Journal of Data Science **10**(2012), 597-617

Obtaining Estimators from Correlation Coefficients: The Correlation Estimation System and R

Rudy A. Gideon
University of Montana

Abstract: Correlation coefficients are generally viewed as summaries, causing them to be underutilized. Creating functions from them leads to their use in diverse areas of statistics. Because there are many correlation coefficients (see, for example, Gideon (2007)) this extension makes possible a very broad range of statistical estimators that rivals least squares. The whole area could be called a "Correlation Estimation System." This paper outlines some of the numerous possibilities for using the system and gives some illustrative examples. Detailed explanations are developed in earlier papers. The formulae to make possible both the estimation and some of the computer coding to implement it are given. This approach has been taken in hopes that this condensed version of the work will make the ideas accessible, show their practicality, and promote further developments.

Key words: Classical regression, correlation coefficient, density parameter estimation, nonlinear regression, rank statistics.

1. Introduction

Amazingly, virtually everything that one does with least squares (LS) or normal theory can be done with any of a multitude of correlation coefficients (CCs) and it can be done in a coherent fashion, with essentially one basic equation. Both continuous and rank based CCs use the same formulae without change of notation. This means that the same computer code could be written to encompass all estimations involving all the CCs. A user could designate which CC was desired and then all computer calculations thereafter would be based on that choice with minimal change to the rest of the computer code, including least squares through Pearson's CC. However, since the degree of robustness of all estimations emanating from a particular CC depends on the degree of robustness of the CC itself, CCs other than Pearson's are usually more desirable. Appendix A.1 gives R-code for five correlation coefficients which are defined in Gideon (2007); three of these are also defined in Section 2.

One focus of this paper is to show how to use any correlation coefficient to estimate location, scale and slope coefficients in simple and multiple linear regression. Once these procedures are developed, the Correlation Estimation System (CES) is extended into nonlinear regression and estimation of parameters for a particular density type. Some of the results are illustrated with a continuous and with a rank based CC using absolute values. Although not done in this paper the CES can be easily extended into time series and general linear models; see Sheng (2002). Many of these areas have been tested using various CCs over 25 years and all results lead one to believe in the value of the approach.

2. Simple Linear Regression

For a random sample (x, y) in a simple linear regression model and for CC r, let b be the estimate of β, i.e. b is the estimated slope of the regression line, which is the line that makes the residuals $y - bx$ uncorrelated with x. In other words, by analogy with the population correlation of the independent random variable with the residual random variable being zero, the estimate of β is found by setting the sample equivalent to zero, that is, by solving the regression equation

$$r(x, y - bx) = 0. \tag{1}$$

The function $r(x, y-bx)$ is non-increasing as b increases, which makes equation (1) easy to solve numerically. The code for this is in Appendix A.2. Of course, for least squares, solving this equation with Pearson's r_p is equivalent to the more familiar minimization process. The median of the uncentered residuals provides a robust estimate of the intercept when used with a robust r. However, it is not necessarily the case that the solution to equation (1) corresponds to the solution to a particular minimization for every correlation coefficient. As an example, two CCs are now introduced – one continuous, r_{av}, and one rank based, r_{mf}. Both of these are examples of CCs in which solving equation (1) does not correspond to a minimization. The general framework for them is found in Gideon (2007). First a continuous absolute value CC is given.

Let $SA_x = \sum |x_i - \bar{x}|$ and similarly for y, and define

$$r_{av} = \frac{1}{2}\Big(\sum \left| \frac{x_i - \bar{x}}{SA_x} + \frac{y_i - \bar{y}}{SA_y} \right| - \sum \left| \frac{x_i - \bar{x}}{SA_x} - \frac{y_i - \bar{y}}{SA_y} \right| \Big).$$

This is the FOURTH CC in Appendix A.1, abscor. For a bivariate normal distribution with CC ρ, the population value of r_{av} is $\rho_{av} = (\sqrt{1+\rho} - \sqrt{1-\rho})/\sqrt{2}$. The inverse is $\rho = \rho_{av}\sqrt{2 - \rho_{av}^2}$, which is needed in multiple linear regression; the R-code is given as MADI in Appendix A.1. (MAD is a Median Absolute Deviation correlation coefficient, Gideon (2007), which extends and is compatible with the existing MAD scale estimator; it has the same inverse as the Absolute Value CC, r_{av}.)

In the same way Spearman's CC is found by substituting ranks in place of the original data in Pearson's CC, substituting ranks in r_{av} gives Gini's CC. However, the formula is simplified by ordering the original data by the x-values so that the data replacement is

$$\left.\begin{array}{l} x_1, y_1 \\ x_2, y_2 \\ \vdots \\ x_i, y_i \\ \vdots \\ x_n, y_n \end{array}\right\} \longrightarrow \begin{array}{ll} 1, & q_1 \\ 2, & q_2 \\ \vdots & \\ i, & q_i \\ \vdots & \\ n, & q_n \end{array}$$

where q_i is the y point that corresponds to the i^{th} smallest x value. Gini's CC is $r_{mf} = (\sum |n+1-q_i-i| - \sum |q_i - i|)/\left[n^2/2\right]$; the subscript mf stands for modified footrule of Spearman and the notation in the denominator is that of greatest integer. The code is given as the THIRD CC, Gini, in Appendix A.1. For the bivariate normal, the population value is $\rho_{mf} = (2/\pi)[\arcsin((1+\rho)/2) - \arcsin((1-\rho)/2)]$. The inverse, denoted GinI, is found in Appendix A.1. In 1906 Spearman attempted to formulate a CC based on just $\sum |q_i - i|$, but Gini's valid version was not formulated until 1914.

