Combining these two partial derivatives leads to:

$$\frac{\partial}{\partial m}\int_{-\infty}^{+\infty}|y-m|f(y)dy = F(m) - (1 - F(m)) = 2F(m) - 1. \quad \text{[A.2]}$$

By setting $2F(m) - 1 = 0$, we solve for the value of $F(m) = 1/2$, that is, the median, to satisfy the minimization problem.

Repeating the above argument for quantiles, the partial derivative for quantiles corresponding to Equation A.2 is:

$$\frac{\partial}{\partial q}E[d_p(Y,q)] = (1-p)F(q) - p(1-F(q)) = F(q) - p. \quad \text{[A.3]}$$

We set the partial derivative $F(q) - p = 0$ and solve for the value of $F(q) = p$ that satisfies the minimization problem.

3. QUANTILE-REGRESSION MODEL AND ESTIMATION

The quantile functions described in Chapter 2 are adequate for describing and comparing univariate distributions. However, when we model the relationship between a response variable and a number of independent variables, it becomes necessary to introduce a regression-type model for the quantile function, the quantile-regression model (QRM). Given a set of covariates, the linear-regression model (LRM) specifies the *conditional-mean* function whereas the QRM specifies the *conditional-quantile* function. Using the LRM as a point of reference, this chapter introduces the QRM and its estimation. It makes comparisons between the basic model setup for the LRM and that for the QRM, a least-squares estimation for the LRM and an analogous estimation approach for the QRM, and the properties of the two types of models. We illustrate our basic points using empirical examples from analyses of household income.[1]

Linear-Regression Modeling and Its Shortcomings

The LRM is a standard statistical method widely used in social-science research, but it focuses on modeling the conditional mean of a response variable without accounting for the full conditional distributional properties of the response variable. In contrast, the QRM facilitates analysis of the full conditional distributional properties of the response variable. The QRM and LRM are similar in certain respects, as both models deal with a continuous response variable that is linear in unknown parameters, but the QRM and LRM model different quantities and rely on different assumptions about error terms. To better understand these similarities and differences, we lay out the LRM as a starting point, and then introduce the QRM. To aid the explication, we focus on the single covariate case. While extending to more than one covariate necessarily introduces additional complexity, the ideas remain essentially the same.

Let y be a continuous response variable depending on x. In our empirical example, the dependent variable is household income. For x, we use an interval variable, *ED* (the household head's years of schooling), or alternatively a dummy variable, *BLACK* (the head's race, 1 for black and 0 for white). We consider data consisting of pairs (x_i, y_i) for $i = 1, \ldots, n$ based on a sample of micro units (households in our example).

By LRM, we mean the standard linear-regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad [3.1]$$

where ε_i is identically, independently, and normally distributed with mean zero and unknown variance σ^2. As a consequence of the mean zero assumption, we see that the function $\beta_0 + \beta_1 x$ being fitted to the data corresponds to the conditional mean of y given x (denoted by $E[y|x]$), which is interpreted as the average in the population of y values corresponding to a fixed value of the covariate x.

For example, when we fit the linear-regression Equation 3.1 using years of schooling as the covariate, we obtain the prediction equation $\hat{y} = -23127 + 5633ED$, so that plugging in selected numbers of years of schooling leads to the following values of conditional means for income.

ED	9	12	16
$E(y \mid ED)$	$27,570	$44,469	$67,001

Assuming a perfect fit, we would interpret these values as the average income for people with a given number of years of schooling. For example, the average income for people with nine years of schooling is $27,570.

Analogously, when we take the covariate to be *BLACK,* the fitted regression equation takes the form $\hat{y} = 53466 - 18268BLACK$, and plugging in the values of this covariate yields the following values.

BLACK	0	1
$E(y \mid BLACK)$	\$53,466	\$35,198

Again assuming the fitted model to be a reflection of what happens at the population level, we would interpret these values as averages in subpopulations, for example, the average income is \$53,466 for whites and \$35,198 for blacks.

Thus, we see that a fundamental aspect of linear-regression models is that they attempt to describe how the location of the conditional distribution behaves by utilizing the mean of a distribution to represent its central tendency. Another key feature of the LRM is that it invokes a homoscedasticity assumption; that is, the conditional variance, $Var(y|x)$, is assumed to be a constant σ^2 for all values of the covariate. When homoscedasticity fails, it is possible to modify LRM by allowing for simultaneous modeling of the conditional mean and the conditional scale. For example, one can modify the model in Equation 3.1 to allow for modeling the conditional scale: $y_i = \beta_0 + \beta_1 x_i + e^{\gamma}\varepsilon_i$, where γ is an additional unknown parameter and we can write $Var(y|x) = \sigma^2 e^{\gamma}$.

Thus, utilizing LRM reveals important aspects of the relationship between covariates and a response variable, and can be adapted to perform the task of modeling what is arguably the most important form of shape change for a conditional distribution: scale change. However, the estimation of conditional scale is not always readily available in statistical software. In addition, linear-regression models impose significant constraints on the modeler, and it is challenging to use LRM to model more complex conditional shape shifts.

To illustrate the kind of shape shift that is difficult to model using LRM, imagine a somewhat extreme situation in which, for some population of interest, we have a response variable y and a covariate x with the property that the conditional distribution of y has the probability density of the form shown in Figure 3.1 for each given value of $x = 1,2,3$. The three probability density functions in this figure have the same mean and standard deviation. Since the conditional mean and scale for the response variable y do not vary with x, there is no information to be gleaned by fitting a linear-regression model to samples from these populations. In order to understand how the covariate affects the response variable, a new tool is required. Quantile regression is an appropriate tool for accomplishing this task.

