The Annals of Statistics
1993, Vol. 21, No. 2, 890–902

CORRELATION CURVES: MEASURES OF ASSOCIATION AS FUNCTIONS OF COVARIATE VALUES

BY STEINAR BJERVE[1] AND KJELL DOKSUM[2]

University of Oslo and University of California, Berkeley

For experiments where the strength of association between a response variable Y and a covariate X is different over different regions of values for the covariate X, we propose local nonparametric dependence functions which measure the strength of association between Y and X as a function of $X = x$. Our dependence functions are extensions of Galton's idea of strength of co-relation from the bivariate normal case to the nonparametric case. In particular, a dependence function is obtained by expressing the usual Galton–Pearson correlation coefficient in terms of the regression line slope β and the residual variance σ^2 and then replacing β and σ^2 by a nonparametric regression slope $\beta(x)$ and a nonparametric residual variance $\sigma^2(x) = \text{var}(Y|x)$, respectively. Our local dependence functions are standardized nonparametric regression curves which provide universal scale-free measures of the strength of the relationship between variables in nonlinear models. They share most of the properties of the correlation coefficient and they reduce to the usual correlation coefficient in the bivariate normal case. For this reason we call them correlation curves. We show that, in a certain sense, they quantify Lehmann's notion of regression dependence. Finally, the correlation curve concept is illustrated using data from a study of the relationship between cholesterol levels x and triglyceride concentrations y of heart patients.

1. Introduction. For bivariate experiments where the contour plots are nearly shaped like lemons or ellipses, the correlation coefficient ρ is a very concise and convenient measure of the strength of the association between the two random variables X and Y. However, in many interesting cases, the contour plots cannot be assumed to be elliptical. For instance, Fisher (1959) reported on studies in psychology and other fields where the association between the response variable Y and covariate X is strong for large values of $X = x$, but the association is weak or nonexistent for small x. In particular, Fisher describes studies where the association between a score X giving level of brain disease is strongly associated with an independently assessed score Y indicating level of pathological behaviour for patients with large values of $X = x$, but the association gets weaker as $X = x$ decreases. Fisher gives an

Received January 1991; revised April 1992.
[1]Partially supported by a grant from the Johan and Mimi Wessmann foundation.
[2]Work partially supported by grant from the Norwegian research foundation for science and the humanities and by NSF Grant DMS-89-01603. This work was partially carried out while visiting the Statistics Department at Harvard University.
AMS 1991 *subject classifications*. 62J02, 62G99
Key words and phrases. Nonparametric regression, nonlinearity, heteroscedasticity, kernel estimation, local correlation.

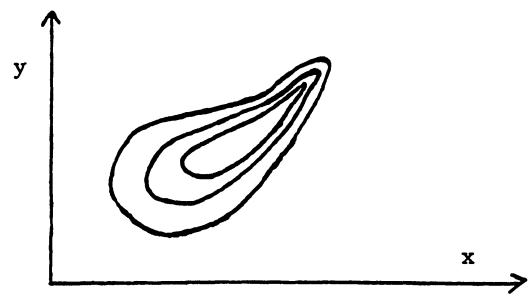

FIG. 1. *A typical twisted pear contour plot. x is level of symptom and y is level of disease.*

associated contour plot and calls it a twisted pear. See Figure 1 which gives a representation of J. Fisher's contour plot.

Our next example is from financial analysis. Here studies [e.g., Karpoff (1987)] of stock market behavior has revealed that the association between change X in prices and volume Y moves from negative to positive as $X = x$ goes from negative to positive. Using Karpoff's plot and data description, we conclude that the contour plot in this case looks somewhat like a twisted sausage or a banana. See Figure 2.

In the statistical literature, there is also an abundance of examples where the strength of association changes with the levels x of the covariate X. See for instance Anscombe (1961), Bickel (1978), Carroll and Ruppert (1982, 1988), Breiman and Friedman (1985) and Silverman (1986). The methods proposed for handling such situations include transformation techniques where the X's and Y's are transformed according to some criteria to the case where the strength of the association does not change with the covariate values. However, in many applications the change in the strength of association is of interest and this change is erased by the transformations. Another approach is nonparametric regression which involves computing estimates of the conditional mean or median of Y given $X = x$. These regression methods only

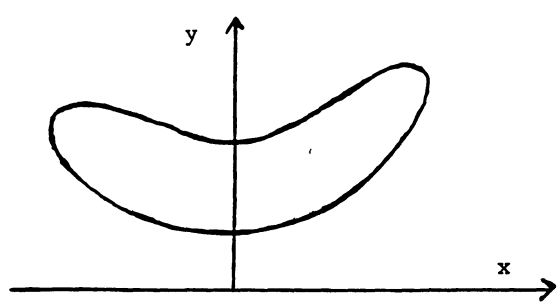

FIG. 2. *A contour adaption of Karpoff's Figure* 1. *x is change in price and y is level of volume.*

consider average (or median) conditional behaviour and do not take into account the width (in the y direction) of the contour plot. From Figure 1 it is clear that the width of the contour in the y direction (heteroscedasticity) is very important for the strength of association. Thus when the strength of the association is of interest, the regression methods need to be supplemented with a measure of spread for Y given $X = x$.

