University of British Columbia
Department of Statistics
Technical Report #184

Comparing the Shapes of Regression Functions

Nancy E. Heckman
Ruben H. Zamar
University of British Columbia
Vancouver BC

January 1999

Abstract

Does a regression function follow a specified shape? Do two regression functions have the same shape? These questions can occur when testing for monotonicity, when counting local maxima, or when studying variation in families of curves. One can address these questions by considering the rank correlation coefficient between two functions. This correlation is a generalization of the rank correlation between two finite sets of numbers and is equal to one if and only if the two functions have the same shape. A sample rank correlation based on smoothed estimates of the regression functions consistently estimates the true correlation. This sample rank correlation can be used as a measure of simlarity between functions in cluster analysis, and as a measure of monotonicity or modality in testing. The manuscript is available as a UBC Statistics Department Technical Report (www.stat.ubc.ca).

Keywords: rank correlation, nonparametric regression, bump-hunting, monotone functions

Acknowledgments. This research is supported by the Natural Sciences and Engineering Research Council grant numbers A7969 and A9276.

1 Introduction

The shape of a function is often of key interest in regression analysis. For instance, local maxima, or "bumps", are scientifically important features in the shape of a regression function. In evolutionary biology, an individual's chance of survival typically depends on the value of a physical trait. If the chance of survival attains a global maximum, then the trait will evolve toward the optimal value. If there are two local maxima, then the trait will evolve differently, with two commonly occurring values emerging. Thus the shape of the curve determines the type of evolution.

The presence of bumps can raise scientific questions. Growth curves of U.S. children clearly show two local maxima in the rate of growth. In Swiss children, however, these growth spurts seem to be either missing or less prominent. (See Ramsay, Bock, and Gasser, 1995.) Are the Swiss and U.S. growth rates similar in shape? If not, how and why do their shapes differ?

Economists often consider the logarithm of a quantity because its derivative gives a relative rate of change or "elasticity". Roy (1998) has considered a panel data study in rural South India of the calorie intake of men and women as a function of income. She finds that the derivative of log intake with respect to income (income elasticity of calorie intake) for men seems to be decreasing, while that for women increases till about the mean household income and then declines. Thus the policy implication is different depending on which is the target group. Roy could study the shape of the elasticity via the methods proposed here or via Harezlak's (1998) proposed test for the number of local maxima in a regression function or its derivatives.

In some applications, it's important to decide if a function is monotone. For example, in calibrating two measurement methods one measures a set of objects by each method. If the corresponding scatterplot of the pairs of measurements is monotone, one measurement method can be substituted for the other. There are already several methods for estimating a monotone regression function (see, e.g., Friedman and Tibshirani, 1984, Mukerjee, 1988, Ramsay, 1988, and Mammen, 1991) but few methods for testing for monotonocity. Bowman, Jones and Gijbels (1998) and Harezlak (1998) propose a test for monotonicity with test statistic equal to a smoothing parameter. Hall and Heckman (1998) propose a test based on the minimum of the slopes of least squares lines, fit over all subintervals of the data.

We are particularly interested in comparing the shapes of regression functions. A common measure of similarity between two functions x and y is the L_2 distance, $\int [x(t) - y(t)]^2 \, dt$. However, this distance doesn't always adequately measure what our eye sees. Marron and Tsybakov (1995) recognized this problem and proposed alternate measures of similarity. Here, we propose a simpler measure. To define this measure suppose that x and y are two functions defined on the unit interval. We'll say that x and y have the same shape if there exists a strictly increasing function g such that $x(t) = g(y(t))$ for all $t \in [0, 1]$, that is, the plot of y versus t is the same as the plot of x versus t after a deformation of the vertical axis. In Section 2 we define the rank correlation, $\rho(x, y)$, between two functions and, in Theorem 2 of that section, we show that x and y have the same shape if and only if their rank correlation is equal to one. Thus, we can measure the similarity in the shapes of x and y by calculating their rank correlation or an estimate of their rank correlation.

We might want to use the rank correlation between two functions in two different situations. In the first, one of the functions is known and the other unknown. For instance, to determine if an unknown function is monotone, we would compare it to a known monotone function. To determine if an unknown function has one mode, we would compare it to a known unimodal function. In the second situation, when neither function is known, we'd like to know if the two functions have similar (unspecified) trends or shapes. A quantification of similarity of shape could be used in testing that the two curves have the same shape or the quantification could be used as a distance measure in a cluster analysis of a collection of curves.

In Section 3, we introduce our estimate of ρ and in Section 4 apply the estimation procedure to cluster analysis of curves. Section 5 contains asymptotic results for our estimates. Finite sample behaviour of our estimate is studied via simulations in Sections 6.

An alternate definition of shape similarity would allow for monotone transformations of both the explanatory and response variables, that is, x and y would be the same shape if there exist g_1 and g_2 strictly increasing with $x(t) = (g_1 \circ y \circ g_2)(t)$ for all $t \in [0,1]$. Analysis using this definition is difficult, and is not studied here. Suggestions for extending our work to this case are discussed in the conclusions.

2 Definition of Rank and Rank Correlation

Let x be a Lebesgue measurable function on the unit interval and μ a probability measure on the Lebesgue sets of $[0,1]$. Define the rank of $x(t)$ among all $x(s)$'s, relative to μ, to be

$$r^{x,\mu}(t) = \mu\{s : x(s) < x(t)\} + \frac{1}{2}\mu\{s : x(s) = x(t)\}.$$

The superscripts x and μ will be dropped when the meaning is clear. The definition of r extends the usual definition of rank. To see this, suppose that μ is the counting measure on $\{t_1, \ldots, t_n\}$. Let k_i be the number of $x(t_j)$'s $< x(t_i)$ and let m_i be the number of $x(t_j)$'s $= x(t_i)$. Then the usual definition of the rank of $x(t_i)$ relative to $x(t_1), \ldots x(t_n)$ is $k_i + (m_i + 1)/2$ (using the midrank adjustment for ties). But $nr^{x,\mu}(t_i) = k_i + m_i/2$, which is equal to the rank of $x(t_i)$ minus $1/2$.

There are several different ways to look at r. The following equivalent definition of r is useful in proving Theorems 1 and 2 below. Let $S \sim \mu$, that is,

$$P\{S \in A\} = \mu\{A\}.$$

Then

$$r(t) = P\{x(S) < x(t)\} + \frac{1}{2}P\{x(S) = x(t)\}. \tag{1}$$

Another view of r shows its connection to the usual definition of rank for a discrete set of n numbers. Let $T \sim \mu$ and suppose that F, the cumulative distribution function of $x(T)$, is continuous (so, for instance, x isn't constant on any set of positive μ-measure). Then $r^{x,\mu}(T) = F(x(T))$, and thus has a uniform distribution. Therefore, its mean is $1/2$ and its variance is $1/12$. Compare to the finite rank situation: let R_i, $i = 1, \cdots, n+1$, be the ranks of $n+1$ distinct numbers and let I be a random variable, uniform on $1, \cdots, n+1$. Then $R_I/(n+1)$ is similar to $r^{x,\mu}(T)$. The expected value of $R_I/(n+1)$ is

1/2 and its variance converges to 1/12 as $n \to \infty$. Theorem 1 below gives an exact value for the mean and variance of $r^{x,\mu}(T)$ for the general case, where F isn't necessarily continuous.

Our definition of the rank correlation between two Lebesgue measurable functions x and y will involve integration of $r^{x,\mu}$ and $r^{y,\mu}$ with respect to μ. The following lemma guarantees that such integration is well-defined.

Lemma 1 *If x is Lebesgue measurable, then $r^{x,\mu}$ is Lebesgue measurable.*

Proof. Let $H_1(w) = \mu\{s : x(s) < w\}$ and $H_2(w) = \mu\{s : x(s) \leq w\}$. Then H_1 and H_2 are monotone, and thus Borel measurable. But $r(\cdot) = H_1(x(\cdot)) + [H_2(x(\cdot)) - H_1(x(\cdot))]/2$, so r is Lebesgue measurable. (See, e.g., Royden, 1968, Chapter 3.)

We can now naturally define the rank correlation between two Lebesque measurable functions x and y with respect to the measure μ as

$$\rho_\mu(x,y) = \frac{\int (r^{x,\mu}(w) - R^{x,\mu})(r^{y,\mu}(w) - R^{y,\mu})d\mu(w)}{\left[\int (r^{x,\mu}(w) - R^{x,\mu})^2 d\mu(w) \int (r^{y,\mu}(w) - R^{y,\mu})^2 d\mu(w)\right]^{1/2}}$$

where

$$R^{x,\mu} = \int r^{x,\mu}(w) d\mu(w)$$

and $R^{y,\mu}$ is defined similarly. Note that, by the Cauchy-Schwarz inequality, $|\rho_\mu(x,y)| \leq 1$.

Theorem 1 *$R^{x,\mu} = 1/2$ and $\int (r^{x,\mu}(w) - R^{x,\mu})^2 d\mu(w) = \Big(1 - P\{x(S_1) = x(S_2) = x(S_3)\}\Big)/12$ where S_1, S_2 and S_3 are independent and identically distributed with $S_i \sim \mu$.*

The theorem is proven in the Appendix.

The following result justifies using ρ to measure similarity between functions.

Theorem 2 *Suppose*

(*) there exists a set A and a function g defined on $y(A)$ with $\mu\{A\} = 1$, $x(w) = g(y(w))$ for all $w \in A$, and g is strictly increasing on $y(A)$.

Then $r^x(w) = r^y(w)$ for all $w \in A$ and so $\rho(x, y) = 1$. Conversely, suppose that $\rho(x, y) = 1$ and either

a. *x and y are continuous and μ is absolutely continuous with respect to Lebesgue measure or*

b. *μ is a discrete measure on $\{s_1, \cdots, s_n\} \subset [0, 1]$ with $\mu\{s_i\} > 0$ for $i = 1, \cdots, n$.*

Then (*) holds.

