An extension of the Galton-Pearson correlation coefficient to the nonlinear case

By

Kjell A. Doksum

Technical Report No. 330
November 27, 1991

Department of Statistics
University of California
Berkeley, California 94720

Nonparametric Statistics and Related Topics (1992) Elsevier, Amsterdam
Edited by A.K.Md.E. Saleh

An extension of the Galton-Pearson correlation coefficient to the nonlinear case

Kjell A. Doksum

Department of Statistics, University of California, Berkeley, CA 94720, U.S.A.

Abstract

This paper addresses the problem of constructing a correlation type measure of association between a response variable Y and a covariate X (possibly vector valued) which is suitable for models where $E(Y \mid X = x)$ is not linear in x and $\mathrm{Var}(Y \mid X = x)$ is not constant in x. A simple local measure is constructed by making use of Galton's ideas that led to the invention of the usual Galton-Pearson correlation coefficient. This measure is justified on several grounds and a related global measure is introduced. The local measure is illustrated on income-potato expenditure data and shows that in low income groups income and potato expenditure is positively correlated while in high income groups the correlation is negative.

1. INTRODUCTION

In this paper I report on some recent results on new measures of correlation. The ideas and results have been developed jointly with Steinar Bjerve (Bjerve and Doksum (1990)), Stephen Blyth, Eric Bradlow and Xiao-Li Meng (Doksum, Blyth, Bradlow and Meng (1991)), as well as Alex Samarov (Doksum and Samarov (1991)). The new measures are designed to measure the strength of the relationship between a response variable Y and a covariate X for experiments where the strength of association between X and Y is different for different values of the covariate X. These dependence measures are extensions of the Galton-Pearson correlation coefficient ρ to the case where $E(Y \mid X = x)$ is non-linear in x and $\mathrm{Var}(Y \mid X = x)$ is non-constant in x. Since the measures are between -1 and 1 and satisfy the invariance properties of the correlation coefficient ρ, we refer to them as *correlation curves*. We also consider global correlation measures obtained by averaging out x.

In linear statistical inference, regression and correlation are treated as complementary topics. Regression measures the average relationship between a response variable Y and a covariate X while correlation measures the strength of the relationship between X and Y as well as the amount of variability that can be explained by regression. In non-linear statistical inference, the regression curve $\mu(x) = E(Y \mid X = x)$ takes the place of the linear regression line; however it is not as obvious what should take the place of the Galton-Pearson correlation coefficient ρ. The problem is that if we apply the usual correlation formula to the conditional distribution of Y given $X = x$, we get the value zero. To see this recall that $\rho^2 \leq \eta^2$, where $\eta^2 = \mathrm{Var}(\mu(X)) / \mathrm{Var}(Y)$ (e.g. Kendall and Stuart (1962)). In the conditional distribution $\mathbf{L}(X, Y \mid X = x)$, η^2 reduces to zero since $\mathrm{Var}(\mu(X) \mid X = x) = 0$, while (except in trivial cases) $\mathrm{Var}(Y \mid x) = \mathrm{Var}(Y \mid X = x) > 0$. If we instead of conditioning on $X = x$ 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.

Our approach is to construct a local measure of correlation by combining ideas from nonparametric regression and Stigler's (1986, 1989) account of Galton's (1888) invention of the correlation coefficient ρ. According to Stigler, Galton realized that in the linear model case the regression slope β is not an appropriate measure of strength of association because of its dependence on the measurement scales selected for X and Y, and because the slope when regressing Y on X is different from the slope when regressing X on Y. Thus, as a measure of the strength of the co-relation between X and Y, Galton proposed using the regression slope computed after X and Y have been converted to the 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. In other words, Galton considered correlation to be the rate of change in the expected value $E(Y \mid X = x)$ measured in units of the Y scale σ_2 as x is increased in units of the X scale σ_1. By defining correlation this way Galton had invented a measure which is independent of the original scales selected for X and Y and which gives the same measure of co-relation when Y is regressed on X as when X is regressed on Y.