2. A correlation curve. Our approach is to construct a measure of local strength of association by combining ideas from nonparametric regression and Galton (1888). According to Galton [see Stigler (1986), page 297; (1989)], the strength of the co-relation between X and Y can be taken as the slope of the regression line computed after X and Y have both been converted to standardized scales $X' = (X - \mu_1)/\sigma_1$ and $Y' = (Y - \mu_2)/\sigma_2$, where (μ_1, σ_1) and (μ_2, σ_2) are location and scale parameters for X and Y, respectively.

When (X, Y) is bivariate normal $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, this leads to the familiar formula

$$\rho = \sigma_1\beta/\sigma_2 \quad \text{(normal case)},$$

where β is the regression slope when Y is regressed on X. Next we introduce the familiar [e.g., Bickel and Doksum (1977), page 36] decomposition

$$\sigma_2^2 = \operatorname{var}(Y) = \text{variance explained} + \text{residual variance}$$

$$= (\sigma_1\beta)^2 + \sigma^2 \quad \text{(normal case)},$$

where $\sigma^2 = \sigma_2^2(1 - \rho^2) = \operatorname{var}(Y|x) = \operatorname{var}(Y|X = x)$ is the variance of Y given $X = x$. We can now write

$$\rho = \frac{\sigma_1\beta}{\left[(\sigma_1\beta)^2 + \sigma^2\right]^{1/2}} \quad \text{(normal case)}. \tag{2.1}$$

In this representation we see how the correlation coefficient ρ is determined by the regression slope β and the residual variance σ^2. The representation also suggests that in the nonnormal world of twisted pears and sausages, a very natural local measure of the strength of the association between Y and X near $X = x$ is the correlation curve

$$\rho(x) = \frac{\sigma_1\beta(x)}{\left[\{\sigma_1\beta(x)\}^2 + \sigma^2(x)\right]^{1/2}} \quad \text{(general case)}, \tag{2.2}$$

where $\beta(x) = \mu'(x)$ is the slope of the nonparametric regression $\mu(x) = E(Y|x) = E(Y|X = x)$, $\sigma^2(x) = \operatorname{var}(Y|x)$ is the nonparametric residual variance and $\sigma_1^2 = \operatorname{var}(X)$ as before. This correlation curve concept makes sense only when X is a continuous random variable, in fact, we assume that $\mu(x) = E(Y|x)$ is differentiable. The distribution of Y can be discrete or continuous. We assume that σ_1^2 and $\sigma^2(x)$ exist.

$\rho(x)$ measures the strength of the association between X and Y locally at $X = x$. Thus, in the price–volume example (Figure 2), the correlation curve

would be negative for x negative and positive for x positive. More generally, for some number x_0, we could have $\rho(x)$ negative for $x < x_0$ and $\rho(x)$ positive for $x > x_0$. On the basis of price–volume data we could find the region "$x < x_1$," where $\rho(x)$ is significantly negative and the region "$x > x_2$," where $\rho(x)$ is significantly positive. In the J. Fisher example where small x has little or no influence on the distribution of Y while large x does (Figure 1), $\rho(x)$ would start out near zero and then increase toward one.

In most applications of linear statistical analysis, the regression slope is of greater interest than the regression line. We are suggesting that the same may be true in nonlinear analysis. The local regression slope, which focuses on change in expected Y as x changes, may be of greater interest than expected Y for a given x. Our approach provides a standardized version of the local regression slope. Moreover, in linear statistical analysis, the Galton–Pearson correlation coefficient, which was obtained as a standardized version of the regression slope, is useful as a universal scale-free measure of the degree of linear relationship. It is used to compare the results from different experiments using different scales when studying the same phenomena, and it facilitates communication between researchers in different fields as well as between statisticians and other scientists. We are suggesting that the concept of correlation similarly can play an important role in nonlinear curve estimation by providing a universal scale-free standardized version of the local regression slope.

