising the Tukey-lambda distribution to the non-symmetrical case. In particular we look at three extensions based on the Power and Pareto forms of the three parameter distributions:

The Generalized Lambda Distribution, GLD

The first natural generalisation is to allow for two shape parameters, α for the left tail and β for the right tail. This gives as the common form:

$$Q(p) = \lambda + \eta[p^{\alpha} - (1 - p)^{\beta}];$$

this distribution was first developed by Ramberg and Schmeiser (1972, 1974.). It was first used as a means of simulating a flexible family of distributions. The distribution was further explored by Ramberg, Tadikamalla, Dudewicz and Mykytka (1979). A detailed study of the GLD distribution is given in the book by Karian and Dudewicz (2000). One feature of the distribution explored in the book is that by suitable choice of parameters the distribution provides a very good approximation to many common distributions. Table 3 gives a summary of some of the main properties of this distribution.

A small problem with the common form of the GLD is that which is the right tail and which the left switches as the parameters become negative. There is also no clear meaning in the limiting case of parameters with zero values. A form that deals with these issues, and gives exponential tails in the limiting cases, was proposed by Freimer et al (1988) as:

$$Q(p) = \lambda + \eta[(p^{\alpha} - 1)/\alpha - \{(1 - p)^{\beta} - 1\}/\beta].$$

It will be seen that the term $(p^{\alpha} - 1)/\alpha$ represents a Power distribution in the left hand tail in the range $(-1/\alpha, 0)$. The term $- \{(1 - p)^{\beta} - 1\}/\beta$ is its reflection to the right with parameter β and support $(0, 1/\beta)$. Thus the two parameters are shape parameters for the left and right hand tails respectively. Negative values still keep the correct left side-right side notation. Table 4 indicates the shapes of the tails for differing tail parameters.

The GLD has had wide use in application areas such as hydrology, meteorology, pharmacology, etc. One particular problem of the GLD is the fact that it does not cover all possibilities of the skewness-kurtosis, though it covers many more than most other families. Karian and Dudewicz (2000) develop a form of the Beta distribution that can be used to cover the missing areas. This does not solve the problem that skewness and kurtosis do not always exist, even though the models do.

The Skew Lambda Distribution, SLD

One aspect of the GLD is that the parameters α and β determine not only the tail shape but the relative weight given to the two tails, i.e. the skewness. An alternative approach is to keep to the single tail shape parameter but introduce a parameter that controls the skewness, via the relative weights of the two tails. A form that does this was proposed by Gilchrist (2000) and is:

$$Q(p) = \lambda + (\eta/2\alpha)[(1 - \delta)\, p^{\alpha} - (1 + \delta)\, (1 - p)^{\alpha}].$$

A special case of this for $\alpha = 0$, was suggested in Gilchrist (1997):

$$Q(p) = \lambda + \eta[(1 - \delta) \ln p - (1 + \delta) \ln(1 - p)],$$

which is a skew form of the logistic distribution. In both these models we have introduced a parameter δ to control the skewness. The above form for the lambda distribution is a small reparameterization of the model:

$$Q(p) = \lambda + \theta p^{\alpha} - \phi(1 - p)^{\alpha}.$$

It will be seen that this comes from the Friemer form with $\beta = \alpha$ and weights applied to the two tails. Note that in this original form there are three linear parameters and one non-linear. Table 2 shows the main properties of this distribution. It will be seen that δ determines the relative weight of the tails, for α <1 this determines the skewness, with positive δ giving positive skewness. For larger α, though the relative weights are correct, the interpretation is normally of a negative skewness. Thus we have four natural parameters governing position, scale, skewness and tail shape.

The Power-Pareto Distribution, Po*PaD

For suitable parameter values the GLD and SLD has support over all the axis. In many situations however variables are inherently positive. For a large mean and a small coefficient of variation (standard deviation/mean) there is sometimes little concern for using the GLD where the distribution covers negative values. However this is often not the case so we need to use models that are inherently positive. The Power and Pareto distributions used in the construction of the GLD and SLD are both inherently positive. By putting constraints on the parameters of the GLD we can make one term a right tailed Power and the other a right tailed Pareto. However another way of combining the two is to replace addition of quantile functions by multiplication, which is legitimate for positive valued distributions. This gives the distribution:

$$Q(p) = \lambda + \eta[p^{\alpha} /(1 - p)^{\beta}]. \quad \alpha, \beta > 0.$$

This distribution we call the Power-Pareto, Po*Pa, see Gilchrist (2000). Though it is often possible to fit the GLD to positive variables it seems that the distributions of such variables must inherently reflect their positive nature, so is more natural to use the Po*Pa distribution. It clearly includes the Power and the Pareto as special cases. Table 5 gives a summary of its main properties.

3.4. The Five Parameter Lambda Distribution, FPLD

In the two main four parameter models we mixed skewness and tail shape with the GLD and we allowed for skewness but forced identical tail shapes in the SLD. Let us now admit that any extensive set of data will enable us to distinguish five basic features: position, spread, skewness, left tail shape and right tail shape. It seems therefore reasonable to use models with five parameters that can be linked with each of these in the population model. The natural form for this is:

$$Q(p) = \lambda + (\eta/2)[(1 - \delta)\, p^{\alpha} - (1 + \delta)\, (1 - p)^{\beta}].$$

The interpretation of the five parameters is as we have seen in previous models. It is a strange phenomena that over the history of statistics statisticians seem to have been quite happy to use many parameters in, for example regression, yet they have kept to two or three parameters in most distributional models. Clearly there are increasing difficulties in estimation as the number of parameters increases, however in our computer age large data sets of over 100 observations are common, so there should be no theoretical reason for not using five parameter distributional models.
It may be useful to use a Friemer form for the distribution giving

$$Q(p) = \lambda + \eta[(1-\delta)(p^{\alpha} - 1)/\alpha - (1+\delta)\{(1 - p)^{\beta} - 1\}/\beta].$$

An alternative to this weights each tail to a median of one, so that the skewness is better reflected by the δ:

$$Q(p) = \lambda + \eta[(1-\delta)(p^{\alpha} - 1)/(1/2^{\alpha} -1)- (1+\delta)\{(1 - p)^{\beta} - 1\}/(1/2^{\beta}-1)].$$

Some properties of the FPLD are given in Table 6. A basic form that we will use later emphasises that we have three linear parameters and two non-linear parameters in the model:

$$Q(p) = \lambda + \theta p^{\alpha} - \phi(1 - p)^{\beta}.$$

For both the main terms to correspond to valid, non-decreasing, quantile functions we require that $\theta\alpha$ and $\phi\beta$ are positive. However this is not always the case. For example in a paper by Tarsitano (2005) this model is fitted to twelve sets of grouped data. Of these all 12 have a valid second term but four have decreasing first terms. The resultant distribution is still valid when the $\theta\alpha$ is negative but such that the rate of decrease of θp^{α} is dominated by the rate of increase of $-\phi(1-p)^{\beta}$. Figure 2 shows the parameters and the shapes of the components for one of these data sets. Thus though model construction suggests that we use all valid components, the evidence from the data used by Tarsitano is that in practice we may only require the resultant quantile function to be valid.

4. SOME OTHER QUANTILE FUNCTION FAMILIES

We have focussed thus far in this paper on the Lambda Family as a wide family of distributions based on a quantile function definition. It should be noted however that there are other families with wide application and simple explicit forms of quantile function, and some overlap, which have been studied in the literature. Johnson, Kotz and Balakrishnan (1994 &95) discuss most of these in terms of their pdfs and moment properties. The quantile functions below give the standard forms leaving out position and scale parameters:

The Wakeby Distribution

This distribution is much used in hydrology and is one of the few five parameter models in common use. It is composed of the linear combination of two Power or Pareto distributions: It can be expressed in a variety of ways:

$$Q(p) = \lambda - \theta(1-p)^{\alpha} - \phi(1-p)^{\beta}.$$

As happens with the Lambda distributions the simplest forms lead to complicated conditions on the parameters to ensure a valid quantile function. See for example the discussions in K and D of the GLD and in Tarsitano (2000) for the Wakeby. In a similar fashion to the GLD we can avoid this problem by generalising to:

$$S(p) = \lambda + \eta[\delta \{1-(1-p)^{\alpha}\}/\alpha + (1-\delta) \{1-(1-p)^{\beta}.\}/\beta]$$

Here it is clear that η provides a scale parameter and δ provides the weighting of the terms. The quantile function is valid for all α and β. The support for the distribution is

(λ , $\lambda + \eta\delta/\alpha + \eta\delta)/\beta$) for positive α, β and (λ ∞) otherwise. Notice that setting λ =0 we have a flexible four parameter model with support (0, ∞) for a negative shape parameter.

The Extreme Value Distributions

As their name implies these are a group of distributions used, for example, for looking at extreme weather or flood data. These have the forms:

Type I	$-\ln[-\ln(p)]$	$-\infty < x < \infty$.
Type II	$[-\ln(p)]^{-\beta}$	$0 < x < \infty$.
Type III	$-[-\ln(p)]^{\beta}$	$-\infty < x < 0$
Generalized EV	$-[(-\ln(p))^{\beta} - 1]/\beta$	$-\infty < x < 1/\beta$

The Generalized Kappa Distribution

This is a distribution that has been used in the study of incomes, its standard form is

$$S(p) = [1 - \{(1 - p^{\alpha})/\alpha\}^{\beta}]/\beta.$$

A reflection of this has been called a General Power distribution. An example of this distribution is shown in Figure 3. It will be seen that it has a Type I Extreme Value Distribution as its limit for zero shape parameters.

The Burr Family

This family is based on the solution of a differential equation. However a sub-set of useful forms have simple and related quantile forms. The basic forms assume positive parameters:

The Power-Pareto $[p^{\alpha} /(1 - p)^{\beta}]$. $0 < x < \infty$.

Special case	$[p/(1-p)]^{\alpha}$.	$0 < x < \infty$.
Burr II	$\ln [p^{\alpha}/(1-p^{\alpha})]$.	$-\infty < x < \infty$.
(Also called a Generalised Logistic).		
Burr III	$[p^{\alpha}/(1-p^{\alpha})]^{\beta}$	$0 < x < \infty$.
Burr IV	$[(1-p^{\alpha})/p^{\alpha})]^{\beta}$	$0 < x < \infty$.
Generalized II/III	$[\{p^{\alpha}/(1-p^{\alpha})\}^{\beta} - 1]/\beta$	$-1/\beta < x < \infty$.
Burr X	$[-\ln(1-p^{\alpha})]^{0.5}$	$0 < x < \infty$.
Burr XII	$[\{1-(1-p)^{\alpha}\}/(1-p)^{\alpha}]^{\beta}$	$0 < x < \infty$.

