11-1-2010

Ridge Regression Based on Some Robust Estimators

Hatice Samkar
Eskisehir Osmangazi University, Turkey, oalpu@ogu.edu.tr

Ozlem Alpu
Eskisehir Osmangazi University, Turkey, hfidan@ogu.edu.tr

Ridge Regression Based on Some Robust Estimators

Hatice Samkar   Ozlem Alpu
Eskisehir Osmangazi University,
Turkey

Robust ridge methods based on M, S, MM and GM estimators are examined in the presence of multicollinearity and outliers. GM_{Walker}, using the LS estimator as the initial estimator is used. S and MM estimators are also used as initial estimators with the aim of evaluating the two alternatives as biased robust methods.

Key words: Multicollinearity, ridge regression, outlier, robust estimation, robust ridge methods.

Introduction

One of the main problems in regression estimation methods is multicollinearity. Multicollinearity is the term used to describe cases in which the regressors are correlated among themselves. The ridge regression model has been advocated in the literature as an alternative to LS estimation for the multicollinearity problem; in this method, which was proposed by Hoerl & Kennard (1970a, b), ridge estimators are used instead of LS estimators.

Another common problem in regression estimation methods is that of non-normal errors. The term simply means that the error distributions have fatter tails than the normal distribution. These fat-tailed distributions are more prone than the normal distribution to produce outliers, or extreme observations in the data. When outliers exist in the data, the use of robust estimators reduces their effects.

Ozlem Alpu is an Assistant Professor of Statistics. Her areas of research are computational statistics, statistical modelling and robust regression. E-mail her at: oalpu@ogu.edu.tr. Hatice Samkar is an Assistant Professor of Statistics. Her research interests include linear models, demography and regression analysis. E-mail her at: hfidan@ogu.edu.tr.

In the case of where both multicollinearity and outliers exist, the use of robust ridge regression is suggested. Robust ridge regression analysis has attracted the attention of some researchers in the literature. Holland (1973) gave the formulas for and derivation of ridge regression methods when weights are associated with each observation, and proposed the combination of ridge regression with robust regression methods. Askin & Montgomery (1980) presented a method based on augmented data sets for combining biased and robust regression techniques. Their estimates were constrained robust estimates, using an appropriately chosen ridge, Stein shrinkage or principal components constraint. Walker (1984) modified Askin and Montgomery's approach to allow the use of GM estimators instead of M estimators (Simpson & Montgomery, 1996). Silvapulle (1991) proposed a new class of ridge type M estimators obtained by using M estimators instead of LS estimators. In addition, he suggested a procedure for choosing the optimal value of the biasing parameter (k) adaptively.

Arslan & Billor (1996) proposed two alternative ridge type GM estimators to handle simultaneously multicollinearity and the existence of outliers. To reduce the effect of outliers, they computed robust estimates for k, and used these estimates to obtain robust ridge estimates for the regression coefficients. Another robust ridge regression estimator was suggested by Pfaffenberger & Dielman (1990). This estimator combines properties of the LAV

(Least Absolute Value) estimator and the ridge estimator, and is called RLAV (Ridge Least Absolute Value) estimator. Simpson & Montgomery (1996) proposed a biased-robust estimator that uses a multistage GM estimator with fully iterated ridge regression to control both influence and collinearity in the regression data set. Simpson & Montgomery (1998a) also evaluated existing and proposed robust methods relative to their performance on a comprehensive group of datasets with and without outliers. In addition, Simpson & Montgomery (1998b) developed and evaluated new robust regression procedures and compared their performance to the best alternatives currently available, in terms of efficiency, breakdown, and bounded influence. They offered the better performing alternatives as possible methods for use in a robust regression scenario.

Wisnowski, Simpson & Montgomery (2002) introduced a robust regression estimator that performs well regardless of the quantity and configuration of outliers. They show that the best available estimators are vulnerable when the outliers are extreme in the regressor space (high leverage). Their proposed compound estimator modified recently published methods with an improved initial estimate and measure of leverage.

