On Least Squares Variogram Fitting

Müller, Werner

DOI:
10.57938/68ff0b2b-09ea-4121-8cc0-a7abbb5e56d1

Published: 01/01/1997

Document Version:
Publisher's PDF, also known as Version of record

Document License:
Unspecified

On Least Squares Variogram Fitting

Werner Müller

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 50
May 1997

http://statmath.wu-wien.ac.at/

On Least Squares Variogram Fitting

Werner G. Müller

No. 50, May 1997

SHORT NOTE

ON LEAST SQUARES VARIOGRAM FITTING

WERNER G. MÜLLER

Department of Statistics, University of Economics, Augasse 2–6, A-1090 Vienna, Austria.

May 6, 1997

Introduction

In geostatistical data analysis second order spatial dependence is usually characterized by a so-called variogram

$$2\gamma(h) = E\{[Z(\mathbf{s}+h) - Z(\mathbf{s})]^2\} \qquad \forall \mathbf{s}, \mathbf{s}+h \in \mathcal{S} \subset \mathbb{R}^2.$$

For the sake of simplicity let us assume that $Z(.)$ is an intrinsically stationary (see Cressie (1991)), isotropic (two-dimensional) random field which is sampled at n locations, yielding data $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n)$. Let us additionally assume that the variogram follows a known parametric form, so that we have the properties

$$E\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]\} = 0,$$

and

$$E\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2\} = 2\gamma(h_{ij}, \theta),$$

where $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ denotes the distance (in a suitable metric) between the sites $\mathbf{s}_i$ and $\mathbf{s}_j$.

It is natural to estimate the variogram by

$$2\hat{\gamma}_k = \frac{1}{N_{H_k}} \sum_{s_i, s_j \in H_k} [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2. \tag{1}$$

For the classical variogram estimator by Matheron (1963) the set (bin) H_k contains all point pairs with distances such that $h_{ij} = h_k$ (h_k being the ordered sequence of all K distinct distances). If the data stems from an irregular grid this set is usually enlarged to contain a certain neighbourhood of k, but always such that $\sum_{k=1}^{K} N_{H_k} = \binom{n}{2}$.

1 Parametric fitting

A parametric fit $\gamma(h, \hat{\theta})$ of the (semi-) variogram is usually found by least squares regression through the points $(h_k, \hat{\gamma}_k)$.

One particular case of (1) is to choose each H_k to contain only one element h_{ij}, i.e. $N_{S_k} \equiv 1$, $k = 1, \ldots, \binom{n}{2} \equiv K$. The collection of pairs $(h_{ij}, \hat{\gamma}_k)$ is then called the (semi-)variogram cloud and we attempt to estimate $\gamma(h, \hat{\theta})$ directly from the (semi-)variogram cloud without prior binning and averaging. Note that although this is a deviation from standard geostatistical practice, it clearly utilizes the maximum available information and is thus, if feasible, preferable.

It is obvious that the 'observations' $\hat{\gamma}_k$ are generally correlated and thus ordinary least squares yields inefficient estimates. The situation requires generalized least squares estimation, i.e.

$$\hat{\theta}_{GLS} = \operatorname{Arg}\min_{\theta}[\hat{\gamma} - \gamma(\theta)]^T \operatorname{Cov}^{-1}(\hat{\gamma})[\hat{\gamma} - \gamma(\theta)], \tag{2}$$

where $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_k)^T$, and $\gamma(\theta) = (\gamma(h_1, \theta), \ldots, \gamma(h_k, \theta))^T$.

If we now assume that $Z(.)$ is a Gaussian random field, Cressie (1985) showed that

$$E(\hat{\gamma}_k) = \gamma(h_k),$$

and

$$\operatorname{Cov}(\hat{\gamma})_{kl} = \frac{1}{2}[\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|) + \gamma(\|\mathbf{s}_{i'} - \mathbf{s}_{j'}\|) - \gamma(\|\mathbf{s}_i - \mathbf{s}_{j'}\|) - \gamma(\|\mathbf{s}_{i'} - \mathbf{s}_j\|)]^2, \tag{3}$$

where $\|\mathbf{s}_{i'} - \mathbf{s}_{j'}\| = h_l$. Thus we have

$$\operatorname{Var}(\hat{\gamma}_k) = 2\gamma^2(h_k).$$

For computational simplicity Cressie (1985) suggested to use the following (weighted least squares) estimator based on (2), with $\operatorname{Cov}(\hat{\gamma}) \simeq \operatorname{diag}(2\gamma^2(\theta))$:

$$\tilde{\theta} = \operatorname{Arg}\min_{\theta} \sum_{k=1}^{K} \gamma^{-2}(h_k, \theta)[\hat{\gamma}_k - \gamma(h_k, \theta)]^2. \tag{4}$$

It has to be noted that for variograms with finite range the assumption of zero off-diagonal elements of $\operatorname{Cov}(\hat{\gamma})$ can be fulfilled by a proper choice of the sampling sites (for a respective algorithm see Müller and Zimmerman (1995)).

