Variance Estimation in Spatial Regression Using a Non-parametric Semivariogram Based on Residuals. — Source link ↗

Hyon-Jung Kim, Dennis D. Boos

Institutions: University of Oulu, North Carolina State University

Related papers:

- An introduction to the bootstrap

- Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis

- Likelihood-Based Inference for Extreme Value Models

- Applications in Statistics

- The Effect of Simulation Order on Level Accuracy and Power of Monte Carlo Tests

Variance Estimation in Spatial Regression Using a Nonparametric Semivariogram Based on Residuals

Hyon-Jung Kim and Dennis D. Boos[1]

Abstract

The empirical semivariogram of residuals from a regression model with stationary errors may be used to estimate the covariance structure of the underlying process. For prediction (kriging) the bias of the semivariogram estimate induced by using residuals instead of errors has only a minor effect because the bias is small for small lags. However, for estimating the variance of estimated regression coefficients and of predictions, the bias due to using residuals can be quite substantial. Thus we propose a method for reducing this bias. The adjusted empirical semivariogram is then isotonized and made conditionally negative-definite and used to estimate the variance of estimated regression coefficients in a general estimating equations setup. Simulation results for least squares and robust regression show that the proposed method works well in linear models with stationary correlated errors.

KEYWORDS: Estimating equations, Matérn family, REML, Sandwich variance estimator.

Institute of Statistics Mimeo Series No. 2524

May 28 2001

---

1   Introduction

Random processes over time or space typically have the property that nearby observations tend to be more alike than observations far apart. Scientific studies of such processes often involve modeling and estimation of a mean response with random errors assumed to be from a stationary process. For example, an additive error linear model is given by

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{x}_i$ may consist of variables that are a function of location as well as other covariates, and the e_i are errors whose correlation depends only on the distance between observations. Two applications of model (1) are as follows.

Example 1. For a project of the USDA Forest Service, the ozone index W126 (Lefohn and Runeckles, 1987) was desired at a number of sites in the northeastern United States where plant damage on bio-indicator plants had been measured. W126 readings and some covariates were available for 174 other sites. Thus a regression model was developed between W126 and the following covariates: x_1 = elevation from sea level in meters, x_2 = mean annual rainfall in inches, x_3 = mean annual temperature in Fahrenheit, x_4 = the Palmer drought severity index, and x_5 = the distance in meters to the closest city with population over 50,000. We focus here on getting standard errors for the regression coefficients, but it was also of interest to get standard errors for the kriging predictions at the sites of the bio-indicator site monitors. The OLS coefficients are *intercept*: -25.2 (23.5), x_1: .0063 (.003), x_2: $-.26$ (.29), x_3: 1.23 (.42), x_4: .60 (1.3), x_5: $-.000064$ (.00002) The standard errors from our proposed method are in parentheses and tell us that x_2 = rainfall and x_4 = Palmer drought severity may be dropped from the model.

Example 2. Global warming has been the focus of many studies, and is still a hotly debated topic. The data in Figure 1, taken from Jones et al. (2000), are global mean annual temperatures for 1856-1999 with the 1961-1990 mean subtracted. The data are overlaid with a simple OLS fit of the temperature deviations to x_1 = year $-$ ymean and x_2 = (year $-$ ymean)^2, where ymean = 1927.5 is the mean of the years. The estimated linear and quadratic coefficients are 0.00425 and .0000421, respectively, with standard errors from our method of 0.000643 and 0.0000167. Since for any stationary error structure these OLS coefficients are uncorrelated (due to centering the years), an approximate Wald test statistic for zero linear and quadratic coefficients is $(0.00425/.000643)^2 + (.0000421/0.0000167)^2 = 50.0$. Although a chisquared distribution with two degrees of freedom may
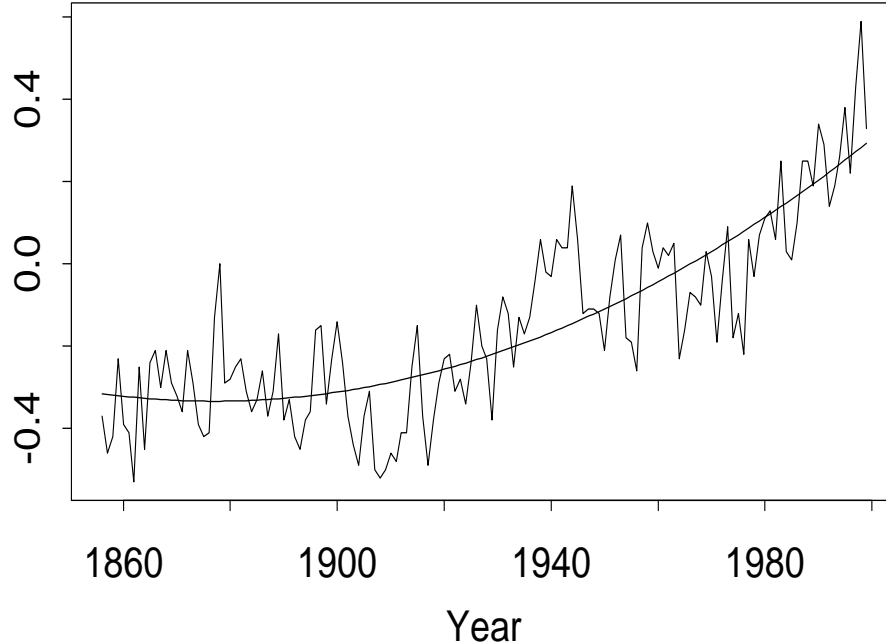
Figure 1: Global mean annual temperatures for 1856-1999 with the 1961-1990 mean subtracted; from Jones et al. (2000).

be a bit optimistic for the null distribution in this situation, the simulations in Table 3 suggest that 50.0 is highly significant. Other approaches using a linear trend may be found in Bloomfield and Nychka (1992) and Sun and Pantula (1999). Wu, Woodroofe, and Mentz (2001) use isotonic regression methods to test for global warming.

One standard approach for model (1) with normally distributed errors is to assume a particular parametric semivariogram model such as an exponential and use restricted maximum likelihood (REML) for estimating the parameters of the model and then to use estimated generalized least squares (EGLS) for estimating $\boldsymbol{\beta}$. Alternatively, one can avoid the normality assumption and use weighted least squares with the empirical semivariogram of the residuals to estimate the parameters of the semivariogram model (see Cressie, 1993, p. 94-99, 165-170). In either case the variance of the EGLS $\widehat{\boldsymbol{\beta}}$ is estimated using standard formulas for generalized least squares with semivariogram estimates inserted where needed. A correct semivariogram model, however, is not always easy to choose, and we may be interested in estimation methods other than least squares.

The purpose of this paper is to give new methods for estimating the variance of $\widehat{\boldsymbol{\beta}}$ when $\widehat{\boldsymbol{\beta}}$ is obtained from a general estimating equations approach, i.e., satisfying $\sum \boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}) = \boldsymbol{0}$, where $\boldsymbol{S}_i$ is usually a function of residuals and weights and $\widehat{\boldsymbol{\sigma}}$ are additional dispersion parameter estimates.

3

This general class includes ordinary least squares (OLS, with $\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}) = (Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}) \boldsymbol{x}_i)$, EGLS, robust regression (e.g., Huber, 1980, Ch. 3, with $\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}) = (\psi[Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}]/\widehat{\sigma}) \boldsymbol{x}_i)$, and generalized estimating equations (GEE, Liang and Zeger, 1986). Although these estimation methods are suitable for a much larger class of models than (1), we will focus on methods for (1). However, a parametric semivariogram model will not be chosen; rather, we will estimate the semivariogram subject only to a monotonicity constraint.

