07-Hao.qxd 3/13/2007 4:30 PM Page 92

7. APPLICATION TO INCOME INEQUALITY IN 1991 AND 2001

The empirical illustrations in previous chapters used oversimplified specifications with one or two covariates. This chapter applies the techniques in the book to a particular topic: the persistence and widening of household income inequality from 1991 to 2001. Our goal is to systematically summarize the techniques developed in this book via a concrete empirical application. Drawing from the U.S. Survey of Income and Program Participation (SIPP), we add the 1991 data to the previously used 2001 data. Household income is adjusted to the 2001 constant dollar. We specify a parsimonious model for household income as a function of five factors (13 covariates): life cycle (age and age-squared), race-ethnicity (white, black, Hispanic, and Asian), education (college graduate, some college, high school graduate, and without high-school education), household types (married couple with children, married couple without children, female head with children, single person, and other), and rural residence. This is the specification used throughout the chapter. Models for both raw-scale income and log-transformed income are fitted. The analyses include (a) assessing the goodness of fit for raw-scale and log-scale income models, (b) comparing ordinary-least-squares (OLS) and median-regression estimates, (c) examining coefficients at the two tails, (d) graphically viewing 19 sets of coefficient estimates and their confidence intervals, and (e) attaining location and shape shifts of conditional quantiles for each covariate in each year and examining the trend over the decade.

Observed Income Disparity

Figure 7.1 shows 99 empirical quantiles for race-ethnicity groups and education groups in 1991 and 2001. One of the most interesting features is the greater spread for the middle 98% of the members in each group in 2001 as compared to 1991.

More detailed comparisons require the actual values of the quantiles. Table 7.1 compares the .025th-quantile, median, and .975th-quantile household incomes (in 2001 constant dollars) for 1991 and 2001. The numbers are weighted to reflect population patterns. A common characteristic is observed for the total and each subgroup: The stretch ($QSC_{.025}$) for the middle 95% households is much wider for 2001 than for 1991, indicating growing total and within-group disparities in income over the decade.

The racial disparity between whites and others in the lower half of the income distribution declined in the last decade. This decline can be seen as

07-Hao.qxd 3/13/2007 4:30 PM Page 93

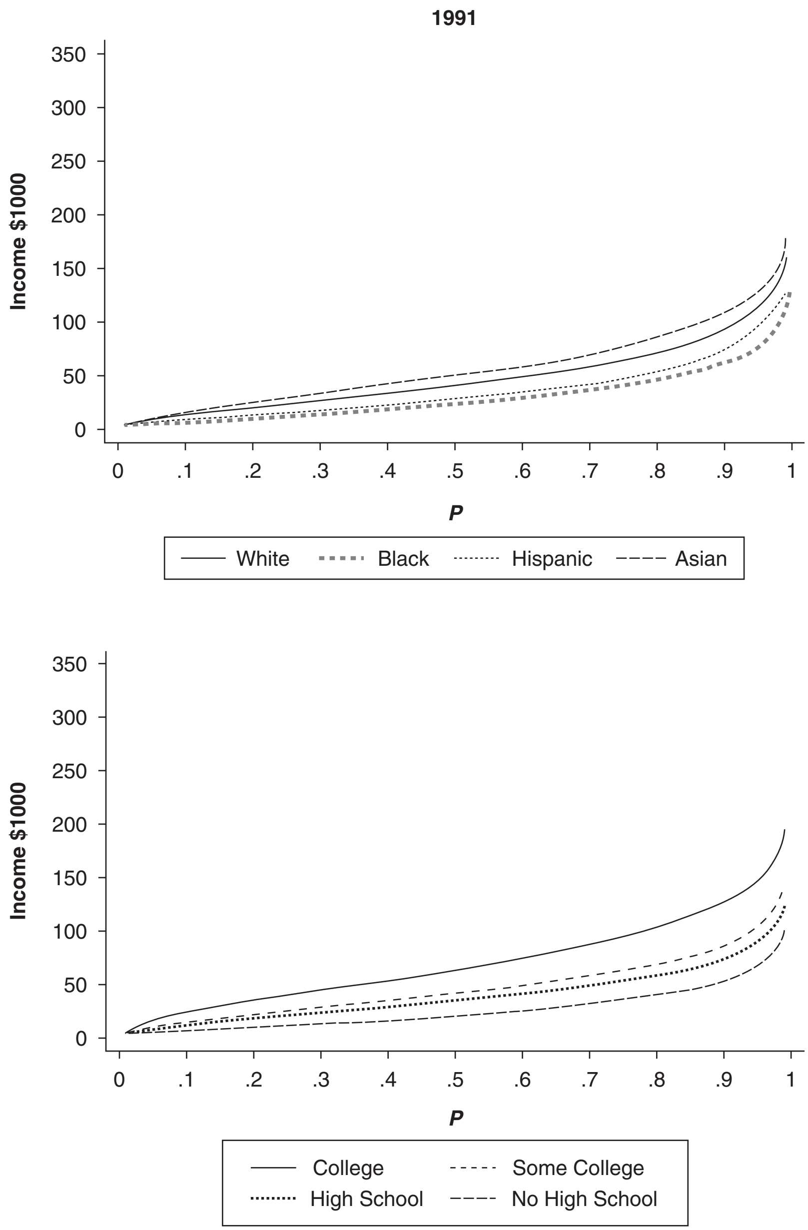

Figure 7.1 Empirical Quantile Functions by Race-Ethnicity and Education Groups

(Continued)