Estimator $\tilde{\theta}$ is one of the most widely used in variogram estimation (see eg. Gotway (1990), Cressie (1991), Lamorey and Jacobson (1995)). However, several authors have noticed poor finite sample behaviour (see McBratney and Webster (1986) and Zhang *et al.* (1995)). A simulation study by Zhang *et al.* (1995) and one in Section 3 backs up this claim.

It is natural to employ asymptotic theory to investigate possible causes of this deficiency. It turns out that if $\tilde{\theta}$ is found by direct minimization of (4) from the (semi-)variogram cloud we have a generally inconsistent estimator (see similar types of estimators in Fedorov (1974)). This is not the case under the assumptions A1 (K fixed) and A2 ($N_{H_k} \to \infty$) of Cressie (1985), but here we have $n \to \infty$ and thus $K \to \infty$ and $N_{H_k} = 1 \quad \forall k$. Moreover, the small sample simulation results in Section 3 indicate that the studied asymptotics may have some explanatory value also for the binned estimator, especially when the number of bins is large. For a particular case of (4) its inconsistency becomes very transparent. Under the assumption of negligible off-diagonal elements of $\text{Cov}(\hat{\gamma})$ (i.e. under certain sampling schemes) it is easy to adopt the proof of inconsistency from Fedorov (1974) to find a simple correction of (4): We can separate the 'observations' $\hat{\gamma}_k = \gamma(h_k, \theta_t) + \epsilon_k$, where θ_t denotes the true parameter. The sum to be minimized from (4) can in the limit also be separated into

$$\sum_{k=1}^{\infty} \gamma^{-2}(h_k, \theta)[\gamma(h_k, \theta) - \gamma(h_k, \theta_t)]^2 + \sum_{k=1}^{\infty} \gamma^{-2}(h_k, \theta)\epsilon_k^2.$$

Due to the law of large numbers this gives

$$\sum_{k=1}^{\infty} \left[\frac{\gamma(h_k, \theta_t)}{\gamma(h_k, \theta)} - 1\right]^2 + \sum_{k=1}^{\infty} \frac{2\gamma^2(h_k, \theta_t)}{\gamma^2(h_k, \gamma)},$$

and after multiplying out and regrouping we find that it is equivalent to minimize

$$\sum_{k=1}^{\infty} \left[\frac{\gamma(h_k, \theta_t)}{\gamma(h_k, \theta)} - \frac{1}{3}\right]^2,$$

which indicates that $\gamma(h, \tilde{\theta})$ consistently estimates $3\gamma(h, \theta_t)$ rather than $\gamma(h, \theta_t)$.

2 Iterative reweighting

For general distributions and a wider class of sampling schemes such a simple proof and a resulting correction is not available and thus an alternative consistent estimator has to be sought. It is a long known fact that for models of the discussed type iteratively reweighted estimators yield consistent estimates. This estimator has been suggested in this context already by Cressie (1985) and used by various authors without justifying the cause (e.g. McBratney and Webster (1986) and Zimmerman and Zimmerman (1991)). It was again Fedorov (1989), who correctly argued for and described an iterative algorithm for the case $\text{Cov}(\hat{\gamma}) = \text{diag}(2\gamma^2(\theta))$. For the more general case (2) this algorithm can be easily adopted. It is however computationally much more intensive and requires the inversion of a $K \times K$ matrix at each step:

$$\hat{\theta}_{r+1} = \text{Arg}\min_{\theta}[\hat{\gamma} - \gamma(\theta)]^T \, \text{Cov}^{-1}(\gamma(\hat{\theta}_r))[\hat{\gamma} - \gamma(\theta)],$$

$$\bar{\theta} = \lim_{r \to \infty} \hat{\theta}_r,$$

where the entries of $\text{Cov}(\gamma(\hat{\theta}_r))$ can be calculated from a parametric version of (3).

In practice the iterations have to be started at an inititial guess (say $\hat{\theta}_0 = \hat{\theta}_{OLS}$). The procedure then yields asymptotically efficient and consistent estimates. For an overview of such properties and further details see del Pino (1989).