Tied value concerns must always be addressed when using rank based CCs. It has been found that producing a unique value for any nonparametric CC using the max-min tied value procedure outlined in Gideon and Hollister (1987) handles this issue. This procedure is found in Appendix A.1 as R- function *fxyrk*. It can be used on all nonparametric CCs, but in the case of the Greatest Deviation Correlation Coefficient (GDCC) it must be used, as it is the only known viable procedure. The definition of GDCC is $r_{gd}(x,y) = (\max_{1\le i\le n}(d_i^-) - \max_{1\le i\le n}(d_i^+))/\,[n/2]$, where $d_i^+ = \sum_{j=1}^{i} I(q_j > i)$, $d_i^- = \sum_{j=1}^{i} I(n+1-q_j > i)$, and I is the indicator function. The same data transformation as above has been used. The code is given as the FIRST CC, GDave, in Appendix A.1. While the CES technique is completely general, the population value is not always known. But, for the elliptically contoured densities, the population value of r_{gd} can be given explicitly as $\rho_{gd} = (2/\pi)\arcsin(\rho)$; the expression for Kendall's Tau is the same. The results for both ρ_{av} and ρ_{mf} are derived in an unpublished paper, Population Values, on the website, www.math.umt.edu/gideon. ρ_{gd} can be found in Gideon and Hollister (1987). The population inverse, $\sin((\pi/2)\rho_{gd})$, is denoted GDI and is found in Appendix A.1.

Note that if the model assumption is the bivariate normal or the bivariate t class of distributions then X and $Y - \beta X$ are uncorrelated so $\rho(X, Y-\beta X)$ is zero; both ρ_{av} and ρ_{mf} are also zero for these random variables as can be checked by substituting $\rho = 0$ into the expressions above.

3. Scale Equations

When working with least squares, finding the variation in the regression residuals via least squares is natural; the two pieces (slope and scale) are naturally connected. When working with these general CCs, we seek the same kind of natural connectivity and so it is inappropriate to use least squares for the scale estimate. For example, in the case of r_{av}, the measure of variation of x or the variation of the residuals from the regression should be based on absolute values, because the original slope calculations were. In general, the same ideas must be employed for both slope and scale to attain the connectivity we desire. It is the author's experience that this natural connectivity requirement is necessary to retain desirable qualities such as the degree of robustness.

With this in mind, the variation in the estimate of residuals is found by solving for s in

$$r(y^0, (y - bx)^0 - sy^0) = 0. \tag{2}$$

Here the superscript means the data are ordered; s estimates the ratio of standard deviations, σ_{res}/σ_y, as one quantity not as the quotient of two quantities. The scale equation (2) is examined in detail in Gideon and Rothan (2009). The quantity σ_{res}/σ_x is estimated using equation (2) with x^0 in place of y^0, leaving the residual term alone. This is needed in estimating the variation of the slopes in the linear regressions. Code for these is in Appendix A.3.

s can also be viewed as the slope, not of the standard regression line, but of a specialized regression, discussed in Section 5. Note that this is essentially the same correlation regression equation (1) applied to ordered data: a numerical routine that solves (1) will also solve (2). Moreover, equation (2) can be used in an entirely different way. Instead of solving (1) for b and utilizing equation (2) to get s (called the Regression Equation Technique or RET), (2) could be used in a minimization: find the b that minimizes s. In this way, equation (2) could subsume equation (1). This minimization is called the Optimization Technique (OT); the idea is exploited in Section 7. Note too that the ordering on the residuals is independent of the ordering on the y variable. This is key because the set of residuals en masse is being measured relative to the set of y values; CES views these globally. The residual for any particular x may be reordered into a different position by the iterative technique used in seeking a minimum. For the normal distribution, the quantity s is estimating $\sigma_{res}/\sigma_y = \sqrt{1 - \rho^2}$, where ρ is the population correlation coefficient.

4. Multiple Linear Regression

For matrix $X_{n\times p}$ (n rows of independent data in which the i^{th} column is the

regressor variable, x_i) and y (the dependent variable), the associated regression equations are

$$r(x_i, y - X_{n\times p}\widehat{\beta}) = 0, \quad i = 1, \cdots, p. \tag{3}$$

The solution is a vector $\widehat{\beta}$ whose i^{th} component is the coefficient of variable x_i. Rummel (1991) shows how to solve these using Gauss-Seidel for the case when r is GDCC, and again, ties are not a problem using the max-min procedure. Every CC has its own set of regression equations and linearity properties. The regression equations correspond to the normal equations in the case of Pearson's r_p. Note that these regression equations give a way to incorporate Kendall's Tau into the realm of multiple regression, which is desirable as Tau is moderately robust and has several remarkable features. (See Sen (1968) or Gideon and Rummel (1992) for the simple linear regression case and for an illustrated look at this work specialized to Kendall's Tau see Correlation and Regression without Sums of Squares on the website.)

The variation in the estimate of residuals, s, which is the slope of a regression line on ordered data (superscript 0) is found as the solution to

$$r(y^0, (y - X_{n\times p}\widehat{\beta})^0 - sy^0) = 0. \tag{4}$$

This is, of course, the multiple regression version of equation (2); it is another instance of the Regression Equation Technique (RET). Again note that a minimization could be used instead: find the $\widehat{\beta}$ that minimizes s; in other words the Optimization Technique (OT) could be employed. The R-program optimize is used for simple linear regression described in Appendix A.4 whereas nlm, described in Appendix A.5, must be used for multiple linear regression.

For completeness the CES also needs a measure of the multiple correlation coefficient. This is provided by the analogue of the coefficient of determination, $1 - s^2$. Thus, a multiple CC can be defined by

$$\sqrt{1 - s^2}. \tag{5}$$

5. A Correlation Coefficient Approach to Minimization of the SD ratio, σ_{res}/σ_y

If one wants a minimal variation estimate s, again equation (4) is used but in the reverse way. Now the coefficients $\widehat{\beta}_i$, $i = 1, \cdots, p$ are chosen to minimize s in (4). Again, for the normal distribution, s is estimating $\sigma_{res}/\sigma_y = \sqrt{1 - \rho^2}$ where now ρ is the multiple correlation coefficient and so minimizing s is equivalent to maximizing ρ. The results from RET and OT are not always the same, but as expected, it has been found that they are usually quite similar. The dearth of linearity properties of some CCs is a major cause of this difference. In addition,

the convergence criteria, as well as the overall computing environment, affects the results. A few examples are given in Section 7 using two of the CCs defined above. All results for each method and all the examples were high quality, leading to an even greater conviction of the value of the methods.

