Correlation Curves as Local Measures of Variance Explained by Regression

Kjell DOKSUM, Stephen BLYTH, Eric BRADLOW, Xiao-Li MENG, and Hongyu ZHAO*

We call (a model for) an experiment *heterocorrelatious* if the strength of the relationship between a response variable Y and a covariate X is different in different regions of the covariate space. For such experiments we introduce a *correlation curve* that measures *heterocorrelaticity* in terms of the variance explained by regression locally at each covariate value. More precisely, the squared correlation curve is obtained by first expressing the usual linear model "variance explained to total variance" formula in terms of the residual variance and the regression slope and then replacing these by the conditional residual variance depending on x and the slope of the conditional mean of Y given $X = x$. The correlation curve $\rho(x)$ satisfies the invariance properties of correlation, it reduces to the Galton–Pearson correlation ρ in linear models, it is between -1 and 1, it is 0 when X and Y are independent, and it is ± 1 when Y is a function of X. We introduce estimates of the correlation curve based on nearest-neighbor estimates of the (conditional) residual variance function and the (conditional) regression slope function, as well as on Gasser–Müller kernel estimates of these functions. We obtain consistency and asymptotic normality results and give simple asymptotic simultaneous confidence intervals for the correlation curve. Real data and simulated data examples are used to illustrate the local correlation procedures.

KEY WORDS: Heterocorrelaticity; Kernel estimates of regression; Local correlation.

1. INTRODUCTION

There is a strong link between regression and correlation in linear statistical theory and methodology. The regression coefficients measure the relationship between the covariates $X = (X_1, \ldots, X_p)$ and the conditional mean $\mu(x) = E(Y|X = x)$ of the response variable Y. More precisely, assuming that a linear model is correct, the regression coefficients $\beta_1, \ldots, \beta_p$ are *slopes* that give the rates of change of $\mu(x) = E(Y|X = x)$. That is, the rate of change $(\partial/\partial x_i)\mu(x)$ is assumed to be a constant, labelled β_i, $i = 1, \ldots, p$. In the linear model, correlation coefficients measure the *strength* of the relationship between the response variable Y and the covariates X_j, $j = 1, \ldots, p$. Moreover, the correlation coefficients measure *variance explained;* that is, how much of the variability of Y can be explained by the covariates.

In this article we consider measures of regression slopes, strength of relationships, and variance explained in the nonlinear case principally for the case of one covariate ($p = 1$). These measures will depend on the values of the covariates; that is, they will be local in nature. To illustrate, in cases where the covariate X represents "level of a symptom" and the response variable Y represents "level of a disease," the regression slope typically is 0 for small $X = x$ and then gradually increases as x increases. Moreover, in such examples the strength of the relationship and the variance explained may also increase with increasing x. We propose to quantify this notion of *heterocorrelaticity* using a notion of local correlation that we call correlation curves.

First, we consider possible shapes of joint densities $f(x, y)$ of X and Y in such symptom–disease examples. As pointed out by Fisher (1959), the contour plots (i.e., plots of (x, y) where $f(x, y)$ is constant) often resemble twisted pears (see also Gjerde, Block, and Block 1988). Examples of such twisted pear models can be generated by bivariate transformation models (see, for instance, Fig. 1).

The Figure 1 contour plot closely resembles Fisher's twisted pear plot of level of symptom x vs. level of disease y and clearly shows how the slope of the regression $E(Y|X = x)$ increases with increasing x, as well as how the strength of the relationship between X and Y increases with increasing $X = x$.

In other examples, the slope of the regression decreases with increasing x, and the strength of the relationship between X and Y similarly decreases with increasing $X = x$. Härdle (1990) gave an example with X = net income and Y = expenditure for food. Figure 2 shows how the strength of the relationship between net income and food expenditure diminishes with increasing net income.

A theoretical model that generates data of this type is given in the following example.

Example 1. We suppose that the response variable Y is related to the covariate X through the relation

$$Y = \mu(X) + \tau(X)\varepsilon,$$

where X and ε are independent and have respective distributions $N(\mu_1, \sigma_1^2)$ and $N(0, \sigma_\varepsilon^2)$. Now (X, Y) have the joint density $f(x, y) = f(x)f(y|x)$, where $f(x)$ is the $N(\mu_1, \sigma_1^2)$ density and $f(y|x)$ is the $N(\mu(x), \tau^2(x)\sigma_\varepsilon^2)$ density (see Fig. 3).

2. MEASURING STRENGTH OF ASSOCIATION. THE CORRELATION CURVE

2.1 Motivation for the Concept and Use of Local Correlation

The original motivation for the concept of local correlation came from Jack Block, Department of Psychology, Univer-

* Kjell Doksum is Professor, Department of Statistics, University of California, Berkeley, CA 94720. Stephen Blyth is Lecturer, Department of Mathematics, Imperial College, London, SW7 2BZ, U.K. Eric Bradlow is a graduate student, Department of Statistics, Harvard University, Cambridge, MA 02138. Xiao-Li Meng is Assistant Professor, Department of Statistics, University of Chicago, IL 60637. Hongyu Zhao is a graduate student, Department of Statistics, Berkeley, CA 94720. This research was supported by National Science Foundation Grants DMS-8801603 and DMS-91106488. The bulk of this work was carried out in the Department of Statistics, Harvard University. The authors thank Wolfgang Härdle for the expenditure data, Alex Samarov and Steve Marron for helpful comments, and the referees and the editors for their comments.

Journal of the American Statistical Association
June 1994, Vol. 89, No. 426, Theory and Methods

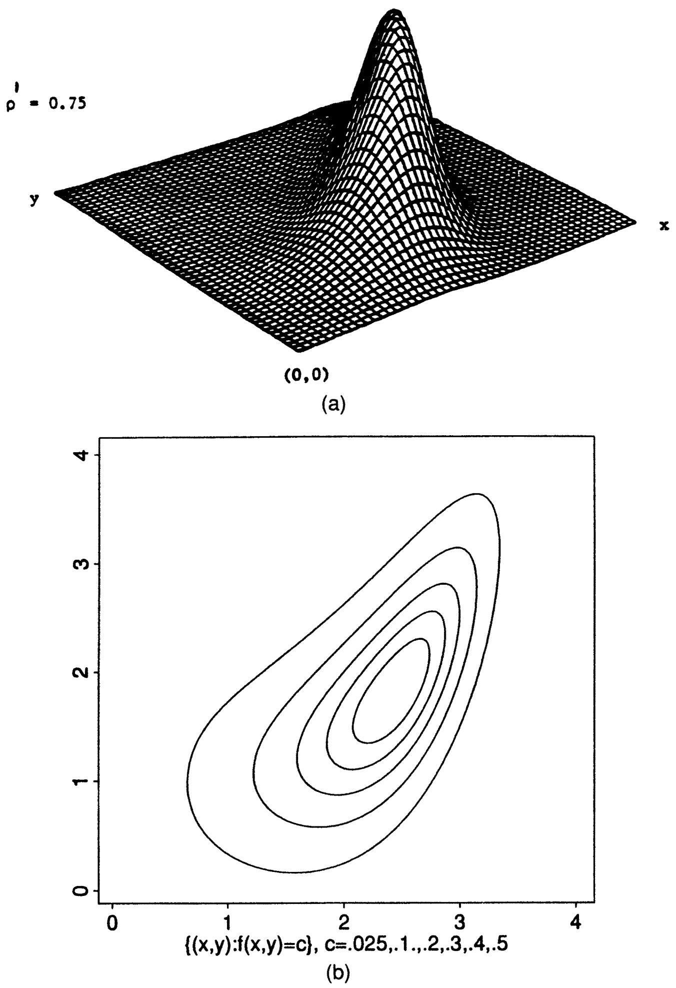


Figure 1. The Joint Density (a) and Contour Plots (b) for (X, Y) in the Transformation Model Where $X = U^{1/3}$, $Y = V$, and (U, V) is Bivariate Normal $N(\mu_1', \mu_2', \sigma_1', \sigma_2', \rho')$, With $\mu_1' = \sigma_1' = 10$, $\mu_2' = 1.55$, $\sigma_2' = .775$, and $\rho' = .75$.

sity of California, Berkeley, and Per Gjerde, Psychology Board, University of California, Santa Cruz. They asked us how to analyze association between X and Y for data where the scatterplot followed a shape similar to the contour plot of Figure 1. Our response was to produce a transformation model where the transformed X and Y, say U and V, appeared to follow (to a close approximation) a linear model. Then we proposed the correlation coefficient between U and V as a measure of the strength of the relationship between X and Y. Block and Gjerde were not pleased, however. The interesting phenomena in their study was that the strength of the relationship was *not* the same for different values of the covariate $X = x$. Our transformations to U and V had produced a constant strength for the relationship and had erased the effect they were trying to capture! Thus in the Figure 1 transformation example, the correlation between $U = X^3$ and $V = Y$ is .75. But we are interested in a measure of the strength of association that quantifies how weak it is for small $X = x$ and how strong it is for large $X = x$.

Once we quantify local strength of association in terms of local correlation, we can develop statistical techniques to analyze this correlation. For instance, for specified quantiles x_q of the X distribution, we can test whether the local correlation is significant at these quantiles. We can also test whether the correlation is stronger at certain quantiles than at other specified quantiles (see Sec. 6).

The preceding transformation model anecdote does not imply that there is no room for transformation models in the study of local correlation. In fact, for small to moderate sample size n, fitting a transformation model and transforming back to the original untransformed scales produces a parametric correlation curve and is one sensible way of studying local correlation, because, even though it may be biased, it will have a much smaller variance than nonparametric estimates. But for moderately large to large sample size n, say $n \geq 200$, the local correlation estimates based on a fitted transformation model have a large mean squared error relative to nonparametric methods due to excessive bias, unless the transformation model provides a good approximation to the true underlying model. In this article we consider the properties of nonparametric methods; however, we give one example of a correlation curve based on a transformation model in Section 6.

