Correlation Curves: Measures of Association as Functions of Covariate Values

By

Steinar Bjerve[1]
Department of Mathematics, University of Oslo

Kjell Doksum [2,3]
Department of Statistics, University of California, Berkeley
and Harvard University

Technical Report No. 280
February 1991

[1]Partially supported by the Johan and Mimi Wessmann foundation
[2]Work partially supported by grant from the Norwegian research foundation for science and the humanities
[3]Research partially supported by NSF Grant DMS-8901603

Department of Statistics
University of California
Berkeley, California 94720

Correlation Curves: Measures of association as functions of covariate values.

Steinar Bjerve[1]
Department of Mathematics, University of Oslo

Kjell Doksum[2,3]
Department of Statistics, University of California, Berkeley
and Harvard University

Abstract.
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 non-parametric 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 $\mathrm{var}(Y \mid X = x)$ and then replacing the regression slope β by a nonparametric regression slope $\beta(x)$. We show that the dependence functions share most of the properties of the correlation coefficient and that 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. Consistency and asymptotic normality results of empirical versions of correlation curves are established. The last two sections present a bootstrap confidence procedure and include a data example and a simulation example.

1. Introduction. For bivariate experiments where the contour plots (plots of (x, y) where the joint density $f(x, y)$ is constant) 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, J. Fisher (1959) reported on studies in psychology and other fields where the association

[1] Partially supported by the Johan and Mimi Wessmann foundation.
[2] Work partially supported by grant from the Norwegian research foundation for science and the humanities.
[3] Research partially supported by NSF Grant DMS-8901603

between the response variable Y and covariate X is strong for large values of X = x, but the association is weak or non-existent 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 associated contour plot and calls it a twisted pear. See Figure 1 which gives a representation of J. Fisher's contour plot.

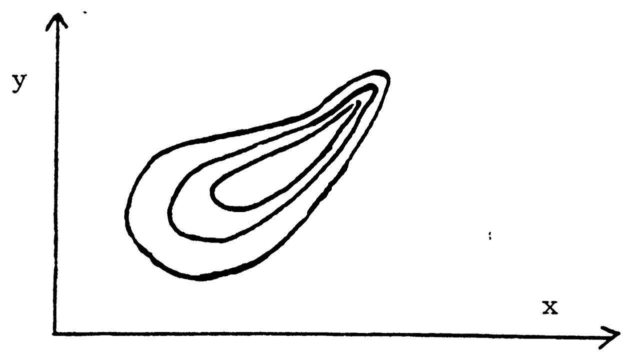

Figure 1. A typical twisted pear contour plot. x is level of symptom and y is level of disease.

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.

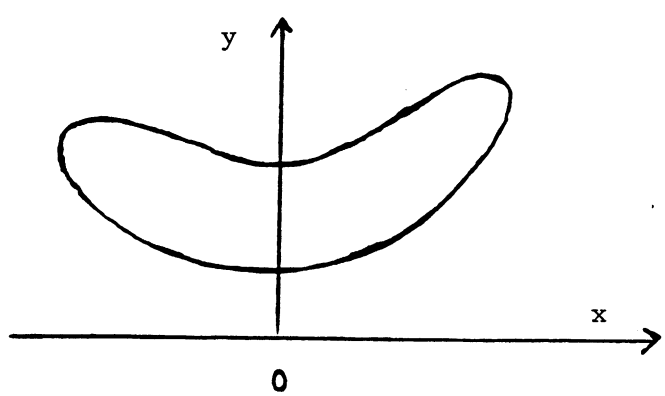

Figure 2. A contour adaption of Karpoff's Figure 1. x is change in price and y is level of volume.

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 (1968), Bickel (1978), Carroll and Ruppert (1982, 1988), Breiman and Friedman (1985), and Silverman (1988). 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 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 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, p.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, p.36)) decomposition

$$\begin{aligned} \sigma_2^2 &= \mathrm{var}(Y) = \text{variance explained} + \text{residual variance} \\ &= (\sigma_1 \beta)^2 + \sigma^2(x) \quad \text{(normal case)} \end{aligned}$$

where $\sigma^2(x) = \mathrm{var}(Y|x) = \mathrm{var}(Y|X = x)$ is the variance of Y given $X = x$. (In the normal case, $\sigma^2(x) = \sigma_2^2(1 - \rho^2)$ does not depend on x, but in non-normal cases it typically does). We can now write

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

In this representation we see how the correlation coefficient ρ is determined by the regression slope β and the residual variance $\sigma^2(x)$. The representation also suggests that in the non-normal 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

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

where $\beta(x) = \mu'(x)$ is the slope of the non-parametric regression $\mu(x) = E(Y|x) = E(Y|X = x)$; and $\sigma_1^2 = \text{var}(X)$ and $\sigma^2(x) = \text{var}(Y|x)$ as before. This correlation curve concept makes sense only when X is a continuous random variable (in fact, $\mu(x) = E(Y|x)$ must be differentiable). The distribution of Y can be discrete or continuous. We have assumed that σ_1^2 and $\sigma^2(x)$ exists.

$\rho(x)$ 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 towards one.

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 $N(\alpha_2 g(x), h^2(x))$ 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

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

If $g(x) = x^2/2$ and $h(x) = 1$, $x > 0$, (which corresponds to a twisted pair 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 this 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 zero to 0.76 as x increases from zero to one. On the other hand, the correlation coefficient between $Z = g(X) = X^2/2$ and Y is $\rho_{ZY} = \dfrac{\alpha_2}{\sqrt{\alpha_2^2 + 45\sigma_\varepsilon^2}}$ which in the case $\alpha_2 = 1$, $\sigma_\varepsilon^2 = 11/180$ equals $\rho_{ZY} = 1/\sqrt{3.75} = 0.52$

3. General correlation curves and their properties. In Section 2 we defined a correlation curve in terms of $\mu(x) = E(Y|x)$, $\sigma_1^2 = \mathrm{var}(X)$, and $\sigma^2(x) = \mathrm{var}(Y|x)$. 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 $\sigma^2(x)$ by other measures of location and scale. This may be desirable since $\mu(x)$, $\sigma_1^2(x)$ 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 $L(Y|X = x)$ 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 $m'(x) = \dfrac{d}{dx} m(x)$, 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

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

It will sometimes be convenient to write (3.1) in the equivalent form

(3.2) $$\rho(x) = \pm\{1 + [\tau_1 m'(x)/\tau(x)]^{-2}\}^{-1/2}$$

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

(i) **Standardization to the unit interval.**

From (3.1), we observe

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

(ii) **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

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 "*" to indicate parameters computed for X^* and Y^*. Using the invariance and equivariance of the location and scale parameters we find $\tau_1^* = |b|\tau_1$, $\tau^*(x^*) = |d|\tau(x)$ 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.

(iii) $\rho(x) = \rho$ **for all x in the bivariate normal case.**

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 $IQR(X) = F_X^{-1}(.75) - F_X^{-1}(.25)$ and let $\tau^2(x) = \text{var}(Y|x)$. In the normal case all measures $m(x)$ of location for $(Y|x)$ equal $E(Y|x)$ and thus

$$(3.3)\quad \rho(x) = \frac{\tau_1 \sigma_2 \rho/\sigma_1}{[(\tau_1\sigma_2\rho/\sigma_1)^2 + \sigma_2^2(1-\rho^2)]^{1/2}} = \frac{\rho\,\tau_1/\sigma_1}{[\rho^2(\tau_1/\sigma_1)^2 + 1 - \rho^2]^{1/2}}$$

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

What goes wrong in this example is that $\tau_1 = IQR(X)$ and $\sigma_1 = \{\text{var}(X)\}^{1/2}$ 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 (3.3).