For later applications, equation (4) is expressed in a more general form. Let the response variable y be modeled by some function, f, with argument x, which relies on vector parameter β so the equation becomes

$$r(y^0, (y - f(x, \widehat{\beta}))^0 - sy^0) = 0. \tag{6}$$

The reason this is more general is that now f can be nonlinear. It is helpful to view this equation geometrically. The ordered residuals are plotted against the ordered ys and a simple linear regression is performed whose slope s is used to ascertain the closeness of the fit. This is done not by focusing on the sum of individual vertical deviations but by forcing the residuals overall to be relatively small, as measured by the slope s of equation (2) or (6); that is, by regressing the sorted residuals on the sorted ys. Thus equation (6) plays the role in CES that the residual sum of squares does in classical least squares analysis. Theoretically s will vary between 0 and 1; a value of 0, or a correlation of 1, denotes an exact fit whereas a value of 1, or a correlation of 0, means there is no information in the model under discussion.

6. Computation of the Standard Errors of the Regression Coefficients

In this section the asymptotic standard errors of the regression coefficients for several CCs are given. These are derived using the asymptotic distributions of the CCs and their population forms, which have only been derived for elliptically contoured distributions such as t distributions. Gideon (2010) illustrates the method. An example is given in the next section. The asymptotic distributions are given for Kendall's Tau, Greatest Deviation, and the absolute value CC. In what follows σ^{ii} is the (i, i) element of matrix Σ^{-1}. It can be shown, in a multiple linear regression setting, using matrix algebra and the relationship between Σ, the covariance matrix, and R, the correlation matrix, that $\sigma^{ii}\sigma^2_{res} = r^{ii}(\sigma^2_{res}/\sigma^2_{x_i})$ where r^{ii} is the (i, i) element of matrix R. The matrix R contains the estimates of the population parameters, ρ. Thus, the relationship between the chosen CC and ρ must be used. For example, $\rho = \rho_{av}\sqrt{2 - \rho^2_{av}}$, which is of course used with the population estimates, r_{av}. Expressions using the other CCs are given in R-code in Appendix A.1. The quantity $\sigma^2_{res}/\sigma^2_x$ is estimated as a unit as explained in Section 3 using a version of equation (2). The term r^{ii} can be interpreted as the inverse of the residual variance of the the i^{th} variable regressed on the other standardized independent variables. Thus, these terms are always one or more

(for one variable, equal to one) and a large size indicates a close relationship to the other variables. See Healy (1986).

The results are that $\widehat{\beta}_i$ is asymptotically distributed:

$$\mathcal{N}\big(\beta, \frac{\pi^2 r^{ii}\sigma^2_{res}}{9(n-1)\sigma^2_{x_i}}\big), \text{ for Kendall; } \mathcal{N}\big(\beta, \frac{\pi^2 r^{ii}\sigma^2_{res}}{4n\sigma^2_{x_i}}\big), \text{ for GDCC; and}$$

$$\mathcal{N}\big(\beta, \frac{(\pi-2) r^{ii}\sigma^2_{res}}{n\sigma^2_{x_i}}\big), \text{ for the absolute value CC.}$$

7. Examples Using CES and Comparisons of the RET and OT Methods

In this section, some examples are given to illustrate the use of the CES method. The examples include simple and multiple linear regression as well as a nonlinear model, done with various CCs and compared to least squares and the Pearson correlation coefficient method using equations (1) through (6). Every example includes r_{av}. Some of the examples use data from Chatterjee and Price (1991, C/P), some from Draper and Smith (1998, D/S), and some use simulation data. Several additional correlation coefficients were used in simple and multiple linear regression examples. Section 8 introduces a specialized version of equation (6) that allows estimation of parameters for univariate distributions.

7.1 Simple Linear Regression

This first example illustrates the broadness of the method and the robustness of several of the CCs by performing a simple linear regression on some data in C/P with several correlation coefficients. The R-programs to do this depend on utilizing the correlation coefficient in a way that the R-routine uniroot can accept. The set of commands that calls uniroot can easily be given a new correlation coefficient as the argument so a new slope is obtained. It is interesting that both continuous and rank based correlation coefficients work equally well. Table 1 is given to allow comparison of the slope estimates for each correlation coefficient. Gideon (2007) should be consulted for illustrative information on the various CCs.

The data from C/P, Television Rating Data, from Chapter 2, Simple Linear Regression, is used to illustrate the results from the RET formulas (1) and (2) for some correlation coefficients, specifically r_{av}, Gini, MAD, GDCC, and Pces. (Pces means using equations (1) and (2) with Pearson's correlation coefficient; Pces stands for Pearson's with CES.) This data had four outliers, two on each side of the bulk of the data that made the regression line steeper than it would otherwise have been. Also the OT using equation (4) is illustrated for this simple linear regression with correlation coefficients r_{av} and MAD.

Table 1: C/P data fitted using various correlation coefficients

Method	Slope from (1) or (4)	Intercept	Relative s from (2) or (4)
r_{av} with RET	0.558	2.329	0.813
r_{mf} (Gini's) with RET	0.583	2.208	0.915
MAD with RET	0.268	3.532	0.908
GDCC with RET	0.383	3.060	0.886
Pces with RET	0.665	1.642	0.737
r_{av} with OT from (4)	0.571	2.264	0.812
MAD with OT from (4)	0.283	3.456	0.871

To compare, the LS values are slope = 0.665, intercept = 1.706, and $s = 0.791$ and LS with 4 outliers deleted gives slope = 0.260, intercept = 3.713, and $s = 0.935$

For the slope calculation, Pces gives the same result as LS, but the median rather than the mean of uncentered residuals was used for all the intercept estimations, including the Pces calculation. Further, equation (2) with Pces was used to calculate s, which estimates the ratio σ_{res}/σ_y directly, whereas LS uses two separate estimates for σ_{res} and σ_y, and then divides to compute this ratio.

