2015
OKSWP1506

Economics Working Paper Series
Department of Economics
OKLAHOMA STATE UNIVERSITY
http://spears.okstate.edu/ecls/

Collinearity Diagnostics in gretl

Lee C. Adkins
Oklahoma State University

Melissa S. Waters
Southern University

R. Carter Hill
Louisiana University

Department of Economics
Oklahoma State
University
Stillwater, Oklahoma

339 BUS, Stillwater, OK 74078, Ph 405-744-5110, Fax 405-744-5180

Collinearity Diagnostics in gretl

Lee C. Adkins, Melissa S. Waters, and R. Carter Hill

July 8, 2015

Abstract

Colinearity is blamed for all sorts of trouble in empirical work: inconclusive or weak results, unexpected signs on coefficients, and general computational mayhem in nonlinear estimators. Collinearity is a matter of degree since perfect collinearity has a perfectly easy solution. Near perfect collinearity can be vexing however since it makes precise measurement of model parameters particulary difficult in some cases.

A number of methods for detecting collinearity have been proposed. Some of these are useful, others not. Hill and Adkins (2001) summarize the good and bad based on much of the relevant literature up to 2001. They also make some recommendations for the detection and amelioration of inadequate variation in the data.

The purpose of our paper is twofold: 1) update any significant findings on collinearity since the Hill and Adkins (2001) survey and 2) to write and document gretl functions that perform several regression diagnostic procedures not already present in the software. These include the diagnostics suggested in Hill and Adkins (2001). In particular, we introduce hansl routines to perform the variance decomposition of Belsely, Kuh, and Welch (1980) for both linear and nonlinear models and provide a function to compute critical values for the Belsley (1982) signal-to-noise ratio test. The use of these is explored in several examples.

1 Anything New?

In terms of advancing the state or collinearity diagnostics has anything new happened? We'd have to say no, although Friendly and Kwan (2009) have suggested a clever method of

visualizing the relationships among the variables. Proposals for dealing with collinearity have been suggested, e.g. Hillebrand and Lee (2012), who shrink OLS to principal components using a Stein estimator and Erkoç et al. (2010), who applies ridge regression to collinearity in nonlinear models. We are still in the process of determining the best way to deal with weak data or weakly identified models. There are proposals to ameliorate the ill-effects of collinearity in linear and non-linear models, but it seems to us that the solutions (mostly biased estimation of parameters) proceed without specific knowledge of whether the problem is data driven, parameter driven, model driven, or a combination of the three. There are several reasons why parameter variances are large and collinearity is only one of the possible causes. The techniques in this paper do not solve these issues. At most, we extend the ways that Belsley et al. (1980) employs to analyze weak data in linear models to nonlinear models. Our experience is that the estimation of nonlinear models suffer from poor data as well as bad model specification, and our goal is to explore how the linear diagnostics can be used in a nonlinear setting. We can definitively say that we know nothing definitive.

In the following sections we review the diagnostics used in linear models and discuss their extension to nonlinear models. All of this is based on the Belsley et al. (1980); Belsley (1982) diagnostics based on condition numbers, variance decompositions, and tests of signal-to-noise in regression. By analogy, we extend these to nonlinear models and propose gretl functions to compute the diagnostics.

2 Linear Model

Denote the linear regression model as

$$y = X\beta + u$$

where y is a $n \times 1$ vector of observations on the dependent variable, X is a $n \times k$ non-stochastic matrix of observations on k explanatory variables, β is a $k \times 1$ vector of unknown parameters, and u is the $n \times 1$ vector of uncorrelated random errors, with zero means and constant variances,σ^2. In the general linear model exact, or perfect, collinearity exists when the columns of X, denoted x_i, $i = 1, \ldots, K$, are linearly dependent. That is, if there is at least one relation of the form $c_1x_1 + c_2x_2 + \cdots + c_Kx_k = 0$, where the c_i are constants, not all equal to zero. In this case the column rank of X is less than k, and the normal equations $X^TX\beta = X^Ty$ do not have a unique solution, and least squares estimation breaks down. Unique best linear unbiased estimators do not exist for all K parameters. However, even in this most severe of cases, all is not lost.

Exact collinearity is rare, and easily recognized. More frequently, one or more linear combinations of explanatory variables are nearly exact, so that $c_1x_1 + c_2x_2 + \cdots + c_Kx_k \approx 0$. We now examine the consequences of such near exact linear dependencies.

2.1 Diagnosing Collinearity using the Eigenvalues and Eigenvectors of X^TX

The $k \times k$ matrix X^TX is symmetric. For symmetric matrices their exists an orthonormal $k \times k$ matrix C such that

$$C^TX^TXC = \Lambda \tag{1}$$

where Λ is a diagonal matrix with the real values $\lambda_1, \lambda_2, \ldots, \lambda_k$ on the diagonal. An orthonormal matrix, sometimes also called an orthogonal matrix, has the property that $C^T = C^{-1}$, so that $C^TC = CC^T = I_k$, where I_k is a $k \times k$ identity matrix. The columns of the matrix C, denoted c_i, are the eigenvectors (or characteristic vectors) of the matrix, and the real values λ_i are the corresponding eigenvalues (or characteristic roots). It is customary to assume that the columns of C are arranged so that the eigenvalues are ordered by magnitude, $\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_k$.

Silvey (1969) popularized the use of eigenvalues to diagnose collinearity, and Belsley et al. (1980) [hereinafter BKW] refined, and improved, the analysis. The $n \times k$ matrix $Z = XC$ is called the matrix of principal components of X. The i^{th} column of Z, z_i, is called the i^{th} principal component. From equation (1) z_i has the property that $z_i^Tz_i = \lambda_i$. If the characteristic root $\lambda_i = 0$, then $z_i = Xc_i = 0$; we have an exact linear relation among the columns of X, and thus exact collinearity. If $\texttt{rank}(X) = \ell < k$, then we will find $k - \ell$ eigenvalues that are zero.