To apply Galton's ideas in non-linear models our approach consists of rewriting ρ in terms of the regression slope β and the residual variance σ_ε^2 and then replacing the regression slope β by the curve $\beta(x) = \mu'(x) = d\mu(x) / dx$ and the residual variance σ_ε^2 by the local residual variance $\sigma^2(x) = \mathrm{Var}(Y \mid x)$. More precisely, consider the linear model

$$Y = \alpha + \beta X + \sigma_\varepsilon \varepsilon, \qquad E(\varepsilon) = 0, \qquad \mathrm{Var}(\varepsilon) = 1 \tag{1}$$

where X and ε are independent. We now specify $\sigma_1^2 = \mathrm{Var}(X)$ and $\sigma_2^2 = \mathrm{Var}(Y)$, as did Pearson, and using $\rho = \sigma_1 \beta / \sigma_2$ and $\sigma_2^2 = \mathrm{Var}(\beta X) + \sigma_\varepsilon^2$, write the Galton-Pearson correlation coefficient ρ as

$$\rho = \frac{\sigma_1 \beta}{\{\sigma_1^2 \beta^2 + \sigma_\varepsilon^2\}^{1/2}} \qquad \text{(linear model)}$$

To obtain our measure of dependence we replace β in the above formula by $\beta(x)$ and σ_ε^2 by $\sigma^2(x)$. That is, we define our correlation curve $\rho(x)$ as

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

Note that $\rho(x)$ is a local measure of dependence. It is large and close to one for a particular covariate value x when the regression slope $\beta(x)$ is positive and large relative to the residual standard deviation $\sigma(x)$. In particular, for the example where X and Y are treatment and response levels, respectively, it is often the case that $\rho(x)$ starts out near zero and increases towards one as the treatment level is increased.

The correlation curve $\rho(x)$ is between -1 and 1 and it satisfies all the equivariance properties of the correlation coefficient ρ. These are properties that Rényi (1959) and Bell (1962) have argued that global correlation measures should have. It reduces to the correlation coefficient ρ in the linear model case.

2. VARIANCE EXPLAINED BY REGRESSION

Next suppose we try to use the coefficient of determination (R^2) idea to motivate a local measure of dependence. This leads to $R^2(x) = 1 - [\mathrm{Var}(Y|x)/\mathrm{Var}(Y)]$, which works well in linear heteroscedastic models where the regression slope $\mu'(x)$ is a constant while $\mathrm{Var}(Y|x)$ changes with x. However when $\beta(x) = \mu'(x)$ is not constant, it takes the same value at values of x where $\beta(x)$ is zero as where $|\beta(x)|$ is large. This conflicts with the idea (Galton (1888), Stigler (1986, 1989)) explained above that a measure of co-relation should measure the rate of change in the mean of Y/σ_2 as x is increased a certain number of σ_1 units. Thus $R^2(x)$ is not adequate as a measure of local correlation since it does not depend on the local values of the slope $\beta(x)$. However, our squared correlation curve $\rho^2(x)$ *does* have an interpretation as a *local* coefficient of determination. To see this suppose that $\mu(x)$ is smooth so that near the point $x_o, \mu(x)$ is nearly linear with slope $\beta(x_o)$. Suppose also that for X close to x_o we can, to a close approximation, write

$$Y \cong Y_o = \alpha_o + \beta(x_o)X + \sigma(x_o)\varepsilon,$$

where ε and X are independent and ε has variance one. For Y_o, the coefficient of determination is

$$1 - \frac{\mathrm{Var}(Y_o|X=x)}{\mathrm{Var}(Y_o)} = 1 - \frac{\sigma^2(x_o)}{\beta^2(x_o)\sigma_1^2 + \sigma^2(x_o)} = \rho^2(x_o).$$

Note that $\rho^2(x_o)$ is just the squared Galton-Pearson correlation coefficient, or equivalently the coefficient of determination, between X and Y_o. This is the key to overcoming the problem that the naive definition of local correlation in terms of conditional correlation always gives the value zero: We use the conditional distribution given $X = x_o$ to compute the slope $\beta(x_o)$ and variance $\sigma^2(x_o)$. Then we compute the correlation for the linear model where $\beta(x_o)$ and $\sigma^2(x_o)$ are the slope and variance, respectively. In other words, the local correlation curve $\rho(x_o)$ is the correlation coefficient for the virtual linear model based on the local regression slope and the local variance.

