

11 Multiple Linear Regression

Multiple linear regression (MLR) is a method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes also called the predictand, and the independent variables the predictors. MLR is based on least squares: the model is fit such that the sum-of-squares of differences of observed and predicted values is minimized. MLR is probably the most widely used method in dendroclimatology for developing models to reconstruct climate variables from tree-ring series. Typically, a climatic variable is defined as the predictand and tree-ring variables from one or more sites are defined as predictors. The model is fit to a period – the *calibration period* – for which climatic and tree-ring data overlap. In the process of fitting, or estimating, the model, statistics are computed that summarize the *accuracy* of the regression model for the calibration period. The performance of the model on data not used to fit the model is usually checked in some way by a process called *validation*. Finally, tree-ring data from before the calibration period are substituted into the prediction equation to get a *reconstruction* of the predictand. The reconstruction is a “prediction” in the sense that the regression model is applied to generate estimates of the predictand variable outside the period used to fit the data. The uncertainty in the reconstruction is summarized by *confidence intervals*, which can be computed by various alternative ways.

Regression has long been used in dendroclimatology for reconstructing climate variables from tree rings. A few examples of dendroclimatic studies using linear regression are reconstruction of annual precipitation in the Pacific Northwest (Graumlich 1987), reconstruction of runoff of the White River, Arkansas (Cleaveland and Stahle 1989), reconstruction of an index of the El Niño Southern Oscillation (Michaelsen 1989), and reconstruction of a drought index for Iowa (Cleaveland and Duvick 1992).

MLR is not strictly a “time series” method. The most important point in application to time series is that observations are typically not independent of one another. As a consequence, special attention must be paid to a regression assumption about the independence of the residuals.

The predictors in any regression problem might be intercorrelated. This so-called *multicollinearity* does not preclude the use of regression, but can make it impossible or difficult to assess the relative importance of individual predictors from the estimated coefficients of the regression equation.

It is ironic that the most interesting periods of dendroclimatic reconstructions derived from regression are often the periods for which application of the regression model is most problematical -- periods whose climatic anomalies are most unlike those of today. The reconstruction for those periods is likely to be more uncertain than implied by regression statistics because the predictors are in a part of the “multivariate predictor space” not sampled by the data used to fit the model. The statistical aspects of this problem can be addressed by distinguishing predictions as *extrapolations*, as opposed to *interpolations*.

The MLR model is reviewed below, with emphasis on topics of particular interest for time series. More detailed information can be found in many standard references – for example, a statistical text on regression (Weisberg 1985), a chapter on regression as applied to the atmospheric sciences (Wilks 1995) and a monograph on regression in a time series context (Ostrom 1990).

11.1 Model

Model equation. The model expresses the value of a predictand variable as a linear function of one or more predictor variables and an error term:

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_K x_{i,K} + e_i$$

$x_{i,k}$ = value of k^{th} predictor in year i
 b_0 = regression constant
 b_k = coefficient on the k^{th} predictor
 K = total number of predictors
 y_i = predictand in year i
 e_i = error term

(1)

Prediction equation. The model (1) is estimated by least squares, which yields parameter estimates such that the sum of squares of errors is minimized. The resulting prediction equation is

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i,1} + \hat{b}_2 x_{i,2} + \dots + \hat{b}_K x_{i,K}$$
(2)

where the variables are defined as in (1) except that “^” denotes estimated values

Residuals. The error term in (1) is unknown because the true model is unknown. Once the model has been estimated, the regression residuals are defined as

$$\hat{e}_i = y_i - \hat{y}_i$$

y_i = observed value of predictand in year i
 \hat{y}_i = predicted value of predictand in year i

(3)

The residuals measure the closeness of fit of the predicted values and actual predictand in the calibration period. The algorithm for estimating the regression equation (solution of the normal equations) guarantees that the residuals have a mean of zero for the calibration period. The variance of the residuals measures the “size” of the error, and is small if the model fits the data well.