EXAMPLE (A generalized linear model [GLM]) Consider the GLM of the form

$$Y = \alpha_1 + \alpha_2 g(X) + h(X)\varepsilon,$$

where X and ε are independent with variances σ_1^2 and σ_ε^2, and where $E(\varepsilon) = 0$. By appropriate choices of g and h as well as distributions of X and ε, the contour plots of the density $f(x, y)$ of (X, Y) will resemble the twisted pear in Figure 1. For instance, if ε has a standard normal distribution, then $(Y|x)$ has $N(\alpha_1 + \alpha_2 g(x), h^2(x))$ distribution, and if the link function $g(x)$ has an increasing derivative $g'(x)$ and if $h(x)$ is constant or decreasing, then the twisted pear model results for most choices of the distribution of X. If $h(x)$ is constant, the correlation coefficient is the appropriate measure of strength of association between $g(X)$ and Y. However, if we are interested in the strength of the relationship between X (the level of the symptom) and Y (the level of the disease), then the correlation curve $\rho(x)$ is the appropriate measure of the strength of the relationship even if $h(x)$ is constant in x. In our GLM with $g(x)$ differentiable, we have

$$\rho(x) = \frac{\alpha_2 \sigma_1 g'(x)}{\left[\{\alpha_2 \sigma_1 g'(x)\}^2 + \sigma_\varepsilon^2 h^2(x)\right]^{1/2}}.$$

If $g(x) = x^2/2$ and $h(x) = 1$, $x > 0$ (which corresponds to a twisted pear model), we find $\rho(x) = \alpha_2 \sigma_1 x / [\{\alpha_2 \sigma_1 x\}^2 + \sigma_\varepsilon^2]^{1/2}$. In this case, the strength of

the association starts out at zero when $x = 0$ and increases until x reaches its largest possible value. To obtain a comparison with the correlation coefficient ρ_{XY} between X and Y, we further assume that X has a uniform distribution on $[0, 1]$. In thise case $\rho(x) = \alpha_2 x/[\{\alpha_2 x\}^2 + 12\sigma_\varepsilon^2]^{1/2}$ and $\rho_{XY} = (1/2)\alpha_2/[(12/45)\alpha_2^2 + 12\sigma_\varepsilon^2]^{1/2}$. A particularly simple and instructive case is $\alpha_2 = 1$ and $\sigma_\varepsilon^2 = 11/180$. In this case $\rho = 0.5$ and $\rho(x) = x/[x^2 + (11/15)]^{1/2}$. Thus $\rho(x)$ increases from 0 to 0.76 as x increases from 0 to 1. On the other hand, the correlation coefficient between $Z = g(X) = X^2/2$ and Y is $\rho_{ZY} = \alpha_2/\sqrt{\alpha_2^2 + 45\sigma_\varepsilon^2}$ which in the case $\alpha_2 = 1, \sigma_\varepsilon^2 = 11/180$ equals $\rho_{ZY} = 1/\sqrt{3.75} = 0.52$.

3. Correlation curves from conditional correlation. If we apply the usual correlation formula to the conditional distribution of (X, Y) given $X = x$, we get the value zero. To see this recall that $\rho^2 \leq \eta^2$, where $\eta^2 = \text{var}(\mu(X))/\text{var}(Y)$ [Cramér (1946)]. In the conditional distribution $\mathbf{L}(X, Y|X = x)$, η^2 reduces to zero since $\text{var}(\mu(X)|X = x) = 0$, while (except in trivial cases) $\text{var}(Y|X = x) > 0$. If instead of conditioning on $X = x$, we condition on X in a neighborhood of x, the conditional versions of ρ^2 and η^2 will be positive but close to zero even when there is a strong relationship between X and Y.

One approach to overcoming the problem that the naive definition of local correlation in terms of conditional correlation gives the value zero is to consider the ratio of the two conditional correlations obtained by conditioning on two small neighborhoods $N_h(x_0) = [x_0 - \sigma_1 h, x_0 + \sigma_1 h]$ and $N_h(x_1) = [x_1 - \sigma_1 h, x_1 + \sigma_1 h]$, $x_0 \neq x_1$. Even though the conditional correlations tend to zero as $h \to 0$, the ratio

$$R_h(x_0, x_1) = \frac{\text{corr}(X, Y|X \in N_h(x_0))}{\text{corr}(X, Y|X \in N_h(x_1))}$$

will have a sensible limit. In fact, $\text{corr}((X, Y|X \in N_h(x_0))$ is to first order $h\sigma_1\beta(x_0)/\sqrt{3}\sigma(x_0)$. This result holds whether we use the Galton–Pearson ρ^2 or the Pearson $\eta^2 = \text{var}(\mu(X))/\text{var}(Y)$ to measure correlation.

Note that this approach is very similar to looking at the rate at which the conditional correlation tends to zero. That is, we could define

$$\xi(x) = \lim_{h \to 0} \frac{\text{corr}(X, Y|X \in N_h(x))}{h/\sqrt{3}} = \frac{\sigma_1\beta(x)}{\sigma(x)}$$

as a local measure of dependence which has the properties of correlation except it is not between -1 and 1 and it does not reduce to ρ in the normal case. Note that when $\sigma(x) > 0$,

$$\rho(x) = \text{sign}\{\xi(x)\}\left[1 + \xi^{-2}(x)\right]^{-1/2}.$$

Thus ρ(x) has an interpretation as the conditional correlation factor $\xi(x)$ mapped onto the interval $[-1, 1]$ in such a way that it coincides with the Galton–Pearson correlation coefficient in the normal model.