Proof. Suppose that $x(w) = g(y(w))$ for g strictly increasing on $y(A)$ and that $\mu\{A\} = 1$. Then obviously $r^x(w) = r^y(w)$ for $w \in A$, so $\rho(x, y) = 1$.

To prove the converse, suppose that $\rho(x, y) = 1$. Then, by the Cauchy-Schwarz inequality, there exists α such that

$$r^x(w) - \frac{1}{2} = \alpha \left(r^y(w) - \frac{1}{2} \right) \quad \mu \text{ almost everywhere.} \tag{2}$$

First note that $\alpha \geq 0$. A negative α would give ρ negative.

Let $A = \{w : (2) \text{ holds}\}$. For case b, A must equal $\{s_1, \cdots, s_n\}$. We'll show that (2) implies the following.

$$\text{For all } t_1, t_2 \in A: \quad x(t_1) < x(t_2) \iff y(t_1) < y(t_2).$$

Suppose not. Suppose that there exist t_1 and t_2 in A with $x(t_1) < x(t_2)$ but $y(t_2) \geq y(t_1)$. Then $r^x(t_1) < r^x(t_2)$ and $r^y(t_1) \geq r^y(t_2)$. But this contradicts (2). Similarly, we can show that, for all t_1 and t_2 in A, $x(t_1) > x(t_2) \iff y(t_1) > y(t_2)$. Therefore, for $t_1, t_2 \in A$, $x(t_1) = x(t_2) \iff y(t_1) = y(t_2)$. So $g(y(t)) = x(t)$ is well-defined for $t \in A$ and strictly increasing on $y(A)$.

Comments.

1. We can generalize Spearman's correlation to the correlation between two functions by
$$\rho^*(x, y) = \frac{\int (x(t) - \bar{x})(y(t) - \bar{y})\, dt}{\sigma_x \sigma_y}$$
where $\bar{x} = \int x(t)\, dt$, $\sigma_x^2 = \int (x(t) - \bar{x})^2\, dt$, and $\bar{y}$ and σ_y^2 are defined similarly. We easily see that ρ^* is close to one if and only if the L^2 distance between $x^* \equiv (x - \bar{x})/\sigma_x$ and $y^* \equiv (y - \bar{y})/\sigma_y$ is small. Thus Spearman's correlation suffers from the same deficiencies as L_2 distance for measuring similarity between two curves.

2. The conditions a and b of Theorem 2 can probably be weakened, but doing so would lead to a more complicated proof. More general conditions are not really needed in practice.

3. From Theorem 2, we see that, if $x^* = g_1 \circ x$ and $y^* = g_2 \circ y$, g_1 and g_2 strictly increasing, then the correlation between x^* and y^* is the same as the correlation between x and y. Thus if we'd like to know if a continuous function y is strictly increasing, we can simply calculate the correlation between y and $x(t) = t$.

4. The measure μ can be chosen to eliminate a subset of the functions' domain or to emphasize the comparision of x and y on a particular subset. In our examples and simulations, we'll just consider μ equal to the Lebesgue measure.

We now consider two examples. In the first, we calculate the correlation between a monotone function x and a parabola y with vertex in $[0, 1]$. As one would hope, the correlation decreases as the vertex increases from 0 to 1/2. The second example illustrates that correlation measures more than just matching of mode position. It also measures the similarity in symmetry about the mode.

Example 1. Suppose that $x(t) = t$ and $y(t) = (t - b)^2$ for $t \in [0, 1]$ and for fixed $b \in [0, 1/2]$. Then $r^x(t) = t$,

$$r^y(t) = \begin{cases} 2(b - w) & \text{for } 0 \le w \le b \\ 2(w - b) & \text{for } b \le w \le 2b \\ w & \text{for } 2b \le w \le 1 \end{cases}$$

and $\rho_\mu(x, y) = 1 - 8b^3$.

Example 2. Suppose that $y(t) = |t - 1/2|$ and

$$x(t) = \begin{cases} 1/2 - t & \text{for } t \in [0, 1/2] \\ m(t - 1/2) & \text{for } t \in [1/2, 1]. \end{cases}$$

for m a positive constant. Then

$$\rho_\mu(x, y) = \begin{cases} (m^2 + 2m - 1)/(2m^2) & \text{for } m > 1 \\ 1 - (m - 1)^2/2 & \text{for } 0 < m \le 1. \end{cases}$$

As one would expect, ρ is an increasing function of $m \in (0,1]$ and a decreasing function of $m \in [1,\infty)$. At $m=0$, $\rho = 1/2$, and ρ approaches $1/2$ as m approaches infinity.

3 Estimation of ρ

Suppose that we have two regression functions, x and y, and we would like to estimate $\rho_\mu(x,y)$. There are two situations: the one-sample problem in which we consider x as a known target function, and the two-sample problem, in which both x and y are unknown. In both situations, we base our estimate of ρ on an estimate $\hat{y}$ (or $\hat{y}$ and $\hat{x}$) of the unknown regression function(s). The function can be estimated via spline smoothing or local linear smoothing. See, for instance, Eubank (1988), Green and Silverman (1994), Wand and Jones (1995), Fan and Gijbels (1996), or Simonoff (1996).

In the one-sample case our estimate of ρ is the rank correlation coefficient between $\hat{y}$ and x, with respect to μ_n, a discrete distribution on a grid $s_i^n, i = 1, \cdots, n$ with, typically n taken much larger than n_y. Our estimate can be easily re-written in terms of R_i^x, the usual rank of $x(s_i^n)$ among the $x(s_j^n)$'s, and $R_i^{\hat{y}}$, the usual rank of $\hat{y}(s_i^n)$ among the $\hat{y}(s_j^n)$'s. For instance, the numerator of $\hat{\rho}$ can be rewritten

$$\begin{aligned}\sum_{i=1}^{n} r^{x,\mu_n}(s_i^n) r^{\hat{y},\mu_n}(s_i^n) - \frac{1}{4} &= \frac{1}{n^3}\sum_{i=1}^{n}\left(R_i^x - \frac{1}{2}\right)\left(R_i^{\hat{y}} - \frac{1}{2}\right) - \frac{1}{4} \\ &= \frac{1}{n^3}\sum_{i=1}^{n} R_i^x R_i^{\hat{y}} - \frac{1}{4}\left(1 + \frac{2}{n} + \frac{1}{n^2}\right). \end{aligned} \tag{3}$$

The "y" part of $\hat{\rho}$'s denominator can be written in terms of the ranks, or by Theorem 1, as

$$\int \left(r^{\hat{y},\mu_n}(s) - R^{\hat{y},\mu_n}\right)^2 d\mu_n(s) = \frac{1}{12}\Big(1 - P\{\hat{y}(s_I^n) = \hat{y}(s_J^n) = \hat{y}(s_K^n) | \hat{y}(\cdot)\}\Big) \tag{4}$$

where I, J, and K are independent and independent of $\hat{y}$, and uniformly distributed on $\{1, \cdots, n\}$. Note that in the probability on the right hand side of (4) $\hat{y}$ is considered as a fixed function.

In the two-sample case we calculate the rank correlation coefficient between $\hat{y}$ and $\hat{x}$ similarly.

In Section 5 we show that $\hat{\rho}$ is consistent if, in the one sample case, y is not constant on any interval and if, in the two sample case, neither x nor y is constant on any interval. Section 6 contains discussion of simulations of $\hat{\rho}$'s distribution. But first, in Section 4, we use rank correlation in a cluster analysis of a set of curves.

4 Cluster Analysis Example

The data, available at J.O. Ramsay's web site www.psych.mcgill.ca/faculty/ramsay, consist of average daily precipitation for thirty-five Canadian weather stations averaged over the years 1961 to 1994. We would like to divide the thirty-five stations into groups with similar daily precipitation patterns. We do this via a cluster analysis, using rank correlation as our similarity measure. Specifically for each station, the 365 data points are smoothed using *smooth.spline* in *Splus*, with smoothing parameters chosen by generalized cross-validation. We then calculate rank correlations among the smoothed curves of the thirty-five stations, using the original time points as our grid. A grid with $n = 365$ is sufficiently large to give a good approximation to the integrals appearing in our definition of rank correlation. We then ran a cluster analysis in *Splus* to divide the stations into four groups. Plots of the smoothed data for the four groups are shown in Figures 1 to 4. The first group has high rainfall in the winter months and low rainfall in the summer months. The second group has fairly variable rainfall, with only a slightly higher level of precipitation in late summer and early fall. The third group's plots have a very large mode in late summer and early fall. It's rather hard to make generalities about the fourth group, since it contains only two stations. But these stations don't seem to fit into any of the three previous groups.

For comparison, we redid the cluster analysis using correlation of the smoothed data instead of correlation of the ranks. The first two groups are identical to those in Figures 1 and 2. The third and fourth groups are almost the same. The only difference is that seven stations, Calgary, Edmonton, Prince Albert, Regina, The Pas, Thunder Bay, and Winnipeg, have been moved from group three to group four. These seven stations are starred in Figure 3. The shapes of these seven stations' curves are roughly highly unimodal and clearly belong in group three. Therefore, we conclude that the rank correlation method has done a better job of grouping stations by *shape* of precipitation curves.

FIGURES 1-4 AROUND HERE

5 Asymptotic Results

The consistency of $\hat{\rho}$ is given for the one-sample problem in Theorem 3 and for the two-sample problem in Theorem 4. The assumptions, given in the Appendix, imply that $\hat{\rho}$ is based on an estimate $\hat{y}$ (or $\hat{x}$ and $\hat{y}$) that converges to y. In contrast, Theorem 5 shows that $\hat{\rho}$ based on the raw data will yield an inconsistent estimate of ρ.