3 Simulation results

To assess the small sample behaviour of the discussed estimators a great number of simulation experiments have been performed. Gaussian random fields with a prespecified variogram were generated on randomly selected sites in the unit square. As the performance criterion we chose the sum of the squared differences from the estimated to the true variogram over all bins. We have compared the OLS and the WLS estimate (both in its binned and its variogram cloud version), the corrected estimate $\frac{\gamma(h_k,\tilde{\theta})}{3}$ and the iterated GLS estimate $\gamma(h_k, \bar{\theta})$.

The perhaps most widely used parametric variogram model, the so-called spherical variogram

$$\gamma_S(h,\theta) = \begin{cases} 0, & h = 0 \\ \theta_1 + \theta_2 \cdot \left\{ \frac{3}{2}\left(\frac{h}{\theta_3}\right) - \frac{1}{2}\left(\frac{h}{\theta_3}\right)^3 \right\}, & 0 < h \leq \theta_3 \\ \theta_1 + \theta_2, & h > \theta_3 \end{cases}$$

was fitted. This variogram increases monotonically from a 'nugget effect' of θ_1 near the origin to a 'sill value' of $\theta_1 + \theta_2$, which is attained at the 'range' θ_3.

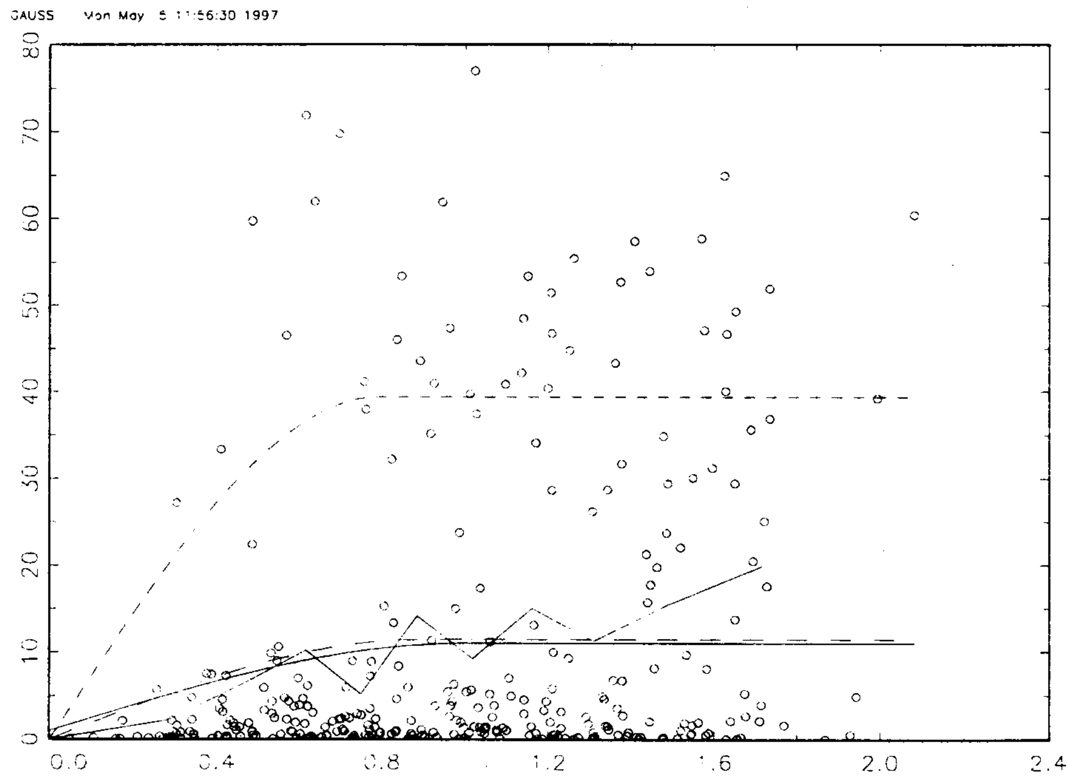

Figure 1: Variogram cloud and various variogram estimators for a typical simulated data set from 25 observations.

Figure 1 represent a typical cases of relative behaviour of the variogram estimates. The smooth solid line gives the true variogram, the unsmooth solid line gives the classical variogram estimate (with $K = 10$), the dotted line is the OLS estimate, the dashed line stands for $\gamma(h_k, \bar{\theta})$ and the closely dashed for $\gamma(h_k, \tilde{\theta})$, the latter three estimated from the variogram cloud.