11.2 Assumptions

The MLR model is based on several assumptions. Provided the assumptions are satisfied, the regression estimators are optimal in the sense that they are *unbiased*, *efficient*, and *consistent*. Unbiased means that the expected value of the estimator is equal to the true value of the parameter. Efficient means that the estimator has a smaller variance than any other estimator. Consistent means that the bias and variance of the estimator approach zero as the sample size approaches infinity. Ostrom (1990, p. 14) lists six basic assumptions for the regression model:

1. **Linearity:** the relationship between the predictand and the predictors is linear. The MLR model applies to **linear** relationships. If relationships are nonlinear, there are two recourses: (1) transform the data to make the relationships linear, or (2) use an alternative statistical

model (e.g., neural networks, binary classification trees). Scatterplots should be checked as an exploratory step in regression to identify possible departures from linearity.

2. **Nonstochastic X:** $E(e_i X_{i,k}) = 0$. The errors are uncorrelated with the individual predictors. This assumption is checked in residuals analysis with scatterplots of the residuals against individual predictors. Violation of the assumption might suggest a transformation of the predictors.
3. **Zero mean:** $E[e_i] = 0$. The expected value of the residuals is zero. This is not a problem because the least squares method of estimating regression equations guarantees that the mean is zero.
4. **Constant variance:** $E[e_i^2] = \sigma^2$. The variance of the residuals is constant. In time series applications, a violation of this assumption is indicated by some organized pattern of dependence of the residuals on time. An example of violation is a pattern of residuals whose scatter (variance) increases over time. Another aspect of this assumption is that the error variance should not change systematically with the size of the predicted values. For example, the variance of errors should not be greater when the predictand is large than when the predictand is small.
5. **Nonautoregression:** $E[e_i e_{i-m}] = 0, m \neq 0$. The residuals are random, or uncorrelated in time. This assumption is one most likely to be violated in time series applications. Several methods of checking the assumption are covered later.
6. **Normality:** the error term is normally distributed. This assumption must be satisfied for conventional tests of significance of coefficients and other statistics of the regression equation to be valid. It is also possible to make no explicit assumption about the form of the distribution and to appeal instead to the Central Limit Theorem to justify the use of such tests. The normality assumption is the least crucial of the regression assumptions.

11.3 Statistics

Sum-of-squares terms. Several regression statistics are computed as functions of the sums-of-squares terms:

$$\begin{aligned}
 \text{SSE} &= \sum_{i=1}^n \hat{e}_i^2 && \text{sum of squares, error} \\
 \text{SST} &= \sum_{i=1}^n (y_i - \bar{y})^2 && \text{sum of squares, total} \\
 \text{SSR} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 && \text{sum of squares, regression}
 \end{aligned}
 \tag{4}$$

$n = \text{sample size (number of observations in calibration period)}$

Partitioning of variation. The regression equation is estimated such that the total sum-of-squares can be partitioned into components due to regression and residuals:

$$\text{SST} = \text{SSR} + \text{SSE} \tag{5}$$

Coefficient of determination. The explanatory power of the regression is summarized by its “R-squared” value, computed from the sums-of-squares terms as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (6)$$

R^2 , also called the *coefficient of determination*, is often described as the proportion of variance “accounted for”, “explained”, or “described” by regression. It is important to keep in mind that a high R^2 does not imply causation. The relative sizes of the sums-of-squares terms indicate how “good” the regression is in terms of fitting the calibration data. If the regression is “perfect”, all residuals are zero, SSE is zero, and R^2 is 1. If the regression is a total failure, the sum-of-squares of residuals equals the total sum-of-squares, **no** variance is accounted for by regression, and R^2 is zero.

ANOVA table and definition of “mean squared” terms. The sums-of-squares terms and related statistics are often summarized in an Analysis of Variance (ANOVA) table:

Source	df	SS	MS
Total	$n - 1$	SST	$MST = SST/(n-1)$
Regression	K	SSR	$MSR = SSR/ K$
Residual	$n - K - 1$	SSE	$MSE = SSE/(n - K - 1)$

Source=source of variation
 SS=sum-of-squares term
 df =degrees of freedom for SS term
 MS=“mean squared” terms

The mean squared terms are the sums-of-squares terms divided by the degrees of freedom.

