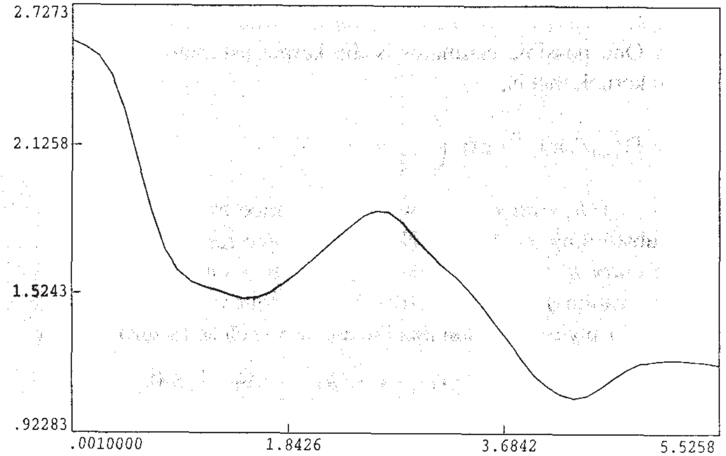

Figure 3.1. Average duration of strikes conditional on activity: Gaussian kernel estimator.

the addition of two trigonometric terms at a time, corresponding to $L = 1, 2, 3$, were 3.79, 2.92, and .4486 respectively, pointing to $L = 2$ as being a reasonable choice. Figure 3.2 therefore plots the estimates made of the conditional mean at x_1 from (3.111) by replacing x_{1i} by x_1, x_{2i} by zero, and β_j, γ_j, α by the ordinary least squares (OLS) estimates. It is striking that there is good agreement between the kernel and Fourier estimators, and both point to a peak at both low and high levels of contraction. However, the 95% confidence intervals associated with the Fourier estimator, presented in Figure 3.3, cast doubt on this feature, as the intervals are very wide for a large contraction. Partly, this could reflect the "boundary value" problem alluded to earlier, but it also reflects the paucity of observations on large contractions.[16] One fact that should be emphasized is that, despite the detection of a relation between strike duration and activity, only a very small fraction of the variation in strike activity is actually explained; the R^2 from the regression (3.111) when $L = 3$ is only .058. Hence, a good model for the duration series really requires a much wider set of determining variables.

3.14.2 *Earnings–Age Profiles*

In the example here we consider the specification of relationships between mean earnings and age (a proxy for experience) and the variability in earnings

[16] Note that the model fitted by Horowitz and Neumann implies that duration declines exponentially with activity; this is a common assumption in the duration literature.

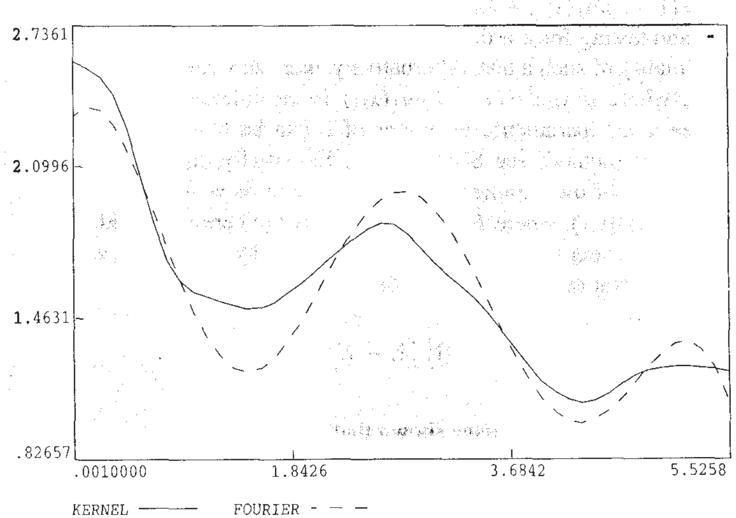

Figure 3.2. Average duration of strikes conditional on activity: kernel and flexible Fourier estimates.