2.2 A Formula for Local Correlation

Note that the naive approach of arguing that if $\mu(x) = E(Y|X = x)$ is smooth, then the relationship between X and Y is locally linear, and we can use the conditional version of the usual correlation, does not provide a formula for local correlation. For instance, in the linear model

$$Y = \alpha + \beta X + \varepsilon; \qquad \varepsilon \text{ and } X \text{ independent,} \tag{1}$$

we find

$$\operatorname{corr}^2(X, Y|X = x) = \frac{\beta^2 \operatorname{var}(X|X = x)}{\operatorname{var}(Y|X = x)} = 0$$

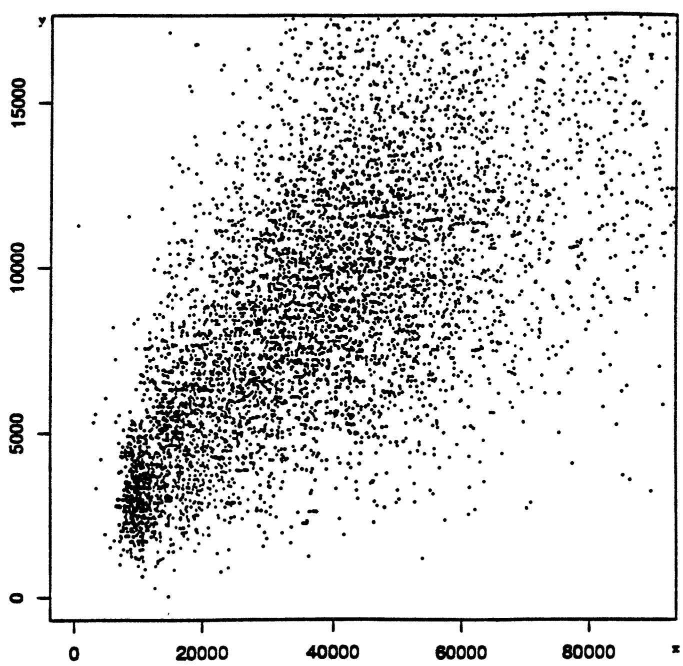


Figure 2. Food Versus Net Income. Scatterplot of Y = expenditure for food versus X = net income (both reported in tenths of pence per week), n = 7,125. The data shown are from the year 1973 of the Family Expenditure Survey, 1968–1983 (from Härdle 1990).

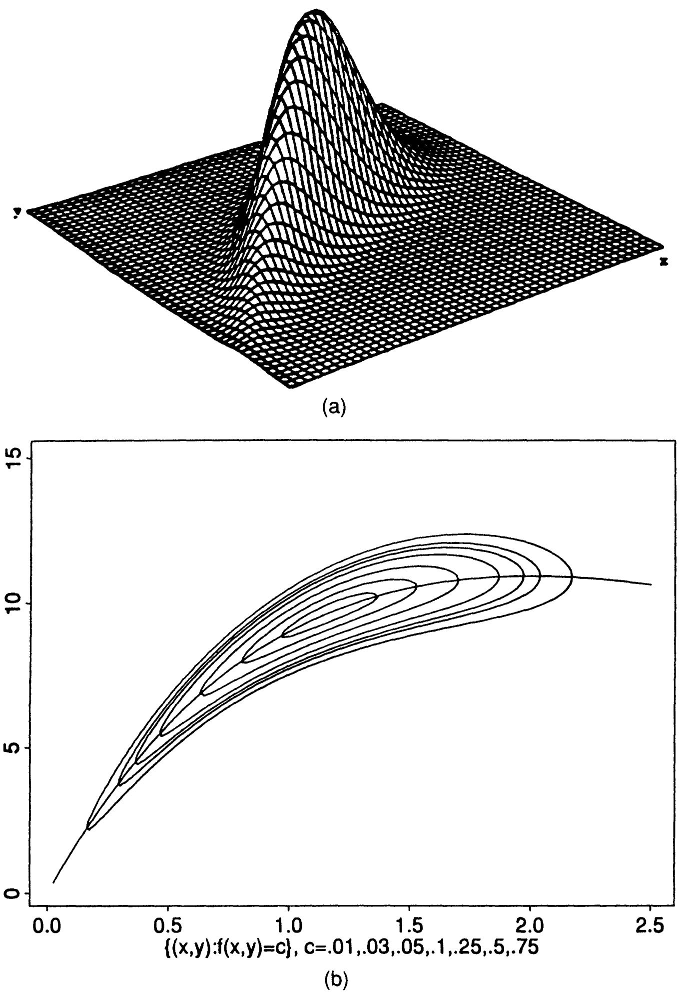


Figure 3. The Joint Density Plot (a) and Contour Plots (b) for f(x, y) = f(x)f(y | x) When f(x) is N(1.2, (1/3)²) and f(y | x) is N(μ(x), σ²(x)), With μ(x) = (x/10) exp{5 − (x/2)} and σ²(x) = τ²(x)σ²_ε = {[1 + (x/2)]/3}². Plot (b) includes a plot of the curve μ(x).

whenever $\text{var}(Y \mid X = x) > 0$.

For this reason, we must be careful when constructing our measure of association by moving from the linear case to the general case. Our approach is to express the linear case correlation in terms of the regression slope β and the residual variance $\sigma_\varepsilon^2 = \text{var}(\varepsilon)$. We then replace the linear case slope β by the general case slope $\beta(x) = (d/dx)E(Y \mid X = x)$ and replace the residual variance σ_ε^2 by the general case residual variance $\sigma^2(x) = \text{var}(Y \mid X = x)$. Thus for the linear model (1), we set $\sigma_1^2 = \text{var}(X)$ and write

$$\rho^2 = \frac{\text{variance explained by regression}}{\text{total variance}}$$

$$= \frac{\text{var}(\alpha + \beta X)}{\text{var}(\alpha + \beta X + \varepsilon)} = \frac{\sigma_1^2 \beta^2}{\sigma_1^2 \beta^2 + \sigma_\varepsilon^2}.$$

This formula shows how in the linear case, the squared correlation coefficient measures the strength of the relationship between X and Y in terms of the regression slope and the residual variance. In the general case where the regression slope and residual variance depend on x, the same formula will still measure the strength of the relationship between X and Y, but now it will measure this strength locally at the given covariate value x. Thus we define the squared correlation curve at x as

$$\rho^2(x) = (\text{corr. curve})^2 = \frac{\sigma_1^2 \beta^2(x)}{\sigma_1^2 \beta^2(x) + \sigma^2(x)}, \qquad x \in S,$$

where S, the support of the distribution of X, is defined by $S = \{x : 0 < F(x) < 1\}$ and F denotes the distribution function of X. Or, because the sign of the correlation is often important,

$$\rho(x) = \frac{\sigma_1 \beta(x)}{\{\sigma_1^2 \beta^2(x) + \sigma^2(x)\}^{1/2}}, \qquad x \in S. \tag{2}$$

Note that this definition makes sense only if F is continuous. In fact, we need to assume that $\beta(x) = (d/dx)E(Y \mid X = x)$ exists. The distribution of Y as well as $Y \mid x$ can be continuous, discrete, or a mixture.

Note that we can also write

$$\rho^2(x) = \{1 + [\sigma_1 \beta(x)/\sigma(x)]^{-2}\}^{-1}.$$

This shows that $\rho^2(x)$ is an increasing function of the standardized regression slope $\sigma_1 \beta(x)/\sigma(x)$. Moreover, $\rho(x)$ is invariant under scale and location changes in X and Y, it reduces to the Galton–Pearson correlation coefficient in the linear model, it is between -1 and 1, and it equals ± 1 whenever Y is a nonconstant function of X. When Y is a constant, $\beta(x) = \sigma(x) = 0$, and we define $\rho(x) = 0$ in this case. When X and Y are independent, $\rho(x) \equiv 0$.

Example 1.1 (continued). Consider the example (Fig. 3) where $Y = \mu(X) + \tau(X)\varepsilon$. In this case,

$$\beta(x) = \mu'(x) = (1/10)[1 - (x/2)]\exp\{5 - (x/2)\}.$$

The resulting curve $\rho(x)$ is plotted in Figure 4.

This plot shows how the strength of the relationship, as measured by $\rho(x)$, starts out very strong for x near 0 and then drops off and approaches 0 as x approaches 2.

2.3 Local Independence and the Correlation Curve

We say that Y is *locally independent* of X at $X = x_0$ if there is a $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta)$, $\mathcal{L}(Y \mid X = x) = \mathcal{L}(Y \mid X = x_0)$. Local independence at x_0 implies that $\mu(x) = \mu(x_0)$ for $x \in (x_0 - \delta, x_0 + \delta)$; thus $\rho(x_0) = 0$, provided that $\sigma^2(x_0) > 0$. But $\rho(x_0) = 0$ does not imply that Y is locally independent of X at x_0. To see this, consider the preceding example where $\rho(2) = 0$, but $\mathcal{L}(Y \mid X = x) \neq \mathcal{L}(Y \mid X$

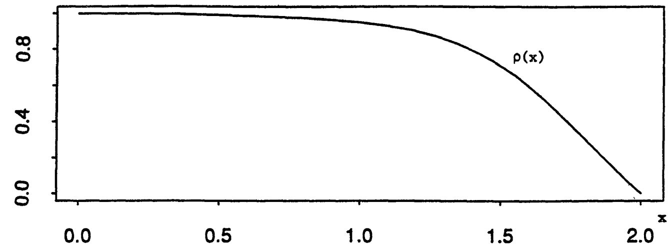


Figure 4. The Correlation Curve ρ(x) for the Model in Example 1.1 and Figure 3.

= 2) for all $x \neq 2$, so Y is *locally dependent* of X at $X = 2$. Our concept of local independence is related to the notion of local independence discussed by Schweder (1970) and Aalen (1987). They defined the time-continuous Markov chain Y to be locally independent of the chain X over some time interval if the transition intensities for changes in Y are independent of the values of X for all times in the interval.