Assuming that the least squares results with the outliers deleted actually give the best estimate, the results in Table 1 make it apparent that MAD and GDCC are by far the most robust methods for this data; they are close to outlier-deleted results. When the outliers were deleted, all the other slopes moved considerably closer to the GDCC and MAD values which hardly changed; exhibiting the exact numbers did not seem necessary. Note that the last column shows that almost all the correlational methods give a better estimate of the s ratio than LS, again assuming that the LS results with the outliers deleted actually give the best estimate. See C/P for their discussion. The two OT rows of Table 1 are produced by minimizing s in equation (2) or (4). Note that the OT estimates of slope for both MAD and r_{av} are close to the corresponding estimates using the RET.

7.2 Multiple Linear Regression

In this section the Attitude Survey Data from the Multiple Regression Model, Chapter 3 of C/P, is used as well as some multivariate normal data that is generated randomly with a random correlation structure. For the C/P data there are six regressor variables used to predict the response variable. In Table 2, all six are used in the fit and in Table 4, the two most important variables, one and three, as determined in C/P, are used. For the RET, the correlation coefficients MAD, r_{av}, r_{mf}, GDCC, and Kendall's Tau are computed for all six variables, as well as for variables one and three. The LS results are also given. The R-instructions using the R-routine nlm for the solutions are sketched in Appendix A.5 under the

heading RET. The standard error material in Section 6 is illustrated with this data in Table 3 and Table 4 for r_{av}, GDCC, and Kendall's Tau and compared to least squares. It is seen that the standard errors for r_{av} and Kendall's Tau are very comparable to the least squares ones. Examination of Tables 2 and 3 indicates that least squares may have misinterpreted the effect of variable five. In analogy with LS the t distribution can probably be used with these standard errors to obtain confidence intervals. GDCC, being the most robust, has somewhat larger SEs.

Table 2: RET for 5 CCs, C/P data, all six variables

Method	No. iter	int	x_1	x_2	x_3	x_4	x_5	x_6	rel s
MAD	250*	31.41	0.667	-0.084	0.300	0.080	-0.329	-0.085	0.650
r_{av}	7	18.82	0.555	-0.029	0.255	0.208	-0.104	-0.210	0.523
r_{mf}	13	24.42	0.500	0.010	0.254	0.218	-0.180	-0.177	0.563
GDCC	26	33.01	0.320	0.113	0.393	0.250	-0.374	-0.131	0.556
Tau	12	21.67	0.475	-0.028	0.350	0.196	-0.133	-0.223	0.561
LS		10.79	0.613	-0.073	0.320	0.082	0.038	-0.217	0.581**

*250 iterations was set as the upper limit and so MAD did not converge. The reason may be that median methods have intervals for the solutions rather than specific points. Even if convergence were near, the solution interval may be just big enough to contain the various iterates, not allowing convergence.

**This was computed by using $s_{res} = 7.068$ and $s_y = 12.173$ so the ratio is 0.581.

Table 3: Standard errors on all six variables

Method	SE x_1	SE x_2	SE x_3	SE x_4	SE x_5	SE x_6
r_{av}	0.167	0.142	0.185	0.235	0.142	0.175
GDCC	0.226	0.188	0.256	0.279	0.250	0.227
Tau	0.157	0.135	0.171	0.215	0.154	0.196
LS	0.161	0.136	0.169	0.221	0.147	0.178

Table 4: RET for 5 CCs, C/P data, variables 1 and 3

Method	No. iter	int	x_1	x_3	rel s
MAD	9	10.362	0.602	0.265	0.531
r_{av}	7	10.320	0.640 (0.134)	0.218 (0.150)	0.531
r_{mf}	5	7.917	0.659	0.243	0.545
GDCC	5	11.141	0.489 (0.222)	0.372 (0.226)	0.618
Tau	4	7.782	0.649 (0.129)	0.255 (0.142)	0.548
LS		9.871	0.642 (0.119)	0.211 (0.134)	0.560*

*This was computed by using $s_{res} = 6.817$ and $s_y = 12.173$ so the ratio is 0.560. For comparison, using the OT (Appendix A.5) on r_{av} took 18 iterations and gave intercept 10.82, coefficients 0.618 and 0.258, and minimum s of 0.527. Values in parentheses are SEs.

The OT results are given in Table 5. The criterion to judge convergence was the smallness of either the sum of the absolute value of the changes in the coefficients or the sum of the absolute values of the correlations in equation (3). As mentioned earlier, the results from the OT method and RET method are not the same. For example, the largest difference is 0.82 of a SE in the coefficient of variable five for r_{av}.

Table 5: OT for 2 CCs, C/P data, all six variables

Method	No. iter	int	x_1	x_2	x_3	x_4	x_5	x_6	rel s
r_{av}	39	8.19	0.594	-0.099	0.350	0.118	0.012	-0.133	0.502
Pces*	35	10.43	0.639	-0.053	0.282	0.065	0.024	-0.157	0.506

*Again Pces is Pearson's correlation coefficient but used with the OT in equation (4). Note that from equation (5), the estimate of the multiple correlation coefficient for the Pces method is 0.862 and for r_{av} is 0.865.

For the second example, the RET for multiple regression was explored using a random generation of seven normal variates with a random correlation structure; one variable was regressed on the other six. Least squares was compared to the Pces and the r_{av} methods. Table 6 contains some rank comparisons on individual $\widehat{\beta}_i$s. The LS method does not seem to be better and many times is worse even for strictly normal data. It is worth noting that this observation does not contradict the Gauss-Markov theorem since the criterion for success in the CES is not that the standard residual variance is a minimum, but rather that the relative ratio (σ_{res}/σ_y) is a minimum. Also note that in the C/P example above, the relative ratio is sometimes smaller for correlation coefficients other than Pearson's.