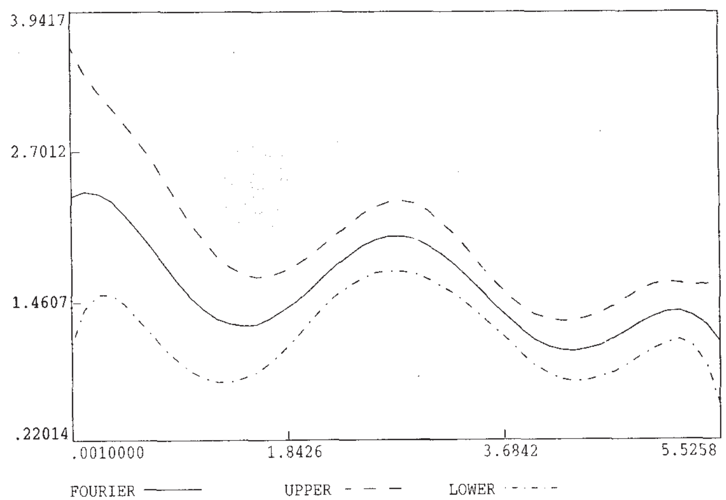

Figure 3.3. Average duration of strikes conditional on activity: flexible Fourier form estimates with upper and lower 95% confidence.

with age. There is an extensive parametric literature in labor economics, where earnings are generally modeled as

$$y_i = \alpha + \beta z_i + \gamma_o x_i + \gamma_1 x_i^2 + u_i,$$

where y_i is logarithm of earnings, z_i is education, and x_i is age (see Mincer, 1974 and Heckman and Polachek, 1974). Mincer finds that earnings increase with age through much of the working life but the rate of increase diminishes with age. He concludes that the concavity of the function is consistent with investment behavior implied by the optimal distribution of human capital investment over the life cycle. The quadratic parametric relationship between earnings and age has been challenged recently by Murphy and Welsch (1990). They have shown that the quadratic specification does not fit the data well, resulting in severely biased estimates of the earnings profile. They conclude that the quartic specification has a sufficiently small bias and could be used as the standard specification. Regarding the profile of the variability of earnings with age, using grouped data Mincer (1974) finds that it changes with the level of schooling.

Our main interest here is to explore the specifications of the mean and variance of earnings with age by nonparametric methods. For this purpose we considered two data sets: Canadian data (1971 Canadian Census Public Use Tapes) on 205 individuals and 1988 Chinese data on 2,449 urban males and 2,342 urban females. The individuals in both data sets were educated to grade 13, and thus schooling was assumed to be constant for simplicity (i.e., z_i does not appear in the relation). For the details on data and various other findings, see Ullah (1985), Singh et al. (1987), Chu and Marron (1991), and Basu and Ullah (1992). The conditional mean and variance of earnings (y) were calculated by using the formulae in Sections 3.2.2 and 3.2.6 with $q = 1$, a normal kernel, and $h = sn^{-1/5}$; $s^2 = \sum(x_i - \bar{x})^2/n$. The use of cross-validated h did not make any difference to the results.

We observe from Figure 3.4 that the quadratic parametric specification provides a smooth concave least squares estimate of the earnings profile. However, the nonparametric specification indicates a "dip" around the mean age of 40. This dip is also found if we use higher order or convolution kernels (Chu and Marron, 1991), and even if we consider the nonparametric regression of earnings on age and schooling. A possible explanation for this dip is the generation effect, because the cross section data represent the earnings of people at a point of time who essentially belong to different generations. Thus the plot of earnings represents the overlap of the earnings trajectories of different generations. Only if the sociopolitical environment of the economy has remained stable intergenerationally can we assume these trajectories to be the same. But this is not the case; one obvious counterexample being the Second World War. Therefore, the dip in the nonparametric regression might be attributed to the generation between 1935 and 1945. This is an important fact about the data