2.4 The Case of Several Covariates

Response variables often depend on more than one covariate. Thus consider an experiment where on each of n subjects, we can measure a response Y and k covariate values $X_1, \ldots, X_k$. Again we will consider first the linear model and then ask what the natural extension to the nonlinear case would be. Our notation for the linear model is

$$Y = \alpha + \mathbf{X}\beta^T + \sigma_\varepsilon \varepsilon, \qquad E(\varepsilon) = 0, \qquad \text{var}(\varepsilon) = 1,$$

where $\beta = (\beta_1, \ldots, \beta_k)$ and ε is independent of $\mathbf{X}$. The covariance matrix

$$\sum = \text{cov}(\mathbf{X})$$

is assumed to be nonsingular. In this setting the natural measure of the strength of the relationship between $\mathbf{X}$ and Y is the coefficient of determination

$$\rho^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{\text{var}(\mathbf{X}\beta^T)}{\text{var}(Y)}.$$

Because in the linear model

$$\text{var}(Y) = \text{var}(\mathbf{X}^T\beta) + \sigma_\varepsilon^2,$$

we can rewrite ρ^2 in terms of the regression coefficient vector β and the residual variance σ_ε^2 as

$$\rho^2 = \frac{\beta \sum \beta^T}{\beta \sum \beta^T + \sigma_\varepsilon^2}. \tag{3}$$

Next, we turn to the nonlinear case where the regression slopes

$$\beta_i(\mathbf{X}) = \frac{\partial}{\partial x_i} E(Y|\mathbf{x})$$

and the residual variance $\sigma^2(\mathbf{x})$ depend on $\mathbf{x}$. In this case we define, in analogy with (3), the local coefficient of determination as

$$\rho^2(\mathbf{x}) = \frac{\text{local variability explained}}{\text{total variability}} = \frac{\beta(\mathbf{x}) \sum \beta^T(\mathbf{x})}{\beta(\mathbf{x}) \sum \beta^T(\mathbf{x}) + \sigma^2(\mathbf{x})},$$

where $\beta(\mathbf{x}) = (\beta_1(\mathbf{x}), \ldots, \beta_k(\mathbf{x}))$.

In the case of $k = 2$ covariates, $\rho^2(\mathbf{x})$ is a correlation surface that shows how the strength of the relationship between $\mathbf{X}$ and Y changes with $\mathbf{X} = \mathbf{x}$. In the case of $k > 2$ covariates, $\rho^2(\mathbf{x})$ can not be plotted, but for a subject with known covariate vector $\mathbf{x}$, an estimate of $\rho^2(\mathbf{x})$ provides a measure of the strength of the relationship to the response variable Y. An interesting application is the following: Suppose that Y is a health variable such as survival time, suppose that x_1 is the level of a treatment, and suppose that x_2, x_3, etc. are health status variables, such as blood pressure and cholesterol level. Now for a given subject the values of x_2, x_3, etc. are known; thus for fixed x_2, x_3, etc., $\rho(\mathbf{x})$ as a function of x_1 measures the strength of the relationship between treatment and survival as a function of treatment level. The case of several covariates will be considered further in a future article and is not dealt with further here.

3. NEIGHBORHOOD ESTIMATES OF THE CORRELATION CURVE

3.1 Neighborhood Estimates and Their Consistency

We assume that we have a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ of independent pairs all having the same distribution as (X, Y). We need estimates of σ_1^2, $\sigma^2(x) = \text{var}(Y|X = x)$, and $\beta(x) = \mu'(x) = (d/dx)\mu(x)$, where $\mu(x) = E(Y|X = x)$. We use $\hat{\sigma}_1^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ as the estimate of σ_1^2. Write $\sigma^2(x) = \mu_2(x) - \mu_1^2(x)$, where $\mu_j(x) = E(Y^j|X = x), j = 1, 2$. There are several classes of estimates of $\mu_j(x)$, $\sigma(x)$, and $\beta(x)$ available (see Härdle 1990, Hastie and Tibshirani 1990, Müller 1988, Nadaraya 1989, and Wahba 1990 for recent treatments with extensive lists of references). Two recent papers that focus on the estimation of the average of derivatives of conditional means and other functionals are those by Härdle and Stoker (1989) and Samarov (1993). Hall and Carroll (1990) analyzed the problem of estimating both $\mu(x)$ and $\sigma(x)$.

In this section we consider empirical neighborhood estimates where population means are estimated using neighborhood averages and population derivatives are estimated using empirical neighborhood derivatives. The empirical derivative estimates are very simple, their properties are transparent, and their asymptotic properties are comparable to those of other proposed estimates.

We will use methods based on local data pairs (X_i, Y_i) that correspond to X_i's close to the given covariate value x we are interested in. Thus we introduce the nearest-neighbor index set $I_k(x)$ as the collection of indices on the k X's among $X_1, \ldots, X_n$ that are closest to x, but with an equal number of X's on either side of x. More precisely, for k even, $I_k(x)$ denote the indices on the set of k values of $X_1, \ldots, X_n$ that have $k/2$ values to the left of x, have $k/2$ values to the right of x, and are closest to x. For k odd, replace $k/2$ by $(k-1)/2$. If $x = X_i$ for some i, then reduce k to $k - 1$ in the aforementioned definition. In all cases let r denote the number of x_i to the left (right) of x. To break ties, when two X's are equally close to x, we choose the smaller index. Now natural estimates of $\mu_j(x)$ and $\sigma^2(x)$ are

$$\hat{\mu}_j(x) = k^{-1} \sum_{i \in I_k(x)} Y_i^j, \; j = 1, 2, \; \hat{\sigma}^2(x) = \hat{\mu}_2(x) - \hat{\mu}_1^2(x).$$

Thus $\hat{\mu}_j(x)$ and $\hat{\sigma}^2(x)$ are the local Y sample moments based on $\{Y_i;\, i \in I_k(x)\}$. Our *empirical derivative* estimate of $\beta(x)$ is

$$\hat{\beta}(x) = \frac{r^{-1} \sum_{i \in I_k^+(x)} Y_i - r^{-1} \sum_{j \in I_k^-(x)} Y_j}{r^{-1} \sum_{i \in I_k^+(x)} X_i - r^{-1} \sum_{j \in I_k^-(x)} X_j},$$

where $I_k^+(x) = \{i\colon i \in I_k(x), X_i > x\}$ and $I_k^-(x) = \{j\colon j$

$\in I_k(x), X_j < x\}$ are the right and left index sets. It is convenient to write

$$\hat{\beta}(x) = \frac{\bar{Y}^+(x) - \bar{Y}^-(x)}{\bar{X}^+(x) - \bar{X}^-(x)},$$

where $\bar{Y}^+(x)(\bar{Y}^-(x))$ is the average of the local Y's corresponding to the X's to the right (left) of the given x and $\bar{X}^+(x)(\bar{X}^-(x))$ is the average of the local X's to the right (left) of x.

We refer to

$$\hat{\rho}(x) = [\hat{\sigma}_1\hat{\beta}(x)]/\{[\hat{\sigma}_1\hat{\beta}(x)]^2 + \hat{\sigma}^2(x)\}^{1/2}$$
$$= \pm\{1 + [\hat{\sigma}_1\hat{\beta}(x)/\hat{\sigma}(x)]^{-2}\}^{-1/2},$$

where $\pm$ denotes the sign of $\hat{\beta}(x)$, as the *nearest-neighbor (sample) correlation curve.*

In what follows we assume that σ_1^2, $\beta(x)$, and $\sigma^2(x)$ exist and that σ_1^2 and $\sigma^2(x)$ are positive. Then $\hat{\rho}(x)$ will be consistent when $\hat{\beta}(x)$ and $\hat{\sigma}^2(x)$ are consistent. We will consider consistency properties with k a function of n such that $k \to \infty$ and $(k/n) \to 0$ as $n \to \infty$.

We consider the conditional model where we condition on the X order statistics. In other words (Bhattacharya 1974), we consider the covariate values to be the nonrandom values $x_1 < \cdots < x_n$ and assume that $Y_1, \ldots, Y_n$ are independent, with Y_i having mean $\mu(x_i)$ and variance $\sigma^2(x_i)$. We assume that each x_i depends on n and that $x_1, \ldots, x_n$ is a *regular sequence* of covariate values in the sense that if $F_n(x) = n^{-1}[\#x_i \le x]$, then $F_n(x) \to F(x)$ for some continuous strictly increasing distribution function $F(x)$ with support (a, b), $a < b$ (a or b could be ∞). We give an account of the pointwise consistency of $\hat{\rho}(x)$ in this model. Throughout, x denotes one fixed point in (a, b) where the strength of the relationship between Y and X is being estimated.

In this setting we can justify our definition of $\hat{\beta}(x)$ by noting that when $\mu''(x)$ exists and is bounded in absolute value by A, $\hat{\beta}(x)$ is "nearly" unbiased; that is,

$$|E(\hat{\beta}(x)) - \beta(x)| \le \frac{1}{2}A \sum_{i\in I_k(x)} (x_i - x)^2 \Big/ \sum_{i\in I_k(x)} |x_i - x|.$$

The variance of $\hat{\beta}(x)$ is also easily bounded. Assume that $\text{var}(Y_i)$ is bounded by B; then

$$\text{var}(\hat{\beta}(x)) \le 4Bk \Big/ \left\{\sum_{i\in I_k(x)} |x_i - x|\right\}^2.$$

These considerations and similar considerations involving $\hat{\mu}_j(x)$, $j = 1, 2$, lead to the following proposition.

Proposition 3.1. Suppose that $\mu_1'(x)$ and $\mu_2''(x)$ exist and are bounded in absolute value, and assume that $\text{var}(Y_i^j)$, $j = 1, 2$, exist and are bounded; then $\hat{\rho}(x)$ converges in probability to $\rho(x)$ provided that

$$k^{1/2} \Big/ \sum_{i\in I_k(x)} |x_i - x| \to 0 \quad (4)$$

and

$$\left\{\sum_{i\in I_k(x)} (x_i - x)^2 \Big/ \sum_{i\in I_k(x)} |x_i - x|\right\} \to 0 \quad (5)$$

as $n \to \infty$ and $k \to \infty$.

Remark 3.1. The condition (5) holds provided we assume that the density $f(x)$ of $F(x)$ exists and

$$x_i = F^{-1}\left(\frac{i - \frac{1}{2}}{n}\right) + o\left(\frac{1}{n}\right), \quad (6)$$

where $F^{-1}(u)$ is differentiable with a continuous positive derivative $1/f(x)$ at the point $u_0 = F(x)$. In fact, under condition (6), the bias of $\hat{\rho}(x)$ is of the order (k/n) and the variance of $\hat{\rho}(x)$ is of the order (n^2/k^3).