Standard error of the estimate. The residual mean square (MSE) is the sample estimate of the variance of the regression residuals. The population value of the error term is sometimes written as σ_e^2 while the sample estimate is given by

$$s_e^2 = MSE \quad (7)$$

where MSE has been defined previously. The square root of the residual mean square is called the *root-mean-square error* (RMSE), or the *standard error of the estimate*

$$s_e = \sqrt{s_e^2} = \sqrt{MSE} = RMSE_c \quad (8)$$

The subscript “c” is attached ($RMSE_c$) in (8) to distinguish the RMSE derived from calibration from the root-mean-square error derived by cross-validation (see later).

F ratio, or “overall F”. Recall that the explanatory power of a regression is given by the regression R^2 , which is computed from sums-of-squares terms. The *F*-ratio, or *overall F*, which is computed from the mean squared terms in the ANOVA table, estimates the statistical significance of the regression equation. The *F*-ratio is given by

$$F = \frac{MSR}{MSE} \quad (9)$$

The advantage of the F - ratio over R^2 is that the F - ratio takes into account the degrees of freedom, which depend on the sample size and the number of predictors in the model. A model can have a high R^2 and still not be statistically significant if the sample size is not large compared with the number of predictors in the model. The F - ratio incorporates sample size and number of predictors in an assessment of significance of the relationship.

The significance of the F - ratio is obtained by referring to a table of the F distribution, using degrees of freedom {df1,df2}, where df1 and df2 are the degrees of freedom for the regression mean square and residual mean square from the ANOVA table.

Adjusted R^2 . The R^2 value for a regression can be made arbitrarily high simply by including more and more predictors in the model. The *adjusted R^2* is one of several statistics that attempts to compensate for this artificial increase in accuracy. The *adjusted R^2* is given by

$$\bar{R}^2 = 1 - \frac{\text{MSE}}{\text{MST}} \quad (10)$$

where MSE and MST are the mean squared terms previously defined in the ANOVA table. Referring to the ANOVA table shows that ratio of mean squared terms is related to the ratio of sum-of-squares terms by

$$\frac{\text{MSE}}{\text{MST}} = \frac{(n-1)\text{SSE}}{(n-K-1)\text{SST}} \quad (11)$$

where n is the number of observations, and K is the number of predictors. Because

$\frac{(n-1)}{(n-K-1)}$ must be greater than zero, it can immediately be seen that adjusted R^2 must be smaller than R^2 , and that the difference in the two statistics depends on both the sample size and the number of predictors in the model.

Confidence interval for estimated coefficients. If the regression assumptions on the residuals are satisfied, including the normality assumption, then the sampling distribution of an estimated regression coefficient is normal with a variance proportional to the residual mean square (MSE). The variance of the estimator also depends on the variances and covariances of the predictors. The idea is best illustrated for the case of simple linear regression, for which the variance of the regression coefficient is given by

$$\text{var}(\hat{b}_1) = \frac{s_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (12)$$

where s_e^2 is the residual mean square, x_i is the value of the predictor in year i , \bar{x} is the mean of the predictor, and the summation is over the n years in the calibration period. The $100(1-\alpha)\%$ confidence interval is $\hat{b}_1 \pm t_{\alpha/2} \sqrt{\text{var}(\hat{b}_1)}$, where $t_{\alpha/2}$ is obtained from a t distribution with $n-2$ degrees of freedom.

For example, if the sample size is $n = 45$ years, the degrees of freedom is 43. If the 95% confidence interval is desired, the appropriate α -level is $\alpha = 0.05$. A t -table for this sample size and α -level gives $t_{0.025,43} = 2.02$. The corresponding confidence interval is

$$\hat{b}_1 \pm 2.02 \sqrt{\text{var}(\hat{b}_1)} \quad (13)$$

To an approximation, the 95% confidence interval for an estimated regression parameter is two standard deviations around the estimate.

