Finance Letters, 2005, **3 (1)**, 64-76

How to Estimate Spatial Contagion between Financial Markets

Brendan O. Bradley [a] and Murad S. Taqqu [b, *]

[a] Acadian Asset Management Inc., USA
[b] Boston University, USA

Abstract

A definition of contagion between financial markets based on local correlation was introduced in Bradley and Taqqu (2004) and a test for contagion was proposed. For the test to be implemented, local correlation must be estimated. This paper describes an estimation procedure based on nonparametric local polynomial regression. The procedure is illustrated on the US and French equity market data.

Keywords: Contagion, Local correlation, Correlation breakdown, Crisis period
JEL classification: C12, C14

1. INTRODUCTION

There is no universally accepted definition of contagion in the financial literature. Typical definitions involve an increase in the cross-market linkages after a market *shock*. The linkage between markets is usually measured by a conditional correlation coefficient, and the conditioning event involves a short *post-shock* or *crisis* time period. Contagion is said to have occurred if there is a significant increase in the correlation coefficient during the crisis period. This phenomenon is also referred to as *correlation breakdown*. Statistically, correlation breakdown corresponds to a change in structure of the underlying probability distribution governing the behavior of the return series. Most tests for contagion attempt to test for such a change in structure, but these tests may be problematic. One difficulty was pointed out by Boyer, Gibson and Loretan (1999) who showed that the choice of conditioning event may lead to spurious conclusions. The reader is referred to Bradley and Taqqu (2004) for an extensive discussion. We proposed in that paper to use local correlation in order to measure contagion. The goal of the present article is to develop the statistical methodology behind such an approach. Applications can be found in (Bradley and Taqqu, 2005).

Suppose that X and Y represent the returns in two different markets. The local correlation provides a measure of dependence for the model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1}$$

where ε is mean zero, unit variance and is independent of X. Thus X affects Y in two ways: through the mean level $m(X)$ and through the standard deviation $\sigma(X)$ associated with the noise ε. If $m(X)$ is linear and $\sigma(X)$ equals the constant σ, one recovers the standard linear regression model

$$Y = \alpha + \beta X + \sigma\varepsilon, \tag{2}$$

where the correlation is

$$\rho = \mathrm{Corr}(X,Y) = \beta\frac{\sigma_X}{\sigma_Y} = \frac{\sigma_X \beta}{\sqrt{\sigma_X^2\beta^2 + \sigma^2}}. \tag{3}$$

---

* *Corresponding author*: Email: *murad@bu.edu* We would like to thank Ashis Gangopadhyay, Vladas Pipiras and Stilian Stoev for many valuable comments which led to an improved exposition of this material. This research was partially supported by the NSF Grants ANI-9805623 and DMS-0102410 at Boston University.

This last formula motivates the following definition of local correlation for the non-linear model (1).

Definition 1.1. *Let X and Y be two random variables with finite variance. The local correlation between Y and X at $X = x$ is given by*

$$\rho(x) = \frac{\sigma_X \beta(x)}{\sqrt{\sigma_X^2 \beta^2(x) + \sigma^2(x)}} \tag{4}$$

where σ_X denotes the standard deviation of X, $\beta(x) = m'(x)$ is the slope of the regression function $m(x) = \mathbb{E}(Y \mid X = x)$ and $\sigma^2(x) = \mathrm{Var}(Y \mid X = x)$ is the nonparametric residual variance.

The local correlation $\rho(x)$ was introduced by Bjerve and Doksum (1993). Since it measures the strength of dependence between X and Y at different points of the distribution of X, we can use it do define (spatial) contagion.

Definition 1.2. *Suppose that X and Y stand for the returns, over some fixed time horizon, of markets X and Y respectively. We say that there is contagion from market X to market Y if*

$$\rho(x_L) > \rho(x_M) \tag{5}$$

where $x_M = F_X^{-1}(0.5)$ is the median of the distribution $F_X(x) = \mathbb{P}\{X \le x\}$ of X and x_L is a low quantile of that distribution.

See Bradley and Taqqu (2004) for a detailed discussion of this definition and of the choice of x_L. Our goal here is to present the theory behind the estimation of $\rho(x_0)$ at a target point x_0. We shall use nonparametric curve estimation techniques to estimate $\rho(x_0)$. The procedure is illustrated on the US and French equity market data. Applications to contagion in financial markets and to flight of quality from the US equity market to the US government bond market can be found in the companion paper Bradley and Taqqu (2005). We make the software written in support of this work freely available and describe its use in the appendix of Bradley and Taqqu (2005).

2. ESTIMATION PROCEDURE

In order to estimate the local correlation measure $\rho(x_0)$ at a target point x_0 we assume that our observations (X_i, Y_i), $i = 1, \ldots n$, are an independent sample from a population (X, Y) [1] and we apply a method similar to those set forth in Bjerve and Doksum (1993) and Mathur (1998). The method consists of estimating the functions $m(x_0)$, $\beta(x_0)$ and $\sigma(x_0)$ through consecutive local polynomial regressions of degrees p_1 and p_2 at x_0. To obtain $\hat{\rho}(x_0)$. Bjerve and Doksum (1993) first use a local linear regression to estimate β with a bandwidth equal to the standard deviation σ_X which has no asymptotically optimal properties), then perform a local linear regression with a bandwidth selection based again on σ_X on the squared residuals to obtain an estimate of $\sigma^2(x_0)$. In contrast, we follow a suggestion of Mathur (1998):

(a) we apply a local quadratic regression to estimate $\beta(x_0)$ using an estimate of the asymptotically optimal bandwidth for that regression (this reduces the bias),

(b) apply a local linear regression on the squared residuals to estimate $\sigma^2(x_0)$ using again an estimate of the asymptotically optimal bandwidth appropriate for this regression (by using techniques developed by Ruppert et al. (1997),

(c) obtain $\hat{\rho}(x_0)$ and show that it is asymptotically normal.

---

[1] When dealing with practical applications, one can first filter the data for heteroscedasticity by assuming $X_i = \sigma_{X,i} \tilde{X}_i$ and $Y_i = \sigma_{Y,i} \tilde{Y}_i$ and perform the local correlation estimation procedure on $(\tilde{X}_i, \tilde{Y}_i)$.

See the monograph of Fan and Gijbels (1996) for details on local polynomial regression. Step (a) is developed in Section 3 and step (b) in Section 4. These steps require the specification of a bandwidth, which is done in Section 5. Step (c) is then presented in Section 5.1. We illustrate the estimation procedure for local correlation using the US and French equity market data in Section 7.