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 (3.1) is called a correlation curve only when τ_1 and $\tau(x)$ are the same type of scale parameters.

(iv) $\rho(x) \equiv 0$ **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.

(v) $\rho(x) \equiv \pm 1$ **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 non-zero. 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 $0/0 = 1$.

(vi) $\rho(x) = \pm 1$ **for almost all** x **implies that** Y **is a function** 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 $\tau(x) = \mathrm{var}(Y|x)$, this condition holds. However when $\tau(x) = \mathrm{IQR}(Y|x)$, it does not hold.

(vii) $\rho(x) \geq 0$ **when** X **and** Y **are regression dependent.**

The pair (X, Y) is **positively regression dependent** if $\Pr(Y \leq y | X = x)$ is non-increasing 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 $\rho(x) \geq 0$.

(viii) $\rho(x)$ **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 (o, ε) 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 (3.2) that $\rho_1(x) \geq \rho_2(x)$ for all x.

(ix) **Interchangeability of X and Y.**

Note that $\rho_{XY}(\cdot) \neq \rho_{YX}(\cdot)$ except in very special cases. If we want a local measure of correlation where X and Y are interchangeable, we can proceed as follows: Let F_X and F_Y denote the distributions of X and Y respectively. Set $x = F_X^{-1}(p)$ and $y = F_Y^{-1}(p)$; thus x and y are both pth quantiles. Define

$$\begin{aligned} \eta_{XY}(x,y) &= [\operatorname{sign}\{\rho_{XY}(x)\}]\{\rho_{XY}(x)\rho_{YX}(y)\}^{1/2} \text{ if } \operatorname{sign}\{\rho_{XY}(x)\} = \operatorname{sign}\{\rho_{YX}(y)\} \\ &= 0 \text{ otherwise}. \end{aligned}$$

We assume that all the conditions of this section are satisfied for (Y,X) as well as (X,Y). Now it is clear that $\eta_{XY}(x,y) = \eta_{YX}(y,x)$. In this paper we prefer the asymmetric situation where how strongly the response variable Y is associated with the covariate variable X locally at $X = x$ is of interest. We will not consider $\eta_{XY}(x,y)$ again in this paper.

Remark 3.1. The joint distribution of $Y(x-\delta)$ and $Y(x+\delta)$, which appear in (viii) above, can be obtained as follows: Suppose $Y(X) = g(X) + h(X)\varepsilon$, where X and ε are independent, then the distribution of $(Y(x-\delta), Y(x+\delta))$ is the distribution of $(g(x-\delta) + h(x-\delta)\varepsilon,\ g(x+\delta) + h(x+\delta)\varepsilon)$. In general, assume we can write $Y = a(X,\varepsilon)$ for some function $a(\cdot,\cdot)$ and repeat the above idea.

Remark 3.2. 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).

4. Smooth correlation curves. Since they depend on the derivative $m'(x)$, the correlation curves $\rho(x)$ considered in Sections 2 and 3 can be erratic and difficult to estimate. Thus, prior to introducing estimates of $\rho(x)$, we pre-smooth $\rho(x)$ by considering the strength of association between X and Y for X in an **interval** containing x rather than for X exactly equal to x. We choose this interval so that there is an equal amount of mass on either side of x. More precisely, we set $x = x_p$ = pth quantile of F_X, where $p = F_X(x)$ and F_X is the distribution function of X. Now our interval is $[x_{p-t}, x_{p+t}]$, where $x_{p-t} = F_X^{-1}(p-t)$ and $x_{p+t} = F_X^{-1}(p+t)$ are the $(p-t)$th and $(p+t)$th quantiles of F_X. Note that this interval has mass t on either side of $x = x_p$. Now rather than using the derivative $\beta(x) = m'(x)$ of the location parameter $m(X)$ for $L(Y \mid X = x)$, we consider the interval slope

$$\beta_t(x) = \frac{m(x_{p+t}) - m(x_{p-t})}{x_{p+t} - x_{p-t}}, \quad x = x_p.$$

In direct analogy with (2.2) and (3.1), we define the smooth correlation curve as

$$\rho_t(x) = \frac{\tau_1 \beta_t(x)}{[\{\tau_1 \beta_t(x)\}^2 + \tau^2(x)]^{1/2}}, \quad x = x_p \tag{4.1}$$

with the convention that $\rho_t(x) = 0$ if both the numerator and denominator equals zero.

Clearly, if $m'(x)$ exists, $\rho_t(x) \to \rho(x)$ as $t \to 0$. Even though $\rho_t(x)$ is an approximation to $\rho(x)$, we prefer to think of it as a correlation curve in its own right: Since $\rho_t(x)$ combines the slope over the interval and the residual variance in accordance with formula (2.2) and Galton's principle of correlation as regression slope on standardized scales, we conclude that $\rho_t(x)$ measures the strength of the association between Y and X for X in the interval $[x_{p-t}, x_{p+t}]$.

Moreover, $\rho_t(x)$ also satisfies the eight basic axioms (i),...,(viii) of correlation curves given in Section 3 and it can be turned into a measure with X and Y interchangeable as in (ix), Section 3.

Remark 4.1. Kowalczyk (1977) and Kowalczyk and Pleszczyńska (1977) considered the functions

$$\mu^{+}_{Y,X}(p) = \frac{E(Y \mid X > x_p) - E(Y)}{E(Y \mid Y > y_p) - E(Y)}, \quad \mu^{-}_{Y,X}(p) = \frac{E(Y \mid X > x_p) - E(Y)}{E(Y) - E(Y \mid Y < y_{1-p})}$$

and defined the monotonic function

$$\begin{aligned} \mu_{Y,X}(p) &= \mu^{+}_{Y,X}(p) \text{ if } \mu^{+}_{Y,X}(p) \geq 0 \\ &= \mu^{-}_{Y,X}(p) \text{ if } \mu^{+}_{Y,X}(p) \leq 0 . \end{aligned}$$

This function measures dependence to the right of x_p. Our function $\rho_t(x_p)$ measures strength of association in a neighborhood of x_p.

5. Estimation of smooth correlation curves. Consistency.

5(a). The general setup. We consider two types of sampling experiments with corresponding models:

I. Random Covariates. In this case, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote the random outcome of an experiment where, for the ith subject in a random sample of size n, Y_i denotes the response and X_i denotes the covariate value. The pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ are assumed to be independent and identically distributed. We assume that the distribution F_X of X_i is continuous while the distribution F_Y of Y_i may be discrete or continuous (or a mixture).

II. Fixed covariates. In this case, the covariate value x_i is fixed and Y_i denotes the random response of a subject selected at random from a population of subjects with covariates value x_i. The distribution function F_{Y_i} of Y_i depends on x_i. We assume that $Y_1, \ldots, Y_n$ are independent and that the x's are distinct. Since the x's are nonrandom we can without loss of generality take $x_1 < \cdots < x_n$.