Our approach is similar in spirit to that of Lumley and Heagerty (1999), who use a weighted empirical variance estimate for the middle part of the "sandwich" asymptotic variance formula that arises naturally from the estimating equations formulation. Their approach is more general because it allows nonstationary errors; in fact Lumley and Heagerty (1999) unify a variety of nonparametric methods including Newey and West (1987), White and Domowitz (1984), Andrews (1991), Lele (1991), and Yasui and Lele (1997). Related nonparametric methods are found in Carlstein (1986), Sherman (1996, 1997), Garcia-Soidan and Hall (1997), and Heagerty and Lumley (2000).

Our approach differs, though, in two ways from these other methods: we use nonparametric semivariogram estimation for the middle part of the "sandwich," and we explicitly remove bias in the semivariogram estimates that would ordinarily result from using residuals rather than errors (which are of course unknown). It is this latter bias issue that we feel is most important. Any nonparametric variance estimation method based on residuals that does not address the bias issue is doomed to underestimate the variance of estimated coefficients and of predictions. The problem with using residuals has long been recognized in the spatial statistics literature (Matheron, 1971, p. 152-155, and Cressie, p. 165-170). However, when fitting parametric variogram models, the bias is by definition not a problem for REML. It is also claimed (see Cressie, 1993, p. 167-169) that the bias from residuals is not a major problem when estimating parametric semivariograms by weighted least squares because the early lags, where the bias in the empirical semivariogram is small, have the largest weights. In general, though, nonparametric semivariogram methods will inherit the bias problem of the empirical semivariogram, especially at moderate to large distances between observations (see Figure 4 in Section 3.2). We have not seen attempts to correct this bias directly as we suggest in Section 3.2, but Matheron (1973) and Cressie (1987) propose a method similar to differencing in time series for handling the bias issue. For visual confirmation of parametric variogram models, Brownie and Gumpertz (1997) suggest adjusting REML estimates of parametric

4

semivariograms so that graphically they are aligned with a plot of the empirical semivariogram of residuals.

Our nonparametric semivariogram is based on several simple ideas. First we construct a semivariogram estimator in the constant mean case by monotonizing the standard empirical semivariogram (see Section 3.1). We also check for positive-definiteness of the resulting covariance matrix and modify our estimate if it is not positive-definite. Next (Section 3.2) we compute the bias of the residuals-based empirical semivariogram in model (1) when $\boldsymbol{\beta}$ is estimated by OLS. Then we correct for this bias by multiplying the empirical semivariogram by estimated factors computed at each distinct distance between observations. Finally, we monotonize and make the adjusted empirical semivariogram conditionally negative-definite. This nonparametric semivariogram is then used to estimate the variance of the estimated regression coefficients. Monte Carlo results for OLS estimates are given in Section 4.1, and results for robust regression estimates are given in Section 4.2. We begin in Section 2 with a general explanation of how to estimate the variance of regression parameter estimates from data with correlated errors.

2   Variance Estimation of Regression Parameter Estimates

The estimating equations approach provides a general framework for deriving the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ that solves

$$\boldsymbol{G}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{0}$$

(and here for simplicity we have dropped the extra dispersion parameters $\boldsymbol{\sigma}$ mentioned in the Introduction). That is, by Taylor series approximation,

$$\boldsymbol{0} = \boldsymbol{G}_n(\widehat{\boldsymbol{\beta}}) \approx \boldsymbol{G}_n(\boldsymbol{\beta}) + \dot{\boldsymbol{G}}_n(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{R}_n,$$

where

$$\dot{vect}\boldsymbol{G}_n(\widehat{\boldsymbol{\beta}}) = \frac{\partial}{\partial \boldsymbol{\beta}^T} \boldsymbol{G}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} \boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta}).$$

Then, under suitable regularity conditions, $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ is asymptotically normally distributed with variance times n

$$A(\boldsymbol{\beta})^{-1} B(\boldsymbol{\beta}) \left\{ A(\boldsymbol{\beta})^{-1} \right\}^T = \lim_{n \to \infty} [-\mathrm{E}\dot{\boldsymbol{G}}_n(\boldsymbol{\beta})]^{-1} \, n\mathrm{Var}[\boldsymbol{G}_n(\boldsymbol{\beta})] \, \left\{ [-\mathrm{E}\dot{\boldsymbol{G}}_n(\boldsymbol{\beta})]^{-1} \right\}^T, \qquad (2)$$

where it is assumed that there are matrices $A(\boldsymbol{\beta})$ and $B(\boldsymbol{\beta})$ such that $-\mathrm{E}\dot{\boldsymbol{G}}_n(\boldsymbol{\beta}) \to A(\boldsymbol{\beta})$ and $n\mathrm{Var}[\boldsymbol{G}_n(\boldsymbol{\beta})] \to B(\boldsymbol{\beta})$ as $n \to \infty$. In likelihood models $-\mathrm{E}\dot{\boldsymbol{G}}_n(\boldsymbol{\beta}) = n\mathrm{Var}[\boldsymbol{G}_n(\boldsymbol{\beta})]$ is the average Fisher information.

Since the estimator

$$\dot{\boldsymbol{G}}_n(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})$$

will typically satisfy $\dot{\boldsymbol{G}}_n(\widehat{\boldsymbol{\beta}}) - \mathrm{E}\dot{\boldsymbol{G}}_n(\boldsymbol{\beta}) \xrightarrow{p} \boldsymbol{0}$ even for correlated data, the problem of finding a consistent estimator of the asymptotic variance (2) is reduced to finding a consistent estimator for the middle term,

$$n\mathrm{Var}\left[\boldsymbol{G}_n(\boldsymbol{\beta})\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta})\boldsymbol{S}_j(Y_j, \boldsymbol{x}_j, \boldsymbol{\beta})^T\right].$$

Lumley and Heagerty (1999) point out that the empirical estimator

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})\boldsymbol{S}_j(Y_j, \boldsymbol{x_j}, \widehat{\boldsymbol{\beta}})^T = \frac{1}{n}\left[\sum_{i=1}^{n}\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})\right]\left[\sum_{j=1}^{n}\boldsymbol{S}_j(Y_j, \boldsymbol{x_j}, \widehat{\boldsymbol{\beta}})^T\right]$$

is identically zero by the definition of $\widehat{\boldsymbol{\beta}}$. This of course contrasts with the common situation where one can average over independent replications. Thus Lumley and Heagerty (1999) suggest estimating the middle term by

$$\widehat{J}_n(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ijn}\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \widehat{\boldsymbol{\beta}})\boldsymbol{S}_j(Y_j, \boldsymbol{x}_j, \widehat{\boldsymbol{\beta}})^T$$

where $w_{ijn} \to 1$ as $n \to \infty$ but $w_{ijn} \to 0$ as the distance between two locations, $d(i, j) \to \infty$ for fixed n. They also show that the methods proposed earlier by Newey and West (1987), White and Domowitz (1984), and Andrews (1991) can all be viewed as weighted empirical estimators with different choices of the weight w_{ijn}. Although these estimators can be used for nonstationary models, one common problem is that they are not very efficient for highly correlated data (Andrews, 1991).