3. LOCAL POLYNOMIAL REGRESSION

Let $(X_i, Y_i), i = 1, \ldots, n$ be the return data for the US and French equity markets respectively. Let $X_p(x_0)$ be any target point at which we would like to know the local correlation $\rho(x_0)$. For our definition of contagion we will use the target points x_L and x_M from Definition 1.2 for x_0. We therefore require estimates of the local slope $\beta(x_0)$ and local residual variance $\sigma^2(x_0)$. To that end, assume the regression function $m(x)$ is $p+1$ times differentiable. Using a Taylor series expansion of the regression function about the target point x_0 we know that

$$m(x) \approx m(x_0) + m^{(1)}(x_0)(x - x_0) + \frac{m^{(2)}(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p. \tag{6}$$

This polynomial estimate of the regression function is fit locally at x_0 using weighted least squares regression. That is, the terms $m^{(k)}(x_0)/k!,\ k = 0, \ldots, p$ are estimated as the coefficients of the weighted least squares problem

$$\min_{(\beta_0(x_0), \ldots, \beta_p(x_0))} \sum_{i=1}^{n} \left\{ Y_i - \sum_{k=0}^{p} \beta_k(x_0)(X_i - x_0)^k \right\}^2 w_i(x_0, h), \tag{7}$$

which yield the estimators

$$\hat{m}^{(k)}(x_0) = k! \hat{\beta}_k(x_0). \tag{8}$$

The weights of the regression at x_0 are given by a kernel function

$$w_i(x_0, h) = K_h(X_i - x_0) = \frac{1}{h} K\left(\frac{X_i - x_0}{h} \right).$$

We will defer discussion of the choice of kernel function K and bandwidth h for the time being. The regression problem (7) may be rewritten in matrix notation. Let

$$X_p(x_0) = \begin{pmatrix} 1 & (X_1 - x_0) & \ldots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_1 - x_0) & \ldots & (X_n - x_0)^p \end{pmatrix}$$

be the $n \times (p+1)$ design matrix for the grid point x_0. Let

$$y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \beta(x_0) = \begin{pmatrix} \beta_0(x_0) \\ \beta_1(x_0) \\ \vdots \\ \beta_p(x_0) \end{pmatrix}$$

be the response and regression parameter vectors respectively. The local polynomial regression problem may then be written as

$$\min_{\beta(x_0)} (y - X_p(x_0)\beta(x_0))^T W_h(x_0)(y - X_p(x_0)\beta(x_0)), \tag{9}$$

where $W_h(x_0) = \text{diag}(w_1(x_0, h), \ldots, w_n(x_0, h))$. The solution to (9) is known to be given by

$$\hat{\beta}(x_0) = (X_p(x_0)^T W_h(x_0) X_p(x_0))^{-1} X_p(x_0)^T W_h(x_0) y. \tag{10}$$

Notice that the estimated value of the regression function at target point x_0 may be written as

$$\hat{m}(x_0) = e_1^T \hat{\beta}(x_0) = e_1^T (X_p(x_0)^T W_h(x_0) X_p(x_0))^{-1} X_p(x_0)^T W_h(x_0) y,$$

where $e_1 = (1,0,...,0)^T$. Because the observations are $\{(X_i, Y_i), i = 1, \ldots, n\}$ and we need to obtain the residuals, $\{\hat{r}_i = Y_i - \hat{Y}_i = Y_i - \hat{m}(X_i), i = 1, \ldots n\}$, we will need $\hat{m}$ evaluated at the observation points $X_1, \ldots, X_n$. Letting $\hat{m} = (\hat{m}(X_1), \ldots, \hat{m}(X_n))^T \in \mathbb{R}^n$ be the vector of estimated values of the regression function at the observed values $\mathbf{x} = (X_1, \ldots, X_n)$ we see that

$$\hat{m} = H_{p,h} y \tag{11}$$

for the smoother matrix $H_{p,h} \in \mathbb{R}^{nxn}$. Its $(i, j)^{th}$ entry is given by

$$(H_{p,h})_{i,j} = e_1^T (X_p(X_i)^T W_h(X_i) X_p(X_i))^{-1} X_p(X_i)^T W_h(X_i) e_j, \tag{12}$$

where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is a unit vector for the i^{th} position of the appropriate dimension. In (12), $X_p(X_i)$ denotes the matrix $X_p(x_0)$ with $x_0 = X_i$. The following result will be used in the sequel and is proved in Section 6.

Proposition 3.1. *The smoother matrix $H_{p,h}$ preserves constant vectors in the sense that $H_{p,h} 1 = 1$.*

Local polynomial regression, aside from being easy to implement, has two additional benefits for our problem. First, the local correlation $\rho(x_0)$ is a function of the local slope $\beta(x_0) = m'(x_0)$ of the regression function $m(x_0)$. By choosing the degree $p \geq 1$ of the polynomial fit in (6), local polynomial regression gives us an immediate estimate of the local slope

$$\hat{m}'(x_0) = \hat{\beta}_1(x_0) \tag{13}$$

in the regression equation (9). The second benefit of locally polynomial regression is a reduction in the bias of the estimated regression function and its derivatives at the boundaries of the support of the distribution of the covariate x. In classical kernel-based nonparametric regression methods, also called *locally constant regression*, the regression function $m(x) = \mathbb{E}(Y \mid X = x)$ is approximated by $m(x_0)$ for x close to x_0, ($p = 0$ in (6)) and $m(x_0)$ is estimated by

$$\hat{m}(x(_0) = \frac{\sum_{i=1}^n w_i(x_0, h) Y_i}{\sum_{i=1}^n w_i(x_0, h)} = \frac{\sum_{i=1}^n K_h(X_i - x_0) Y_i}{\sum_{i=1}^n K_h(X_i - x_0) Y_i},$$

that is, by a weighted average about the target point x_0. If the target point x_0 is near the boundary of the support X the weighted average may be strongly biased, even when the kernel has compact support, since more interior points than exterior points may be used in computing the local average. This bias may be reduced by fitting locally a polynomial in x_0 instead of a constant.

Using the local polynomial regression above, Fan and Gijbels (1996) show[2] that under certain non-restrictive regularity conditions the asymptotic conditional bias and variance of the local derivative estimator

[2] See Theorem 3.1 of Fan and Gijbels (1996) or Theorem 4.2 of Ruppert and Wand (1994). Its proof may be found in Ruppert and Wand (1994) or Fan and Gijbels (1996). The regularity conditions require that $f_X(x_0) > 0$ and that $f_X(\cdot)$, $m^{(p+1)}(\cdot)$ and $\sigma^2(\cdot)$ are continuous in a neighborhood of x_0. Additionally, we require that $n \to \infty$, $h \to 0$ such that $nh \to \infty$.