The Asymmetric Laplace Distribution

One of the earliest defined distributions is the Laplace. This is formed of an Exponential and a Reflected Exponential, back to back, its PDF has the form:

$$f(x) = (1/2) \exp - 2|x|/\theta.$$

In quantile form this is:

$$Q(p) = \begin{cases} (\theta./2) \ln 2p & 0 \le p \le ½ \\ -(\theta./2) \ln 2(1-p) & ½ \le p \le 1. \end{cases}$$

This is generalised to give an asymmetric form:

$$Q(p) = \begin{cases} s\theta. \ln p/s & 0 \le p \le s \\ -(1-s)\theta \ln \{(1-p)/(1-s)\} & s \le p \le 1. \end{cases}$$

This still has back to back exponential form, but with probabilities s and 1-s to the left and right of the peak. It may be noted for later use that the pdf of this form may be written as:

$$f(x) = (1//\theta) \exp - [\rho_s(x)/\{s(1-s)\theta\}].$$

Where $\rho_s(x)$ is a function called the 'check' function defined as

$$\rho_s(x) = sxI_{(-\infty,0)}(x) - (1-s)xI_{(0,\infty)}(x) \text{ , where } \quad I_{(A)}(x) = 1, x \subset A, = 0 \text{ otherwise.}$$

The Kamps Distributions

This family was first considered by Kamps (1991) and is of interest because it is defined in terms of its quantile density function, q(p). This is given by

$$q(p) = \eta p^{\alpha}(1-p)^{\beta}.$$

It will be seen that setting the values α=0, β=-1; α=0, β<-1; $\alpha \neq$-1, β=0; α= β= -1 and integrating, to get Q(p), we derive the quantile functions for the Exponential, Pareto, Power and Logistic distributions as special cases.

5 CONSTRUCTING DISTRIBUTIONS

The classical approach to handling distributions is to choose an appropriate distribution from the large and ever expanding catalogue of distributions available, e.g. in Johnsom, Kotz and Balakrishnan (1994,1995). If this does not work conveniently the data is transformed before it is used. It will be noticed however that in sections 3 and 4 we have repeatedly seen distributions whose form is a development of some simpler distribution. Implicitly we have started with simple distributions and constructed more complex forms. This has been done by addition, as in the lambda distributions, by multiplication, as in the Power-Pareto, by raising the quantile function to a power, as in going from exponential to Weibull, and by raising p to a power, as in going from the Logistic to the Burr II. When using powers we have also made use of the generalising technique of replace terms like p^{α} by $(p^{\alpha} - 1)/\alpha$ so that the model can deal with the case where α becomes negative or zero, when this expression become ln p. Let us look briefly at each of these model construction techniques.

Adding Quantile Functions

The simplest approach to this is illustrated by the Skew Lambda distribution:

$$Q(p) = \lambda + (\eta/2\alpha)[(1 - \delta)\, p^{\alpha} - (1 + \delta)\,(1 - p)^{\alpha}],$$

which is based on the Power distribution, for α positive, or the Pareto distribution for α negative. Replacing these by some general right tailed distribution on $(0,\infty)$, $S(p;\alpha)$ we have the general form made by combining a distribution with its reflection with a weighting controlled by δ, which we attach as $(1+ \delta)$ to the right tailed form:

$$Q(p) = \lambda + (\eta/2)[(1 + \delta)\, S(p;\alpha) - (1 - \delta)\, S(1-p;\alpha)].$$

A little algebra will show that if S(p) has Median M_S, inter-quartile range, IQR_S , a positive skewness coefficient G_S and a Moor's Kurtosis K_S then the constructed distribution has:

$$M = \lambda + \eta\delta M_S,\ IQR = \eta IQR_S,\ G = \delta G_S \text{ and } K = K_S.$$

Thus, noting that this last depends only on the shape parameter α, each of the four parameters directly determine the corresponding shape measure.

Multiplying Positive Quantile Functions

A further group of quantile functions can be constructed by the multiplication of the quantile functions, H(p) and K(p), for two positive distributions. It makes some sense to standardise the two in some way to make them comparable before multiplication, which is easily done by requiring that for each the median is one. The effect of the multiplication is to stretch the distributions relative to this median.

Example

Suppose $H(p) = (2p)^{\alpha}$ and $K(p) = 1/[2(1 - p)]^{\beta}$, standardised power and Pareto distributions then the product is the Power-Pareto, discussed above, with a minor scale change, i.e. $Q(p) = (2p)^{\alpha} /[2(1 - p)]^{\beta}$, which has median 1 and support $(0, \infty)$.

Transforming p

As both p and p^{α} lie in (0, 1) and for positive α the transformation is monotone the distribution $Q(p^{\alpha})$ is a valid distribution with the same support as Q(p). For $\alpha > 1$ it will be seen that the effect of the transformation is to move probability to the right within the support. For $\alpha < 1$ probability is moved to the left.

Example

For the Logistic distribution this becomes

$$Q_T(p) = [-\ln\{ p^{\alpha}/(1 - p^{\alpha})\}],$$

See Figure 4 for an example.

Example

For the Weibull Distribution we have: $Q(p) = [-\ln(1-p)]^{\beta}$ which transforms to

$$Q_T(p) = [-\ln(1 - p^{\alpha})]^{\beta} \quad ,$$

which is a general version of the Burr X distribution, referred to above. See Figure 5 for an example.

Transforming Q(p)

As for any non-decreasing transformation T(.) the distribution of T(X) is simply T[Q(p)], we have a simple way to carry out modifications. In classical statistics this compares with transforming the data. However there are important differences.

(a) Using quantiles we seek to keep to the distribution of the original variable in its natural metric. For example one frequently sees economic data transformed with a logarithm so that the Normal distribution may be used. The natural log-Normal distribution in quantile form is exp{N(p)}, where N(p) is the Normal quantile function. This distribution usually occurs because of the operation of the multiplicative version of the Central Limit theorem in which the many natural innovations multiply each others effects.

(b) When transforming data one has the effect of transforming the complete model, which in many regression situations may already be quite complex. Trying to separate the various effects on the raw data using the resultant fitted transformed model is often almost impossible. With quantile models we need only transform the elements of concern. Thus we may use the transform as $Q(p) = \lambda + \eta T\{S(p)\}$, keeping to the natural position and scale parameters and a model for the raw data.

A look at the distributions mentioned above will show that many relate to others by straight transformations such as ln( ) or $[\ ]^{\beta}$ for positive β.

Example

The Pareto distribution has quantile function in standard form $Q(p) = 1/(1-p)^{\beta}$.

Taking log transformation gives $Q_T(p) = -\beta\ln(1-p)$, which is the Exponential distribution.

Using Graphics to Guide

It is hoped that at this stage it will be clear that a quantile definition of distributions moves the focus from the classical one of choosing a distribution from a catalogue to that of thinking about the construction of appropriate distributions. As a first guide to doing this we might consider the answers to the following two questions:

(a) What do we know conceptually about the variable(s) being modelled? For example is it on a doubly infinite range or is it essentially positive? Is anything known of the mechanism that generates it?

(b) What do we know empirically about the distribution of the data? What are its main characteristics? For example it presumably has a central position and a spread but does it show skewness? Has it long or short tails? What shapes are the tails of the data?

Finding the answers to these questions is frequently aided by graphical analysis. To look at the data graphically there are a range of plots that can be used.

(a) Plot the ordered data against the order number, $x_{(r)}$ v r. This often confirms the type of data being handled. For the benefit of later analysis it is best, for data with a doubly infinite support, to standardise by subtracting the sample median, m and dividing by 2IQR. For data with support on $(0, \infty)$ then we divide by the median, as for these distributions this acts as a scale factor. When comparing with models this enables us to ignore the λ and η terms and focus on the form of S(p). Often plots become clearer when such standardisations has been carried out.

(b) A plot of $x_{(r)}$ against some simple quantile function, $S(p_{(r)})$. The $p_{(r)}$ are the rankits, i.e. the expected values of the order statistics $E(x_{(r)})$, e.g. for the uniform distribution, r/(n+1). Using the argument that medians are easier to judge by eye, one could use the 'median rankits', i.e. the population medians of the order statistics, $M_{(r)}$ These are given, exactly, by:

$$M_{(r)} = Q(p_{(r)}) = Q(\text{Betainv}(0.5,r,n+1-r)),$$

where Betainv(.) is the inverse Beta function. A correct choice for the distribution and its parameters for S(.) will lead to a straight line plot of $x_{(r)}$ against $M_{(r)}$. Figure 6 shows a set of soil data used for illustration below plotted against the fitted values for the Wakeby distribution. Figure 7 shows the corresponding distributional residuals, i.e. $e_{(r)} = x_{(r)} - M_{(r)}$.

(c) If the data show a clear concave or convex curve when plotted against a simple S(p) then this quantile function can be replaced by $S(p_{(r)}{}^{\alpha})$. The value α can be chosen by maximising the correlation between $x_{(r)}$ and $S(p_{(r)}{}^{\alpha})$. Analysis of this nature is best done on a spreadsheet that enables the user to easily explore different models and parameter values with the appropriate graphs alongside. The objective of such study is to get the model S(.) to lead to a straight line plot of $x_{(r)}$ against $M_{(r)}$, the Q-Q or fit-observation plot. For positive distributions the transformation $Q^{\beta}(p)$ provides a further option to explore.

(d) Plots of the empirical p-pdf, $f(p)=dp/dx$, provide a rough plot that can be compared with the theoretical forms for standard distributions. A simple way to derive this is by approximating dp/dx by

$$d_{(r)} = (r/(n+1)-(r-1)/(n+1))/(x_{(r)} - x_{(r-1)}) = 1/[(n+1)(x_{(r)} - x_{(r-1)})].$$

The $d_{(r)}$ can then be smoothed and plotted against say $(r+0.5)/(n+1)$. To look at the corresponding population pdf we simply plot $1/q(p)$ against $Q(p)$ for a range of values of p in (0,1).

6. METHODS OF FITTING

6.1 Introduction

Most discussions of estimation for distributions have focussed on the standard textbook distributions, defined by their density functions, that typically have only one or two parameters. The literature contains little discussion of distributional models with four or five parameters. Yet as we have seen the descriptors of position, spread, skewness and tail shapes lead naturally to the consideration of models with more parameters. Little of the literature deals with fitting models defined by their quantile functions. It is therefore worth looking at several approaches to estimation in quantile defined models, see Gilchrist (2000) for details, and to discuss their relevance to models with four or five parameters. As indicated at the beginning of this paper we will consider particularly those methods that make use of the ordered data.

The Method of Moments

Though the Method of Moments has been used for fitting the GLD distribution, the high variability in the third and fourth moments, and even worse if requiring a fifth moment, suggests that the estimators obtained will have poor efficiency. To make matters worse for the GLD the relation between the third and fourth moment and the α and β parameters is a complex one leading to multiple solutions and in some regions of the moment space very flat surfaces, making for poor estimates. Though Tables are available in K and D the Method of Moments is not an appropriate method for models with these larger numbers of parameters. A further weakness of the method of moments is the non-existence of higher moments for some parameter values of many of the distributions considered here.

The Method of Percentiles.

The Method of Percentiles is based on matching population and sample percentiles. However there is no universal method for doing this matching. For example it could be done for the Five Parameter Lambda by choosing the parameters to equalise the median, quantiles and deciles for population and sample. The problem with this approach is that each of the five points depends on all five parameters. The alternative approach depends on the development of measures of position, spread, skewness and tail shape so that one can equate population to sample measures. The books by WGG and K & D both discuss this approach but use slightly different measures as shown in Table 1. The population values of these various quantities are given in the tables for some of the distributions we have been discussing. Sample values are found using the ordered data. The books by Harter and Balakrishnan (1996) and David and Nagaraja (2003) cover many aspects of using order statistics in estimation.