If we assume the existence of $\mu_j'''(x)$, $j = 1, 2$, we can reduce the asymptotic bias further.

Proposition 3.2. Suppose that x_i satisfies (6). If we assume that $f'(x)$, $\mu_1'''(x)$, and $\mu_2'''(x)$ exist and are continuous at x, then the bias of $\hat{\rho}(x)$ is of the order $(k/n)^2$. In fact,

$$(n/k)^2[E(\hat{\rho}(x)) - \rho(x)] = \sigma_1[1 - \rho^2(x)]^{3/2}$$
$$\times [2\beta''(x)f(x) - 3\beta'(x)f'(x)]/96f^3(x)\sigma(x) + o(1).$$

Under the conditions of Proposition 3.2, we similarly find that

$$(k^3/n^2)\text{var}(\hat{\rho}(x)) = 16\sigma_1^2 f^2(x)[1 - \rho^2(x)]^3 + o(1),$$

and that the asymptotic mean squared error (AMSE) of $\hat{\rho}(x)$ is

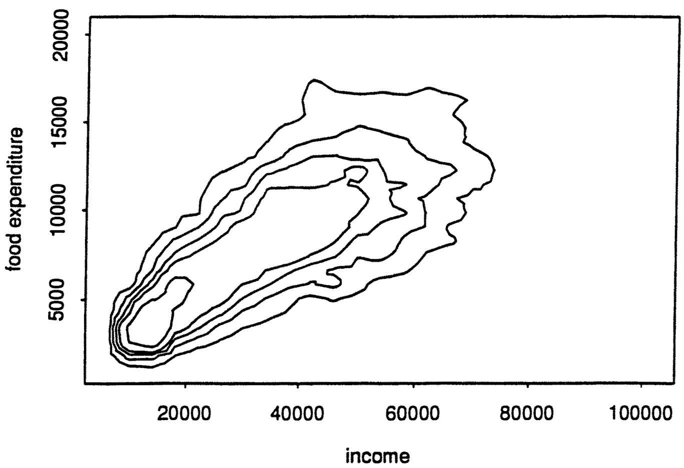
$$\text{AMSE}(\hat{\rho}(x)) = \sigma_1^2[1 - \rho^2(x)]^3$$
$$\times \left\{\left[\frac{2\beta''(x)f(x) - 3\beta'(x)f'(x)}{96f^3(x)\sigma(x)}\right]^2\left(\frac{k}{n}\right)^4 + 16f^2(x)\left(\frac{n^2}{k^3}\right)\right\}.$$

Remark 3.2. Note that the AMSE tends to 0 and $\hat{\rho}(x)$ is consistent provided that (k/n) and (n^2/k^3) both tend to 0. If k is of the form $k = cn^\delta$, $c > 0$, then this holds provided that $\frac{2}{3} < \delta < 1$.

To find the order of the nearest-neighborhood size k that minimizes the AMSE, we find the $k = cn^\delta$ for which the squared asymptotic bias and large sample variance have the same order. That is, we solve $(n^2/k^3) = (k/n)^4$ for δ. This gives $\delta = \frac{6}{7}$ and $k = cn^{6/7}$, and $\hat{\rho}(x)$ has an AMSE of the order $n^{-4/7}$. This is the same AMSE order that Gasser and Müller (1984) found for the convolution kernel estimate of $\beta(x)$ under the assumption of homoscedasticity and three continuous derivatives for $\mu(x)$. The rate $n^{-4/7}$ is the best possible for estimates of $\beta(x)$ and $\rho(x)$ under the conditions given (Stone 1980).

3.2 Asymptotic Normality of $\hat{\rho}(x)$: Confidence Intervals

Using a δ-method expansion of $\hat{\rho}(x) - \rho(x)$ in terms of $\hat{\sigma}_1 - \sigma_1$, $\hat{\sigma}(x) - \sigma(x)$, and $\hat{\beta}(x) - \beta(x)$, we find that only the term involving $\hat{\beta}(x) - \beta(x)$ contributes to the asymptotics. Note that $\hat{\beta}(x)$ is a sum of independent variables. Thus we can apply the Lindeberg–Feller central limit theorem. This gives



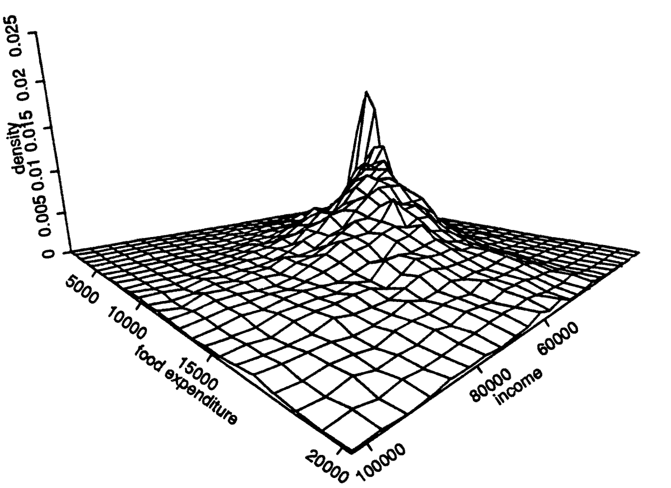


Figure 5. Empirical Contour Plots and Density Plot for the Net Income–Food Expenditure Data of Figure 2.

$$\frac{(n/k)^2[\hat{\rho}(x) - \rho(x)] - m(x)}{\tau_n(x)} \xrightarrow{\mathcal{L}} N(0, 1),$$

where

$$m(x) = \sigma_1[1 - \rho^2(x)]^{3/2} \times [2\beta''(x)f(x) - 3\beta'(x)f'(x)]/96f^3(x)\sigma(x)$$

and

$$\tau_n(x) = 4\sigma_1[1 - \rho^2(x)]^{3/2}f(x)(n^3/k^{7/2}).$$

This result holds when the ratio of the $(k/n)^3$ order term in the bias expansion of $\hat{\beta}(x)$ to the asymptotic standard deviation of $\hat{\beta}(x)$ tends to 0; that is, when $(k/n)^3/(n/k^{3/2}) = (k^{9/2}/n^4) \to 0$. When $k = cn^\delta$, this leads to the condition $\delta < \frac{8}{9}$.

The $\tau_n(x)$ term is simple and easy to estimate, whereas the $m(x)$ term is complicated. We can modify our asymptotics to eliminate the bias term $m(x)$ by requiring that the ratio of the $(k/n)^2$ term in the bias expansion of $\hat{\beta}(x)$ to the asymptotic standard deviation of $\hat{\beta}(x)$ tend to 0. This requires $(k^{7/2}/n^3) \to 0$, which in the case $k = cn^\delta$ means that we select $\delta < \frac{6}{7}$. In view of Remark 3.2, this leads to the restriction $\frac{2}{3} < \delta < \frac{6}{7}$. Under this restriction, assuming the Lindeberg condition, we find that

$$\frac{(k^{3/2}/n)[\hat{\rho}(x) - \rho(x)]}{a(x)} \xrightarrow{\mathcal{L}} N(0, 1),$$

where $a(x) = 4\sigma_1[1 - \rho^2(x)]^{3/2}f(x)$. A consistent estimate of $a(x)$ is $\hat{a}(x) = 4\hat{\sigma}_1[1 - \hat{\rho}^2(x)]^{3/2}(\bar{x}^+ - \bar{x}^-)^{-1}(k/2n)$. Thus an asymptotic $100(1 - \alpha)\%$ confidence interval for $\rho(x)$ is

$$\rho(x) = \hat{\rho}(x) \pm z\left(1 - \frac{1}{2}\alpha\right)2\hat{\sigma}_1[1 - \hat{\rho}^2(x)]^{3/2}(\bar{x}^+ - \bar{x}^-)^{-1}/\sqrt{k},$$

where $z(1 - \frac{1}{2}\alpha)$ is the $1 - \frac{1}{2}\alpha$ quantile of the standard normal distribution.

4. KERNEL ESTIMATES

4.1 Kernel Estimates and Their Consistency

As we do for nearest-neighbor estimates, we consider the covariate values $x_1 < \cdots < x_n$ to be fixed and form a regular sequence and assume that Y_i $(i = 1, \ldots, n)$ are independently distributed with mean $\mu(x_i) = E(Y|x_i)$ and variance $\sigma^2(x_i) = \text{var}(Y|x_i)$. Let $w(x)$ be a bounded integrable function with finite support $[-\tau, \tau]$ for some $\tau > 0$. We also assume that

$$\int_{-\tau}^{\tau} w(t)\, dt = 1 \quad \text{and} \quad w(\tau) = w(-\tau) = 0.$$

Such $w(x)$ is often referred to as a *kernel* (notice that it is not restricted to be nonnegative). Its integral, $W(x) = \int_{-\tau}^{x} w(t)\, dt$, is called an *integrated kernel.*

Let $s_i = (x_i + x_{i+1})/2$, $i = 1, \ldots, n$, $s_0 = x_1$, $s_n = x_n$, and let w_j, $j = 0, 1, 2$, be three kernels. Then the Gasser–Müller (1984) kernel estimates for $\mu_1(x) = \mu(x)$, $\mu_2(x) = E(Y^2|x)$, and $\beta(x) = \mu'(x)$ can be constructed as

$$\tilde{\mu}_j(x) = -\sum_{i=1}^{n}\left[W_j\left(\frac{x - s_i}{b_{jn}}\right) - W_j\left(\frac{x - s_{i-1}}{b_{jn}}\right)\right]Y_i^j, \quad j = 1, 2,$$

and

$$\tilde{\beta}(x) = \frac{-1}{b_{0n}}\sum_{i=1}^{n}\left[w_0\left(\frac{x - s_i}{b_{0n}}\right) - w_0\left(\frac{x - s_{i-1}}{b_{0n}}\right)\right]Y_i,$$

where $\{b_{jn}, n \geq 1, j = 0, 1, 2\}$ are sequences of bandwidths and $W_j(x) = \int_{-\infty}^{x} w_j(x)\, dx$. The corresponding estimate for the correlation curve $\rho(x)$ is then obtained as

$$\tilde{\rho}(x) = \frac{\hat{\sigma}_1\tilde{\beta}(x)}{[\hat{\sigma}_1^2\tilde{\beta}^2(x) + \tilde{\sigma}^2(x)]^{1/2}},$$

where $\tilde{\sigma}^2(x) = \tilde{\mu}_2(x) - \tilde{\mu}_1^2(x)$ and $\hat{\sigma}_1^2$ is the sample variance of the x_i's. We will call $\tilde{\rho}(x)$ the *kernel correlation curve* (with respect to the kernels w_j and bandwidth $\{b_{jn}, n \geq 1\}$, $j = 0, 1, 2$).