$\hat{m}^{(\upsilon)}(x_0), \upsilon \le p$ are given by

$$\text{Bias}\left(\hat{m}^{(\upsilon)}(x_0) \mid \boldsymbol{\mathcal{X}}\right) = e_{\upsilon+1}^T S^{-1} c_p \frac{\upsilon!}{(p+1)!} m^{(p+1)}(x_0) h^{p+1-\upsilon} + o_p(h^{p+1-\upsilon}) \tag{14}$$

$$\text{Var}\left(\hat{m}^{(\upsilon)}(x_0) \mid \boldsymbol{\mathcal{X}}\right) = e_{\upsilon+1}^T S^{-1} S^* S^{-1} e_{\upsilon+1} \frac{\upsilon! \sigma^2(x_0)}{f_X(x_0) n h^{1+2\upsilon}} + o_p\left(\frac{1}{nh^{1+2\upsilon}}\right). \tag{15}$$

Equation (14) for the conditional bias is valid for the case $p-\nu$ odd and has a slightly different form otherwise. For our purposes, we will use $p=2$ and $\nu=1$ and so we concentrate on this case. The vectors and matrices in the expressions for the bias and variance above are either constants or functions of the kernel function. To define them, let $\mu_j = \int u^j K(u)\,du$ and $\upsilon_j = \int u^j K^2(u)\,du$ be moments of K and K^2 respectively. Then the vector $c_p = (\mu_{p+1}, \dots, \mu_{2p+1})^T \in \mathbb{R}^{(p+1)}$ and the matrices $S \in \mathbb{R}^{(p+1)x(p+1)}$ and $S^* \in \mathbb{R}^{(p+1)x(p+1)}$ are given by $S = (\mu_{j+l})_{0 \le j,l \le p}$ and $S^* = (\upsilon_{j+l})_{0 \le j,l \le p}$.

Consistent with Bjerve and Doksum (1993), we choose the Epanechnikov kernel

$$K(u) = \frac{3}{4}\left(1-u^2\right)_+.$$

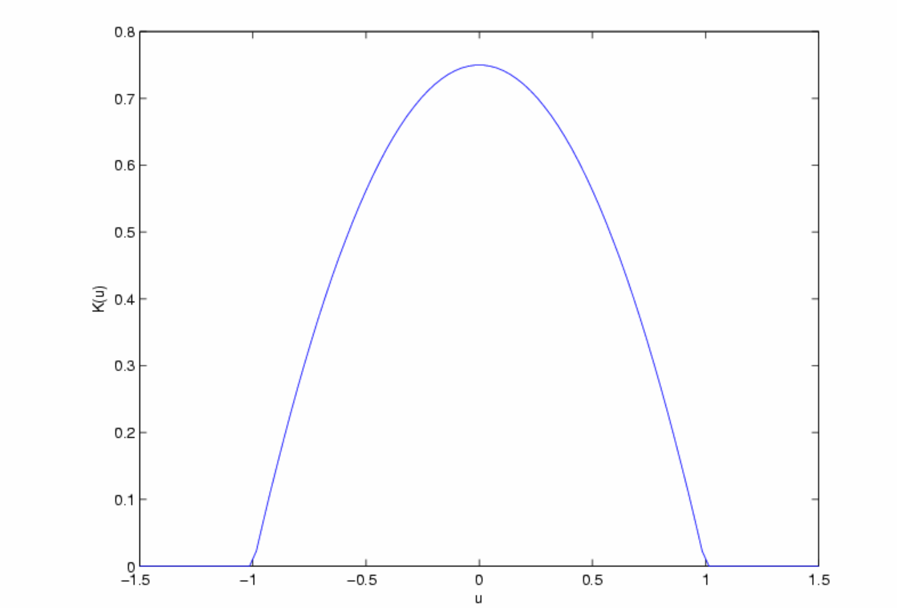

Figure 1. The Epanechnikov kernel $K(u) = \dfrac{3}{4}(1-u^2)_+$

The kernel is plotted in Figure 1. This choice of kernel is typical in local polynomial modelling. In fact, for local polynomial estimators it may be shown that the Epanechnikov kernel is optimal in the sense that for all choices of p and ν it minimizes the asymptotic mean squared error. See Theorem 3.4 of Fan and Gijbels (1996) for a more detailed discussion of this point[3].

The choice of the degree of the polynomial is typically taken to be $p = \nu + 1$. This choice gives a first order reduction in the bias of $\hat{m}^{(\upsilon)}$ without substantially increasing its variance. Since we are primarily concerned with reducing the bias of the local slope estimate we choose $\nu = 1$, $p = 2$ and the Epanechnikov kernel. This

[3]The proof may be found in Fan et al. (1997). The minimization is over all non-negative, symmetric and Lipschitz continuous functions.

yields

$$S=\begin{pmatrix}1 & 0 & 1/5\\ 0 & 1/5 & 0\\ 1/5 & 0 & 3/35\end{pmatrix},\quad S^{*}=\begin{pmatrix}3/5 & 0 & 3/35\\ 0 & 3/35 & 0\\ 3/35 & 0 & 1/35\end{pmatrix},$$

and $c_2=(0,3/35,0)^T$. For our problem, we are interested in the local slope, $\hat{\beta}(x_0)\equiv\hat{m}'(x_0)=\hat{\beta}_1(x_0)$, (see Definition1.1 and (8)). Applying (14) and (15), we obtain that the asymptotic conditional bias and variance of the local slope are given by

$$\text{Bias}\left(\hat{\beta}(x_0)\,|\,\boldsymbol{\mathcal{X}}\right)=\frac{1}{14}m^{(3)}(x_0)h^2+o_p(h^2) \tag{16}$$

$$\text{Var}\left(\hat{\beta}(x_0)\,|\,\boldsymbol{\mathcal{X}}\right)=\frac{15}{7}\frac{\sigma^2(x_0)}{f_X(x_0)nh^3}+o_p\left(\frac{1}{nh^3}\right). \tag{17}$$

In fact, Fan and Gijbels (1996) show that under certain non-restrictive regularity conditions, as the number of data points $n\to\infty$, the bandwidth $h\to 0$ and $nh\to\infty$, conditional on $\boldsymbol{\mathcal{X}}$, the above estimator of the local slope is asymptotically normal. Applying their Theorem 5.2, one gets

Theorem 3.1. *Suppose $\hat{\beta}$ is the estimator described above*[4]. *Suppose also that the following regularity conditions hold: $f_X(x)$, $m^{(3)}(x)$ and $(d/dx)\sigma^2(x)$ are continuous, the residual variance $\sigma^2(x)$ is positive and finite and $\mathbb{E}(Y^4\,|\,X=x)$ is bounded. Then for $f_X(x_0)>0$, we have*

$$\left(\frac{7f_X(x_0)nh^3}{15\sigma^2(x_0)}\right)^{1/2}[\hat{\beta}(x_0)-\beta(x_0)-(h^2m^3(x_0)/14+o(h^2)]\to\boldsymbol{\mathcal{N}}(0,1), \tag{18}$$

as $n\to\infty$, $h\to 0$ and $nh\to\infty$.