A third distinctive feature of the LRM is its *normality assumption.* Because the LRM ensures that the ordinary least squares provide the best possible fit for the data, we use the LRM without making the normality assumption for purely descriptive purposes. However, in social-science research, the LRM is used primarily to test whether an explanatory variable

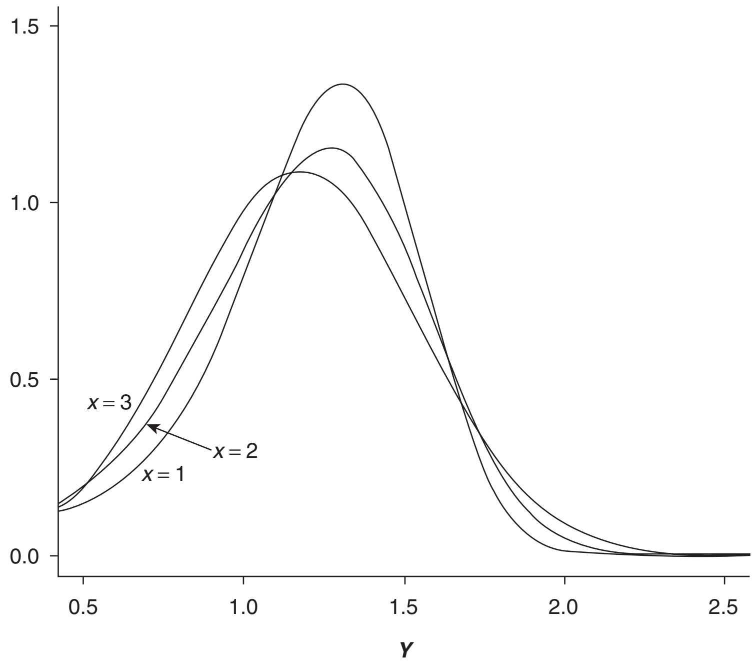

Figure 3.1 Conditional Distributions With the Same Mean and Standard Deviation but Different Skewness

significantly affects the dependent variable. Hypothesis testing goes beyond parameter estimation and requires determination of the sampling variability of estimators. Calculated *p*-values rely on the normality assumption or on large-sample approximation. Violation of these conditions may cause biases in *p*-values, thus leading to invalid hypothesis testing.

A related assumption made in the LRM is that the regression model used is appropriate for all data, which we call the *one-model assumption.* Outliers (cases that do not follow the relationship for the majority of the data) in the LRM tend to have undue influence on the fitted regression line. The usual practice used in the LRM is to identify outliers and eliminate them. Both the notion of outliers and the practice of eliminating outliers undermine much social-science research, particularly studies on social stratification and inequality, as outliers and their relative positions to those of the majority are important aspects of inquiry. In terms of modeling, one would simultaneously need to model the relationship for the majority cases and for the outlier cases, a task the LRM cannot accomplish.

All of the features just mentioned are exemplified in our household income data: the inadequacy of the conditional mean from a distributional point of view and violations of the homoscedasticity assumption, the normality assumption, and the one-model assumption. Figure 3.2 shows the distributions of income by education groups and racial groups. The location shifts among the three education groups and between blacks and whites are obvious, and their shape shifts are substantial. Therefore, the conditional mean from the LRM fails to capture the shape shifts caused by changes in the covariate (education or race). In addition, since the spreads differ substantially among the education groups and between the two racial groups, the homoscedasticity assumption is violated, and the standard errors are not estimated precisely. All box graphs in Figure 3.2 are right-skewed. Conditional-mean and conditional-scale models are not able to detect these kinds of shape changes.

By examining residual plots, we have identified seven outliers, including three cases with 18 years of schooling having an income of more than \$505,215 and four cases with 20 years of schooling having an income of more than \$471,572. When we add a dummy variable indicating membership in this outlier class to the regression model of income on education, we find that these cases contribute an additional \$483,544 to the intercept.

These results show that the LRM approach can be inadequate for a variety of reasons, including heteroscedasticity and outlier assumptions and the failure to detect multiple forms of shape shifts. These inadequacies are not restricted to the study of household income but also appear when other measures are considered. Therefore, it is desirable to have an alternative approach that is built to handle heteroscedasticity and outliers and detect various forms of shape changes.

As pointed out above, the conditional mean fails to identify shape shifts. The conditional-mean models also do not always correctly model central location shifts if the response distribution is asymmetric. For a symmetric distribution, the mean and median coincide, but the mean of a skewed distribution is no longer the same as the median (the .5th quantile). Table 3.1 shows a set of brief statistics describing the household income distribution. The right-skewness of the distribution makes the mean considerably larger than the median for both the total sample and for education and racial groups (see the first two rows of Table 3.1). When the mean and the median of a distribution do not coincide, the median may be more appropriate to capture the *central tendency* of the distribution. The location shifts among the three education groups and between blacks and whites are considerably smaller when we examine the median rather than the mean. This difference raises concerns about using the conditional mean as an appropriate measure for modeling the location shift of asymmetric distributions.