To implement our approach we first make the simplifying assumption that $\boldsymbol{S}_i$ has the form $\boldsymbol{S}_i(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) = w_i(\boldsymbol{\beta})S(e_i(\boldsymbol{\beta}))\boldsymbol{x}_i$, where S is now a real-valued function and the $e_i(\boldsymbol{\beta})$ are from a strictly stationary process. Thus, the $S(e_i(\boldsymbol{\beta}))$ are also from a strictly stationary process, and we will estimate the latter process using semivariogram techniques. We then employ it to estimate the middle term of (2) utilizing the relationship between semivariogram and covariance functions under second-order stationarity. We present the details of estimating the semivariogram in the next section.

3 Nonparametric Semivariogram Estimation

A process $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D, D \subset \mathbb{R}^d\}$ is called intrinsically stationary when it has constant expectation and the variance of the increments depends only on the difference of locations. For an intrinsically stationary process, the semivariogram is defined at locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ by

$$\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2) = \frac{1}{2}\text{Var}[Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2)].$$

It is called isotropic if $\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2)$ is only a function of the Euclidean distance $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ between locations. We then use the simpler notation $\gamma(h)$ where h is the distance between locations. The function $2\gamma(\cdot)$ is called the variogram. A valid semivariogram needs to be conditionally negative-definite; that is, it must satisfy

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\boldsymbol{s_i} - \boldsymbol{s_j}) \leq 0$$

for each set of locations $\boldsymbol{s_1}, ..., \boldsymbol{s_n}$ and all $\lambda_1, \ldots, \lambda_n$ such that $\sum_{i=1}^{n} \lambda_i = 0$ (see Cressie, 1993, p. 86). For a sample of given realizations from $Z(\cdot)$, the empirical variogram is the unbiased estimator of an isotropic variogram given by

$$2\widehat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \{Z(\boldsymbol{s_i}) - Z(\boldsymbol{s_j})\}^2$$

where

$$N(h) = \{(\boldsymbol{s_i}, \boldsymbol{s_j}) : \|\boldsymbol{s_i} - \boldsymbol{s_j}\| = h : i, j = 1, 2, ..., n\}$$

and $|N(h)|$ is the number of distinct pairs in $N(h)$. Although the empirical variogram is unbiased for the variogram, it cannot be used directly in procedures such as kriging (spatial prediction) because it may not be conditionally negative-definite.

One standard approach has been to choose a parametric variogram model (which by definition is conditionally negative-definite) and fit it by restricted maximum likelihood (REML), maximum likelihood (ML), or weighted nonlinear least squares. There are a number of widely used parametric variogram models based on isotropic processes such as the exponential, spherical, and Gaussian. The Matérn class, originally given by Matérn (1960), allows a wide range of flexibility in that it has a parameter which controls the smoothness of a random field. The class can be defined by its isotropic autocovariance function:

$$C(h) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}h}{\rho}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{2\nu^{1/2}h}{\rho}\right)$$

where σ is the scale parameter, ν the shape parameter, and ρ measures how quickly the correlation of the random field decays with distance. The function $\Gamma(\cdot)$ is the gamma function and $\mathcal{K}_\nu$ is the modified Bessel function of the third kind of order ν (Stein 1999). When $\nu = \frac{1}{2}$, the model becomes the exponential model, and the limit $\nu \to \infty$ corresponds to a Gaussian model. The underlying true variogram is rarely known, and selection of a variogram model is quite arbitrary in practice. The Matérn class appears to be the best choice of the present parametric models to estimate the dependence structure of a process since it includes or approximates a number of common models.

There have been several attempts to avoid selecting variogram models via nonparametric variogram estimation (Shapiro and Botha, 1991; Cherry et al., 1996; Lele, 1995; Hall et al., 1994; Barry and Ver Hoef, 1996; and Gorsich and Genton, 2000). Shapiro and Botha(1991) appear to be the first to consider a nonparametric semivariogram estimate based on Bochner's theorem. Cherry et al. (1996) implemented the Shapiro-Botha estimator using the statistical package S-plus and compared its performance to parametric estimation of the semivariogram using nonlinear least squares. They found good performance of their nonparametric estimates and sometimes better performance than the traditional parametric approach. One problem with their method is that the sill estimates tend to be biased and highly variable. Cherry (1997) suggested a simple remedy for this sill problem, but semivariogram values other than the sill also seem to have high variability as well. Lele (1995) provided a nonparametric estimator of the semivariogram using a spline function and included a study of the performance of his estimator in terms of prediction and prediction error.

3.1 Monotone Semivariogram Estimation

Most parametric covariance models used in spatial analyses have correlations that are monotone decreasing with distance. Moreover, most physical processes exhibit this monotone behavior as well. Thus, the centerpiece of our approach is to assume that the correlations are monotone decreasing with distance. This is a much weaker assumption than any of the common parametric semivariogram models.

Our basic approach to produce a nonparametric monotone semivariogram estimator is to apply the pool adjacent violators algorithm (PAVA, Barlow et al. 1972, p.13) to empirical semivariogram estimates. Although the variance of the empirical semivariogram at a lag h is not exactly proportional to the inverse of the number of pairs (e.g., Genton, 1998, p. 328), we use that weighting in

the PAVA routine. This is similar to Cressie's (1993, p. 96) weighting in nonlinear least squares fitting of parametric semivariograms except that for simplicity we ignore the extra weighting factor due to the value of the semivariogram.

The basic idea of PAVA is the following: starting with y_1, move to the right and stop if $y_i > y_{i+1}$. In that case replace y_i and y_{i+1} by their weighted average y_i^*. Then move to the left to make sure that $y_{i-1} \leq y_i^*$. If $y_{i-1} > y_i^*$, then replace y_{i-1} with the weighted average of all three. Continue moving to the left and checking the monotone requirement, and then proceed again to the right. This process of averaging and back-averaging is continued until the right end point is reached.

Consistency of the monotonized empirical semivariogram follows simply from the consistency of the empirical semivariogram as long as the underlying semivariogram is monotone. For example, suppose that the empirical semivariogram $\hat{\gamma}(h)$ is strongly consistent for a strictly monotone semivariogram $\gamma(h)$ at the points $h_1, \ldots, h_k$. Then for a sufficiently large sample size with probability one, $\hat{\gamma}(h_1) < \hat{\gamma}(h_2) < \cdots < \hat{\gamma}(h_k)$, and the monotonized version is the same as the empirical semivariogram. This same idea extends to cases that are not strictly monotone.

In many situations, isotonic estimators have improved mean squared error properties over the original estimators used in their construction. We will illustrate that property with Monte Carlo simulation.

Since the variance of the empirical semivariogram is large for large lags, we follow the common practice of truncating the isotonized semivariogram at half the maximum distance found in the data set. In order not to induce upward bias at that distance, we monotonize a larger set of lags before truncating. For example, for a time series on the equally spaced time points of $1, \ldots, 100$ we monotonize on the lags $1, \ldots, 70$ (number of pairs ≥ 30) and then truncate at 50.