By a result of Bhattacharya (1974), there is a strong connection between models I and II: In model I, let $X_{(1)} < \cdots < X_{(n)}$ denote the X-order statistics and let $Y_{[1]}, \ldots, Y_{[n]}$ denote the Y-statistics induced by the X-order, that is $Y_{[i]} = Y_{k_i}$, where k_i denotes the subscript on the X with rank i among $X_1, \ldots, X_n$ ($X_{k_i} = X_{(i)}$). These $Y_{[\ ]}$'s are called concomitants. Conditionally on $X_{(1)} = x_1, \ldots, x_{(n)} = x_n$, $Y_{[1]}, \ldots, Y_{[n]}$ are independent, and the conditional distribution of $Y_{[i]}$ depends only on x_i. Thus model II is a conditional version of model I. Conversely, if in model II we replace $x_1 < \cdots < x_n$ by the order statistics $X_{(1)}, \ldots, X_{(n)}$ of a random sample from F_X, we let $(r_1, \ldots, r_n)$ denote a random permutation of $(1, \ldots, n)$, and we set $X_i = X_{(r_i)}$, $Y_i' = Y_{r_i}$, then $(X_1, Y_1'), \ldots, (X_n, Y_n')$ are independent identically distributed random pairs. Thus model II is a randomized (anti-conditional) version of model I.

In Sections 2 and 3 we formulated correlation curves in terms of the random covariate model. To formulate the correlation curve in terms of the fixed covariate model, we rewrite this model as

(5.1) $$Y_i = m(x_i) + \tau(x_i)\varepsilon_i, \quad i = 1, \ldots, n; \quad x_1 < \cdots < x_n$$

where $m(x_i)$ and $\tau(x_i)$ are location and scale parameters for the distribution of Y_i, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent with location and scale parameters zero and one respectively. With this notation the smooth correlation curve is

$$\rho_n(x) = \frac{\tau_{1n}\beta_t(x)}{[\{\tau_{1n}\beta_t(x)\}^2 + \tau^2(x)]^{1/2}}, \quad x = x_p$$

as before except τ_{1n} is now a known scale value computed from the given x's. Moreover, in $\beta_t(x) = \beta_t(x_p) = \{m(x_{p+t}) - m(x_{p-t})\}/(x_{p+t} - x_{p-t})$, we take $p = F_n(x)$, $x_{p-t} = F_n^{-1}(p-t)$ and $x_{p+t} = F_n^{-1}(p+t)$ where $F_n(x) = n^{-1}[\#x_i \le x]$ and $F_n^{-1}(u) = \min\{x : F_n(x) \ge u\}$. $m(x)$ and $\tau(x)$ are unknown functions defined on $[x_1, x_n]$. Here and throughout, the dependence of x_i, x_{p-t}, etc. on n is suppressed.

If $\hat{\rho}_n(x)$ is an estimate of $\rho_n(x)$ obtained by replacing $m(x_{p-t})$, $m(x_{p+t})$ and $\tau^2(x)$ by consistent estimates, then it follows from (5.3) below that $\hat{\rho}(x)$ is consistent in the sense that $|\hat{\rho}_n(x) - \rho_n(x)|$ tends to zero in probability. See Härdle (1990) for a recent survey of estimates of $m(x)$, and see Müller and Stadtmüller (1987), Hall and Carroll

(1989), and Hall, Kay and Titterington (1990) for the consistent estimation of $\tau^2(x)$ as well as $m(x)$ under certain regularity conditions. Here we use Kolmogorov's inequality to give an elementary argument for uniform consistency of nearest neighbor estimates of $\rho^2(x)$ that require a minimum of smoothness conditions.

5(b). Nearest Neighbor Estimation. For the rest of this section we will consider the correlation curve $\rho(x)$ based on means and variances. Thus we let τ_{1n} be the standard deviation σ_{1n} of $x_1 < \cdots < x_n$, we take $m(x_i) = \mu(x_i) = E(Y_i)$ and $\tau^2(x_i) = \sigma^2(x_i) = \mathrm{var}(Y_i)$. The basic model is

$$Y = \mu(x) + \sigma(x)\varepsilon$$

where $\mu(x)$ and $\sigma^2(x)$ are the unknown mean and variance functions and ε is a random variable with mean zero and variance one. We will also assume the existence of the residual variance function

$$\sigma_2^2(x) = \mathrm{Var}(Y - \mu(x))^2 = \sigma^4(x)[E(\varepsilon^4) - 1].$$

The data is generated according to the model

$$Y_i = \mu(x_i) + \sigma(x_i)\varepsilon_i, \quad i = 1, \ldots, n; \quad x_1 < \cdots < x_n \tag{5.2}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent with mean and variance zero and one, respectively. In the asymptotics each $x_i = x_{in}$ depends on n, the second subscript on x_i having been omitted. We assume that $x_1, \ldots, x_n$ is a *regular sequence of covariate values* in the sense that $\{x_i, i = 1, \ldots, n\}$ is dense on some interval $[a, b]$ with $a < b$ (possibly infinite), $F(x) = \lim F_n(x)$ exists for each $x \in [a, b]$, and $F(x)$ is a continuous and strictly increasing distribution function on $[a, b]$.

Let I_{p-t}, I_p and I_{p+t} be the sets of indices on the k values of $x_1, \ldots, x_n$ closest to x_{p-t}, x_p and x_{p+t}, respectively. (In case of ties, choose the smaller index). Define

$$\hat\mu(x_{p-t}) = k^{-1} \sum_{i \in I_{p-t}} Y_i, \quad \hat\mu(x_{p+t}) = k^{-1} \sum_{i \in I_{p+t}} Y_i, \quad \hat\sigma^2(x_p) = k^{-1} \sum_{i \in I_p} [Y_i - \hat\mu(x_p)]^2.$$

Let $\hat\rho_n(x)$ denote the estimate of $\rho_n(x)$ obtained by replacing $\mu(x_{p-t})$, $\mu(x_{p+t})$ and $\sigma^2(x_p)$ by $\hat\mu(x_{p-t})$, $\hat\mu(x_{p+t})$ and $\hat\sigma^2(x_p)$.

In the following $k = k_n$ is a function of n tending to infinity as $n \to \infty$.

Theorem 5.1. Suppose that $x_1, \ldots, x_n$ is a regular sequence of covariate values and suppose that $\max_{1 \le i \le n-k} |x_{i+k} - x_i| \to 0$ as $n \to \infty$. Assume that $k^{-2} \sum_{i=1}^{n} \sigma^2(x_i) \to 0$ and $k^{-2} \sum_{i=1}^{n} \sigma_2^2(x_i) \to 0$ as $n \to \infty$. Assume that $\inf_{x_t \le x \le x_{1-t}} \{\sigma^2(x)\} > 0$ and that $\limsup_{n \to \infty} \sigma_{1n}^2 < \infty$. Then for each $\delta > 0$,

$$P\{\sup_{x_t \le x \le x_{1-t}} |\hat{\rho}_n^2(x) - \rho_n^2(x)| > \delta\} \to 0 \text{ as } n \to \infty.$$

Proof: Let

$$A_n(x) = [\mu(x_{p+t}) - \mu(x_{p-t})]^2 \sigma_{1n}^2, \quad B_n(x) = (x_{p+t} - x_{p-t})^2 \sigma^2(x)$$
$$\hat{A}_n(x) = [\hat{\mu}(x_{p+t}) - \hat{\mu}(x_{p-t})]^2 \sigma_{1n}^2, \quad \hat{B}_n(x) = (x_{p+t} - x_{p-t})^2 \hat{\sigma}^2(x).$$

Then, using a little algebra, we can write

(5.3) $\hat{\rho}_n^2(x) - \rho_n^2(x) = \hat{A}_n(x)[\hat{A}_n(x) + \hat{B}_n(x)]^{-1} - A_n(x)[A_n(x) + B_n(x)]^{-1}$

$= \{[1 - \rho_n^2(x)][\hat{A}_n(x) - A_n(x)] + \rho_n^2(x)[\hat{B}(x) - B(x)]\}[\hat{A}_n(x) + \hat{B}_n(x)]^{-1}.$