4. General correlation curves and their properties. In Section 2 we defined a correlation curve in terms of $\mu(x) = E(Y|x)$, $\sigma_1^2 = \text{var}(X)$ and $\sigma^2(x) = \text{var}(Y|x)$. However, just as there are many measures of location and scale, there are many correlation curves. These are obtained by replacing $\mu(x)$, σ_1^2 and $\sigma_1^2(x)$ by other measures of location and scale. This may be desirable since $\mu(x)$, σ_1^2 and $\sigma_1^2(x)$ do not always exist. Moreover, they are very sensitive to the tail behaviour of the distributions of X and $(Y|x)$. Thus, in our definition of the correlation curve $\rho(x)$, we replace $\mu(x)$ and $\sigma(x)$ by measures $m(x)$ and $\tau(x)$ of location and scale in the distribution $\mathbf{L}(Y|X = x)$ of Y given $X = x$. We assume only that $m(x)$ and $\tau(x)$ are location and scale parameters in the sense that they satisfy the usual equivariance and invariance properties. Similarly, we replace σ_1 by a scale parameter τ_1 for the distribution of X. Our basic assumption is that $m'(x) = (d/dx)m(x)$, τ_1 and $\tau(x)$ exist. Thus X has a continuous distribution while the distribution of Y may be discrete or continuous. Each time we specify $m(x)$, τ_1 and $\tau(x)$ we get a correlation curve whose formula is

$$\rho(x) = \rho_{XY}(x) = \frac{\tau_1 m'(x)}{\left[\{\tau_1 m'(x)\}^2 + \tau^2(x)\right]^{1/2}}. \tag{4.1}$$

It will sometimes be convenient to write (4.1) in the equivalent form

$$\rho(x) = \pm\left\{1 + \left[\tau_1 m'(x)/\tau(x)\right]^{-2}\right\}^{-1/2}, \tag{4.2}$$

where the sign $\pm$ is the same as the sign of $m'(x)$.

Rényi (1959) and Bell (1962) have discussed axioms that global correlation measures should satisfy. Local correlation measures should also satisfy such axioms. Under appropriate conditions, the correlation curves satisfy the following eight basic properties (axioms) of correlation. [In these axioms, the expression "for all x" means for all x in the support $S = \{x: 0 < F_X(x) < 1\}$ of the distribution $F_X(x)$ of X.]

AXIOM 1. *Standardization to the unit interval.* From (4.1), we observe

$$-1 \le \rho(x) \le 1 \quad \text{for all } x.$$

AXIOM 2. *Invariance and equivariance.* Each correlation curve $\rho(x)$ has invariance and equivariance properties that are direct analogs of those of the correlation coefficient ρ, that is, the following proposition holds.

PROPOSITION 1. *If $X^* = a + bX$ and $Y^* = c + dY$ with $bd \neq 0$, then, for all x^* in the support of the distribution of X^*, $\rho_{X^*Y^*}(x^*) = \text{sign}(bd)\rho_{XY}(x)$, where $x = (x^* - a)/b$.*

PROOF. In the proof we use an asterisk to indicate parameters computed for X^* and Y^*. Using the invariance and equivariance of the location and

scale parameters we find

$$\tau_1^* = |b|\tau_1, \qquad \tau^*(x^*) = |d|\tau(x)$$

and

$$\frac{d}{dx^*} m^*(x^*) = d\left\{ \frac{d}{dx^*} m\left(\frac{x^* - a}{b} \right) \right\} = d\{m'(x)/b\};$$

thus the result follows. □

AXIOM 3. *$\rho(x) = \rho$ for all x in the bivariate normal case.* This axiom is important since we want to connect our local correlation concept to a notion that people are familiar with. It gives the sense that $\rho(x)$ measures local strength of association in familiar correlation units. It turns out that in order to achieve $\rho(x) \equiv \rho$ in the bivariate normal $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ case, we need to add the condition that τ_1 and $\tau(x)$ are scale parameters of the "same type." We give an example where $\rho(x) \neq \rho$, and then explain the term "same type."

EXAMPLE. Let τ_1 be the interquartile range $\mathrm{IQR}(X) = F_X^{-1}(0.75) - F_X^{-1}(0.25)$ and let $\tau^2(x) = \mathrm{var}(Y|x)$. In the normal case all measures $m(x)$ of location for $(Y|x)$ equal $E(Y|x)$ and thus

$$\rho(x) = \frac{\tau_1 \sigma_2 \rho/\sigma_1}{\left[(\tau_1 \sigma_2 \rho/\sigma_1)^2 + \sigma_2^2(1 - \rho^2) \right]^{1/2}} = \frac{\rho \tau_1/\sigma_1}{\left[\rho^2(\tau_1/\sigma_1)^2 + 1 - \rho^2 \right]^{1/2}}. \tag{4.3}$$

Now $\rho(x) \neq \rho$ since $\tau_1/\sigma_1 = 1.348 \neq 1$.