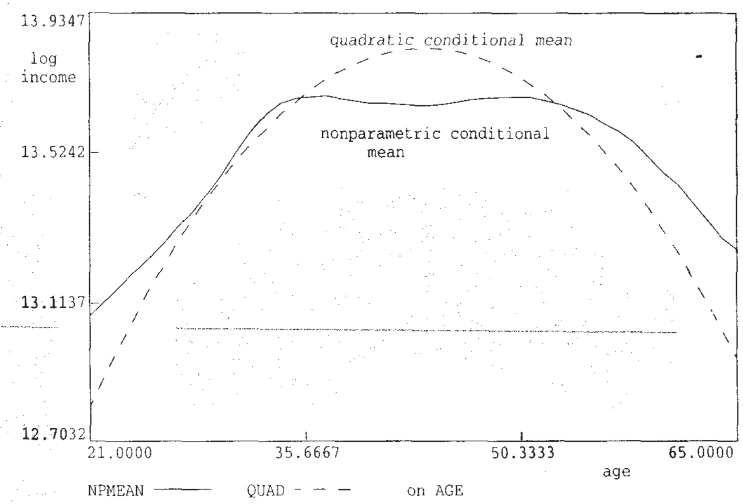

Figure 3.4. Nonparametric and quadratic fits, income/age relation (Canadian data).

observed from the nonparametric specification, but it is missed by a parametric specification. The dip also indicates the lack of global concavity and that the quartic relationship may be more appropriate than the quadratic specification; also see Murphy and Welch's (1990) parametric findings on this point. The question of whether the dip is significant or not can be answered by using the parametric and nonparametric specification testing procedures described in Section 3.13.

Looking at Figure 3.5 we find that, for Chinese data, male earnings increase at an increasing rate in the early part of their career. The peak is not found at the mean age, 37, but much later after retirement (65–66), and then it declines; the retirement age is 55. This is in contrast to the Canadian case where the decline begins around the mean age. Overall, the earnings profile appears to be quadratic. For females, however, the relationship is peculiar since between the ages 44 and 50 there is a decline in earnings and then an increase till around 60 after which it falls again. One explanation for the dip around ages 44–50 is again the cohort effect. This time it might be attributed to the cultural revolution in the 1950s. However, if true, the reason why the cultural revolution affected females ages 44–50, but not males, is unclear and needs further institutional details on the role of women in China. Finally, the observation that, for the same level of schooling, male earnings are uniformly higher than those for the female earnings indicates the possiblity of "wage discrimination." The test for the significant difference between male and female earning functions can be done by using the test due to Lavergne (1998).

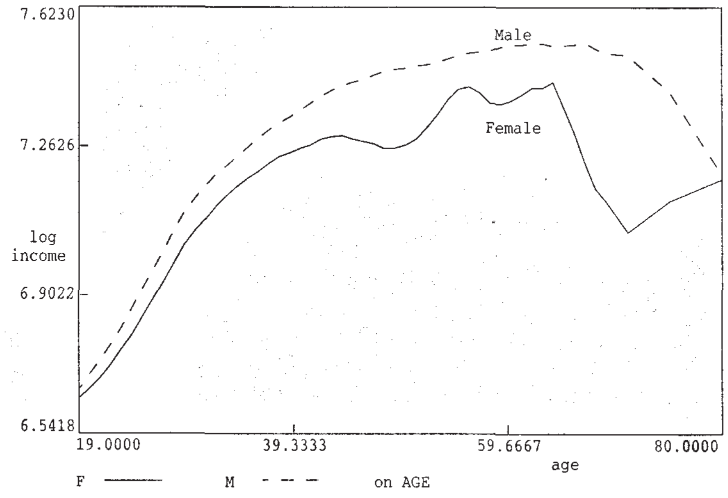

Figure 3.5. Nonparametric conditional mean, age/income relation (Chinese data, male and female).

