STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, February 2025, pp 611–615.
Published online in International Academic Press (www.IAPress.org)

Geometric Weighted Least Squares Estimation

Reinhard Oldenburg *

Department of Mathematics, Augsburg University, Germany

Abstract Optimal efficiency of least squares (LS) estimation requires that the error variables (residuals) have equal variance (homoscedasticity). In LS applications with multiple output variables, heteroscedasticity can even cause bias. In weighted LS, weights are chosen to compensate for differences in variance. The selection of these weights can be challenging, depending on the specific application. This paper introduces a general method, Geometric Weighted Least Squares (GWLS) estimation, which estimates weights using the inequality between the geometric and arithmetic means. A simulation study explores the performance of the method.

Keywords Weighted least squares, Geometric mean, Latent variables

AMS 2010 subject classifications 62J02

DOI: 10.19139/soic-2310-5070-2324

1. Introduction

Least squares is a method that can be applied to a very broad class of statistical models. In models with multiple output variables (multi-response models), it is particularly important to have equal error variances, as otherwise not only is efficiency affected, but estimates may also be biased. Weighted least squares provides a general framework for compensating for heteroscedasticity (see, e.g., [5], ch. 11). Naturally, it works best when the correct error variances are known. If this is not the case, several techniques may be employed, with a general one being iteratively reweighted least squares. For SEM models, the inverse of the covariance matrix is a natural candidate as weight matrix and is used, for example, in GLS estimation of SEM models [1], p. 114. Other approaches to weight selection investigate higher moments to go beyond normality assumptions, e.g. Browne [2].

The purpose of this article is to introduce a non-iterative method for weight estimation that can be applied to a wide class of models. It is based on a very simple and elementary idea, yet it appears that it has not been applied and published before. With perfectly chosen weights, the product of the weight and the corresponding error variance should be equal across all variables. The well-known inequality between the geometric and arithmetic means implies that a sum is minimized under the constraint that the product is constant, if all the summands are equal. This is exactly the condition needed for weight selection. Hence, the method presented here will be called Geometric Weighted Least Squares (GWLS).

The paper introduces the method and applies it to a simple non-linear latent variable model that shows bias when estimated naively by unweighted least squares. This particular model is a special case of models that are relevant, for instance, in educational studies [4].

*Correspondence to: Reinhard Oldenburg (Email: reinhard.oldenburg@math.uni-augsburg.de). Department of Mathematics, Augsburg University. Universitätsstr. 14, 86163 Augsburg, Germany.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

2. GWLS

For models with a single outcome variable, unweighted least squares estimation of linear models remains unbiased, though it may be inefficient. However, for models with multiple outcome variables, even this property no longer holds. Weighted least squares address this issue but introduce the new challenge of selecting appropriate weights. A common method is to begin with unweighted least squares (ULS), then calculate estimates of the error variances from the solution, which can be used as weights in the next iteration.

In this paper, we focus on regression models with multiple outcome variables. For the sake of generality, we use the following notation for the model:

$$g(\theta, \mathbf{X}, \mathbf{Z}) = \mathbf{E} \tag{1}$$

Here $\mathbf{X}$ is a vector of k observable random variables, $\mathbf{Z}$ is a vector of p latent random variables, θ is a vector of model parameters, and $\mathbf{E} = (\epsilon_1, \ldots, \epsilon_m)$ is a vector of m error variables that are assumed to have mean 0. The model defining vector-valued function $g = (g_1, \ldots, g_m)$ may be non-linear.

When a sample of size n is drawn for this model, one has a $n \times k$ matrix of observations, and for given parameter values, one can calculate the matrix of errors, which I denote, by slight abuse of notation, as $\mathbf{E}$ as well.

For the special case without latent variables (i.e. $p = 0$) and under the assumption that the covariance matrix Σ of $\mathbf{E}$ is non-singular, it can be shown ([5]) that minimizing $\mathbf{E}'.\Sigma^{-1}.\mathbf{E}$ gives unbiased estimates of the parameters θ.