In this study, robust ridge regression methods based on M, S, MM and GM estimators are examined in the presence of both outliers and multicollinearity. The computation of GM estimates requires two stages of parameter estimation, an initial estimate that provides a good starting point and a secondary estimate with iterations to a final estimate. LS is used as the initial estimator of GM in the study. In addition, S and MM estimators are used as initial estimators, with the aim of evaluating two alternatives as biased robust methods, as they are the top two robust estimation methods and are also highly efficient and effective against most types of outlier configurations. The performance of the robust ridge estimators is examined by using mean square error (MSE) on a hospital manpower dataset (Myers, 1990).

Methodology
Ridge Regression

Consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2.1)$$

where $\mathbf{y}$ is a vector of n response values, $\mathbf{X}$ is an $n \times p$ matrix of rank p, $\boldsymbol{\beta}$ is a vector such that $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$, and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I_n}$. All variables in this model are corrected for their means and scaled to unit length, so that $\mathbf{X'X}$ is in correlation form.

If the columns of $\mathbf{X}$ are multicollinear, then the least-squares estimator of $\boldsymbol{\beta}$, namely $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$, is an unreliable estimator due to the large variances associated with its elements. The most popular of the methods that can be used to cope with multicollinearity is ridge regression. This method, developed by Hoerl & Kennard (1970a, b), is based on adding a positive constant k to the diagonal element of $\mathbf{X'X}$. This leads to a biased estimator $\boldsymbol{\beta_R}$ of $\boldsymbol{\beta}$, called the ridge estimator and given by:

$$\hat{\boldsymbol{\beta}}_{\mathbf{R}} = (\mathbf{X'X} + k\mathbf{I_n})^{-1}\mathbf{X'y} \qquad (2.2)$$

When both outliers and multicollinearity occur in a dataset, it would seem beneficial to combine methods designed to deal with these problems individually. Thus, robust ridge estimators will be resistant to multicollinearity problems and will be less affected by outliers.

Robust Ridge Regression

The following formula is used to compute robust ridge estimates:

$$\hat{\boldsymbol{\beta}}_{RobustRidge} = (\mathbf{X'X} + k^*\mathbf{I})^{-1}\mathbf{X'X}\hat{\boldsymbol{\beta}}_{Robust}, \qquad (2.3)$$

where $\hat{\boldsymbol{\beta}}_{Robust}$ denotes the coefficient estimates from the robust estimators. Many methods of selecting appropriate k^* values have been proposed in the literature. In this study, the method proposed by Hoerl, Kennard & Baldwin (HKB) (1975), based on LS estimators, has been

used for the selection of the k^* value, building on robust estimators:

$$k^* = \frac{p.\hat{\sigma}^2_{Robust}}{\hat{\boldsymbol{\beta}}'_{Robust}\hat{\boldsymbol{\beta}}_{Robust}}, \qquad (2.4)$$

Where p is the number of regressors, and $\hat{\sigma}^2_{Robust}$ is the robust scale estimator.

Robust Estimations

The most popular of all robust estimation techniques is M estimation, proposed by Huber (1964). The M estimator minimizes the objective function

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}}}{s}\right).$$

Differentiating the objective function with respect to the coefficients β, defining $\psi = \rho'$, and setting the partial derivates to 0, the system of equations can be written

$$\min_{\beta} \sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}}}{s}\right).\mathbf{x_i} = 0 ,$$

where s is a robust estimate of scale.

GM estimators are a natural extension of M estimators (Walker, 1984). GM estimation is multistage estimation with two desirable properties, efficiency and bounded influence. These estimators bound the influence of the observations both in the x and y direction by using weight functions. The GM estimators are solutions to the normal equations

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}}}{s\pi_i}\right)\mathbf{x_i} = 0 ,$$

where the π_i denote the weights. This estimator was developed by Schweppe (Simpson & Montgomery, 1998a).

In the literature, several GM estimation approaches are suggested using various combinations of GM components (objective function, initial estimate, scale estimate, π-weight function, etc.). The GM estimation approach of Walker (1984) is one of the approaches. The GM approach of Walker uses the Schweppe objective function that downweights outliers with high leverage points only if the corresponding residual is large. It is recommended to use the LS as the initial estimator and a non-iterated MAD as the estimate of scale. Convergence to the final estimate is obtained by using iteratively reweighted LS (Wisnowski, Montgomery & Simpson, 2001).