Table 6: Total ranks of coefficients from 16 simulations of a 7-variate normal distribution

Method	x_1	x_2	x_3	x_4	x_5	x_6
Pces	32.5	33	34.5	33	34.5	32.5
r_{av}	25	27	26	31	28	31
LS	38.5	36	35.5	32	33.5	32.5

Recall that the OT relies on a geometric approach in which the slope of a simple linear regression estimates directly the relative ratio. In the RET, this relative ratio is calculated after estimating the βs. In LS theory, the two approaches (minimizing the standard variance or using RET with Pearson's correlation coefficient or Pces) are identical, whereas in general in CES, the two approaches (RET and OT) give reasonably close results, but are not usually identical. Since any method would win a comparison within its own measurement technique, to give a valid comparison, a rank counting procedure was used. To compare Pces, r_{av}, and LS,

16 runs were made of the 7-variate normal and the closest to the true regression coefficient was recorded by ranks. This time, for each of the 16 random correlation structures, a data set was generated with known population values. The comparison is shown in Table 6; rank 1 was closest to the true parameter, and so forth. So $32 = (16)(2)$ is the expected total sum of the ranks for each column if all three methods are equally good. Note that the r_{av} method was best because estimates for all six coefficients were under the expected 32.

7.3 Nonlinear Regression

This section gives two examples of estimating the parameters in a nonlinear situation. The illustrations are kept simple by using the exponential distribution but generally any nonlinear model could be considered. Equation (6) is used first with $f(x, a, b) = a \exp(-bx)$ where parameters $a, b > 0$; data was randomly generated by adding normally distributed error to the model. An example from the nonlinear regression chapter in D/S uses an actual data set with the model $f(x, a, b) = a + (0.49 - a) \exp(-b(x - 8))$. R-coding is found in Appendix A.6.

In both examples, a and b are varied in order to minimize s; in other words the OT is being employed. Theoretically r can be any correlation coefficient, but for computational purposes the nlm routine in R works only on continuous functions in its minimization technique so only continuous correlation coefficients could be tried. Thus only r_{av} was employed for the randomization example; r_{av} and Pces were used on the D/S example.

The Randomization Example

Many simulations were run, but only one result is given; the sample size is 45, $a = 1$ and $b = 0.5$. The graphs show the two most basic concepts: first, the ordered residuals plotted against the ordered response variable with a regression line, Figure 1, i.e. results from equation (6) and second, the actual fit, Figure 2, with estimated values of 1.012 for a and 0.495 for b. Any curve fitting method is good only when there are sufficient data points throughout the essential range of the model; this was certainly observed in these simulations. With this understanding of having adequate data, very good fits were obtained as illustrated in Figures 1 and 2, showing again the viability of the CES and the usefulness of r_{av}.

The Draper/Smith Example

In this section the chemical industry example of available chlorine at the time of manufacture from D/S illustrates nonlinear fitting. D/S as well as other practitioners show various sophisticated methods for dealing with the problem of non-linear curve fitting. The procedure indicated here gives a simple alternative

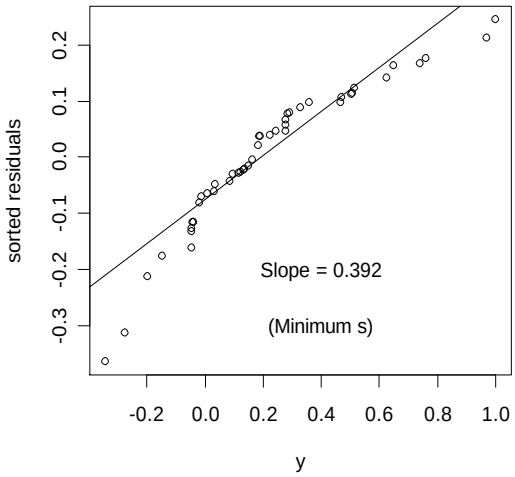

Figure 1: Exponential inference

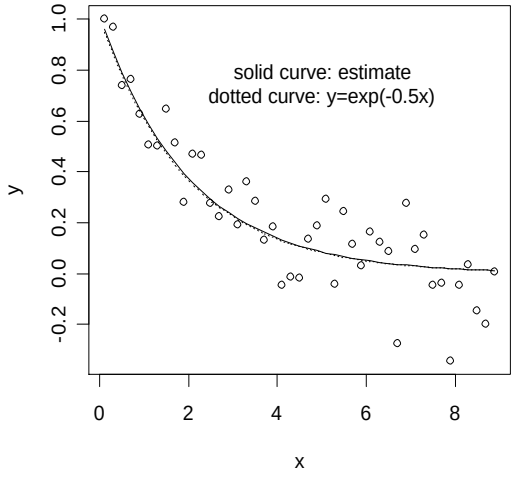

Figure 2: Exponential curve fit

way to get a feasible fit. When the Pces correlation coefficient was used (usually meaning results close to least squares), the methods of this paper gave essentially the same result that D/S obtained, as desired. The fit from D/S was very good, so CES passed its "stress test". After 12 iterations, the convergence criteria were satisfied giving final estimates of $a = 0.392$ and $b = 0.103$ for Pces and 0.391, 0.107 for r_{av}. For comparison, the D/S results were $a = 0.39$, and $b = 0.102$. Because the fit was so close no additional figures are shown.