Let's now assume that errors are uncorrelated, i.e. that $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$ is diagonal, with $\sigma_j^2 = \mathbb{E}[\mathbf{E}_j^2] = \text{var}(\mathbf{E}_j)$. Then the weighted least square problem is to solve

$$\text{argmin}_{\theta,\mathbf{Z}} \sum_{j=1}^{m} w_j g_j(\theta, \mathbf{X}, \mathbf{Z})^2 \tag{2}$$

with weights that are chosen as $w_j := \sigma_j^{-2}$. Then, for the correct parameters θ, one has $\mathbb{E}[g_j] = 0, \mathbb{E}[g_j^2] = \sigma_j^2$. Thus, the minimum value of the objective function of (2) is m and all the $w_k g_k(\theta, \mathbf{X}, \mathbf{Z})^2 = 1$ are equal.

When the error variances are unknown, we propose to use the fact that the correct weights make the summands in (2) equal. The well-known theorem between the geometric and arithmetic means states that for non-negative numbers $x_1, \ldots, x_m \geq 0$ one has

$$\frac{1}{m} \sum_{j=1}^{m} x_j \geq \left(\prod_{j=1}^{m} x_j \right)^{\frac{1}{m}} \tag{3}$$

and equality holds iff all the numbers are equal $x_1 = \ldots = x_m$.

Setting $x_j = w_j g_j^2$ one gets the method of GWLS, i.e.

$$\text{argmin}_{w,\theta,\mathbf{Z}} \sum_{j=1}^{m} w_j g_j(\theta, \mathbf{X}, \mathbf{Z})^2, \text{ subject to } \prod_{j=1}^{m} w_j g_j^2 \geq 1 \tag{4}$$

The solution to this problem will make the constraint sharp, i.e., $\prod_{j=1}^{m} w_j g_j^2 = 1$ and is a self-consistent approximation of weights and error variances. Therefore, it inherits all theoretical properties of weighted LS with the correctly chosen weights, at least in the limit of large n and when minimization finds the correct local linear minimum. The correct solution is one of the minimizers, but, unfortunately, there is no proof of the converse.

For practical implementations, one should reformulate the constraint as $\sum_{j=1}^{m} \log(w_j) + \log(g_j^2) \geq 0$. Moreover, it is advisable to start the numerical methods to solve (4) from good initial values, e.g., from a solution of the unweighted problem. This was done in the simulation studies reported below.

Experiments (reported below) indicate that this estimation technique is not very efficient for smaller samples, i.e. the variation of estimation errors is relatively large. Inspection shows that estimates that are far off the expected values are often characterized by large correlations between errors or by correlations between errors and latent variables. When zero correlations $\text{cor}(\epsilon_j, \epsilon_{j'}) = 0, j \neq j'$ are estimated from a sample of size n, it is well known that the estimates of the correlation have mean 0 and standard deviation of $(n-1)^{-0.5}$. Therefore, the terms

$(n-1)\text{cor}(\epsilon_j,\epsilon_{j'})^2$ have variance of 1 just as $w_j g_j^2$ and hence these expressions can be added to the objective function without violating homoscedasticity. The same holds true for the correlations between errors and latent variables. This leads to an objective function F_c, which gives rise to a modified estimation method, GWLSc (for controlled GWLS):

$$F_c := \sum_{j=1}^{m} w_j g_j(\theta, \mathbf{X}, \mathbf{Z})^2 + F_e + F_l \tag{5}$$

$$F_e := \sum_{j<j'} (n-1)\text{cor}(\epsilon_j, \epsilon_{j'})^2 \tag{6}$$

$$F_l := \sum_{j=1}^{m} \sum_{s=1}^{p} (n-1)\text{cor}(\epsilon_j, Z_s)^2 \tag{7}$$

This objective function is minimized subject to $\prod_{j=1}^{m} w_j g_j^2 \geq 1$.