In this study, Walker's (1984) GM method and two alternative GM estimation approaches have been used. In the first approach, the Schweppe function, Huber's Ψ, $\min(1,c|\text{DFFITS}|^{-1})$ and S estimation are used instead of LS for the objective function, leverage function, π-weight function, initial and scale estimation, respectively. Final parameter estimates are found by iteratively reweighted LS.

S estimators developed by Rousseeuw & Yohai (1984) are based on the minimization of the dispersion of the residuals. The S estimator is given by

$$\min_{\hat{\beta}} s(e_1(\beta),...,e_n(\beta)) ,$$

and the scale estimator is

$$\hat{\sigma} = s(e_1(\hat{\beta}),...,e_n(\hat{\beta})) .$$

The dispersion function $s(e_1(\beta),...,e_n(\beta))$ is found as the solution to $\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right)$=K, where K is a constant and $\rho(.)$ is the residual function. Rousseeuw & Yohai (1984) suggest using the following function:

$$\rho(x) = \begin{cases} \dfrac{x^2}{2} - \dfrac{x^4}{2c^2} + \dfrac{x^6}{6c^4} & |x| \le c \\[2ex] \dfrac{c^2}{6} & |x| > c \end{cases} .$$

The 50% breakdown point of the S estimators is achieved by taking c=1.548 and K=0.1995 (Rousseeuw & Leroy, 1987).

In the second GM approach considered, the objective, leverage and π-weight functions are calculated as in the first GM approach, and the MM estimator is used for the initial and scale estimation. The MM estimator is a high breakdown and high efficiency estimator with three stages. The initial estimate is a high breakdown estimate using an S estimate. The second stage computes an M estimate of the scale of the errors from the initial S estimate residuals.

The last step is an M estimate of the regression parameters using a redescending ψ function that assigns a weight of 0.0 to abnormally large residuals (Wisnowski, Montgomery & Simpson, 2001). Because MM estimation combines high breakdown value estimation and M estimation, it has both a high breakdown property and a higher statistical efficiency than S estimation (Chen, 2002). Although MM estimation does not theoretically bound the possible influence, it performs very well in some high leverage outlier situations (Simpson & Montgomery, 1998).

MSE Criterion for Robust Ridge Estimators

To illustrate the performance of robust ridge estimators, the MSE criterion proposed by Silvapulle (1991) is used for M estimation and that of Arslan & Billor (1996) for GM estimation. The MSE of the robust ridge estimators based on the M and GM estimators is as follows:

$$MSE(\hat{\mathbf{\alpha}}_{Robust}(k^*)) =$$

$$\sum_{i=1}^{n} \lambda_i (\lambda_i + k^*)^{-2} \Omega_{ii} + \sum_{i=1}^{n} \left(\frac{k^* \alpha_i}{\lambda_i + k^*} \right)^2 ,$$

$$MSE(\hat{\mathbf{\alpha}}_{Robust}) = \sum_{i=1}^{p} \Omega_{ii} ,$$

where $\mathbf{\Omega}$ is a *(pxp)* $\mathrm{cov}(\hat{\mathbf{\alpha}}_{Robust})$ matrix, and λ_i are the eigenvalues of $\mathbf{X'X}$. Any estimator $\hat{\mathbf{\alpha}}$ of $\mathbf{\alpha}$ has a corresponding estimator $\hat{\mathbf{\beta}}(=\mathbf{P}\hat{\mathbf{\alpha}})$,

such that $MSE(\hat{\mathbf{\alpha}}) = MSE(\hat{\mathbf{\beta}})$, where $MSE(\hat{\mathbf{\beta}})$ refers to the total MSE, $E\left\{ (\hat{\mathbf{\beta}} - \mathbf{\beta})'(\hat{\mathbf{\beta}} - \mathbf{\beta}) \right\}$ (Silvapulle, 1991; Arslan & Billor, 1996).