8. Correlation Coefficient Estimation of the Parameters of Univariate Distributions

This section shows the generality of the correlation coefficient method by adjusting equation (6) for use in the estimation of the parameters of univariate distributions. The response variable is replaced by some form of the empirical distribution function F_n and the estimating function F is the theoretical cumulative distribution function. The parameter β of the distribution function F is varied to find the minimum s. In addition, here the residuals en masse are minimized relative to the edf $F_n(x)$, so $F_n(x)$ appears in place of the earlier y. The $F_n(x)$ needs no superscript, of course, as it is intrinsically ordered. R-coding is found in Appendix A.7. The adjusted equation is

$$r(F_n(x), (F_n(x) - F(x, \widehat{\beta}))^0 - sF_n(x)) = 0. \tag{7}$$

It has been shown that the solutions to equation (7) behave reasonably with respect to location and scale changes when a distribution that can be standardized, such as the normal, is used. For such distributions, however, the parameter estimation technique related to equation (2) is an alternative. A paper on this idea is available in Gideon and Rothan (2009) and is also on the website. However, for distributions like the gamma, the proposed method of equation (7) is appropriate. One example is given for the gamma distribution with 25 randomly generated observations with parameters scale = 2 and shape = 3. The estimates were 1.19 for scale and 4.16 for shape. The results are summarized in Figures 3 and 4. Figure 3 is a geometrical picture of the fitting process using equation (7) while Figure 4 plots x versus $F_n(x)$, $F(x)$, and the estimated $F(x)$. It is probably worth noting that nlm has some trouble staying in the appropriate solution space when working on certain non-linear problems. It seems that choosing a suitable starting value is critical. The example presented gave a good fit immediately.

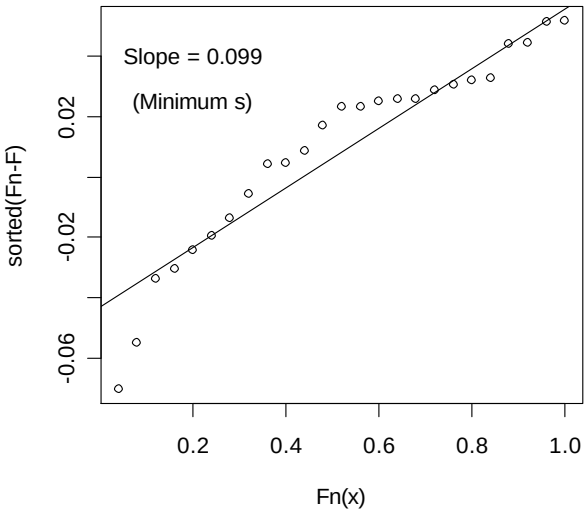

Figure 3: Gamma inference

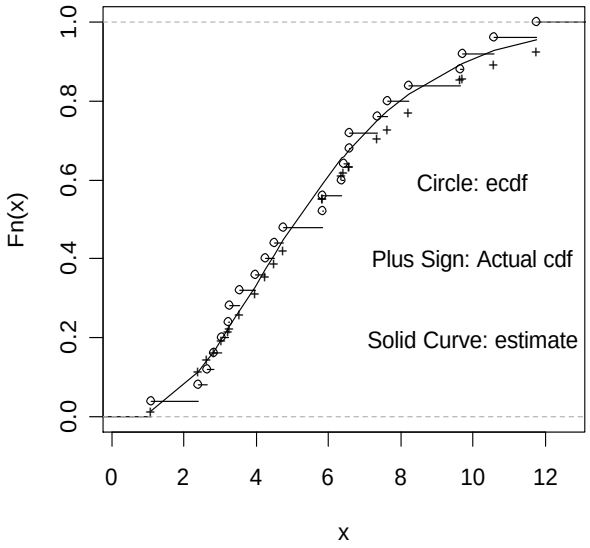

Figure 4: Gamma estimation

9. Conclusion

The CES provides a very general method to estimate parameters in a number of different settings and with different estimation criteria. The CES has a multitude of possibilities; many have presented themselves just in putting together this paper. Certainly further study needs to be undertaken, but the area is so broad that definitive study by a single person is virtually impossible. A profitable study also needs better computational ability in R for the implicit equations of CES. This idea is discussed in Appendix A.8.

It is apparent that CES with just Pces rivals least squares, but the method can be used with all correlation coefficients (both continuous and nonparametric) yielding a unified general estimation system applicable in many diverse areas. One of the most appealing features of CES is its coherence, in that all CCs - even robust ones - are treated the same, which makes it particularly easy to understand and apply. Additionally, when a robust CC is used, the system obviates the need for identifying outliers, which is notoriously difficult in the multivariate case. Over the years GDCC, which displays robustness, was successfully incorporated into many areas of estimation, such as time series, general linear models, and of course nonlinear regression and estimation of parameters for a particular density type. See Sheng (2002) for a general discussion of these areas. This leads one to believe that any correlation coefficient could be similarly profitably employed as shown in this paper by r_{av} and the results in Table 1. All the necessary machinery involving R and the estimations of this paper are included so that

anyone could reproduce this work. Further asymptotic inference on the RET for multiple regression is given in the papers Gideon (2010) and Gideon, Prentice, and Pyke (1989).