It follows that $\hat{\rho}_n^2(x) - \rho_n^2(x)$ converges uniformly in probability to zero provided $\hat{\mu}(x_{p-t}) - \mu(x_{p-t})$, $\hat{\mu}(x_{p+t}) - \hat{\mu}(x_{p+t})$ and $\hat{\sigma}^2(x) - \sigma^2(x)$ converge uniformly in probability to zero and provided $\inf_{x_t \le x \le x_{1-t}} \{B_n(x)\}$ is bounded away from zero as $n \to \infty$. By assumption, $\sigma^2(x)$ is bounded away from zero. Since $F(x) = \lim F_n(x)$ is continuous, then $x_{p+t} - x_{p-t} = F_n^{-1}(p+t) - F_n^{-1}(p-t)$ is bounded away from zero as $n \to \infty$. Thus $\inf_{x_t \le x \le x_{1-t}} \{B_n(x)\}$ is bounded away from zero. Next we show that $\hat{\mu}(x_{p-t})$, $\hat{\mu}(x_{p+t})$ and $\hat{\sigma}^2(x_p)$ converge uniformly in probability. We start with $\hat{\mu}(x) = k^{-1} \sum_{i \in I_k(x)} Y_i$, where $I_k(x)$ set of indices on the k values of $x_1, \ldots, x_n$ closest to x. The deviation $\hat{\mu}(x) - \mu(x)$ has the random part $\hat{\mu}(x) - E(\hat{\mu}(x))$ and the deterministic part $E(\hat{\mu}(x)) - \mu(x)$. The random part is taken care of by the following Lemma.

Lemma 5.1. In the fixed covariate model, assuming only that $\sigma^2(x_i)$ exists for $i = 1, \ldots, n$,

(a) $$P(\sup_{-\infty<x<\infty} |\hat{\mu}(x) - E(\hat{\mu}(x))| > \delta) \le \frac{4}{(k\delta)^2} \sum_{i=1}^{n} \sigma^2(x_i)$$

(b) $$\sup_{-\infty<x<\infty} |\hat{\mu}(x) - E(\hat{\mu}(x))| \text{ tends to zero in probability as } n \to \infty$$

provided $\sigma^2(x)$ is bounded above and $(n/k^2) \to 0$ as $n \to \infty$.

Proof. (The proof given here leads to an upper bound sharper than the bound given in Bjerve, Doksum and Yandell (1985)).

$\hat{\mu}(x) - E(\hat{\mu}(x)) = k^{-1} \sum_{i \in I_k(x)} [Y_i - E(Y_i)]$ is a step function which is constant on each of the intervals

(5.4) $J_i = ((x_i + x_{i+k})/2, (x_{i+k} + x_{i+k+1})/2], \quad i = 0, \ldots, n-k, \quad x_0 = -\infty, x_{n+1} = \infty.$

In fact, if we set $W_j = k^{-1}[Y_j - E(Y_j)]$, then we can write

$$\hat{\mu}(x) - E(\hat{\mu}(x)) = \sum_{j=i+1}^{i+k} W_j, \quad x \in J_i, \quad i = 0, \ldots, n-k.$$

Set $S_i = \sum_{j=1}^{i} W_j$, $i = 1, \ldots, n$; $S_0 = 0$, then

$$\sup_{-\infty<x<\infty} |\hat{\mu}(x) - E(\hat{\mu}(x))| = \max_{0\le i\le n-k} |\sum_{j=i+1}^{i+k} W_j|$$

$$= \max_{0\le i\le n-k} |S_{i+k} - S_i| \le \max_{0\le i\le n-k} \{|S_{i+k}| + |S_i|\} \le 2 \max_{1\le i\le n} |S_i|.$$

Since $\mathrm{var}(W_i) = k^{-1}\sigma^2(x_i)$, result (a) follows from Kolmogorov's inequality. Result (b) follows from (a).

Returning to the proof of Theorem 5.1 we have now shown that $\hat{\mu}(x_{p-t}) - E(\hat{\mu}(x_{p-t}))$ and $\hat{\mu}(x_{p+t}) - E(\hat{\mu}(x_{p+t}))$ converge to zero in probability uniformly over the respective sets $\{x_{p-t} : t \le p < 1\}$ and $\{x_{p+t} : 0 < p \le 1-t\}$. Next we turn to the deterministic part represented by $E(\hat{\mu}(x)) - \mu(x)$.

Lemma 5.2. If $\mu(x)$ satisfies the Lipschitz condition

$$(5.5) \qquad |\mu(x) - \mu(y)| \le c\,|x-y| \text{ for all } x, y \in [x_1, x_n], \text{ some } c > 0,$$

then

(a) $$\sup_{x_1\le x\le x_n} |E(\hat{\mu}(x)) - \mu(x)| \le c \max_{1\le i\le n-k} |x_{i+k} - x_i|.$$

(b) $$\sup_{x_1\le x\le x_n} |E(\hat{\mu}(x)) - \mu(x)| \text{ tends to zero as } n \to \infty$$

if x_i can be written as $x_i = F^{-1}((i-0.5)/n) + o(n^{-1})$ where F^{-1} satisfies the Lipschitz condition (5.5), and if $(k/n) \to 0$ as $n \to \infty$.

Proof. Note that, using (5.4), $|E(\hat{\mu}(x)) - \mu(x)| = |k^{-1} \sum_{j\in I_k(x)} \mu(x_j) - \mu(x)|$ $\le k^{-1} \sum_{j\in I_k(x)} |\mu(x_j) - \mu(x)| \le ck^{-1} \sum_{j\in I_k(x)} |x_j - x|$. Let J_i be as defined in (5.4) except $x_0 = x_1$ and $x_{n+1} = x_n$. For $x \in J_i = ((x_i + x_{i+k})/2, \ (x_{i+1} + x_{i+k+1})/2]$, $\sum_{j\in I_k(x)} |x_j - x| = \sum_{j=i+1}^{i+k} |x_j - x|$. Next note that for $x \in J_i$, $i = 1, \ldots, n-k-1$, $\max\{|x_j - x|;\ i+1 \le j \le i+k\}$ is bounded above by the larger of $x_{i+k} - 0.5(x_i + x_{i+k}) = 0.5(x_{i+k} - x_i)$ and $0.5(x_{i+k+1} + x_{i+1}) - x_{i+1} = 0.5(x_{i+k+1} - x_{i+1})$. For $x \in J_0 = (x_1, (x_1 + x_{k+1})/2]$, $\max\{|x_j - x|;\ 1 \le j \le k\}$ is bounded above by the larger of $x_k - x_1$ and $0.5(x_1 + x_{k+1}) - x_1 = 0.5(x_{k+1} - x_1)$; while for $x \in J_{n-k} = ((x_{n-k} + x_n)/2, x_n]$, $\max\{|x_j - x|;\ n-k+1 \le j \le n\}$ is bounded above by the larger of $x_n - 0.5(x_{n-k} + x_n) = 0.5(x_n - x_{n-k})$ and $x_n - x_{n-k+1}$. Result (a) follows. Result (b) follows since

$$|x_{i+k} - x_i| = |F^{-1}\left(\frac{i+k-0.5}{n}\right) - F^{-1}\left(\frac{i-0.5}{n}\right) + o(n^{-1})|$$

$$\leq c\,\left|\frac{i+k-0.5}{n} - \frac{i-0.5}{n} + o(n^{-1})\right| = c\,\left|\frac{k}{n} + o(n^{-1})\right|$$

Returning again to the proof of Theorem 5.1 we have now shown that $E(\hat{\mu}(x_{p-t})) - \mu(x_{p-t})$ and $E(\hat{\mu}(x_{p+t})) - \mu(x_{p+t})$ converge uniformly to zero over the respective sets $\{x_{p-t} : t < p < 1\}$ and $\{x_{p+t} : 0 < p < 1-t\}$

Finally we turn to $\hat{\sigma}^2(x) - \sigma^2(x)$.