In the one-sample problem, we have a known function x, we observe data

$$Y_i = y(t_i) + \eta_i \quad i = 1, \cdots, n_y \tag{5}$$

with the η_i's independent and identically distributed with mean zero. We choose gridpoints $s_i^n, i = 1, \cdots, n$, let μ_n be uniform discrete measure on the s_i^n's, and estimate $\rho_\mu(x, y)$ by $\hat{\rho} = \rho_{\mu_n}(x, \hat{y})$.

Theorem 3 *Suppose A1-A6 of the Appendix hold. Then*

$$\hat{\rho} \to \rho_\mu(x, y)\sqrt{1 - P\{y(S_1) = y(S_2) = y(S_3)\}}$$

in probability, where the S_i's are independent and identically distributed $\sim \mu$.

The proof is given in the Appendix.

The estimate $\hat{\rho}$ is consistent if $P\{y(S_1) = y(S_2) = y(S_3)\}$ is equal to zero, that is if y is not constant on any set of positive μ measure. For general y, consistent estimation of ρ probably requires consistent estimation of $P\{y(S_1) = y(S_2) = y(S_3)\}$. Unfortunately, this is very difficult since it entails estimating on which intervals, if any, y is constant.

Now consider the two-sample problem. We observe data as in (5) and, independently,

$$X_i = x(u_i) + \epsilon_i \quad i = 1, \cdots, n_x, \tag{6}$$

and estimate x and y by $\hat{x}$ and $\hat{y}$. Our estimate of $\rho_\mu(x, y)$ for a given μ is $\hat{\rho} = \rho_{\mu_n}(\hat{x}, \hat{y})$, for μ_n as described above.

Theorem 4 *Assume that A1-A6 of the Appendix hold, not only for $\hat{y}$ and y but also for $\hat{x}$ and x. Then*

$$\hat{\rho} \to \rho_\mu(x,y)\sqrt{1 - P\{y(S_1) = y(S_2) = y(S_3)\}}\sqrt{1 - P\{x(S_1) = x(S_2) = x(S_3)\}}$$

in probability, where the S_i's are independent and identically distributed $\sim \mu$.

The proof is similar to that of Theorem 3 and is omitted.

The following Theorem shows that an estimate of ρ based on the raw data is not consistent.

Theorem 5 *For the one-sample problem, suppose that $n_y = n$ and $t_i = s_i^n$. Assume A1, A6 of the Appendix and that the joint distribution of $Y_1, \cdots, Y_n$ is continuous. We can consider the vector Y of observations as a function, with $Y(s_i^n) = Y_i$. Then*

$$\hat{\rho} \equiv \rho_{\mu_n}(x, Y) \to \sqrt{12}\frac{\int_{s,w}\left[r^{x,\mu}(w) - \frac{1}{2}\right]\left[H_{\eta_1-\eta_2}(y(s) - y(w)) - \frac{1}{2}\right]\, d\mu(s)\, d\mu(w)}{\left[\int (r^{x,\mu}(w) - \frac{1}{2})^2\, d\mu(w)\right]^{1/2}}$$

where

$$H_\eta(y) = I\{y \neq 0\}P\{y < \eta\} + \frac{1}{2}I\{y = 0\}.$$

For the two-sample problem, suppose that $n_x = n_y = n$ and $t_i = u_i = s_i^n$. Assume A1, A6 (for both x and y) and that the joint distributions of $Y_1, \cdots, Y_n$ and of $X_1, \cdots, X_n$ are continuous. Then

$$\hat{\rho} = \rho_{\mu_n}(X, Y) \to$$

$$12\int_{s,w}\left[H_{\epsilon_1-\epsilon_2}(x(s) - x(w)) - \frac{1}{2}\right]\ \left[H_{\eta_1-\eta_2}(y(s) - y(w)) - \frac{1}{2}\right]\, d\mu(s)\, d\mu(w).$$

The proof is given in the Appendix.

The following heuristic argument shows the connection between Theorem 5 and Theorems 3 and 4. Consider the one sample case. Suppose that, in Theorem 5, the η_i's actually converge to zero, that is, we're basing our ranks on Y_i's that approach the $y(s_i^n)$'s. Thus we're in the situation of Theorem 3. Then $\eta_1 - \eta_2$ will converge to a point mass at zero. Suppose that we can substitute this point mass distribution into the Theorem 5 result. Then $H_{\eta_1-\eta_2}(y) \to I\{y<0\}+\frac{1}{2}I\{y=0\}$ and so

$$\int_s H_{\eta_1-\eta_2}(y(s)-y(w))\,d\mu(s) \sim \int_s \left[I\{y(s)<y(w)\}+\frac{1}{2}I\{y(s)=y(w)\}\right]d\mu(s) = r^{y,\mu}(w).$$

Thus

$$\hat{\rho} \to \sqrt{12}\;\frac{\int_w \left[r^{x,\mu}(w)-\frac{1}{2}\right] r^{y,\mu}(w)\;d\mu(w)}{\left[\int (r^{x,\mu}(w)-\frac{1}{2})^2\;d\mu(w)\right]^{1/2}}$$

which, by Theorem 1, is equal to

$$\rho_\mu(x,y)\sqrt{1-P\{y(S_1)=y(S_2)=y(S_3)\}}$$

with S_1, S_2, S_3 independently and identically distributed according to μ. This is just the result of Theorem 3.

Note that, in the one-sample problem, if we test for monotonicity by taking $x(t)=t$ and using the raw data, then $\hat{\rho}$ is a function of $\sum iR_i^Y$, the usual rank correlation statistic for testing $E(Y_i)=E(Y_j)$ for all i and j.

6 Simulation of $\hat{\rho}$'s Distribution

Theorems 3 and 4 guarantee that, under regularity conditions, $\hat{\rho}$ is consistent. We also studied the finite sample behaviour of $\hat{\rho}$ via simulation. In what follows we discuss the simulation results for the one-sample case. The results for the two-sample case were similar.

Each simulation run involved two functions, x and y, with one thousand data sets. For each data set the Y_i's were generated as in (5), with $t_i=i/n$, $i=1,\cdots,n$. We considered $n=50$ and $n=100$. We used one of the following distributions for the η_i's.

a. the η_i's are normally distributed with zero mean and standard deviation $\sigma_y=0.1$;

b. $10\eta_i$ has a t distributions with 3 degrees of freedom.

We used the following functions:

1. $x(t) = t$ and $y(t) = t$ ($\rho = 1$).

2. $x(t) = t$ and $y(t) = t^2$ ($\rho = 1$).

3. $x(t) = t$ and $y(t) = (t-b)^2$, $b = 0.1, 0.2, 0.3, 0.4, 0.5$ ($\rho = 0.992, 0.936, 0.784, 0.488, 0$).

Figure 5 contains two randomly generated data sets.

FIGURE 5 HERE

For each data set a cubic spline estimate, $\hat{y}$, of the regression function was calculated using *Splus*'s *smooth.spline*, with smoothing parameter chosen automatically by generalized cross-validation. We then calculated $\hat{\rho}$, the rank correlation coefficient based on the pairs $x(i/500)$, $\hat{y}(i/500)$, $i = 1, \cdots, 500$. Although smoothing splines do not satisfy A3 of the Appendix typically the dependence between $\hat{y}(s)$ and $\hat{y}(t)$ is small for $|s - t|$ large, so A3 is approximately satisfied. We've used smoothing spline estimation because it is the most widely available smoothing technique with automatic smoothing parameter selection.

Results are presented in Tables 1 and 2. Our estimate of ρ does very well, particularly when regression errors are normally distributed. However the estimates are slightly biased. When $\rho = 1$, some bias is to be expected since $\hat{\rho}$ never exceeds one.

TABLES 1, 2 HERE

From our tables, it seems that $P\{\hat{\rho} = 1\} > 0$, that is, that with positive probability $\hat{y}$ is a monotone function. This isn't suprising. Most smoothing methods will produce a monotone function (for *any* given data set) provided that the smoothing parameter is sufficiently large. Even if the underlying regression function is not monotone, there is a positive probability that the estimate will be monotone.

In Figure 6 we've plotted the densities of the $\hat{\rho}$'s for some of our one-sample simulations ($n = 50$), but first omitting the values of $\hat{\rho}$ that are equal to one. The true value of ρ is denoted by a vertical line. The bias appears to be due to $\hat{\rho}$ values equal to one.

FIGURE 6 HERE

7 Conclusions

We have defined $\rho_\mu(x, y)$, the rank correlation between two functions x and y, and have successfully estimated it. We have shown that ρ is a useful measure of similarity in a cluster analysis of precipitation data.

Unfortunately, the question of how to use $\hat{\rho}$ for testing remains unsolved. Suppose that we would like to test H_0 : x and y have the same shape, that is, that $\rho_\mu(x, y)$, the rank correlation between x and y with respect to a specified measure μ, is equal to one. We would reject H_0 if our estimate of $\rho(x, y)$ were too small. We could estimate ρ as in Section 3. If we could calculate the null distribution of $\hat{\rho}$, we could easily calculate our p-value. However, our estimate is very complicated, in particular because it is not linear in the data. Even if we could determine its asymptotic distribution, it would undoubtedly depend on unknowns, such as the curvature of the regression functions and the standard deviation(s) of the regression error(s). We studied some bootstrap procedures. Unfortunately, these did not prove too successful, probably because $\hat{\rho}$ has positive probability of equaling 1.