Because $\hat{\sigma}_1 \to \sigma_1$ and $\tilde{\rho}$ is a continuous function of $\tilde{\mu}_j$ $(j = 1, 2)$ and $\tilde{\beta}$ (for fixed x), the pointwise consistency (in both the weak and strong sense) of $\tilde{\rho}(x)$ is guaranteed once the pointwise consistency of $\tilde{\mu}_j(x)$ $(j = 1, 2)$ and $\tilde{\beta}(x)$ is established. Under the homoscedastic model

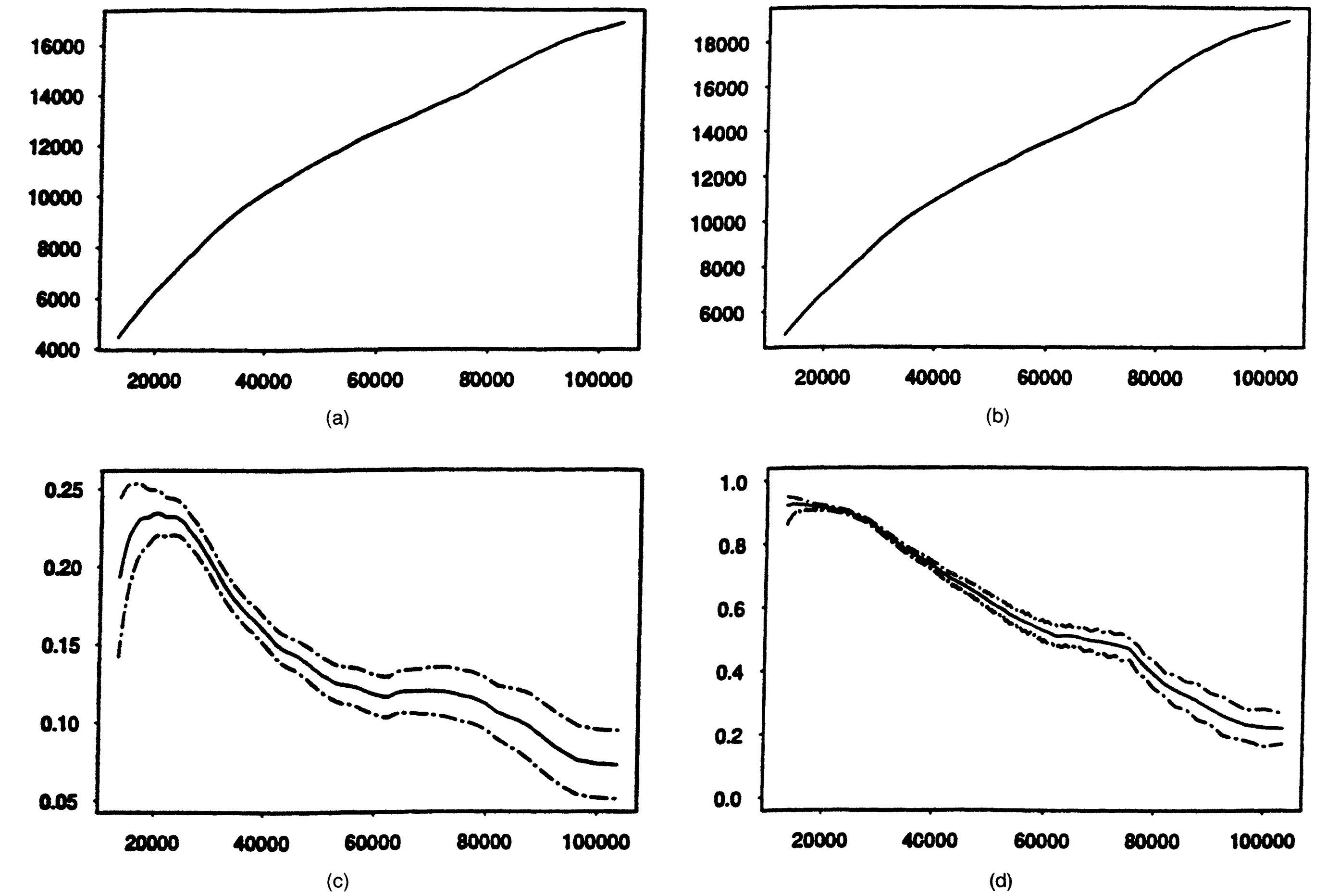


Figure 6. The Kernel Estimates $\tilde{\mu}(x)$, $\tilde{\sigma}(x)$, $\tilde{\beta}(x)$, and $\tilde{\rho}(x)$ of the Local (a) Mean, (b) Standard Deviation, (c) Slope, and (d) Correlation for the Net Income–Food Expenditure Data of Figure 2. The bands in (c) and (d) are 90% pointwise confidence bands.

$$Y_i = \mu(x_i) + \varepsilon_i,$$

$$\varepsilon_i \sim \text{iid}, \qquad E(\varepsilon_i) = 0, \qquad 0 < \text{var}(\varepsilon_i) \equiv \sigma^2 < \infty$$

and the assumption

$$\max_{1 \le i \le n} |s_i - s_{i-1}| = O\left(\frac{1}{n}\right), \tag{7}$$

Gasser and Müller (1984) established some general results for kernel estimates (of the aforementioned type) of regression functions, as well as for their derivatives. We noticed that with little modifications to their proofs, their theorems 1 and 2 can be extended to the general heteroscedastic model

$$Y_i = \mu(x_i) + \varepsilon_i,$$

$$\varepsilon_i \sim \text{indep.}, \qquad E(\varepsilon_i) = 0, \qquad 0 < \text{var}(\varepsilon_i) < \infty,$$

as long as

$$\max_{1 \le i \le n} \text{var}(\varepsilon_i) \le B < \infty, \tag{8}$$

where B is a constant that does not depend on the sample size n. These results then can be used to establish the consistency of our kernel correlation curve $\tilde{\rho}(x)$. The regularity condition (7) is assumed for all the following propositions.

Proposition 4.1 (Weak Consistency of $\tilde{\rho}(x)$). Suppose that

$$\max_{1 \le i \le n} E(Y^4 | x_i) \le B < \infty, \tag{9}$$

and as $n \to \infty$,

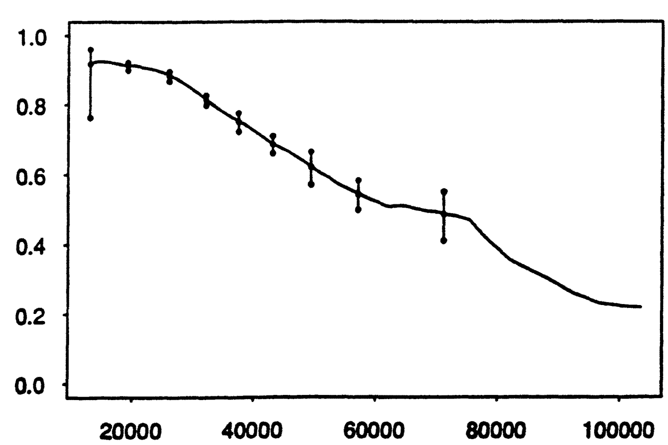


Figure 7. The Estimated Correlation Curve $\tilde{\rho}(x)$ With 90% Simultaneous Confidence Intervals at Each of the Quantiles $x_{.1}, x_{.2}, \ldots, x_{.9}$ for the Net Income–Food Expenditure Data of Figure 2.