Results

A hospital manpower dataset taken from Myers (1990) was examined as an example to compare the performance of the considered estimators. This example contains five regressors and one response variable. Because the data have been standardized, the model does not include the intercept term, thus, the $\mathbf{X'X}$ matrix is in the form of a correlation matrix:

$$\mathbf{X'X} = \begin{bmatrix} 1.000 & 0.907 & 0.999 & 0.936 & 0.671 \\ 0.907 & 1.000 & 0.907 & 0.910 & 0.447 \\ 0.999 & 0.907 & 1.000 & 0.933 & 0.671 \\ 0.936 & 0.910 & 0.933 & 1.000 & 0.463 \\ 0.671 & 0.447 & 0.671 & 0.623 & 1.000 \end{bmatrix}.$$

The matrix $\mathbf{X'X}$ has eigenvalues $\lambda_1 = 4.197$, $\lambda_2 = 0.667$, $\lambda_3 = 0.095$, $\lambda_4 = 0.041$, $\lambda_5 = 0.0001$. It is observed that the regressors are moderately to highly correlated. Moreover, $(\lambda_1 / \lambda_5) = (4.1971 / 0.0001) = 83942$, which implies the existence of multicollinearity in the dataset.

In addition, in Figure 1, $x_6, x_7, x_8, x_{10}, x_{12}, x_{15}, x_{16}$ and x_{17} are flagged as outliers or leverage points, and the points $(y_6, x_6), (y_7, x_7)$, $(y_8, x_8), (y_{10}, x_{10})$, $(y_{12}, x_{12}), (y_{15}, x_{15})$, (y_{16}, x_{16}) and (y_{17}, x_{17}) are regression outliers. The points $(y_8, x_8), (y_{10}, x_{10})$, $(y_{12}, x_{12}), (y_{15}, x_{15})$, (y_{16}, x_{16}) and (y_{17}, x_{17}) are called bad leverage points. Regression outliers for which *x* values are not leverage points are called outliers in the *y* direction. In Figure 1, the points $(y_6, x_6), (y_7, x_7)$ are outliers in the *y* direction as well.

In the presence of outliers in the data, the use of robust methods provides more stable parameter estimates. With this aim, initial robust regression estimates have first been calculated to

Figure 1: Robust Residuals versus Distances for Hospital Manpower Data
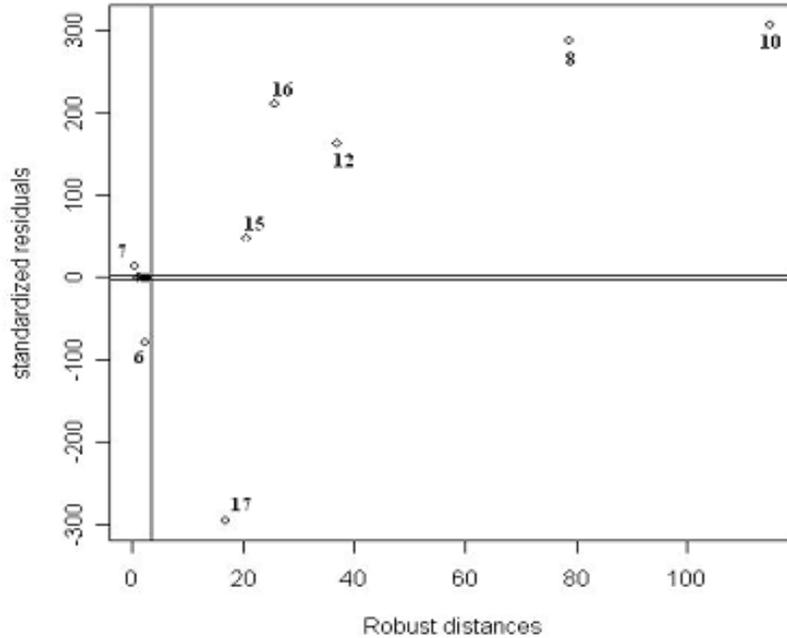
obtain robust ridge estimates in the presence of both multicollinearity and outliers; these estimates are given in Table 1.