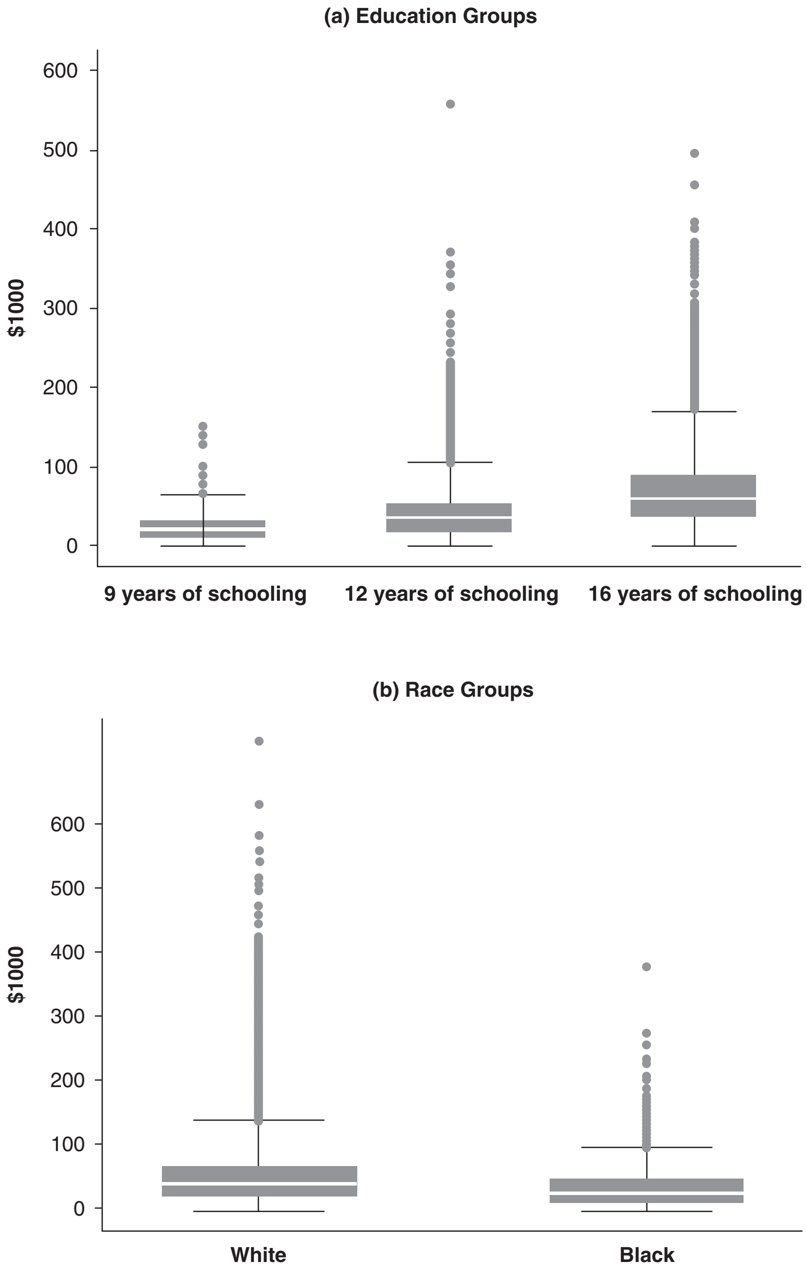

Figure 3.2 Box Graphs of Household Income

TABLE 3.1
Household Income Distribution:
Total, Education Groups, and Racial Groups

	Total	*ED = 9*	*ED = 12*	*ED = 16*	*WHITE*	*BLACK*
Mean	50,334	27,841	40,233	71,833	53,466	35,198
Quantile						
Median (.50th Quantile)	39,165	22,146	32,803	60,545	41,997	26,763
.10th Quantile	11,022	8,001	10,510	21,654	12,486	6,837
.25th Quantile	20,940	12,329	18,730	36,802	23,198	13,412
.75th Quantile	65,793	36,850	53,075	90,448	69,680	47,798
.90th Quantile	98,313	54,370	77,506	130,981	102,981	73,030
Quantile-Based Scale						
$(Q_{.75}-Q_{.25})$	44,853	24,521	34,344	53,646	46,482	34,386
$(Q_{.90}-Q_{.10})$	87,291	46,369	66,996	109,327	90,495	66,193
Quantile-Based Skewness						
$\frac{(Q_{.75}-Q_{.50})}{(Q_{.50}-Q_{.25})}-1$	.46	.50	.44	.26	.47	.58
$\frac{(Q_{.90}-Q_{.50})}{(Q_{.50}-Q_{.10})}-1$	1.10	1.28	1.01	.81	1.07	1.32

Conditional-Median and Quantile-Regression Models

With a skewed distribution, the median may become the more appropriate measure of central tendency; therefore, conditional-*median* regression, rather than conditional-*mean* regression, should be considered for the purpose of modeling location shifts. Conditional-median regression was proposed by Boscovich in the mid-18th century and was subsequently investigated by Laplace and Edgeworth. The median-regression model addresses the problematic conditional-mean estimates of the LRM. Median regression estimates the effect of a covariate on the conditional median, so it represents the central location even when the distribution is skewed.

To model both location shifts and shape shifts, Koenker and Bassett (1978) proposed a more general form than the median-regression model, the quantile-regression model (QRM). The QRM estimates the potential differential effect of a covariate on various quantiles in the conditional distribution, for example, a sequence of 19 equally distanced quantiles from the .05th quantile to the .95th quantile. With the median and the off-median quantiles, these 19 fitted regression lines capture the location shift (the line for the median), as well as scale and more complex shape shifts (the lines for off-median quantiles). In this way, the QRM estimates the differential effect of a covariate on the full distribution and accommodates heteroscedasticity.

Following Koenker and Bassett (1978), the QRM corresponding to the LRM in Equation 3.1 can be expressed as:

$$y_i = \beta_0^{(p)} + \beta_1^{(p)} x_i + \varepsilon_i^{(p)}, \quad [3.2]$$

where $0 < p < 1$ indicates the proportion of the population having scores below the quantile at p. Recall that for LRM, the conditional mean of y_i given x_i is $E(y_i|x_i) = \beta_0 + \beta_1 x_i$, and this is equivalent to requiring that the error term ε_i have zero expectation. In contrast, for the corresponding QRM, we specify that the pth conditional *quantile* given x_i is $Q^{(p)}(y_i|x_i) = \beta_0^{(p)} + \beta_1^{(p)} x_i$. Thus, the conditional pth quantile is determined by the quantile-specific parameters, $\beta_0^{(p)}$ and $\beta_1^{(p)}$, and a specific value of the covariate x_i. As for the LRM, the QRM can be formulated equivalently with a statement about the error terms ε_i. Since the term $\beta_0^{(p)} + \beta_1^{(p)} x_i$ is a constant, we have $Q^{(p)}(y_i|x_i) = \beta_0^{(p)} + \beta_1^{(p)} x_i + Q^{(p)}(\varepsilon_i) = \beta_0^{(p)} + \beta_1^{(p)} x_i$, so an equivalent formulation of QRM requires that the pth quantile of the error term be zero.

