Weighted Least Squares

Jeff Gill and Lucas Leemann

Washington University, St. Louis

Weighted least squares is a standard compensation technique for non-constant error variance (heteroscedasticity), which is common in political science data. By assigning individual weights to the observations the heterescedasticity can be removed by design. The square root of the inverse of the error variance of the observation is typically used as weight. The key idea is that less weight is given to those observations with a large error variance. This forces the variance of the residuals to be constant. Weighted least squares is an example of the broader class of generalized least squares estimators. The idea was first presented by Aitken (1935).

Theory

The ordinary linear model has the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is a $n \times 1$ outcome vector with continuous measure, $\mathbf{X}$ is a $n \times k$ invertible matrix with explanatory variables down the columns and a leading column of ones, $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector to be estimated, and $\boldsymbol{\varepsilon}$ is a $n \times 1$ error vector with assumed mean zero. The ordinary least squares estimator of $\boldsymbol{\beta}$ is achieved by minimizing the squared error terms and is produced by: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. In presence of heteroscedasticity the ordinary least squares estimator of $\boldsymbol{\beta}$ is not BLUE: the best linear unbiased estimator. The term "best" here means it achieves the minimum possible variance.

Weighted least squares allows one to reformulate the model and generate estimators which are in principle BLUE. The introduction of a weight matrix $\mathbf{\Omega}$ into the calculation of $\hat{\boldsymbol{\beta}}$ removes the heteroscedasticity from the model. The $\mathbf{\Omega}$ matrix is created by taking the error variance of the i^{th} case (estimated or known), ν_i, and assigning it to the i^{th} diagonal $\mathbf{\Omega}_{ii} = \frac{1}{\nu_i} = \omega_i$, leaving the off-diagonal elements as zero. So large error variances are reduced by premultiplying the model terms by this reciprocal.

We can premultiply each term in the standard linear model setup by the square root of the $\mathbf{\Omega}$ matrix (that is, by the standard deviation). This "square root" is actually produced from a Cholesky factorization: if $\mathbf{A}$ is a positive definite symmetric ($\mathbf{A}' = \mathbf{A}$) matrix, then there must exist a matrix $\mathbf{G}$ such that: $\mathbf{A} = \mathbf{G}\mathbf{G}'$. A matrix, $\mathbf{A}$, is positive definite if for any nonzero $p \times 1$ vector $\mathbf{x}$, $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$. In our case, this decomposition is greatly simplified because the $\mathbf{\Omega}$ matrix has only diagonal values (all off-diagonal values equal to zero). Therefore the Cholesky factorization is produced simply from the square root of these diagonal values. Premultiplying gives:

$$\mathbf{\Omega}^{\frac{1}{2}}\mathbf{y} = \mathbf{\Omega}^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\varepsilon}. \tag{1}$$

Instead of minimizing squared errors in the usual manner, we now minimize $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{\Omega}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, and the subsequent weighted least squares estimator is found by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Omega}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{y}$. The weighted least squares estimator gives theoretically the best linear unbiased estimate (BLUE) of the coefficient estimator in the presence of heteroscedasticity.

Weighted Least Squares and Feasible Weighted Least Squares

In this setup it is required that the variance of the error, ν_i, has to be known. In principle there are two possibilities; ν_i is derived from the underlying data generating process or ν_i is estimated.

An example for the first is the linear probability model in which the structure of heteroscedasticity is known. In a binary model the variance is $\text{Var}(\varepsilon_i) = \mathbf{X}_i\boldsymbol{\beta}(1 - \mathbf{X}_i\boldsymbol{\beta})$. This gives an expression of the form of heteroscedasticity and allows to estimate a linear model with WLS. The weights are directly computed by using the OLS estimates of $\boldsymbol{\beta}$ to compute $\widehat{\text{Var}}(\varepsilon_i)$. Because $\hat{\boldsymbol{\beta}}_{OLS}$ is an unbiased estimator $\widehat{\text{Var}}(\varepsilon_i)$ is also unbiased. A possible obstacle here can be that the linear probability model may produce a $\hat{y}_i$ which lies outside of the $[0,1]-$interval and therefore produces negative weights (Goldberger, 1964).