Table 1: Initial Robust Parameter Estimates

$\hat{\beta}_{Robust}$	M	S	MM
$\hat{\beta}_1$	-0.2159	0.4036	0.5948
$\hat{\beta}_2$	0.1782	0.6739	0.6543
$\hat{\beta}_3$	1.1891	0.3053	0.1099
$\hat{\beta}_4$	-0.0759	-0.3054	-0.2582
$\hat{\beta}_5$	-0.1281	-0.0909	-0.0853
$\hat{\sigma}$	0.0264	0.0199	0.0199

As shown in Table 1, the $\hat{\beta}_1$ value of the M estimator is found to have a negative sign. This value is inconsistent with the $\hat{\beta}_1$ values obtained from the S and MM estimators. In the presence of multicollinearity in a dataset, the signs of parameters can be found to be different from expectations. The sign of the $\hat{\beta}_1$ value can be said to occur inversely due to the potential effects of multicollinearity. In addition, the magnitudes of the parameter values for the M estimator are fairly different from those of the S and MM estimators. It is thought that the S and MM estimates are better than the M estimates because the scale estimates of S and MM are more efficient than the M estimates.

Second, biasing parameters (k^*) have been found by using the estimates in Table 1. Robust ridge estimates via the biasing parameters are calculated and shown in Table 2. In Table 2, the sign of $\hat{\beta}_1$ value of the M estimate is the same as that of the other robust ridge estimates. The effect of multicollinearity on the sign of the $\hat{\beta}_1$ value is removed by using ridge regression. The magnitudes of the parameter estimates are coherent with each other, except ridge regression estimates based on M estimates.

Table 2: Robust Ridge Parameter Estimates

$\hat{\beta}_{Robust\ Ridge}$	M	S	MM	GM$_{Standard}$	GM$_S$	GM$_{MM}$
$\hat{\beta}_1$	0.4621	0.3329	0.3343	0.3927	0.3600	0.3426
$\hat{\beta}_2$	0.1755	0.6663	0.6472	0.4351	0.4913	0.5117
$\hat{\beta}_3$	0.5239	0.3540	0.3444	0.4162	0.3874	0.3683
$\hat{\beta}_4$	-0.0857	-0.2801	-0.2299	-0.1491	-0.1408	-0.1263
$\hat{\beta}_5$	-0.1325	-0.0842	-0.0776	-0.0738	-0.0678	-0.0681
k^*	0.0032	0.0024	0.0023	0.0055	0.0030	0.0031
MSE	0.0398	0.0249	0.0242	0.0314	0.0209	0.0200

From Table 2, it is observed that the result of MSE based on M estimation is the worst among other robust methods. The worst value of the scale estimates in Table 1 belongs to the M estimates; thus, the results of Table 1 are consistent with those of Table 2. The result of MSE for GM$_{Walker}$ is the second worst value. GM estimators were developed to overcome the deficiency of M estimators; Table 2 shows that the MSE value of GM is better than that of M. It has been noted that GM estimation is multistage, while the initial estimates of GM$_{Walker}$ are based on LS. The method of LS is not robust in the presence of outliers in the data. For this reason, the MSE of GM$_S$ and GM$_{MM}$, proposed in the study as alternatives to GM$_{Walker}$, are less than that of GM$_{Walker}$. The MSEs of the GM$_S$ and GM$_{MM}$ estimates are significantly less than the MSEs of the other robust ridge estimates. Furthermore, the results of MSEs for robust ridge estimates based on MM are less than those of the S estimates.

Conclusion

In this study, in the presence of both multicollinearity and outliers in a dataset, a biasing parameter k^* is calculated using the $\hat{\beta}_{Robust}$ and $\hat{\sigma}_{Robust}$ values obtained from several robust methods (M, S, MM, GM$_{Walker}$, GM$_S$ and GM$_{MM}$), robust ridge estimates are then obtained. The performance of the robust estimators is affected by the percentage of data that are outliers, the location of the outliers in the x and y directions and their magnitudes. For this reason, the performance of the estimators considered must be interpreted in terms of these components.