Our definition of $\rho_\mu(x, y)$ may be useful in other contexts where shape is of primary interest. One might use $\rho_\mu(x, y)$ for aligning two curves. When studying families of regression functions such as growth curves, one typically aligns the curves so that features "match up". A "typical" curve is then a pointwise average of the aligned curves. To align two curves x and y, one would find two increasing functions g_1 and g_2 so that x is close to $g_1 \circ y \circ g_2$. We could first find a monotone g_2 so that x and $y \circ g_2$ have the same shape, that is, maximal correlation. We could search for such a g_2 in the parametric family of scale and location changes: $g_2(t) = \beta_0 + \beta_1 t$. Or we could take a nonparametric approach, searching for some smooth monotone g_2. Ramsay and Li, 1998, have developed a method to find g_2, smooth and monotone, to make x and $y \circ g_2$ close in L_2 norm.

Rank functions may also be used in a principal component analysis of a family of curves $\{x_j(\cdot)\}$, to study how the curves most differ in shape. Ramsay and Silverman (1997) use smoothed principal components analysis on data $\{x_j(t_i)\}$ to characterize how curves differ. A similar computation based on the $r^{\hat{x}_j}(\cdot)$'s might focus the analysis on curve shape.

8 Appendix

Suppose that our gridpoints $s_i^n, i = 1, \cdots, n$ are *regular* according to μ, that is, suppose that

A1. $\mu[0,t] = \int_0^t f(u)\, du$ for f a strictly positive continuous density on $[0,1]$, and
$$\int_0^{s_i^n} f(t)\ dt = \frac{i}{n}.$$

We make the following additional assumptions on μ, $\hat{y}$, and y, for the one sample problem, and add the conditions on $\hat{x}$ and x, for the two sample problem.

A2. $(\hat{y}(s_1^n), \cdots, \hat{y}(s_n^n))$ has a continuous (joint) distribution.

A3. There exists a bandwidth $h_y = h_y^n \to 0$ such that $\hat{y}(s)$ depends only on those Y_i's with $|t_i - s| \leq h_y$.

A4. There exists a sequence $\Delta_n > 0$, converging to 0, such that
$$\lim_{n\to\infty} \sup_{k=1,\cdots,n} P\{|\hat{y}(s_k^n) - y(s_k)| > \Delta_n\} = 0.$$

A5. Suppose that $E(Y_i) = v$ for all i with $|t_i - s| \leq h_y$. Then $E(\hat{y}(s)) = v$. If, in addition, $E(Y_i) = v$ for all i with $|t_i - t| \leq h_y$, then the distribution of $\hat{y}(s) - \hat{y}(t)$ is symmetric about 0.

A6. y is continuous on $[0,1]$ and there exists a partition of $[0,1]$ into disjoint intervals $\mathcal{I}_1, \cdots, \mathcal{I}_K$ such that, on each interval, y is either strictly increasing, strictly decreasing, or constant.

Condition A2 guarantees that, with probability one, there are no ties among the $\hat{y}(s_i^n)$'s. Presumably this condition can be dropped but proofs of the theorems would become more complicated.

Conditions A2-A4 are satisfied by many popular smoothing methods provided that the Y_i's have continuous distributions, y is continuous, the t_i's form a regular grid with respect to some positive continuous density, and $n_y h_y \to \infty$. For instance, we could use the Nadaraya-Watson estimate (Nadaraya, 1964, Watson, 1964)

$$\hat{y}(s) = \text{the average of the } Y_i\text{'s with } |t_i - s| \le h_y,$$

or the locally linear estimate with compact kernel. In this case, $\hat{y}(s)$ is the estimate of $y(s)$ gotten by fitting a least squares line to the subset of the data, (t_i, Y_i) with $|t_i - s| \le h_y$. See, e.g., Wand and Jones, 1995, or Fan and Gijbels, 1996.

Condition A5 is typically satisfied when the distributions of the Y_i's are symmetric about their expected values. Without condition A5, we can't guarantee that, when $E(Y_i) \equiv v$, a constant, $E(R_i^{\hat{y}})/n \to 1/2$. However condition A5 can be replaced by

A5*. For all t, $\mu\{s : y(s) = y(t)\} = 0$.

Condition A6 allows us to exclude y's that have rank functions that are difficult to approximate by a finite grid. Consider, for example, $y(t) = t\sin(1/t)$.

Proof of Theorem 1. The superscripts x and μ will be dropped. To prove the first statement, let S_1 and S_2 be independent $\sim \mu$. Then

$$R = P\{x(S_1) < x(S_2)\} + \frac{1}{2}P\{x(S_1) = x(S_2)\} = P\{x(S_2) < x(S_1)\} + \frac{1}{2}P\{x(S_1) = x(S_2)\}.$$

But

$$P\{x(S_1) < x(S_2)\} + \frac{1}{2}P\{x(S_1) = x(S_2)\} + P\{x(S_2) < x(S_1)\} + \frac{1}{2}P\{x(S_1) = x(S_2)\} = 1$$

so $R = 1/2$.

To prove the second statement, write

$$\int (r(w) - R)^2 d\mu(w) = \int (r(w) - 1/2)^2 d\mu(w) = \int r^2(w) d\mu(w) - 1/4$$

and study

$$\int r^2(w)\ d\mu(w) =$$

$$\int \left[P\{x(S_1) < x(w)\} + \frac{1}{2}P\{x(S_1) = x(w)\}\right]\left[P\{x(S_2) < x(w)\} + \frac{1}{2}P\{x(S_2) = x(w)\}\right] d\mu(w)$$

$$= P\{x(S_1) < x(S_3) \text{ and } x(S_2) < x(S_3)\} + \frac{1}{2}P\{x(S_1) < x(S_3) \text{ and } x(S_2) = x(S_3)\}$$

$$+\frac{1}{2}P\{x(S_2) < x(S_3) \text{ and } x(S_1) = x(S_3)\} + \frac{1}{4}P\{x(S_1) = x(S_2) = x(S_3)\}.$$

But

$$P\{x(S_1) < x(S_3) \text{ and } x(S_2) < x(S_3)\} = P\{\max\{x(S_1), x(S_2)\} < x(S_3)\}$$

$$= \frac{1}{3}P\{\text{the maximum of } x(S_1), x(S_2), x(S_3) \text{ is attained at exactly one } x(S_i)\}$$

$$\equiv \frac{1}{3}P_1$$

and

$$P\{x(S_1) < x(S_3) \text{ and } x(S_2) = x(S_3)\} =$$

$$\frac{1}{3}P\{\text{the maximum of } x(S_1), x(S_2), x(S_3) \text{ is attained at exactly 2 } x(S_i)\text{`s}\}$$

$$\equiv \frac{1}{3}P_2$$

$$= P\{x(S_2) < x(S_3) \text{ and } x(S_1) = x(S_3)\}.$$

So

$$\int (r(w) - R)^2 d\mu(w) = \frac{1}{3}(P_1 + P_2) + \frac{1}{4}P\{x(S_1) = x(S_2) = x(S_3)\} - \frac{1}{4}$$

$$= \frac{1}{3}\Big[1 - P\{x(S_1) = x(S_2) = x(S_3)\}\Big] + \frac{1}{4}P\{x(S_1) = x(S_2) = x(S_3)\} - \frac{1}{4}$$

$$= \frac{1}{12}\Big[1 - P\{x(S_1) = x(S_2) = x(S_3)\}\Big].$$

The proofs of Theorems 3-5 depend on a sequence of Lemmas. Convergence of μ_n to μ and of r^{y,μ_n} to $r^{y,\mu}$ is studied in Lemmas 2 and 3. Lemma 4 gives continuity and Riemann integrability properties of $r^{y,\mu}$. Lemma 5 gives asymptotic properties of $R_i^{\hat{y}}$.

Lemma 2 *For $\{s_i^n\}$ satisfying (A1), there exists $0 < C_1 < C_2 < \infty$ such that*

$$\frac{C_1}{n} \leq s_i^n - s_{i-1}^n \leq \frac{C_2}{n} \tag{7}$$

for all $i = 1, \cdots, n$ and n. (Take $s_0^n \equiv 0$.) Furthermore

$$\sup_{t\in[0,1]} \max\Big\{|\mu_n[0,t] - \mu[0,t]|, |\mu_n[0,t) - \mu[0,t)|\Big\} \to 0 \quad \text{as } n \to \infty. \tag{8}$$

Proof. Let $C_2 = 1/\min\{f(t), 0 \le t \le 1\}$ and let $C_1 = 1/\max\{f(t), 0 \le t \le 1\}$. Then

$$\frac{1}{n} \int_{s_{i-1}^n}^{s_i^n} f(t)\ dt \le \frac{s_i^n - s_{i-1}^n}{C_1}.$$

The integral is bounded below by $(s_i^n - s_{i-1}^n)/C_2$.

To prove (8), note that, for $t \in [s_i^n, s_{i+1}^n)$,

$$|\mu_n[0,t] - \mu[0,t]| = \left| \frac{i}{n} - \int_0^{s_i^n} f(u)\ du - \int_{s_i^n}^{t} f(u)\ du \right|$$

$$= \int_{s_i^n}^{t} f(u)\ du \le \max\{f(t), 0 \le t \le 1\}\ \frac{C_2}{n}.$$

The difference $|\mu_n[0,t) - \mu[0,t)|$ is bounded similarly.

Lemma 3 *Assume A1 and A6. Then*

$$\lim_{n\to\infty} \sup_C |\mu_n\{s : y(s) < C\} - \mu\{s : y(s) < C\}| = 0$$

and

$$\lim_{n\to\infty} \sup_C |\mu_n\{s : y(s) = C\} - \mu\{s : y(s) = C\}| = 0.$$

Therefore

$$\lim_{n\to\infty} \sup_{t\in[0,1]} |r^{y,\mu_n}(t) - r^{y,\mu}(t)| = 0.$$

Proof. The first statement follows from (8) and the fact that, by A6, $\{s : y(s) < C\}$ is the union of at most K intervals.

Again by A6, $\{s : y(s) = C\}$ is the union of at most K s's and at most K intervals. Therefore, by (8), the second statement holds.