For more than one predictor, the confidence intervals for regression can be computed similarly, but the equation is more complicated. The equation for the variances and covariances of estimated coefficients is expressed in matrix terms by

$$\text{var}(\hat{\boldsymbol{\beta}}) = s_e^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (14)$$

where \mathbf{X} is the time series matrix of predictors. Equation (14) returns a matrix, with the variances of the parameters along the diagonal, and the covariances as the off-diagonal elements (Weisberg 1985, p. 44). The appropriate degrees of freedom of the t distribution is $df = n - K - 1$, where K is the number of predictors in the model, and n is the sample size.

11.4 Selecting predictors

General guidelines. The predictors for an MLR model are sometimes specified beforehand, and are sometimes selected by some automated procedure from a pool of *potential predictors*. Various schemes for automated variable screening are available. For example, the *forward stepwise* method enters additional predictors one by one depending on maximum reduction of the residual variance (variance not accounted for by predictors already in the model). The *best subsets* method tests all possible sets of 1, 2, 3, ... predictors and selects the set giving the best value of accuracy adjusted for loss of degrees of freedom as measured by any of several possible statistics.

It is generally a good idea to restrict the pool of potential predictors to variables with some plausible physical link to the predictand. For example, if the predictand is Tucson seasonal precipitation, tree-ring indices from the Santa Catalina Mountains are reasonable predictors, but tree-ring indices from Mongolia are not.

It is also important that the predictor pool not be made unnecessarily large – for example by including all sorts of transformations of the original variables in the pool. This is important because R^2 can be seriously biased when the pool includes a large number of potential predictors, even if only a few of those predictors are selected for the final model (Rencher and Pun 1980).

Lagged predictors and prewhitened predictors. *Lagged* regression models refer to models in which the relationship between the predictors and predictand is not constrained to be contemporaneous. In a *distributed lag* model, the predictors of y_t include $x_{t\pm m}$, where m might take on some value besides zero. In econometrics terminology, this particular model is a distributed lag model with lagged *exogenous* variables (variables “outside of” or not dependent on the model). In dendroclimatology, positive lags on the predictors in distributed-lag models are easy to rationalize: the climate in year t affects tree growth not just in year t , but in succeeding years; thus the tree-ring indices for years $t+m$ might inform on what the climate was in year t . The case for negative lags is less obvious, but plausible: the tree-ring in year t holds information on the climate in year t , but the information is clouded by the preconditioning of the ring in year t by climate or biology of earlier years; thus including negative lags on the tree-ring index lets the model compensate or remove the confounding effects of prior years’ climate on the current year’s ring.

The dendroclimatic strategy of *prewhitening* tree-ring indices as predictors in regression models to reconstruct climate is grounded in similar rationale to that for using negatively lagged predictors in distributed-lag models. With prewhitening, a time series model (say, an AR model)

is fit to the full length of tree-ring series and the model residuals are regarded as *residual indices*. These residual indices are then used as predictors in the climate reconstruction model. The rationale is that the current year's index is preconditioned to some extent by past conditions, including biology and possibly climate. The time-series modeling presumably adjusts for this distortion by removing the linear dependence of the tree-ring index on its past values.