If X is of full column rank k, so that there are no exact linear dependencies among the columns of X, then is a positive definite and symmetric matrix, and all its eigenvalues are not only real but also positive. If we find a "small" eigenvalue, $\lambda_i \approx 0$, then

$$\lambda_i = z_i^Tz_i = c_i^TX^TXc_i \approx 0$$

and there is a near exact linear dependency among the columns of X. If there is a single small eigenvalue, then the linear relation indicates the form of the linear dependency, and which of the explanatory variables are involved in the relationship. If there are two (or more) small eigenvalues, then we have two (or more) near exact linear relations. Multiple linear relationships do not necessarily indicate the form of the linear dependencies. The eigenvectors associated with the near zero eigenvalues define a 2-dimensional vector space in which the two near exact linear dependencies exist. While we may not be able to identify the individual relationships among the explanatory variables that are causing the collinearity, we can identify the variables that appear in the two (or more) relations.

The singular-value decomposition of X is an alternative technique that achieves the same goals as the analysis of eigenvalues. For computational reasons there are reasons to prefer the singular-value decomposition, and the literature on collinearity is divided between the two approaches. The matrix X may be decomposed as $X = UDV^T$, where $U^TU = V^TV = I_k$ and

D is a diagonal matrix with non-negative diagonal values $\mu_1, \mu_2, \ldots, \mu_k$, called the singular values of X.

The relation to eigen analysis is that the singular values are the positive square roots of the eigenvalues of X^TX, and the matrix $V = C$. A small singular value implies a near exact linear dependence among the columns of X, just as does a small eigenvalue. We will ignore the computational issues and treat these two approaches as equivalent.

2.2 Collinearity and the Least Squares Estimator

Using equation (1) and the properties of the matrix of eigenvectors C, we can write $X^TX = X\Lambda C^T$, and therefore

$$(X^TX)^{-1} = C\Lambda^{-1}C^T = \sum_{i=1}^{k} \lambda_i^{-1} c_i c_i^T \tag{2}$$

defining $C = \{c_1, c_2, \ldots, c_k\}$ to be the matrix of characteristic vectors. The covariance matrix of the least squares estimator b is $cov(b) = \sigma^2(X^TX)^{-1}$, and using equation (2) the variance of b_j is

$$var(b_j) = \sigma^2 \left(\frac{c_{j1}^2}{\lambda_1} + \frac{c_{j2}^2}{\lambda_2} + \ldots + \frac{c_{jk}^2}{\lambda_k} \right) \tag{3}$$

The orthonormality of C implies that $\sum_{\ell=1}^{k} c_{j\ell}^2 = 1$, so variance of b_j depends upon three distinct factors. First, the magnitude of the error variance, σ^2; second, the magnitudes of the constants c_{jk}; and third, the magnitude of the eigenvalues, λ_ℓ. A small eigenvalue may cause a large variance for b_j if it is paired with a constant $c_{j\ell}$ that is not close to zero. The constants $c_{j\ell} = 0$ when x_j and x_ℓ, are orthogonal. This fact is an important one for it will allow us to determine which variables are not involved in collinear relationships.

Suppose β_j is a critical parameter in your model, and there is one small eigenvalue, $\lambda_k \approx 0$. If x_j is not involved in the corresponding linear dependency , then c_{jk} will be small, and the fact that will not adversely affect the precision of estimation of β_j. The presence of collinearity in the data does not automatically mean that "all is lost." If X^TX has one or more small eigenvalues, then you must think clearly about the objectives of your research, and determine if the collinearity reduces the precision of estimation of your key parameters by an unacceptable amount. This leads us to the next question, "What is a small eigenvalue?"

2.3 Variance Decomposition of Belsley et al. (1980)

A useful property of eigenvalues is that $tr(X^TX) = \sum_{i=1}^{k} \lambda_i$. This implies that the size of the eigenvalues is determined in part by the scaling of the data. Data matrices consisting of

Condition Index	Variance Proportions of OLS $var(b_1)$	$var(b_2)$	$\cdots$	$var(b_k)$
η_1	ϕ_{11}	ϕ_{12}	$\cdots$	ϕ_{1k}
η_1	ϕ_{21}	ϕ_{22}	$\cdots$	ϕ_{2k}
.				
.	$\cdots$	$\cdots$	$\cdots$	$\cdots$
.				
η_k	ϕ_{k1}	ϕ_{k2}		ϕ_{kk}

Table 1: Matrix of Variance Proportions

large numbers will have larger eigenvalues, in total, than data matrices with small numbers. To remove the effect of scaling Belsley et al. (1980)), whose collinearity diagnostic procedure is proposed here, suggest scaling the columns of X to unit length. This scaling is only for the purpose of diagnosing collinearity, not for model estimation or interpretation.

To diagnose collinearity, examine the proportion of the variance of each least squares coefficient contributed by each individual eigenvalue. Define $\phi_{jk} = c_{jk}^2/\lambda_k$, and let ϕ_j be the variance of b_j, apart from the error variance, σ^2.

$$\phi_j = \left(\frac{c_{j1}^2}{\lambda_1} + \frac{c_{j2}^2}{\lambda_2} + \ldots + \frac{c_{jk}^2}{\lambda_k} \right)$$

Then, the proportion of the variance of b_j associated with the k^{th} eigenvalue λ_k is $\frac{\phi_{jk}}{\phi_j}$. Note the reversal of the subscripts. This is convenient for tabling the variance proportions, which has a now standard format. The columns of the table correspond to the variances of individual least squares coefficients, and the sum of each column is one. The rows of this matrix correspond to the different eigenvalues, which have been scaled in a certain way. The "condition index" is the square root of the ratio of the largest eigenvalue, λ_1, to the ℓ^{th} largest, λ_ℓ, that is,

$$\eta_\ell = \left(\frac{\lambda_1}{\lambda_\ell} \right)^{\frac{1}{2}}.$$

The condition indices are ordered in magnitude, with $\eta_1 = 1$ and η_k being the largest, since its denominator is the smallest eigenvalue.

