An Automatic Procedure for Fitting Variograms by Cressie's Approximate Weighted Least Squares Criterion

Robert M. Brunell
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275-0332

October 27, 1992

Acknowledgements

This research was supported by contract DE FG05 90ER61015 from the U.S. Department of Energy.

An Automatic Procedure for Fitting Variograms by Cressie's Approximate Weighted Least Squares Criterion

Abstract

Cressie (1985) proposed weighted least squares as a reasonable compromise between the efficiency of generalized least squares and the simplicity of ordinary least squares for fitting a variogram model to a sample variogram. He introduced an approximation to the weighted least squares criterion which reduced the estimation problem to iteratively reweighted least squares. Golway (1991) provided some details as to how the minimization could be achieved, but still left a large number of confusing choices to be made by the user. In this paper we present an algorithm for minimizing Cressie's approximate weighted least squares criterion with little intervention on the part of the user.

1. Introduction. A standard problem in spatial statistics is given a second-order stationary random process $\{Z(\underline{s}) : \underline{s} \in D\}$ where D is a subset of Euclidean space, estimate its variogram

$$2\gamma(\underline{h};\theta) = \mathrm{Var}(Z(\underline{s}_1) - Z(\underline{s}_2))$$

where $\underline{h} = \underline{s}_1 - \underline{s}_2$. Typically, the first step is to estimate the variogram by either the method-of-moments estimator

$$2\hat{\gamma}(\underline{h};\theta) = \frac{1}{|N(\underline{h})|} \sum_{N(\underline{h})} (Z(\underline{s}_i) - Z(\underline{s}_j))^2, \quad \underline{h} \in \Re^d$$

where $N(\underline{h}) = \{(\underline{s}_i, \underline{s}_j) : \underline{s}_i - \underline{s}_j = \underline{h}; i,j = 1,\ldots,n\}$ and $|N(\underline{h})|$ is the number of distinct pairs in $N(\underline{h})$, or by Cressie's robust estimator

$$2\bar{\gamma}(\underline{h};\theta) = \left\{ \frac{1}{|N(\underline{h})|} \sum_{N(\underline{h})} |Z(\underline{s}_i) - Z(\underline{s}_j)|^{1/2} \right\}^4 (0.457 + 0.494/|N(\underline{h})|)^{-1}, \quad \underline{h} \in \Re^d.$$

Unfortunately, the estimates are rarely conditional non-negative definite. This shortcoming is often overcome by fitting one of the isotropic variogram models below:

1. White noise (pure nugget effect):

$$\gamma(\underline{h};\theta_0) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0, & \underline{h} \neq \underline{0}, \end{cases}$$

   for $\theta_0 \geq 0$.

2. Linear model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1 \|\underline{h}\|, & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$ and $\theta_1 \geq 0$.

3. De Wijs (logarithmic) model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1 \log \|\underline{h}\|, & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$ and $\theta_1 \geq 0$.

4. Power model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1 \|\underline{h}\|^{\theta_2}, & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $0 \leq \theta_2 < 2$.

5. Exponential model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1[1 - \exp(-\|\underline{h}\|/\theta_2)], & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$, and $\theta_2 \geq 0$.

6. Gaussian model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1[1 - \exp(-\{\|\underline{h}\|/\theta_2\}^2)], & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$, and $\theta_2 \geq 0$.

7. Rational Quadratic model

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1[\|\underline{h}\|/\theta_2]^2/(1 + [\|\underline{h}\|/\theta_2]^2), & \underline{h} \neq \underline{0}, \end{cases}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$, and $\theta_2 \geq 0$.

8. Spherical model:

$$\gamma(\underline{h};\underline{\theta}) = \begin{cases} 0, & \underline{h} = \underline{0}, \\ \theta_0 + \theta_1(\frac{3}{2}\|\underline{h}\|/\theta_2 - \frac{1}{2}[\|\underline{h}\|/\theta_2]^3), & 0 < \|\underline{h}\| \leq \theta_2, \\ \theta_0 + \theta_1, & \|\underline{h}\| > \theta_2, \end{cases}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$, and $\theta_2 \geq 0$.