3. A SIGNAL-NOISE INTERPRETATION OF LOCAL CORRELATION

Consider the linear model (1) with $\mathrm{Var}(X) = 1$. The βX part of the model is the part that relates X to Y and it is called the *signal* part of model while the disturbance term $\sigma_\varepsilon \varepsilon$ is called the *noise* part. We can think of $[dE(Y|x)/dx]^2 = \beta^2$ as a measure of the strength of the signal and $[dE(Y|x,\varepsilon)/d\varepsilon]^2 = \sigma_\varepsilon^2$ as a measure of the strength of the noise. With this terminology, we can write the usual correlation coefficient as

$$\rho^2 = \frac{\text{signal}}{\text{signal + noise}}.$$

Next consider the nonlinear heteroscedastic model

$$Y = \mu(X) + \sigma(X)\varepsilon, \quad E(\varepsilon) = 0, \quad \mathrm{Var}(\varepsilon) = 1 \tag{2}$$

where X and ε are independent and $\mathrm{Var}(X) = 1$. In analogy with the linear model case, we use $[dE(Y|x)/dx]^2 = \beta^2(x)$ as a local measure of the strength of the signal, and we use $[dE(Y|x,\varepsilon)/d\varepsilon]^2 = \sigma^2(x)$ as a local measure of the strength of the noise. Thus we can write

$$\rho^2(x) = \frac{\text{local signal}}{\text{local signal + local noise}}.$$

When X does not have variance 1, the map $X \to X/\sigma_1$, will give the formula of the previous sections.

4. WHY CORRELATION?

Is there a place for correlation methods in nonlinear analysis? Why not use the regression slope $\beta(x) = \mu'(x)$ and the residual standard deviation $\sigma(x)$ separately rather than combining them into a correlation measure $\rho(x)$? To answer this, we can ask why it is that Galton, Pearson and Fisher, as well as a great number of researchers and statisticians since their time has put such a great emphasis on correlation? Moreover, why is it that correlation plays such an important role in the scientific disciplines that use statistics? The answer is that relating variables is one of the most important and interesting activities in scientific disciplines and correlation provides a universal scale-free measure of the strength of relationships between variables. Thus, in linear models, $100\rho^2$ gives the percentage of the variation of the response variable that can be explained by the regressors. Or equivalently, it gives the percentage of the signal plus noise sum that comes from the signal. This notion is universal and scale free. It can be used to compare the results of two laboratories 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 analysis. Our attempt at providing a correlation measure in the nonlinear case is such that $100\rho^2(x)$ gives the percentage of the local signal plus local noise sum that comes from the local signal.

5. STANDARDIZED REGRESSION

Consider again the idea (Galton (1888), Stigler (1986), 1989)) that a measure of corelation should represent the rate of change of $E(Y|x)$ measured in units of Y-standard deviations as X is increased in units of X standard deviations. When $\sigma^2(x) = \mathrm{Var}(Y|x)$ is not constant, and local dependence near x is of interest, it makes sense to use the conditional $(Y|x)$ standard deviation $\sigma(x)$ rather than the overall Y standard deviation σ_2. This leads to

$$\xi(x) = \sigma_1 \beta(x) / \sigma(x)$$

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

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

Thus $\rho(x)$ has an interpretation as the standardized regression slope $\xi(x)$ mapped onto the interval $[-1, 1]$ in such a way that it coincides with the Galton-Pearson correlation coefficient in the linear model. Moreover, the problem of estimating $\rho(x)$ essentially reduces to the problem of estimating $\xi(x)$.

5. RELATIVE LOCAL CORRELATION

Another 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

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

will have a sensible limit. In fact, $\mathrm{corr}((X, Y | X \in N_h(x_0))$ is to first order $h\sigma_1 \beta(x_0) / \sqrt{3}\,\sigma(x_0) = h\xi(x_0)/\sqrt{3}$. Thus this approach leads back to the standardized regression slope of the previous section. This result holds whether we use the Galton-Pearson ρ^2 or the Pearson $\eta^2 = \mathrm{Var}(\mu(X)) / \mathrm{Var}(Y)$ 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 consider

$$\lim_{h \to 0} \frac{\mathrm{corr}((X, Y | X \in N_h(x_0))}{h / \sqrt{3}} = \frac{\sigma_1 \beta(x_0)}{\sigma(x_0)} = \xi(x_0).$$

As pointed out earlier, $\rho(x)$ is $\xi(x)$ mapped onto $[-1, 1]$ in such a way that it coincides with the Galton-Pearson correlation coefficient in linear models.