Table 1 summarizes much of what we can learn about collinearity in data. BKW carried out extensive simulations to determine how large condition indices affect the variances of the least squares estimators. Their diagnostic procedures, also summarized in Belsley (1991, Chapter 5), are these:

Step 1 Begin by identifying large condition indices. A small eigenvalue and a near exact linear dependency among the columns of X is associated with each large condition index. BKWs experiments lead them to the general guidelines that indices in the range

0-10 indicate weak near dependencies, 10-30 indicate moderately strong near dependencies, 30-100 is a large condition index, associated with a strong near dependency, and indices in excess of 100 are very strong. Thus when examining condition indexes values of 30 and higher should immediately attract attention.

Step 2 If there is a single large condition number : Examine the variance-decomposition proportions. If there is a single large condition number, indicating a single near dependency associated with one small eigenvalue, collinearity adversely affects estimation when two or more coefficients have 50% or more of their variance associated with the large condition index, in the last row of Table 1. The variables involved in the near dependency have coefficients with large variance proportions.

If there are two or more large condition numbers of relatively equal magnitude If there are $J \geq 2$ large and roughly equal condition numbers, then $X^T X$ has J eigenvalues that are near zero and J and there are J near exact linear dependencies among the columns of X. Since the J corresponding eigenvectors span the space containing the coefficients of the true linear dependence, the "50% rule" for identifying the variables involved in the near dependencies must be modified. In this case, sum the variance proportions for the coefficients across the J large condition number rows in Table 1. The variables involved in the (set of) near linear dependencies are identified by summed coefficient variance proportions of greater than 50%. The variance proportions in a single row do not identify specific linear dependencies, as they did when there was but one large condition number.

If there are two or more large condition numbers with one extremely large An extremely large condition index, arising from a very small eigenvalue, can "mask" the variables involved in other near exact linear dependencies. For example, if one condition index is 500 and another is 50, then there are two near exact linear dependencies among the columns of X. However, the variance decompositions associated with the condition number of 50 may not indicate that there are two or more variables involved in a relationship. Identify the variables involved in the set of near linear dependencies by summing the coefficient variance proportions in the last J rows of Table 1, and locating the sums greater than 50%.

Step 3 Perhaps the most important step in the diagnostic process is determining which coefficients are not affected by collinearity. If there is a single large condition number, coefficients with variance proportions less than 50% in the last row of Table 1 are not adversely affected by the collinear relationship in the data. If there are $J \geq 2$ large condition numbers, then sum the last J rows of variance proportions. Coefficients with summed variance proportions of less than 50% are not adversely affected by the collinear relationships. If the parameters of interest have coefficients unaffected by collinearity, then small eigenvalues and large condition numbers are not a problem.

Step 4 If key parameter estimates are adversely affected by collinearity, further diagnostic steps may be taken. If there is a single large condition index the variance proportions

identify the variables involved in the near dependency. If there are multiple large condition numbers, auxiliary regressions may be used to further study the nature of the relationships between the columns of X. In these regressions one variable in a near dependency is regressed upon the other variables in the identified set. The usual t-statistics may be used as diagnostic tools to determine which variables are involved in specific linear dependencies. See Belsley (1991, p. 144) for suggestions. Unfortunately, these auxiliary regressions may also be confounded by collinearity, and thus they may not be informative.

3 Signal-to-Noise Tests

Belsley (1982) considers a method for determining the presence of "weak data" using a test that considers the size of a coefficient relative to its variability, that is its signal-to-noise ratio (s/n). Combined with the condition number analysis and variance decomposition one can diagnose whether a regression suffers from collinearity and/or from "short data." Signal-to-noise is defined as

$$\tau_k \equiv \beta_k / \sigma_{b_k} \tag{4}$$

where β_k is the parameter value of the k^{th} coefficient in the model and σ_{b_k} is β_k's estimator's standard error; both are population parameters. The inverse of τ_k is often called the coefficient of variation. Note that τ_k can be small either because the parameter (low signal) β_k is small or because the variance of its estimator is large (high noise). As Belsley points out, there is a superficial resemblance to the usual t-ratio

$$t_k = b_k / s_{b_k}. \tag{5}$$

The difference is that τ_k is a parameter and t_k is a statistic that has a noncentral t-distribution. Thus, t_k is used as an estimator of τ_k. One should consult Belsley (1982) for the rationale used for the test. For our purposes, we will state how to conduct the test and to interpret the results.

In Belsley (1982, Section 3.2) the hypothesis to be tested is that the signal-to-noise level exceeds some threshold. If it does, then the data are adequate and if the threshold is not exceeded then the data are weak. The weakness may be due to collinearity or because it is "short," a concept that will be defined below. The null and alternative hypotheses are:

$$\begin{array}{ll} A_0: & \tau^2 = \tau_*^2 \\ A_1: & \tau^2 > \tau_*^2 \end{array} \tag{6}$$

where the k subscript is dropped for notational simplicity and the hypothesized threshold is subscripted with $*$, i.e., τ_*^2. The test statistic is based on the usual Wald test statistic used for testing hypotheses of the form H_0: $R\beta = r$ against H_1: $R\beta \neq r$, where R is $J \times k$

matrix of known constants, Rank(R)= $J \leq k$, and r is also known and $J \times 1$. This represents J linear hypotheses about the unknown parameters of the model. Let $b \sim N(\beta, \Sigma)$ and Σ consistently estimated via $\hat{\Sigma}$. Then the Wald statistic is:

$$\phi^2 = (Rb - r)^T (R\hat{\Sigma}R^T)^{-1}(Rb - r) \sim F_{J,n-k}(\lambda) \tag{7}$$

where λ is the non-centrality parameter associated with statistic; its value depends on the unknowns β and Σ. Belsley argues that the non-centrality of this statistic is equivalent to the s/n parameter τ^2 and under the null hypothesis that the data are less than or equal to the desired threshold

$$\phi^2 \sim F_{J,n-k}(\tau_*^2) \tag{8}$$

To test the hypothesis in (6), adopt a test size α (such as 0.05 or 0.1) and calculate

$$F_\alpha \equiv_{1-\alpha} F_{J,n-k}(\tau_*^2) \tag{9}$$

which is the $(1-\alpha)$ critical value for the non-central F with J and $n-k$ degrees of freedom and non-centrality parameter τ_*^2. Note that if $J = 1$ then $\phi^2 = t^2$. In the absence of specific knowledge of the parameters, it is reasonable to consider how far β_j are from zero using the t-ratio squared, i.e., $t^2 = (b_j/s_j)^2$ where s_j is the estimated standard error of the estimator b_j.