The last statement follows from the first two.

Lemma 4 *Assume A1 and A6. Then $r^{y,\mu}$ is continuous except at a finite number of points, and is thus Riemann integrable.*

Proof. Write the closure of $\mathcal{I}_j$ as $[c_{j-1}, c_j]$, $0 = c_0 < c_1 \cdots < c_K = 1$. Suppose that y is constant on $\mathcal{I}_j$ $j \in J^*$, with $y(s) \equiv y_j$ on $\mathcal{I}_j$, and let $\mathcal{S}_j = \{s : y(s) = y_j\}$. Then $\mathcal{S}_j$ is made up of some union of the $\mathcal{I}_j$'s, $j \in J^*$, and (possibly) a finite number of points in the $\mathcal{I}_j$'s, $j \notin J^*$. Call this finite collection of points $\mathcal{A}_j$.

We'll show that $r^{y,\mu}$ is continuous at

$$t \notin \mathcal{B} \equiv \{c_0, \cdots, c_K\} \cup \left(\bigcup_{j \in J^*} \mathcal{A}_j \right).$$

Note that $\mu\{\mathcal{B}\} = 0$ and that if $t \notin \mathcal{B} \cup \left(\bigcup_{j \in J^*} \mathcal{I}_j \right)$, then $\mu\{s : y(s) = y(t)\} = 0$.

Fix $t \notin \mathcal{B}$ and let $t_n \to t$. Since $\mathcal{B}^c$ is open, we can assume that the t_n's are also in $\mathcal{B}^c$.

Suppose that $t \in (c_{j^*-1}, c_{j^*})$ some $j^* \in J^*$. Then we can assume that the t_n's are also in (c_{j^*-1}, c_{j^*}). So $y(t_n) = y(t)$ and thus $r^{y,\mu}(t_n) = r^{y,\mu}(t)$.

Now suppose that $t \in (c_{j-1}, c_j)$, $j \notin J^*$. Then the set $\{s : y(s) \le y(t)\}$ is made up of a finite number of intervals with endpoints $\{w_1, \cdots, w_\ell\}$ and, in an open interval about each w_i, y is strictly monotone. Let $\mathcal{O}$ be the union of such a collection of open intervals. We can assume that the t_n's are in $\mathcal{O}$. Since the t_n's are in (c_{j-1}, c_j) and not in $\mathcal{B}$, the sets $\{s : y(s) \le y(t_n)\}$ are made up of a finite number of intervals with endpoints $\{w_1^n, \cdots, w_\ell^n\}$ converging to $\{w_1, \cdots, w_\ell\}$. So

$$r^{y,\mu}(t_n) = \mu\{s : y(s) \le y(t_n)\} \to \mu\{s : y(s) \le y(t)\} = r^{y,\mu}(t).$$

Lemma 5 *1. Suppose that A1, A2, and A3 hold. Let $R_i^{\hat{y}}$ denote the usual rank of $\hat{y}(s_i^n)$ among the $\hat{y}(s_j^n)$'s. Then there exist constants D_1 and D_2 not depending on n such that $|cov(R_i^{\hat{y}}, R_j^{\hat{y}})| \le D_1(nh_y)^2$ whenever $|i-j| > D_2 n h_y$.*

2. Suppose that A1-A6 hold. Then

$$\lim_{n\to\infty} \sup_{1 \le i \le n} |E(R_i^{\hat{y}})/n - r^{y,\mu}(s_i^n)| = 0.$$

Proof of 1. The superscript $\hat{y}$ will be dropped. First write

$$cov(R_i, R_j) = \sum_{k,\ell=1}^{n} cov\Big(I\{\hat{y}(s_k^n) \le \hat{y}(s_i^n)\}, I\{\hat{y}(s_\ell^n) \le \hat{y}(s_j^n)\} \Big).$$

By A3, we know that $\hat{y}(s)$ depends on Y_i's with $|s - t_i^n| \le h_y$. Thus $\hat{y}(s)$ and $\hat{y}(t)$ are independent if $|s-t| > 2h_y$. By Lemma 2, there exists $C_1 > 0$

such that $|s_i^n - s_j^n| > C_1|i-j|/n$. So if $|i-j| > 2nh_y/C_1 \equiv D_2nh_y$, then $|s_i^n - s_j^n| > 2h_y$, and so $\hat{y}(s_i^n)$ and $\hat{y}(s_j^n)$ are independent. So suppose that $|i-j| > D_2nh_y$. Then, by the same reasoning, the covariance between two indicator functions in the above summation is zero unless one of the following holds: $|k-\ell| \le D_2nh_y$, $|k-j| \le D_2nh_y$, or $|i-\ell| \le D_2nh_y$. So the covariance between R_i and R_j is the sum of $O((nh_y)^2)$ terms, all of which are bounded in absolute value by one.

Proof of 2. Write

$$E\,(R_i/n) = \frac{1}{n}\sum_k P\{\hat{y}(s_k^n) \le \hat{y}(s_i^n)\}.$$

First suppose that

$$\mu\{s : y(s) = y(s_i^n)\} = 0 \tag{9}$$

Let Δ_n be as in A4 and let

$$A_i^n = \{k : |y(s_i^n) - y(s_k^n)| > 2\Delta_n\}.$$

Then, by A4,

$$\sup_{k \in A_i^n} \left| P\{\hat{y}(s_k^n) \le \hat{y}(s_i^n)\} - I\{y(s_k^n) \le y(s_i^n)\} \right| \to 0$$

as $n \to \infty$, uniformly in i satisfying (9). So, summing over k, we can write

$$\left| E\,(R_i/n) - \mu_n\{s : y(s) \le y(s_i^n)\} \right| \le 2\mu_n\{(A_i^n)^c\} + o(1).$$

But $\mu_n\{(A_i^n)^c\} \to 0$ uniformly in i satisfying (9) since, by Lemma 3,

$$\begin{aligned} |\mu_n\{(A_i^n)^c\} - \mu\{(A_i^n)^c\}| &\le \sup_C \left| \mu_n\{s : |y(s) - C| \le \Delta_n\} - \mu\{s : |y(s) - C| \le \Delta_n\} \right| \\ &\to 0 \end{aligned}$$

and $\mu\{(A_i^n)^c\} \to 0$. Therefore

$$\lim_{n\to\infty} \sup_{i:(9)\text{ holds}} |E(R_i^{\hat{y}}/n) - r^{y,\mu}(s_i^n)| = 0.$$

Now suppose that

$$\mu\{s : y(s) = y(s_i^n)\} \neq 0$$

Then, by A1 and A6, $y(s_i^n)$ can take on at most K values. Let v denote one of these values and write $\hat{y}(s_i^n)$ as $v + e_i^n$. We must show that

$$\frac{1}{n}\sum_k P\{\hat{y}(s_k^n) \le v + e_i^n\} \to \mu\{s : y(s) < v\} + \frac{1}{2}\mu\{s : y(s) = v\}.$$

Suppose that $y(s) \equiv v$ for $s \in \cup_{j\in J^*}\mathcal{I}_j$ and consider the following partition of $\{s : y(s) = v\}$:

$$A_n = \cup_{j\in J^*}(c_{j-1} + h_y, c_j - h_y)$$

and

$$B_n = \{s \notin A_n : y(s) = v\}.$$

By an argument similar to that following (9),

$$\frac{1}{n}\sum_{s_k^n \notin A_n \cup B_n} P\{\hat{y}(s_k^n) \le v + e_i^n)\} - \mu\{s \notin A_n \cup B_n : y(s) \le v\} \to 0.$$

By A2 and A5, for $s_k^n \in A_n$, $P\{\hat{y}(s_k^n) \le v + e_i^n\} = P\{e_k^n \le e_i^n\} = 1/2$, so

$$\frac{1}{n}\sum_{s_k^n \in A_n} P\{\hat{y}(s_k^n) \le v + e_i^n)\} = \frac{1}{2}\mu_n\{A_n\}.$$

The lemma follows since $\mu_n\{A_n\} \to \mu\{s : y(s) = v\}$, $\mu\{B_n\} \to 0$ and

$$\mu\{s \notin A_n \cup B_n : y(s) \le v\} = \mu\{s \notin A_n \cup B_n : y(s) < v\} \to \mu\{s : y(s) < v\}.$$

Proof of Theorem 3. First consider the random portion of the denominator of $\hat{\rho}$. Using (4) and condition A2,

$$\int \left(r^{\hat{y},\mu_n}(s) - R^{\hat{y},\mu_n}\right)^2 d\mu_n(s) = \frac{1}{12}\Big(1 - P\{I = J = K\}\Big) \to \frac{1}{12}.$$

So, by (3), we must show that

$$\frac{1}{n^3}\Sigma R_i^x R_i^{\hat{y}} \to \int r^{x,\mu}(t) r^{y,\mu}(t)\ d\mu(t).$$

First let D_1 and D_2 be as in Lemma 5 and write

$$\mathrm{var}(\Sigma R_i^x R_i^{\hat{y}}) = \sum_{i,j:|i-j|\le D_2 n h_y} R_i^x R_j^x \mathrm{cov}(R_i^{\hat{y}}, R_j^{\hat{y}}) + \sum_{i,j:|i-j|> D_2 n h_y} R_i^x R_j^x \mathrm{cov}(R_i^{\hat{y}}, R_j^{\hat{y}}).$$

The first term is bounded by $(D_2 n h_y)^2 n^4 = D_2^2 n^6 h_y^2$. By Lemma 5, the second term is bounded by

$$\sum_{i,j:|i-j|>D_2 n h_y} n^2 D_1 (n h_y)^2 \le D_1 n^6 h_y^2.$$

Thus the variance of $\Sigma R_i^x R_i^{\hat{y}}/n^3$ converges to zero.