Appendix: R-program Outline

A.1 Definitions of max-min procedure and five CCs

```
# min-max procedure, fxyrk, produces two unique sets of ranks that allow the computation of the most positive and most negative correlations. w is used in all NPCCs below.
fxyrk = function(x, y) {n = length(x)
# most positive computation
xt = x[order(y, x)] # x order by y with y ties ordered by x
rky = 1:n
rky1 = rky[order(xt, rky)] # ranks of y ordered by x
# most negative computation
xrr = n+1−rank(x) # reverse ranks on the x
xt = x[order(y, xrr)] # x ordered by y with y ties ordered by rev(x)
rky2 = order(xt, n:1) # ranks of y ordered by x with y ties ordered by rev(y)
w = matrix(c(rky1, rky2), n, 2, byrow = FALSE)}
# FIRST GDCC
GDave = function(x, y) {w = fxyrk(x, y)
n = length(w[, 1]); n1 = n−1; k = 1; cc = NULL
if(sum(abs(w[, 1]−w[, 2])) == 0) nave = 1 else nave = 2
while(k <= nave) {rky = w[, k]
ryr = n+1−rky
dy = NULL; dyn = NULL
for (i in 1:n1) {
dy = c(dy, sum(rky[1:i]−i>0))
dyn = c(dyn, sum(ryr[1:i]−i>0))}
mdyr = max(dyn)
mdy = max(dy)
cc[k] = (mdyr−mdy)/(n%/%2)
cc[2] = cc[k]
k = k+1}
GDcor = (cc[1]+cc[2])/2
GDcor}
# SECOND Kendall's Tau
KENtau = function(x, y) {w = fxyrk(x, y)
n = length(x); n1 = n−1
```

```
rky = w[, 1]; rky2 = w[, 2]
dy = 0; dy2 = 0
for(i in 1:n1) {i1 = i+1
dy = dy+sum(rky[i]<rky[i1:n])
dy2 = dy2+sum(rky2[i]<rky2[i1:n])}
KT = ((dy+dy2)/choose(n, 2))−1
KT}
# THIRD Gini or Modified Footrule
Gini = function(x, y) {w = fxyrk(x, y)
n = length(x); ident = 1:n
rky = w[, 1]; rky2 = w[, 2]
dpnc = sum(abs(n+1−rky−ident))
dppc = sum(abs(rky−ident))
dpnc2 = sum(abs(n+1−rky2−ident))
dppc2 = sum(abs(rky2−ident))
den = n^2%/%2
Gcor = ((dpnc+dpnc2)−(dppc+dppc2))/(den*2)
Gcor}
# FOURTH Absolute Value CC, the continuous version of Gini
abscor = function(x, y) {
ym = mean(y); xm = mean(x)
SAx = sum(abs(x−xm)); SAy = sum(abs(y−ym))
dpnc = sum(abs((x−xm)/SAx+(y−ym)/SAy))
dppc = sum(abs((x−xm)/SAx−(y−ym)/SAy))
rav = (dpnc−dppc)/2
rav}
# FIFTH MAD CC
MADcor = function(x, y) {xm = median(x)
ym = median(y)
madx = median(abs(x−xm))
mady = median(abs(y−ym))
dpnc = median(abs((x−xm)/madx+(y−ym)/mady))
dppc = median(abs((x−xm)/madx−(y−ym)/mady))
rmad = (dpnc−dppc)/2
rmad}
# the population inverses for elliptical contoured populations
GinI = function(y) {tan(pi*y/4)*sqrt(1+2*cos(pi*y/2))} # Gini inv
GDI = function(y) sin(pi*y/2) # GD or Kendall inverse
MADI = function(y) y*sqrt(2−y^2) # MAD or AV (abs) CC inverse
# setting up functions for CES regression
```

```
GDslp = function(b, x, y) {GDave(x, y−b*x)}
GDslp2 = uniroot(GDslp, c(−10, 10), x = x, y = y)$root
Ginslp = function(b, x, y) {Gini(x, y−b*x)}
Ginslp2 = uniroot(Ginslp, c(−2, +2), x = x, y = y)$root
Kenslp = function(b, x, y) {KENtau(x, y−b*x)}
Kenslp2 = uniroot(Kenslp, c(−2, +2), x = x, y = y)$root
abcslp = function(b, x, y) {abscor(x, y−b*x)}
abcslp2 = uniroot(abcslp, c(−2, +2), x = x, y = y)$root
madslp = function(b, x, y) {MADcor(x, y−b*x)}
bslp = uniroot(madslp, c(−5, 7.0), x = x, y = y)$root
```

A.2 Simple linear regression using uniroot

```
Let f = function(x, y) { ... }
# in curly brackets choose one of the CCs from A.1 on data (x, y)
fslp = function(b, x, y) f(x, y−b*x) # solve for b in this function
slp = uniroot(fslp, c(l, u), x = x1, y = y1)$root # the slope of the regression
int = median(y1−slp*x1) # the intercept of the regression
```

A.3 Estimate of scale or error of the regression, also using uniroot

```
# The estimate of (1) σ_res/σ_y and (2) σ_res/σ_x as entities.
res = y1−(int+slp*x1); (1) y1s = sort(y1) or (2) yls = sort(x1)
# Only case (1) is shown in the next line.
s = uniroot(fslp, c(l, u), x = y1s, y = sort(res))$root
# s is the slope of the regression on ordered data. For case (1), sqrt(1 − s^2) estimates
the regression correlation coefficient, and for case (2), s is used in the estimate of
the variation of the slope parameter.
```

A.4 The minimum SD program using optimize (selects *b* to minimize *s* for a simple linear regression)

```
ftest = function(b, x, y) {y3 = sort(y−b*x)
      s = uniroot(fslp, c(−1, 2), x = sort(y), y = y3)$root
      return(s)}
out1 = optimize(ftest, c(0, 1), x = x1, y = y1)
```

A.5 Multiple linear regression, using uniroot and nlm

```
# Let y1 be the response data, and XM the n×k matrix of regressor variable
data where there are k variables and the sample size is n. Again let f and fslp be
```

as above in A.1 and let b be the notation for the vector of regression coefficients not including intercept.

Optimization Technique (*OT*) *using nlm*

```
# define a function g of the regression coefficients to be used with R-routine nlm.
g = function(b) {
    s = uniroot(fslp, c(l, u), x = sort(y1), y = sort(y1−XM%*%b))$root
    return(s)}
# Note: b was 6 dimensional in the simulations and 2 and 6 in the C/P analysis
# The output for the multiple regression is obtained by
out = nlm(g, initialb)
int = median(y1−XM%*%b)
```

Regression Equation Technique (*RET*) *using uniroot and Gauss-Seidel*

```
# can use least squares method to compute an initial b value
while (de>0.005 & ct<250 & ctcor>0.01) {bp = b
     for(i in 1:k) {XMS = XM[, −i]
         bs = b[−i]
          ys = y1−XMS%*%bs
          b[i] = uniroot(fslp, c(bl[i], bu[i]), x = XM[, i], y = ys)$root
          }
          de = sum(abs(bp−b)) # the total change in the coefficients
          ct = ct+1 # a counter that is initially zero
          yres = y1−XM%*%b # the updated resdiduals
          for(i in 1:k) bcor[i] = f(XM[, i], yres)
          ctcor = sum(abs(bcor))}
int = median(y1−XM%*%b) # the intercept of the fit
yhat = int+XM%*%b # the predicted values of the model
```

The fitted model estimates are in b and int. Generally the regression equations (3) (as all numerical calculations) are only solvable to within some tolerance. The convergence measures used here are (1) ct, upper bound on total number of iterations, (2) de, the smallness of the sum total of the absolute value of the changes in the slopes, and (3) ctcor, the smallness of the sums of the absolute values of the correlations of the regressor variables with the residuals. The necessity of each of these has been observed; there may be some overarching convergence measure that is yet to be found.