To conduct the test, a suitable non-centrality parameter has to be determined. Using iso-density ellipsoids

$$(R\beta - r)^T (R\Sigma R^T)^{-1}(R\beta - r) \sim {}_\gamma\chi_J^2 \tag{10}$$

as a measure of adequate s/n, where ${}_\gamma\chi_J^2$ is the γ-critical value from the central chi-square with J degrees of freedom. The parameter γ determines the size of the iso-ellipsoid that captures the relationship between the true value of the parameters and their hypothesized values. Larger values of γ increase the size of the ellipsoid and reduces the probability that $Rb = r$.

The test then proceeds as:

1. Choose a level $0 \leq \gamma < 1$ to define the desired adequacy level for your test. Higher levels of γ increase the stringency of evidence required for the s/n to be adequate.
2. Choose a test size α for the s/n test statistic. Then, compute the relevant critical value using
$$F_\alpha = {}_{1-\alpha}F_{J,n-k}({}_\gamma\chi_J^2) \tag{11}$$
3. If $\phi_2 > F_\alpha$ reject A_0 in favor of A_1.

There are four reasons why A_0 may not be rejected (and the data deemed to be weak).

1. $R\beta \approx r$. For instance, suppose a coefficient is close to zero, $\beta_i \approx 0$. In this case there is very little signal.

2. The greater the inherent noise, the smaller the statistic. In a linear model this might be due to $\sigma 2 \gg 0$.

3. Greater collinearity.

4. Short data. For instance, a particular variable in a linear model, x_j, might be short in the sense that $x_j^T x_j \approx 0$. Rescaling x_j won't help since it changes the signal by an equivalent amount.

4 Collinearity in Nonlinear Models

Assessing the severity and consequences of collinearity in nonlinear models is more complicated than in linear models. To illustrate the difficulties, we will first discuss its detection in a nonlinear least squares model and then in the context of maximum likelihood estimation and generalized linear models. The basic BKW variance decomposition analysis extends easily to these situation.

4.1 Nonlinear Least Squares

Consider the nonlinear model

$$y = f(X, \beta) + e \tag{12}$$

where $e \sim (0, \sigma^2 I_T)$ and $f(X, \beta)$ is some nonlinear function that relates the independent variables and parameters to form the systematic portion of the model. The nonlinear least squares estimator chooses $\hat{\beta}$ to minimize $S(\beta) = e^T e$. The least squares solution is

$$Z(\beta)^T [y - f(X, \beta)] = 0 \tag{13}$$

where $Z(\beta) = \partial f(X, \beta)/\partial \beta$. The matrix of second derivatives is referred to as the Hessian and is $H(\beta) = \partial^2 f(X, \beta)/\partial \beta \beta^T$. If there is more than one value of β that minimizes S, then the parameters of the model are *unidentified* and cannot be estimated. This occurs when the Hessian is singular and corresponds to perfect collinearity in the linear model. When the Hessian is nearly singular, then the model is poorly identified and reliable estimates may be difficult to obtain.

A useful algorithm for finding the minimum of $S(\beta)$ is the Gauss-Newton. The Gauss-Newton algorithm is based on a first order Taylor's series expansion of $f(X, \beta)$ around an

initial guess, β_1, for the parameters, β. From that a pseudo-linear model is constructed

$$\bar{y}(\beta) = Z(\beta_1)\beta + e \tag{14}$$

where $\bar{y}(\beta) = y - f(x, \beta_1) + Z(\beta_1)\beta_1$. Notice that the depedent variable, $\bar{y}(\beta)$ and the regressors, $Z(\beta_1)$ are completely determined given β_1. The next round estimate, β_2 is obtained by using ordinary least squares on the pseudo-linear model, $\beta_2 = [Z(\beta_1)^T Z(\beta_1)]^{-1} Z(\beta_1)^T \bar{y}(\beta_1)$, on equation (14). The iterations continue until $\beta_{n=1} \approx \beta_n$.

It can be shown that asymptotically

$$Z(\beta)^T Z(\beta)/2T \doteq H(\beta)/T. \tag{15}$$

Therefore, if H is nearly singular, then $Z(\beta)^T Z(\beta)$ will be as well. This implies that the columns of $Z(\beta)$ can be treated as regressors and analyzed using the diagnostic procedures discussed in the preceding sections.

The Gauss-Newton algorithm is affected by collinearity when $[Z(\beta_n)^T Z(\beta_n)]$ becomes singular for any of its iterations. In fact, the model could be well conditioned at the final solution, but be nearly singular at one of the many intermediate points visited by the Gauss-Newton algorithm. Unfortunately, when a near singularity is encountered the algorithm becomes numerically unstable and it often fails to converge. A solution here is to pick better starting values that avoid regions of the parameter space for which the function is ill conditioned.

A more common scenario is that the function itself is badly behaved for many points in the parameter space, including the actual minimum. In this instance, the collinearity problem is very similar to that in linear models and can be examined by using the collinearity diagnostics discussed above on the matrix of pseudo-regressors, $Z(\beta_n)$.

The conditioning of the data can be influenced to some degree by rescaling the data. Many convergence problems can be solved simply by scaling your variables in the appropriate way. On the other hand, the ill-effects of collinearity may persist regarless of scaling. By this we mean that precise estimates of the parameters are just not possible with the given data no matter how they are scaled. To detect collinearity in this setup it is suggested that the the columns of $Z(\beta)$ be rescaled to have the same length before computing the collinearity diagnostics. Large condition numbers indicate collinearity that cannot be further mitigated by scaling.

Although there are other algorithms for finding the minimum of $S(\beta)$ they are all likely to suffer the same ill-effects from collinearity.[1] It is possible that some may be better behaved in the intermediate steps of the iterative solution. Nevertheless, the asymptotic result in

[1]For instance, the Newton-Raphson, which is based on the second order Taylor's series approximation, uses the hessian computed at each round.

equation (15) suggests that in the end, it is unlikely that the ill-effects of collinearity can be manipulated in a material way by using another estimator of the asymptotic covariance matrix.