What goes wrong in this example is that IQR and var are different "types" of scale parameters. We say that two *scale parameters* are of the *same type* if they are equal when applied to the same distribution.

PROPOSITION 2. *If τ_1 and $\tau(x)$ are the same type of scale parameters, and if (X, Y) is bivariate normal with Galton–Pearson correlation coefficient ρ, then $\rho(x) \equiv \rho$ for all x.*

PROOF. Since $(Y|x)$ is normal with variance $\sigma_2^2(1 - \rho^2)$, we can write $\tau(x)$ as $\tau(x) = \tau_2\sqrt{1 - \rho^2}$ where τ_2 is the scale parameter $\tau(x)$ applied to $\mathbf{L}(Y)$. Since X and Y both have normal distributions, invariance and equivariance yields $(\sigma_2/\sigma_1) = (\tau_2/\tau_1)$. The result now follows from (4.3). □

It follows that if $\tau_1^2 = \mathrm{var}(X) = \sigma_1^2$ and $\tau^2(x) = \mathrm{var}(Y|x)$, then $\rho(x) \equiv \rho$. Similarly, $\rho(x) \equiv \rho$ when $\tau_1 = \mathrm{IQR}(X)$ and $\tau(x) = \mathrm{IQR}(Y|x)$.

$\rho(x)$ as defined by (4.1) is called a correlation curve only when τ_1 and $\tau(x)$ are the same type of scale parameters.

AXIOM 4. *$\rho(x) \equiv 0$ for all x when X and Y are independent.* Since in this case $m'(x) \equiv 0$, the only condition needed for this result to hold is that $\tau(x) > 0$ for all x.

AXIOM 5. $\rho(x) \equiv \pm 1$ *for all* x *when* Y *is a function of* X. Suppose $Y = g(X)$, then, since $m(x)$ is a location parameter, $m(x) = g(x)$, and since $\tau(x)$ is a scale parameter for $Y|x$, then $\tau(x) = 0$. It follows that $\rho(x) = \tau_1 g'(x)/\{[\tau_1 g'(x)]^2 = \pm 1$ provided that τ_1 and $g'(x)$ exists and are nonzero. Moreover, $\rho(x) = 1$ when $g'(x) > 0$ and $\rho(x) = -1$ when $g'(x) < 0$. The case $g'(x) = 0$ is handled by defining $0/0 = 1$.

AXIOM 6. $\rho(x) = \pm 1$ *for almost all* x *implies that* Y *is a function of* x. Note that $\rho(x) = \pm 1$ implies that $\tau(x) = 0$. Thus the result holds provided $\tau(x) = 0$ for almost all x implies that $Y = g(x)$ for almost all x for some function g. When $\tau(x) = \text{var}(Y|x)$, this condition holds. However when $\tau(x) = \text{IQR}(Y|x)$, it does not hold.

AXIOM 7. $\rho(x) \geq 0$ *when* X *and* Y *are regression dependent.* The pair (X, Y) is *positively regression dependent* if $\Pr(Y \leq y|X = x)$ is nonincreasing in x [Lehmann (1966)]. Let $Y(x)$ denote a random variable with distribution $\Pr(Y \leq y|X = x)$. Then regression dependence means that for $x_1 < x_2$, $Y(x_1)$ is stochastically smaller than $Y(x_2)$. It follows that if the location parameter $m(x)$ for $Y(x)$ has a derivative $m'(x)$, then $m'(x) \geq 0$ and $\rho(x) \geq 0$.

AXIOM 8. $\rho(x)$ *increases with increasing regression dependence.* Let (X, Y_1) and (X, Y_2) be two pairs of random variables, let $Y_1(x)$ and $Y_2(x)$ denote random variables with distributions $\mathbf{L}(Y_1|x)$ and $\mathbf{L}(Y_2|x)$, and let $(m_1(x), \tau_1(x))$ and $(m_2(x), \tau_2(x))$ denote location and scale parameters of the same type for $Y_1(x)$ and $Y_2(x)$, respectively. The pair (X, Y_1) is said to be *more regression dependent* than the pair (X, Y_2) if $Y_1(x)/\tau_1(x)$ is stochastically more increasing than $Y_2(x)/\tau_2(x)$ in the sense that for each δ in some neighborhood $(0, \varepsilon)$ of zero, $\{Y_1(x + \delta) - Y_1(x - \delta)\}/\tau_1(x)$ is stochastically larger than $\{Y_2(x + \delta) - Y_2(x - \delta)\}/\tau_2(x)$. It follows that if $m_1(x)$ and $m_2(x)$ are location parameters such that the location of a difference is the difference of the locations and if $m'_1(x)$ and $m'_2(x)$ exist, then $\{m'_1(x)/\tau_1(x)\} \geq \{m'_2(x)/\tau_2(x)\}$. Thus, if we let $\rho_1(x)$ and $\rho_2(x)$ denote the correlation curves corresponding to (X, Y_1) and (X, Y_2), then it follows from (4.2) that $\rho_1(x) \geq \rho_2(x)$ for all x.