Figure 2 shows the empirical semivariogram and the isotonized semivariogram for one data set on a 10 x 10 grid generated from an exponential semivariogram with range parameter $= 1$ and sill $= 3$, $\gamma(h) = 3[1 - \exp(-h/3)]$. Note that the two estimates only coincide at a few points. Figure 3 gives averages of 1000 replications of the situation in Figure 2. We can see that the empirical semivariogram is unbiased as advertised, and that the monotone estimate is somewhat biased downward in the middle. Table 1 shows, however, that the monotone estimate has lower mean squared error on the log scale for distances beyond 3. We use the log scale because the mean squared error for variance-like quantities is well-known to reward negatively biased estimators too
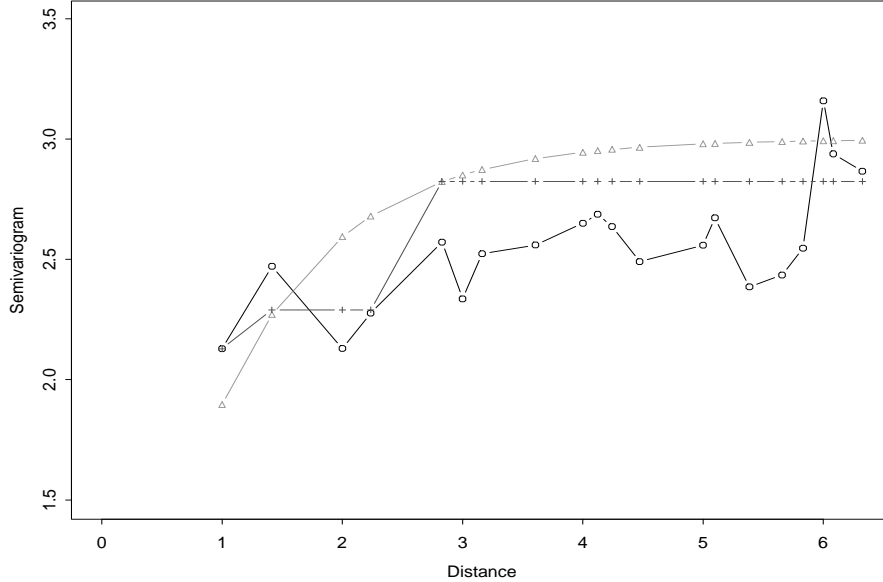
Figure 2: Empirical semivariograms for data on a 10 x 10 grid generated from a constant mean process with an exponential (sill=1,range=3) semivariogram (Δ): empirical semivariogram (O) and isotonized version (+).

much. Note that the standard errors of differences of estimates in Table 1 are lower than the reported standard errors for individual entries.

Table 1. Mean squared errors of the logarithm of empirical and monotone semivariogram estimates for constant mean data on a 10×10 grid with exponential semivariogram. Results based on 1000 replications. Average standard errors are in the last row.

	Distance	1.00	1.41	3.00	4.12	6.32
Range=1, Sill=3	Empirical	0.023	0.029	0.043	0.046	0.053
	Monotone	0.023	0.028	0.039	0.040	0.043
Range=2, Sill=3	Empirical	0.022	0.030	0.060	0.078	0.110
	Monotone	0.022	0.030	0.058	0.075	0.096
	Std. Errors	0.001	0.001	0.003	0.003	0.004

Very often, the isotonized semivariogram is already conditionally negative-definite. Sometimes it is not, and then we suggest using the spectral decomposition of the covariance matrix followed by
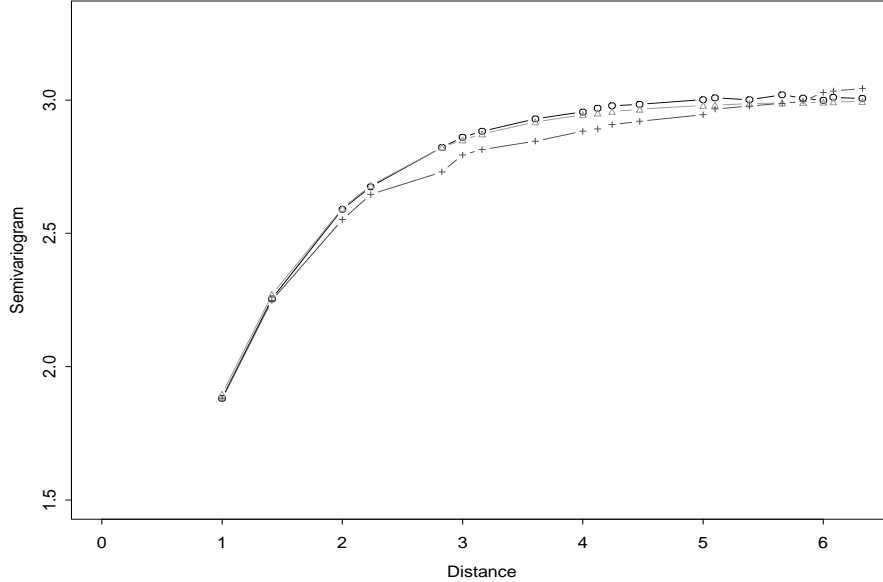
Figure 3: Average of 1000 empirical semivariograms for data on a 10 x 10 grid generated from a constant mean process with an exponential (sill=1,range=3) semivariogram (Δ): empirical semivariogram (O) and isotonized version (+). Standard deviations of estimates are approximately bounded by .02.

replacement of the negative eigenvalues by small positive ones (see Rousseeuw and Molenberghs, 1993, for other methods). For the variance estimation discussed in Sections 4 and 5, this has virtually no effect. However, for other purposes such as kriging, it may be required. Also, the isotonic estimators have a "boxy" appearance, and sometimes we find it more appealing to smooth the isotonized semivariogram using a spline or other smoother.

3.2   Bias Correction of Residuals-Based Semivariogram

Consider model (1) where the errors e_i are drawn from a mean zero, second-order stationary random error process. We shall assume that the unknown regression coefficients $\boldsymbol{\beta}$ are estimated by ordinary least squares. A natural approach is then to construct the empirical semivariogram from the residuals. Unfortunately, the distribution of the residuals is not the same as the errors, and the empirical semivariogram is seriously biased downward. Figure 4 shows the average of the empirical semivariogram for 1000 replications of an exponential semivariogram with range parameter = 1
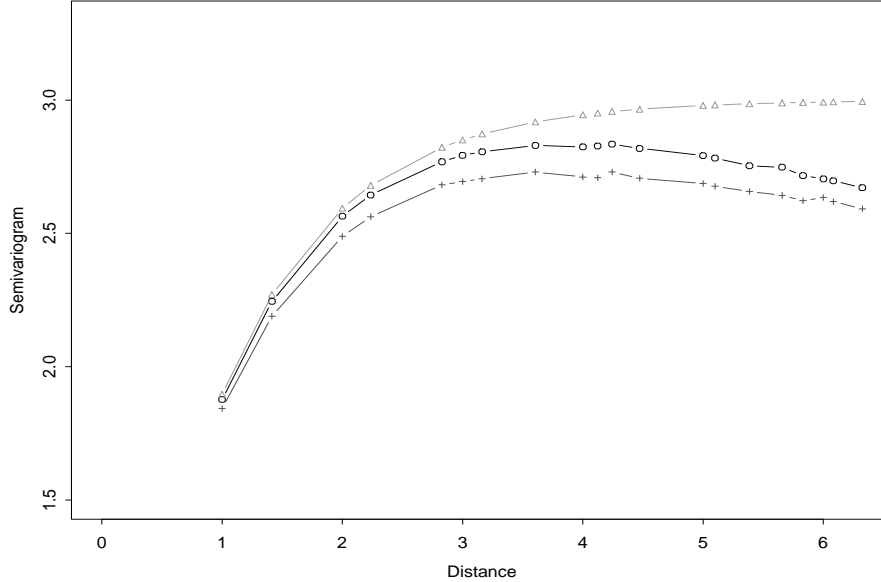
Figure 4: Exponential(sill=1,range=3) semivariogram (Δ) for data on a 10 by 10 grid, average of 1000 empirical semivariograms from residuals with $p = 3$ estimated parameters (O) and with $p = 6$ estimated parameters (+). Standard deviations of estimates are approximately bounded by .02.

and sill $= 3$ for a 10 by 10 grid. The middle curve is for residuals based on fitting an intercept and the location coordinates (x and y). The lower curve is for residuals from a fit with the same three variables and with three more independent variables (randomly generated standard normal variables).