Example

For the symmetric lambda

$$Q(p) = \lambda + (\eta/\alpha)[p^{\alpha} -(1 - p)^{\alpha}]$$

we may equate population and sample values, indicated with upper case and lower case letters, for three simple percentile based measures. Thus we choose the parameters so that

$$M = \lambda = m\ ,\ IQR = 2\eta(3^{\alpha} -1)/\alpha 4^{\alpha} = iqr,\ T^* = (2.5)^{\alpha}\ [(3^{\alpha} -1)/(9^{\alpha} -1)] = t^*.$$

Solving the last equation gives the α to be used in the second equation.

Example

It is evident that the form of the skew lambda lends itself to this method of estimation. It is seen from Table 2 that equating the sample k to the population K, or t* to the population T*, which depend only on α, give estimates of the shape parameter. Equating measures of the sample and population skewness leads to the skewness parameter δ, the scale parameter η comes from the IQR and finally the position parameter λ comes from the equation that equates sample and population median.

It will be seen that for those distributions with a skewness parameter, GQ and GD give a more natural basis for estimation than TR. As a way of investigating the choice between the use of quartiles or deciles the Logistic distribution was used: $Q(p) = \lambda + \eta \ln(p/(1-p))$. Clearly the inter-decile range is more variable than the inter-quartile range. However the estimate of η requires division by a factor which is larger for the decile estimate, see Table 7, An approximation for the means and variances of the estimate of η showed that the decile based estimate was more efficient. To confirm this a study of 2500 samples of 20 simulated observations from a Logistic with parameters (0, 1) gave the standard errors for the estimates by the various methods shown in the Table 7. This confirms that the deciles provide the more efficient estimate.

A further issue that is explored in detail in Karian and Dudewicz (2000) is that in solving the percentile based equations for the GLD, and indeed for most multi-parameter distributions, there may be multiple solutions. Though it is rarely mentioned this also applies to other methods of estimation, including maximum likelihood. Where such multiple solutions are found the choice between them will depend on the practical interpretation of the parameters for the data and a graphical comparison of fit and observations. It may well be clear that some of the fitted models make no practical sense and show little resemblance to the data. It may also be that some simple parameter values give a fit comparable to more extreme values, so there is no loss in using the simple values in practice. As Tarsitano (2005) notes multiple solutions may be an opportunity as well as a problem.

The Methods of Distributional Least Squares and Least Absolutes

A visual check on the quality of fit of a distribution was obtained by plotting the ordered observations against some measure of their predicted positions using the fitted model, the Q-Q plot. This suggests that a measure of the quality of this fit is actually used for the estimation, Thus we might plot the ordered observations $x_{(r)}$ against $\mu_{(r)} = E(X_{(r)}$; parameters) and choose the parameters to get the best Least Squares fit, i.e we minimise

$$C_{DLS} = \sum(x_{(r)} - \mu_{(r)})^2$$

We call this method of estimation Distributional Least Squares. The expectation, $\mu_{(r)}$, the rankit, is available as an exact function for some distributions. Thus for the GLD the rankits are:

$$\mu_{(r)} = \lambda + \eta[B(n+1,\alpha)/ B(r,\alpha)- B(n+1,\beta)/ B(n-r+1,\beta)]$$

For many distributions there is no exact form, so an approximation could be used. For example $\mu_{(r)} = Q(r/(n+1))$, since r/(n+1) is the expectation for the r th uniform order statistic. For some distributions alternative functions of r and n give better estimates of the expectation. Notice that for many of the distributions discussed above the model contains both linear and non-linear parameters. Thus the use of DLS may involve a combination of explicit solutions with optimisation searches. This method has some drawbacks, as does the plot of $x_{(r)}$ against the fitted $\mu_{(r)}$ as a check on the distribution. The main obvious criticism is that as it stands it ignores the high correlations between order statistics and the heterogeneity of their variances. A paper by Lloyd (1959) applies Generalised Least Squares to deal with this in the case where there are only position and scale parameters and there are exact formulae for the variances and co-variances. However we will look at this in detail later and see that the naive approach above has some merits

A problem with the above DLS approach is that the order statistics in the tails have heavily skewed distributions, which makes a visual inspection of the graph potentially misleading and the slope of the fitted line biased. A way to partially avoid this is to use not expectation but the medians of the order statistics. The median rankit, $M_{(r)}$ is given by Q(BETAINV(0.5, r, n+1-r)). This is an exact result. Use of the median rankit has the benefit that the median, even of highly skewed distributions, is an intuitive quantity reasonably seen on a plot. The least squares criterion leads naturally to expectations. However it is the Least Absolute criterion that leads naturally to the use of medians. Thus the natural partner to the plot of $x_{(r)}$ against a fitted $M_{(r)}$ is the use of Distributional Least Absolutes, DLA. Thus we choose the parameters to minimise

$$C_{DLA} = \sum \left| (x_{(r)} - M_{(r)} \right|$$

An optimisation procedure easily leads to solutions, though it needs to be remembered that one does not always get a unique solution. Thus when using search techniques it is worth exploring around automatically found solutions. For large samples this is not really an issue. Table 8 shows the tabular layout of the steps for fitting by DLS and DLA.

Before leaving these two methods it is worth commenting again on their usual neglect of the heterogeneity and correlations in the ordered data. In classic statistics we start with the criterion C_{DLS} but then transform the data so that the new model has errors that are independent and have constant variance. Then least squares is applied and lo and behold we have parameter estimates, at least for linear models, that are unbiased and have optimum variances. Notice however that what has happened is that the statistician has in effect told us that our criteria should not have been C_{DLS}, which has been rejected, but the criteria of seeking unbiased and minimum variance estimators. The approach advocated here is that we reject this argument and say 'no' we really think C_{DLS} and C_{DLA} make sense for the purpose of getting the closest fitting lines in the fit–observation plot. We would rather have a good fit than good estimators.

The actual calculation for DLS and DLA is easily carried out using statistical or spreadsheet software.

Example

The methods of DLS and DLA have the great merit that it is almost a nonsense to use them without doing the associated plot of residuals and fit v observation. They are also of such a nature that alternative models can be explored by just changing the Q(p) column in a spreadsheet. As an example Table 9 shows the values of the optimum DLA criteria, the Sum of Absolute Deviations, SAD, from a variety of distributional models when applied to a set of soil data given in Karian and Dudewicz (p106). When these models were fitted it was noticed that the distributional residual for the largest observation was uniformly an ‘outlier’. Making use of the fact that the distribution of the largest observation of a sample of n has quantile function $Q(p^{1/n})$, it was evident that this observation was significantly small for a variety of different models. There are some good explanations for the phenomena of low extremes. The SAD summation therefore excluded this term, though the observation was still left in the data set and n was unaltered in the calculations. In comparing the models it is observed that the Wakeby gave the best fit. Figure 6 shows the Fit-observation plot for the fitted Wakeby distribution and Figure 7 the corresponding distributional residuals.

The Method of Probability Weighted Moments (PWM)

This method is analogous to the Method of Moments but uses special moments defined to avoid the high powers of x that were the drawback of that method. The simplest Population PWM are defined by

$$\omega_{r,s} = \int_0^1 Q(p)p^r(1-p)^s \, dp.$$

The sample equivalent will be

$$w_{r,s} = \sum_1^n x_{(i)} \, p^*{}_i{}^r(1-p^*{}_i)^s \, ,$$

where p^*_I =BETAINV(0.5,r,n+1-r) or some other suitable value.

Example

The PWM are linear functions and the Lambda distributions are also linear sums of distributions. Thus the forms of the PWM for the lambda family are of fairly straightforward form.

For example the FPLD has PWM of

$$\varpi_{00} = \lambda^* + (\eta/2)[(1-\delta)/\alpha(\alpha+1) - (1+\delta)/\beta(\beta+1)]$$

$$\varpi_{10} = (1/2)\lambda^* + (\eta/2)[(1-\delta)/\alpha(\alpha+2) - (1+\delta)/\beta(\beta+1)(\beta+2)]$$

$$\varpi_{20} = (1/3)\lambda^* + (\eta/2)[(1-\delta)/\alpha(\alpha+3) - 2(1+\delta)/\beta(\beta+1)(\beta+2)(\beta+3)]$$

$$\varpi_{30} = (1/4)\lambda^* + (\eta/2)[(1-\delta)/\alpha(\alpha+4) - 6(1+\delta)/\beta(\beta+1)(\beta+2)(\beta+3)(\beta+4)]$$

$$\varpi_{40} = (1/5)\lambda^* + (\eta/2)[(1-\delta)/\alpha(\alpha+5) - 24(1+\delta)/\beta(\beta+1)(\beta+2)(\beta+3)(\beta+4)(\beta+5)]$$

The symmetry of these expressions enables simple ratios of differences of the ϖ to be constructed that relate in direct fashions to the skewness and shape parameters. Thus there are fairly direct ways of solving for the five parameters. This method is used extensively for the extrme distributions and other positive models. However what little evidence there is suggests that, with the complex forms for the FPLD models, the estimation procedure leads to rather poorly performing estimates, e.g. Tarsitano (2005).

Maximum Likelihood

It is possible, but not natural, to use the method of maximum Likelihood for estimation with quantile functions. For each ordered observation $x_{(r)}$ we can use estimated parameters, θ^, to obtain $p_{(r)}$ from $x_{(r)} = Q(p_{(r)};\theta\hat{})$, solved by some iterative procedure. The log likelihood is then

$$L(\theta\hat{}) = -\sum \ln q(p_{(r)};\theta\hat{}).$$

The parameters can then be chosen to maximise the log likelihood. If a distribution, like the GLD, has no simple explicit pdf then this is the best way to carry out the ML optimisation, but if there is an explicit pdf then the usual ML method provides a more straightforward approach. For details and examples see Gilchrist (2000). Tarsitano (2000) looks at applying ML to the Wakeby distribution, which like the GLD only has explicit quantile form. A rather different approach, for the case of explicit pdf, is given by Heathcote, Brown and Mewhort (2002). In the method called quantile maximum likelihood they use quantiles to break the data into sections as in a frequency table. The frequency table enables the data to be modelled by a multinomial distribution and the parameters chosen using maximum likelihood for this distribution.

Fitting Distributions from Frequency Table Data

Data sometimes appears in the form of frequency tables. To fit these using DLS or DLA we note that if u_j is the upper boundary of an class containing n_j observations out of a total of N, with p_j being the total proportion up to u_j then the distributional least squares criterion, called Pseudo Least Squares in Tarsitano (2005), takes the form

$$\sum_{j=2}^{m} [u_j - Q(p_j;\theta))]^2 f_j, \text{ where } f_j = n_j/N.$$

Example

The paper by Tarstiano (2005) applies this approach to the GLD distribution. Giving as criterion:

$$\sum_{j=2}^{m} [u_j - \lambda - \eta[p_j^{\ \alpha} - (1 - p_j)^{\beta}]^2 f_j .$$

λ and η can be obtained explicitly by simple weighted least squares in terms of the pairs (u_j , y_j), where $y_j = [p_j^{\ \alpha} - (1 - p_j)^{\beta}]$. The minimum value for the criterion function depends on minus the correlation $-r(u_j , y_j)$, which is a function of α and β. We can therefore choose these parameters to maximise the correlation and thence obtain all four parameters. One small variation on this is to replace the f_j by above the form $f_j=(n_j+n_{(j+1)})/2(N-n_1/2_{-} - n_m/2)$. This weights according to the numbers both sides of the boundary.