07-Hao.qxd 3/13/2007 4:30 PM Page 94

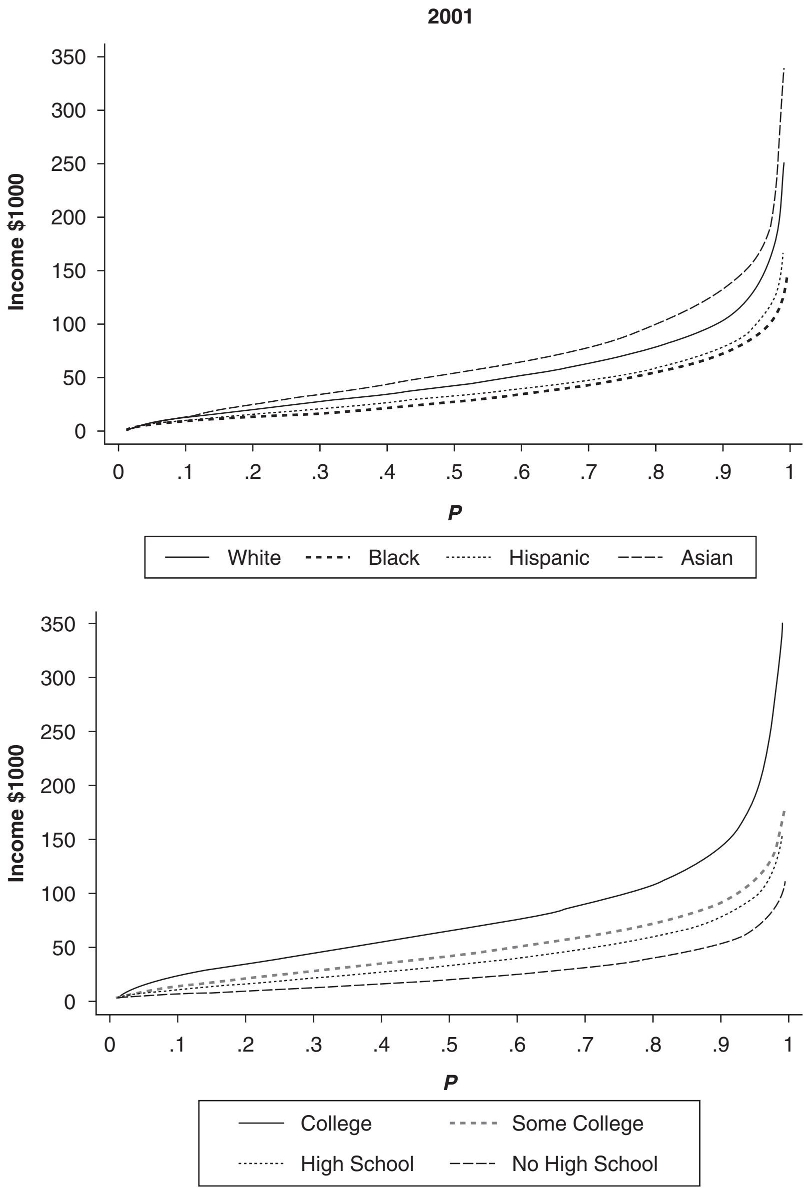

Figure 7.1 (Continued)

TABLE 7.1
Household Income Distribution by Groups: 1991 and 2001

	Quantile					
	1991			*2001*		
Group	*.025*	*.500*	*.975*	*.025*	*.500*	*.975*
Total	6256	38324	131352	6000	40212	164323
Race-Ethnicity						
White	6765	40949	135443	6600	42878	172784
Black	3773	23624	101160	3788	27858	113124
Hispanic	5342	28851	114138	5600	33144	119454
Asian	5241	49354	149357	4800	55286	211112
Education						
College graduate	11196	64688	168912	10910	65298	263796
Some college	8059	42082	120316	6364	41901	134796
High school grad.	6392	35723	104102	5347	33246	118162
No high school	4918	20827	80603	4408	20319	79515
Household Type						
Married w/ children	12896	55653	143343	14193	61636	204608
Married w/o children	11621	43473	146580	10860	47665	176375
Female head	3666	23420	94114	3653	27690	96650
Single person	4884	20906	83213	3977	21369	91551
Other household type	7301	37896	115069	6600	41580	150123
Residence						
Urban	6330	40732	137574	6199	42504	174733
Rural	6122	32874	111891	5419	33505	118079

the fall in the .025th-quantile income of white households in contrast with a moderate gain for the black and Hispanic counterparts. Asians made greater headway than whites at the median and at the .975th quantile, but the lowest 2.5% of Asian households were left behind.

An important change in income inequality is the change in returns to education for the top tail. While most college graduates gained an ample amount of income over the decade, more than half of the people with a below-college education saw their income actually decline. In particular, more than 97.5% of high school dropouts in 2001 had a notably lower income than their 1991 counterparts.

Consideration of household type, defined by marriage and presence of children, leads us to another arena where social stratification reshapes the income distribution. Progress is seen for married couples with children, whereas the income of single-mother families and single-person households is stagnant. Inequality between urban and rural areas and inequality within both urban and rural areas intensified over the decade studied.

07-Hao.qxd 3/13/2007 4:30 PM Page 96

Descriptive Statistics

Table 7.2 presents the weighted mean and standard deviation for variables used in the analyses. We see that mean income increased by nearly $5,000 from 1991 to 2001, a much higher figure than the growth in median income observed in the previous table. The small increase in log income reminds us that the log transformation contracts the right tail of the distribution. We observe greater diversity in the race-ethnicity structure and considerable improvement in the population's education. However, the number of households of married couples with children decreased, whereas "other" types and single-person households were on the rise. The United States continued the urbanization and suburbanization seen in previous decades.