It is important to note that for different values of the quantile p of interest, the error terms $\varepsilon_i^{(p)}$ for fixed i are related. In fact, replacing p by q in Equation 3.2 gives $y_i = \beta_0^{(q)} + \beta_1^{(q)} x_i + \varepsilon_i^{(q)}$, which leads to $\varepsilon_i^{(p)} - \varepsilon_i^{(q)} = (\beta_0^{(q)} - \beta_0^{(p)}) + x_i(\beta_1^{(q)} - \beta_1^{(p)})$, so that the two error terms differ by a constant given x_i. In other words, the distributions of $\varepsilon_i^{(p)}$ and $\varepsilon_i^{(q)}$ are shifts of one another. An important special case of QRM to consider is one in which the $\varepsilon_i^{(p)}$ for $i = 1, \ldots, n$ are independent and identically distributed; we refer to this as the i.i.d. case. In this situation, the qth quantile of $\varepsilon_i^{(p)}$ is a constant $c_{p,q}$ depending on p and q and not on i. Using Equation 3.2, we can express the qth conditional-quantile function as $Q^{(q)}(y_i|x_i) = Q^{(p)}(y_i|x_i) + c_{p,q}$.[2] We conclude that in the i.i.d. case, the conditional-quantile functions are simple shifts of one another, with the slopes $\beta_1^{(p)}$ taking a common value β_1. In other words, the i.i.d. assumption says that there are no shape shifts in the response variable.

Equation 3.2 dictates that unlike the LRM in Equation 3.1, which has only one conditional mean expressed by one equation, the QRM can have numerous conditional quantiles. Thus, numerous equations can be expressed in the form of Equation 3.2.[3] For example, if the QRM specifies 19 quantiles, the 19 equations yield 19 coefficients for x_i, one at each of the 19 conditional quantiles ($\beta_1^{.05}, \beta_1^{.10}, \ldots, \beta_1^{.95}$). The quantiles do not have to be equidistant, but in practice, having them at equal intervals makes them easier to interpret.

Fitting Equation 3.2 in our example yields estimates for the 19 conditional quantiles of income given education or race (see Tables 3.2 and 3.3). The coefficient for education grows monotonically from \$1,019 at the .05th quantile to \$8,385 at the .95th quantile. Similarly, the black effect is weaker at the lower quantiles than at the higher quantiles.

The selected conditional quantiles on 12 years of schooling are:

p	.05	.50	.95
$E(y_i \mid ED_i = 12)$	\$7,976	\$36,727	\$111,268

and the selected conditional quantiles on blacks are:

p	.05	.50	.95
$E(y_i \mid BLACK_i = 1)$	\$5,432	\$26,764	\$91,761

These results are very different from the conditional mean of the LRM. The conditional quantiles describe a conditional distribution, which can be used to summarize the location and shape shifts. Interpreting QRM estimates is a topic of Chapters 5 and 6.

Using a random sample of 1,000 households from the total sample and the fitted line based on the LRM, the left panel of Figure 3.3 presents the scatterplot of household income against the head of household's years of schooling. The single regression line indicates mean shifts, for example, a mean shift of \$22,532 from 12 years of schooling to 16 years of schooling

TABLE 3.2
Quantile-Regression Estimates for Household Income on Education

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*	*(14)*	*(15)*	*(16)*	*(17)*	*(18)*	*(19)*
ED	1,019	1,617	2,023	2,434	2,750	3,107	3,397	3,657	3,948	4,208	4,418	4,676	4,905	5,214	5,557	5,870	6,373	6,885	8,385
	(28)	(31)	(40)	(39)	(44)	(51)	(57)	(64)	(66)	(72)	(81)	(92)	(88)	(102)	(127)	(138)	(195)	(274)	(463)
Constant	−4,252	−7,648	−9,170	−11,160	−12,056	−13,308	−13,783	−13,726	−14,026	−13,769	−12,546	−11,557	−9,914	−8,760	−7,371	−4,227	−1,748	4,755	10,648
	(380)	(424)	(547)	(527)	(593)	(693)	(764)	(866)	(884)	(969)	(1,084)	(1,226)	(1,169)	(1,358)	(1,690)	(1,828)	(2,582)	(3,619)	(6,101)

NOTE: Standard errors in parentheses.

TABLE 3.3
Quantile-Regression Estimates for Household Income on Race

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*	*(14)*	*(15)*	*(16)*	*(17)*	*(18)*	*(19)*
BLACK	−3,124	−5,649	−7,376	−8,848	−9,767	−11,232	−12,344	−13,349	−14,655	−15,233	−16,459	−17,417	−19,053	−20,314	−21,879	−22,914	−26,063	−29,951	−40,639
	(304)	(306)	(421)	(485)	(584)	(536)	(609)	(708)	(781)	(765)	(847)	(887)	(1,050)	(1,038)	(1,191)	(1,221)	(1,435)	(1,993)	(3,573)
Constant	8,556	12,486	16,088	19,718	23,198	26,832	30,354	34,024	38,047	41,997	46,635	51,515	56,613	62,738	69,680	77,870	87,996	102,981	132,400
	(115)	(116)	(159)	(183)	(220)	(202)	(230)	(268)	(295)	(289)	(320)	(335)	(397)	(392)	(450)	(461)	(542)	(753)	(1,350)

NOTE: Standard errors in parentheses.

(5633 · (16 – 12)). However, this regression line does not capture shape shifts.

The right panel of Figure 3.3 shows the same scatterplot as in the left panel and the 19 quantile-regression lines. The .5th quantile (the median) fit captures the central location shifts, indicating a positive relationship between conditional-median income and education. The slope is $4,208, shifting $16,832 from 12 years of schooling to 16 years of schooling (4208 · (16 – 12)). This shift is lower than the LRM mean shift.