7. GENERAL COMMENTS ON QUANTILE DISTRIBUTION MODELLING

In our discussions of distributions there are some general matters that warrant further comment:

(a) Several general forms have been illustrated by our examples. For example the five parameter form based on $Q(p) = \lambda + \theta S(p;\alpha) - \phi S(1 - p;\beta)$. In theory both tail terms should be increasing functions of p. A look at the many articles on the generalised lambda, with $\theta = \phi$ and right tail $S(p;\alpha) = p^{\alpha}$ or $1/(1-p)^{\alpha}$ depending on signs of the parameters, indicate that this is not always the case. Q(p) can still be increasing as long as one increasing term dominates. However it pays to look carefully at the various elements in such complex models, as with all models, since perfectly reasonable methods of estimation can give misleading estimates. It is always advisable to examine the fit-observation, Q-Q, plot, the distributional residuals and also look at the form of the fitted f(p).

(b) There have been a few studies of the properties of the various estimation methods for fitting quantile models and in particular the GLD. For example Tarsitano (2005) looks at fitting the GLD to some dozen data sets in the form of frequency tables and compares estimates. As suggested in the previous discussion the method of moments performs badly with distributions with 4 or more parameters. A careful look at some of the fitted models shows the type of problem referred to in (a) above. Though the method of probability weighted moments is better it is still not as good as the method of percentiles. The methods of Distributional Least Squares, called Pseudo Least Squares by Tarsitano, and Distributional Least Absolutes perform best.

(c) It is worth emphasising the feature used in the approach to transforming distributions above. Namely that when looking at the distributions discussed here in quantile form the many relationships between them is evident in their formulae, which is rarely true when expressed in classical CDF and pdf forms.

(d) The link between quantile functions and order statistics pays dividends in terms of the identification of outliers. Suppose that Q(p) is the fitted model for a set of n observations $x_{(1)}, \ldots, x_{(n)}$ then the distributions of the smallest and largest observations are $Q_{X(1)}(p) = Q(1 - (1 - p)^{1/n})$ and $Q_{X(n)}(p) = Q(p^{1/n})$. Thus $Q_{X(1)}(0.05)$ and $Q_{X(n)}(0.95)$ give limits to the whole data and observations outside these limits become suspect.

(e) It may also be observed that for many of the distributions that deal with highly skew or heavy tailed distributions the quantile form of the model is simpler in expression than the pdf form.

8. THE STANDARD QUANTILE REGRESSION APPROACH

One criticism that can be levelled at many old statistical textbooks is their failure to distinguish clearly between creating or identifying a model and the separate issue of how to estimate its parameters. This occurs particularly in the chapters, and indeed books, labelled 'Regression'. Often here the linear regression model and least squares estimation are presented as though they inherently went together. In one classic book of several hundred pages just two lines are devoted to saying that these models are fitted by Least Squares. Yet in practice there are many situations where the two issues of model choice and choice of method of estimation need to be clearly separated. In particular a criteria based on the squares of residuals is often inappropriate as a measure of the quality of fit, see for example the discussion in Gilchrist (1984). A recent occurrence of failure to separate the two issues is in the many articles that have recently appeared on the topic that is now called 'Quantile Regression'. This was originally developed by Koenker and Bassett (1978) and has had many useful and valid applications, see for example Koenker and Gelling (2001). The great contribution that this approach has made has been to draw attention away from simply looking at the mean regression curve and to emphasise the study of the whole shape of the distribution about such mean (or median) line. As with ordinary regression, quantile regression seems to have moved into an approach where the model and the estimation method are locked together. The purpose here is to seek to show that, as with ordinary regression, the consideration of the model and of the estimation of the parameters are distinct statistical operations and a choice of estimation method should be consciously made. To explore briefly the method of quantile regression as it is usually presented we need first to look at some basic properties of sample statistics. In particular we note that, given a set of data y_i, i= 1,..., n.:

The sample mean, a, is the value that minimises $\sum(y_i - a)^2$

The sample median, b, is the value that minimises $\sum|y_i - b|$

The sample s quantile, c, is the value that minimises $\sum\rho_s(y_i - c)$, where $\rho_s(z)$ is the piecewise linear 'check' function, as defined above in the discussion of the Asymmetric Laplace distribution. This is a loss-function with lines of slope s and 90^0+s meeting at $\rho_s(0) = 0$.

These three approaches correspond respectively to directly estimating the population mean, median and a specific s quantile by the methods of least squares, least absolutes and a linear programming specification respectively. In the first two the sample mean and median provide explicit solutions for the minima. In none of the approaches is there any recognition made of the distribution of the population from which the sample comes. They can be treated and generalised to be ways of estimating the population quantities corresponding the sample statistics, they, as they stand, are unrelated to issues of estimating <u>model</u> parameters. For estimating model parameters we obviously have to specify the distributional model. For Normally distributed errors the Method of Least Squares is equivalent to the Method of Maximum Likelihood, and hence the common assumption of Normality, or the use of transformations to achieve Normality. It should also be noted that the approach to estimating sample quantiles gives only one way of deriving a specific s Quantile.

Koenker and Bassett (1978) extend the above ideas to cover regression by replacing c by the conditional mean function x'β and minimising the criterion $\sum\rho_s(y_i - x'\beta)$, Thus though no distributional model is specified we are estimating the parameters of a deterministic regression lines. But the regression line is an assumed linear model for the s quantile, each quantile having a different line. This assumption is not realistic in many situations, so the model has been developed by allowing the specified quantile to be an factor in the parameters of the x'β model, giving x'β(s). This approach in general gives a model that is applied separately to each required quantile. In the paper by Koenker and Geling (2001) the models are extended to allow for not only the linear parameters to depend on s but also the variability. This is achieved by allowing the regressor to influence the error term thus the s quantile is modelled by x'β(s) + x'γS(s), where γ provides further parameters and S(p) is the form of the distributional quantile function. If, for a given quantile, this is represented as x'β*(s), then the regression parameters can all be estimated for each s by minimising $\sum\rho_s(y_i - x'\beta^*(s))$. The linearity assumption for the regression line is here, and in many other papers on quantile regression, relaxed by the use of smoothed lines, in particular by linear B-splines. In the paper a range of models are considered, such as Cox's proportional hazard model, the Bennet proportional odds model, and accelerated failure time models. Using the last of these in regression form the authors use a very large set of data to estimate five assumed parameters separately for each of 42 specified quantiles. Thus in effect 42 different semi-parametric models are used, one for each quantile. There are thus over 200 estimated parameters in the model. There are applications where one quantile is the central interest and others where looking at several may be a useful part of data analysis. However there are a number of concerns about this approach and a regret that the term Quantile Regression has been attached to it, as if that were all there is to using quantile methods in regression. There are a number of concerns that need expressing:

(a) One purely technical problem is that an approach to fitting separate models for each quantile is that without special modification there is no guarantee that the conditional quantile function, $Q(s|x)$, is actually a non decreasing function of s for all x. There are ways of addressing this problem in the literature, but it emphasises the piecemeal nature of theapproach.

(b) There is the concern at the shear numbers of parameters involved if many differing quantiles are studied. The fact that each will have some sampling distribution means that some of the large number will be misleadingly large or small.

(c) We have noted that the use of least squares is often favoured as the criteria in regression since is equivalent to applying maximum likelihood when the errors of the regression are normally distributed. This implies that, if this assumption is true, the estimates have many optimum properties. It has been pointed out that the use of the check function is equivalent to using maximum likelihood and assuming that the regression errors have asymmetric laplace distributions. But unfortunately this cannot truly be seen as an application of maximum likelihood as the assumed error distribution is different for each quantile,s.

(d) As with the discussion at the beginning of this section it should be noted that in many applications the form of $S(\theta)$ and its parameters are effectively ignored. The approach is essentially non-parametric. Yet in many regression problems understanding the nature of the distributions of the stochastic element is central.

To gain real understanding in modelling the models should seek to give pictures of all the underlying structure, stochastic as well as deterministic. The quantile regression approach does do this but only by separately looking at different quantiles. This is clearly an advance on just looking at the average behaviour as given by least squares. Koenker (2000) has drawn attention to the origin of these ideas going back to Edgeworth (1887) and the importance of the quantiles in providing vital information on the stochastic structure in regression. However to explore such structure fully a fully parametric approach is surely appropriate. Such an approach and the appropriate models are only hinted at or passed over without comment in most papers labelled 'quantile regression'. In the next section we explore what a fully parametric quantile approach to regression looks like.

9. REGRESSION MODELS AS CONDITIONAL QUANTILE FUNCTIONS.

In Gilchrist (2000) an approach is advocated that clearly separates the modelling and the estimation and advocates different approaches to estimation as being possible for models in quantile form. Unlike the above approach it takes the traditional regression modelling objective of developing a single overall parametric model and then estimating its parameters. In effect this means that S(p) is estimated as a function of p, rather than as a set of specific values. The model basis is similar to the classical linear regression model with the errors expressed in quantile function form. Thus the model specifies the deterministic element of the conditional quantile, $Q_y(p|x)$ of y, perhaps following Galton, as the median regression line $x'\beta$. We thus have as the simplest form the p conditional quantile

$$Q_y(p|x) = x'\beta + \eta S(p;\phi).\ 0 \le p \le 1.$$

This emphasises that S(.) is no longer a set of specific values, but as a quantile distribution is a function of p usually containing its own shape parameters, ϕ. The parameter η is a scale parameter. For median regression we set $S(1/2;\phi) = 0$.. As a further generalisation all parameters can be made functions of x and/or p, by the use of appropriate functional forms. Thus for example we might generalize to:

$$Q_y(p|x) = x'\beta(p) + \eta(x)S(p; \phi(x)).$$

We thus parallel Koenker and Bassett by expressing regression type models in terms of quantile functions. It might be noted that though this formulation of regression is often presented as relatively new, the formulation for median regression and Normal errors was first given by Sir Francis Galton over 100 years ago. Indeed Galton implicitly used the whole quantile regression formulation (Gilchrist (2005)).

We now turn to the question of how to estimate the parameters of regression quantile models. As with Koenker and Bassett's approach we generalise the distributional methods to cover regression and autoregression by replacing the expectations and medians of the distributions by the expressions needed to include the effects of the regressor variables. Thus for example in the method of distributional least absolutes in Section 6 we need to modify the criteria to reflect the presence of regressor variables and the need to use the residuals, not the data, as the basis for the ordering. Thus

$$C_{DLA} = \sum_i |(y_{i(r)} - M_{i(r)}|$$

Here $y_{i(r)}$ is the observation on the dependent variable for which the distributional residual, $e_{i(r)} = y_{i(r)} - M_{i(r)}$ is the rth in order of magnitude. The value of r can for example be given immediately as a function within Excel. We develop the median values to give for the regression quantile model

$$M_{(i)} = Q_y(p^*_{(r)} | x_{i(r)}).$$

Here $p^*_{(r)}$ = BETAINV(0.5, r, n+1-r). Table 10 gives a suitable layout for this form of calculation, following a spreadsheet format.