The graph apparently indicates that $\gamma(h_k, \tilde{\theta})$ is considerably biased also in small samples, which can be ameliorated by multiplying the correction factor $\frac{1}{3}$. In fact, it turned out as an overall result that also the binned version of $\gamma(h_k, \tilde{\theta})$ is considerably biased, even for moderate bin sizes. The difference in performance of the other considered estimators is more pronounced the smaller the influence of the nugget effect is. Therefore, and for brevity we report in Table 3 only the results for $\theta_t = 1, 10, 1$, other settings can be recalculated using the GAUSS-386 code (see Aptech, Inc. (1993)), which can be downloaded from **http:**

eeyore.wu-wien.ac.at
stat4
vario.sim given in the appendix.

n	binned OLS ($K = 10$)	OLS	binned WLS ($K = 10$)	WLS	corrected WLS	iterated GLS
5	627.8	627.8	2403.5	2403.5	367.2	434.6
	256.0	256.0	443.6	443.6	327.0	238.9
10	351.4	349.6	611.8	3263.2	235.7	251.5
	147.7	146.3	183.7	1253.7	163.8	133.5
15	272.4	271.4	356.5	3495.1	186.2	191.6
	128.1	128.3	125.3	1384.6	139.4	109.3
20	231.5	231.3	266.0	3456.7	166.8	168.0
	106.8	107.9	100.6	1900.0	107.6	85.8
25	184.5	184.3	205.6	3539.8	145.2	148.1
	86.1	85.0	79.9	2115.0	89.9	69.7

Table 1: Mean (upper line) and Median (lower line) of the performance criterion of various variogram estimators over 1001 runs each ($\theta_t = 1, 10, 1$).

The distribution of any performance measure in this context is highly skewed. Therefore the mean and the median over 1001 runs is reported. It can be seen that as n grows the performance of the fitted γ get better except for the 'pure' WLS method, which gives very poor results for large sample sizes. However, it is evident that the WLS estimator maintains its deficiency, when the number of bins is relatively large compared to the sample size. Furthermore, it is remarkable that the corrected WLS estimator has excellent mean behaviour, although the considered designs, may well have departed from the ones assumed in calculating the correction (its median performance, however, is worse than the one from the OLS estimators).

4 Conclusions

Once more the poor finite sample behaviour of the (direct) weighted least squares variogram estimator was identified. It was attempted to explain it by applying a different kind of asymptotics as considered usually. The presented asymptotic results are therefore not in conflict with standard geostatistical theory.

These asymptotics and the simulation results, however, indicate that it is preferable to employ an iteratively reweighted generalized least squares variogram estimator based on the fully specified covariance matrix, which is not standard geostatistical practice. With the increased computing power there is less and less reason not to do so. All other least squares based methods are clearly less efficient or even, as shown, potentially defective.

If the number of observation sites is not very large it is also recommendable to renounce grouping (and therewith loss of information) and work directly with the semivariogram cloud.

Acknowledgements

I am most grateful to V.V. Fedorov for pointing out his 1974 paper and for many constructive remarks on this note. Thanks go also to D.L. Zimmerman and a referee for identifying several inaccuracies in an earlier version.

References

Aptech, Inc., . (1993). *GAUSS-386 Command Reference.* Washington.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17:563–586.

Cressie, N. (1991). *Statistics for Spatial Data.* John Wiley and Sons, New York.

del Pino, G. (1989). The unifying role of iterative generalized least squares in statistical algorithms. *Statistical Science*, 4:394–408.

Fedorov, V.V. (1974). Regression problems with controllable variables subject to error. *Biometrika*, 61:49–56.

Fedorov, V.V. (1989). Kriging and other estimators of spatial field characteristics. *Atmospheric Environment*, 23(1):175–184.

Gotway, C.A. (1990). Fitting semivariogram models by weighted least squares. *Computers & Geosciences*, 17:171–172.

Lamorey, G. and Jacobson, E. (1995). Estimation of semivariogram parameters and evaluation of the effects of data sparsity. *Mathematical Geology*, 27(3):327–358.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

McBratney, A.B. and Webster, R. (1986). Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science*, 37:617–639.

Müller, W.G. and Zimmerman, D.L. (1995). An algorithm for sampling optimization for semivariogram estimation. In Kitsos, C.P. and Müller, W.G., editors, *Model-Oriented Data Analysis 4*, pages 173–178, Heidelberg. Physica.

Zhang, X.F., van Eijkeren, J.C.H. and Heemink, A.W. (1995). On the weighted least-squares method for fitting a semivariogram model. *Computers & Geosciences*, 21(4):605–606.

Zimmerman, D.L. and Zimmerman, M.B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33(1):77–91.