TABLE 7.2
Descriptive Statistics of Variables Used in Analysis

	1991		*2001*	
Variable	*Mean*	*SD*	*Mean*	*SD*
Response				
Income ($)	46168	33858	51460	46111
Log income	10.451	0.843	10.506	0.909
Age	49	17	49	17
Age-squared	2652	1798	2700	1786
Covariate				
Race-Ethnicity				
White	.795	.404	.755	.430
Black	.101	.301	.118	.322
Hispanic	.079	.269	.094	.292
Asian	.025	.157	.033	.177
Education				
College graduate	.230	.421	.261	.439
Some college	.210	.407	.296	.457
High school grad.	.341	.474	.302	.459
No high school	.219	.414	.141	.348
Household Type				
Married w/ children	.330	.470	.287	.452
Married w/o children	.224	.417	.233	.423
Female head	.108	.310	.104	.305
Single person	.257	.437	.267	.442
Other household type	.082	.274	.110	.313
Residence				
Urban	.732	.443	.773	.419
Rural	.268	.443	.227	.419
Sample Size	10111		25891	

Notes on Survey Income Data

Two characteristics of survey income data make the QRM approach a better strategy for analysis than the LRM. First, only 0.2% of the households have incomes over a million dollars, whereas for over 96% of the population, income is less than $100,000. Thus, data for the very rich profoundly influence the OLS coefficient estimates. Second, survey income is often top-coded for each income source; thus, it is not straightforward to assess at which level a household's *total* income is trimmed. In addition, surveys in different years may use different top-coding criteria, resulting in a tedious process to make the data from different years comparable. These problems are not concerns in quantile-regression modeling owing to the robustness property of the QRM described in Chapter 3. In this example, we choose the two extremes to be the .025th and .975th quantiles, thus focusing on modeling the middle 95% of the population. Since data points that have been top-coded tend to be associated with *positive* residuals for the fitted .975th QRM, the effect on the QRM estimates of replacing the (unknown) income values with top-coded values tends to be minimal. This simplifies data management since we can include in the analysis all survey data points, top-coded or not.

Throughout this example, each covariate is centered at its mean. Consequently, the constant term from the income OLS regression represents the mean income of the population, whereas the constant term from the log-income OLS regression represents the mean log income. For the fitted QRM models based on centered covariates, the constant term for the income quantile regression represents the conditional quantile for income at the typical setting, and the constant term for the log income represents the conditional quantile for log income at the typical setting.

Goodness of Fit

Because the QRM no longer makes linear-regression assumptions, raw-scale income can be used without transformation. Nevertheless, we would like to choose a better-fitting model if log transformation can achieve it. We thus perform comparisons of goodness of fit between the income equation and the log-income equation. We fit separate QRMs at the 19 equally spaced quantiles (a total of $2 \times 19 = 38$ fits), using Stata's "qreg" command. Although the qreg command produces the asymptotic standard errors (which can be biased), we are only interested in the goodness-of-fit statistics, the QRM Rs. Table 7.3 shows the QRM's *R*s (defined in Chapter 5) for the raw- and log-scale response.

In general, log transformation yields a better fit of the model to the data than the raw scale. For the 1991 data, the *R* of log income is higher for

07-Hao.qxd 3/13/2007 4:30 PM Page 98

$0 < p < .65$—nearly two thirds of the 19 quantiles examined gain a better fit. For the 2001 data, the R of log income is higher for $0 < p < .85$, presenting a stronger case for using log transformation for the 2001 data than for the 1991 data. However, the log scale does not fit as well at the top tail. If the top-tail behavior and stratification are the major concern, the raw-scale income should be used. For this reason, we will illustrate analyses of both scales.

Conditional-Mean Versus Conditional-Median Regression

We model the conditional median to represent the relationship between the central location of income and the covariates. By contrast, conditional-mean models, such as the OLS, estimate the conditional mean, which tends to capture the upper tail of the (right-skewed) income distribution. The median regression was estimated using the Stata "qreg" command. This command was also used on 500 bootstrap samples of the original sample so

TABLE 7.3
Goodness of Fit: Raw-Scale Versus Log-Scale Income QRM

	1991			*2001*		
	Income	*Log Income*	*Difference*	*Income*	*Log Income*	*Difference*
Quantile	*(1)*	*(2)*	*(2) – (1)*	*(1)*	*(2)*	*(2) – (1)*
.05	.110	.218	.109	.093	.194	.101
.10	.155	.264	.109	.130	.237	.107
.15	.181	.281	.099	.154	.255	.101
.20	.198	.286	.088	.173	.265	.091
.25	.212	.290	.078	.188	.270	.083
.30	.224	.290	.067	.200	.274	.074
.35	.233	.290	.057	.209	.276	.066
.40	.242	.289	.048	.218	.277	.059
.45	.249	.288	.039	.225	.276	.051
.50	.256	.286	.029	.231	.275	.044
.55	.264	.282	.019	.236	.273	.037
.60	.270	.279	.009	.240	.270	.030
.65	.275	.275	−.001	.243	.266	.023
.70	.280	.270	−.010	.246	.262	.015
.75	.285	.264	−.021	.249	.256	.008
.80	.291	.258	−.032	.249	.250	.000
.85	.296	.250	−.047	.250	.242	−.008
.90	.298	.237	−.061	.252	.233	−.019
.95	.293	.213	−.080	.258	.222	−.036

NOTE: Presented are R-squared of QRM.

as to obtain the bootstrap standard error (see Appendix for Stata codes for this computing task). Table 7.4 lists the OLS estimates and median-regression estimates for raw-scale and log-scale income in 2001. We expect that the effects based on OLS would appear stronger than effects based on median regression because of the influence of the data in the upper-income tail on OLS coefficients.

While the coefficients of the income equation are in absolute terms, the log-income coefficients are in relative terms. With a few exceptions, the