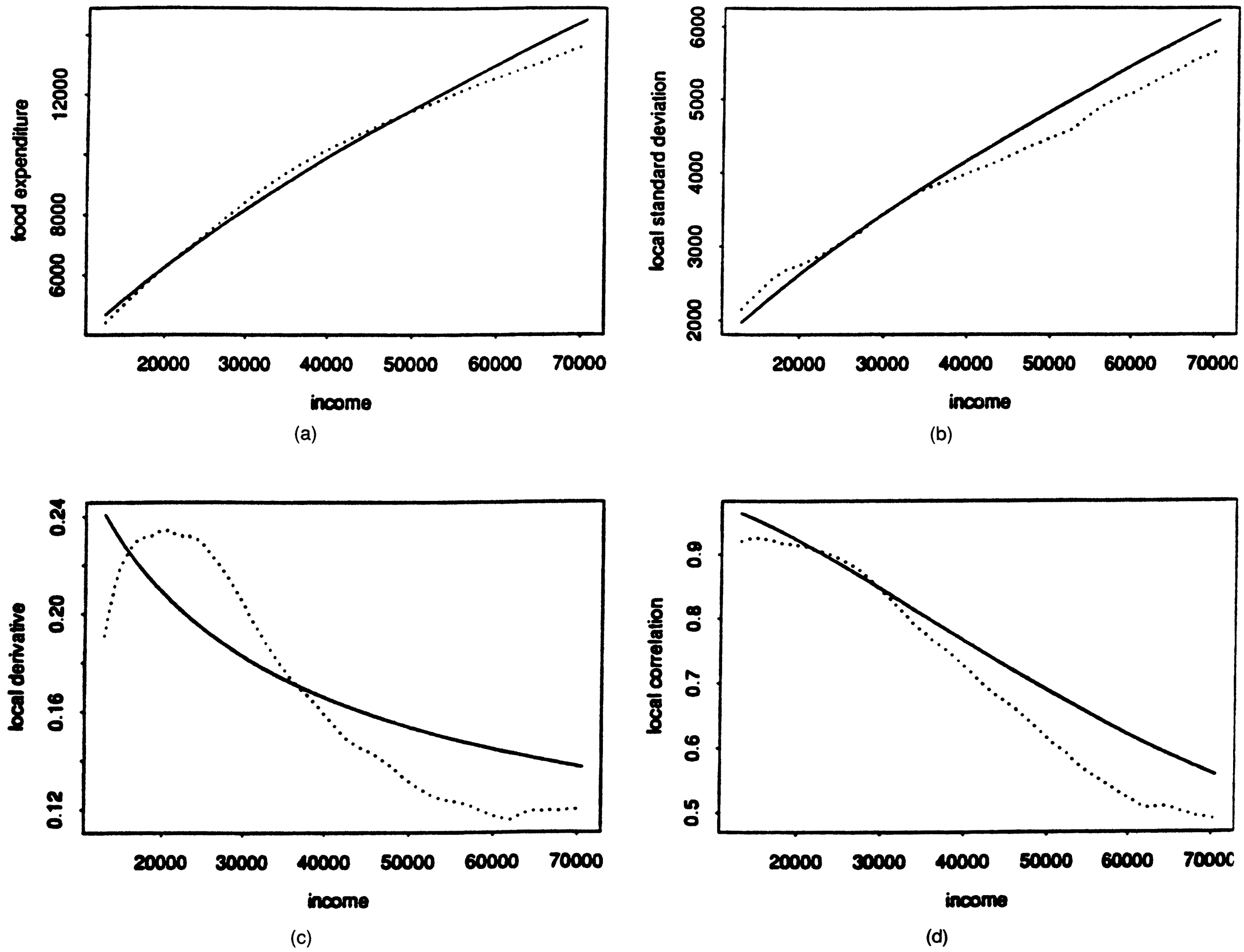


Figure 8. Parametric (Solid Curves) and Nonparametric (Dotted Curves) Estimates for the Net Income–Food Expenditure Data. The curves estimated are (a) $\mu(x)$, (b) $\sigma(x)$, (c) $\beta(x)$, and (d) $\rho(x)$.

(a) $b_{jn} \to 0$ and

(b) $nb_{jn}^3 \to \infty, \qquad j = 0, 1, 2.$

Then $\tilde{\rho}(x) \xrightarrow{p} \rho(x)$ at any x such that $\mu_j(x)$ ($j = 1, 2$) and $\beta(x)$ are continuous.

Proposition 4.2 (Strong Consistency of $\tilde{\rho}(x)$). Suppose that

$$\max_{1 \le i \le n} E(|Y|^{2p}|x_i) \le B < \infty \tag{10}$$

for some $p \ge 2$ and, in addition to (a) and (b),

(c) $n^{1/p} b_{jn} \to \infty$, and

(d) $\sum_{n=1}^{\infty} \exp\{-n^{1/2(1-p^{-1})} b_{jn}\} < \infty, \qquad j = 0, 1, 2.$

Then $\tilde{\rho}(x) \xrightarrow{\text{a.s.}} \rho(x)$ at any x such that $\mu_j(x)$ ($j = 1, 2$) and $\beta(x)$ are continuous.

Remark 4.1. For bandwidth sequences of the form $b_{jn} = c_j n^{-\alpha_j}$, $c_j > 0$, $\alpha_j > 0$, the necessary and sufficient condition for b_{jn} to satisfy all (a)–(d) is

$$0 < \alpha_j < \min\left\{\frac{1}{3}, \frac{1}{p}, \frac{1}{2}\left(1 - \frac{1}{p}\right)\right\}.$$

For example, when $p = 2$ (in this case (9) and (10) are equivalent), we have strong consistency when α_j is any value between 0 and $\frac{1}{4}$, $j = 0, 1, 2$. (a) and (b) imply that we need $0 < \alpha_j < \frac{1}{3}$ to have weak consistency. If we compare this with Remark 3.2 using the kernel-neighbor connecting formula $b_{jn} = (k/n)(2f(x))^{-1}$, we find that $0 < \alpha_j < \frac{1}{3}$ corresponds exactly to $\frac{2}{3} < \delta < 1$ when $k = cn^{\delta}$.

4.2 Asymptotic Normality: Confidence Intervals

As we did in Section 3, we assume that $\mu'''(x)$ exists. We consider kernels with support $[-1, 1]$. In this case the optimal kernel for the estimation of $\mu_j(x)$, $j = 1, 2$, is the Epanechnikov kernel $.75(1 - t^2)I[|t| \le 1]$, and the optimal kernel for the estimation of $\beta(x)$ is the GMM kernel (Gasser, Müller, and Mammitzsch 1985)

$$w_0(t) = 3.75(t^3 - t)I[|t| \le 1].$$

With these kernels, $\tilde{\rho}(x)$ has an AMSE of

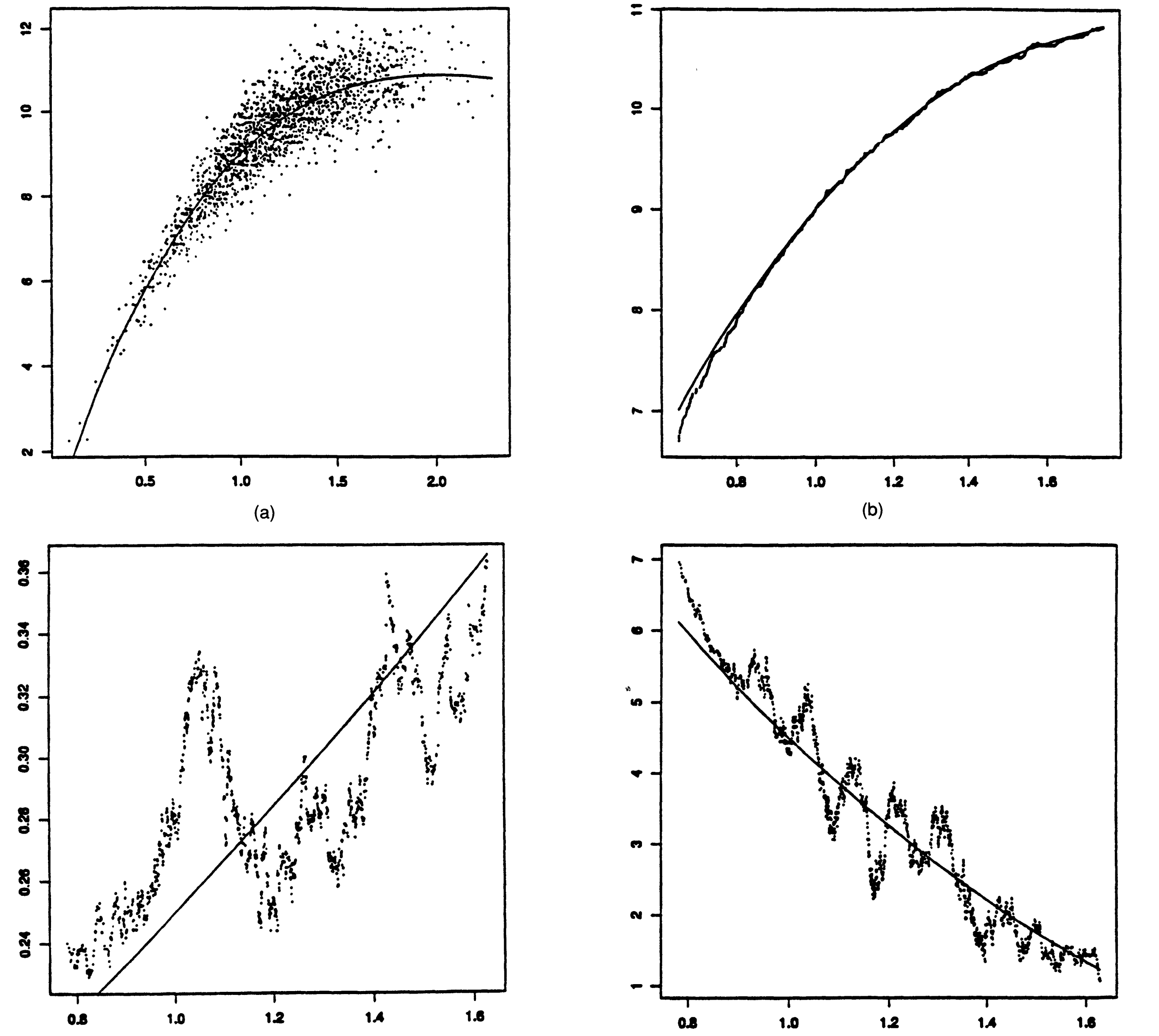


Figure 9. Nearest-Neighbor Estimates With k = 60 Based on n = 2,000 Observations From the Twisted Pear Model of Example 1.1 and Figure 4. (a) Scatterplot and True Local Mean $\mu(x)$. (b) True and Estimated Local Mean. (c) True and Estimated Local Residual Variance. (d) True and Estimated Local Regression Slope.

AMSE$(\tilde{\rho}(x))$

$$= \sigma_1^2[1-\rho^2(x)]^3\left\{\left[\frac{.0714\beta''(x)}{\sigma(x)}\right]^2 b_{on}^4 + \frac{1}{f(x)}\frac{1}{nb_{0n}^3}\right\}.$$

Moreover,

$$\frac{b_{0n}^{-2}[\tilde{\rho}(x)-\rho(x)] - \tilde{m}(x)}{\tilde{\tau}_n(x)} \xrightarrow{\mathcal{L}} N(0,1),$$

where

$$\tilde{m}(x) = \sigma_1[1-\rho^2(x)]^{3/2}.0714\beta''(x)/\sigma(x)$$

and

$$\tilde{\tau}_n(x) = \sigma_1[1-\rho^2(x)]^{3/2}/[nb_{0n}^7 f(x)]^{1/2}.$$

The $\tilde{m}(x)$ term in this result can be eliminated by requiring

7

Figure 10. True and k = 60 Nearest-Neighbor Estimated Correlation Curves Based on 2,000 observations from the Model of Example 1.1 and Figure 3.

that $(nb_{0n}) \to 0$. When $b_{0n} = cn^{-\alpha}$, $c > 0$, $\alpha > 0$, this means that $\alpha > \frac{1}{7}$. In view of Remark 4.1, this leads to the restriction $\frac{1}{7} < \alpha < \frac{1}{3}$. Under this restriction, assuming Lindeberg's condition, we find that

$$(nb_{0n}^3)^{1/2}[\tilde{\rho}(x) - \rho(x)]/b(x) \to N(0, 1),$$

where $b(x) = \sigma_1[1 - \rho^2(x)]^{3/2}/f^{1/2}(x)$.