AXIOM 9. *Interchangeability of* X *and* Y. Note that $\rho_{XY}(\cdot) \neq \rho_{YX}(\cdot)$ except in very special cases. However, we get a local measure of correlation where X and Y are interchangeable by setting

$$\eta_{XY}(x, y) = [\text{sign}\{\rho_{XY}(x)\}]\{\rho_{XY}(x)\rho_{YX}(y)\}^{1/2}, \quad \text{if sign}\{\rho_{XY}(x)\} \text{ equals sign}\{\rho_{YX}(y)\}$$
$$= 0, \quad \text{otherwise.}$$

REMARK 4.1. A definition of "more regression dependent" based on comparing the Kolmogorov distance between $Y_1(x_1)$ and $Y_1(x_2)$ to the Kolmogorov

distance between $Y_2(x_1)$ and $Y_2(x_2)$ was considered by Bell and Doksum (1967).

REMARK 4.2. Suppose we consider using a local measure of the scale of X rather than τ_1 in our formula (4.1) for $\rho(x)$. Thus we could consider using $1/f(x)$, where f is the density of X, as a local measure of scale. In this case (4.3) shows that if we want to satisfy axiom 3, we need to standardize $1/f(x)$ by dividing it by $1/\phi((x - \mu_1)/\sigma_1)$, where ϕ is the $N(0, 1)$ density. This leads to replacing τ_1 in (4.1) by $\tau_1(x) = \phi((x - \mu_1)/\sigma_1)/f(x)$. The resulting more complicated correlation curve is not very different from $\rho(x)$ when $f(x)$ is nearly bell-shaped.

5. A data example. We will illustrate the correlation curve (2.2) using readings of plasma lipid concentrations taken on 371 patients in a heart study; see Scott, Gotto, Cole and Gorry (1978). For each patient we have the levels of cholesterol x and triglyceride y. This data set has also been analysed by Silverman [(1986), pages 81–83].

Local weighted linear regression will be used to estimate the functions $\mu(x) = E(Y|x)$, $\beta(x) = d\mu(x)/dx$ and $\sigma(x) = \{E[Y - \mu(x)]^2|x\}^{1/2}$. The methods for $\mu(x)$ and $\beta(x)$ used here are from Fan (1993). They are similar to methods considered by Stone (1977), Cleveland (1979) and Cleveland and Devlin (1988). The methods are as follows: Let $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ denote the Epanechnikov kernel. Consider 100 grid points along the x axis. Let x_0 denote any one of the grid points and let $y = a(x_0) + b(x_0)x$ be the weighted least squares line computed from the data $(x_1, Y_1), \ldots, (x_n, Y_n)$ with weights $w_1, \ldots, w_n$, where $w_i = K((x_i - x_0)/h)$, $h = s_1$, and s_1 is the sample standard deviation of $x_1, \ldots, x_n$. The estimates $\hat{\mu}(x_0)$ and $\hat{\beta}(x_0)$ of $\mu(x_0)$ and $\beta(x_0)$ are now $a(x_0) + b(x_0)x_0$ and $b(x_0)$, respectively. Similarly, to estimate $\sigma^2(x_0) = E([Y - \mu(x_0)]^2|x_0)$, let $y = c(x_0) + d(x_0)x$ be the weighted least squares line computed from the data $(x_1, \tilde{\varepsilon}_1^2), \ldots, (x_n, \tilde{\varepsilon}_n^2)$ with weights $w_1, \ldots, w_n$ as before, where $\tilde{\varepsilon}_i = [Y_i - \hat{\mu}(X_i)]$ is the ith residual, $i = 1, \ldots, n$. The estimate $\hat{\sigma}^2(x_0)$ of $\sigma^2(x_0)$ is now $c(x_0) + d(x_0)x_0$ and the estimate of the local correlation $\rho(x_0)$ at x_0 is $\hat{\rho}(x_0) = s_1\hat{\beta}(x_0)/\{s_1^2\hat{\beta}^2(x_0) + \hat{\sigma}^2(x_0)\}^{1/2}$. Finally, the above procedures are repeated for the 100 grid points and the curves $\hat{\mu}(x)$, $\hat{\sigma}(x)$, $\hat{\beta}(x)$ and $\hat{\rho}(x)$ are completed by using standard software to "connect the dots". The curves are plotted only for the central 90% of the x values due to the large uncertainly, as expressed by the mean squared error, in the tails.