Example

An example of bivariate data given in D and K (p148) gives diameters, d, and heights, h, of trees. A distributional study of the data suggests at least one outlier. Thus in using the SAD criterion one observation was not included in the sum. Applying a simple regression with Normal errors

$$Q_h(p|x) = \lambda + \beta d + \eta N(p)$$

Fitting this with DLA gave SAD = 45.8. A look at the residuals suggests some skewness. Repeating the analysis with a Skew Lambda, $S(p;\delta,\alpha)$, error we have

$$Q_h(p|d) = \lambda + \beta d + \eta S(p;\delta,\alpha)$$

This gave SAD = 31.9.

Example

Figure 8 shows the weight gain, w, of four groups of animals receiving different doses, d, of a drug. The drug is seen to reduce weight gain. The linear form of model of the last example with either Normal or Logistic errors gives a good first fit (Using DLS gave the minimum sum of squared deviations, SSD, of 0.731 and 0.736 respectively). There is some evidence that the variance reduces as dose increases. So the scale term η was replaced by $\eta(\gamma/(\gamma+d))$, with anew parameter γ, giving a final model for the logistic

$$Q_w(p|d) = \lambda + \beta d + \eta(\gamma/(\gamma+d)) \ln\{p/(1-p)\}$$

This model gave the minimum SSD = 0.548. Figure 9 gives the distributional fit-observation diagram for this model.

Example

An example in Gilchrist (2000) relates the road distance between two points ,y, to the linear distance, x, by the model

$$Q_y(p|x) = x+[\phi x/\{\ln(2)\}^{\beta}][-\ln(1-p)]^{\beta}.$$

This corresponds to a Weibull Distribution whose median is a linear function of x going through the origin and with slope $1 + \phi$. The standard deviation is also a linear function of x. This not only gives a good fit to the data but can be argued for as a conceptually reasonable model for the situation described.

Example

The case study used by Koenker and Geling (2001) uses the log transformation of the data and the model:

$$Q_{\ln T}\,(p \mid x) = x'\beta + x'\eta S(p).$$

As a result of the methodology used in quantile regression there is no need to assign a form for S(p). As previously commented aA feature of quantile functions is that the quantile function and monotone transformations are invertible. Hence the model of the previous example can be expressed for the raw data as

$$Q_T\,(p \mid x) = \exp[x'\beta + x'\eta S(p)].$$

Issues of interpretation often support the use of models that are based on the raw data and avoid transformations. For example the transformation usually mixes the deterministic and stochastic elements and makes them difficult to separate again. The possible value of transformations lies at the initial estimation stage, for example the linear form of the log data enables good initial estimates to be found easily before a non-linear optimisation. The avoidance of transformations using quantile forms can give dividends without any change in the basic model as is illustrated in the following example.

Example

For one set of data a log transformation was originally used and a quadratic in x fitted. The classical regression approach applies least squares to the transformed data and then estimates the spread using the standard deviation of the residuals in this model. If one now transforms back to the original data and calculates the sum of squares of errors for the fitted curve one obtains Residual sum of squares of 2.594 and for the transformed distributional residuals allowing for the fitted Normal one obtained SSD = 0.4526. However we may keep to the raw data and use the model

$$y_p = \alpha\exp[\beta x + \gamma x^2 + \eta N(p)].$$

If we follow the classical approach using two stages so apply LS to the deterministic element and then fit the stochastic element we get RSS = 0.268 and SSD = 0.034, clearly a significant improvement. However a major feature of this approach is that we can fit both deterministic and stochastic terms in one operation. Applying least squares in one operation leads to a SSD of 0.0188, a further significant improvement. As a further development the logistic, with its longer tails, was used in place of the normal and gave a SSD of 0.0160.

Before leaving regression it is worth noting very briefly that time series provides a specific area of application. Classical time series and forecasting have put great emphasis on the use of Normally distributed innovations in modelling and on the expectation of future values as the target of forecasting. It is also clear however from the literature that Normality is not universal, with many situations giving rise to long tailed distributions. Further there are many applications where the whole distribution of the future values is of prime interest, for example within studies of Value-at-risk.

As an example for a simple time series an autoregressive model is defined by

$$Q_{x(t)|\mathbf{x}(t-1)}(p|\mathbf{x}_{t-1}) = \beta_0 + \beta_1 x_{t-1} + \ldots + \beta_i x_{t-i} + \eta S(p:\phi).$$

Experience with such models suggests that where S(p) is a non-symmetric distribution, with a long tail controlled by the shape parameter ϕ, the series generated looks very similar to those where non-linear models are often used to represent jumps in the mean.

11 CONCLUSIONS

An inherent assumption of most statistical literature is that statistics is based on the classical definitions of the pdf, f(x), and cdf, F(x). The inverse of the cdf, $Q(p)=F^{-1}(x)$, happens to occur in some calculations but is otherwise of no great interest. A basic tenet of this paper is that there is in a sense an alternative statistical universe in which the quantile function is the foundation definition, which happens to have an inverse called the cdf. These two universes turn out to have different properties. What one does elegantly, the other struggles with and vice versa. Where one has a simple explicit expression and straightforward calculation the other gets lost in approximations and tables. Each suggests its own appropriate methodology. Table 11 seeks to set out the comparison of results and methods, showing some of the strengths and weaknesses of the two approaches. This paper, one hopes, has provided illustration for the strengths and some details of the quantile based approach.

For distributional modelling the illustrations in this paper show the flexibility of quantile functions as the defining form of distributions. They also underline the concept that we need to consider not just the selection of pdf and cdf from the library of distributions, but also consider using quantile approaches to construct distributions to have the features that are important in the application and the data.. We have seen that the basic forms of the quantile functions are often simple in form and show clearly the relationships between distributions. We have also seen that with careful construction the parameters can be made to relate to specific shape properties of distributions. The fact that some distributions have five such distinct features led to five parameter models. However in terms of building, choosing and fitting distributions one should not forget the basic principle that simplicity is preferable to complexity.

For regression situations the standard approaches of quantile regression may be criticised on the failure to separate the definition of the model and the method of estimation chosen, no option being given, and also on the confusion caused by commandeering the general term quantile regression to describe one very limited aspect of the use of regression involving quantile functions, namely non-parametric fitting. We have emphasised that we start with a regression model with the stochastic element described by the quantile function, so both deterministic and stochastic element are explicit in the model. The vast emphasis on least squares regression with implicit normality has led to neglect of the many situations where non-normality is an important feature. We then have to choose a method of estimation to fit the parameters of both deterministic and stochastic aspects of the model, which may or may not be clearly distinct. Of the range available we have suggested distributional least squares or least absolutes as being particularly appropriate.

13 REFERENCES

David, H. A. and Nagaraja, H. N. (2003) Order Statistics (3rd Edition).. Hoboken. New Jersey. John Wiley and Sons Inc.

Edgeworth, F. Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine* **16**, 360-375.

Friemer, M., Mudholker, G. S., Kollia, G. and Lin, C. T. (1988). A study of the generalized Tukey lambda distribution. *Comm. Statist. Theory Meth.* **17,** (10) 3547-3567.

Galton, F. (1883). *Enquiries into Human Faculty and Its Development.* London. Macmillan,

Gilchrist, W. G. (1979) *Statistical Modelling*. John Wiley and Sons. Chichester.

Gilchrist, W. G. (1997) Modelling with quantile distribution functions. *Journal of Applied Statistics.* **24** (1) 113-122.

Gilchrist, W. G. (2000) *Statistical Modelling with Quantile Functions*. Bocca Raton. Fl. Chapman and Hall / CRC Press.

Gilchrist, W.G. (2005) Galton misrepresented. Significance. **2** (3).

Harter H. L. and Balakrishnan, N. (1996). *The CRC Handbook of Tables for the use of Order Statistics in Estimation.* Boca Raton. Fl. CRC Press.

Hastings, C., Mosteller, F., Tukey, J. W. and Winsor, C. P. (1947) Low moments for small samples: a comparative study of statistics. *Ann. Math. Statist.* **18**, 413-426.

Heathcote, A. Brown, S. and Mewhort, D. J. K. (2002) Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review,* **9**, 394-401.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994 and 1995). *Continuous Univariate Distributions, 2nd Edition, Vols 1 and 2.* New York. John Wiley and Sons.

Karian, Z.A. and Dudewicz, E.J. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods..* Boca Raton. Fl. CRC Press.

Koenker, R.W. (2000). Galton, Edgeworth, Frisch, and prospects for quantile regression in economics. *Jnl of Econometrics*, **95,** 347-374.

Koenker, R.W. and Bassett, G (1978) Regression Quantiles. *Econometrica,* **46**, (1) 33-50.

Lloyd, E. H. (1952). Least Squares estimation of location and scale parameters using order statistics. *Biometrica.* **38**, 88-95.

Parzen, E. (1979) Nonparametric statistical data modelling. *J. Amer. Statist. Assoc.* **74**,105-121.

Ramberg, J. S. and Schmeiser, B. W. (1972). An approximate method for generating symmetric random variables. *Comm. ACM.* **15,** *987-990.*

Ramberg, J. S. and Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Comm. ACM.* **17,** 78-82.

Ramberg, J. S., Tadikamala, P. R., Dudewicz, E. J., and Mytytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics,* **21,** 201-214.

Tarsitano, A. (2000). Fitting the Wakeby Model using maximum likelihood. (Paper, and others, available at www.ecostat.unical.it/Tarsitano/atarsita.htm)

Tarsitano, A. (2004). Fitting the Generalized lambda Distribution to Income Data. C*OMPSTAT' 2004 1862-1861.*

Tarsitano, A. (2005). Estimation of GLD parameters for grouped data. *Communications in Stat. – Theory and Method.* **34,** (8) 1689-1710.

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33,** 1-67.

Table 1. Measures of Shape.

Measure of	WGG	K & D
Position	Median	Median
Spread	IQR	IDR
Skewness	GQ = [UQ+LQ-2M]/IQR GD= [UD+LD-2M]/IDR	TR=[M-LD]/[UD-M]
Tail Shape - Symmetrical	T =IDR/IQR	T*=IQR/IDR
Tail Shape - General	UT=[UD-M}/[UQ-M], LT=[M-LD]/[M-LQ]	
Moor's Kurtosis	[(Q(7/8)-Q(5/8))+(Q(3/8)-Q(1/8))]/IQR	

Table 2. Shape measures for the Symmetric and Skew Lambda distributions.