Lemma 5.3. Suppose that as $n \to \infty$, $\max_{1 \leq i \leq n-k} |x_{i+k} - x_i| \to 0$, $k^{-2} \sum_{i=1}^{n} \sigma^2(x_i) \to 0$, and $k^{-2} \sum_{i=1}^{n} \sigma_2^2(x_i) \to 0$. Assume that $\mu(x)$ and $\sigma^2(x)$ satisfy the Lipschitz condition (5.5). Then, for each $\delta > 0$, $P\{\sup_{x_1 \leq x \leq x_n} |\hat{\sigma}^2(x) - \sigma^2(x)| > \delta\} \to 0$ as $n \to \infty$.

Proof.

$$\text{(5.6)} \qquad \hat{\sigma}^2(x) - \sigma^2(x) = k^{-1} \sum_{i \in I_k(x)} [Y_i - \hat{\mu}(x)]^2 - \sigma^2(x)$$

$$= k^{-1} \sum_{i \in I_k(x)} [Y_i - \mu(x)]^2 - \sigma^2(x) - [\hat{\mu}(x) - \mu(x)]^2 .$$

The third term converges uniformly to zero in probability by Lemmas 5.1 and 5.2. Similarly, if we let $\tilde{\sigma}^2(x)$ denote the expected value of the first term, the difference between the first term and $\tilde{\sigma}^2(x)$ tends uniformly to zero in probability by Lemma 5.1. It remains to show that $\tilde{\sigma}^2(x) - \sigma^2(x)$ tends uniformly to zero. Note that

$$\text{(5.7)} \qquad \tilde{\sigma}^2(x) - \sigma^2(x) = k^{-1} \sum_{i \in I_k(x)} E[Y_i - \mu(x)]^2 - \sigma^2(x), \text{ where}$$

$$E[Y_i - \mu(x)]^2 = E[Y_i - \mu(x_i)]^2 + [\mu(x_i) - \mu(x)]^2$$

$$= \sigma^2(x_i) + [\mu(x_i) - \mu(x)]^2 .$$

By Lemma 5.2, $k^{-1} \sum_{i \in I_k(x)} \sigma^2(x_i) - \sigma^2(x)$ tends uniformly to zero. Similarly, the proof of Lemma 5.2 shows that $\{c \max_{1 \leq i \leq n-k} |x_{i+k} - x_i|\}^2$ is a uniform upper bound on $k^{-1} \sum_{i \in N_k(x)} [\mu(x_i) - \mu(x)]^2$. The result follows.

This completes the proof of Theorem 5.1.

6. Asymptotic Normality of Estimated Correlation Curves.

6(a). General Correlation Curves. Suppose that $\hat{m}(x_{p+t})$, $\hat{m}(x_{p-t})$ and $\hat{\tau}^2(x)$ are consistent estimates of $m(x_{p+t})$, $m(x_{p-t})$ and $\tau^2(x)$ respectively. Let

$$a_n(x) = [m(x_{p+t}) - m(x_{p-t})]^2 \tau_{1n}^2, \quad b_n(x) = (x_{p+t} - x_{p-t})^2 \tau^2(x)$$

$$\hat{a}_n(x) = [\hat{m}(x_{p-t}) - \hat{m}(x_{p-t})]^2 \tau_{1n}^2, \quad \hat{b}_n(x) = (x_{p+t} - x_{p-t})^2 \hat{\tau}^2(x).$$

We consider the fixed covariates case with $x_1, \ldots, x_n$ a regular sequence of covariates such that $\tau_{1n}^2 \to \tau_1^2$ as $n \to \infty$ where τ_1^2 is a scale parameter for $F(x) = \lim F_n(x)$, and where $F_n(x) = n^{-1}[\# x_i \le x]$ as before. The key to obtaining asymptotic normality of estimated correlation curves is to use a little algebra to rewrite $\sqrt{k}[\hat{\rho}_n^2(x) - \rho_n^2(x)]$ as

$$\sqrt{k}[\hat{\rho}_n^2(x) - \rho_n^2(x)] = \{[1 - \rho_n^2(x)]\sqrt{k}[\hat{a}_n(x) - a_n(x)] + \rho_n^2(x)\sqrt{k}[\hat{b}_n(x) - b_n(x)]\}/[\hat{a}_n(x) + \hat{b}_n(x)]. \tag{6.1}$$

Now we can use Slutsky's Theorem to conclude that if we replace the denominator on the right hand side of (6.1) by its limit $a(x) + b(x)$, then the limiting distribution of the resulting quantity will be the limiting distribution of $\sqrt{k}[\hat{\rho}_n^2(x) - \rho^2(x)]$. Similarly, we can replace $x_q = F_n^{-1}(q)$ by $F^{-1}(q)$. By abuse of notation we from now on use x_q to denote $F^{-1}(q)$ rather than $F_n^{-1}(q)$. Thus $a(x) = [m(x_{p+t}) - m(x_{p-t})]^2 \tau_1^2$ and $b(x) = (x_{p+t} - x_{p-t})^2 \tau^2(x)$ are limiting versions of $b_n(x)$ and $a_n(x)$.

Proposition 6.1. Suppose that $k \to \infty$ and $(k/n) \to 0$ as $n \to \infty$. Suppose that for each $x \in [x_t, x_{1-t}]$, $\sqrt{k}[\hat{a}_n(x) - a_n(x)]$ and $\sqrt{k}[\hat{b}_n(x) - b(x)]$ are asymptotically normal $N(\mu_a(x), \sigma_a^2(x))$ and $N(\mu_b(x), \sigma_b^2(x))$, respectively. Then, for each $x \in [x_t, x_{1-t}]$, $\sqrt{k}[\hat{\rho}_n^2(x) - \rho_n^2(x)]$, converges to a normal $N(\mu_\rho(x), \sigma_\rho^2(x))$, distribution, where

$$\mu_\rho(x) = b(x)\mu_a(x) + a(x)\mu_b(x), \quad \sigma_\rho^2(x) = b^2(x)\sigma_a^2(x) + a^2(x)\sigma_b^2(x).$$

Moreover, if $\{\sqrt{k}[\hat{a}_n(x) - a_n(x)];\ x \in [x_t, x_{1-t}]\}$ and $\{\sqrt{k}[\hat{b}_n(x) - b_n(x)];\ x \in [x_t, x_{1-t}]\}$ converge weakly in $D[x_t, x_{1-t}]$ to the respective processes $W_1(x)$ and $W_2(x)$, then $\sqrt{k}[\rho_n^2(x) - \rho^2(x)]$ converges weakly in $D[x_t, x_{1-t}]$ to the process $\{[1 - \rho^2(x)]W_1(x) + \rho^2(x)W_2(x)\}/[a(x) + b(x)]$, where W_1 and W_2 are independent.

Proof. Since $(k/n) \to 0$ as $n \to \infty$, then there exists n_0 such that for all $n \ge n_0$, the three sets I_{p-t}, I_p and I_{p+t} do not intersect. It follows that $\hat{a}_n(x)$ and $\hat{b}_n(x)$ are independent for $n \ge n_0$, and the results follows.