TABLE 7.4
OLS and Median Regression: 2001 Raw and Log Income

	OLS		*Median*	
Variable	*Coeff.*	*SE*	*Coeff.*	*BSE*
Income				
Age	2191	(84.1)	1491	(51.4)
Age-squared	–22	(.8)	–15	(.5)
Black	–9800	(742.9)	–7515	(420.7)
Hispanic	–9221	(859.3)	–7620	(551.3)
Asian	–764	(1369.3)	–3080	(1347.9)
Some college	–24996	(643.7)	–18551	(612.5)
High school grad.	–32281	(647.4)	–24939	(585.6)
No high school	–38817	(830.0)	–30355	(616.4)
Married w/o children	–11227	(698.5)	–11505	(559.6)
Female head	–28697	(851.1)	–25887	(580.2)
Single person	–37780	(684.3)	–32012	(504.8)
Other household type	–14256	(837.3)	–13588	(672.8)
Rural residence	–10391	(560.7)	–6693	(344.1)
Constant	50431	(235.2)	43627	(185.5)
Log Income				
Age	0.0500	(.0016)	0.0515	(.0016)
Age-squared	–0.0005	(.00002)	–0.0005	(.00001)
Black	–0.2740	(.0140)	–0.2497	(.0145)
Hispanic	–0.1665	(.0162)	–0.1840	(.0185)
Asian	–0.1371	(.0258)	–0.0841	(.0340)
Some college	–0.3744	(.0121)	–0.3407	(.0122)
High school grad.	–0.5593	(.0122)	–0.5244	(.0123)
No high school	–0.8283	(.0156)	–0.8011	(.0177)
Married w/o children	–0.1859	(.0132)	–0.1452	(.0124)
Female head	–0.6579	(.0160)	–0.6214	(.0167)
Single person	–0.9392	(.0129)	–0.8462	(.0136)
Other household type	–0.2631	(.0158)	–0.2307	(.0166)
Rural residence	–0.1980	(.0106)	–0.1944	(.0100)
Constant	10.4807	(.0044)	10.5441	(.0045)

NOTE: BSE is bootstrap standard error based on 500 replicates.

07-Hao.qxd 3/13/2007 4:30 PM Page 100

OLS coefficients for log income are larger in magnitude than for median regression. For example, compared with being white, being black decreases the conditional-median income by $100(e^{-.274} - 1) = -24\%$ according to the OLS results, but by $100(e^{-.2497} - 1) = -22\%$ according to the median-regression results. In other words, mean income for blacks is 24% lower than it is for whites, and blacks' median income is 22% lower than whites', all else being equal. We note that while we can determine the effect of being black in absolute terms on the conditional median because of the monotonic equivariance property of the QRM, we cannot do so with the conditional-mean log-scale estimates because the LRM does not have the monotonic equivariance property. We will later return to attaining effects in absolute terms from log-income-equation estimates.

Graphical View of QRM Estimates From Income and Log-Income Equations

An important departure of the QRM from the LRM is that there are numerous sets of quantile coefficients being estimated. We use Stata's "sqreg" command for fitting the QRM with 19 equally spaced quantiles (.05th, . . . , .95th) simultaneously. The sqreg command uses the bootstrap method to estimate the standard errors of these coefficients. We specified 500 replicates to ensure a large enough number of bootstrap samples for stable estimates of the standard errors and 95% confidence intervals. The sqreg command does not save estimates from each bootstrap but only presents a summary of the results. We perform this bootstrapping for raw-scale income and log-transformed income. Results from the sqreg are used to make graphical presentations of coefficients.

Using such a large number of estimates results in a trade-off between complexity and parsimony. On the one hand, the large numbers of parameter estimates are capable of capturing complex and subtle changes in the distribution shape, which is exactly the advantage of using the QRM. On the other hand, this complexity is not without costs, as we may be confronted with an unwieldy collection of coefficient estimates to interpret. Thus, a graphical view of QRM estimates, previously optional, becomes a necessary step in interpreting QRM results.

We are particularly interested in how the effect of a covariate varies with the quantiles of interest. The graphical view in which we plot how the estimated QRM coefficients vary with p is valuable for highlighting trends in these coefficients. For raw-scale coefficients, a horizontal line indicates that the coefficient does not vary with p, so that the effect of a constant change in the covariate on the quantile of the response is the same for all quantiles.

In other words, with all the other covariates fixed, the covariate change produces a pure location shift: a positive shift if the line is above the horizontal zero line and a negative shift if the line is below the zero line. On the other hand, a straight nonhorizontal line indicates both location and scale shifts. In this case, the location shift is determined by the quantile coefficient at the median: A positive median coefficient indicates a rightward location shift and a negative median coefficient indicates a leftward location shift. An upward-sloping straight line indicates a positive scale shift (the scale becomes wider). By contrast, a downward-sloping straight line indicates a negative scale shift (the scale becomes narrower). Any nonlinear appearance in the curve implies the presence of a more complex shape shift, for example, in the form of a skewness shift. These graphs, however, provide neither exact quantities of shape shifts nor their statistical significance. We will examine their significance later using shape-shift quantities.

To illustrate how to identify the location and shape shifts using a graphical view, we examine closely the age effect on raw-scale income in Figure 7.2. As the coefficients and the confidence envelope are above 0 (the thick horizontal line), the age effects on various quantiles of raw-scale income are all positive and significant. The age coefficients form an upward-sloping, generally straight line, indicating that an increase in age shifts the location of the income distribution rightward and expands the scale of the income distribution.

The plots in Figure 7.3 show results for raw-scale income. Coefficient point estimates and 95% confidence intervals based on bootstrap standard errors are plotted against $p \in (0,1)$. The shaded area indicates that the effect of a covariate is significant for particular quantiles if the area does not cross zero. For example, the Asian effect is insignificant beyond $p > .45$ because the confidence envelope crosses 0 beyond that point. Chapter 4 summarizes some basic patterns that provide hints as to location shifts and scale shifts for raw- and log-scale coefficients. Below we discuss patterns emerging from our example.