4.2 Maximum Likelihood

Maximum likelihood estimation can be approached in a similar fashion. Instead of minimizing the sum-of-squared errors function the goal is to choose parameter values that maximize the log-likelhihood function, $\ell(\beta, X)$. The algorithms use either first derivatives of ℓ, the second, or both. As in the Gauss-Newton algorithm for nonlinear least squares, each of the algorithms involves inversion of the hessian (e.g., Newton-Raphson), its negative expectation (the negative information matrix used in the method-of-scoring), or a cross-products matrix of partial first derivatives (e.g. the method of Berndt, Hall, Hall, and Hausman). In any of these instances, the inverted matrix evaluated at the each round of estimates is instrumental in solving for the parameter values that maximize the likelihood function. If at any point in the process it becomes singular or nearly so, estimation fails. If convergence occurs, then the inverse of the estimated asymptotic covariance matrix can be subjected to conditioning diagnostics in the same manner as the NLLS estimator.

4.3 Generalized Linear Models

This basic approach has been used in other contexts. Weissfeld and Sereika (1991) explore the detection of collinearity in the class of generalized linear models (GLM). This broad class of models includes the linear regression model, binary choice models like logit and probit, polychotomous choice models, the Poisson regression model, the cox proportional hazard model, and others (see McCullagh and Nelder, 1989 for discussion). In the generalized linear models the information matrix associated with the log-likelihood function can be expressed generally as

$$I(\beta) = X^T W X \tag{16}$$

where W is a $T \times T$ diagonal weight matrix that often is a function of the unknown parameters, β, the independent variables, and the responses, y. In this form, Segerstedt and Nyquist (1992) observe that ill-conditioning in these models can be due to collinearity of the variables, X, the influence of the weights, W, or both. They suggest a transformation of the data that, when plotted in the same diagram with the original data, can illuminate the change in conditioning that occurs due to the weights. Unfortunately, the method is manageable only in a few dimensions.

In GLM, Weissfeld and Sereika (1991) suggest applying the BKW condition number diag-

nostics to the the scaled information matrix ($-E[H(\beta)]$). Lee and Weissfeld (1996) do the same for the Cox regression model with time dependent regressors. Although the variance decompositions can be computed in these instances, their interpretation is not as straightforward since collinearity can also be due to the way the weights interact with the explanatory variables.

Lesaffre and Marx (1993) also investigate the problem of ill-conditioning in generalized linear models and take a slightly different approach. Following Mackinnon and Puterman (1989) they suggest that only the columns of X be standardized to unit length, forming X_1. Then, conditioning diagnostics are computed on $X_1\hat{W}X_1$, where $\hat{W}$ is the estimated weight matrix based on the rescaled data.[2] The square root of the ratio of largest to smallest eigenvalue describes the worst relative precision with which linear combinations of the location parameters can be estimated. Thus, this scaling gives a structural interpretation to the conditioning diagnostic. One problem with this scaling is that $X_1\hat{W}X_1$ could be ill-conditioned because of the effects of $\hat{W}$ which could either cause the algorithm to fail or result in very large estimated variances for the parameters of the model.

4.4 BKW diagnostics based on covariance

All of the approaches to diagnosing poorly identified models can be subjected to the condition number, variance decomposition of BKW. Tests for adequate signal-to-noise may also prove useful, even though the Belsley's theory relied on the exact normality of the least squares estimator of the classical normal linear regression model.

Even though the BKW and s/n diagnostic can identify weaknesses of the data or model, they cannot distinguish problems in the data from problems with the parameters, since the two interact often in a nonseparable way in the estimator's covariance. Despite these problems, we compute condition numbers and perform the BKW decomposition on the scaled estimated inverse of the variance-covariance matrix. This is convenient in gretl since the variance-covariance of on estimated model can be retrieved after estimation using the accessor `$vcv`. The inverse variance-covariance matrix is scaled so that the principal diagonal has one in each element.

Suppose an estimator $b \sim (\beta, \Sigma^{-1})$. Then Σ is the inverse of the estimated covariance matrix and let s_i be the i^{th} diagonal element. The matrix S be a $k \times k$ diagonal matrix with the s_i, $i = 1, \ldots, k$ on the diagonal. Then, the inverse covariance is scaled

$$\Sigma_s = S^{-1/2}\Sigma S^{-1/2}. \tag{17}$$

[2]Note, $X_1\hat{W}X_1$ is not rescaled. This is not the same as finding the condition number of the scaled estimated inverse of the information matrix.

For the least squares estimator, the condition numbers and variance decomposition based on Σ_s are the same as those based on the scaled, uncentered regressors in the original BKW diagnostic. In other words, it has no effect on the BKW table used for diagnostics. The advantage is that it can be computed almost automatically for any linear or nonlinear estimator that yields a nonsingular covariance upon convergence. Given that there are different ways to estimate covariance in nonlinear models, it is possible that some methods (i.e, outer product of gradients) might be (slightly) better behaved in small samples than others (inverse information or the negative inverse of the hessian in ML estimation).

Finally, another diagnostic that could be useful is the relative condition number. Gretl can compute this using the internal function `rcond`. This function takes a matrix A as an argument and returns the reciprocal condition number for A with respect to the 1-norm. The value is computed as the reciprocal of the product, 1-norm of A times 1-norm of A-inverse, i.e.

$$r_c = \frac{1}{\|A\|_1 \|A^{-1}\|_1} \tag{18}$$

where $\|A\|_1$ is the maximum of the absolute values of the column sums of A. The result is a number between 1 and 0, with 1 being perfectly conditioned (orthogonal regressors) and 0 being perfectly ill-conditioned. As a collinearity diagnostic, we would suggest that r_c be computed based on the scaled inverse covariance rather than unscaled version as gretl does by default using `vif`.

5 Examples

In this section, several examples are considered.