A.6 Nonlinear estimation using uniroot and nlm

Let the data be in x and y and define a function, g2, for the estimation.

```
ysort = sort(y)
g2 = function(b) {
     s = uniroot(fslp, c(0, 1), x = ysort, y = sort(y−b[1]*exp(−b[2]*x)))$root
     return(s)}
out = nlm(g2, c(μ, η), steptol = 0.001)
# μ and η are the initial values of b[1] and b[2]
```

A.7 Univariate distribution estimation of parameters using uniroot and nlm

```
fn = ecdf(x) # the empirical distribution function of data x
g1 = function(b) {
     s = uniroot(fslp, c(0, 1), x = fn(x), y = sort(fn(x)−F(x, b[1], b[2])))$root
     return(s)} # F is the theoretical d.f. under consideration assum-
     ing two parameters
out = nlm(g1, c(μ, η)) # μ and η are the initial values of b[1] and b[2]
plot(fn); lines(x, F(x, b[1], b[2]), type = "l")
# Plot the outcome of the minimization, i.e. sorted residuals
(fn-F)^0, versus fn. The slope of the fit is the minimum s.
yres = sort(fn(x)−F(x, b[1], b[2]))
ss = out$minimum
int = median(yres−ss*fn(x))
plot(fn(x), yres); abline(int, ss) # the final iteration plot
```

A.8 Suggestions for broadening the functionality of R

It is apparent that the R-routine nlm needs to be fine-tuned (or a new routine created) for solving implicit equations involving non-linear functions, such as most distribution functions. The current form does not allow the CES method of estimation to work flawlessly when a location parameter is part of the minimization of equation (6). In running many simulations it was clear that a simple shift in location would have given the minimization technique a better solution. A work around is to include a constraint that allows the zero on the vertical axis of the residual plot to be centered within the residuals. Observe that this is the case for Figures 1 and 3. Also there were problems with nlm keeping the iterated values of the parameters within a feasible solution space; it is very sensitive to initial values. No problems seem to occur with the R-routines and the fitting of linear models when no location parameters were involved in the minimization.

A second improvement would be for the nlm to generalize its technique so that nonparametric correlation coefficients can be included as estimators. Estimation with GDCC was run for many years with a numerical system using a C program

that obtained the centered point of a solution interval with never a problem of convergence. So the preferable nlm would also include centered solution points. This most likely would allow convergence of the MAD method as used in Table 2.

Acknowledgements

Special thanks to the numerous students who have participated in the CES development at the University of Montana, especially Ph.D. students Bob Hollister, Steve Rummel, Adele Rothan, and HuaiQing Sheng and to editorial assistant Carol Ulsafer. Many thanks also to the referee. The clarity and thoroughness of the paper were greatly improved by his or her suggestions and perseverance.

References

Chatterjee, S. and Price, B. (1991). *Regression Analysis by Example*, 2nd edition. Wiley & Sons, New York.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edition. Wiley & Sons, New York.

Gideon, R. A. (2007). The correlation coefficients. *Journal of Modern Applied Statistical Methods* **6**, 517-529.

Gideon, R. A. (2010). Using correlation coefficients to estimate slopes in multiple linear regression. *Sankhya* **72-B**, 96-106.

Gideon, R. A. and Hollister, R. A. (1987). A rank correlation coefficient resistant to outliers. *Journal of the American Statistical Association* **82**, 656-666.

Gideon, R. A., Prentice, M. J. and Pyke, R. (1989). The limiting distribution of the rank correlation coefficient r_{gd}. In *Contributions to Probability and Statistics*: *Essays in Honor of Ingram Olki* (Edited by L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson), 217-226. Springer, New York.

Gideon, R. A. and Rothan, A. M., CSJ (2009). Location and scale estimation with correlation coefficients. *Communications in Statistics - Theory and Methods* **40**, 1561-1572.

Gideon, R. A. and Rummel, S. E. (1992). Correlation in simple linear regression. University of Montana, Department of Mathematical Sciences. http://www.math.umt.edu/gideon/CORR-N-SPACE-REG.pdf.

Gini, C. (1914). *L'Ammontare e la Composizione della Ricchezza delle Nazioni.* Fratelli Bocca, Torino.

Healy, M. J. R. (1986). *Matrices for Statistics.* Oxford University Press, New York.

Kendall, M. G. and Gibbons, J. D. (1990). *Rank Correlation Methods*, 5th edition. Oxford University Press, New York.

Pearson, K. (1911). On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika* **8**, 250-254.

Rummel, S. E. (1991). *A Procedure for Obtaining a Robust Regression Employing the Greatest Deviation Correlation Coefficient.* Ph.D. Dissertation, University of Montana, Missoula, Montana.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* **63**, 1379-1389.

Sheng, HuaiQing (2002). *Estimation in Generalized Linear Models and Time Series Models with Nonparmetric Correlation Coefficients.* Ph.D. Dissertation, University of Montana, Missoula, Montana. http://wwwlib.umi.com/dissertations/fullcit/3041406.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72-101.

Received April 14, 2011; accepted May 9, 2012.

Rudy A. Gideon
Emeritus
Department of Mathematical Sciences
University of Montana
Missoula, MT 59812, USA
gideon@mso.umt.edu