The graph for the constant coefficient is a predicted quantile function for income for the *typical* household (i.e., the income of a fictional household based on the mean of all covariates) and serves as the baseline. This quantile function indicates that for the typical household, income has a right-skewed distribution. This skewness is less pronounced than the skewness observed for the income data without taking into account the effects of the covariates. Among the 13 covariates, only "age" has positive effects. The middle 70% of the population is estimated to have a proportional increase in income with age. The bottom-tail rates of the age effect are disproportionately lower, whereas the upper-tail rates are disproportionately higher. However, this non-proportionality is not sufficient to allow conclusions about skewness, because

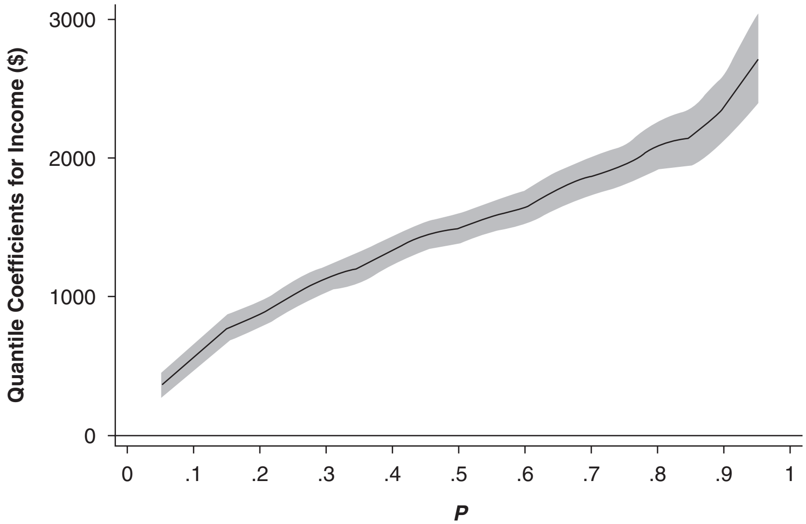

Figure 7.2 Age Effect: Raw-Scale QRM Coefficient and Bootstrap Confidence Envelope, 2001

the baseline skewness (represented by the constant term) must be taken into account. All other covariates have negative effects. As mentioned earlier, the Asian effect is significant for the lower tail of the conditional distribution. This segment of the curves is quite flat, suggesting a pure location shift for the lower half. A few covariates have close-to-flat curves; for example, compared with whites, Hispanics' income is lower by a similar amount at almost all quantiles, making the curve flat. However, most covariates appear to produce not only location shifts but also substantial shape shifts.

The graphs for log coefficients are presented in Figure 7.4. We note that log transformation contracts the right-skewed distribution to give approximate normality. Thus, the graph of the constant coefficients resembles the quantile function of a normal distribution. As discussed in Chapter 4, the log coefficient approximates proportional change in relative terms; straight flat lines indicate location shifts and scale shifts without changing the skewness. Any departure from the straight flat line becomes difficult to interpret as it tends to indicate combinations of location, scale, and skewness shifts. In addition, because on the log scale a tiny amount of log income above or below a straight flat line at the upper quantiles translates to a large amount of income, we should be cautious in claiming a close-to-flat curve. For example, the curves for the three lowest categories of education appear quite flat, but we do not claim them as close-to-flat because