Looking at Figure 3.6, we see that the Canadian data clearly exhibit variability in earnings that is a convex function of age; in particular, it appears to be inversely related to the mean earnings, that is, $\hat{V}(y|x) \propto (\hat{E}(y|x))^{-1}$ (this is also true of the Chinese data). The result is consistent with Mincer's (1974, p. 101) finding based on U.S. data. However, we should note that, although Mincer calculates a variance based on an arbitrary grouping of age, the nonparametric analysis does not need such a grouping. A useful implication of the above finding is that, in practice, one can circumvent the need for parametric specifications of $\hat{V}(y \mid x)$, the conditional heteroskedasticity – see Chapter 5 for more on this point.

Another observation from Figure 3.6 is that, for the early ages, the income is high and the variability (uncertainty) is low, but for the later ages the income is low and the variability is high. This implies that the income inequality may be influenced by life cycle effects. To analyze this hypothesis one can estimate several inequality measures conditional on age. Usually, this is done by calculating interage group inequality, which is then interpreted as a measure of inequality caused by life-cycle effects. The main criticism of this approach is that the results may be sensitive to the choice of the arbitrarily chosen age groups and their widths. Alternatively, one can calculate various inequality measures conditional on age. Some well-known inequality measures that satisfy the principle of transfers are Atkinson's (A) inequality measure, Theil's entropy (TE)

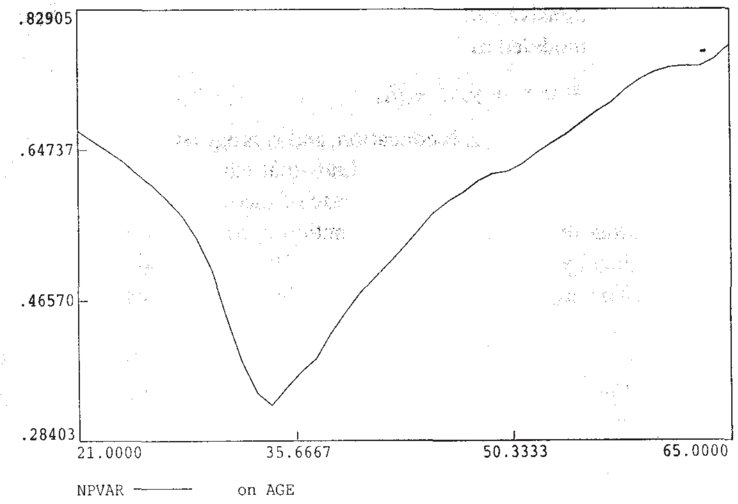

Figure 3.6. Nonparametric conditional variance for income/age relation (Canadian data).

measure, and the coefficient of variation (C.V.); see Kakwani (1980) for an excellent description. The conditional versions of these measures can be written as

$$A(x) = 1 - [E(y^{1-\epsilon}|x)]^{1/(1-\epsilon)} (E(y|x))^{-1},$$
$$TE(x) = (E(y|x))^{-1} E[y \log y|x] - \log E(y|x),$$
$$C.V.(x) = \sqrt{V(y|x)}/E(y|x),$$

where the parameter $\epsilon \geq 0$ controls the degree of inequality aversion or perception. When $\epsilon = 0$ there is no perceived inequality and when $\epsilon = \infty$ there is concern for the poor only. In practice ϵ is usually considered to be between 0 and 2.

Though we do not perform such a calculation here, the inequality measures $A(x)$ and $TE(x)$ can easily be calculated by using the results of Section 3.2.6. For example, the $C.V.(x)$ follows directly from the calculation of $V(y|x)$ and $E(y|x)$. In fact, since in the above study, $\hat{V}(y|x) \propto \left(\hat{E}(y|x)\right)^{-1}$, it will be the case that $C\hat{.}V.(x) \propto \left(\hat{E}(y|x)\right)^{-3/2}$.

3.14.3 *Review of Applied Work on Nonparametric Regression*

There has been an increasing use of nonparametric methods in applied econometrics in recent years and an exhaustive treatment is impossible. Consequently, we limit ourselves to an interesting selection of the papers that have made