6(b). Nearest neighbor correlation curves. As in Section 5(b), let $\hat{\rho}_n(x)$ be the estimated correlation curve based on the nearest neighbor estimates $\hat{\mu}(x_{p-t})$, $\hat{\mu}(x_{p+t})$ and $\hat{\sigma}^2(x_p)$, respectively. Then we are interested in the asymptotic normality of

$$\sqrt{k}[\hat{A}_n(x) - A_n(x)] = \sqrt{k}\{[\hat{\mu}(x_{p+t}) - \hat{\mu}(x_{p-t})]^2 - [\mu(x_{p+t}) - \mu(x_{p-t})]^2\}\sigma_{1n}^2.$$

Assuming that $\sigma_{1n}^2 \to \sigma_1^2$, where σ_1^2 is the variance in the distribution function F, we can use Slutsky's Theorem to replace σ_{1n}^2 by σ_1^2. Similarly, using the expression

$a^2 - b^2 \cong 2b(a-b)$ for a close to b we find that $\sqrt{k}[\hat{A}_n(x) - A(x)]$ has the same asymptotic distribution as

(6.2) $$2\sqrt{k}[\mu(x_{p+t}) - \mu(x_{p-t})][\hat{\mu}(x_{p+t}) - \mu(x_{p+t}) - \{\hat{\mu}(x_{p-t}) - \mu(x_{p-t})\}]$$

Proposition 6.2. Assume that $k \to \infty$ and $(k/n) \to 0$ as $n \to \infty$. Suppose that $\{Y_i - \mu(x_i) : i \in I_q\}$ satisfy the Lindeberg - Feller Central Limit Theorem conditions for $q = p - t$ and $q = p + t$. Assume that $x_1, \ldots, x_n$ is a regular sequence of covariates such that

(6.3) $\sqrt{k}\max\{|x_i - x_q| : i \in I_q\} \to 0$ as $n \to \infty$ for $q = p - t$ and $q = p+t$.

Suppose that $|\mu(x_i) - \mu(x_q)| \le c|x_i - x_q|$ and that $|\sigma^2(x_i) - \sigma^2(x_q)| \le c|x_i - x_q|$ for $i \in I_q$, $q = p - t$ and $q = p + t$, and some $c > 0$, then $\sqrt{k}[\hat{A}_n(x) - A(x)]$ is asymptotically normal, $N(0, \sigma_A^2(x))$, where

$$\sigma_A^2(x) = 4[\mu(x_{p+t}) - \mu(x_{p-t})]^2[\sigma^2(x_{p+t}) + \sigma^2(x_{p-t})].$$

Proof. The proof follows from the expression (6.2), Lemma 5.2, the Lindeberg - Feller Central Limit Theorem, and the fact that under the conditions given, $k^{-1}\sum_{i \in I_q} \sigma^2(x_i) \to \sigma^2(x_q)$ for $q = p - t$ and $q = p + t$.

Remark 6.1. The Lindeberg - Feller conditions are satisfied if in model (5.2) we assume that $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. and

$$\frac{\{\max \sigma^2(x_i) : i \in I_q\}}{\sum_{i \in I_q} \sigma^2(x_i)} \to 0 \quad \text{as} \quad n \to \infty$$

for $q = p - t$ and $q = p + t$.

Remark 6.2. The condition (6.3) is satisfied if x_i can be written as $x_i = F^{-1}(\frac{i - 0.5}{n}) + o(n^{-1})$ with F^{-1} satisfying the Lipschitz condition (5.5) and provided $(k^{3/2}/n) \to 0$ as $n \to \infty$.

To find the asymptotic distribution of the nearest neighbor correlation curve $\hat{\rho}(x)$, it remains to find the asymptotic distribution of

$$\sqrt{k}[\hat{B}_n(x) - B_n(x)] = \sqrt{k}(x_{p+t} - x_{p-t})^2[\hat{\sigma}^2(x) - \sigma^2(x)].$$

Proposition 6.3. Assume that the conditions of Proposition 6.2 are satisfied when $t = 0$. In addition assume that $|\sigma_2^2(x_i) - \sigma_2^2(x)| \le c|x_i - x|$ for $i \in I_p$, some $c > 0$ and that $\{[Y_i - \mu(x_i)]^2 - \sigma^2(x_i) : i \in I_p\}$ satisfy the Lindeberg-Feller Central Limit Theorem conditions. Then $\sqrt{k}[\hat{B}_n(x) - B(x)]$ is asymptotically normal, $N(0, \sigma_B^2(x))$, where

$$\sigma_B^2(x) = (x_{p+t} - x_{p-t})^4 \sigma_2^2(x)$$

Proof. Using the proof of Lemma 5.3, we can write

$$\sqrt{k}\,[\hat{\sigma}^2(x) - \sigma^2(x)] = k^{-1/2} \sum_{i \in I_k(x)} \{[Y_i - \mu(x_i)]^2 - \sigma^2(x_i)\} + R_{n,k}$$

where $R_{n,k}$ is a remainder term which tends to zero as $n \to \infty$, $k \to \infty$, $(k/n) \to 0$. Now the result follows by applying the Lindeberg-Feller Theorem to $\{[Y_i - \mu(x_i)]^2 - \sigma^2(x_i) : i \in I_k(x)\}$.

Remark 6.3. The Proposition 6.3 Lindeberg-Feller conditions are satisfied if in model (5.2) we assume that $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. and

$$\frac{\{\max \sigma_2^2(x_i) : i \in I_k(x)\}}{\sum_{i \in I_k(x)} \sigma_2^2(x_i)} \to 0 \text{ as } n \to \infty.$$

Combining Propositions 6.3 and 6.4 we arrive at

Theorem 6.1. Under the conditions of Propositions 6.3 and 6.4, the nearest neighbor correlation curve $\hat{\rho}(x)$ is asymptotically normal in the sense that as $n \to \infty$ $\sqrt{k}\,[\hat{\rho}(x) - \rho(x)]$ tends in law to $N(0, \sigma_\rho^2(x))$, where

$$\sigma_\rho^2(x) = \{[1 - \rho^2(x)]^2 \sigma_A^2(x) + \rho^4(x)\sigma_B^2(x)\} / [a(x) + b(x)]^2$$

7. Computing the Estimated Correlation Curve and a Bootstrap Confidence Procedure.

7(a). The Estimate. We let $(x_{(1)}, y_1), \ldots, (x_{(n)}, y_n)$ denote the observed data. We assume that these have been generated by the fixed covariate model (5.2) where the x's are nonrandom and ordered. We will describe an algorithm for estimating $\hat{\rho}(x_p)$, where $x_p = x_{(m)}$ with $m = [np] + 1$ and where $[x]$ denotes the largest integer less than or equal to x. We need to define three disjoint and adjoining neighbourhoods of size k, neighbourhoods about the points x_{p-t}, x_p and x_{p+t}. We will define them in terms of the indexes of x's closest to x_q, $q = p - t, p, p + t$. Denote these neighbourhoods as NI, NII and NIII respectively. NII is then seen to be

$$\text{NII} = \{[np] + 1 - (k-1)/2, \ldots, [np] + 1 + (k-1)/2\}.$$

This neighborhood is of size k when k is odd and k + 1 otherwise. For simplicity, let us assume that k is odd. Likewise,

$$\text{NI} = \{[n(p-t)] + 1 - (k-1)/2, \ldots, [n(p-t)] + 1 + (k-1)/2\}$$

and

$$NIII = \{[n(p+t)] + 1 - (k-1)/2, \ldots, [n(p+t)] + 1 + (k-1)/2\},$$

Choose t so that $n \cdot t$ is integer valued, say $n \cdot t = l$. We require that the index-sets NI, NII and NIII are adjoining, non-overlapping and that they only contain positive integers. These requirements lead to the following (recall that $m = [n \cdot p] + 1$):