Observe that (18) involves the leading terms of the variance in (17) and of the bias in (16). Relationship (18) implies that if $h=o\left(n^{-1/7}\right)$, then

$$\left(\frac{7f_X(x_0)nh^3}{15\sigma^2(x_0)}\right)^{1/2}[\hat{\beta}(x_0)-\beta(x_0)]\to\boldsymbol{\mathcal{N}}(0,1) \tag{19}$$

and $\hat{\beta}(x_0)$ is asymptotically unbiased. Observe that asymptotic unbiasedness is not necessarily optimal because the asymptotic variance may be large. In section 5 we will choose a bandwidth h_1 for which $h_1=O(n^{-1/7})$ but which optimizes the bias-variance tradeoff.

4. RESIDUAL VARIANCE ESTIMATION

Our estimate of the local correlation in (4) still requires an estimate of the local residual variance $\sigma^2(x_0)$. The estimation procedure is similar to the one used above. It was first introduced by Mathur (1995) and its asymptotic properties were established by Ruppert et al. (1997). Let p_1 and h_1 denote the degree of the polynomial and bandwidth for the smooth of the *y* vector used above, namely the values of p and h used to get

---

[4] $\hat{\beta}$ is the local slope estimator of a local quadratic regression using the Epanechnikov kernel.

$\hat{m}$ (see (11)). Let $\hat{r} = (Y_1 - \hat{m}(X_1), \ldots, Y_n - \hat{m}(X_n))^T$ be the vector of estimated residuals from the above estimation of the regression function. Note that $\hat{r} = (I - H_{p_1,h_1})y$ for the smoother matrix H_{p_1,h_1} in (11). Following Fan and Yao (1998) and Ruppert et al. (1997) we propose to estimate $\sigma^2(x_0)$ in a manner analogous to a second smooth of the estimated squared residuals $\hat{r}^2$ by $H_{p_2,h_2}\hat{r}^2$. The matrix H_{p_2,h_2} here is as above with elements given by (12), but the values for the degree of the polynomial $p = p_2$ and bandwidth $h = h_2$ may be different from the values p_1 and h_1 used for $\hat{m}$.

A natural requirement is that the estimator $\hat{\sigma}^2(x)$ be unbiased in the case of homoscedastic regression error $\sigma^2(x) = \sigma^2$. As shown in Section 6, this implies the following proposition.

Proposition 4.1. *Let* $\hat{r} = (I - H_{p_1,h_1})y \in \mathbb{R}^{\mathrm{n}}$ *be the vector of residuals from an initial smooth* H_{p_1,h_1} *of the data and let* $H_{p_2,h_2} \in \mathbb{R}^{nxn}$ *be a second smoother matrix. If the residual variance* $\sigma^2(x)$ *is constant, that is* $\sigma^2(x) = \sigma^2$, *then*

$$\mathbb{E}(H_{p_2,h_2}\hat{r}^2 \mid \boldsymbol{\mathcal{X}}) = H_{p_2,h_2}[\mathrm{Bias}^2(\hat{m} \mid \boldsymbol{\mathcal{X}}) + \sigma^2(1+\Delta)], \tag{20}$$

where $\mathrm{Bias}(\hat{m} \mid \boldsymbol{\mathcal{X}}) = \mathbb{E}(\hat{r} \mid \boldsymbol{\mathcal{X}})$ *and*

$$\Delta = \mathrm{diag}(H_{p_1,h_1} H_{p_1,h_1}^T - 2H_{p_1,h_1}) \tag{21}$$

is the vector of diagonal elements of the matrix.

Recall from Proposition 3.1. that $H_{p,h}1 = 1$ for all polynomial smoothers $H_{p,h}$. If $\hat{m}$ is unbiased for m then $\mathbb{E}(H_{p_2,h_2}\hat{r}^2 \mid \boldsymbol{\mathcal{X}}) = \sigma^2(1 + H_{p_2,h_2}\Delta)$, which suggests the following estimator for the residual variance:

$$\hat{\sigma}^2 = \frac{H_{p_2,h_2}\hat{r}^2}{1 + H_{p_2,h_2}\Delta}, \tag{22}$$

where multiplication and division are taken componentwise. The estimator is unbiased at each of the observation points X_i, that is $\mathbb{E}(\hat{\sigma}^2(X_i) \mid \boldsymbol{\mathcal{X}}) = \sigma^2$.

Even though our estimator $\hat{m}$ is biased (see (14)), the estimator of the residual variance at the observation points $X_1, \ldots, X_n$ given by (22) and the structure of the smoother matrix given by (12) motivate the following residual variance estimator at target point x_0:

$$\hat{\sigma}^2(x_0) = \frac{e_1^T(X_{p_2}(x_0)^T W_{h_2}(x_0)(X_{p_2}(x_0))^{-1} X_{p_2}(x_0)^T W_{h_2}(x_0)\hat{r}^2}{1 + e_1^T(X_{p_2}(x_0)^T W_{h_2}(x_0)(X_{p_2}(x_0))^{-1} X_{p_2}(x_0)^T W_{h_2}(x_0)\Delta}. \tag{23}$$

Recall the vectors $\hat{r}$ and Δ are functions of the degree p_1 of the initial polynomial fit with bandwidth h_1. The asymptotic properties of the estimator (23) are established in Theorem 2 of Ruppert et al. (1997). They show, under certain regularity conditions and for p_2 odd, that if

$$[h_1^{2(p_1+1)} + (nh_1)^{-1}] = o\left(h_2^{(p_2+1)}\right) \tag{24}$$

then

$$\hat{\sigma}^2(x_0) = \sigma^2(x_0) + O_p\left(h_2^{(p_2+1)} + (nh_2)^{-1/2}\right). \tag{25}$$

A similar result hold for p_2 even. We will use this result in Section 5.1, along with the asymptotic normality result (18) from above, to show asymptotic normality of our estimator of local correlation. When estimating the residual variance we take $p_2 = 1$.