07-Hao.qxd 3/13/2007 4:30 PM Page 103

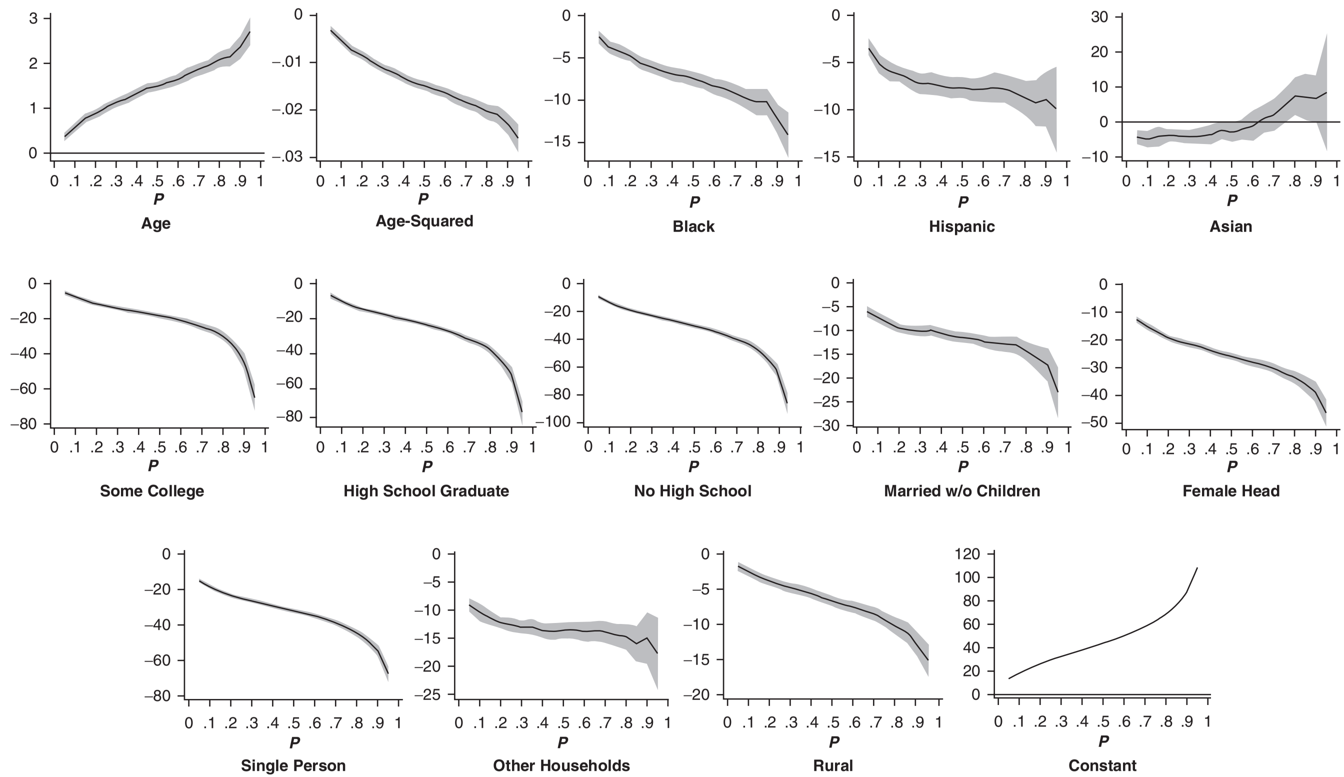

Figure 7.3 Raw-Scale QRM Coefficients and Bootstrap Confidence Envelopes: 2001

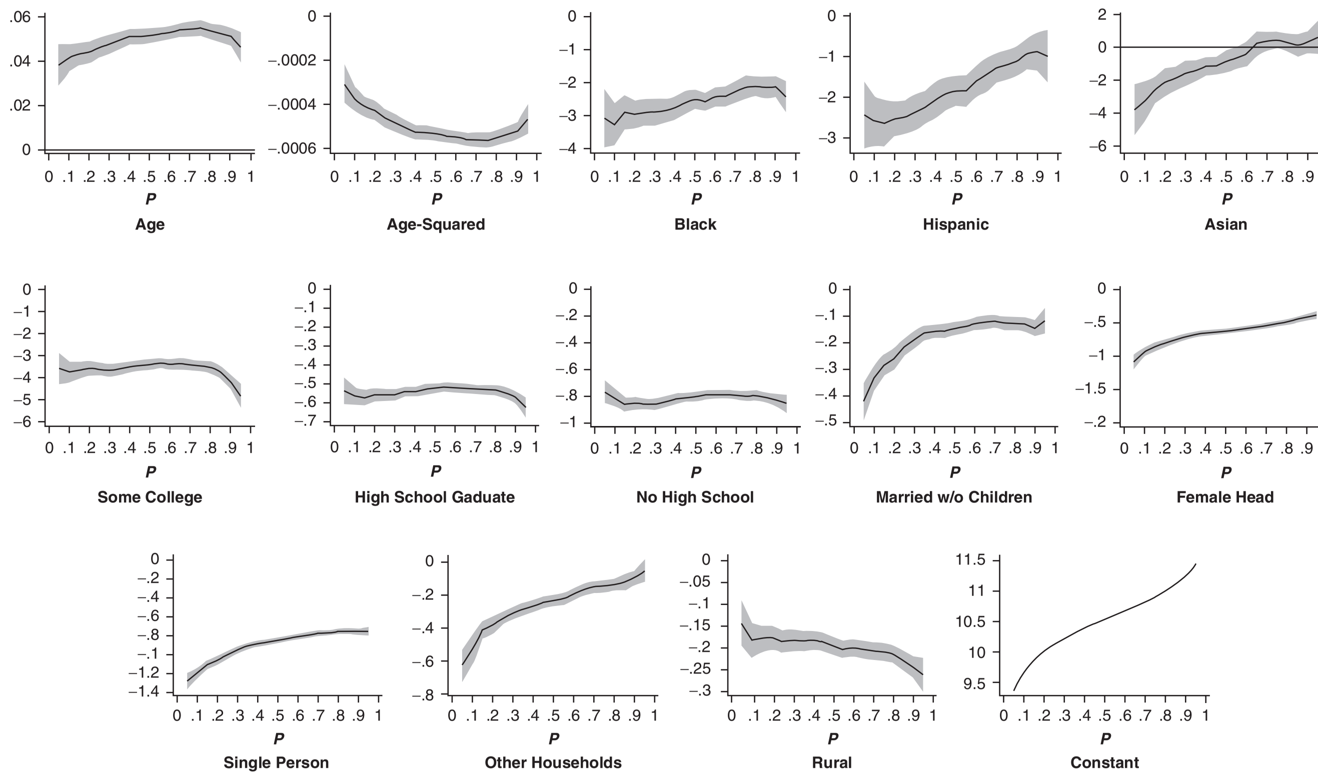

Figure 7.4 Log-Scale QRM Coefficients and Bootstrap Confidence Envelopes: 2001

their upper tail above the .8th quantile drops discernibly. In short, graphs of log coefficients are less telling and require greater caution in interpretation than graphs of raw-scale coefficients.

Quantile Regressions at Noncentral Positions: Effects in Absolute Terms

Graphical views offer an overview of the covariates' impact on the shape of the conditional-income distribution. We now complement the graphical view with a closer look at some of the off-central positions. We choose two extremes that fall outside of the graphs we just examined: the .025th and .975th quantiles. In order to obtain coefficient standard errors for these additional .025th- and .975th-quantile regressions of raw-scale income, we can either use "sqreg" with 500 replicates or manually perform the bootstrap method for 500 replicates, saving all 500 sets of resulting coefficient estimates. The conditional shape-shift quantities require programming based on each of the bootstrap replicates of these two quantile estimates, so we present the manual bootstrap results here. With the 500 sets of coefficient estimates, we use the median as the point estimate and the middle 95% as the confidence interval. If the confidence interval does not cross 0, the coefficient is significant at the $p = .05$ level. These results are almost identical to the sqreq outputs.

Estimates for the log-income equations are not in absolute terms. Because effects in absolute terms are essential to understanding the impact of a covariate on the shape of the distribution, we need to find the effect in absolute terms, evaluated at the typical setting (the mean of all covariates). As for the raw income, we save 500 sets of log-scale coefficients from bootstrap samples. For each covariate in the estimation based on a bootstrap sample, we

- Obtain the log-conditional quantile of one unit increase from the mean of the covariate by adding the coefficient to the constant term.
- Take the exponential of this log-conditional quantile and the exponential of the constant term to yield two raw-scale conditional quantiles.
- Take the difference between these two raw-scale-conditional quantiles, which becomes the effect of the covariate in absolute terms, evaluated at the typical setting, the TSE.