Figure 3 gives the cholesterol and triglyceride data together with the mean curve $\hat{\mu}(x_0)$. Figure 4 gives the estimated standard deviation curve $\hat{\sigma}(x)$. Both $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are increasing as x increases from small to moderate levels and both level off as x approaches higher cholesterol levels thereby illustrating both nonlinearity and heteroscedasticity.

Figure 5 gives the slope curve $\hat{\beta}(x)$ and shows how the estimated local regression coefficient drops from about one for the low cholesterol group to about zero for $x = 245$ and increases to about 0.8 for $x = 285$. Figure 6

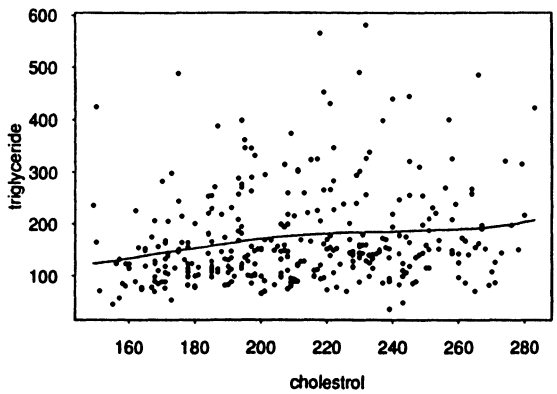

FIG. 3. *Scatter plot of plasma lipid concentrations with estimated local mean regression.*

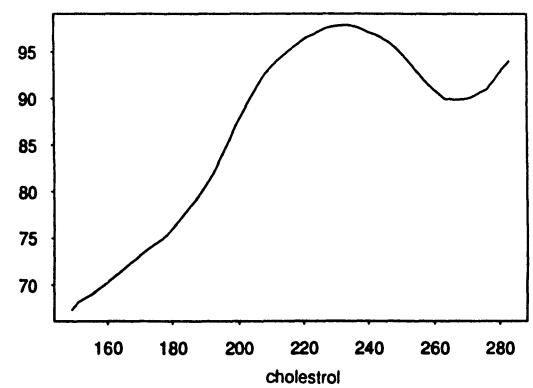

FIG. 4. *Estimated local standard deviation.*

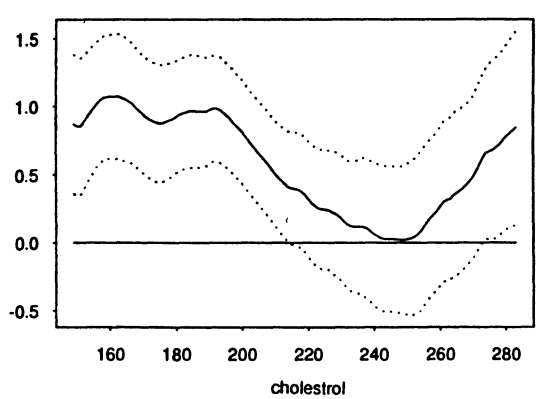

FIG. 5. *Estimated local regression slope with 90% pointwise confidence band.*

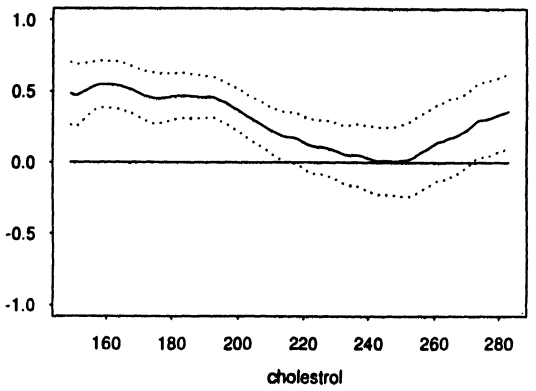

FIG. 6. *Estimated local correlation with 90% pointwise confidence band.*

combines $\hat{\beta}(x)$ and $\hat{\sigma}(x)$ into a measure of the local strength of the relationship between X and Y in terms of correlation units on the interval $[-1, 1]$. The estimated local correlation starts out about 0.5 for the low cholesterol group, drops off to a value close to zero around the cholesterol level 245 and reaches the value about 0.35 for the moderately high cholesterol level 285. The dotted line gives an approximate 90% pointwise confidence band for $\rho(x)$ obtained by using the δ method and weighted least squares standard error software appropriate for the fixed design points case where $x_1, x_2, \ldots, x_n$ are regarded as nonrandom. Finally, Figure 7 gives approximate 90% Bonferroni simultaneous confidence intervals at the 10th, 20th, ..., 90th percentiles $\hat{x}_{.1}, \hat{x}_{.2}, \ldots, \hat{x}_{.9}$ of $x_1, x_2, \ldots, x_n$. Since only the first four of these intervals are