In addition to the estimated location shifts, the other 18 quantile-regression lines provide information about shape shifts. These regression lines are positive, but with different slopes. The regression lines cluster tightly at low levels of education (e.g., 0–5 years of schooling) but deviate from each other more widely at higher levels of education (e.g., 16–20 years of

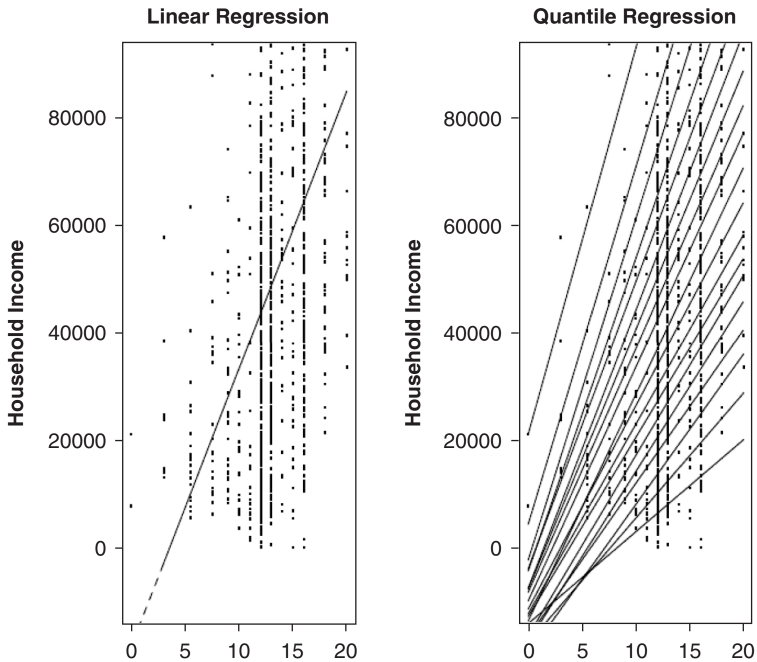

Figure 3.3 Effects of Education on the Conditional Mean and Conditional Quantiles of Household Income: A Random Sample of 1,000 Households

schooling). A shape shift is described by the tight cluster of the slopes at lower levels of education and the scattering of slopes at higher levels of education. For instance, the spread of the conditional income on 16 years of schooling (from $12,052 for the .05th conditional quantile to $144,808 for the .95th conditional quantile) is much wider than that on 12 years of schooling (from $7,976 for the .05th conditional quantile to $111,268 for the .95th conditional quantile). Thus, the off-median conditional quantiles isolate the location shift from the shape shift. This feature is crucial for determining the impact of a covariate on the location and shape shifts of the conditional distribution of the response, a topic discussed in Chapter 5 with the interpretation of the QRM results.

QR Estimation

We review least-squares estimation so as to place QR estimation in a familiar context. The least-squares estimator solves for the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ by taking those values of the parameters that minimize the sum of squared residuals:

$$\min \textstyle\sum_i (y_i - (\beta_0 + \beta_1 x_i))^2. \qquad [3.3]$$

If the LRM assumptions are correct, the fitted response function $\hat{\beta}_0 + \hat{\beta}_1$ approaches the population conditional mean $E(y\,|\,x)$ as the sample size goes to infinity. In Equation 3.3, the expression minimized is the sum of squared vertical distances between data points (x_i, y_i) and the fitted line $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

A closed-form solution to the minimization problem is obtained by (a) taking partial derivatives of Equation 3.3 with respect to β_0 and β_1, respectively; (b) setting each partial derivative equal to zero; and (c) solving the resulting system of two equations with two unknowns. We then arrive at the two estimators:

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}, \ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

A significant departure of the QR estimator from the LR estimator is that in the QR, the distance of points from a line is measured using a weighted sum of vertical distances (without squaring), where the weight is $1 - p$ for points below the fitted line and p for points above the line. Each choice for this proportion p, for example, $p = .10, .25, .50$, gives rise to a different fitted conditional-quantile function. The task is to find an estimator with the desired property for each possible p. The reader is reminded of the discussion in Chapter 2 where it was indicated that the mean of a distribution can be viewed as the point that minimizes the average squared distance over the population, whereas a quantile q can be viewed as the point that minimizes an average weighted distance, with weights depending on whether the point is above or below the value q.

For concreteness, we first consider the estimator for the median-regression model. In Chapter 2, we described how the median (m) of y can be viewed as the minimizing value of $E|y - m|$. For an analogous prescription in the median-regression case, we choose to minimize the sum of absolute residuals. In other words, we find the coefficients that minimize the sum of absolute residuals (the absolute distance from an observed value to its fitted value). The estimator solves for the βs by minimizing Equation 3.4:

$$\Sigma_i | y_i - \beta_0 - \beta_1 x_i |. \qquad [3.4]$$

Under appropriate model assumptions, as the sample size goes to infinity, we obtain the conditional median of y given x at the population level.

When expression Equation 3.4 is minimized, the resulting solution, which we refer to as the *median-regression line,* must pass through a pair of data points with half of the remaining data lying above the regression line and the other half falling below. That is, roughly half of the residuals are positive and half are negative. There are typically multiple lines with this property, and among these lines, the one that minimizes Equation 3.4 is the solution.

Algorithmic Details

In this subsection, we describe how the structure of the function Equation 3.4 makes it amenable to finding an algorithm for its minimization. Readers who are not interested in this topic can skip this section.

The left panel of Figure 3.4 shows eight hypothetical pairs of data points (x_i, y_i) and the 28 lines $(8(8-1)/2 = 28)$ connecting a pair of these points is plotted. The dashed line is the fitted median-regression line, that is, the line that minimizes the sum of absolute vertical distance from all data points. Observe that of the six points not falling on the median-regression line, half of the points are below it and the other half are above it. Every line in the (x, y) plane takes the form $y = \beta_0 + \beta_1 x$ for some choice of intercept-slope pair (β_0, β_1), so that we have a correspondence between *lines* in the (x, y) plane and *points* in the (β_0, β_1) plane. The right panel of Figure 3.4 shows a plot in the (β_0, β_1) plane that contains a point corresponding to every line in the left panel. In particular, the solid circle shown in the right panel corresponds to the median-regression line in the left panel.

In addition, if a line with intercept and slope (β_0, β_1) passes through a given point (x_i, y_i), then $y_i = \beta_0 + \beta_1 x_i$, so that (β_0, β_1) lies on the line $\beta_1 = (y_i/x_i) - (1/x_i)\beta_0$. Thus, we have established a correspondence between points in the (x, y) plane and lines in the (β_0, β_1) plane and vice versa, a phenomenon referred to as *point/line duality* (Edgeworth, 1888).