$$[n(p-t)] + 1 - (k-1)/2 \geq 1 \iff m \geq l + (k+1)/2,$$

$$[n(p-t)] + 1 + (k-1)/2 = [np] - (k-1)/2 \Rightarrow l = k.$$

Thus,

$$(3k+1)/2 \leq m \leq n - (3k+1)/2$$

and

$$(3k+1)/2n \leq p \leq n + 1 - (3k+1)/2n.$$

The computer program Mathematica, which is widely available, can conveniently be used to compute the dependence function and produce a plot of the function. We assume that we have two functions to our disposal, Median[y] and IQR[y], that returns the median and the interquartile range respectively (or, if the mean and standard deviation are preferred, Mean[y] and SD[y]). In the notation of Mathematica, the sets above are called lists. Let J denote the list $\{1, \ldots, k\}$. Then NII = m + J − 1 − Floor[(k − 1)/2] will denote the list of integers in NII. Correspondingly,

$$NI = m - 1 - \mathrm{Floor}[(k-1)/2] + J - 1$$

and

$$NIII = m + 1 + \mathrm{Floor}[(k-1)/2] + J - 1.$$

In the notation of Mathematica, if y is the list containing the Y-observations, then y[[N]] is the list of Y-observations with indexes in the list N.

An expression that in Mathematica defines a function that will return the value of $\hat{\rho}(x_p)$, is now given by

```
Ro[m] := deltay IQR[x]/Sqrt[(deltay IQR[x])^2
  + ((x[[m + 1]] - x[[m - 1]]) IQR[y[[NII]]])^2]
```

where deltay has to be given the value deltay = Median[y[[NIII]]] − Median[y[[NI]]]. A plot of the dependence function is obtained through the following Mathematica statements:

```
R = Range[k + (k + 1)/2, n - k - (k + 1)/2]
ListPlot[Transpose[{x[[R]], Map[Ro, R]}]]
```

7(b). A Bootstrap Simultaneous Confidence Procedure. From the collection $\{(x_j, y_j), j \in NI\}$ of pairs corresponding to the first index set, we select k pairs with replacement. Then, independently, we do the same for the second and third index sets. This procedure is repeated B times resulting in B independent triples of independent bivariate samples. At the ith stage we have three independent bivariate samples whose index sets are denoted by NI_i^*, NII_i^* and $NIII_i^*$; $i = 1, \ldots, B$. We let $\hat{\rho}_i^*(\hat{x}_p)$ denote the estimate correlation curve based on $\{(x_j, y_j) : j \in NI_i^*\}$, $\{(x_j, y_j) : j \in NII_i^*\}$ and $\{(x_j, y_j) : j \in NIII_i^*\}$. More precisely, $\hat{\rho}_i^*(\hat{x}_p)$ is obtained by computing the formula in Section 5b with σ_{1n}, x_{p-t}, x_p, x_{p+t} unchanged but $\hat{\mu}(x_{p+t}) - \hat{\mu}(x_{p-t})$ replaced by

$$\hat{\mu}_i^*(x_{p+t}) - \hat{\mu}_i^*(x_{p-t}) = k^{-1} \sum_{j \in NIII_i^*} y_j - k^{-1} \sum_{j \in NI_i^*} y_j$$

and $\hat{\sigma}^2(x_p)$ replaced by

$$\hat{\sigma}_i^{2*}(x_p) = k^{-1} \sum_{j \in NII_i^*} [y_j - \hat{\mu}_i^*(x_p)]^2, \quad \text{where} \quad \hat{\mu}_i^*(x_p) = k^{-1} \sum_{j \in NII_i^*} y_j .$$

Now we approximate the distribution of $\sqrt{k}\,|\hat{\rho}(x_p) - \rho(x_p)|$ with the empirical distribution of $\{\sqrt{k}\,|\hat{\rho}_i^*(x_p) - \hat{\rho}(x_p)|, i = 1, \ldots, B\}$. Let k_α^* denote the $(1-\alpha)$th quantile of this distribution, then our level $(1-\alpha)$ confidence interval for $\rho(x_p)$ is

$$\rho(x_p) = \hat{\rho}(x_p) \pm k_\alpha^* / \sqrt{k}$$

Suppose we are interested in the strength of the relationship between X and Y at several quantiles of the covariate, say at $q_1, \ldots, q_a$ where q_i denote the p_ith quantile of X. To get simultaneous confidence intervals for $\rho(q_1), \ldots, \rho(q_a)$, we consider the empirical distribution of $\{\sqrt{k} \max_{1 \le j \le a} |\hat{\rho}_i^*(q_j) - \hat{\rho}(q_j)|, i = 1, \ldots, B\}$. Let c_α^* denote the $(1-\alpha)$th quantile of this distribution, then our level $(1-\alpha)$ simultaneous confidence intervals for $\rho(q_1), \ldots, \rho(q_a)$ are

$$\rho(q_j) = \hat{\rho}(q_j) \pm c_\alpha^* / \sqrt{k}, \quad j = 1, \ldots, a$$

8. Examples.

8(a). A Data Example. Figure 3 below gives the scatter plot for pairs (x, y) of readings of plasma lipid concentrations taken on 371 diseased patients in a heart study; see Scott, Gotto, Cole and Gorry (1978). This data set has also been analysed by Silverman (1986, pp.81-83). Figure 4 gives the corresponding empirical correlation curve with $t = 0.15$.

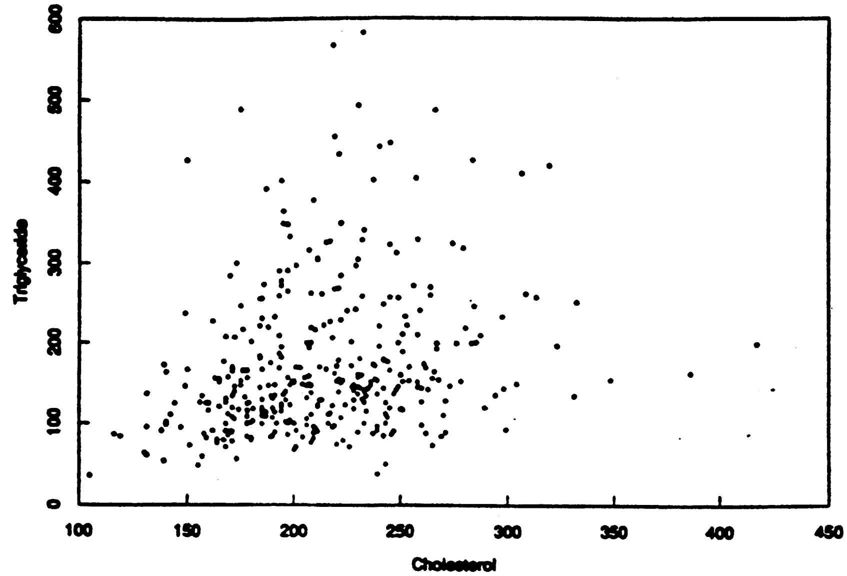

Figure 3. Scatter plot of plasma lipid concentrations.
x = plasma cholesterol concentration (mg/100ml),
y = plasma triglyceride concentration (mg/100ml)