6. REGRESSION DEPENDENCE

Lehmann (1966) analyzed the concept of regression dependence where Y is said to be regression dependent on X if $P(Y \le y | X = x)$ is decreasing in x. It turns out that the correlation curve $\rho(x)$ measures the strength of regression dependence. To make this precise, let (X, Y_1) and (X, Y_2) denote two pairs of random variables, and let $Y_1(x)$ and $Y_2(x)$ denote random variables with respective distribution functions $P(Y_1 \le y | X = x)$ and $P(Y_2 \le y | X = x)$. Let $\sigma_1^2(x) = \mathrm{Var}(Y_1 | x)$ and $\sigma_2^2(x) = \mathrm{Var}(Y_2 | x)$. The pair (X, Y_1) is said to be *more regression dependent* than the pair (X, Y_2) if $Y_1(x)/\sigma_1(x)$ is stochastically more increasing than $Y_2(x)/\sigma_2(x)$ in the sense that for δ in some neighborhood $(0, \varepsilon)$ of zero, $[Y_1(x+\delta) - Y_1(x-\delta)]/\sigma_1(x)$ is stochastically larger $[Y_2(x+\delta) - Y_2(x-\delta)]/\sigma_2(x)$. To make the meaning of this definition precise, we need to specify what we mean by $Y_i(x+\delta) - Y_i(x-\delta)$. One way to do this is to restrict attention to the model

$$Y_i(X) = \mu_i(X) + \sigma_i(X)\varepsilon_i, \quad i = 1,2, \tag{3}$$

where X and $(\varepsilon_1, \varepsilon_2)$ are independent and ε_i has mean zero and variance one. The joint distribution of ε_1 and ε_2 is otherwise arbitrary. Without loss of generality, we assume that X has variance one. With this model we can write

$$Y_i(x) = \mu_i(x) + \sigma_i(x)\varepsilon_i, \quad i = 1,2.$$

Let $\xi_i(x) = \sigma_i^{-1}(x) d\mu_i(x)/dx$ denote the standardized regression slope for (X, Y_i), $i = 1,2$, and let

$$Z_i(x, \delta) = [Y_i(x+\delta) - Y_i(x-\delta)]/2\delta\sigma_i(x), \quad i = 1,2.$$

Now it is clear that $E(Z_i(x, \delta)) \to \xi_i(x)$ as $\delta \to 0$. Thus, for the model (3) with $\mu_i(x)$ differentiable, if (X, Y_1) is more regression dependent than (X, Y_2), and if we let $\rho_1(x)$ and $\rho_2(x)$ denote the correlation curves for (X, Y_1) and (X, Y_2), respectively, then $\rho_1(x) \ge \rho_2(x)$ for all x.

7. GENERAL CORRELATION CURVES AND THEIR PROPERTIES

Earlier 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 $\mu(x), \sigma_1^2$ and $\sigma^2(x)$ by other measures of location and scale. This may be desirable since $\mu(x), \sigma_1^2$ and $\sigma^2(x)$ do not always exist. Moreover, they are very sensitive to the tail behavior 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 conditional distribution $L(Y \mid X = x)$ 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 $m'(x) = dm(x)/dx, \tau_1$ and $\tau(x)$ exist. Thus X has a continuous distribution while the distribution of Y may be discrete or continuous. Each time we specify $m(x), \tau_1$ and $\tau(x)$ we get a correlation curve whose formula is

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

It will sometimes be convenient to write $\rho(x)$ in the equivalent form

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

where the sign $\pm$ is the same as the sign of $m'(x)$. Under appropriate conditions, the correlation curves satisfy the basic properties (axioms) of correlation (Rényi (1959), Bell (1962)). This is the case, for instance, if $m(x)$ and $\tau(x)$ are chosen as the median and interquartile range of the distribution $L(Y \mid X = x)$, respectively, and τ_1^2 is chosen as the interquartile range of the distribution of X.

8. THE CASE OF SEVERAL COVARIATES. 8(a). THE LOCAL CASE.

Consider an experiment where on each of n independent subjects we can measure a response Y and k covariate values $X_1, \ldots, X_k$. In a non-linear setting, a measure of variance explained by regression (a coefficient of determination) is

$$\eta^2 = \frac{\operatorname{Var}(\mu(\mathbf{X}))}{\operatorname{Var}(Y)} = 1 - \frac{E(\operatorname{Var}(Y \mid \mathbf{X}))}{\operatorname{Var}(Y)}$$

where $\mu(\mathbf{X}) = E(Y \mid \mathbf{X})$ and $\mathbf{X} = (X_1, \ldots, X_k)$ (e.g. Kendall and Stuart (1962, p. 196)). As a local version of η^2 we could try $1 - \operatorname{Var}(Y \mid \mathbf{x})/\operatorname{Var}(Y)$, however, as in the case of one covariate, this is not sensitive to local changes in $E(Y \mid \mathbf{X} = \mathbf{x})$. Thus, instead, we will consider first the linear model and then ask what the natural extension to the non-linear case would be. Our notation for the linear model is