To understand more clearly the effect of the residuals, let V denote the covariance matrix of the errors for a sample with n data points. Then simple calculations show that the covariance matrix of the residuals $\widehat{e}_i = Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}$ is given by $(I - P)V(I - P)$, where I is the n dimensional identity matrix and $P = X(X^T X)^{-1} X^T$ is the projection matrix of X, where X is formed from the row vectors $\boldsymbol{x}_i^T$.

The expected value of the empirical semivariogram of the residuals at lag h is

$$
\begin{aligned}
\mathrm{E}_{res}(h) &= \frac{1}{2|N(h)|} \sum_{N(h)} \mathrm{E}(\widehat{e}_{i+h} - \widehat{e}_i)^2 \\
&= \frac{1}{2|N(h)|} \sum_{N(h)} \left[\mathrm{E}(\widehat{e}_{i+h}^2) + \mathrm{E}(\widehat{e}_i^2) - 2\mathrm{E}(\widehat{e}_{i+h}\widehat{e}_i) \right]
\end{aligned}
$$

12

$$\approx \quad \frac{1}{n}\text{trace}(I - P)V(I - P) - \frac{1}{|N(h)|}\sum_{N(h)}[(I - P)V(I - P)]_{i,i+h}\,,$$

where the approximation comes in by taking the average of all of the diagonal elements of $(I - P)V(I - P)$ instead of the subset of the diagonal elements implied by summing over elements h distance apart. This approximation is not necessary but makes the computing considerably easier.

The expected value of the empirical semivariogram of the errors at lag h is of course $\gamma(h)$, but we can write it in a form similar to the above:

$$
\begin{aligned}
\text{E}_{err}(h) &= \frac{1}{2|N(h)|}\sum_{N(h)}\text{E}(e_{i+h} - e_i)^2 \\
&= \sigma^2 - \frac{1}{|N(h)|}\sum_{N(h)}[V]_{i,i+h}\,,
\end{aligned}
$$

where σ^2 is the variance of the errors. Now the ratio of these quantities, $\text{fac}(h) = \text{E}_{err}(h)/\text{E}_{res}(h)$, is essentially the ratio of the true semivariogram $\gamma(h)$ in Figure 4 to the average of the empirical semivariograms.

We feel that the downward bias in Figure 4 in unacceptable. Moreover, applying PAVA to the residuals-based empirical semivariogram will certainly improve it, but the scope for improvement is quite limited. Thus we feel it is important to first adjust the empirical semivariogram before monotonizing. Our approach then is to estimate $\text{fac}(h)$ by using estimated covariances obtained from the monotonized version of the residuals-based empirical semivariogram. We multiply the original residuals-based empirical semivariogram by these estimated factors resulting in a bias-reduced empirical semivariogram. Finally we monotonize this bias-reduced empirical semivariogram to obtain our estimated semivariogram. It is also possible to iterate the factor estimation step, but we found that it did not make much difference. Figure 5 shows the average of the residuals-based empirical semivariogram, the bias-reduced empirical semivariogram, and the monotonized version for 1000 samples from a linear model with intercept and locations fitted ($p = 3$ case) and errors generated from an exponential semivariogram with range parameter $= 2$ and sill $= 3$. Comparing Figure 5 with Figure 3, one can see that monotonizing has a similar effect; that is, it tends to pull down on average the bias-reduced empirical semivariogram. Table 2 shows that the monotonized version has improved mean squared error properties relative to the original residuals-based empirical semivariogram.

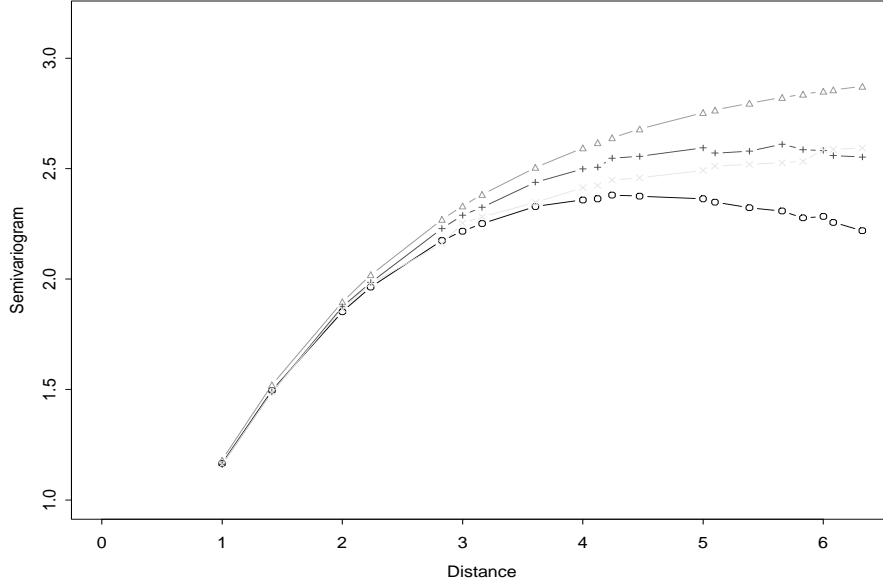A number of authors (see Cressie, 1993, p. $167-169$) have shown concern about the bias

Figure 5: Exponential(sill=2,range=3) semivariogram (Δ) on a 10 x 10 grid and the average of 1000 replications of semivariogram estimates based on residuals from $p = 3$ estimated parameters: empirical semivariogram from residuals (O), bias-reduced empirical variogram (+), monotonized bias-reduced empirical variogram ($\times$). Standard deviations of estimates are approximately bounded by .02.

Table 2. Mean squared errors of logarithms of semivariogram estimates: empirical, bias-reduced empirical and monotone. Data are ¿from a 10×10 grid with exponential semivariogram. Results based on 1000 replications. Average standard errors are in the last row.

	Distance	1.00	1.41	3.00	4.12	6.32
Range=1, Sill=3	Empirical	0.023	0.029	0.046	0.052	0.067
	Bias-reduced	0.023	0.029	0.046	0.051	0.060
	Monotone	0.023	0.029	0.042	0.045	0.047
Range=2, Sill=3	Empirical	0.023	0.032	0.069	0.098	0.173
	Bias-reduced	0.023	0.032	0.067	0.092	0.139
	Monotone	0.023	0.032	0.067	0.091	0.116
	Std. Errors	0.001	0.001	0.003	0.003	0.004

problems when the variogram is based on least squares residuals. However, to our knowledge, no methods have successfully corrected the bias of residuals-based empirical variograms. In general, the bias of such estimators is small at lags near the origin but more substantial at distant lags (see Figure 4). Cressie (1993, p. 167-168) concludes that the effect of bias on kriging will be small if
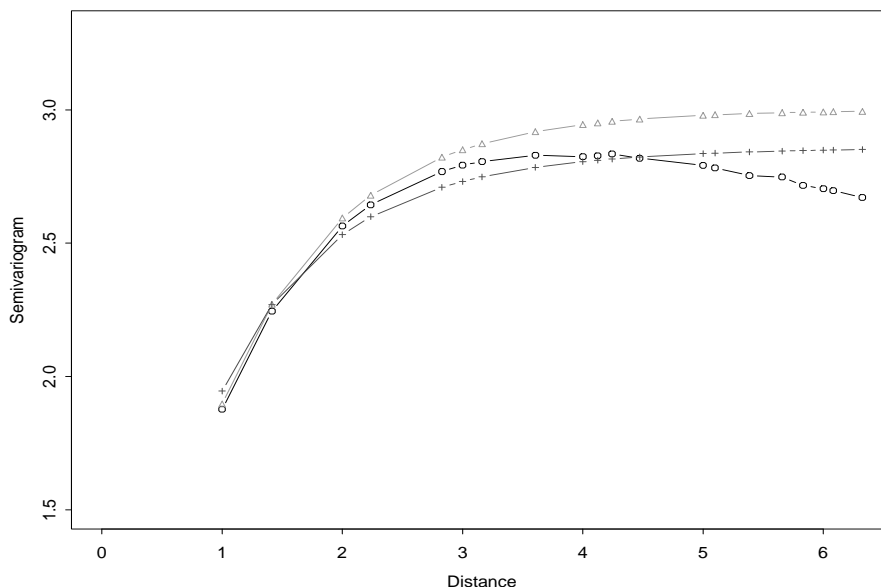

