4.10 Quantile Regression

4.10.1 Background and Motivation

Standard regression approaches effectively model the (conditional) mean of the dependent variable – that is, they capture the average value of *y* given the average values of all of the explanatory variables. We could of course calculate from the fitted regression line the value that *y* would take for any values of the explanatory variables, but this would essentially be an extrapolation of the behaviour of the relationship between *y* and *x* at the mean to the remainder of the data.

As a motivational example of why this approach will often be sub-optimal, suppose that it is of interest to capture the cross-sectional relationship across countries between the degree of regulation of banks and gross domestic product (GDP). Starting from a very low level of regulation (or no regulation), an increase in regulation is likely to encourage a rise in economic activity as the banking system functions better as a result of more trust and stability in the financial environment. However, there is likely to come a point where further increasing the amount of regulation may impede economic growth by stifling innovation and the responsiveness of the banking sector to the needs of the industries it serves. Thus we may think of there being a non-linear (∩-shaped) relationship between regulation and GDP growth, and estimating a standard linear regression model may lead to seriously misleading estimates of this relationship as it will 'average' the positive and negative effects from very low and very high regulation.

Of course, in this situation it would be possible to include non-linear (i.e., polynomial) terms in the regression model (for example, squared, cubic, … terms of regulation in the equation). But *quantile regressions*, developed by Koenker and Bassett (1978), represent a more natural and flexible way to capture the complexities inherent in the relationship by estimating models for the conditional quantile functions. Quantile regressions can be conducted in both time-series and cross-sectional contexts, although the latter are more common. It is usually assumed that the dependent variable, often called the *response variable* in the literature on quantile regressions, is independently distributed and homoscedastic; these assumptions can of course be relaxed but at the cost of additional complexity. Quantile regressions represent a comprehensive way to analyse the relationships between a set of variables, and are far more robust to outliers and non-normality than OLS regressions, in the same fashion that the median is often a better measure of average or 'typical' behaviour than the mean when the distribution is considerably skewed by a few large outliers. Quantile regression is a non-parametric technique since no distributional assumptions are required to optimally estimate the parameters.

The notation and approaches commonly used in quantile regression modelling are different to those that we are familiar with in financial econometrics, and this probably limited the early take up of the technique, which was historically more widely used in other disciplines. Numerous applications in labour economics were developed, for example. However, the more recent availability of the techniques in econometric software packages and increased interest in modelling the 'tail behaviour' of series have spurred applications of quantile regression in finance. The most common use of the technique here is to value at risk modelling. This seems natural given that the models are based on estimating the quantile of a distribution of possible losses – see, for example, the study by Chernozhukov and Umantsev (2001) and the development of the CaViaR model by Engle and Manganelli (2004).[1]

Quantiles, denoted τ, refer to the position where an observation falls within an ordered series for y – for example, the median is the observation in the very middle; the (lower) tenth percentile is the value that places 10% of observations below it (and therefore 90% of observations above), and so on. More precisely, we can define the τ-th quantile, $Q(\tau)$, of a random variable y having cumulative distribution $F(y)$ as

$$Q(\tau) = \inf y : F(y) \geq \tau \qquad \textbf{(4.52)}$$

where inf refers to the infimum, or the 'greatest lower bound' which is the smallest value of y satisfying the inequality. By definition, quantiles must lie between zero and one.

Quantile regressions take the concept of quantiles a stage further and effectively model the entire conditional distribution of y given the explanatory variables (rather than only the mean as is the case for OLS) – thus they examine their impact on not only the location and scale of the distribution of y, but also on the shape of the distribution as well. So we can determine how the explanatory variables affect the fifth or ninetieth percentiles of the distribution of y or its median and so on.

4.10.2 Estimation of Quantile Functions

In the same fashion as the OLS estimator finds the mean value that minimises the sum of the squared residuals, minimising the sum of the absolute values of the residuals will yield the median value. By definition, the absolute value function is symmetrical so that the median always has the same number of data points above it as below it. But if instead the absolute residuals are weighted differently depending on whether they are positive or negative, we can calculate the quantiles of the distribution. To estimate the τ-th quantile, we would set the weight on positive observations to τ, which is the quantile of interest, and that on negative observations to $1 - \tau$. We can select the quantiles of interest (or the software might do this for us), but common choices would be 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95. The fit is not always good for values of τ too close to its limits of 0 and 1, so it is advisable to avoid such values.

We could write the minimisation problem for a set of quantile regression parameters $\hat{\beta}_\tau$, each element of which is a $k \times 1$ vector, as

$$\hat{\beta}_\tau = \text{argmin}_\beta \left(\sum_{i:y_i > \beta x_i} \tau |y_i - \beta x_i| + \sum_{i:y_i < \beta x_i} (1 - \tau)|y_i - \beta x_i| \right) \quad \textbf{(4.53)}$$

This equation makes it clear where the weighting enters into the optimisation. As above, for the median, $\tau = 0.5$ and the weights are symmetric, but for all other quantiles they will be asymmetric. This optimisation problem can be solved using a linear programming representation via the simplex algorithm or it can be cast within the generalised method of moments (GMM) framework.

As an alternative to quantile regression, it would be tempting to think of partitioning the data and running separate regressions on each of them – for example, dropping the top 90% of the observations on y and the corresponding data points for the xs, and running a regression on the remainder. However, this process, tantamount to truncating the dependent variable, would be wholly inappropriate and could lead to potentially severe sample selection biases of the sort discussed in Chapter 12 and highlighted by Heckman (1976). In fact, quantile regression does not partition the data – all observations are used in the estimation of the parameters for every quantile.