The performance of ridge estimators based on M, S, MM, GM$_{Walker}$, GM$_S$ and GM$_{MM}$ estimation methods have been considered for the dataset in terms of the MSE criterion. For this dataset, the result of MSE from robust ridge regression based on M estimation is the worst among all robust techniques. Because the data includes outliers in both the x and y directions, the M estimators cannot bound the outliers in the x direction. In this situation, GM estimators, which bound the effects of outliers in both the x and y directions, are expected to have better performance than M estimators. Thus, under these circumstances, it has been shown that the ridge GM estimators would be preferred.

However, the result of MSE for GM$_{Walker}$ is the second worst value. There are several outliers in the x-direction in the data and a few of them are extreme. On the other hand, the GM$_{Walker}$ method uses LS estimates, which are not robust, as initial estimates. In this situation, it is expected that the MM and S estimators should have better performance in

terms of MSE, because MM and S estimators are high breakdown estimators.

GM$_S$ and GM$_{MM}$ estimates combine the properties of high breakdown, efficiency and robustness against outliers in the x and y directions. Consequently, the MSE of the GM$_S$ and GM$_{MM}$ estimates are somewhat smaller than that of the GM$_{Walker}$ ridge estimates. According to this result, it can be said that using robust estimation methods as an initial estimator for GM give more efficient and high breakdown estimates when the dataset contains outliers in the x and y directions. As the result, the performance of robust ridge regression estimates based on GM$_S$ and GM$_{MM}$ estimators met expectations in terms of the MSE criterion in this dataset.

References

Arslan, O., & Billor, N. (1996). Robust ridge regression estimation based on the GM-estimators. *Journal of Mathematical and Computational Science*, *9*(*1*), 1-9.

Askin, G. R., & Montgomery, D. C. (1980). Augmented robust estimators. *Techonometrics*, *22*, 333-341.

Chen, C. (2002). Robust regression and outlier detection with the ROBUSTREG procedure. *Proceedings of the Twenty-seventh Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc.

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55-67.

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, *12*, 69-82.

Hoerl, A. E, Kennard, R. W., & Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistical Theory*, *4*(*2*), 105-123.

Holland, P. W. (1973). Weighted ridge regression: Combining ridge and robust regression methods. *NBER Working Paper Series*, *#11*, 1-19.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*, 73-101.

Myers, R. (1990). *Classical and Modern Regression with Applications*. Boston, MA: Duxburry.

Pfaffenberger, R. C., & Dielman, T. E. (1990). A comparison of regression estimators when both multicollinearity and outliers are present. In *Robust regression: Analysis and applications*, K. Lawrence & J. Arthur (Eds.), 243-270. New York: Marcel Dekker.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

Rousseeuw, P. J., & Yohai, V. J. (1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis*, J. Franke, W. Hardle & D. Martin (Eds.), 256-272. Heidelberg, Germany: Springer-Verlag.

Silvapulle, M. J. (1991). Robust ridge regression based on an M estimator. *Austrailian Journal of Statistics*, *33*, 319-333.

Simpson, J. R., & Montgomery, D. C. (1996). A biased robust regression technique for combined outlier-multicollinearity problem. *Journal of Statistical Computation Simulation*, *56*, 1-22.

Simpson, J. R., & Montgomery, D. C. (1998a). A performance based assessment of robust regression methods. *Communications in Statistical Simulations*, *27*(*4*), 1031-1049.

Simpson, J. R., & Montgomery, D. C. (1998b).The development and evaluation of alternative generalized M estimation techniques. *Communications in Statistical Simulations*, *27*(*4*), 999-1018.

Walker, E. (1984). *Influence, collinearity and robust estimation in regression*. Unpublished Ph.D. dissertation, Department of Statistics, Virginia Polytechnic Institute.

Wisnowski, J. W., Montgomery, D. C., & Simpson, J. R. (2001).A comparative analysis of multiple outlier detection procedures in the linear regression model. *Computational Statistics and Data Analysis*, *36*, 351-382.

Wisnowski, J. W., Simpson, J. R., & Montgomery, D. C. (2002). An improved compound estimator for robust regression. *Communications in Statistical Simulations*, *31*(*4*), 653-672.