5.1 Klein-Goldberger Consumption

The first example provides a benchmark by which we can judge whether the gretl functions are working properly. The example is contained in Chapter 21 of ITPE II where Judge et al. (1988) employ the BKW collinearity analysis to the Klein-Goldberger consumption function model (Judge et al., 1988, Chapter 21).

$$C = \beta_1 + W\beta_2 + P\beta_3 + A\beta_4 + u \tag{19}$$

where the regressors include a constant, wage income (W), price level (P), and farm income (A). The BKW variance decomposition produced using the user written hansl functions

```
The BKW variance decomposition
          cond     const         W         P         A
l1       1.000     0.001     0.001     0.000     0.002
l2       6.076     0.042     0.008     0.009     0.112
l3      20.553     0.207     0.654     0.025     0.811
l4      29.255     0.750     0.338     0.966     0.075
```

and the reciprocal condition number, $r_c = 9.0552e-4$, which is very small. Note, these results match those from (Judge et al., 1988, page 873). Based on the BKW analysis the largest condition number of 29.255 is at the threshold of indicating a strong near linear dependency and probably deserves some attention. The variance decomposition proportion is greater than 0.5 for the variable P and the constant. The variation of the data in this direction is seriously impeding estimation of β_3 and the intercept β_1.

The signal-to-noise tests reveal that the data are in fact very weak. The BKW diagnostic indicates this is due to collinearity in estimation of β_3 but that the low s/n impedes estimation of the coefficients for A and P. Collinearity is the problem for β_3 and the data are short for estimation of β_4. Choosing $\gamma = 0.9$ and testing at the 5% level the t^2 s/n statistics are

```
    The t-squared signal-to-noise statistics are:

                const       W         P         A
t-squared       0.831    37.207     0.476     0.012
```

The output from computation of the critical values, which has to be done using R at this point is:

```
The signal-to-noise parameters and critical value are:
       gamma       alpha           J         n-k    critical
       0.900       0.050       1.000      16.000      13.275
```

The critical value is 13.275 and on the the t^2 for W is apparently strong enough. Note that it is also the only variable that is significantly different from zero. The very large condition number from the BKW diagnostic (29.255) and the large variance component for P suggests that this variable is highly collinear with the others. Moderately bad collinearity associated with the third eigenvalue, which predominately affect the precision of the coefficient for A is showing signs of being short as well. Its s/n is very low but collinearity is only moderately severe.

A joint test of the weakness of A and P yields a smaller critical value for the Wald statistic yields:

```
Test on Model 1:

  Null hypothesis: the regression parameters are zero for the variables
    P, A
  Test statistic: F(2, 16) = 0.238842, p-value 0.790297

The signal-to-noise parameters and critical value are:
       gamma       alpha           J          n-k    critical
       0.900       0.050       2.000      16.000      10.002
```

The weak data hypothesis cannot be rejected for $\gamma = 0.9$ at 5% since $0.2388 < 10.002$.

Gretl's internal collinearity diagnostic command is `vif`. The VIF for regressor j is defined as

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \tag{20}$$

where R_j^2 is the coefficient of multiple correlation between regressor j and the other regressors. The factor has a minimum value of 1.0 when the variable in question is orthogonal to the other independent variables. Neter et al. (1990) suggest inspecting the largest VIF as a diagnostic for collinearity; a value greater than 10 is sometimes taken as indicating a problematic degree of collinearity.

For the Klein-Goldberger model the VIFs are:

```
Variance Inflation Factors

Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

           W    7.735
           P    2.086
           A    6.213
```

By this criterion, the collinearity is bad, but does not cross the threshold of 10 suggested by Neter et al. (1990). Also, we note that the reciprocal condition number reported by VIF is based on the unscaled regressor cross products matrix, X^TX. Its value is 2.74e-6. Since the r_c is used more as an indicator of the near singularity X^TX, which in turn creates problems for performing matrix inversion, this may make sense. However, as a useful collinearity diagnostic we wonder whether the columns of the covariance should first be scaled to unit length. This would at least make inter-model comparisons more useful.

The reciprocal condition number computed for the scaled covariance is 9.05e-4, larger than the one based on unscaled covariance, but still very near zero.

5.2 Longley

One of the benchmarks used for ill-conditioned regressors is the badly collinear Longley model of employment. There are only sixteen observations and seven regressors: constant, GNP deflator, GNP, unemployment, size of the armed forces, population, and year. Gretl reports a near singularity when the linear model is estimated by `ols`. The BKW variance decomposition diagnostics for this model are:

```
The BKW variance decomposition
         cond    const   prdefl      gnp    unemp   armfrc      pop     year
l1      1.000    0.000    0.000    0.000    0.000    0.000    0.000    0.000
l2      9.142    0.000    0.000    0.000    0.014    0.092    0.000    0.000
l3     12.256    0.000    0.000    0.000    0.001    0.064    0.000    0.000
l4     25.337    0.000    0.000    0.001    0.065    0.427    0.000    0.000
l5    230.424    0.000    0.457    0.016    0.006    0.115    0.010    0.000
l6   1048.080    0.000    0.505    0.328    0.225    0.000    0.831    0.000
l7  43275.043    1.000    0.038    0.655    0.689    0.302    0.160    1.000
```

Rounding has caused the variance proportion for year due to the smallest eigenvalue to be 1. There are three condition numbers greater than 230 and clearly the collinearity situation is

very bad. Interestingly, `armfrc` is not showing high collinearity. `year` is actually orthogonal to the other variables and hence, not collinear. The other variables are showing signs of being involved in at least one of the three highly collinear relationships.

The signal-to-noise statistics are:

```
            const    prdefl      gnp    unemp   armfrc      pop     year
t-squared   15.294     0.031    1.144   17.110   23.252    0.051   16.127

The signal-to-noise parameters and critical value are:
       gamma       alpha           J          n-k    critical
       0.900       0.050       1.000        9.000      15.650
```

For this rather low standard ($\gamma = 0.9$ unemployment, armed forces, and year are showing acceptable s/n despite high collinearity. Population GNP, and price deflator have very weak s/n which we would attribute to the collinearity of the data.