Now consider the expectation and use Lemma 5.

$$\begin{aligned} n^{-3} E(\Sigma R_i^x R_i^{\hat{y}}) &= \frac{1}{n}\sum_i \frac{R_i^x}{n}\frac{E(R_i^{\hat{y}})}{n} \\ &= \frac{1}{n}\sum_i \frac{R_i^x}{n}\left(r^{y,\mu}(s_i^n) + o(1)\right) \\ &= \frac{1}{n}\sum_i r^{x,\mu}(s_i^n) r^{y,\mu}(s_i^n) \quad + o(1) \end{aligned}$$

by Lemma 3. Since $r^{x,\mu}$ and $r^{y,\mu}$ are Reimann integrable, the above converges to $\int r^{x,\mu}(t) r^{y,\mu}(t)\ d\mu(t)$.

Proof of Theorem 5. The proof will be presented for the one-sample case. The two-sample case is similar. Consider the random part of the denominator of $\hat{\rho}$. Using (4) and the fact that the Y_i's have a continuous distribution

$$\int \left(r^{Y,\mu_n}(s) - R^{Y,\mu_n}\right)^2 d\mu_n = \frac{1}{12}\left(1 - P\{I = J = K | Y_1, \cdots, Y_{n_y}\}\right) \to \frac{1}{12}.$$

As in the proof of Theorem 3, one shows that the variance of $\sum R_i^x R_i^Y / n^3$ converges to zero. To study its expected value, we study $E(R_i^Y/n)$.

$$\begin{aligned} E(R_i^Y/n) &= \frac{1}{n}\sum P\{y(s_k^n) - y(s_i^n) \le \eta_i - \eta_k\} \\ &= \frac{1}{n}\sum P\{y(s_k^n) - y(s_i^n) \le \eta_1 - \eta_2\} + o(1) \\ &= \frac{1}{n}\sum_{k:y(s_k^n)\neq y(s_i^n)} P\{y(s_k^n) - y(s_i^n) < \eta_1 - \eta_2\} + \frac{1}{n}\sum_{k:y(s_k^n) = y(s_i^n)} \frac{1}{2} + o(1) \\ &= \int_s \Big[I\{y(s) \neq y(s_i^n)\} P\{y(s) - y(s_i^n) < \eta_1 - \eta_2\} \\ &\qquad + \frac{1}{2} I\{y(s) = y(s_i^n)\}\Big]\ d\mu(s) + o(1) \\ &= \int H_{\eta_1 - \eta_2}(y(s) - y(s_i^n))\ d\mu(s) + o(1) \end{aligned}$$

So, in the one-sample case,

$$
\begin{aligned}
E\left(\frac{1}{n^3}\sum_i R_i^x R_i^Y\right) &= \frac{1}{n}\sum_i r^{x,\mu_n}(s_i^n) E\left(\frac{R_i^Y}{n}\right) \\
&= \frac{1}{n}\sum_i r^{x,\mu}(s_i^n)\int_s H_{\eta_1-\eta_2}(y(s)-y(s_i^n))\ d\mu(s) + o(1) \\
&\to \int_{s,w} r^{x,\mu}(s) H_{\eta_1-\eta_2}(y(s)-y(w))\ d\mu(s)\ d\mu(w).
\end{aligned}
$$

The proof of the two-sample case is similar.

9 References

Bowman, A.W., Jones, M.C., and Gijbels, I., 1998. "Testing monotonicity of regression", *Journal of Computational and Graphical Statistics*, to appear.

Eubank, R. L., 1988. *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.

Fan, J. and Gijbels, I., 1996. *Local Polynomial Modeling and its Applications*, Chapman and Hall, London.

Friedman, J.H. and Tibshirani, R.J., 1984. The monotone smoothing of scatterplots, *Technometrics*, **26**, 243-50.

Green, P. J. and Silverman, B. W., 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London.

Hall, P. and Heckman, N.E., 1998. "Testing for monotonicity of a regression mean without selecting a bandwidth", Statistics Department, University of British Columbia Technical Report 178.

Harezlak, Jarek (1998). *Bump Hunting Revisited.* MSc Thesis, Statistics Department, University of British Columbia.

Mammen, E. (1991). "Estimating a smooth monotone regression function", *Annals of Statistics*, **19**, 724-40.

Marron, J.S. and Tsybakov, A.B. (1995). "Visual error criteria for qualitative smoothing", *Journal of the American Statistical Association*, **90**, 499-507.

Mukerjee, H., 1988. "Monotone nonparametric regression", *Annals of Statistics*, **16**: 741-50.

Nadaraya, E.A., 1964. "On estimating regression", *Theory of Probability and its Applications*, **9**: 141-142.

Ramsay, J.O., 1988. "Monotone regression splines in action" (with comments), *Statistical Science*, **3**: 425-461.

Ramsay, J.O., Bock, R.D., Gasser, T., 1995. "Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data", *Annals of Human Biology*, **22**: 413-426.

Ramsay, J.O. and Li, X., 1998. "Curve registration", *Journal of the Royal Statistical Society B*, **60**: 351-363.

Ramsay, J.O. and Silverman, B.W., 1997. *Functional Data Analysis*, Springer, New York.

Roy, N., 1998. "Is there variation in calorie response to income change across income groups and gender? A study of rural south India", manuscript.

Royden, H.L., 1968. *Real Analysis*, MacMillan Publishing Company, New York.

Simonoff, J.S., 1996. *Smoothing Methods in Statistics*, Springer-Verlag, New York.

Wand, M.P. and Jones, M.C., 1995. *Kernel Smoothing*, Chapman and Hall, London.

Watson, G.S., 1964. "Smooth regression analysis", *Sankhya A*: **26**: 359-372.

Simulated Distribution of $\hat{\rho}$: One-sample Case: $n = 50$

Functions	ρ	Dist	Mean (SE*)	% of 1's
$x(t) = t$	1	Normal	0.999 (0.000)	93.3
$y(t) = t$		t_3	0.996 (0.001)	91.5
$x(t) = t$	1	Normal	0.995 (0.000)	79.6
$y(t) = t^2$		t_3	0.990 (0.001)	79.2
$x(t) = t$	0.992	Normal	0.986 (0.001)	68.6
$y(t) = (t - .1)^2$		t_3	0.978 (0.002)	68.5
$x(t) = t$	0.936	Normal	0.952 (0.002)	40.0
$y(t) = (t - .2)^2$		t_3	0.940 (0.003)	48.9
$x(t) = t$	0.784	Normal	0.829 (0.003)	8.8
$y(t) = (t - .3)^2$		t_3	0.832 (0.005)	25.5
$x(t) = t$	0.488	Normal	0.537 (0.005)	0.2
$y(t) = (t - .4)^2$		t_3	0.570 (0.008)	7.8
$x(t) = t$	0.0	Normal	0.028 (0.005)	0.0
$y(t) = (t - .5)^2$		t_3	0.038 (0.010)	1.9

Table 1: A thousand simulations were carried out for each combination of functions and error distributions. Our sample size was $n_y = 50$, with $t_i = i/50$. We estimated y at $i/500, i = 1, \cdots, 500$, using a cubic smoothing spline, with the smoothing parameter chosen by generalized cross-validation.
* Our standard errors (SE's) have been rounded to the nearest one thousandth.

Simulated Distribution of $\hat{\rho}$: One-sample Case: $n = 100$

Functions	ρ	Dist	Mean (SE*)	% of 1's
$x(t) = t$	1	Normal	1.000 (.000)	96.1
$y(t) = t$		t_3	.998 (.009)	92
$x(t) = t$	1	Normal	.998 (.000)	83.4
$y(t) = t^2$		t_3	.994 (.001)	79.6
$x(t) = t$	0.992	Normal	.991 (.001)	66.4
$y(t) = (t - .1)^2$		t_3	.984 (.001)	67.5
$x(t) = t$	0.936	Normal	.952 (.002)	29.1
$y(t) = (t - .2)^2$		t_3	.947 (.002)	39.6
$x(t) = t$	0.784	Normal	.810 (.003)	1.7
$y(t) = (t - .3)^2$		t_3	.819 (.004)	12.6
$x(t) = t$	0.488	Normal	.510 (.003)	0
$y(t) = (t - .4)^2$		t_3	.528 (.005)	1.6
$x(t) = t$	0.0	Normal	.008 (.004)	0
$y(t) = (t - .5)^2$		t_3	.010 (.006)	.3

Table 2: A thousand simulations were carried out for each combination of functions and error distributions. Our sample size was $n_y = 100$, with $t_i = i/100$. We estimated y at $i/500, i = 1, \cdots, 500$, using a cubic smoothing spline, with the smoothing parameter chosen by generalized cross-validation.
* Our standard errors (SE's) have been rounded to the nearest one thousandth.