The eight lines shown in the right panel of Figure 3.4 correspond to the eight data points in the left panel. These lines divide the (β_0, β_1) plane into polygonal regions. An example of such a region is shaded in Figure 3.4. In any one of these regions, the points correspond to a family of lines in the (x, y) plane, all of which divide the data set into two sets in exactly the same way (meaning that the data points above one line are the same as the points above the other). Consequently, the function of (β_0, β_1) that we seek to minimize in Equation 3.4 is linear in each region, so that this function is convex with a graph that forms a polyhedral surface, which is plotted from two different angles in Figure 3.5 for our example. The vertices,

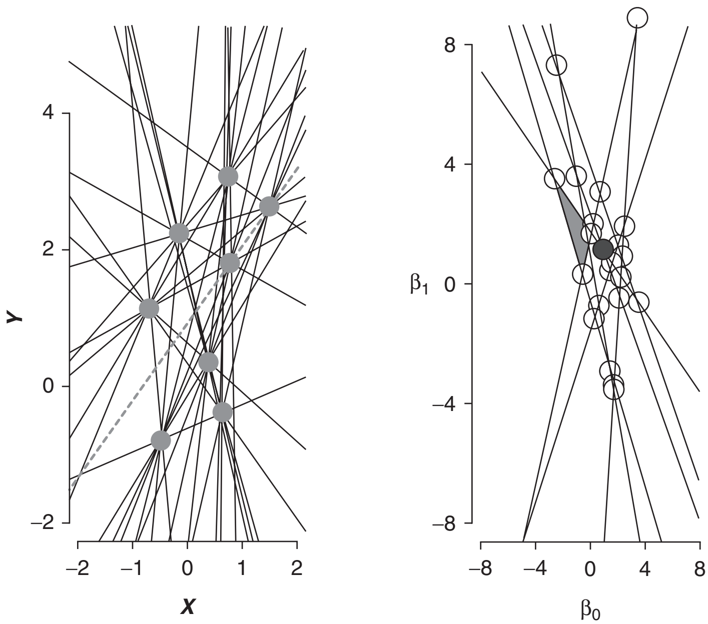

Figure 3.4 An Illustration of Point/Line Duality

edges, and facets of the surface project to points, line segments, and regions, respectively, in the (β_0, β_1) plane shown in the right-hand panel of Figure 3.4. Using the point/line duality correspondence, each vertex corresponds to a line connecting a pair of data points. An edge connecting two vertices in the surface corresponds to a pair of such lines, where one of the data points defining the first line is replaced by another data point, and the remaining points maintain their position (above or below) relative to both lines.

An algorithm for minimization of the sum of absolute distances in Equation 3.4, one thus leading to the median-regression coefficients $(\hat{\beta}_0, \hat{\beta}_1)$, can be based on exterior-point algorithms for solving linear-programming problems. Starting at any one of the points (β_0, β_1) corresponding to a vertex, the minimization is achieved by iteratively moving from vertex to

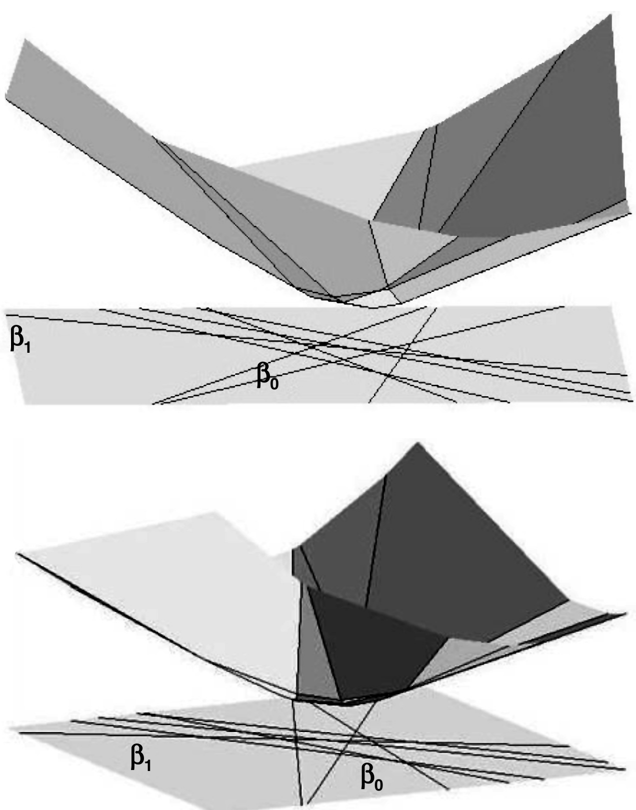

Figure 3.5 Polyhedral Surface and Its Projection

vertex along the edges of the polyhedral surface, choosing at each vertex the path of the steepest descent until arriving at the minimum. Using the correspondence described in the previous paragraph, we iteratively move from line to line defined by pairs of data points, at each step deciding which new data point to swap with one of the two current ones by picking the one that leads to the smallest value in Equation 3.4. The minimum sum of absolute errors is attained at the point in the (β_0, β_1) plane below the lowest vertex of the surface. A simple argument involving the directional derivative with respect to β_0 (similar to the one in Chapter 2 showing that the median is the solution to a minimization problem) leads to the conclusion that the same number of data points lie above the median-regression line as lie below it.

The median-regression estimator can be generalized to allow for *p*th quantile-regression estimators (Koenker & d'Orey, 1987). Recall from the discussion in Chapter 2 that the *p*th quantile of a univariate sample $y_1, \ldots, y_n$ distribution is the value q that minimizes the sum of weighted distances from the sample points, where points below q receive a weight of $1 - p$ and points above q receive a weight of p. In a similar manner, we define the *p*th quantile-regression estimators $\hat{\beta}_0^{(p)}$ and $\hat{\beta}_1^{(p)}$ as the values that minimize the weighted sum of distances between fitted values $\hat{y}_i = \hat{\beta}_0^{(p)} + \hat{\beta}_1^{(p)} x_i$ and the y_i, where we use a weight of $1 - p$ if the fitted value underpredicts the observed value y_i and a weight of p otherwise. In other words, we seek to minimize a weighted sum of residuals $y_i - \hat{y}_i$ where positive residuals receive a weight of p and negative residuals receive a weight of $1 - p$. Formally, the *p*th quantile-regression estimators $\hat{\beta}_0^{(p)}$ and $\hat{\beta}_1^{(p)}$ are chosen to minimize