Figure 6: Exponential(sill=1,range=3) semivariogram (Δ), average of 1000 replications of semivariogram estimates based on residuals from $p = 3$ estimated parameters: empirical semivariogram from residuals (O) and weighted nonlinear least squares parametric estimate (+). Standard deviations of estimates are approximately bounded by .02.

a parametric variogram is fitted with more weight given to the estimates at small lags, such as by the weighted least squares method. However, he notes that a kriging variance can be more influenced by the residuals-based variogram estimates. Here our main concern is with estimation of the variance of estimated regression parameters, and the substantial bias at large lags should not be ignored.

Figure 6 shows that a parametric semivariogram estimated by WNLS fitted to the residuals-based empirical semivariogram has considerable bias. The average values of the range and sill using the Monte Carlo results that produced Figure 6 are 0.85 and 2.82, respectively (true values are 1.0 and 3.0). If instead one uses the bias-reduced empirical semivariogram, then the WNLS method produces nearly unbiased results: the corresponding average range and sill estimates are 1.01 and

3.03, respectively.

The estimation approach least likely to be affected by residual bias is restricted maximum likelihood (REML). REML maximizes the likelihood of error contrasts that do not depend on estimated regression parameters. The REML method is a useful tool for analyzing data with spatial variation since it does not suffer from the often severe underestimation of the parameters that regular maximum likelihood does.

One disadvantage of likelihood estimation procedures is that they rely on the Gaussian assumptions, and such assumptions are often inappropriate; the underlying distribution of most processes is not known or contamination of the distribution can occur by a few errant observations. Nonparametric estimation of the variogram may provide less sensitive estimates for some non-Gaussian models. Of course, least squares estimation of regression coefficients is also questionable in the face of non-Gaussian errors. In Section 4.2 we consider using robust regression estimates.

For confirmation of parametric semivariogram models, Brownie and Gumpertz (1997) suggest adjusting parametric semivariograms fitted by REML so that plots of the adjusted semivariograms will be consistent with the residual-based empirical semivariogram. Their approach is similar in spirit to our bias-reduced empirical semivariogram; the difference is they adjust the fitted semivariogram instead of the empirical semivariogram.

4 Variance Estimation of Regression Parameter Estimates Based on Semivariograms

4.1 Ordinary Least Squares Estimates of $\boldsymbol{\beta}$

Under the model (1) with second order stationary errors and $\boldsymbol{\beta}$ estimated by ordinary least squares, the variance of $\widehat{\boldsymbol{\beta}}$ is given by

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = (X^T X)^{-1} X^T V X (X^T X)^{-1},$$

where V is the covariance matrix of the errors. For estimating this variance, we utilize the relationship between a semivariogram function $\gamma(\boldsymbol{h})$ and a covariance function $C(\boldsymbol{h})$ under second-order stationarity, $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$ and $C(\boldsymbol{h}) = \gamma(\infty) - \gamma(\boldsymbol{h})$, $C(\boldsymbol{0}) = \gamma(\infty)$ is the variance of the

errors. Then we use the monotone semivariogram estimates proposed in Section 3.2 to estimate V and substitute in the above expression.

Recall, though, that Figure 5 shows how the correlations tend to be overestimated by our monotone semivariogram estimate. This happens because on average in the middle distances the monotone estimate is biased downward. On a sample by sample case, this can be seen when the middle part of the monotone estimate is flat but rises near the half maximum distance. In such a case, the correlations are positive for a large number of distances h. This results in variance estimates that can be too large. Thus we use a cutoff rule for correlations: we set any estimated correlation to 0 if its value is smaller than $1/\sqrt{n}$. This is a fairly arbitrary rule, but it is similar in spirit to the WEAVE weight functions of Lumley and Heagerty (1999). Basically, the number of nonzero correlations cannot be growing too quickly with n if we are to have good variance estimates.

Table 3 shows results for errors from an ar(1) process with $\rho = .1$, .5, and .9. The X matrix consists of an intercept term, a linear trend, and a seasonal term and was taken from the simulation study of Lumley and Heagerty (1999, p. 469). For comparison, we also report results using the approach of Lumley and Heagerty (1999) denoted by "WEAVE" in Table 3. We obtained their results using a program from their website with default values. Perhaps other settings would have performed better for the $\rho = .9$ case. Our results, denoted by "Monotone ESV" for monotone empirical semivariogram, are reasonably on target (values close to 1.00), but more variable than the WEAVE results. Part of that higher variability can be explained by the fact that our estimates are larger on average than the WEAVE estimates.

Table 4 shows results for Gaussian spatial processes observed on a 10×10 grid, that are generated using an exponential variogram model with the sill = 3 and range parameter = 1 and 2, respectively. The average correlation is 0.04 for the model with the range parameter equal to 1, and is 0.14 for the model with the range parameter 2. (For comparison with the time series simulation in Table 3, note that the average correlation for $\rho = .1$, .5, and .9 are .002, .02, and .16, respectively.) A mean surface was added to these Gaussian processes to describe a spatial trend: $f(\mu) = \beta_0 + \beta_1 x + \beta_2 y$, where (x, y) defines a point in $\mathbb{R}^2$ with $\beta_0 = 0.0$, $\beta_1 = 0.9$, $\beta_2 = 0.06$. Three covariates, a column of ones for the intercept and x (longitude) and y (latitude), were fitted to the model to estimate the mean surface. We then compared our methods to REML estimation using an assumed Matérn variogram model (without nugget, which is correct here). We see in Table 4 that our estimates are a little closer to the target value and a little more variable than the

Table 3. Average of Standard Deviation Estimates (divided by
true standard deviation) for Regression from Time Series Data
of Size 100 with Autoregressive Errors. Based on 100 repli-
cations. Standard errors are in parentheses.

	ρ	intercept	trend	season
Monotone ESV	0.1	1.10 (0.036)	1.08 (0.031)	1.04 (0.019)
WEAVE		0.97 (0.009)	0.95 (0.016)	0.94 (0.013)
Monotone ESV	0.5	1.19 (0.051)	1.15 (0.042)	1.00 (0.026)
WEAVE		0.87 (0.015)	0.85 (0.018)	0.84 (0.018)
Monotone ESV	0.9	0.97 (0.054)	0.96 (0.048)	0.84 (0.033)
WEAVE		0.54 (0.021)	0.53 (0.022)	0.65 (0.026)

REML-Matérn estimates. The entry labeled "WNLS (exp.)" denotes using weighted least squares
to fit an exponential variogram (without nugget) to the empirical variogram of the residuals. This
entry is similar to the REML-Matérn estimates and is perhaps better than we might have expected
based on Figure 6. Our estimates have one large advantage not illustrated by these results: our
approach is invariant to the presence or absence of a nugget effect. In contrast, with a nugget effect
the Matérn or exponential models would need to be fit with an added nugget parameter and results
then would be more variable.