Often the form of heteroscedasticity is not known and rather $\hat{\nu}_i$ than ν_i is used. By relying on an estimate of ν_i the WLS estimator is no longer unbiased (Wooldridge 2003: 268). But it is still a consistent estimate and asymptotically more efficient than the OLS estimator. This is often referred to as *feasible weighted least squares.* This implies a two-step procedure. In a first step a linear model is estimated using OLS and based on $\hat{\boldsymbol{\varepsilon}}_{OLS}$ one can derive $\hat{\boldsymbol{\nu}}$ and therefore $\hat{\mathbf{\Omega}}$. The FWLS estimate is obtained by minimizing $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\hat{\mathbf{\Omega}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$. The next section illustrates a feasible weighted least squares estimation.

Example

This illustration is based on political data from Swiss cantons in 1990 (Vatter et al. 2004). The outcome variable is the number of cantonal employees per 1000 inhabitants. The two predictor variables are the degree of proportionality in the electoral system and the cantonal GDP. The results from standard OLS estimation of the linear model do not allow rejection of the null-hypothesis for either of the explanatory variables at standard thresholds.

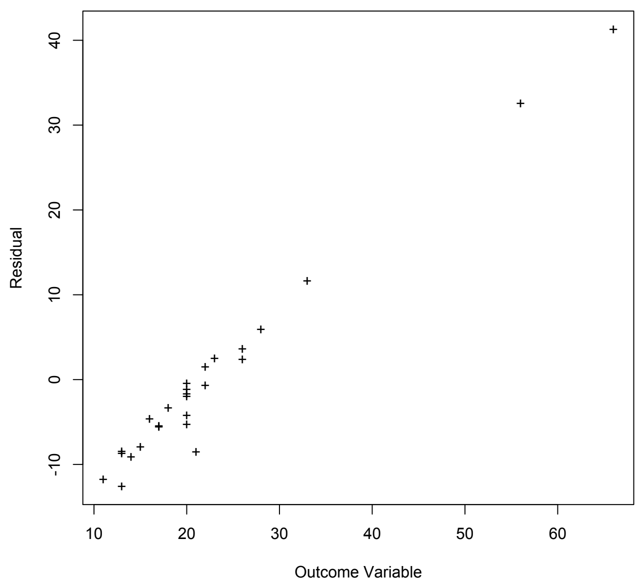
Table 1: Comparison of OLS and WLS

Variable	**OLS**	**WLS**
GDP	0.00014	0.00007
	(0.74)	(1.13)
PR	-1.95035	-1.47071
	(0.46)	(2.15)
Intercept	26.13945	24.45942
	(2.40)	(12.74)
N	26	26
Breusch-Pagan	5.68	.
($p-$value)	(0.017)	.

Note: Outcome variable is number of public employees per 1000 habitants. Absolute $t-$values are in parentheses.

Based on the results presented in Table 1 the conclusion is that PR has no effect on the number of public employees. It is possible to test for heteroscedasticity by using e.g. the Breusch-Pagan test (Breusch and Pagan 1979; Cook and Weisberg, 1983). In this example the squared residuals ($\hat{\varepsilon}_i^2$) are regressed on the predicted values of the outcome ($\hat{y}_i$). If the residuals have a common variance, the explanatory power of the regression is low. The null-hypothesis of the test states a constant error variance. In the example here the test value is 5.68 (χ^2 with one degree of freedom) what corresponds to a $p-$value of 0.017. Based on this the null-hypothesis can be rejected.

Figure 1: Outcome Variable vs. OLS Errors



In Figure 1 the residuals from the OLS estimation are plotted against the outcome variable. There is a strong positive trend which visually confirms the Breusch-Pagan test result.

The WLS approach is an effective way to address the heteroskedasticity in such cases. Based on the estimation results of the OLS procedure it is possible to derive $\hat{\boldsymbol{\Omega}}$, where $\hat{\boldsymbol{\Omega}}_{ii} = \frac{1}{\hat{\varepsilon}_i^2}$ and $\hat{\boldsymbol{\varepsilon}}$ is the estimated error of the OLS procedure. The second column in Table 1 reports the results from the WLS procedure $\hat{\boldsymbol{\Omega}}^{\frac{1}{2}}\mathbf{y} = \hat{\boldsymbol{\Omega}}^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \hat{\boldsymbol{\Omega}}^{\frac{1}{2}}\boldsymbol{\varepsilon}$. The GDP of a canton still does not have a reliable effect on the outcome variable. But in the WLS procedure the effect of the degree of PR on the number of cantonal employees is significant at conventional levels and negative.