Table 7.5 shows the effects in absolute terms from income and log-income equations at the .025th and .975th quantiles. The top panel is from the income equation. The constant term represents the estimated value of

TABLE 7.5
Effects in Absolute Terms on Tail Quantiles:
2001 Raw and Log Income

	0.025th Quantile	*0.975th Quantile*
Variable	*Coeff.*	*Coeff.*
Income Model		
Age	248**	3103**
Age-squared	–2**	–29**
Black	–1991**	–17380**
Hispanic	–2495**	–7418
Asian	–4221**	16235
Some college	–2607**	–105858**
High school grad.	–4332**	–119924**
No high school	–6211**	–129464**
Married w/o children	–4761**	–18878**
Female head	–10193**	–50465**
Single person	–12257**	–78570**
Other household type	–7734**	–16876**
Rural residence	–943**	–18654**
Constant	10156**	137561**
Log-Income Model		
Age	396**	5409**
Age-squared	–3**	–53**
Black	–2341**	–28867**
Hispanic	–1835**	–8032
Asian	–3259**	8636
Some college	–1916**	–49898**
High school grad.	–2932**	–57557**
No high school	–4095**	–70006**
Married w/o children	–3149**	–12471**
Female head	–5875**	–33219**
Single person	–6409**	–63176**
Other household type	–4382**	–5282**
Rural residence	–938**	–26742**
Constant	8457**	115804**

NOTE: Significance (** $p < .05$) is determined by 95% confidence interval based on 500 bootstrap replicates.

the .025th and .975th quantiles, respectively, when all covariates are at their mean values: about $10,000 at the bottom and about $137,000 at the top. The most striking pattern is the huge difference in the effect of a covariate on the two ends. For example, being black reduces income by $1,991 at the .025th quantile and by $17,380 at the .975th quantile. In addition, Hispanics and Asians have significantly lower income than whites at the .025th quantile but not at the .975th percentile.

The lower panel shows the TSEs based on the log-income equation. The constant term represents the .025th and .975th conditional quantiles at the typical setting. The TSEs are quite similar to those estimated from the income equation. They are not exactly the same, because the log-income model fits better than the income model and because the log-income equation estimates are evaluated at the typical setting.

Assessing a Covariate's Effect on Location and Shape Shifts

QRM estimates can be used to calculate precisely how a covariate shifts the location and shape of the conditional distribution. To do such an assessment, we compare two groups: a reference group and a comparison group. In the case of a continuous covariate, the reference group is defined by equating the covariate to some value, and the comparison group is defined by increasing the covariate by one unit, holding other covariates constant. For a dichotomous covariate, we change its value from 0 to 1, holding other covariates constant. All comparisons are made in absolute terms to reveal the raw-scale distribution. Thus, if log-income regression is used to fit the data, the coefficient in absolute terms for a covariate is obtained first (as in the previous section). Location shifts are captured by the coefficients at the median. Shape (scale and skewness) shifts are based on a combination of a number of coefficients. Their significance levels are determined using the bootstrap method.

Table 7.6 shows the results from the income model for 1991 and 2001, with location shifts in the top panel, scale shifts in the middle, and skewness shifts in the bottom. In 1991, all covariates except Asian significantly shift the comparison group's location from the reference group. Some of these effects change noticeably from 1991 to 2001. The Asian location shift, insignificant in 1991, becomes significantly negative in 2001, suggesting the absolute advantage of whites over minorities. Other racial and ethnic groups' location shifts, however, appear to become weaker. Age's location shift is less important in 2001 than in 1991. The same is true for having less education. However, the negative location shifts for household types that are not "married with children" become stronger, as does rural residence.

Location shifts capture between-group differences. As previously discussed for Table 7.4, median-regression coefficients are weaker than OLS coefficients. For the highly right-skewed income distribution, median-regression coefficients capture the central location shift, whereas the OLS coefficients are influenced more heavily by the right tail. Using location shifts (median regression), our findings regarding education groups suggest that the

TABLE 7.6
Location and Shape Shifts of Conditional Quantiles: From Raw-Scale QRM

Shift	*1991*	*2001*
Location Shift		
Age	1801**	1501**
Age-squared	–169**	–149**
Black	–7878**	–7473**
Hispanic	–8692**	–7616**
Asian	–1231	–2850**
Some college	–19173**	–18588**
High school grad.	–25452**	–24926**
No high school	–32595**	–30345**
Married w/o children	–9562**	–11501**
Female head	–22366**	–25862**
Single person	–27866**	–32039**
Other household type	–11716**	–13659**
Rural residence	–5284**	–6698**
Scale Shift (middle 95% of population)		
Age	3393**	2852**
Age-squared	–305**	–272**
Black	–14617**	–15378**
Hispanic	–3027	–4893
Asian	11425	20842
Some college	–34212**	–103245**
High school grad.	–49002**	–115600**
No high school	–63477**	–123369**
Married w/o children	3708	–14001**
Female head	–9177	–40290**
Single person	–32482**	–66374**
Other household type	–8220	–8819**
Rural residence	–9817**	–17693**
Skewness Shift (middle 95% of population)		
Age	–0.0200**	–0.0195**
Age-squared	0.0003**	0.0002**
Black	0.0242	0.0713
Hispanic	0.2374**	0.1833**
Asian	0.0395	0.1571
Some college	0.3524**	–0.8572
High school grad.	0.5245**	–1.0263
No high school	0.7447**	–1.1890
Married w/o children	0.4344**	0.1514
Female head	0.8493**	0.3781**
Single person	0.5229**	0.2184
Other household type	0.1748	0.1714
Rural residence	0.0446	0.0541

NOTE: Significance (** $p < .05$) is determined by 95% confidence interval based on 500 bootstrap replicates.

07-Hao.qxd 3/13/2007 4:30 PM Page 109

education effect in terms of location shifts is not as strong as indicated in the literature. The change in location shift, or between-group difference, is only one part of the story about how inequality changed over the decade; the other is the shape change, or relative within-group differences. The advantage of the QRM is that they disentangle the between- and within-group differences, advancing our understanding of changes in inequality.

Scale shifts are one type of shape changes. Among the three racial and ethnic minority groups, only blacks have a shorter conditional-income distribution scale than do whites. The scale for the income of the middle 95% of blacks is much narrower than it is for whites, suggesting greater homogeneity among blacks than whites and the significance of race in determining income. This scale shift becomes stronger in 2001. The same is seen in the three less-educated groups. The education scale shift offers a consistent and refined finding about the increasing importance of education in determining income: It is the shape shift, rather than the location shift, that indicates the rising importance of education.