$$Y = \alpha + \mathbf{X}\beta^T + \sigma_\varepsilon \varepsilon, \;\; E(\varepsilon) = 0, \;\; \operatorname{Var}(\varepsilon) = 1 \tag{4}$$

where $\beta = (\beta_1, \ldots, \beta_k)$ and ε is independent of $\mathbf{X}$. Here $\mathbf{X}\beta^T$ is the part of the model that relates $\mathbf{X}$ to Y and it is the signal part of the model while $\sigma_\varepsilon \varepsilon$ is the noise part. If the covariance matrix Σ of $\mathbf{X}$ is the identity $\mathbf{I}$, then the X's are independent and have variance one. Thus a measure of the strength of the relationship between Y and $\mathbf{X}$ is $\| dE(Y \mid \mathbf{x})/d\mathbf{x} \|^2 = [(\mathbf{x}\beta^T)']^2 = \beta\beta^T$, where $\|\cdot\|$ denotes the Euclidean norm. Similarly a measure of the strength of the noise is $[dE(Y \mid \mathbf{x}, \varepsilon) \mid d\varepsilon]^2 = \sigma_\varepsilon^2$. If $\Sigma \neq \mathbf{I}$, we use the transformation $\mathbf{X} \rightarrow \Sigma^{-1/2}(\mathbf{X} - \mu)$, where μ is the mean vector of $\mathbf{X}$, and find that the strength of the signal is $\beta \Sigma \beta^T$. This leads to the following interpretation of the coefficient of determination:

$$\rho^2 = \frac{\text{signal}}{\text{signal+noise}} = \frac{\beta \Sigma \beta^T}{\beta \Sigma \beta^T + \sigma_\varepsilon^2}$$

Consider next the nonlinear heteroscedastic model

$$Y = \mu(X) + \sigma(X)\varepsilon, \ E(\varepsilon) = 0, \ \ \mathrm{Var}(\varepsilon) = 1 \tag{5}$$

where $\mathbf{X}$ and ε are independent. Let $\beta(\mathbf{x}) = (d\mu(\mathbf{x})/d\mathbf{x})$ and suppose $\mathbf{\Sigma} = \mathrm{cov}(\mathbf{X}) = \mathbf{I} =$ identity matrix, then $\|dE(Y|\mathbf{x})/d\mathbf{x}\|^2 = \beta(\mathbf{x})\beta^T(\mathbf{x})$ is a local measure of the strength of the signal. Moreover $\sigma^2(\mathbf{x}) = [dE(Y|\mathbf{x},\varepsilon)/d\varepsilon]^2$ is a local measure of the strength of the noise. If $\mathbf{\Sigma} \neq \mathbf{I}$, we standardize $\mathbf{X}$ by mapping it to $\Sigma^{-1/2}(\mathbf{X} - \mu)$ and the local signal becomes $\beta(\mathbf{x})\Sigma\beta^T(\mathbf{x})$. This leads to the following local coefficient of determination for the nonlinear heteroscedastic model (5):

$$\rho^2(\mathbf{x}) = \frac{\text{local signal}}{\text{local signal + local noise}} = \frac{\beta(\mathbf{x})\Sigma\beta^T(\mathbf{x})}{\beta(\mathbf{x})\Sigma\beta^T(\mathbf{x}) + \sigma^2(\mathbf{x})}$$

8(b) GLOBAL MEASURES OF CORRELATION IN THE NONLINEAR CASE

In experiments where Y denotes the response of a person and $\mathbf{x}$ is a set of covariates for that person, $\rho^2(\mathbf{x})$ will give the strength of the relationship between the response and the covariates for that person. However, since $\rho^2(\mathbf{x})$ is a map from R^k to R, it may be cumbersome to work with, and it may be useful to have a global measure of correlation. One useful such measure is given by η^2 as defined in Section 8(a). Another measure which is more sensitive to changes in $E(Y|\mathbf{X} = \mathbf{x})$ is obtained by taking expected values in the local signal plus noise sum expression of Section 8(a) preceeding. This leads to the following global measure of association.

$$\frac{E(\text{local signal})}{E(\text{local signal}) + E(\text{local noise})} = \frac{E[\beta(\mathbf{X})\Sigma\beta^T(\mathbf{X})]}{E[\beta(\mathbf{X})\Sigma\beta^T(\mathbf{X})] + E[\sigma^2(\mathbf{X})]}$$

9. LOCAL INCOME-POTATO EXPENDITURE CORRELATION

The concept of local correlation will be illustrated using the $n = 7125$ pairs of income-potato expenditure values from the Family Expenditure Survey (1973) as described by Härdle (1990). Following Härdle, we divide income by average income and potato expenditure by average potato expenditure so that both variables x and y have mean one.