Until recently, the most common methods of fitting variogram models to sample variogram models were by eye or by ordinary least squares. While ordinary least squares is an improvement over visual fitting, it ignores the covariances between variogram estimates. Fitting by generalized least squares requires calculating the variances of the estimates of the sample variogram at each lag and covariances between them, which is very complicated (Cressie 1985). Weighted least squares is less efficient, but only requires calculating the

variances. The loss of efficiency is small (Zimmerman and Zimmerman 1991), so Cressie argued that weighted least squares is an acceptable compromise between efficiency and simplicity. Cressie simplified the fitting process even further by proving that

$$C(\underline{\theta}) = \sum_{j=1}^{K} |N(\underline{h}_j)| \left(\frac{g(\underline{h}_j)}{\gamma(\underline{h}_j; \underline{\theta})} - 1 \right)^2$$

is a good approximation to the weighted least squares criterion, where $\underline{h}_1, \ldots, \underline{h}_K$ are equally spaced lags at which the variogram is estimated and $g(\cdot)$ is either the method-of-moments or robust estimator.

Minimizing $C(\underline{\theta})$ is not entirely straightforward. Examining the first-order conditions

$$\nabla C(\underline{\theta}) = 2 \sum_{j=1}^{K} \frac{|N(\underline{h}_j)| g(\underline{h}_j)}{\gamma^3(\underline{h}_j; \underline{\theta})} \left(\gamma(\underline{h}_j; \underline{\theta}) - g(\underline{h}_j) \right) \frac{\partial \gamma(\underline{h}_j; \underline{\theta})}{\partial \underline{\theta}}$$

reveals that the problem can be recast as iteratively reweighted least squares (Cressie 1991) with weights

$$w_j = \frac{|N(\underline{h}_j)| g(\underline{h}_j)}{\gamma^3(\underline{h}_j; \underline{\theta})}.$$

However, models (4)-(8) above are not linear in $\underline{\theta}$. Golway (1991) explored minimizing $C(\underline{\theta})$ by using PROC NLIN in SAS without resolving the choice of minimization method (e.g. steepest descent, Newton, modified Gauss-Newton, Marquardt, secant), step size, starting values or grid size. In this paper we propose an automatic procedure for variogram fitting that relieves the user of these many confusing choices.

2. Exact Solutions. In a few cases, it is possible to calculate exact solutions to the minimization problem. For the white noise model (1)

$$\hat{\underline{\theta}} = \frac{\sum_{j=1}^{K} |N(\underline{h}_j)| g^2(\underline{h}_j)}{\sum_{j=1}^{K} |N(\underline{h}_j)| g(\underline{h}_j)}.$$

When the nugget θ_0 is constrained to be zero, one may calculate an exact solution for the linear model (2)

$$\hat{\underline{\theta}} = \frac{\sum_{j=1}^{K} |N(\underline{h}_j)| [g(\underline{h}_j)/\underline{h}_j]^2}{\sum_{j=1}^{K} |N(\underline{h}_j)| [g(\underline{h}_j)/\underline{h}_j]}.$$

as well as for the De Wijs model (3)

$$\hat{\underline{\theta}} = \frac{\sum_{j=1}^{K} |N(\underline{h}_j)| [g(\underline{h}_j)/\log \underline{h}_j]^2}{\sum_{j=1}^{K} |N(\underline{h}_j)| [g(\underline{h}_j)/\log \underline{h}_j]}.$$

3. Iteratively Reweighted Linear Least Squares. Each of the first three models (white noise, linear, and De Wijs) are linear in their parameters. Therefore, they may be fit by iteratively reweighted linear least squares. This avoids calculating the Hessian of the criterion function. We recommend using the ordinary least squares estimates as the starting values. Experience has shown that the white noise model always converges, and the linear and De Wijs models fit converge whenever the functional form is a reasonable candidate for the sample variogram.