$$\sum_{i=1}^{n} d_p(y_i, \hat{y}_i) = p \sum_{y_i \geq \beta_0^{(p)} + \beta_1^{(p)} x_i} |y_i - \beta_0^{(p)} - \beta_1^{(p)} x_i| + (1 - p) \sum_{y_i < \beta_0^{(p)} + \beta_1^{(p)} x_i} |y_i - \beta_0^{(p)} - \beta_1^{(p)} x_i|, \qquad [3.5]$$

where d_p is the distance introduced in Chapter 2. Thus, unlike Equation 3.4, which states that the negative residuals are given the same importance as the positive residuals, Equation 3.5 assigns different weights to positive and negative residuals. Observe that in Equation 3.5, the first sum is the sum of vertical distances of data points from the line $y = \beta_0^{(p)} + \beta_1^{(p)} x$, for points lying above the line. The second is a similar sum over all data points lying below the line.

Observe that, contrary to a common misconception, the estimation of coefficients for each quantile regression is based on the weighted data of the whole sample, not just the portion of the sample at that quantile.

An algorithm for computing the quantile-regression coefficients $\hat{\beta}_0^{(p)}$ and $\hat{\beta}_1^{(p)}$ can be developed along lines similar to those outlined for the median-regression coefficients. The pth quantile-regression estimator has a similar property to one stated for the median-regression estimator: The proportion of data points lying below the fitted line $y = \hat{\beta}_0^{(p)} + \hat{\beta}_1^{(p)}x$ is p, and the proportion lying above is $1 - p$.

For example, when we estimate the coefficients for the .10th quantile-regression line, the observations below the line are given a weight of .90 and the ones above the line receive a smaller weight of .10. As a result, 90% of the data points (x_i, y_i) lie above the fitted line leading to positive residuals, and 10% lie below the line and thus have negative residuals. Conversely, to estimate the coefficients for the .90th quantile regression, points below the line are given a weight of .10, and the rest have a weight of .90; as a result, 90% of observations have negative residuals and the remaining 10% have positive residuals.

Transformation and Equivariance

In analyzing a response variable, researchers often transform the scale to aid interpretation or to attain a better model fit. Equivariance properties of models and estimates refer to situations when, if the data are transformed, the models or estimates undergo the same transformation. Knowledge of equivariance properties helps us to reinterpret fitted models when we transform the response variable.

For any linear transformation of the response variable, that is, the addition of a constant to y or the multiplication of y by a constant, the conditional mean of the LRM can be exactly transformed. The basis for this statement is the fact that for any choice of constants a and c, we can write

$$E(c + ay|x) = c + aE(y|x). \quad [3.6]$$

For example, if every household in the population received \$500 from the government, the conditional mean would also be increased by \$500 (the new intercept would be increased by \$500). When the \$1 unit of income is transformed to the \$1,000 unit, the conditional mean in the \$1 unit is increased by 1,000 times as well (the intercept and the slope are both multiplied by 1,000 to be on the dollar scale). Similarly, if the dollar unit for wage rate is transformed to the cent unit, the conditional mean (the intercept and the slope) is divided by 100 to be on the dollar scale again. This property is termed *linear equivariance* because the linear transformation is the same for the dependent variable and the conditional mean. The QRM also has this property:

$$Q^{(p)}(c + ay|x) = c + a(Q^{(p)}[y \mid x]), \qquad [3.7]$$

provided that a is a positive constant. If a is negative, we have $Q^{(p)}(c + ay \mid x) = c + a(Q^{(1-p)}[y|x])$ because the order is reversed.

Situations often arise in which nonlinear transformation is desired. Log transformations are frequently used to address the right-skewness of a distribution. Other transformations are considered in order to make a distribution appear more normal or to achieve a better model fit.

Log transformations are also introduced in order to model a covariate's effect in relative terms (e.g., percentage changes). In other words, the effect of a covariate is viewed on a multiplicative scale rather than on an additive one. In our example, the effects of education or race were previously expressed in additive terms (the dollar unit), and it may be desirable to measure an effect in multiplicative terms, for example, in terms of percentage changes. For example, we can ask: What is the percentage change in conditional-mean income brought about by one more year of schooling? The coefficient for education in a log income equation (multiplied by 100) approximates the percentage change in conditional-mean income brought about by one more year of schooling. However, under the LRM, the conditional mean of log income is not the same as the log of conditional-mean income. Estimating two LRMs using income and log income yields two fitted models:

$$\hat{y} = -23{,}127 + 5{,}633ED, \ \log \hat{y} = 8.982 + .115ED.$$

The result from the log income model suggests that one more year of education increases the conditional-mean income by about 11.5%.[4] The conditional mean of the income model at 10 years of schooling is \$33,203, the log of which becomes 8.108. The conditional mean of the log income model at the same schooling level is 10.062, a much larger figure than the log of the conditional mean of income (8.108). While the log transformation of a response in the LRM allows an interpretation of LRM estimates as a percentage change, the conditional mean of the response in absolute terms is impossible to obtain from the conditional mean on the log scale:

$$E(\log y \mid x) \neq \log[E(y \mid x)] \text{ and } E(y_i \mid x_i) \neq e^{E[\log y_i \mid x_i]}. \qquad [3.8]$$

Specifically, if our aim is to estimate the education effect in absolute terms, we use the income model, whereas for the impact of education in relative terms, we use the log income model. Although the two objectives are related to each other, the conditional means of the two models are not related through any simple transformation.[5] Thus, it would be a mistake to use the log income results to make conclusions about the distribution of income (though this is a widely used practice).

The log transformation is one member of the family of *monotone* transformations, that is, transformations that preserve order. Formally, a transformation h is a monotone if $h(y) < h(y')$ whenever $y < y'$. For variables taking positive values, the power transformation $h(y) = y^{\phi}$ is monotone for a fixed positive value of the constant ϕ. As a result of nonlinearity, when we apply a monotone transformation, the degree to which the transformation changes the value of y can differ from one value of y to the next. While the property in Equation 3.6 holds for linear functions, it is not the case for general monotone functions, that is, $E(h(y)|x) \neq h(E(y_i|x_i))$. Generally speaking, the "monotone equivariance" property fails to hold for conditional means, so that LRMs do not possess monotone equivariance.