Skewness shifts are another type of shape change. An increase in the skewness of a conditional quantile indicates uneven within-group differentiation that favors the top-tail members. The 1991 results show that many disadvantaged groups experience this uneven within-group differentiation, including Hispanics, the three less-educated groups, and disadvantaged household types (single-mother, single-person, and "other" households). Some of these shifts disappear in 2001, particularly those of the education groups. This finding further reveals the mechanism by which society rewards college graduates and limits upward mobility for the most able among the less educated.

Results on the raw scale from the log-income model are shown in Table 7.7. These results capture the same trends for life cycle, racial and ethnic groups, education groups, household types, and rural residence. The location shifts and scale shifts in each year, as well as their decade trends, are similar whether income or log income is fitted. Discrepancies are found for skewness shifts. In particular, skewness is reduced significantly for less-educated groups in 2001; this finding is significant based on the log-income model but insignificant based on the income model. It is not surprising that such discrepancies should appear when examining the two model fits (income and log income). They represent fundamentally distinct models, with one of them (log income) providing a better fit. On the other hand, if qualitative conclusions differ, it may indicate that the results are sensitive. We determine whether this is the case by looking at the overall evaluation of a covariate's role in inequality.

We develop an overall evaluation of a covariate's impact on inequality, which examines the alignment of the signs of location and shape shifts.

TABLE 7.7
Location and Shape Shifts of
Conditional Quantiles: From Log-Scale QRM

Shift	*1991*	*2001*
Location Shift		
Age	2456**	1994**
Age-squared	–24**	–20**
Black	–9759**	–8386**
Hispanic	–7645**	–6300**
Asian	–1419	–3146**
Some college	–10635**	–11012**
High school grad.	–14476**	–15485**
No high school	–20891**	–20892**
Married w/o children	–3879**	–5103**
Female head	–15815**	–17506**
Single person	–19599**	–21658**
Other household type	–6509**	–7734**
Rural residence	–4931**	–6725**
Scale Shift (middle 95% of population)		
Age	4595**	5008**
Age-squared	–41**	–50**
Black	–17244**	–26509**
Hispanic	–2503	–6017
Asian	4290	12705
Some college	–22809**	–47992**
High school grad.	–32675**	–54434**
No high school	–44457**	–65956**
Married w/o children	77	–9264**
Female head	–10269	–27272**
Single person	–32576**	–56791**
Other household type	–7535	–906
Rural residence	–12218**	–25760**
Skewness Shift (middle 95% of population)		
Age	–0.0417**	–0.0100
Age-squared	0.0005**	0.0002
Black	0.1127	–0.0682
Hispanic	0.2745**	0.1565**
Asian	–0.0383	0.1469
Some college	0.0655	–0.2775**
High school grad.	0.0934	–0.2027**
No high school	0.2742**	–0.1456**
Married w/o children	0.0890	–0.0272
Female head	0.5404**	0.3193**
Single person	0.2805**	–0.0331
Other household type	0.0164	0.1640**
Rural residence	0.0012	–0.0740

NOTE: Significance (** $p < .05$) is determined by 95% confidence interval based on 500 bootstrap replicates.

Only significant shifts are counted. For a covariate, in-sync signs in the three shifts indicate that the covariate exacerbates inequality; the larger the number of significant signs, the stronger the exacerbating effect becomes. Out-of-sync signs indicate that the covariate may increase between-group inequality while decreasing within-group inequality, or vice versa. The left panel of Table 7.8 for the income model shows that none of the covariates have in-sync effects on inequality in 1991, but many do in 2001. These in-sync covariates are education groups, household types (except female heads), and rural residence. The right panel shows the corresponding results from the log-income model. We see little contradiction in the overall evaluation. For example, for education groups, the pattern changes from out of sync in 1991 to in sync in 2001 in both models. Thus, American society in 2001 was more unequal and its social stratification more salient by education, marriage, presence of children, and rural residence than was the case a decade earlier.

In this example, we use the middle 95% population to calculate the shape-shift quantities. Researchers can design their own shape-shift definitions according to their research questions. It is possible to design corresponding shape shifts for the middle 99%, 98%, 90%, 80%, or 50% of the population. We leave this to our readers to undertake.

TABLE 7.8
Overall Evaluation of Covariates' Role in Inequality:
Synchronicity Patterns in Coefficients

	Income Equation		*Log-Income Equation*	
Variable	*1991*	*2001*	*1991*	*2001*
Age	+ + −	+ + −	+ + −	+ + 0
Age-squared	− − +	− − +	− − +	− − 0
Black	− − 0	− − 0	− − 0	− − 0
Hispanic	− 0 +	− 0 +	− 0 +	− 0 +
Asian	0 0 0	− 0 0	0 0 0	− 0 0
Some college	− − +	− − 0	− − 0	− − −
High school grad.	− − +	− − 0	− − 0	− − −
No high school	− − +	− − 0	− − +	− − −
Married w/o children	− 0 +	− − 0	− 0 0	− − 0
Female head	− 0 +	− − +	− 0 +	− − +
Single person	− − +	− − 0	− − +	− − 0
Other household type	− 0 0	− − 0	− 0 0	− 0 +
Rural residence	− − 0	− − 0	− − 0	− − 0

07-Hao.qxd 3/13/2007 4:30 PM Page 112

Summary

What are the sources of the persistent and widening income inequality in the recent decade? To address this research question, we apply the techniques developed in this book. We start with a descriptive analysis using the notion of quantiles introduced in Chapter 2. For income data, we discuss the issues of right-skewed distributions and top-coding and explain why and how the QRM can accommodate these features. The analyses follow the steps discussed in Chapters 3 through 6: defining and fitting models, assessing the goodness of fit, estimating the inference of parameters, making graphs of coefficients and their confidence intervals, and calculating location and shape shifts and their inferences. We describe the income and log-income models, paying special attention to the reconstructed raw-scale coefficients. Along with our description of the steps, we demonstrate the utility of the QRM techniques in addressing the research question through interpretations of the results. It is our hope that this systematic summarization of the application procedures will provide clear guidance for empirical research.