5.3 Ordered Probit

The third example considers a nonlinear model, ordered probit. The data are from Mroz and the dependent variable is the number of children less than six years of age (kidsl6) and the regressors are a constant, mother's education, mother's experience, and mother's age. The model is rather dumb, but the dependent variable takes integer values 0, 1, 2, and 3, which serves our purpose. Collinearity in this case is not so much about linear relationships among the regressors, but the conditioning of the variance covariance matrix that also includes terms for the estimation of the cutoff points in the ordered probit model.

The results:

Model 2: Ordered Probit, using observations 1–753
Dependent variable: kidsl6
Standard errors based on Hessian

	Coefficient	Std. Error	z	p-value
educ	0.0437362	0.0265277	1.6487	0.0992
exper	−0.0282272	0.0101281	−2.7870	0.0053
age	−0.102429	0.00984913	−10.3999	0.0000
cut1	−2.91206	0.516607	−5.6369	0.0000
cut2	−1.74208	0.512837	−3.3970	0.0007
cut3	−0.692334	0.547788	−1.2639	0.2063

Mean dependent var	0.237716	S.D. dependent var	0.523959
Log-likelihood	−356.2304	Akaike criterion	724.4609
Schwarz criterion	752.2053	Hannan–Quinn	735.1494

Number of cases 'correctly predicted' = 610 (81.0 percent)
Likelihood ratio test: $\chi^2(3) = 196.349$ [0.0000]

Test for normality of residual –
Null hypothesis: error is normally distributed
Test statistic: $\chi^2(2) = 2.42029$
with p-value = 0.298154

The collinearity diagnostics are:[3]

```
The BKW variance decomposition
          cond      educ     exper       age      cut1      cut2      cut3
l1       1.000     0.002     0.022     0.002     0.001     0.000     0.000
l2       1.595     0.000     0.001     0.000     0.001     0.013     0.033
l3       1.974     0.000     0.002     0.000     0.002     0.005     0.110
l4       3.673     0.010     0.963     0.007     0.002     0.002     0.002
l5      10.833     0.505     0.002     0.385     0.000     0.002     0.005
l6      23.627     0.482     0.012     0.605     0.994     0.977     0.850
```

The second column includes the condition numbers and the last six are the variance decompositions for the 3 variables and the 3 cutoff points. Overall, the conditioning is not too bad since the largest condition number is 23.627, which is below the extreme threshold of 30. On the other hand, the extremely high proportion of variance of the cutoff points associated with the smallest eigenvalue suggests how fragile estimation of these parameters might be.

[3]Notice that there are no column headings since there is not a one-to-one mapping of variable to parameter in this model. A means of automating this will take some thought... This 'feature' also applies to the t^2 statistics as well.

The usefulness of the variance decomposition in this context is speculative and a remedy is not easily discerned. Since the model is nonlinear, one could not simply eliminate a principal component associated with the smallest eigenvalue, estimate the model, and perform the inverse transformation and expect proper results. The insignificance of the third cutoff is more easily explained by the fact that there are relatively few women in the sample with three kids younger than six! Also, its value is getting closer to zero, so the signal is becoming weaker. The other thing to note is that there are no column headings for the BKW table.

Other than the absence of finite sample results, we see no reason why the s/n diagnostics couldn't be employed in a nonlinear model like this one. The statistic is based on the normality of the estimator, which is asymptotic in this instance. The primary problem is in interpreting the results. For instance, it is not clear what the concept of short data mean when the underlying "variables" in the nonlinear model are functions of parameters. For instance, the columns of $Z(\beta)$ in NLS depend on β and estimates of it an the final round. The columns of the covariance matrix in the MLE are not linear functions of X; they depend on parameter values is a way that makes assigning blame for weak data equally difficult, if not impossible. However, if s/n is weak for relationships that according to the BKW diagnostics are not highly collinear, we can conclude that the identification problem is due to low signal ($\beta \approx 0$) or that the implied scores are short in the same sense that data are. This is speculative and deserves more thought.

Finally, we try our hand at computing the s/n diagnostic on the nonlinear model.

The signal-to-noise statistics are:

```
                educ     exper       age      cut1      cut2      cut3
t-squared      2.718     7.767   108.157    31.775    11.539     1.597

The signal-to-noise parameters and critical value are:
       gamma       alpha           J         n-p     critical
       0.900       0.050       1.000     747.000      10.869
```

Signal-to-noise is adequate for the age coefficient and the first two cut-offs. Inadequate for the other parameters. This is despite the fact that experience is significantly different from zero in the regression. With one fairly high condition number we'd say that there is one collinear relationship that involves education and age, made worse by the existence of the cutoff parameters. Surprisingly, of the four variables only education seems to fall victim to weakness of the data. Experience does not appear to be collinear, but is showing signs of being short.

5.4 Linear Instrumental Variables

The final example considered is that of a model with an endogenous regressor estimated using linear instrumental variables. The model is:

$$ln(\text{wage}) = \beta_1 + \text{educ}\beta_2 + \text{exper}\beta_3 + \text{age}\beta_4 + u \tag{21}$$

Here, the years of schooling, educ, is endogenous. We use mother's (mothereduc) and father's (fathereduc) years of schooling as instruments. First, the model is estimated using OLS.

```
The BKW variance decomposition
         cond     const      educ     exper       age
l1      1.000     0.001     0.002     0.013     0.002
l2      4.265     0.007     0.021     0.769     0.002
l3     11.418     0.005     0.507     0.156     0.437
l4     19.668     0.987     0.471     0.063     0.559

The signal-to-noise statistics are:
               const      educ     exper       age
t-squared      1.735    59.211    12.613     0.086

The signal-to-noise parameters and critical value are:
       gamma       alpha           J         n-k    critical
       0.900       0.050       1.000     424.000      10.904
```

Collinearity is bad, but not awful. Signal-to-noise is low for the constant and age, but adequate for education and experience.