It is quite useful to plot each of the estimated parameters, $\hat{\beta}_{i,\tau}$ (for $i = 1, \ldots, k$), against the quantile, τ (from 0 to 1) so that we can see whether the estimates vary across the quantiles or are roughly constant. Sometimes ±2 standard error bars are also included on the plot, and these tend to widen as the limits of τ are approached. Producing these standard errors for the quantile regression parameters is unfortunately more complex conceptually than estimating the parameters themselves and thus a discussion of these is beyond the scope of this book. Under some assumptions, Koenker (2005) demonstrates that the quantile regression parameters are asymptotically normally distributed. A number of approaches have been proposed for estimating the variance-covariance matrix of the parameters, including one based on a bootstrap – see Chapter 13 for a discussion of this.

4.10.3 An Application of Quantile Regression: Evaluating Fund Performance

A study by Bassett and Chen (2001) performs a style attribution analysis for a mutual fund and, for comparison, the S&P500 index. In order to examine how a portfolio's exposure to various styles varies with performance, they use a quantile regression approach.

Effectively evaluating the performance of mutual fund managers is made difficult by the observation that certain investment styles – notably, value and small cap – yield higher returns on average than the equity market as a whole. In response to this, factor models such as those of Fama and French (1993) have been employed to remove the impact of these characteristics – see Chapter 14 for a detailed presentation of these models. The use of such models also ensures that fund manager skill in picking highly performing stocks is not confused with randomly investing within value and small cap styles that will outperform the market in the long run. For example, if a manager invests a relatively high proportion of his portfolio in small firms, we would expect to observe higher returns than average from this manager because of the firm size effect alone.

Bassett and Chen (2001) conduct a style analysis in this spirit by regressing the returns of a fund on the returns of a large growth portfolio, the returns of a large value portfolio, the returns of a small growth portfolio, and the returns of a small value portfolio. These style portfolio returns are based on the Russell style indices. In this way, the parameter estimates on each of these style-mimicking portfolio returns will measure the extent to which the fund is exposed to that style. Thus we can determine the actual investment style of a fund without knowing anything about its holdings purely based on an analysis of its returns *ex post* and their relationships with the returns of style indices. Table 4.2 presents the results from a standard OLS regression and quintile regressions for $\tau = 0.1$, 0.3, 0.5 (i.e., the median), 0.7 and 0.9. The data are observed over the five years to December 1997 and the standard errors are based on a bootstrapping procedure.

Table 4.2 OLS and quantile regression results for the Magellan fund

	OLS	***Q*(0.1)**	***Q*(0.3)**	***Q*(0.5)**	***Q*(0.7)**	***Q*(0.9)**
Large growth	0.14	0.35	0.19	0.01	0.12	0.01
	(0.15)	(0.31)	(0.22)	(0.16)	(0.20)	(0.22)
Large value	0.69	0.31	0.75	0.83	0.85	0.82
	(0.20)	(0.38)	(0.30)	(0.25)	(0.30)	(0.36)
Small growth	0.21	–0.01	0.10	0.14	0.27	0.53
	(0.11)	(0.15)	(0.16)	(0.17)	(0.17)	(0.15)
Small value	–0.03	0.31	0.08	0.07	–0.31	–0.51
	(0.20)	(0.31)	(0.27)	(0.29)	(0.32)	(0.35)
Constant	–0.05	–1.90	–1.11	–0.30	0.89	2.31
	(0.25)	(0.39)	(0.27)	(0.38)	(0.40)	(0.57)

Notes: Standard errors in parentheses.

Source: Bassett and Chen (2001). Reprinted with the permission of Springer-Verlag.

Notice that the sum of the style parameters for a given regression is always one (except for rounding errors). To conserve space, I only present the results for the Magellan active fund and not those for the S&P – the latter exhibit very little variation in the estimates across the quantiles. The OLS results (column 2) show that the mean return has by far its biggest exposure to large value stocks (and this parameter estimate is also statistically significant), but it also exposed to small growth and, to a lesser extent, large growth stocks. It is of interest to compare the mean (OLS) results with those for the median, *Q*(0.5). The latter show much higher exposure to large value, less to small growth and none at all to large growth.

It is also of interest to examine the factor tilts as we move through the quantiles from left ($Q(0.1)$) to right ($Q(0.9)$). We can see that the loading on large growth monotonically falls from 0.31 at $Q(0.1)$ to 0.01 at $Q(0.9)$ while the loadings on large value and small growth substantially increase. The loading on small value falls from 0.31 at $Q(0.1)$ to -0.51 at $Q(0.9)$. A way to interpret (those of the current authors rather than those of Bassett and Chen) these results is to say that when the fund has historically performed poorly, this has resulted in equal amounts from its overweight exposure to large value and growth, and small growth. On the other hand, when it has historically performed well, this is a result of its exposure to large value and small growth but it was underweight small value stocks. Finally, it is obvious that the intercept (coefficient on the constant) estimates should be monotonically increasing from left to right since the quantile regression effectively sorts on average performance and the intercept can be interpreted as the performance expected if the fund had zero exposure to all of the styles.

KEY CONCEPTS

The key terms to be able to define and explain from this chapter are

- multiple regression model
- restricted regression
- residual sum of squares
- multiple hypothesis test
- R^2
- hedonic model
- data mining
- dummy variables
- variance-covariance matrix
- F-distribution
- total sum of squares
- non-nested hypotheses
- $\bar{R}^2$
- encompassing regression
- quantile regression
- qualitative data

Appendix 4.1 Mathematical Derivations of CLRM