Table 5 contains results similar to Table 4 but for a 16 by 16 lattice. Comparing Table 4 and
Table 5 shows that performance is improving with sample size.

Table 6 displays results for Gaussian data generated from the Matérn model with smoothing
parameter 1, often called the Whittle model (Whittle, 1954). Whittle suggests that this model is
natural for agricultural field trials. Results here are similar to those in Tables 4 and 5.

4.2   Robust Regression Estimates of $\boldsymbol{\beta}$

The OLS estimators minimize the sum of residual squares $\sum_{i=1}^{n} e_i(\boldsymbol{\beta})^2$, and efficiency losses may
arise since this sum of squares is sensitive to large values that occur more frequently with non-
Gaussian data. Robust M-estimators minimize an objective function $\sum_{i=1}^{n} \rho(e_i(\boldsymbol{\beta}))$ that is less
sensitive to large values. Equivalently by taking derivatives, one solves $\sum_{i=1}^{n} \psi(e_i(\boldsymbol{\beta}))\boldsymbol{x}_i = 0$,

Table 4. Average of Standard Deviation Estimates (divided by true
standard deviation) for Regression from 10×10 Spatial Data of Size
100 with Exponential Variogram Errors. Based on 100 replications.
Standard errors are in parentheses.

	range	sill	intercept	x coord.	y coord.
Monotone ESV	1	3	0.89 (0.033)	0.90 (0.034)	0.90 (0.034)
REML(Matérn)			0.86 (0.031)	0.86 (0.025)	0.86 (0.025)
WNLS (exp.)			0.87 (0.025)	0.87 (0.024)	0.87 (0.024)
Monotone ESV	2	3	0.73 (0.029)	0.76 (0.030)	0.76 (0.030)
REML(Matérn)			0.71 (0.025)	0.75 (0.024)	0.75 (0.024)
WNLS (exp.)			0.73 (0.029)	0.74 (0.026)	0.74 (0.026)

Table 5. Average of Standard Deviation Estimates (divided by true
standard deviation) for Regression from 16×16 Spatial Data of Size
100 with Exponential Variogram Errors. Based on 100 replications.
Standard errors are in parentheses.

	range	sill	intercept	x coord.	y coord.
Monotone ESV	1	3	1.03 (0.032)	1.03 (0.039)	1.03 (0.039)
REML(Matérn)			0.94 (0.020)	0.94 (0.020)	0.94 (0.020)
WNLS (exp.)			0.93 (0.017)	0.94 (0.017)	0.94 (0.017)
Monotone ESV	2	3	0.86 (0.026)	0.87 (0.024)	0.87 (0.024)
REML(Matérn)			0.84 (0.026)	0.84 (0.024)	0.84 (0.024)
WNLS (exp.)			0.82 (0.024)	0.83 (0.022)	0.83 (0.022)

where $\psi = \rho'$. In order for the estimators to be scale invariant, one needs to alter the above equation to

$$\sum_{i=1}^{n} \psi \left(\frac{e_i(\boldsymbol{\beta})}{\widehat{\sigma}} \right) \boldsymbol{x}_i = 0,$$

where $\widehat{\sigma}$ is a scale estimate. There are different options for selecting ρ and $\widehat{\sigma}$, but a common approach is to use

$$\rho(t) = \begin{cases} \frac{t^2}{2} & \text{if } |t| \le k, \\ k|t| - \frac{k^2}{2} & \text{otherwise.} \end{cases}$$

Table 6. Average of Standard Deviation Estimates (divided by true standard deviation) for Regression from 16×16 Spatial Data of Size 100 with Whittle Variogram Errors. Based on 100 replications. Standard errors are in parentheses.

	range	sill	intercept	x coord.	y coord.
Monotone ESV	1	3	1.10 (0.035)	1.08 (0.025)	1.08 (0.025)
REML(Matérn)			0.98 (0.015)	0.96 (0.025)	0.96 (0.025)
WNLS (Whittle)			0.98 (0.013)	0.96 (0.012)	0.96 (0.012)
Monotone ESV	2	3	1.00 (0.029)	0.99 (0.029)	0.99 (0.029)
REML(Matérn)			0.93 (0.019)	0.93 (0.015)	0.93 (0.015)
WNLS (Whittle)			1.01 (0.029)	0.99 (0.027)	0.99 (0.027)

$$\psi(t) = \begin{cases} -k & \text{if } t < k, \\ t & \text{if } -k \le t \le k, \\ k & \text{if } t > k. \end{cases}$$

where k is a tuning constant, usually selected to give an appropriate asymptotic efficiency when the data are Gaussian. "Huber's Proposal 2" finds the scale estimate $\widehat{\sigma}$ by solving simultaneously

$$\sum_{i=1}^{n} \psi^2 \left(\frac{e_i(\boldsymbol{\beta})}{\widehat{\sigma}} \right) = C_k,$$

where C_k is a constant chosen so that $\widehat{\sigma}$ is consistent for the standard deviation when the data are from a Gaussian distribution. For example, when $k = 1$, $C_1 = .516$, and the asymptotic relative efficiency of $\widehat{\boldsymbol{\beta}}$ to least squares for independent Gaussian data is close to 90%.

Asymptotic normality of $\widehat{\boldsymbol{\beta}}$ was proved by Koul (1977, p. 688). One version of the asymptotic variance is given by

$$D(X^T X)^{-1} X^T V_\psi X (X^T X)^{-1},$$

where $D = \sigma / \mathrm{E}\psi'(e_1/\sigma)$ and V_ψ is the covariance matrix of $\psi(e_1/\sigma), \ldots, \psi(e_n/\sigma)$. We should also mention that we have ignored the role of estimating σ in the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ because for errors with symmetric marginal distributions, the asymptotic correlation between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ is zero.

These M-estimators fit into the general scheme outlined in Section 2. That is, we assume that $\psi(e_1/\sigma), \ldots, \psi(e_n/\sigma)$ are second-order stationary and estimate their covariance matrix using our

Table 7. Average of Standard Deviation Estimates (divided by true standard deviation) for Regression from 10×10 Spatial Data of Size 100 with t-transformed Exponential Variogram Errors. Based on 100 replications. Standard errors are in parentheses. True standard deviations are based on 10,000 replications and have estimated coefficient of variation .007.

	range	sill	intercept	x coord.	y coord.
True Standard Dev. (LS)	1	3	0.84	0.118	0.118
True Standard Dev. (Robust)			0.75	0.106	0.106
Monotone ESV (LS)	1	3	0.91 (0.036)	0.91 (0.034)	0.91 (0.034)
Monotone ESV (Robust)			0.89 (0.035)	0.90 (0.032)	0.90 (0.040)
True Standard Dev. (LS)	2	3	1.25	0.168	0.168
True Standard Dev. (Robust)			1.14	0.154	0.154
Monotone ESV (LS)	2	3	0.73 (0.031)	0.75 (0.031)	0.75 (0.031)
Monotone ESV (Robust)			0.73 (0.032)	0.76 (0.032)	0.77 (0.036)

nonparametric semivariogram estimator based on treating $\psi(\widehat{e}_1/\widehat{\sigma}), \ldots, \psi(\widehat{e}_n/\widehat{\sigma})$ as residuals ¿from least squares regression. The estimated factors for bias reduction are not quite correct since they officially only apply to least squares residuals. However, one way to motivate their use is by defining pseudo-observations $\widetilde{Y}_i = \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + d\psi(\widehat{e}_i/\widehat{\sigma})$ for any constant d. Then the least squares estimate based on $\widetilde{Y}_i$ is just $\widehat{\boldsymbol{\beta}}$, and the residuals are $d\psi(\widehat{e}_i/\widehat{\sigma})$.