Local weighted linear regression will be used to estimate the functions $\mu(x) = E(Y|x)$, $\beta(x) = d\mu(x)/dx$ and $\sigma(x) = \{E[X - \mu(x)]^2|x\}^{1/2}$. The methods for $\mu(x)$ and $\beta(x)$ used here are essentially from Fan (1991). They are similar to methods proposed by Stone (1977) and Cleveland (1979). The methods are as follows: Let $K(u) = 0.75\,(1 - u^2)\,I(|u| \le 1)$ 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 $w_i = K((x_i - x_0)/h)$, $h = 0.5s_1$, and s_1 is the sample standard deviation of $x_1, \ldots, x_n$, $n = 7125$. 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".

Figure 1 gives the income-potato expenditure data together with the mean curve $\hat{\mu}(x)$. Figure 2 gives the estimated standard deviation curve $\hat{\sigma}(x)$. Both $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are increasing for the lower income group, they level off at the middle income group and become decreasing in the moderately high income group, thereby illustrating both nonlinearity and heteroscedasticity (compare Härdle (1990, pages 102-103))

Figure 3 gives the slope curve $\hat{\beta}(x)$ and shows how the local regression coefficient drops from about 1.4 for the low income group to about -1.2 for the moderately high income group. Finally, Figure 4 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 local correlation starts out about .9 for the low income group, drops off to a value close to zero around the middle income group and reaches the value -.6 for the moderately high income group near $x = 2.4$. Thus income and potato expenditure is strongly positively correlated in the low income group and it is moderately negatively correlated in the higher income group.

Acknowledgments. I am indebted to Ashwini Mathur and Eric Vittinghoff for computing the graphs in Section 9 and to Wolfgang Härdle for helping obtain the data used in this section. Jack Block and Per Gjerde contributed to the ideas in this paper. Peter Bickel suggested the relative local correlation interpretation of Section 5 and Jeff Wu suggested the interpretation in terms of conditional correlation rates. This research was partially supported by NSF grant DMS-8901603 and by the Norwegian Research Foundation for Science and Humanities. Part of the work was carried out while the author was visiting the Department of Statistics, Harvard University.