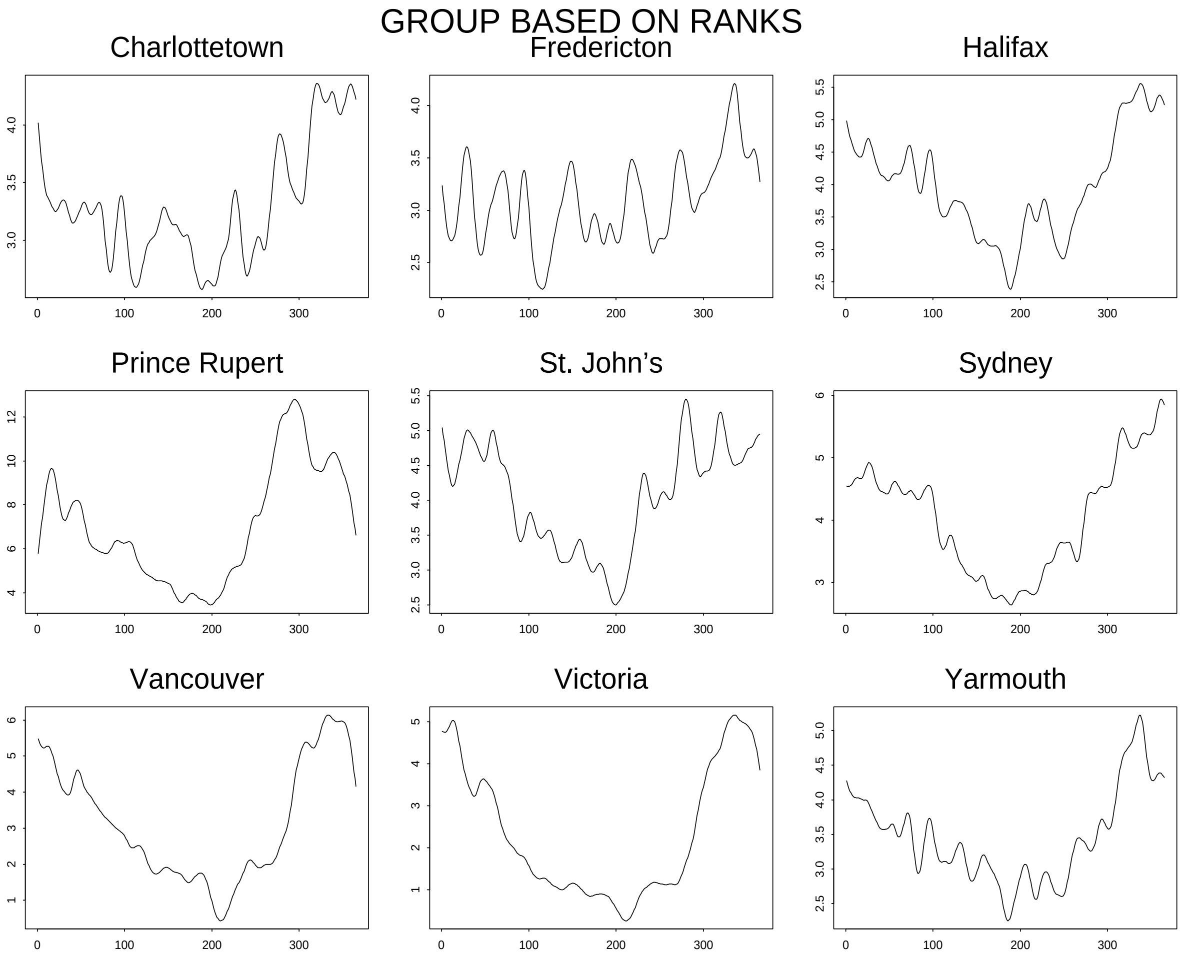

Figure 1: One of four groups from cluster analysis of precipitation data. Precipitation is high in the winter, low in the summer.

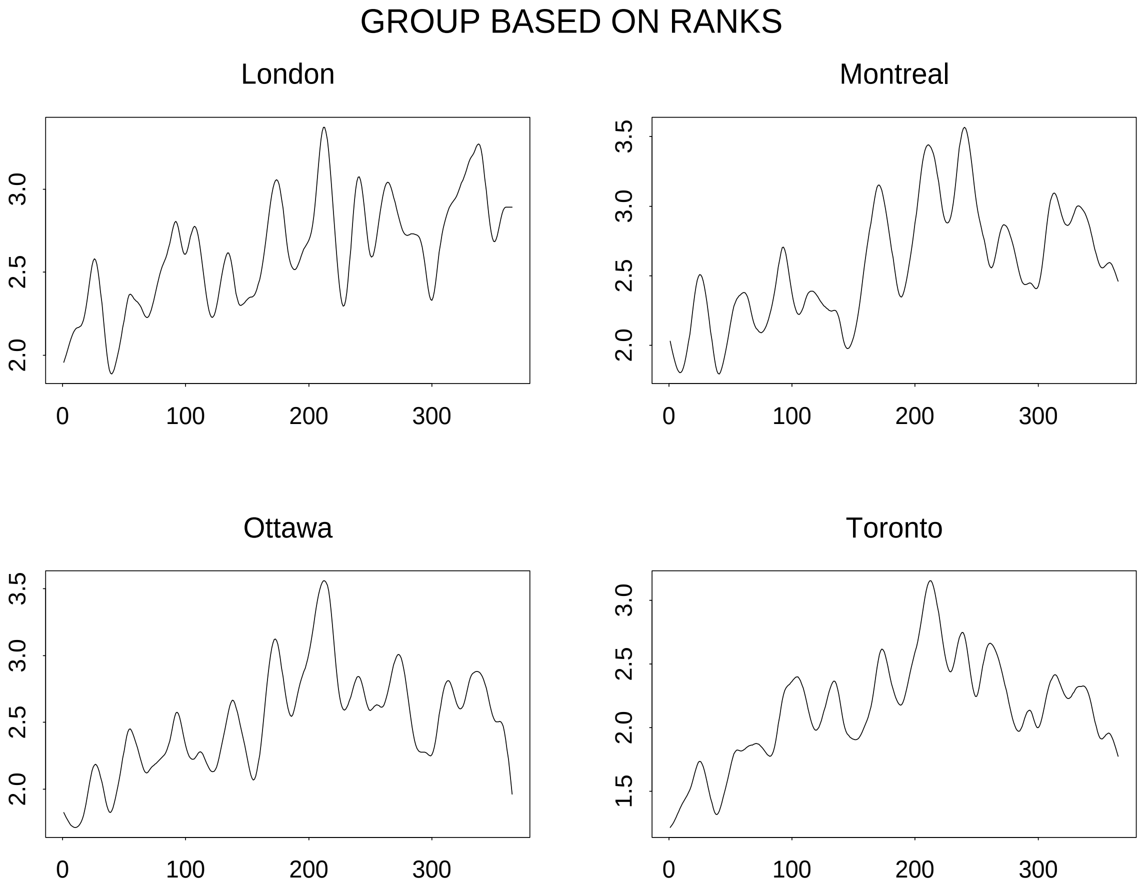

Figure 2: One of four groups from cluster analysis of precipitation data.

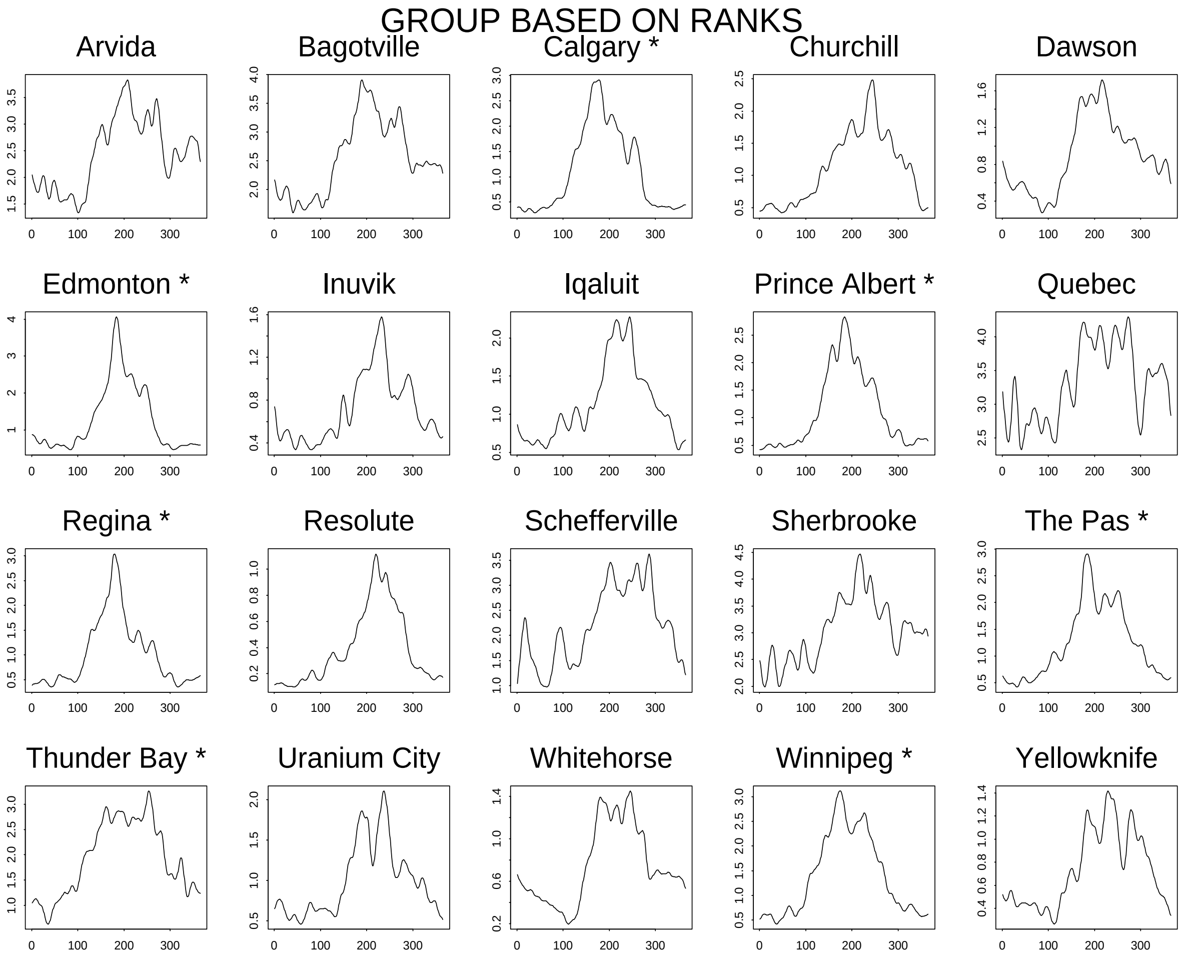

Figure 3: One of four groups from cluster analysis of precipitation data. There is high precipitation in late summer, early fall. The starred plots denote those placed in group four in a cluster analysis using correlation, not rank correlation.