Statistic	Tukey-Lambda	Skew Lambda
Q(p)	$\lambda + (\eta/\alpha)[p^{\alpha} -(1 - p)^{\alpha}]$.	$\lambda+(\eta/2\alpha)[(1-\delta)p^{\alpha} -(1+\delta) (1-p)^{\alpha}]$.
q(p)	$\eta[p^{\alpha-1} +(1 - p)^{\alpha-1}]$.	$(\eta/2)[(1-\delta)p^{\alpha-1} +(1+\delta) (1-p)^{\alpha-1}]$.
M	λ	$\lambda+\eta\delta/2^{\alpha}\alpha$
IQR	$2\eta(3^{\alpha} -1)/\alpha 4^{\alpha}$	$\eta(3^{\alpha} - 1)/4^{\alpha}\alpha$
IDR	$2\eta(9^{\alpha} -1)/\alpha 10^{\alpha}$	$\eta(9^{\alpha} - 1)/10^{\alpha}\alpha$
GQ	0	$\delta (2.2^{\alpha} -3^{\alpha} -1)/(3^{\alpha}-1)$
GD	0	$\delta (2.5^{\alpha} -9^{\alpha} -1)/(9^{\alpha}-1)$
TR	1	$(1+\delta c)/(1-\delta c)$ $c=(9^{\alpha}+1-2.5^{\alpha})/(9^{\alpha}-1)$
T*	$\frac{(3^{\alpha} -1)\, 10^{\alpha}}{(9^{\alpha} -1)\, 4^{\alpha}}$	$\frac{(3^{\alpha} -1)\, 10^{\alpha}}{(9^{\alpha} -1)\, 4^{\alpha}}$
K	$(7^{\alpha}+3^{\alpha}-5^{\alpha}-1)/[2^{\alpha}(3^{\alpha}-1)]$	$(7^{\alpha}+3^{\alpha}-5^{\alpha}-1)/[2^{\alpha}(3^{\alpha}-1)]$

Table 3 The main properties of the Generalized Lambda Distribution

Standard form	$p^{\alpha}-(1-p)^{\beta}$
Support	Depends on α,β see K and D
Median	$1/2^{\alpha}-1/2^{\beta}$
Quartiles	$LQ = (1/4)^{\alpha}-(3/4)^{\beta}$. $UQ =(3/4)^{\alpha}-(1/4)^{\beta}$.
IQR	$(1/4^{\alpha})(3^{\alpha}-1)+(1/4^{\beta})(3^{\beta}-1)$
IDR	$(1/10^{\alpha})(9^{\alpha}-1)+(1/10^{\beta})(9^{\beta}-1)$
Skewness GQ	$\frac{[(1/4^{\alpha})(3^{\alpha}+1)-(1/4^{\beta})(3^{\beta}+1)-1/2^{\alpha}+1/2^{\beta}]}{(1/4^{\alpha})(3^{\alpha}-1)+(1/4^{\beta})(3^{\beta}-1)}$
T*	$\frac{(1/4^{\alpha})(3^{\alpha}-1)+(1/4^{\beta})(3^{\beta}-1)}{(1/10^{\alpha})(9^{\alpha}-1)+(1/10^{\beta})(9^{\beta}-1)}$

Table 4 The tail shapes of the Lambda Distributions

Parameter values	$(p^{\alpha}-1)/\alpha$	$-[(1-p)^{\beta}-1]/\beta$	Individual support
<0	Reflected Centred Pareto	Centred Pareto	$(-\infty, 0), (0, \infty)$
0	Reflected Exponential	Exponential	$(-\infty, 0), (0, \infty)$
(0 to 1)	Shifted Power Decreasing to left	Shifted Power Decreasing to right	$(-1/\alpha, 0), (o, 1/\beta)$
1	Uniform	Uniform	(-1, 0), (0, 1)
>1	Shifted Power Increasing to left	Shifted Power Increasing to right	$(-1/\alpha, 0), (o, 1/\beta)$

Table 5 Properties of the Power*Pareto Distribution

Standard form	$p^{\alpha}/(1-p)^{\beta}$, $\alpha,\beta > 0$
Support	$(0, \infty)$
Median	$2^{\beta-\alpha}$
Quartiles	$LQ = 4^{\beta}/(4^{\alpha}3^{\beta})$. $UQ = (3^{\alpha}4^{\beta})/4^{\alpha}$
IQR	$4^{\beta-\alpha}/[3^{\alpha}-(1/3)^{\beta}]$.
Skewness G	$[[3^{\alpha} + (1/3)^{\beta} - 2^{\alpha-\beta}.]/ [3^{\alpha}-(1/3)^{\beta}]$.

Table 6 The main properties of the Five Parameter Lambda Distribution

Standard form	$\phi p^{\alpha}-\theta(1-p)^{\beta}$
Support	Depends on α,β
Median	$\phi/2^{\alpha} - \theta/2^{\beta}$
Quartiles	$LQ = \phi(1/4)^{\alpha} - \theta(3/4)^{\beta}$. $UQ = \phi(3/4)^{\alpha} - \theta(1/4)^{\beta}$.
IQR	$(\phi/4^{\alpha})(3^{\alpha}-1)+(\theta/4^{\beta})(3^{\beta}-1)$
IDR	$(\phi/10^{\alpha})(9^{\alpha}-1)+(\theta/10^{\beta})(9^{\beta}-1)$
Skewness GQ	$\dfrac{[(\phi/4^{\alpha})(3^{\alpha}+1)-(\theta/4^{\beta})(3^{\beta}+1)-\phi/2^{\alpha}+\theta/2^{\beta}]}{(\phi/4^{\alpha})(3^{\alpha}-1)+(\theta/4^{\beta})(3^{\beta}-1)}$
T*	$\dfrac{(\phi/4^{\alpha})(3^{\alpha}-1)+(\theta/4^{\beta})(3^{\beta}-1)}{(\phi/10^{\alpha})(9^{\alpha}-1)+(\theta/10^{\beta})(9^{\beta}-1)}$

Table 7 Simulation estimates for the Logistic Distribution:
$Q(p) = \lambda + \eta\ln(p/(1-p))$. $\lambda=0$, $\eta=1$

Parameter	Method	Estimate	Standard Error
λ	MoM	x	0.4055
λ	MoP	m	0.4298
λ	Dist L S	intercept	0.4137
η	MoM	$\sqrt{3}s/\pi$	0.1921
η	MoP –IQR	IQR/2ln3	0.2592
η	MoP -IDR	IDR/4ln3	0.2121
η	Dist L S	slope	0.2213

Table 8 Example of tabular layout for fitting quantile distributions by DLS or DLA
Notes: Rows and columns interchanged for ease of presentation. Table needs fully setting up, with 'sensible' estimates, before formal estimation carried out.

Initial and current estimates of parameters	λ, η,α, β, (A)				Final Column
Original data z_s		z_1	......	z_n	
Ordered data x_r	1, 2 for smallest.	x_1	......	x_n	
r		1	2,3,...	n	
p^*_r	BETAINV(0.5,r,n+1-r)	p^*_1	......	p^*_n	
Linear parameter fit To put in (A)	$Q(p^*) = \lambda + \eta S(p^*;\ \alpha, \beta)$ λ = Intercept(x against S(p*)). η = Slope(x against S(p*))	$S(p^*_1)$	......	$S(p^*_n)$	
Whole model fit $Q(p^*;\ \alpha, \beta, \gamma, \eta)$	x_r^ = $Q(p^*_r\ ; \lambda, \eta, \alpha, \beta)$ Estimates for (α,β) by optimisation to minimise B below	x_1^	......	x_n^	
Distributional Residuals	$f_r = x - x$^	f_1	......	f_n	
Squares of distributional residuals (or Abs)	f_r^2 ($\lvert f \rvert$)	f_1^2	......	f_n^2	Sum =SDD (or SAD) (B)

Now plot f_r against x_r^ for a distributional residual plot and x_r against x_r^ for a fit-observation diagram (Q-Q plot).

Table 9 Values of the minimum DLA criterion for the soil data

Distributions (k Parameters)	SAD = $\sum \lvert (x_{(r)} - M_{(r)} \rvert$
(Generalized Pareto)^power (3)	5318
Power*Pareto (3)	5478
Skew Lambda (4)	5877
GLD (4) using MoP estimates	5802
GLD (4)	4947
Wakeby (4)	3192
FPLD (5)	3758

Table 10. Illustrative tabular layout for fitting regression quantile models of the form $y = Q(p; x, \alpha, \beta, \gamma, \eta)$

Notes: Rows and columns interchanged for ease of presentation. Table needs fully setting up, with 'sensible' estimates, before formal estimation carried out.

Current estimates of	$\alpha, \beta, \gamma, \eta$				Final Column
s raw data order		1	2.. n-1	n	
x_s		x_1	......	x_n	
y_s		y_1	.....	y_n	
Regression Fit (deterministic)	$y\wedge = E[Q(p; x, \alpha, \beta, \gamma, \eta)]$ (or $Q(1/2; x, \alpha, \beta, \gamma, \eta)$) Initial estimates using LS or optimisation of A below	$y_1\wedge$	......	$y_n\wedge$	
Regression residual	$e = y - y\wedge$	e_1	......	e_n	
Squares residuals (or Abs residuals)	e^2 ($\lvert e \rvert$)	e_1^2	......	e_n^2	Sum = RSS (or RSA) (A)
r = rank order of regression residual.	1, 2 for smallest.	r_1	......	r_n	
$p^*_{r,s}$	BETAINV(0.5,r,n+1-r)	$p^*_{r1,1}$	......	$p^*_{rn,n}$	
Whole model fit $Q(p^*; x_s, \alpha, \beta, \gamma, \eta)$	$y_s\wedge\wedge = Q(p^*_{r,s}; x_s, \alpha, \beta, \gamma, \eta)$ Estimates using optimisation of B below	$y_1\wedge\wedge$	......	$y_n\wedge\wedge$	
Distributional Residuals	$f_s = y - y\wedge\wedge$	f_1	......	f_n	
Squares of distributional residuals, deviations (or Abs)	f_s^2 ($\lvert f \rvert$)	f_1^2	......	f_n^2	Sum =SSD (or SAD) (B)

Table 11. Comparison of the parallel universes of classical statistics and quantile modelling.