11.5 Multicollinearity

The predictors in a regression model are often called the “independent variables”, but this term does not imply that the predictors are themselves independent statistically from one another. In fact, for natural systems, the predictors can be highly intercorrelated. “Multicollinearity” is a term reserved to describe the case when the intercorrelation of predictor variables is high. It has been noted that the variance of the estimated regression coefficients depends on the intercorrelation of predictors (equation (14)). Haan (2002) concisely summarizes the effects of multicollinearity on the regression model. Multicollinearity does not invalidate the regression model in the sense that the predictive value of the equation may still be good as long as the prediction are based on combinations of predictors within the same multivariate space used to calibrate the equation. But there are several negative effects of multicollinearity. First, the variance of the regression coefficients can be inflated so much that the individual coefficients are not statistically significant – even though the overall regression equation is strong and the predictive ability good. Second, the relative magnitudes and even the signs of the coefficients may defy interpretation. For example, the regression weight on a tree-ring index in a multivariate regression equation to predict precipitation might be negative even though the tree-ring index by itself is positively correlated with precipitation. Third, the values of the individual regression coefficients may change radically with the removal or addition of a predictor variable in the equation. In fact, the sign of the coefficient might even switch.

Signs of multicollinearity. Signs of multicollinearity include 1) high correlation between pairs of predictor variables, 2) regression coefficients whose signs or magnitudes do not make good physical sense, 3) statistically nonsignificant regression coefficients on important predictors, and 4) extreme sensitivity of sign or magnitude of regression coefficients to insertion or deletion of a predictor variable.

Variance Inflation Factor (VIF). The Variance Inflation Factor (VIF) is a statistic that can be used to identify multicollinearity in a matrix of predictor variables. “Variance Inflation” refers here to the mentioned effect of multicollinearity on the variance of estimated regression coefficients. Multicollinearity depends not just on the bivariate correlations between pairs of predictors, but on the multivariate predictability of any one predictor from the other predictors. Accordingly, the VIF is based on the multiple coefficient of determination in regression of each predictor in multivariate linear regression on all the other predictors:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (15)$$

where R_i^2 is the multiple coefficient of determination in a regression of the i th predictor on all other predictors, and VIF_i is the variance inflation factor associated with the i th predictor. Note that if the i th predictor is independent of the other predictors, the variance inflation factor is one, while if the i th predictor can be almost perfectly predicted from the other predictors, the variance inflation factor approaches infinity. In that case the variance of the estimated regression coefficients is unbounded.

Multicollinearity is said to be a problem when the variance inflation factors of one or more predictors becomes large. How large appears to be a subjective judgement. According to Haan (2002), some researchers use a VIF of 5 and others use a VIF of 10 as a critical threshold. These VIF values correspond, respectively, to R_i^2 values of 0.80 and 0.90. Some compute the average VIF for all predictors and declare that an average “considerably” larger than one indicates multicollinearity (Haan, 2002). At any rate, it is important to keep in mind that multicollinearity requires strong intercorrelation of predictors, not just non-zero intercorrelation. The VIF is closely related to a statistic call the tolerance, which is $1/\text{VIF}$. Some statistics packages report the VIF and some report the tolerance (Haan 2002).

11.6 Analysis of Residuals

Analysis of residuals consists of examining graphs and statistics of the regression residuals to check that model assumptions are satisfied. Some frequently used residuals tests are listed below.

Time series plot of residuals. The time series plot of residuals can indicate such problems as non-constant variance of residuals, and trend or autocorrelation in residuals. A time-dependent variance might show, say, as an increasing scatter of the residuals about the zero line with time. The slope of the scatter plot of residuals on time can be tested for significance to identify trend in residuals.

Scatterplot of residuals against predicted values.. The residuals are assumed to be uncorrelated with the predicted values of the predictand. Violation is indicated by some noticeable pattern of dependence in the scatterplots. For example, the residual might flare out (increased scatter) with increasing value of the predictand; the remedy might be a transformation (e.g., log transform) of the predictand.

Scatterplots of residuals against individual predictors. The residuals are assumed to be uncorrelated with the individual predictors. Violation of these assumptions would be indicated by some noticeable pattern of dependence in the scatterplots, and might suggest transformation of the predictors.

Histogram of residuals. The residuals are assumed to be normally distributed. Accordingly, the histogram of residuals should resemble a normal pdf. But keep in mind that a random **sample** from a normal distribution will be only approximately normal, and so the some departures from normality in the appearance of the histogram are expected – especially for small sample size.