4. Minimizing the Profile Criterion Function. The remaining models (exponential, Gaussian, rational quadratic, and spherical) are linear in the first two parameters, but not in the third. This suggests that we may proceed by minimizing the profile criterion function

$$P(\theta_2) = \min_{\theta_0,\theta_1} C(\theta_0, \theta_1, \theta_2)$$

to obtain the optimal value of θ_2 and then use iteratively reweighted linear least squares to obtain the best θ_0 and θ_1. When θ_2 is positive, $P(\theta_2)$ may be evaluated by iteratively reweighted linear least squares starting at the ordinary least squares estimate. When $\theta_2 = 0$, the profile criterion function is set to the minimized value of $C(\theta_0)$ for the white noise model (1). In evaluating $P(\theta_2)$ we again avoid calculating the Hessian, and experience has shown that the iteration converges if the functional form of the model is a reasonable candidate for the sample variogram and θ_2 is not too distant from the optimal θ_2. When the iteration fails to converge we set $P(\theta_2)$ equal to a huge number.

Examining the functional form of models (4)-(8) suggests that the profile criterion function $P(\theta_2)$ could be multimodal or wildly fluctuating. Furthermore, the calculation of $P'(\theta_2)$ is far from straightforward. Therefore, we will find its minimum by a golden section search (Kennedy and Gentle 1980, pp. 432-3). Given an initial bracket, the golden section search reduces the size of the bracket by a factor of $(3 - \sqrt{5})/2 \approx 0.382$ on each the step. The profile criterion function is evaluated at about 38.2% of the way from the left endpoint to the right endpoint and also at about 38.2% of the way from the right endpoint to the left endpoint. The interval is then revised to reflect where the minimum must lie in light of the profile criterion function values.

It remains, then, to choose the initial bracket. For the power model (4), the permissible values of θ_2 form a bounded set $[0, 2)$. Therefore, we perform a grid search over the set with a grid size equal to one-fifth. The initial bracket is the value of θ_2 that provides the smallest value of $P(\theta_2)$ plus and minus one-fifth, subject to the constraint that the initial bracket must be a subset of $[0, 2]$.

For the last four models (exponential, Gaussian, rational quadratic, and spherical), the third parameter may be interpreted as the range of the variogram, that is, the value

of the argument at which the variogram levels off. Hence the profile criterion function is evaluated at each of the lags at which the variogram is estimated. If the largest lag produces the smallest value of $P(\theta_2)$, more equally spaced lags are generated and evaluated until the profile criterion function begins to rise. The initial bracket is then the value of θ_2 that corresponds to the smallest value of $P(\theta_2)$ plus and minus the (common) distance between adjacent lags, subject to the constraint that the left endpoint may not be negative.

5. Discussion and Summary. Some caution is recommended before employing the algorithms described above. Each of the last four models is convex. When fitting one of these models to a sample variogram that is concave, the optimal value for the range is clearly infinite, so the algorithm to search for the initial bracket will usually be caught in an infinite loop. Also, when the optimum value of θ_2 is larger than the largest experimental lag, the algorithm implicitly assumes that the first minimum after the largest experimental lag is the global minimum, although this may not be true. Finally, all of the linear least squares calculations are unconstrained, and so they may produce invalid values for the first two parameters. If θ_0 is negative, the user should fit a model with no nugget (i.e. set θ_0 to zero). If θ_1 is negative, the sample variogram is decreasing and the user should fit the white noise model (1).

We have described algorithms for fitting experimental variograms to variogram models by Cressie's approximate weighted least squares criterion. The only input required from the user is the convergence criterion for the iteratively reweighted least squares and the number of iterations permitted. There is no theory as to why the procedures described in sections 3 and 4 should produce a global minimum, however, the procedures work well nevertheless. A program that implements the algorithms detailed above is available from the author.

References

Cressie, N.A.C. (1985), "Fitting Variogram Models by Weighted Least Squares," *Mathematical Geology*, **17**, 563-586.

Cressie, N.A.C (1991), *Statistics for Spatial Data*, Wiley: New York.

Golway, C.A. (1991), "Fitting Semivariogram Models by Weighted Least Squares," *Computers and Geosciences*, **17**, 171-172.

Green, P.J. (1984) "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society*, **B46**, 149-192.

Kennedy, W.J. and Gentle, J.E. (1980), *Statistical Computing*, Marcel Dekker : New York.

Zimmerman, D.L., and Zimmerman, M.B. (1991) "A Monte Carlo Comparison of Spatial Semivariogram Estimators, and Corresponding Kriging Predictors," *Technometrics*, **33**, 77-91.