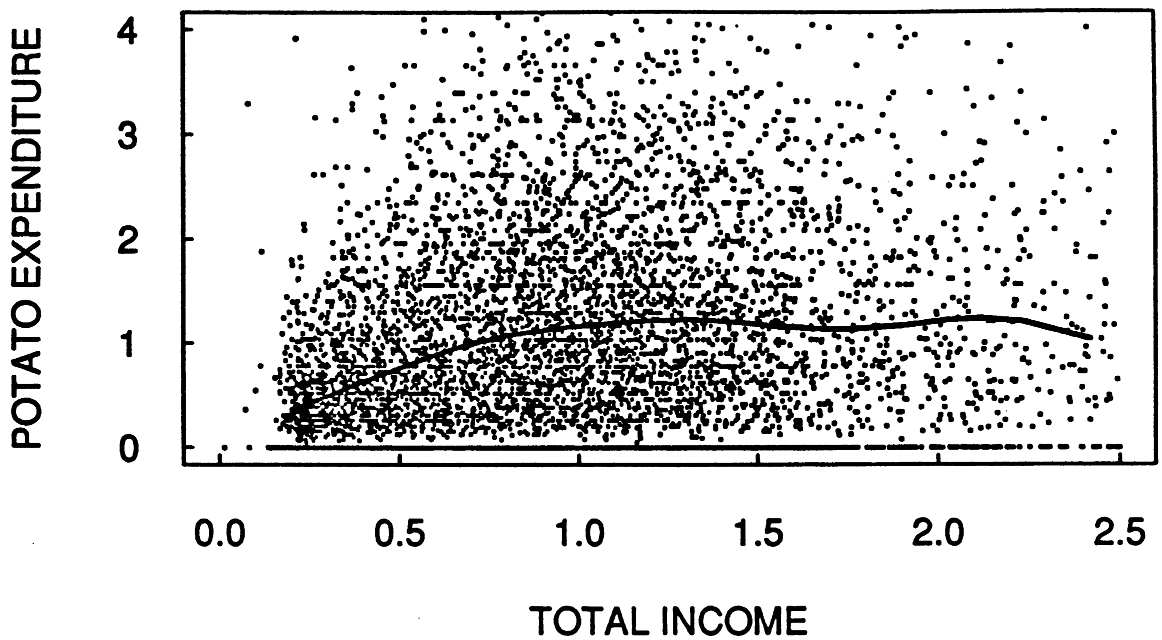

Figure 1. Data with estimated regression function

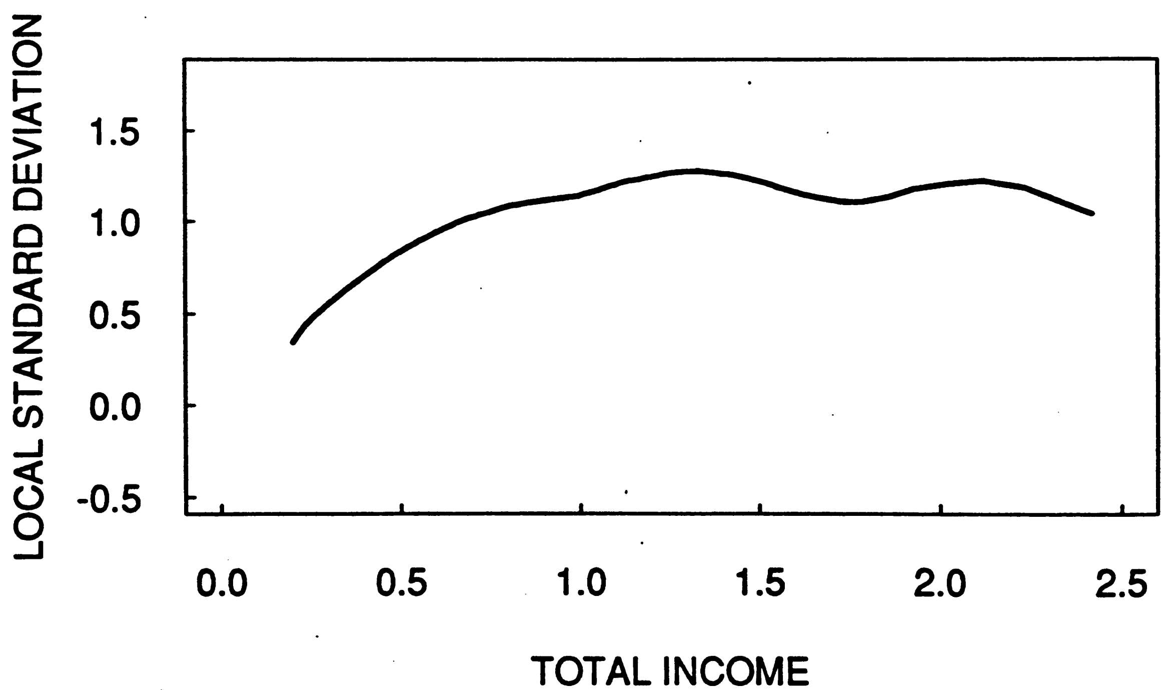

Figure 2. Estimated standard deviation

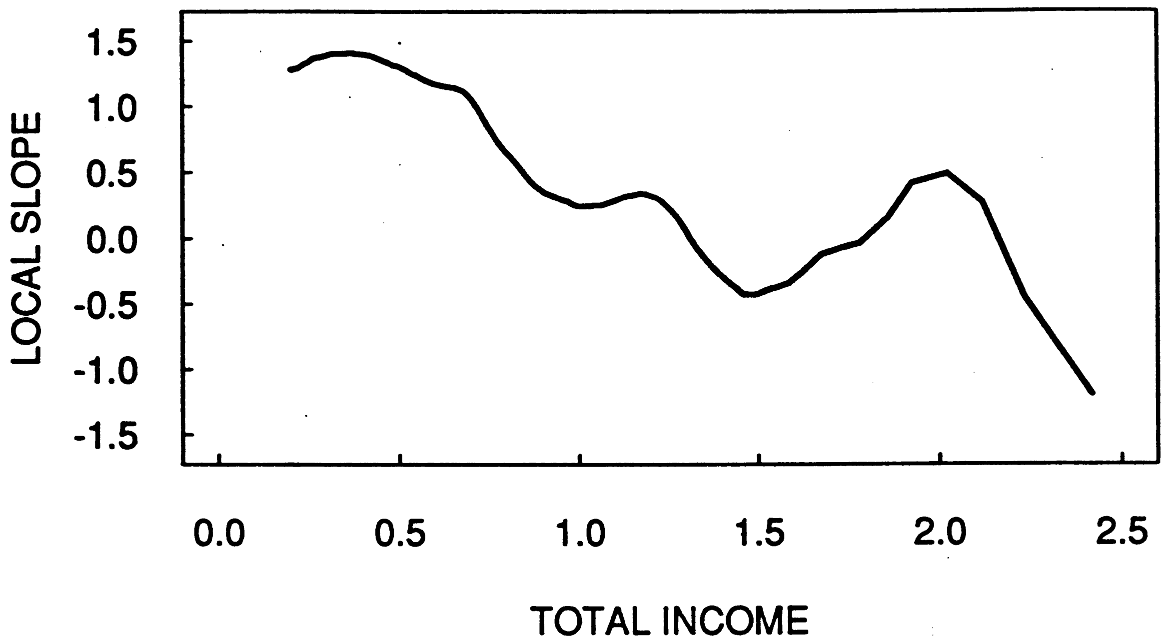

Figure 3. Estimated local slope

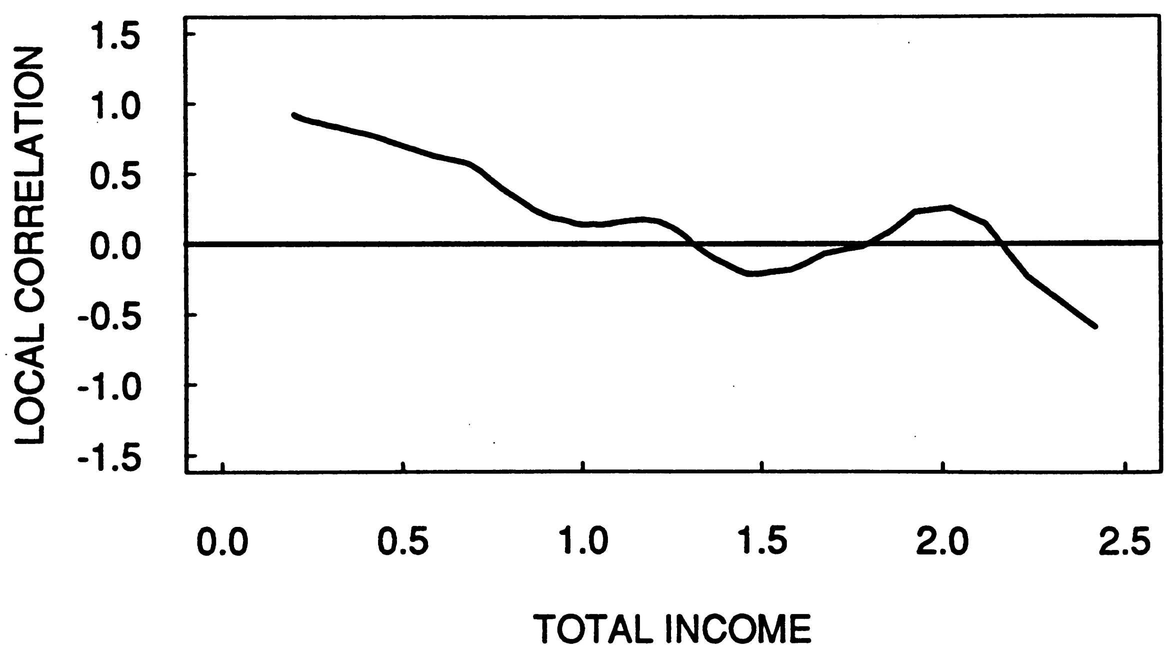

Figure 4. Estimated local correlation

10. REFERENCES

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

Bjerve, S. and Doksum, K. A. (1990). Correlation curves: Measures of association as functions of covariate values. Technical Report No. 280, Department of Statistics, U.C. Berkeley.

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

Doksum, K. A., Blyth, S., Bradlow, E. and Meng, X.-L. (1991). Correlation curves as local measures of variance explained by regression. Technical Report. Department of Statistics, Harvard University.

Doksum, K.A., and Samarov, A. (1991). Measures of association in nonlinear nonparametric regression models. Technical Report. Department of Statistics. Harvard University.

Family Expenditure Survey (1973). Department of Employment, Statistics Division, Her Majesty's Stationary Office, London; and ESRC Data Archive, University of Essex.

Fan, J. (1992). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, **20**. To appear.

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

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

Kendall, M. G., and Stuart A. (1962). *The Advanced Theory of Statistics*, Vol. 2. Griffin, London.

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. Hung.*, **10**, 441-451.

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. *Statist. Science*, **4**, 73-86.

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