A consistent estimate of $b^2(x)$ is

$$\hat{b}^2(x) = \hat{\sigma}_1^2[1 - \tilde{\rho}^2(x)]^3 nb_{0n}^3 \times \sum_{i=1}^{n} \left[w_0\left(\frac{x - s_i}{b_{0n}}\right) - w_0\left(\frac{x - s_{i-1}}{b_{0n}}\right) \right]^2.$$

It follows that an asymptotic $100(1 - \alpha)\%$ confidence interval for $\rho(x)$ is

$$\rho(x) = \tilde{\rho}(x) \pm z\left(1 - \frac{1}{2}\alpha\right)\hat{\sigma}_1[1 - \tilde{\rho}^2(x)]^{3/2}\hat{b}(x).$$

5. CHOOSING THE BANDWIDTHS

The neighborhood size k and the bandwidths b_{jn} should be chosen to make the mean squared error of the estimator small. There are three ways to try to accomplish this: cross-validation, the plug-in method, and the reference distribution approach. We will use the third method to choose the bandwidths in $\tilde{\mu}_1(x)$, $\tilde{\mu}_2(x)$, and $\tilde{\beta}(x)$.

Table 1. Monte Carlo Bias, Standard Deviation, and Root Mean Squared Error of $\hat{\rho}(x)$ for 500 Simulations from Model (12)

	Fixed design			Random design		
Quantile	$x_{.25}$	$x_{.50}$	$x_{.75}$	$x_{.25}$	$x_{.50}$	$x_{.75}$
Bias	−.0057	−.0129	−.0156	−.0042	−.0098	−.0148
Standard deviation	.0081	.0131	.0220	.0121	.0187	.0288
Root MSE	.0098	.0183	.0271	.0128	.0211	.0323

NOTE: The sample size is n = 2,000.

The idea behind the reference distribution approach is to approximate the joint distribution of (X, Y) by a relatively simple parametric model and then use the bandwidth that is optimal for this model. The chosen bandwidth will not converge to the optimal bandwidth unless the parametric model coincides with the true model, but it will have a relatively small variance. A natural model would be the power transformation model

$$h(Y; \lambda) = \alpha + \theta h(x; \gamma) + \varepsilon, \qquad (11)$$

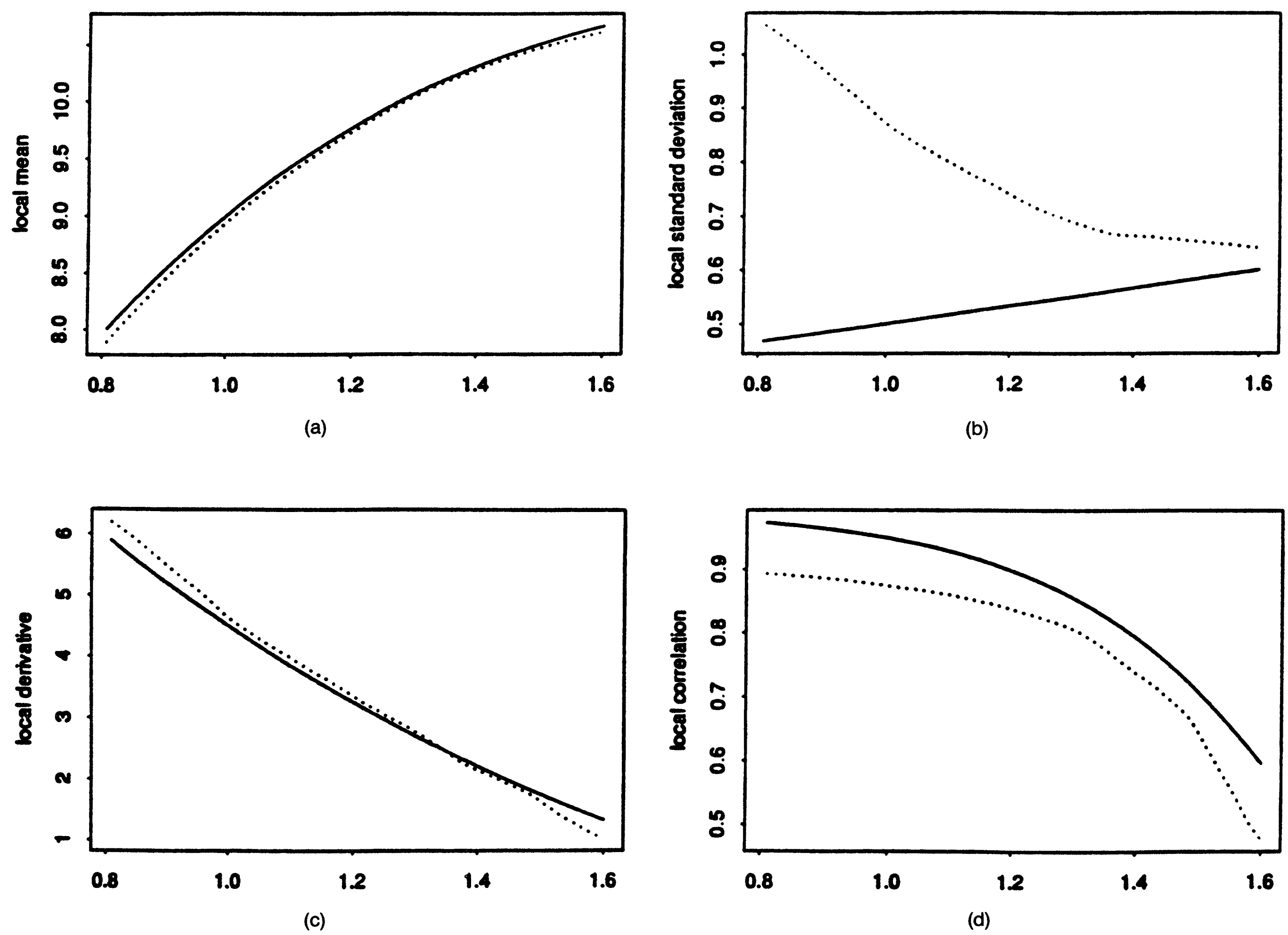


Figure 11. The True Curves (Solid Lines) and Kernel Estimates With Data-Based Bandwidths for (a) $\mu(x)$, (b) $\sigma(x)$, (c) $\beta(x)$, and (d) $\rho(x)$. The estimates are based on 2,000 observations from the model of Example 1.1 and Figure 3.

where X and ε are independent and normally distributed, $E(\varepsilon) = 0$, $\text{var}(\varepsilon) = \sigma_\varepsilon^2$, and

$$h(t; \lambda) = \frac{t^\lambda - 1}{\lambda}, \qquad \lambda \neq 0$$
$$= \log(t), \qquad \lambda = 0.$$

The optimal bandwidths for the model (11) can be computed and will be functions of α, θ, λ, and γ. Next, α, θ, λ, and γ are estimated by maximizing the likelihood for the model (11).

For the income–food expenditure data of Figure 2, a log-log model seemed particularly appropriate; thus we consider the model

$$\log Y = \alpha + \theta \log X + \varepsilon. \tag{12}$$

We find, from Müller (1988), that the locally optimal bandwidths for $\tilde{\mu}_j(x)$, $j = 1, 2$, are

$$b_{jn} = \{(3/20)\sigma_j^2(x)/n(.1)^2 f(x)[\mu_j''(x)]^2\}^{1/5}, \qquad j = 1, 2,$$

where $\sigma_j^2(x) = \text{var}(Y^j \mid x)$, $j = 1, 2$. Similarly, the locally optimal bandwidth for $\tilde{\beta}(x)$ is

$$b_{0n} = \{.75(2.143)\sigma^2(x)/n(.0714)^2 f(x)[\mu'''(x)]^2\}^{1/7}$$

Note that the locally optimal bandwidths depend on x. Thus we write $b_{jn} = b_{jn}(x)$. For the model (12),

$$\mu_j(x) = \exp\{j(\alpha + \theta \log x) + \frac{1}{2}(j\sigma_\varepsilon)^2\}, \qquad j = 1, 2,$$

and

$$\sigma_j^2(x) = [\exp\{2j(\alpha + \theta \log x) + (j\sigma_\varepsilon^2)\}][\exp(j\sigma_\varepsilon)^2 - 1], \qquad j = 1, 2.$$

Thus $b_{jn}(x)$, $j = 0, 1, 2$, can be readily expressed in terms of α, θ, σ_1^2, and σ_ε^2. Finally, α, θ, σ_1^2, and σ_ε^2 are estimated using a least squares package on the data ($\log X_i$, $\log Y_i$), $i = 1, \ldots, n$. For the model (12), this yields the maximum likelihood estimates $\hat{b}_{jn}$ of b_{jn}, $j = 0, 1, 2$. In the next section we use these local bandwidths when illustrating the correlation curve concept.

6. REAL DATA ILLUSTRATIONS

Our methods are useful only if the number of (x, y) pairs in the sample is 150 or more. Fortunately, there is an abundance of such data sets in computer data bases.

Consider again the data set of Figure 2, which gives the scatterplot of x = net income and y = expenditure for food for 7,125 households in Great Britain. In Figure 5 we give S-produced empirical contour plots and an empirical density plot that clearly demonstrate the "pear-shape" aspect of this data set. These plots were computed by averaging five S-produced histograms based on grids of 25 rectangles in the (x, y) plane. This smoothing technique for bivariate densities was suggested to us by Steve Marron.