3. A toy model

This section investigates model with two latent variables X, Y that take values in the interval $[-1, 1]$. Each latent variable has two indicator variables. Between the latent variables an implicative relation is assumed in the following form: Negative X gives no information about Y, but for positive values there is linear prediction possible. Using the unit step function $\Theta(x) := (x + |x|)/2$ the full model is

$$\begin{aligned} x_1 &= X + u_1 + \epsilon_1 && (8)\\ x_2 &= c_2 X + u_2 + \epsilon_2 && (9)\\ y_1 &= Y + u_3 + \epsilon_3 && (10)\\ y_2 &= c_4 Y + u_4 + \epsilon_4 && (11)\\ 0 &= \Theta(X)(Y - aX - u_5 + \epsilon_5) && (12) \end{aligned}$$

Data for this model has been simulated with the choices $a = 0.9, c_2 = 0.7, c_4 = 0.4, u_j = 0$. Furthermore, normal errors with $\text{sd}(\epsilon_1) = 0.3, \text{sd}(\epsilon_2) = 0.4, \text{sd}(\epsilon_3) = 0.1, \text{sd}(\epsilon_4) = 0.2$ have been used. Additionally, ϵ_5 was uniformly distributed from $[-0.1, 0.1]$.

The simulated data have been estimated using uniform LS ULS (i.e. all weights fixed to 1), GWLS and GWLSc as explained above, and furthermore some iterative versions: ULS1 used the inverse variances of ϵ_j as estimated by ULS as weights. ULS2 iterates this by taking estimates from ULS1. Similarly, GLS1, GLS2, GLS3 are iterates, but these methods don't assume a diagonal covariance matrix but instead minimize $E'\Sigma^{-1}E$ where Σ is the estimate of the error covariance matrix estimated in the previous iteration step. In principle, one may hope that taking this iteration further will lead to convergence. However, for the model above this behavior was not observed, but some of these early iterates give rather good estimates, although it is unsatisfactory from a theoretical point of view, that there is no criterion to choose which iterate is best.

The method has been prototyped in Wolfram Mathematica, but for the sake of accessibility, the code has been converted to Python and is available at https://myweb.rz.uni-augsburg.de/~oldenbre/GW5.py. The following simulations have been run on a MacBook pro with M1 processor, using Python 3.12 with Numpy 2.1.1 and Casadi 3.6.7 (https://web.casadi.org/). For each of the sample sizes $n = 100, 200, 500$, I drew 250 simulations. The most interesting and demanding parameter is a and only results for this parameter are displayed in tables 1–3. The tables give the mean error, the standard deviation of estimation, and the root mean square error RMSE. The last column reports run-time in seconds (for ULSi and GLSi only the times of the last iteration step is given). Optimization fails in approx 2% of the cases for ULSi and GLSIi, and in about 5% of the cases for GWLSc. These cases have been eliminated from the results given in the tables. However, including them (Casadi gives a result, although with a warning) has almost no impact on the results. Best values for error estimates and RMSE are typeset in bold.

Table 1. Simulation results for 250 samples of size $n = 100$.

Method	Mean Err.	Std. dev.	RMSE	time
ULS	-0.212	0.173	0.273	0.149
ULS1	0.146	0.230	0.272	0.169
ULS2	0.456	0.333	0.565	0.187
GLS1	-0.025	0.269	0.271	0.228
GLS2	0.076	0.317	0.326	0.229
GLS3	0.137	0.330	0.357	0.233
GWLS	0.175	0.626	0.650	0.322
GWLSc	**-0.023**	0.187	**0.187**	7.13

Table 2. Simulation results for 250 samples of size $n = 200$.