5. CHOICE OF BANDWIDTH

In order to carry out the local regressions of degrees $p_1 = 2$ and $p_2 = 1$ described above we need to choose the appropriate bandwidths h_1 and h_2. The choice of bandwidth is crucial to local polynomial modelling. A bandwidth too small results in our under-smoothing the data. Since in this case only data points X_i close to the target point x_0 are used in the fit, the resulting estimator has a small bias but large variance. When the bandwidth is too large, we have an over-smoothing of the data and an estimator with small variance but large bias. This is the typical bias versus variance tradeoff in statistics. We will use a data-driven bandwidth selection rule from Section 4.2 of Fan and Gijbels (1996) which, as we will see, is asymptotically optimal in the sense that it minimizes the weighted Mean Integrated Squared Error (MISE)

$$\int [\mathrm{Bias}^2(\hat{m}^{(\upsilon)}(x)) + \mathrm{Var}(\hat{m}^{(\upsilon)}(x))]\tilde{w}(x)dx \tag{26}$$

for some non-negative weight function $\tilde{w}(x)$. Equations (14) and (15) give the asymptotic conditional bias and variance of $\hat{m}^{(\upsilon)}$ respectively as a function of bandwidth h. Expressing (26) as MISE*(h)*, we get MISE*(h)* $= ah^{2p+2-2\upsilon} + bn^{-1}h^{-1-2\upsilon} + o_p\left(h^{2(p+1-\upsilon)} + (nh^{1+2\upsilon})^{-1}\right)$ for some constants a and b. This implies that the optimal choice of bandwidth is $h_{\mathtt{opt}} = O\left(n^{-1/(2p+3)}\right)$. In fact, it is straightforward to verify that (14) and (15) imply

$$h_{\mathtt{opt}} = C_{\nu,p}(K)\left[\frac{\int \sigma^2(x)\tilde{w}(x)/f_X(x)dx}{\int \left\{m^{(p+1)}(x)\right\}^2 \tilde{w}(x)dx}\right]^{1/(2p+3)} n^{-1/(2p+3)}. \tag{27}$$

The constant $C_{\nu,p}(K)$ is a function of the kernel K, the degree of fit p and the order of the derivative ν. It is given by

$$C_{\nu,p}(K) = \left[\frac{(p+1)!^2(2\nu+1)\int K_\nu^*(t)dt}{2(p+1-\nu)\left\{\int t^{p+1}K_\nu^*(t)dt\right\}^2}\right]^{-1/(2p+3)}$$

where $K_\upsilon^*(t) = e_{\upsilon+1}^T S^{-1}(1, t, \ldots, t^p)^T K(t)$ (see Section 3.2.2 of Fan and Gijbels, 1996). K_ν^* is called the equivalent kernel.

The optimal bandwidth $h_{\mathtt{opt}}$ in (27) depends on unknown quantities and must be estimated. In fact, one must do this before going through the steps described above in Sections 3 and 4. In order to estimate $h_{\mathtt{opt}}$, we start with a preliminary and rough estimators $\breve{m}(x)$ for $m(x)$ and $\breve{\sigma}^2(x)$ for $\sigma^2(x)$. This is because our goal here is not to estimate the parameters $m(x)$ and $\sigma^2(x)$, but only to obtain an estimate of the optimal bandwidth. We obtain $\breve{m}(x)$ by fitting a polynomial of order $p+3$ to $m(x)$[5]. This is done using a global least squares, that is, by choosing the α_k, $k = 0, \ldots, p+3$ which minimize $\sum_{i=1}^n \left(Y_i - \sum_{k=0}^{p+3} \alpha_k X_i^k\right)^2$. This yields the estimator $\breve{m}(x) = \breve{\alpha}_0 + \breve{\alpha}_1 x + \cdots + \breve{\alpha}_{p+3}x^{p+3}$. The estimator for the $p+1$ st derivative of the regression function is then given by

$$\breve{m}^{(p+1)}(x) = (p+1)!\breve{\alpha}_{p+1} + (p+2)!\breve{\alpha}_{p+2}x + \frac{(p+3)!}{2!}\breve{\alpha}_{p+3}x^2.$$

The residuals $Y_i - \breve{m}(X_i)$ of this fit are used to obtain the usual global sample variance estimator $\breve{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \breve{m}(X_i))^2$ for σ^2. Now let $w_0(x)$ be some specified weight function. After the change of variables $\tilde{w}(x) = w_0(x) f_X(x)$ and assuming a constant residual variance $\breve{\sigma}^2$ the optimal bandwidth (27) may be written

[5] We use a $p+3$ degree fit in order to obtain a quadratic fit for the $p+1$ *st* order derivative of m.

$$h = C_{\upsilon,p}(K) = \left[\frac{\breve{\sigma}^2 \int w_0(x)dx}{n \int \left\{ m^{(p+1)}(x) \right\}^2 w_0(x) f(x) dx} \right]^{1/(2p+3)} \tag{28}$$

The denominator of (28) may be estimated by $\sum_{i=1}^{n} \left\{ \breve{m}^{(p+1)}(X_i) \right\}^2 w_0(X_i)$ which yields the estimator

$$\hat{h}_{opt} = C_{\upsilon,p}(K) \left[\frac{\breve{\sigma}_2 \int w_0(x)dx}{\sum_{i=1}^{n} \left\{ \breve{m}^{(p+1)}(X_i) \right\}^2 w_0(X_i)} \right]^{1/(2p+3)} \tag{29}$$

We choose w_0 to give equal weight to all data points in the central 95% of the empirical distribution of X .

5.1. Asymptotic Normality of the Local Correlation Estimator

The estimation procedure outlined above results in an estimator of the local correlation of the form

$$\hat{\rho}(x_0) = \frac{s_X \hat{\beta}(x_0)}{\sqrt{s_X^2 \hat{\beta}^2(x_0) + \hat{\sigma}^2(x_0)}}. \tag{30}$$

The estimator $s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is the sample estimator of variance σ_X^2 . Recall that the estimator $\hat{\beta}(x_0)$ is the result of the local quadratic regression using a bandwidth $h_1 = O\left(n^{-1/(2p_1+3)}\right)$ (see (27)) with $p_1 = 2$, that is $h_1 = O\left(n^{-1/7}\right)$. The estimator $\hat{\sigma}^2(x_0)$ is given in (23) and is the result of a local linear regression ($p_2 = 1$) with bandwidth $h_2 = O\left(n^{-1/(2p_2+3)}\right) = O\left(n^{-1/5}\right)$. Notice that in this case Relation (24) holds and therefore so does (25). In fact, the following result holds.