Acf of residuals. The residuals are assumed to be nonautocorrelated. If the assumption is satisfied, the acf of residuals should not be large at any non-zero lag. Special interest should be attached to the lowest lags, since physical systems are characterized by persistence from year to year.

Lag-1 scatterplot of residuals. This plot also deals with the assumption of independence of residuals. The residuals at time t should be independent of the residuals at time $t-1$. The scatterplot should therefore resemble a formless cluster of points. Alignment in some direction might be evidence of autocorrelation of residuals at lag 1.

Durbin-Watson. The Durbin-Watson (D-W) statistic tests for autocorrelation of residuals, specifically lag-1 autocorrelation. The D-W statistic tests the null hypothesis of no first-order autocorrelation against the alternative hypothesis of positive first-order autocorrelation. The alternative hypothesis might also be negative first-order autocorrelation. Assume the residuals follow a first-order autoregressive process $e_t = pe_{t-1} + n_t$, where n_t is random and p is the first-order autocorrelation coefficient of the residuals. If the test is for positive autocorrelation of residuals, the hypotheses for the D-W test can be written

$$\begin{aligned} H_0 : p &= 0 \\ H_A : p &> 0 \end{aligned} \tag{16}$$

where p is the population value of the first-order autocorrelation coefficient of residuals. The D-W statistic is given by

$$d = \frac{\sum_{i=1}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2}$$

For positive autocorrelation, the residuals at successive times will tend to be similar, such that the numerator will be small relative to the denominator; in the limit, as the first-order autocorrelation approaches 1, the numerator goes to zero and the d goes to zero. It can be shown (Ostrom, 1990, p. 28), that if the residuals follow a first-order autoregressive process, d is related to the first-order autocorrelation coefficient, p , as follows

$$d = 2(1 - p) \tag{17}$$

The above equation implies that

- $d = 2$ if no autocorrelation ($p = 0$)
- $d = 0$ if the first-order autocorrelation is 1.0
- $d = 4$ if the first-order autocorrelation is -1.0

Durbin and Watson established upper (d_u) and lower (d_l) limits for the significance levels of a computed d . These limits are available in tables in many statistics texts, and are stored in a user-written lookup table in MATLAB. The application of the D-W tests is to compute d from the regression residuals, choose a significance level (e.g., 99%), look up the upper and lower limits from the table, use a decision rule depending on the alternative hypothesis. For positive autocorrelation, the decision rule is

if $d < d_l$ reject H_0
 if $d > d_u$ do not reject H_0
 if $d_l \leq d \leq d_u$ inconclusive

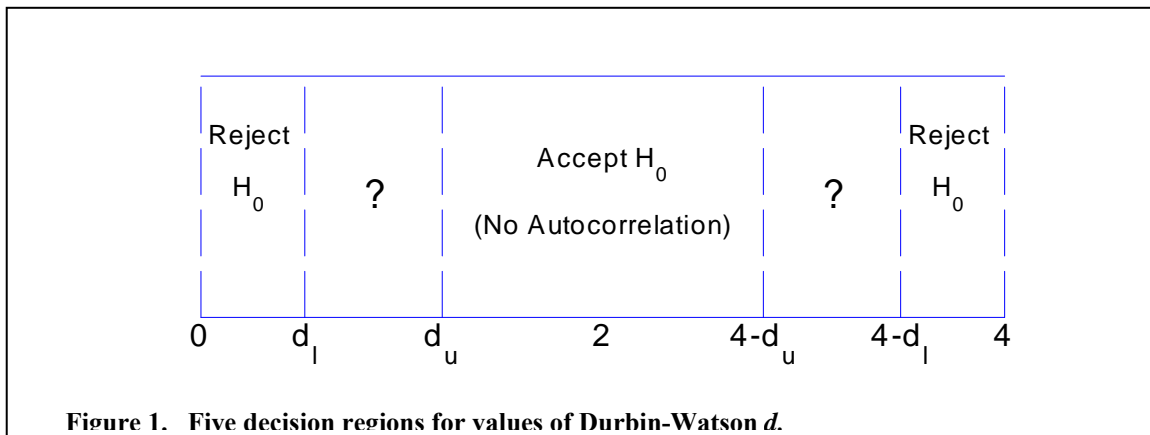


Figure 1. Five decision regions for values of Durbin-Watson d .