Aspect	Classical Statistics	Quantile Modelling
Defining functions	$f(x)$, $F(x)$	$Q(p)$
The Expectation, μ'_k	$E(x^k) = \int_{support} x^k f(x)dx$	$= \int_0^1 [Q(p)]^k dp$
The Median	M st $\int^M f(x)dx = 1/2$	$Q(0.5)$
Linear functions	$E(aX + bY)= aE(X)+bE(Y)$	$M(aX + bY)$ no general result.
T(.) non decreasing transform.	No general result for distribution. Many approximations.	$Q_T(p) = T(Q(p))$ $M(T(X)) = T(M(X))$
Constructing distributions	Not possible in general	Can +, x (for +ve dists) and T(Q(p)) to create new distributions
Order statistics - distributions	Form of $f(X_{(r)})$. No explicit general result. Approximations.	Q[BETAINV(p, r, n-r+1)]
Order statistics -Rankits	$E(X_{(r)})$. No explicit general result. Approximations.	$M(X_{(r)}) =$ Q[BETAINV(0.5, r, n-r+1)]
Sampling distributions $t(x_1, x_2, .. , x_n)$	f(t) sometimes explicitly obtainable.	$t(Q(u_1), Q(u_2), .. , Q(u_n))$ may simulate from uniform random numbers.
Estimation	Moments, ML,	+ Percentiles Distributional LS and LA
Regression - form	Conditional means	Conditional medians
Regression - handling	Transform to Normality	Use in natural form
Regression - fitting	LS looks only at deterministic element	DLS and DLA estimate both deterministic and stochastic elements
Quantile Regression	Is a non-parametric method of estimation	Uses full parametric models

Zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz

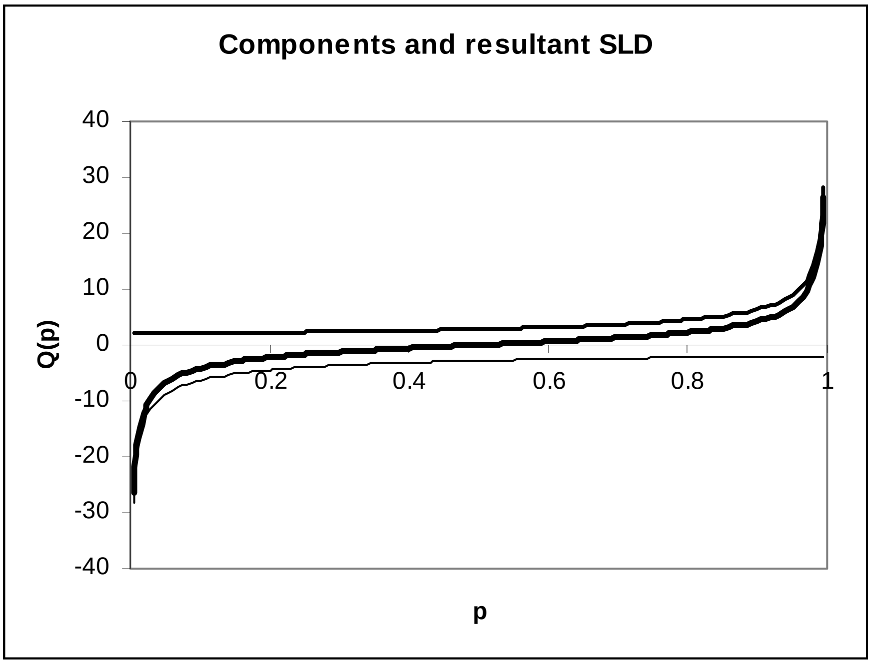
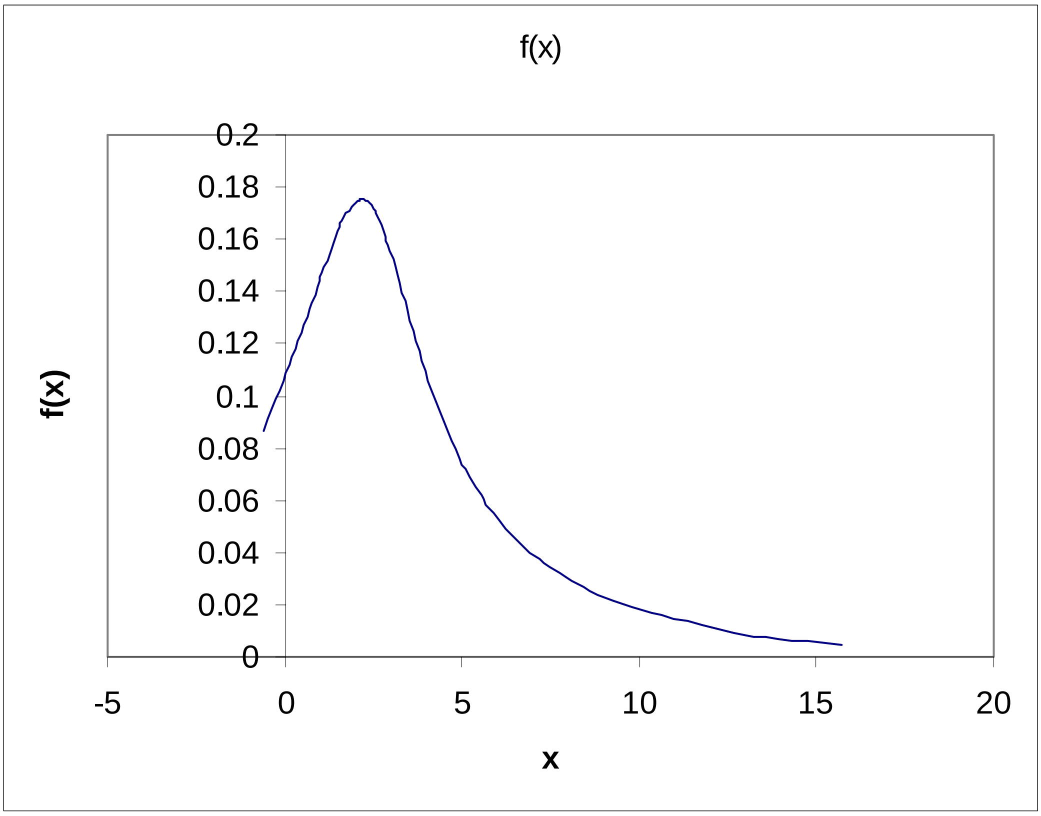


Figure 1 The quantile functions of a Pareto, its reflection and sum (SLD), α = -0.2.

Zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz



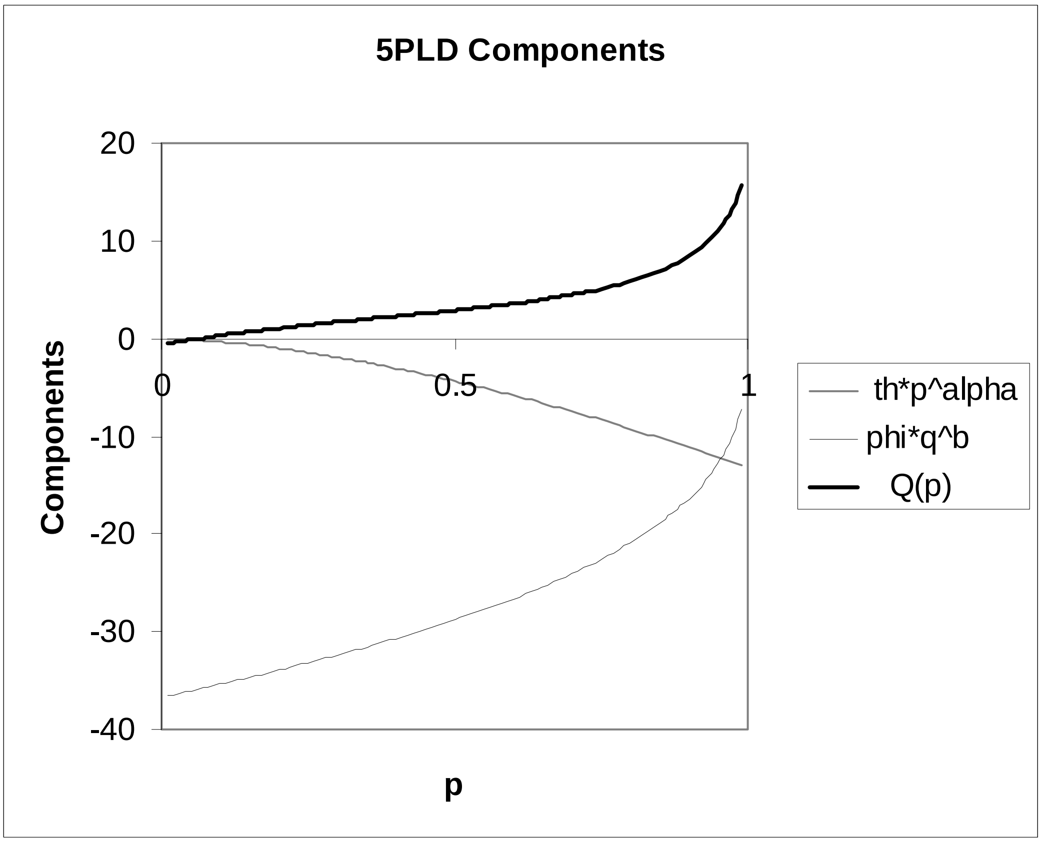


Figure 2. FPLD involving one invalid component.

ZZZZZZZZZZZZZZZzzzzzzzzzzzz

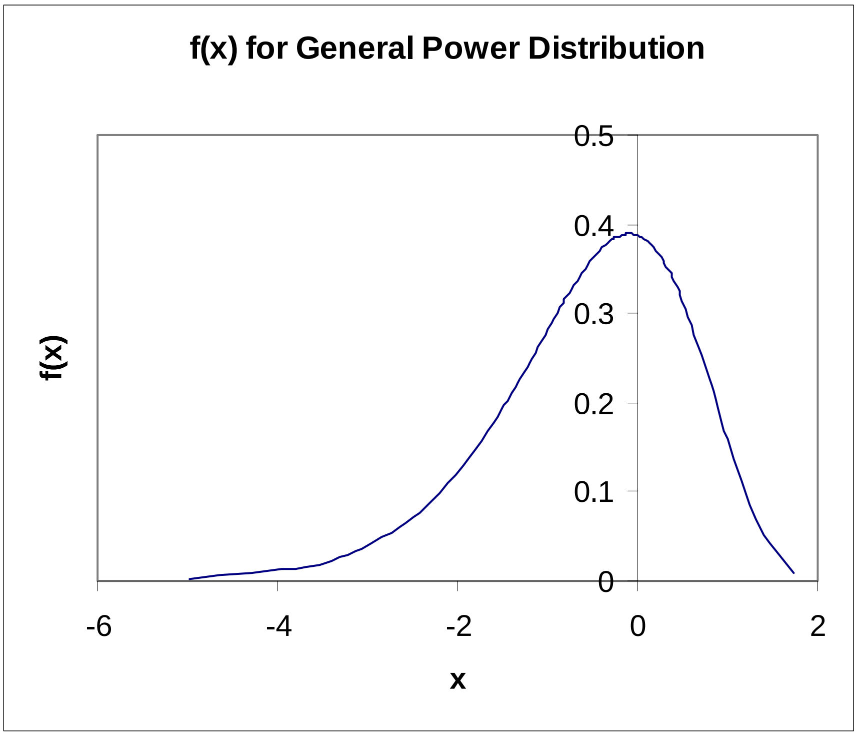


Figure 3. The General Power Distribution, α = 0.1, β= 0.1.