Theorem 5.1. *Suppose that (i) x_0 is an interior point of the support of $f_X(x)$, (ii) $m(x)$ has 4 continuous derivatives in a neighborhood of x_0, (ii) $\sigma^2(x)$ has 3 continuous derivatives in a neighborhood of x_0, (iii) $f_X(x)$ and $\sigma^4(x)$ are differentiable in a neighborhood of x_0 where the innovations ε in (1) have finite fourth moment, (iv) the local regressions are performed with $p_1 = 2$ and $p_2 = 1$ and (v) $h_1 \to 0$, $h_2 \to 0$, $nh_1 \to \infty$, $nh_2 \to \infty$ such that $h_1 = o(n^{-1/7})$ and $h_2 = o(n^{-1/5})$. Then for the estimators described in (30) above, we have*

$$\left[\frac{7 f_X(x_0) n h_1^3}{15 \sigma^2{}_X} \right]^{1/2} \left[1 - \rho(x_0)^2 \right]^{-3/2} \left[\hat{\rho}(x_0) - \rho(x_0) \right] \to \mathcal{N}(0,1). \tag{31}$$

The proof is given in Section 6.

Equations (31) and (18) relate the conditional asymptotic variance of the estimator $\hat{\rho}(x_0)$ to that of $\hat{\beta}(x_0)$:

$$\mathrm{Var}(\hat{\rho}(x_0) \,|\, \boldsymbol{\mathcal{X}}) = \mathrm{Var}(\hat{\beta}(x_0) \,|\, \boldsymbol{\mathcal{X}}) \ \frac{\sigma^2{}_X}{\sigma^2(x_0)} \left[1 - \rho(x_0)^2 \right]^3. \tag{32}$$

Let $\hat{\sigma}^2_{\hat{\rho}(x_0)}$ and $\hat{\sigma}^2_{\hat{\beta}(x_0)}$ denote estimators of the conditional variance of $\hat{\rho}(x_0)$ and $\hat{\beta}(x_0)$ respectively. Relation (10) gives $\hat{\beta}(x_0)$ and implies that its conditional covariance matrix is given by

$$\begin{aligned}\mathrm{Cov}(\hat\beta(x_0)\mid \boldsymbol{\mathcal{X}}) &= \mathrm{Cov}\left(\left(X_p(x_0)^T W_h(x_0) X_p(x_0)\right)^{-1} X_p(x_0)^T W_h(x_0) y \mid \boldsymbol{\mathcal{X}}\right)\\ &= \left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} X_{p1}^T W_{h1} \mathrm{Cov}\left(y\mid \boldsymbol{\mathcal{X}}\right) W_{h1} X_{p1}\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1}\\ &= \left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} X_{p1}^T \Sigma X_{p1}\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} \end{aligned} \tag{33}$$

where the dependence of X_{p_1} and W_{h_1} on x_0 has been dropped and $\Sigma = \mathrm{diag}\,(w_i^2(x_0,h_1)\sigma^2(X_i))$ for $i=1,\ldots,n$. Since $\sigma^2(X_i)$ is unknown, instead of estimating it as in (23), it is sufficient in this context to estimate it by $\hat\sigma^2(x_0)$, that is to assume that it is locally homoscedastic. Hence Σ is estimated by $\mathrm{diag}(w_i^2(x_0,h_1)\hat\sigma^2(X_i))$ and

$$\hat\sigma^2_{\hat\beta(X_0)} = e_2^T\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} X_{p1}^T W_{h1}^2 X_{p1}\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} e_2 \hat\sigma^2(x_0). \tag{34}$$

The vectors e_2 pick off the second diagonal element of the covariance matrix of $\hat\beta(x_0)$ since this is the term related to the local slope $\beta(x_0)$ of the local regression. In view of (32) this gives the following estimator of the conditional variance of $\hat\rho(x_0)$:

$$\begin{aligned}\hat\sigma^2_{\hat\rho(x_0)} &= \hat\sigma^2_{\hat\beta(x_0)} \frac{s_X^2}{\hat\sigma^2(x_0)}\left[1-\hat\rho(x_0)^2\right]^3\\ &= e_2^T\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} X_{p1}^T W_{h1}^2 X_{p1}\left(X_{p1}^T W_{h1} X_{p1}\right)^{-1} e_2 s_X^2\left[1-\hat\rho(x_0)^2\right]^3, \end{aligned} \tag{35}$$

which does not involve $\hat\sigma^2(x_0)$ anymore.

6. PROOFS

Proof of Proposition 3.1. The form of the $i,\, j^{\text{th}}$ element of the smoother matrix $H_{p,h}$ is

$$(H_{p,h})_{i,j} = e_1^T (X_p(X_i)^T W_h(X_i) X_p(X_i))^{-1} X_p(X_i)^T W_h(X_i) e_j .$$

Let H_i represent the i^{th} row of $H_{p,h}$. Dropping for notational simplicity the dependence on the target point X_i, we get

$$H_i = e_1^T\left(X_p^T W_h X_p\right)^{-1} X_p^T W_h .$$

It suffices to show that $H_i 1 = 1$, that is, $H_i(1,...,1) = 1$.

Let $S_n = X_p^T W_h X_p \in \mathbb{R}^{(p+1)x(p+1)}$. Then $H_i = e_1^T S_n^{-1} X_p^T W_h$ where the matrix S_n has the form

$$(S_n)_{i,j} = S_{n,i+j-2}, \quad 1\le i,j\le p+1, \quad \text{where } S_{n,k} = \sum_{l=1}^{n} \underbrace{K_h(X_l - X_i)}_{w_l}(X_l - X_i)^k .$$