Method	Mean Err.	Std. dev.	RMSE	time
ULS	-0.219	0.115	0.247	0.279
ULS1	0.099	0.135	0.164	0.688
ULS2	0.437	0.231	0.484	0.339
GLS1	-0.086	0.154	0.176	0.439
GLS2	0.062	0.187	0.196	0.482
GLS3	0.152	0.232	0.276	0.416
GWLS	0.076	0.548	0.552	0.569
GWLSc	**-0.045**	0.132	**0.135**	30.04

The results show that GWLS and GWLSc work quite well in this example. However, the best traditional method in this comparison, GWLS2, scored only slightly lower. Nevertheless, it is unsatisfactory that there is no theoretical argument explaining why GLS2 is better than GLS3. Among the new methods, GWLS without the penalty for error correlations needs larger samples to show its potential, while GWLSc also works well with smaller samples. However, run-time for GWLSc is considerably larger than for the other methods. This is partly due to the fact that the calculation of the Jacobi matrix by automatic differentiation in the present code is not optimal. The systematic structure of the Jacobi matrix could be used to provide explicit derivatives. However, implementing this would take some care and time and would make the code harder to read. Thus, it has not been realized yet.

This toy model is used to evaluate the methods in a transparent situation. The method has also been applied to more complex models, such as a non-linear variation of Bollen's model [1], Fig. 8.1, p. 324. In this variation, the linear equation $\eta_2 = \xi_1 + \beta_{21}\eta_1 + \gamma_{21}\xi_2 + \zeta_2$ was replaced with $\eta_2 = \xi_1 + \beta_{21}\eta_1 \cdot \xi_2 + \gamma_{21}\xi_2^3 + \zeta_2$. It turned out that estimates for the parameters β_{21} and γ_{21} were significantly improved by GWLSc. The true values in the simulation were chosen to be 0.5 and 0.8, respectively. ULS1 estimated these values as 0.75 and 0.88, GLS2 estimated them as 0.65 and 0.82, and GWLSc estimated them as 0.59 and 0.78. This provides further evidence that the method has potential, but of course, many more models need to be investigated.

4. Conclusion

The paper has shown that the inequality between the geometric and arithmetic means can be used to estimate parameters and weights in a least squares problem in one consistent step. The results are therefore self-consistent: a priori weights and a posteriori weights are equal. However, the example model shows that the standard deviations of the estimates are rather large, although there is no apparent bias. The controlled version, which penalizes error correlations in a consistent way, performs much better in this respect—especially for small sample sizes—but it is much slower. A practical advice is thus to try the controlled version first. If run-time is excessive try GWLS and GWLSc on random sub-samples. If they differ little, the faster method can be safely applied to the full data set. In

Table 3. Simulation results for 250 samples of size $n = 500$.

Method	Mean Err.	Std. dev.	RMSE	time
ULS	-0.245	0.074	0.256	0.76
ULS1	0.095	0.082	0.123	1.64
ULS2	0.404	0.110	0.411	1.10
GLS1	-0.095	0.106	0.142	1.38
GLS2	0.067	0.130	0.145	1.35
GLS3	0.165	0.137	0.212	1.17
GWLS	**0.003**	0.274	0.274	1.94
GWLSc	-0.063	0.083	**0.100**	328.4

the derivation of the objective function above, normality of errors was assumed. In fact, the method is not robust against violations of this assumptions, e.g. replacing normal errors by uniform errors eliminates the advantage of GWLS and GWLSc.

There are models motivated from educational studies (see, e.g. [3]) that involve nonlinear relationships between latent variables and furthermore constraints on some of the parameters. Simulation studies have shown that GWLSc gives good results in these cases, too, but the publication will be postponed to another publication.

Taken together, this paper presents a method for weight selection that seems to be new, although rather simple. The method obviously has potential, but further research is needed to explore this in more detail.

REFERENCES

1. K. Bollen, *Structural Equations with Latent Variables*, Wiley, 1989.
2. M. W. Browne, *Asymptotically distribution-free methods for the analysis of covariance structures*, British Journal of Mathematical and Statistical Psychology, 37(1), pp. 62-83, 1984.
3. R. Oldenburg, *Do fuzzy-logic non-linear models provide a benefit for the modelling of algebraic competency?*, International Journal of Research in Education Methodology, **13**, pp 1–10, 2022.
4. R. Oldenburg, *Least square estimation of non-linear structural models*, Statistics, Optimization & Information Computing, **12**, pp 281–297, 2024.
5. G.A.F.Seber, and C. J. Wild, *Nonlinear Regression*, Wiley, 2003.