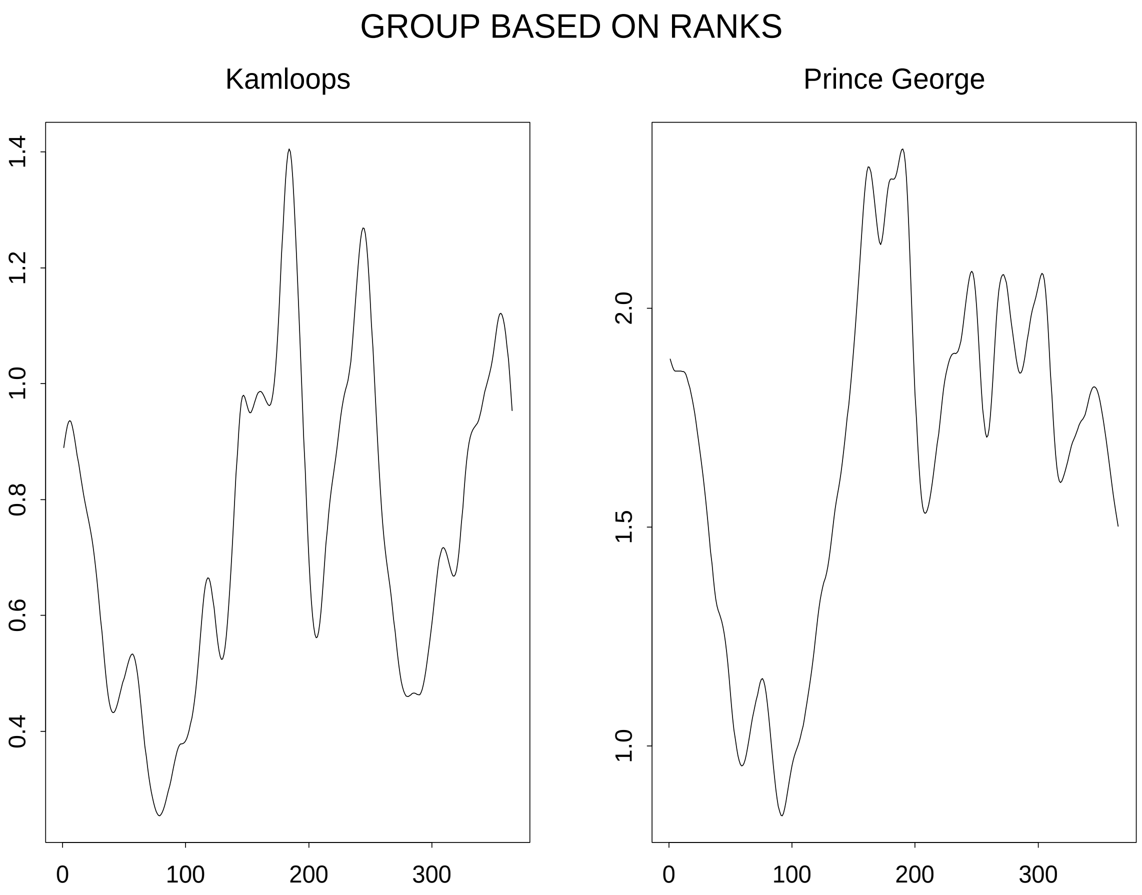

Figure 4: One of four groups from cluster analysis of precipitation data.

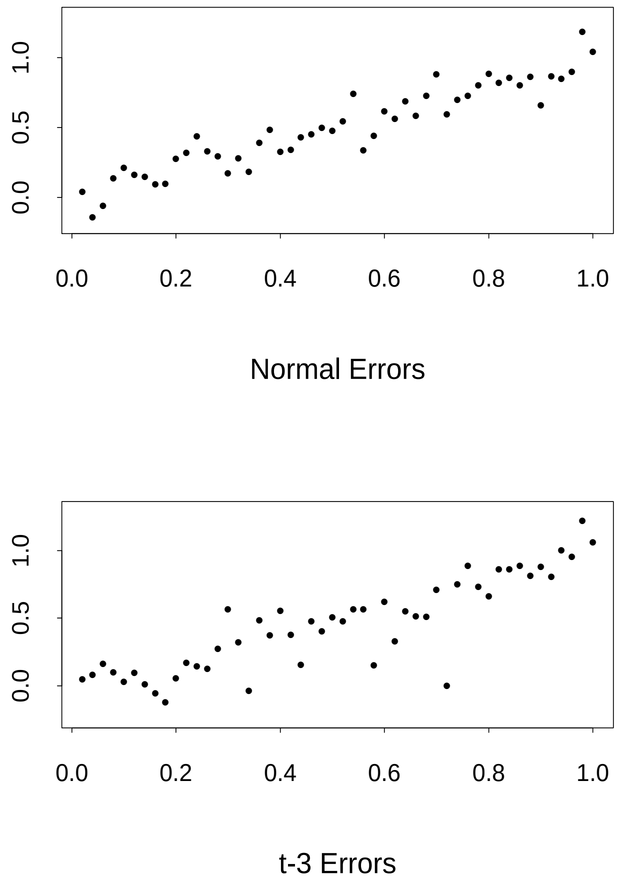

Figure 5: Data sets generated from $y(t) = t$ with normal errors (top) and t_3 errors (bottom).

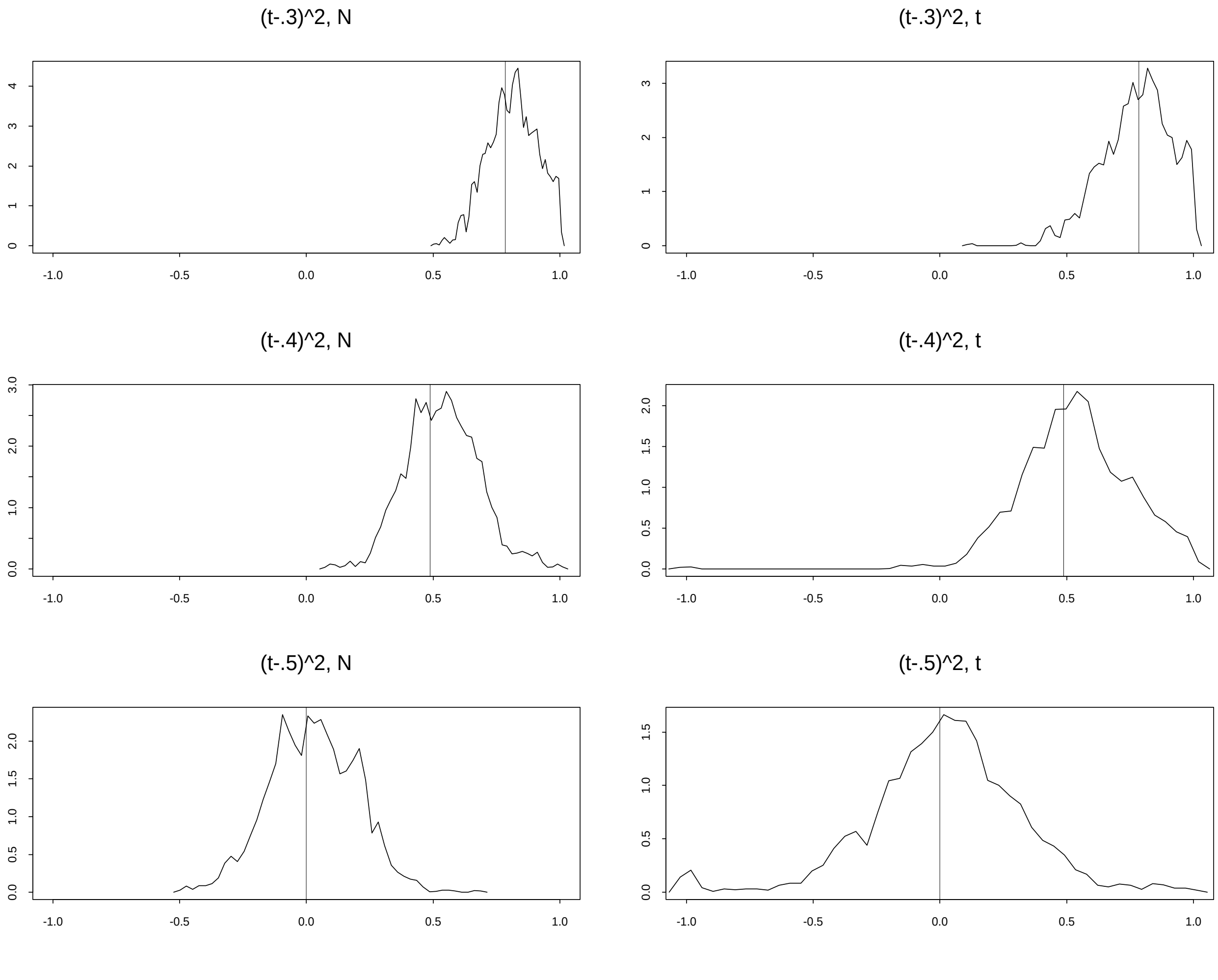

Figure 6: Densities of simulated $\hat{\rho}$'s, as described in Section 5 for the one-sample problem, $n = 50$. Values of $\hat{\rho}$ equal to one have been removed. In the left column, errors were normally distributed, in the right, t_3 distributed. All $\hat{\rho}$'s were based on $x(t) = t$, known. The first row contains results with $y(t) = (t - .3)^2$, the second row contains results with $y(t) = (t - .4)^2$, and the third row contains results with $y(t) = (t - .5)^2$.