For negative autocorrelation, the decision rule is

if $d > 4 - d_l$	reject H_0
if $d < 4 - d_u$	do not reject H_0
if $4 - d_u \leq d \leq 4 - d_l$	inconclusive

According to Ostrom (1990, p. 29), some researchers reject the D-W statistic in favor of a simple rule of thumb for how much autocorrelation of residuals is tolerable in regression. This rule of thumb is that an alternative method to regression should be used if the first-order autocorrelation of residuals is greater than 0.30.

Portmanteau test. The portmanteau statistic, or Q statistic, is designed to test whether the regression residuals are purely random, or white noise (Ostrom 1990, p. 50). Unlike the D-W test, the portmanteau test does not restrict the possible form of autocorrelation to first-order autoregressive. The null hypothesis for the test is that the residuals are completely random; the alternative hypothesis is that the residuals are generated by an autoregressive or moving average model of some order. If the residuals are random, the acf of residuals should be zero at all nonzero lags. The Q statistic is computed as

$$Q = N \sum_{k=1}^K r_k^2 \quad (18)$$

where r_k is the lag-k autocorrelation coefficient of the regression residuals, N is the length of the time series of residuals, and K is chosen as the maximum anticipated order of autoregressive or moving-average process hypothesized under the alternative hypothesis to have generated the residuals. As a rule of thumb, K should be chosen as no larger than about $N/4$ or $N/5$, where N is the length of the time series.

If the null hypothesis is true, Q is distributed as chi-square with K degrees of freedom. Large acf coefficients lead to a high computed Q . A high Q therefore indicates significant autocorrelation and rejection of the null hypothesis. The p -value for Q is the probability of obtaining as high a Q as computed when the null hypothesis is true. The p -value for a computed Q can be obtained from a chi-square table. In summary, rejection of the null hypothesis is indicated by large acf coefficients and high computed Q . The more significant the computed Q , the lower its p -value.

References

- Cleaveland, M., and DuVick, D.N., 1992, Iowa climate reconstructed from tree rings, 1640-1982, *Water Resources Research* 28(10), 2607-2615.
- Cleaveland, M.K., and Stahle, D.W., 1989, Tree ring analysis of surplus and deficit runoff in the White River, Arkansas, *Water Resources Research* 25 (6), 1391-1401.
- Graumlich, L.J., 1987, Precipitation variation in the Pacific Northwest (1675-1975) as reconstructed from tree rings, *Annals of the Association of American Geographers* 77(1), 19-29.
- Haan C. T. (2002) *Statistical methods in Hydrology*, second edition. Iowa State University Press, Ames, Iowa.
- Michaelsen, J., 1989, Long-period fluctuations in El Nino amplitude and frequency reconstructed from tree-rings, *Geophysical Monograph* 55, American Geophysical Union, 69-74.
- Ostrom, C.W., Jr., 1990, *Time Series Analysis, Regression Techniques*, Second Edition: Quantitative Applications in the Social Sciences, v. 07-009: Newbury Park, Sage Publications.
- Rencher, A.C., and Pun, Fu Ceayong, 1980, Inflation of R^2 in best subset regression, *Technometrics* 22 (1), 49-53.

- Weisberg, S., 1985, *Applied Linear Regression*, 2nd ed., John Wiley, New York, 324 pp.
- Wilks, D.S., 1995, *Statistical methods in the atmospheric sciences*: Academic Press, 467 p.
- Woodhouse, C.A., 1999, Artificial neural networks and dendroclimatic reconstructions: An example from the Front Range, Colorado, USA: *The Holocene*, v. 9, no. 5, p. 521-529.