Since on the one hand,

$$X_p^T W_h 1 = \begin{pmatrix} 1 & \cdots & 1\\ X_1 - X_i & \cdots & X_n - X_i\\ \vdots & & \vdots\\ (X_1-X_i)^p & \cdots & (X_n - X_i)^p \end{pmatrix}\begin{pmatrix} \mathrm{w}_1\\ \mathrm{w}_2\\ \vdots\\ \mathrm{w}_n\end{pmatrix} = \begin{pmatrix} S_{n,0}\\ S_{n,1}\\ \vdots\\ S_{n,p}\end{pmatrix},$$

and on the other,

$$S_n e_1 = \begin{pmatrix} S_{n,0} & S_{n,1} & \cdots & S_{n,p} \\ S_{n,1} & S_{n,2} & \cdots & S_{n,p+1} \\ \vdots & \vdots & & \vdots \\ S_{n,p} & S_{n,p+1} & \cdots & S_{n,2p} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} S_{n,0} \\ S_{n,1} \\ \vdots \\ S_{n,p} \end{pmatrix},$$

we have

$$X_p^T W_h 1 = S_n e_1 .$$

This implies that

$$H_i 1 = e_1^T S_n^{-1} X_p^T W_h 1 = e_1^T S_n^{-1} S_n e_1 = e_1^T e_1 = 1 ,$$

which concludes the proof. □

Proof of Proposition 4.1. Assume all vector multiplications, including powers, are taken component-wise. Let $m = (m(X_1), \dots, m(X_n))^T$, $\hat{m} = H_{p1,h1} y$, $\sigma^2 = (\sigma^2(X_1), \dots, \sigma^2(X_n))^T = \sigma^2 1$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_\mathrm{n})^T$ and let $\hat{r}^2 = (y - \hat{m})^2$. Then

$$\begin{aligned} \mathbb{E}\left(\hat{r}^2 \mid \mathcal{X}\right) &= \mathbb{E}\left([y-\hat{m}]^2 \mid \mathcal{X}\right) \\ &= \mathbb{E}\left([m+\sigma\varepsilon-\hat{m}]^2 \mid \mathcal{X}\right) \\ &= \mathbb{E}\left([m-\hat{m}]^2 + \sigma^2\varepsilon^2 + 2\sigma\varepsilon m - 2\sigma\varepsilon\hat{m} \mid \mathcal{X}\right) \\ &= \mathrm{MSE}\left(\hat{m} \mid \mathcal{X}\right) + \sigma^2 1 - 2H_{p1,h1}\,\mathbb{E}\left(\sigma m\varepsilon + \sigma^2\varepsilon^2 \mid \mathcal{X}\right) \end{aligned}$$

since $\mathbb{E}(\varepsilon^2) = 1$, $\mathbb{E}(\varepsilon) = 0$ and $\hat{m} = H_{p1,h1} y = H_{p1,h1}(m + \sigma\varepsilon)$. hence

$$\mathbb{E}\left(\hat{r}^2 \mid \mathcal{X}\right) = \mathrm{Bias}^2\left(\hat{m} \mid \mathcal{X}\right) + \mathrm{Var}\left(\hat{m} \mid \mathcal{X}\right) + \sigma^2(1 - 2\mathrm{diag}(H_{p1,h1})).$$

Now, note that $\mathrm{Var}\left(\hat{m} \mid \mathcal{X}\right) = \mathrm{diag}\left(\mathrm{Cov}\left(\hat{m} \mid \mathcal{X}\right)\right) = \mathrm{diag}\left(H_{p1,h1}\,\mathrm{Cov}\left(y, \mid \mathcal{X}\right) H_{p1,h1}^T\right) = \sigma^2 \mathrm{diag}\left(H_{p1,h1} H_{p1,h1}^T\right)$ since, by assumption, $\mathrm{Cov}\left(y, \mid \mathcal{X}\right) = \sigma^2 I_n$. Letting $\Delta = \mathrm{diag}\left(H_{p1,h1} H_{p1,h1}^T - 2H_{p1,h1}\right)$, we get

$$\mathbb{E}\left(\hat{r}^2 \mid \mathcal{X}\right) = \mathrm{Bias}^2\left(\hat{m} \mid \mathcal{X}\right) + \sigma^2(1 + \Delta)$$

and the result follows. □

Proof of Theorem 5.1. First note that the regularity conditions of Theorem 5.1 are those of Theorem 3.1 and of Theorem 2 of Ruppert et al. (1997). We have

$$\hat{\rho}(x_0) = \frac{s_X \hat{\beta}(x_0)}{\sqrt{s_X^2 \hat{\beta}^2(x_0) + \hat{\sigma}^2(x_0)}} \stackrel{\text{def}}{=} g(\theta_n),$$

where $\theta_n = (s_X, \hat{\beta}(x_0), \hat{\sigma}^2(x_0))^T$. Observe that $\rho(x_0) = g(\theta_0)$ where $\theta_0 = (s_X, \beta(x_0), \sigma^2(x_0))^T$. Expanding $g(\theta_n)$ in a Taylor series about θ_0 we get

$$\hat{\rho}(x_0) - \rho(x_0) = \left[1 - \rho^2(x_0)\right]^{3/2} \left\{ \underbrace{\frac{\beta(x_0)}{\sigma(x_0)}[s_X - \sigma_X]}_{I} + \underbrace{\frac{\sigma_X}{\sigma(x_0)}\left[\hat{\beta}(x_0) - \beta(x_0)\right]}_{II} - \dots\dots \right.$$

$$\left. \underbrace{\frac{\sigma_X \beta(x_0)}{\sigma^{3/2}(x_0)}\left[\hat{\sigma}^2(x_0) - \sigma^2(x_0)\right]}_{\mathrm{III}} \right\} + R(\theta_n - \theta), \tag{36}$$

and so

$$\left[1-\rho^2(x_0)\right]^{-3/2}\left[\hat{\rho}(x_0)-\rho(x_0)\right]=\mathrm{I}+\mathrm{II}-\mathrm{III}+R(\theta_n-\theta). \tag{37}$$

When multiplied by

$$r_n(x_0) \stackrel{\text{def}}{=} \left[\frac{7 f_X(x_0) n h_1^3}{15\sigma^2{}_X}\right]^{1/2}, \tag{38}$$

only term II contributes to the asymptotics. Indeed, $h_1 = o(n^{-1/7})$ implies $r_n(x_0) = o(n^{2/7})$. For term *I* we know that $n^{1/2}\left[s_X-\sigma_X\right]\to\mathcal{N}(0,V(X))$ and so $r_n(x_0)\left[s_X-\sigma_X\right]=o_p(1)$.The asymptotics of term III are determined by equation (25). For $p_2=1$ and $h_2=O(n^{-1/5})$ we have that $\hat{\sigma}^2(x_0)-\sigma^2(x_0)=O_p(n^{-2/5})$.This implies that $r_n(x_0)\left[\hat{\sigma}_2(x_0)-\sigma^2(x_0)\right]=o_p(1)$. The contribution of term II is determined as follows. In view of (38), equation (19), which is an immediate corollary of Theorem 3.1, implies that

$$r_n(x_0)\frac{\sigma_X}{\sigma(x_0)}\left[\hat{\beta}(x_0)-\beta(x_0)\right]\to\mathcal{N}(0,1).$$