Zzzzzzzzzzzzzzzzzzzzzzzzzzzz

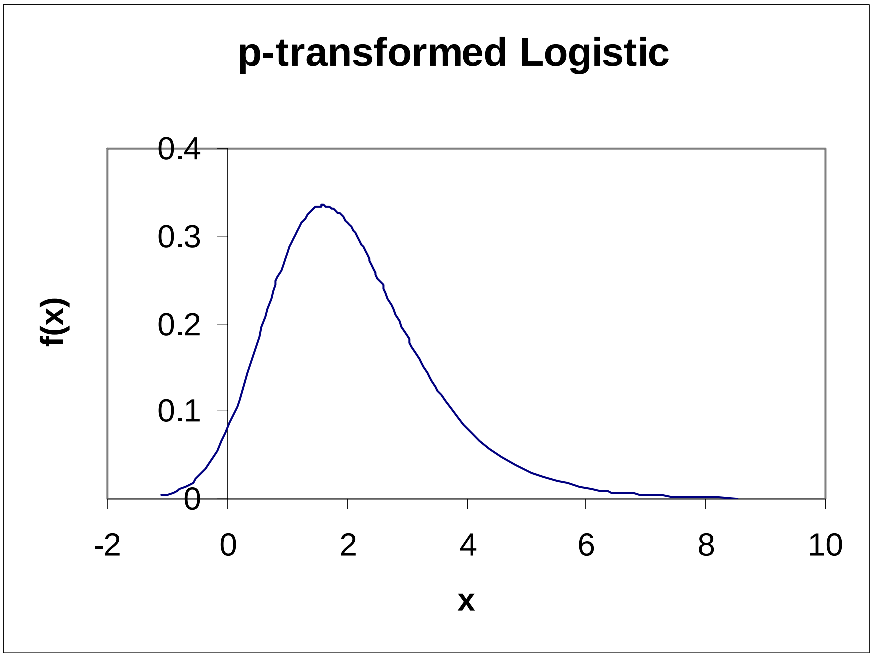


Figure 4. Example of the p-transformed Logistic. α = 0.2

Zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz

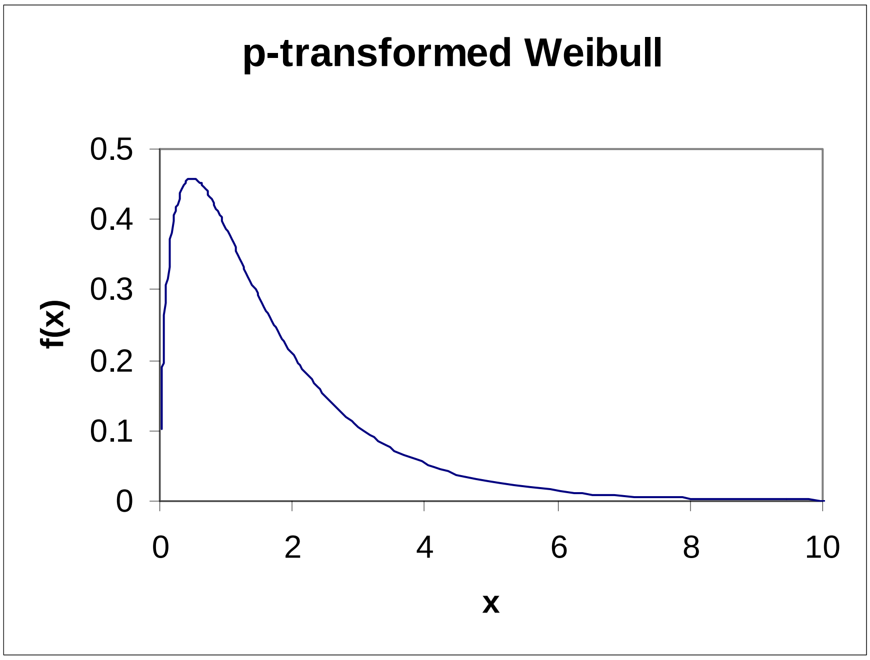


Figure 5. Example of the p-transformed Weibull. $\alpha = 0.5$, $\beta = 1.2$.

Zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz

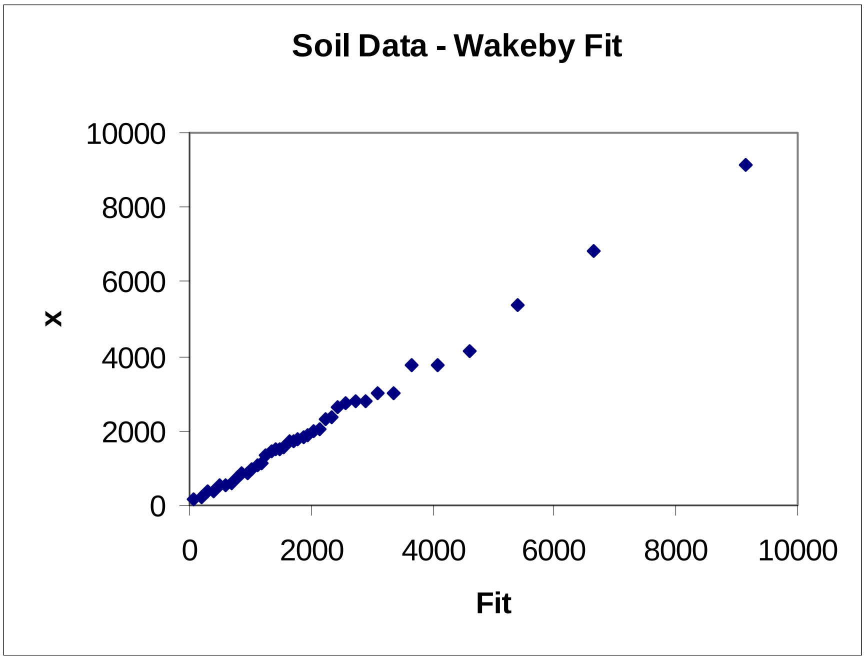


Figure 6. Fit-observation plot for the soil data.

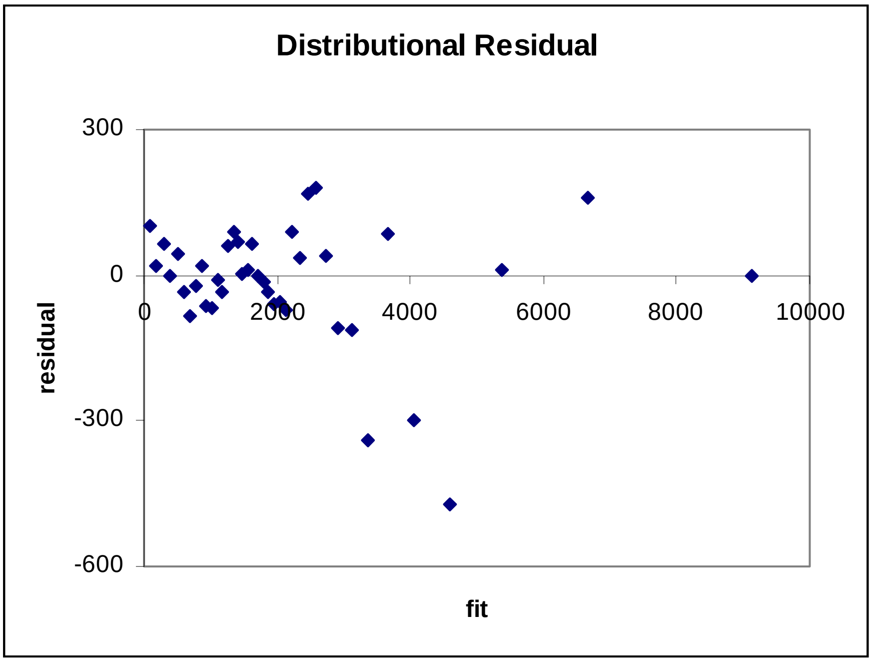


Figure 7. Distributional residuals v fit for soil Wakeby fit.

Zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz

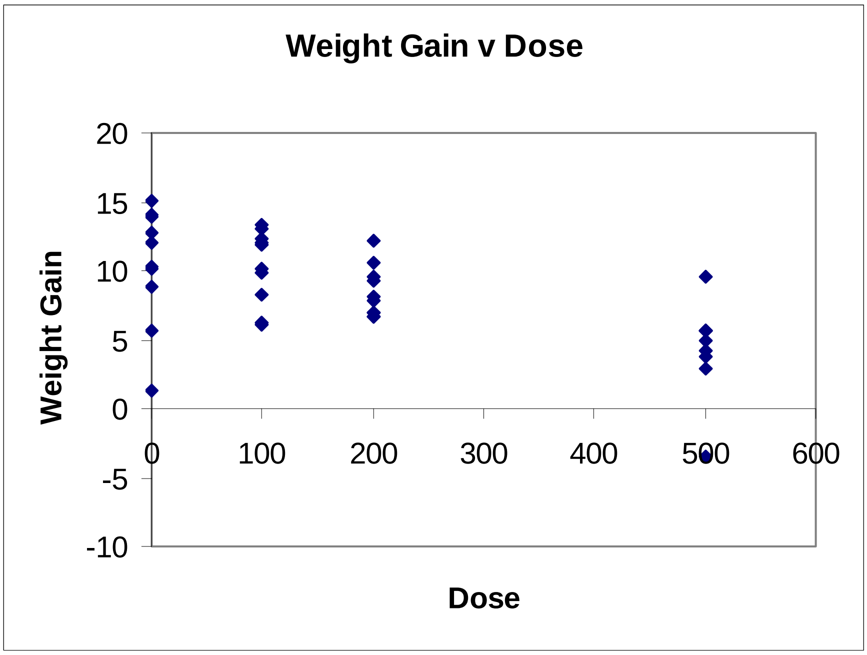


Figure 8. Data for weight gain v dose example.

zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz

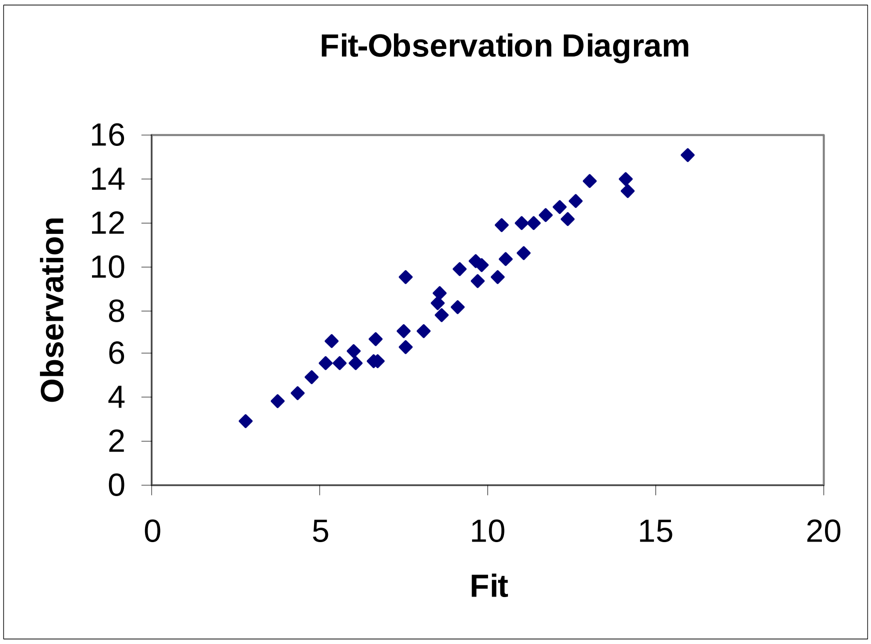


Figure 9. Distributional fit-observation for weight gain regression.