Table 7 shows results for situations similar to Table 4 but with errors from t-transformed exponential variogram models. That is, we transformed the exponential errors used in Table 4 using the transformation $F_5^{-1}(\Phi(Y_i/\sqrt{3}))$, where Φ is the standard normal distribution function and F_5^{-1} is the inverse distribution function of a t distribution with 5 degrees of freedom. Thus the marginal distribution of Y_i is a t distribution with 5 degrees of freedom.

The entries in rows 1 and 2 and 5 and 6 of Table 7 are Monte Carlo estimates of the variance of $\widehat{\boldsymbol{\beta}}$ based on 10,000 replications. These show that the robust regression estimates have less variability than the least squares estimates as expected. The other rows are our standard deviation estimates for the regression parameter estimates divided by the estimated true standard deviations. The results are fairly similar to Table 4 and show that the estimates are negatively biased but less biased for the data with less correlation.

Splus version 3.4 along with the spatial module was used for all calculations. The robust estimates were calculated with the rlm function of Venables and Ripley (1997, p. 260) with $k = 1$ and $sw = 3$. The PAVA fortran routine in Cran (1980) was called from Splus.

5   Discussion

Our main goal has been to estimate the variance of regression coefficients obtained from a general estimating equations approach without making parametric semivariogram assumptions. Thus we have proposed a new method for nonparametric semivariogram estimation based on monotonizing a reduced-bias empirical semivariogram based on residuals.

The simulations of Section 4 indicate that the resulting variance estimates are reasonably unbiased and are converging as the sample size grows. Comparison with REML using the true Matérn model and WNLS with the correct model shows that the new method is competitive with parametric methods. It might be useful to make comparisons when the parametric methods are misspecified.

REFERENCES

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, Vol. 59, No.3 817-858.

Barlow, R., Bartholemew, D., Bremner, J., and Brunk, H. (1972) *Statistical Inference Under Order Restrictions*, John Wiley, New York.

Barry J.P. and Ver Hoef, J. M. (1996), " Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 297-322.

Bloomfield, P., and Nychka, D. (1992), "Climate Spectra and Detecting Climate Change," *Climatic*

Change, 21, 275-287.

Brownie, C. and Gumpertz, M. (1997), "Validity of spatial analyses for large field trials," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 1-23.

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence," *The Annals of Statistics*, 14, 1171-1179.

Cran, G. W. (1980), "[Algorithm AS 149] Amalgamation of Means in the Case of Simple Ordering," *Applied Statistics*, 29, 209-211.

Cherry, S., Banfield, J., and Quimby, W.F. (1996), "An Evaluation of a Nonparametric Method of Estimating Semi-variograms of Isotropic Spatial Processes," *Journal of Applied Statistics*, Vol. 23, No. 4, 435-449.

Cherry, S. (1997), "Non-parametric Estimation of the Sill in Geostatistics," *Environmetrics*, 8, 13-27.

Cressie, N. (1987), "A Nonparametric View of Generalized Covariances for Kriging," *Mathematical Geology*, Vol. 19, 425-449.

Cressie, N. (1993), *Statistics for Spatial Data* John Wiley, New York.

Garcia-Soidan, P. H. and Hall, P. (1997), "On Sample Reuse Methods for Spatial Data," *Journal of the American Statistical Association, Biometrics*, 53, 273-281.

Genton, M. G. (1998) "Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure," *Mathematical Geology*, Vol. 30, No. 4, 323-345.

Gorsich, D. J., and Genton, M. G.(2000) "Variogram Model Selection via Nonparametric Derivative Estimation," *Mathematical Geology*, Vol. 32, No. 3, 249-270.

Hall, P., Fisher, N. I., and Hoffman, B. (1994), "On the Nonparametric Estimation of Covariance Functions," *The Annals of Statistics*, 22, 2115-2134.

Heagerty, P.J. and Lumley, T. (2000), "Window Subsampling of Estimating Functions With Application to Regression Models," *Journal of the American Statistical Association*, 95, 197-211.

Huber, P. (1980), *Robust Statistics* John Wiley, New York.

Jones, P. D., Parker, D. E., Osborn, T. J., and Briffa, K. R. (2000), "Global and Hemispheric

Temperature Anomalies—Land and Marine Instrumental Records," in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, TN.
`http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html`

Koul, H. L. (1977), "Behavior of Robust Estimators in the Regression Model with Dependent Errors," *Annals of Statistics*, 5, 681-699.

Lele, S. (1991), "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes," *Journal of the Royal Statistical Society*, Ser. B, 53, 253-267.

Lefohn, A. S., and Runeckles, V. C. (1987), "Establishing a Standard to Protect Vegetation—Ozone Exposure/dose Considerations," *Atmospheric Environment*, 21, 561-568.

Lele, S. (1995), "Inner Product Matrices, Kriging, and Nonparametric Estimation of the Variogram," *Mathematical Geology*, 27, 673-692.

Liang, K.-Y., and Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.

Lumley, T. and Heagerty, P.J. (1999), "Weighted Empirical Adaptive Variance Estimators for Correlated data regression," *Journal of the Royal Statistical Society, Ser. B*, 61, 459-477.

Matérn, B. (1960), Spatial Variation. *Meddelanden fran Stantens Skogsforskniningsinstitut*, 49, No. 5.

Matheron, G. (1971), The Theory of Regionalized Variables and Its Applications, *Cahiers du Centre de Morphologie Mathematique*, No. 5, Fontainebleau, France.

Matheron, G. (1973), "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, 5, 439-468.

Newey, W.K., and West, K.D. (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Rousseeuw, P. J., and Molenberghs, G. (1993), "Transformation of Non Positive Semidefinite Correlation Matrices," *Communications in Statistics—Theory and Methods*, 22, 965-984.

Shapiro, A. and Botha, J.D. (1991), "Variogram Fitting with a General Class of Conditionally

Nonnegative Definite Functions," *Computational Statistics and Data Analysis*, 11, 87-96.

Sherman, M. (1996), "Variance Estimation for Statistics Computed from Spatial Lattice Data," *Journal of the Royal Statistical Society*, Ser. B, 58, 509-523.

Sherman, M. (1997), "Subseries Methods in Regression," *Journal of the American Statistical Association*, 92, 1041-1048.

Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer.

Sun, H., and Pantula, S. G. (1999), "Testing for Trends in Correlated Data," *Statistics and Probability Letters*, 41, 87-95.

Venables, W. N., and Ripley, B. D. (1997), *Modern Applied Statistics with S-Plus*, New York: Springer.

White, H., and Domowitz, I. (1984), "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143-161.

Whittle, P. (1954), "On stationary processes in the plane," *Biometrika*, 41, 434-449.

Wu, W. B., Woodroofe, M., and Mentz, G. (2001), "Isotonic Regression: Another Look at the Change Point Problem," preprint, Department of Statistics, University of Michigan.

Yasui and Lele, S. (1997), "A Regression Method for Spatial Disease Rates: an Estimating Function Approach," *Journal of the American Statistical Association*, 92, 21-32.