The remainder term *R* is handled in the usual way. Note that

$$R(\theta_n-\theta_0)=g(\theta_n)-g(\theta_0)-\nabla g(s)^T\big|_{s=\theta_0}(\theta_n-\theta_0). \tag{39}$$

By the differentiability of g at θ we know that $R(\theta_n-\theta_0)=o_p(\|\theta_n-\theta_0\|)$. Now, since $r_n(x_0)(\theta_n-\theta_0)$ converges in distribution, it is uniformly tight (Prohorov's theorem). Multiplying both sides of (39) by $r_n(x_0)$, this implies that $r_n(x_0)R(\theta_n-\theta)=o_p\left(r_n(x_0)\|\theta_n-\theta\|\right)$. The tightness of $r_n(x_0)(\theta_n-\theta)$ implies that $r_n(x_0)\|\theta_n-\theta\|=O_p(1)$ and since $o_p(O_p(1))=o_p(1)$, the theorem follows. □

7. ILLUSTRATION

Figure 2 illustrates the procedure for French equity returns Y as a function of the US equity returns X. The data are described in Bradley and Taqqu (2005). The procedure is applied to 101 equidistant target points x_0 located in the central 95% of the empirical distribution of the US equity returns. The correlation curve plot shows a clear increase in the local correlation between the French and US equity markets as the US market does poorly. That is, when the US market is doing badly (negative x_0), the corresponding local correlation is high. Additionally, the plots indicate an increase in both the local slope $\hat{\beta}(x_0)$ and local residual standard deviation $\hat{\sigma}(x_0)$. In this case, the increase in the local residual standard deviation is not sufficient to overcome the increase in the local slope and the local correlation increases as a result. Had the model been $Y=m(X)+\varepsilon$ instead of (1), then the residual standard deviation $\hat{\sigma}(x)$ would be assumed constant and the large increase in the local slope $\hat{\beta}(x)$ would have contributed (recall the definition of local correlation in (4)) to a large increase in the local correlation $\hat{\rho}(x)$. That increase, which would not have been mitigated by the increase in $\hat{\sigma}(x)$ (now assumed constant), would have been dramatic and spurious. However, in accordance with our intuition, we see that the residual variance $\hat{\sigma}(x)$ is roughly an increasing function of $|x|$, the absolute value of the returns of the US equity market. That is, conditional upon large (absolute value) returns *x* in the US market, the variance of the French market increases as $|x|$ increases.

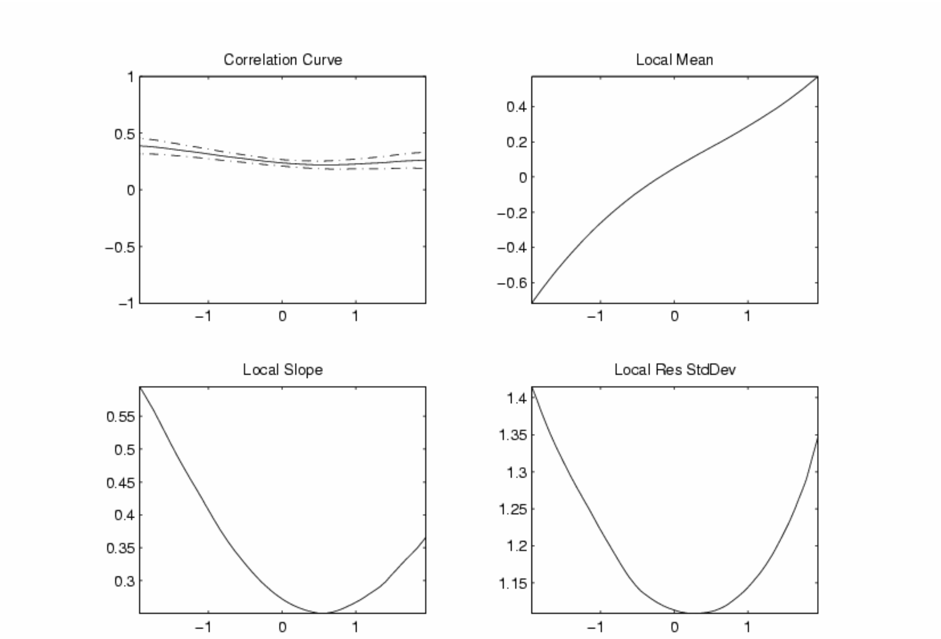

Figure 2. The correlation curve, local mean, slope, and residual standard deviation for the French equity market as a function of the (log) returns, expressed as a percent, of the US equity market.

95% confidence levels are attached using normality of the estimator and equation (35).

REFERENCES

Bjerve, S. and K. Doksum (1993) Correlation curves: measures of association as functions of covariate values, *Annals of Statistics*, **21**, 890-902.

Boyer, B., M. Gibson and M. Loretan, M. (1999) Pitfalls in tests for changes in correlations, *Technical Report*, Board of Governors of the Federal Reserve System, International Finance Discussion Paper.

Bradley, B. and M. Taqqu (2004) Framework for analyzing spatial contagion between financial markets, *Finance Letters,* **2 (6)***,* 8-15.

Bradley, B. and M. Taqqu (2005) Empirical evidence on spatial contagion between financial markets, *Finance Letters,* **3 (1)**, to appear.

Fan, J., T. Gasser, I. Gijbels, M. Brockmann and J. Engel (1997) Local polynomial regression: optimal kernels and asymptotic minimax efficiency, *Annals of the Institute of Mathematical Statistics,* **49**, 79-99.

Fan, J. and I. Gijbels (1996) *Local polynomial modelling and its applications*, Chapman & Hall.

Fan, J., I. Gijbels, T. Hu and L. Huang (1996) A study of variable bandwidth selection for local polynomial modelling, *Statistica Sinica,* **6(1)**, 113--127.

Fan, J. and Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression', *Biometrika,* **85**, 645-660.

Mathur, A. (1995) On the estimation of residual variance function, *Paper presented at the Joint Statistical Meetings*.

Mathur, A. (1998) Partial correlation curves, *PhD Thesis*, University of California, Berkeley.

Ruppert, D. and M. Wand (1994) Multivariate locally weighted least squares regression, *Annals of Statistics,* **22 (3)**, 346-1370.

Ruppert, D., M. Wand, U. Holst and O. Hössjer (1997) Local polynomial variance function estimation, *Technometrics,* **39**, 262-273.