Next we give the mean function kernel estimate $\tilde{\mu}(x)$ and the standard deviation function estimate $\tilde{\sigma}(x)$, as well as the kernel estimate $\tilde{\beta}(x)$ of the local regression coefficient and the kernel estimate $\tilde{\rho}(x)$ of the correlation curve for the net income–food expenditure data (Fig. 6). In this plot the bandwidth is chosen using the reference distribution approach described in Section 5. In particular, the reference distribution is determined by (12). We also computed nearest-neighbor curves, as described in Section 3. These curves produced similar results and are not given here. Note that $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$ both increase steadily. $\tilde{\beta}(x)$ decreases fast in the region from 25,000 to 55,000, then levels off before resuming its decline around x = 77,000. There is a curious increase in $\tilde{\beta}(x)$ at the low end, which may be due to a boundary effect. In fact, the pointwise confidence band shows the high uncertainty in this region. The correlation curve estimate $\tilde{\rho}(x)$ shows a steady decline in the strength of the relationship between income x and food expenditure; it starts at .92 for the low income levels and then drops steadily. But there is still some correlation at the higher income levels. In fact, the estimated local correlation never dips below .22. For this data set, the Galton–Pearson correlation coefficient takes the value r = .602.

Next, in Figure 7 we give the 90% Bonferroni simultaneous confidence intervals for $\rho(x)$ at the quantiles $x_{.1}$, $x_{.2}$, $\ldots$, $x_{.90}$. Because of the large sample size, simultaneous significance obtains. Not only can we conclude that the correlations at these quantiles are significant at the 10% level, we can also conclude that the correlation is significantly larger at the 10th percentile than at the 60th percentile, significantly higher at the 20th percentile than at the 40th percentile, and so on.

Finally, we compare parametric and nonparametric correlation curves. Thus we compute $\mu^*(x)$, $\sigma^*(x)$, $\beta^*(x)$, and $\rho^*(x)$ for the parametric model (12) with the parameters estimated by maximum likelihood. In Figure 8 we see that the parametric curves $\mu^*(x)$ and $\sigma^*(x)$ follow the nonparametric curves $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$ well, indicating that the model (12) gives a good first-order approximation to the data. But $\beta^*(x)$ and $\rho^*(x)$ are quite different from the empirical kernel curves $\tilde{\beta}(x)$ and $\tilde{\rho}(x)$, indicating that the parametric curves are severely biased for functions involving derivatives, because simulation studies show that $\tilde{\beta}(x)$ and $\tilde{\rho}(x)$ have small bias for large sample sizes (see Müller 1988 and Table 1 of this article).

7. SIMULATED DATA

7.1 Simulations

We did a number of simulations to determine how close our empirical location, slope, and correlation curves fall to the true curves for various models. Here we present the results for the model $Y = \mu(X) + \tau(X)\varepsilon$ as given in Example 1.1 and Figure 3. X and ε are generated independently as $X \sim N(1.2, (1/3)^2)$ and $\varepsilon \sim N(0, (1/3)^2)$. We use n = 2,000 and k = 60. Figure 9a shows the scatterplot and the true local mean $\mu(x)$, and Figure 9b shows the true curve $\mu(x)$ and the estimate $\hat{\mu}(x)$. Similarly, Figures 9c and 9d, shows the true and estimated local variance and slope.

Next, Figure 10 shows the true and estimated correlation slopes. Figures 9 and 10 indicate that with neighborhood size k = 60, the estimates are somewhat erratic but are, on the average, nearly unbiased. By choosing a larger k, we would obtain smoother curves with a larger bias.

Finally, in Figure 11 we give the kernel estimates of $\mu(x)$, $\sigma(x)$, $\beta(x)$, and $\rho(x)$ for the model of Example 1.1. We use the reference distribution approach with the reference model

$$\log Y = \alpha + \beta X + \varepsilon,$$

where X and ε are independent and normally distributed. Note that even with the wrong reference distribution, the reference distribution approach yields good results. The estimate of $\rho(x)$ is very smooth but a bit negatively biased. Note that $\tilde{\sigma}(x)$ overestimates $\sigma(x)$ in the region where $\mu(x)$ has high curvature (cf. Hall and Carroll 1990).

7.2 Simulation Results

We did a simulation study to evaluate the accuracy of the kernel estimate with the bandwidth selected using the reference distribution approach of Section 5. We considered the model

$$\log Y = \alpha + \beta \log X + \varepsilon,$$

with $\log X$ and ε independent and normally distributed with distributions $N(10.44, .41)$ and $N(0, .16)$. The true values of α and β are 2.05 and .67. This model was chosen because it gives a crude approximation to net income–food expenditure data. We obtained the distribution of $\tilde{\rho}(x)$ in 500 simulations from this model for x equal to the 25th, 50th, and 75th percentiles. We considered both the random design case, where X is random with a log normal distribution F, and the fixed design case with $x_i = F^{-1}((i - \frac{1}{2})/n)$, F lognormal as before. The results, presented in Table 1, show that the kernel estimate of $\rho(x)$ with the bandwidth chosen by the reference method is very accurate at the sample size $n = 2{,}000$. The bias is small, and absolute bias is much smaller than the standard deviation. The root mean squared error is smaller near the first quantile $x_{.25}$, which reflects the fact that for our lognormal X distribution, the data are more concentrated near this value.

8. SUMMARY

We have addressed the question of how to measure the strength of the relationship between a regressor x and a response Y in heterocorrelatious experiments, where the strength of this relationship depends on the level of x. We proposed a correlation curve that measures this relationship as a ratio $\rho(x)$ of the local variability $\sigma_1\beta(x)$ explained by regression to the total local variability $\{\sigma_1^2\beta^2(x) + \sigma^2(x)\}^{1/2}$, where $\beta(x)$ is the derivative of $\mu(x)$, $\mu(x) = E(Y|x)$, $\sigma^2(x) = \text{var}(Y|x)$, and $\sigma_1^2 = \text{var}(X)$.

We considered two classes of estimates of the correlation curve $\rho(x)$. The first is based on nearest-neighbor–type estimates of $\beta(x)$ and $\sigma^2(x)$, and the second is based on a Gasser–Müller-type kernel estimate of $\beta(x)$ and $\sigma^2(x)$. We gave conditions under which these estimates are consistent and asymptotically normal. To choose the bandwidth b in the kernel estimates, we fit a power transformation model to the data and used the optimal plug-in estimate of b for this model. We showed this approach to work well in examples and Monte Carlo simulations that give the mean squared error of the estimated correlation curve.

APPENDIX: PROOFS

Proof of Proposition 3.2

Use the δ method for moments to express $E(\hat{\rho}(x))$ in terms of $\hat{\sigma}_1^2$, $E(\hat{\sigma}^2(x))$, and $E(\hat{\beta}(x))$. Note that in the δ method, only the term involving $E(\hat{\beta}(x)) - \beta(x)$ contributes to the asymptotic bias of $\hat{\rho}(x)$, because the terms involving $E(\hat{\sigma}_1^2) - \sigma_1^2$ and $E(\hat{\sigma}^2(x)) - \sigma^2(x)$ converge to 0 at a faster rate than do the terms involving $E(\hat{\beta}(x)) - \beta(x)$. Now use the Taylor expansion

$$\mu(x_i) = \mu(x) + \sum_{j=1}^{3} [\mu^{(j)}(x)/j!](x_i - x)^j + o(x_i - x)^3$$

to complete the calculation.

Proof of Proposition 4.1

Applying Theorem 1 of Gasser and Müller (1984), we have $\tilde{\mu}_1(x) \xrightarrow{p} \mu_1(x)$ and $\tilde{\beta}(x) = \tilde{\mu}_1'(x) \xrightarrow{p} \mu_1'(x) = \beta(x)$. Applying the same theorem to the model $Y_i^2 = E(Y^2|x_i) + \varepsilon_i' = \mu_2(x_i) + \varepsilon_i'$, we obtain $\tilde{\mu}_2(x) \xrightarrow{p} \mu_2(x)$. Notice that (9) guarantees that (8) is satisfied for both ε_i and ε_i'.

Proof of Proposition 4.2

The proof is the same as for Proposition 4.1, except for applying theorem 2 of Gasser and Müller (1984) instead of theorem 1. Notice that (10) is stronger than (9) when $p > 2$.

[Received May 1991. Revised July 1993.]

REFERENCES

Aalen, O. O. (1987), "Dynamic Modeling and Causality," *Scandinavian Actuarial Journal,* 4, 177–190.
Bhattacharya, P. K. (1974), "Convergence of Sample Paths of Normalized Sums of Induced Order Statistics," *The Annals of Statistics,* 2, 1034–1039.
Fisher, J. (1959), "The Twisted Pear and the Prediction of Behavior," *Journal of Consulting Psychology,* 23, 400–405.
Gasser, T., and Müller, H. G. (1984), "Estimating Regression Functions and Their Derivatives With Kernel Methods," *Scandinavian Journal of Statistics,* 11, 171–185.
Gasser, T., Müller, H. G., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society,* Ser. B, 47, 283–52.
Gjerde, P. F., Block, J., and Block, J. H. (1988), "Depressive Symptoms and Personality During Late Adolescence: Gender Differences in the Externalization-Internalization of Symptom Expression," *Journal of Abnormal Psychology,* 97, 475–586.
Hall, P., and Carroll, R. J. (1990), "Variance Function Estimation in Regression: The Effect of Estimating the Mean," *Journal of the Royal Statistical Society,* Ser. B, 51, 3–14.
Härdle, W. (1990), *Applied Nonparametric Regression* (Econometric Society Monographs), Cambridge, U.K.: Cambridge University Press.
Härdle, W., and Stoker, T. (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association,* 84, 986–995.
Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models,* London: Chapman and Hall.
Müller, H. G. (1988), *Nonparametric Regression Analysis of Longitudinal Data* (Lecture Notes in Statistics) New York: Springer.
Nadaraya, E. A. (1989), *Nonparametric Estimation of Probability Densities and Regression Curves,* Dordrecht: Kluwer.
Samarov, A. M. (1993), "Exploring Regression Structure Using Nonparametric Functional Estimation," *Journal of the American Statistical Association,* 88, 836–847.
Stone, C. J. (1980), "Optimal Rates of Nonparametric Estimation," *The Annals of Statistics,* 8, 1348–1360.
Schweder, T. (1970), "Composable Markov Processes," *Journal of Applied Probability,* 7, 400–410.
Wahba, G. (1990), "Spline Models for Observational Data," Philadelphia: SIAM.