By contrast, the conditional quantiles do possess monotone equivariance; that is, for a monotone function h, we have

$$Q^{(p)}(h(y)\,|\,x) = h(Q^{(p)}[y\,|\,x]). \qquad [3.9]$$

This property follows immediately from the version of monotone equivariance stated for univariate quantiles in Chapter 2. In particular, a conditional quantile of log y is the log of the conditional quantile of y:

$$Q^{(p)}(\log(y)\,|\,x) = \log(Q^{(p)}[y\,|\,x]), \qquad [3.10]$$

and equivalently,

$$Q^{(p)}(y\,|\,x) = e^{Q^{(p)}[\log(y)\,|\,x]}, \qquad [3.11]$$

so that we are able to reinterpret fitted quantile-regression models for untransformed variables to quantile-regression models for transformed variables. In other words, assuming a perfect fit for the pth quantile function of the form $Q^{(p)}(y\,|\,x) = \beta_0 + \beta_1 x$, we have $Q^{(p)}(\log\ y\,|\,x) = \log(\beta_0 + \beta_1 x)$, so that we can use the impact of a covariate expressed in absolute terms to describe the impact of a covariate in relative terms and vice versa.

Take the conditional median as an example:

$$Q^{(.50)}(y_i\,|\,ED_i) = -13769 + 4208ED_i,\ Q^{(.50)}(\log(y_i)\,|\,ED_i) = 8.966 + .123ED_i.$$

The conditional median of income at 10 years of schooling is \$28,311. The log of this conditional median, 10.251, is similar to the conditional median of the log income equation at the same education level, 10.196. Correspondingly, when moving from log to raw scale, in absolute terms, the conditional median at 10 years of schooling from the log income equation is $e^{10.916} = 28{,}481$.

The QRM's monotone equivariance is particularly important for research involving skewed distributions. While the original distribution is distorted by the reverse transformation of log-scale estimates if the LRM is used, the original distribution is preserved if the QRM is used. A covariate's effect on the response variable in terms of percentage change is often used in inequality research. Hence, the monotone equivariance property allows researchers to achieve both goals: measuring percentage change caused by a unit change in the covariate and measuring the impact of this change on the location and shape of the raw-scale conditional distribution.

Robustness

Robustness refers to insensitivity to outliers and to the violation of model assumptions concerning the data *y*. Outliers are defined as some values of *y* that do not follow the relationship for the majority values. Under the LRM, estimates can be sensitive to outliers. Earlier in the first section of this chapter, we presented an example showing how outliers of income distribution distort the mean and the conditional mean. The high sensitivity of the LRM to outliers has been widely recognized. However, the practice of eliminating outliers does not satisfy the objective of much social-science research, particularly inequality research.

In contrast, the QRM estimates are not sensitive to outliers.[6] This robustness arises because of the nature of the distance function in Equation 3.5 that is minimized, and we can state a property of quantile-regression estimates that is similar to a statement made in Chapter 2 about univariate quantiles. If we modify the value of the response variable for a data point lying above (or below) the fitted quantile-regression line, as long as that data point remains above (or below) the line, the fitted quantile-regression line remains unchanged. Stated another way, if we modify values of the response variable without changing the sign of the residual, the fitted line remains the same. In this way, as for univariate quantiles, the influence of outliers is quite limited.

In addition, since the covariance matrix of the estimates is calculated under the normality assumption, the LRM's normality assumption is necessary for obtaining the inferential statistics of the LRM. Violation of the normality assumption can cause inaccuracy in standard errors. The QRM is robust to distributional assumptions because the estimator weighs the local behavior of the distribution near the specific quantile more than the remote behavior of the distribution. The QRM's inferential statistics can be distribution free (a topic discussed in Chapter 4). This robustness is important in studying phenomena of highly skewed distributions such as income, wealth, educational, and health outcomes.

Summary

This chapter introduces the basics of the quantile-regression model in comparison with the linear-regression model, including the model setup, the estimation, and the properties of estimates. The QRM inherits many of the properties of sample quantiles introduced in Chapter 2. We explain how LRM is inadequate for revealing certain types of effects of covariates on the distribution of a response variable. We also highlight some of the key features of QRM. We present many of the important differences between the QRM and the LRM, namely, (a) multiple-quantile-regression fits versus single-linear-regression fits to data; (b) quantile-regression estimation that minimizes a weighted sum of absolute values of residuals as opposed to minimizing the sum of squares in least-squares estimation; and (c) the monotone equivariance and robustness to distributional assumptions in conditional quantiles versus the lack of these properties in the conditional mean. With these basics, we are now ready to move on to the topic of QRM inference.

Notes

1. The data are drawn from the 2001 panel of the Survey of Income and Program Participation (SIPP). Household income is the annual income in 2001. The analytic sample for Chapters 3 through 5 includes 19,390 white households and 3,243 black households.

2. $Q^{(q)}(y_i \mid x_i) = Q^{(q)}(\beta_0^{(p)} + x_i\beta_1^{(p)} + \varepsilon_i^{(p)}) = \beta_0^{(p)} + x_i\beta_1^{(p)} + Q^{(q)}(\varepsilon_i^{(p)}) = Q^{(p)}(y_i \mid x_i) + c_{p,q}$.

3. The number of distinct quantile solutions, however, is bounded by the finite sample size.

4. Precisely, the percentage change is $100(e^{.115}-1) = 12.2\%$.

5. The conditional mean is proportional to the exponential of the linear predictor (Manning, 1998). For example, if the errors are normally distributed $N(0, \sigma_\varepsilon^2)$, then $E(y_i \mid x_i) = e^{\beta_0 + \beta_1 x_i + 0.5\sigma_\varepsilon^2}$. The term $e^{0.5\sigma_\varepsilon^2}$ is sometimes called the smearing factor.

6. Note that this robustness does not apply to outliers of covariates.