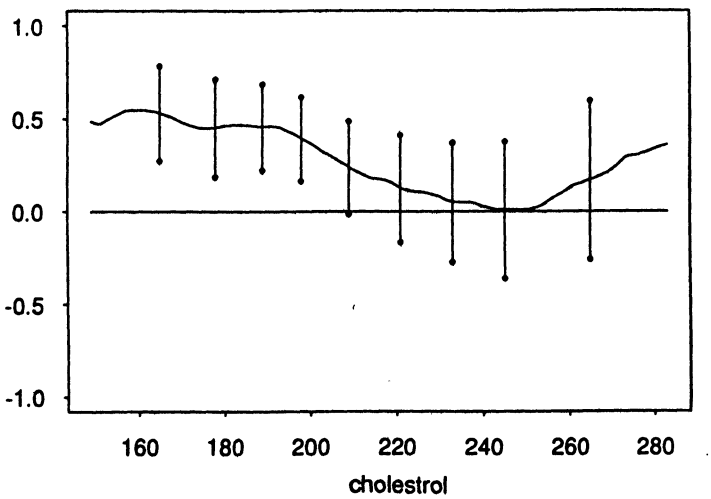

FIG. 7. *Estimated local correlation with 90% simultaneous confidence intervals.*

above the horizontal $\rho(x) = 0$ axis, we conclude that the local correlation between cholesterol and triglyceride is significantly positive at the lower cholesterol level $\hat{x}_{0.10} = 165$, $\hat{x}_{0.20} = 178$, $x_{0.3} = 189$ and $\hat{x}_{0.4} = 189$ while there is no significant association at the higher cholesterol levels.

We consulted a medical expert (Jon Bremer) on cholesterol and fatty substances who said that measurements on cholesterol and triglyceride are known to be positively correlated but that it is thought that this positive correlation does not include individuals with high values of cholesterol. Our results give a statistical confirmation of this statement: At cholesterol levels $x = 165$, 178, 189 and 198, the estimated correlations are 0.530, 0.453, 0.456 and 0.392, respectively. They are significantly different from zero at level $\alpha = 0.10$. At cholesterol levels $x = 209$, 221, 233, 245 and 265 the estimated correlations are 0.239, 0.127, 0.052, 0.011 and 0.171, respectively. They are not significant at level of significance $\alpha = 0.10$. High values of triglyceride is not considered to be a risk factor for heart disease to the same extent as high values of cholesterol are.

The significance claims made in this section are based on approximations whose closeness to the actual probabilities will be the subject of a future study. In particular, it is conjectured that closer approximations can be obtained by using variance stabilizing transformations of $\hat{\rho}(x)$.

Acknowledgments. Jack Block, Per Gjerde, Steve Blyth and Xiao-Li Meng contributed to the ideas in this paper. Peter Bickel suggested the relative local correlation interpretation of Section 3 and Jeff Wu suggested the interpretation in terms of conditional correlation rates. We are grateful to David W. Scott for making the cholesterol data available and to Hongyu Zhao for computing the graphs in Section 5.

REFERENCES

ANSCOMBE, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 1–36. Univ. California Press, Berkeley.

BELL C. B. (1962). Mutual information and maximal correlation as measures of dependence. *Ann. Math. Statist.* **33** 587–595.

BELL C. B. and DOKSUM, K. A. (1967). Distribution-free tests of independence. *Ann. Math. Statist.* **38** 429–446.

BICKEL, P. J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Ann. Statist.* **6** 266–291.

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Oakland, CA.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–598.

CARROLL, R. J. and RUPPERT, D. (1982). Robust estimation in heteroscedastic linear models. *Ann. Statist.* **10** 429–441.

CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

CLEVELAND, W. S. and DEVLIN, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 596–610.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216.

FISHER, J. (1959). The twisted pear and the prediction of behaviour. *Journal of Consulting Psychology* **23** 400–405.

GALTON, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proc. Roy. Soc. London* **45** 135–145.

KARPOFF, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* **22** 109–126.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

RÉNYI, A. (1959). On measures of dependence. *Acta. Math. Acad. Sci. Hungar.* **10** 441–451.

SCOTT, D. W., GOTTO, A. M., COLE, J. S. and GORRY, G. A. (1978). Plasma lipids as collateral risk factors in coronary heart disease—a study of 371 males with chest pain. *Journal of Chronic Diseases* **31** 337–345.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press, Cambridge, MA.

STIGLER, S. M. (1989). Galton's account of the invention of correlation. *Statist. Sci.* **4** 73–86.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OSLO
P.O. BOX 1053
BLINDERN 0316, OSLO 3
NORWAY

DEPARTMENT OF STATISTICS
STATISTICAL LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720