Software Issues

Weighted least squares or actually feasible weighted least squares is implemented in all major software packages. Here, the three most common packages, R, Stata, and SAS are considered. At this point a general cautionary note is due. Throughout this text the weight matrix was defined as $\mathbf{\Omega}$ and its diagonal elements as $\frac{1}{\nu_i}$, where ν_i was the variance of the error of observation i. Unfortunately some authors as well as some programs refer to the variance-covariance matrix as $\mathbf{\Omega}$ and therefore their weight matrix is $\mathbf{\Omega}^{-1}$. Therefore it is important to carefully inspect the associated documentation, including the use of the variance versus standard deviation as weights.

In R, the package MASS offers the function `lm.gls` which will estimate a model using FWLS. The required weight matrix is $\mathbf{\Omega}$ but there is an option to use the variance-covariance matrix (in our notation $\mathbf{\Omega}^{-1}$) as the weight matrix (`inverse=TRUE`).

In Stata the command `vwls` allows to estimate a WLS. The weights are here $\sqrt{\nu_i}$ and have to be specified (option `, sd(...)`). This can be easily done by estimating a model with OLS, saving the absolute value of the residuals and then using them as weights.

Finally, in SAS the command `REG` in connection with the statement `WEIGHTS` estimates a model with WLS. Here, the weights are the reciprocal of the variance of the residuals of each observation from the OLS estimation. Therefore the weights are the $\mathbf{\Omega}_{ii}$.

Remarks

Weighted least squares allows to estimate linear models in the presence of heteroscedasticity. By pre-multiplying the observations by a weight matrix, $\mathbf{\Omega}$, the error variance becomes constant.

There are also other possible remedies to heteroscedasticity. First, heteroscedasticity may be the result of a mis-specified model and may require the researcher to change the model. Second, the OLS estimates are still unbiased and it is possible to use robust standard errors such as the Huber-White sandwhich estimator (Huber, 1967; White, 1980) to correct for heteroscedasticity. Weighted least squares estimation is a standard regression tool for social scientists and others, and is used in iteratively weighted least squares to estimate generalized linear models (Gill, 2007).

Cross-References:

TBD

Further Readings:

Aitken, Alexander C. 1935. "On the least-squares and linear combinations of observations" in *Proceedings of the Royal Society of Edingburgh* 55: 42-48.

Breusch, Trevor S., and Adrian R. Pagan. 1979. "A simple test for heteroscedasticity and random coefficient variation" in *Econometrica* 47: 1287-1294.

Cook, R. Dennis, and Sanford Weisberg. 1983. "Diagnostics for heteroscedasticity in regression" in *Biometrika* 70: 1-10.

Faraway, Julien J. 2003. *Linear Models with R.* Boca Raton, FL: Chapman and Hall/CRC.

Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, FL: Chapman and Hall/CRC.

Goldberger, Arthur S. 1964. *Econometric Theory.* New York: John Wiley & Sons.

Greene, William H. 2003. *Econometric Analysis.* Englewood Cliffs: Prentice Hall.

Huber, Peter J. (1967). "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions" in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* vol. I: 221–33.

McCulloch, Charles E. and Shayle R. Searle. 2001. *Generalized, Linear and Mixed Models.* New York: John Wiley & Sons.

Rao, C. Radhakrishna. 1995. *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons.

Rao, C. Radhakrishna and Helge Toutenburg. 1995. *Linear Models. Least*

Squares and Alternatives. New York: Springer.

Ravishanker, Nalini and Dipak K. Dey. 2002. *A First Course in Linear Model Theory.* Boca Raton, FL: Chapman and Hall/CRC.

Vatter, Adrian, Markus Freitag, Christoph Müller and Marc Bühlmann. 2004. *Political, Social, and Economic Data of the Swiss Cantons 1983-2002*, University of Bern, Bern.

White, Halbert (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity" in *Econometrica* 48: 817–38.

Wooldridge, Jeffrey M. 2003. *Introductory Econometrics.* Cincinnati, OH: South-Western College Publishing.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.