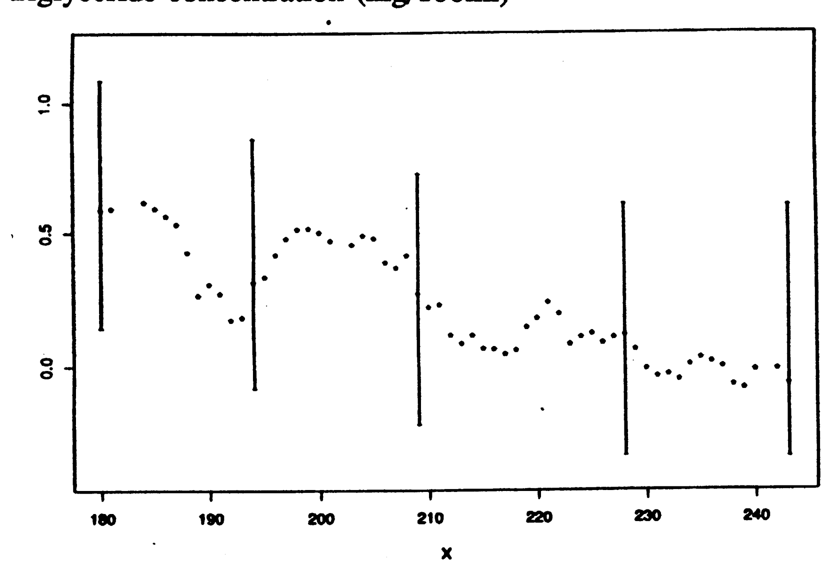

Figure 4. The correlation curve for the data in Figure 3. Here $t = 0.15$, $k = 55$ and $0.225 \le p \le 0.775$. The vertical bars indicate 90% simultaneous bootstrap confidence intervals. They are based in 1000 bootstrap simulations.

The empirical correlation curve indicates a strong to moderate association between cholesterol and triglyceride concentration for small to moderate values of cholesterol concentration. The correlation curve is nearly zero for x larger than the 64th quantile.

The simultaneous confidence intervals show that at $x = 180$, the hypothesis of no association between tryglycerite and cholesterol can be rejected at the 10% level of significance. The hypothesis of no association is not rejected at the values x = (194, 209, 228 and 243.

We consulted a medical expert (Jon Bjerve) 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 level $x = 180$, the estimated correlation is 0.67 which is significantly different from zero at level $\alpha = .10$ (as well as level $\alpha = .05$). At cholesterol level $x = 243$, the estimated correlation is 0.067 which is not close to significant at any reasonable level of significance. 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.

8(b). A Simulation Example. Figure 5 below gives the contour plots for the model $(X, Y) = (20 - \exp(S), T)$, where (S, T) has the bivariate normal distribution $N(2, 4, 1, 2, -0.5)$. Clearly a twisted pear effect is evident. Figure 6 below shows the true correlation curve $\rho(x)$ and the empirical correlation curve $\hat{\rho}(x)$ based on $n = 1000$ observations drawn from the given (X, Y) distribution as well as vertical bars indicating 90% simultaneous bootstrap confidence intervals based on 1000 bootstrap simulations. We used $t = 0.15$ and $k = 150$.

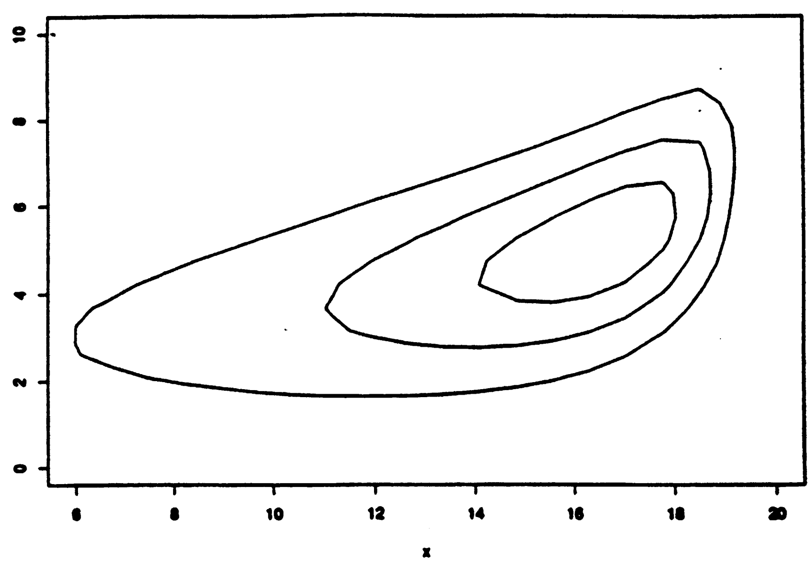

Figure 5. Contour plots $\{(x, y)\colon f(x, y) = c\}$ with $c = 0.05$, 0.10 and 0.15 for the transformed bivariate normal model.

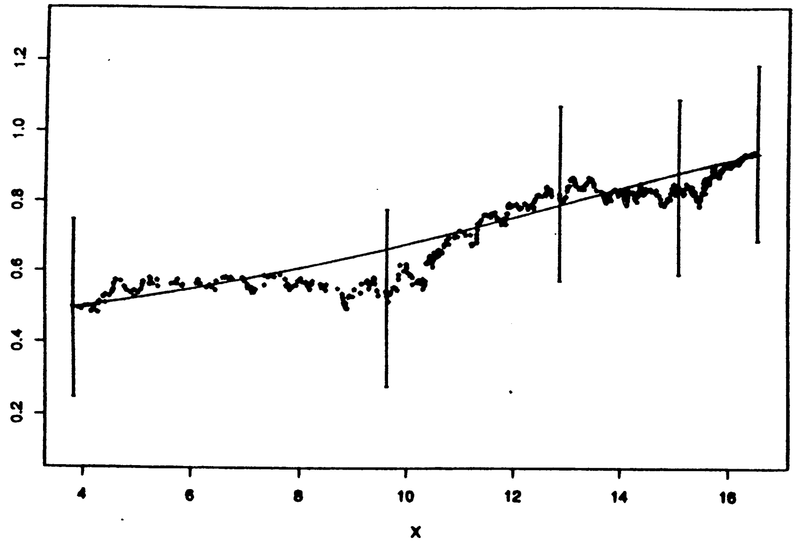

Figure 6. The true correlation curve ρ (x) and estimated correlation curve $\hat{\rho}$ (x) for the transformed bivariate normal model. The vertical bars are 90% simultaneous bootstrap confidence intervals for the correlation curve ρ (x).

Acknowledgements. We are grateful to Jack Block and Per Gjerde for comments that contributed to the ideas in this paper. We are grateful to Thomas Scheike for help with the programming, and to David W. Scott for making the cholesterol data available.

References

Anscombe, F.J. (1961). Examination of Residuals. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, Vol I, pp 1-36. University of California Press, Berkeley.

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

Bhattacharya, P.K. (1974). Convergence of sample paths of normalised sums of induced order statistics. *Ann. Statist.* **2**, 1034-1039.

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, Inc., Oakland, California.

Bjerve, S., Doksum, K.A. and Yandell, B.S. (1985). Uniform confidence bounds for regression based on a simple moving average. *Scand. J. of Statist.* **12**, 159-170.

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, London.

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

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

Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: the effect of estimating the mean. *J. R. Statist. Soc. Ser. B* **51**, 3-14.

Hall, P., Kay, J.W., and Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521-528.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, England; New York.

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

Kowalczyk, T. (1977). General Definition and Sample Counterparts of Monotonic Dependence Functions of bivariate distributions *MaOpfStS* **8**, 351-369.

Kowalczyk, T. and Pleszczyńska, E. (1977). Monotonic Dependence Functions of bivariate distributions. *Ann. Statist.* **5**, 1221-1227.

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

Müller, H.G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **12**, 221-232.

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, New York.

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

Stigler, S.M. (1989). Galton's account of the invention of correlation. *Statistical Science* **4**, 73-86.