Two-stage least squares estimation is quite different. First, the instruments appear to be quite strong.

```
Weak instrument test -
  First-stage F-statistic (2, 423) = 55.5516
```

However, collinearity is induced by using imperfect instruments.

```
The BKW variance decomposition
         cond     const      educ     exper       age
l1      1.000     0.000     0.000     0.013     0.002
```

```
l2      4.302     0.003     0.004     0.784     0.003
l3     13.661     0.016     0.072     0.194     0.859
l4     35.625     0.981     0.924     0.009     0.137

 The signal-to-noise statistics are:
               const      educ     exper       age
t-squared      0.300     4.180    12.569     0.200
```

The critical value for the $\gamma = .9$ 5% test is still 10.904. Note that the largest condition number has nearly doubled in size to 35.6 and education is clearly involved in this near singularity. The signal-to-noise has become inadequate for education, even when the instruments are considered strong.

6 Conclusion

We have briefly summarized know results for collinearity diagnostics in linear and nonlinear regression models. We base our main analysis on the condition number and variance decomposition analysis suggested by Belsley et al. (1980). By analogy, we extend that analysis to nonlinear models based on the scaled inverse of the covariance matrix. Although this introduces some noise into the diagnostics, we find them very similar in magnitude to their linear model counterparts.

Signal-to-noise statistics are a useful supplement and allow us to identify whether the data are weak due to collinearity or because they are "short". This also seems to translate, probably better in fact, to the nonlinear environment. The gretl functions written for these statistics are straightforward to use and relatively easy to interpret. Being based on the variance covariance matrix, the BKW can be applied to nearly any model desired.

We had to resort to R to compute the noncentral F critical values. This if fiddley in the sense that it can not be done automatically, requiring hard coding of the locations of reading and writing of matrices. We can hope that gretl finds her own native functions soon.

Finally, we took a stab at using the diagnostics in an instrumental variables setting. The results were quite interesting in that collinearity induced by the use of instruments becomes apparent when compared to the OLS results. We think this deserves further investigation.

7 Code

Below you will find an example script that can be used as a model for your own diagnostic purposes. This is follwed by the complete set of code for all of the functions used. The examples in the preceding sections can be reproduced using the `hansl` code at the bottom. These functions have been packaged into a function `bkw.gfn`.

The function takes two inputs: 1) the estimated variance covariance matrix produced by estimation and 2) a string consisting of either the variable names (linear model) or a user defined string that will serve as headings for the columns of the BKW table. The output is the BKW diagnostics matrix.

7.1 Example Script

```
open mroz
list xlist = const educ exper age

probit kidsl6 xlist
string vn = "educ exper age cut1 cut2 cut3"
matrix co = $vcv
matrix bb = bkw(co, vn)
```

In this example we use a nonlinear model to illustrate the manual construction of the string that populates the column names. The ordered probit has 4 categories, thus in the parameterization without a constant there are three cutoff points to estimate. In gretl's parameterization of the model, there is no constant so the three regressors are named in the string.

The signal-to-noise diagnostic `signal_noise` simply computes the critical value from the non-central F and puts all of the relevant information into a table. The computation of the noncentral critical values for the hypothesis tested is all that is really required. Specify the degree of evidence you desire (γ) and choose a test size (α) and feed degrees of freedom (`J` and `n-k`).

We have also written a function `t2` that computes the signal-to-noise diagnostics for each parameter with the null being that no signal is present, i.e., this is the standard t-ratio. It is not really necessary to compute `t2` since gretl produces t-ratios for every model it estimates. The `t2` requires the variable names and the coefficients and standard errors from the estimated model.

```
scalar g = .90
scalar a = .05
matrix sn = signal_noise(g,a,1,$df)

t2(vn,$coeff,$stderr)
```

7.2 The complete set of code for all functions

Below you will find the complete listing of code used in the functions.

```
clear
set messages off
set echo off

# This creates a string to name rows of BKW matrix
# Input: a scalar that indicates number of eigenvalues
function string lam_names(scalar k)
    string lall = null
    loop for i=1..k
        sprintf l "l%d",i
        string ls = "@l "
        string lall += ls
    endloop
    return lall
end function

# This performs the variance decomp of BKW
# Input:  L = vector of Eigenvalues based on X'X or covariance
#         V = matrix of Eigenvectors
# Note: If using eigenvalues, L, from scaled X, then square them
function matrix bkw_var_decomp(matrix L, matrix V)
    scalar k = cols(V)
  # Variance Decomposition
    matrix u2 = (V.^2)' .* (1/L ** ones(k,1))
    matrix vard = sumr(u2)
    matrix p = u2 ./ (vard ** ones(1,k))
    matrix p = p'
    return p
end function
```

```
# This scales the diagonals of the inverse covariance matrix to unit length
# Input:  covariance matrix of estimator
# Output:  vector of condition numbers
function matrix scale_cov(matrix vc)
    matrix vc = inv(vc)
    matrix xs = 1/sqrt(diag(vc))
    matrix S=I(cols(vc))
    matrix S[diag] = xs
    matrix scaled_cov = qform(S,vc)
    return scaled_cov
end function

# This computes condition numbers
# Input: matrix of variables, scores, or covariance
# BKW recommend scaling before using this
function matrix cond_num(matrix vc)
    matrix lambda = svd(vc)
    scalar ll = max(lambda)
    matrix condnumber = sqrt(ll./lambda)
    matrix rc = rcond(vc)
    printf "Relative condition number \n%10.4e\n\n",rc
    printf "Condition numbers based on normalized covariance \n%10.0f\n",condnumber
    return condnumber
end function

#Assembles the BKW table and prints it to screen
#Inputs:  l = matrix of eigenvalues
#         p = matrix of eigenvectors
#         vn = string to use for column names
function matrix bkw_table(matrix l, matrix p, string vn)
    string lnames = lam_names(nelem(l))
    string lname = strsub(lnames, ",", " ")
    string vname = strsub(vn, ",", " ")
    matrix p = l'~p
    colnames(p, " cond @vname " )
    rownames(p, " @lname ")
    printf "\nThe BKW variance decomposition \n%10.3f\n ",p
    return p
end function

#Computes the BKW table and outputs it to matrix
#Inputs:  co = scaled inverse covariance matrix
```

```
#         vn = string to use for column names
function matrix bkw(matrix co, string vn)
    matrix co = scale_cov(co)
    matrix V U
    matrix L = svd(co, &U , &V)
    matrix VD = bkw_var_decomp(L, V)
    matrix cond_n = cond_num(co)
    matrix bkw = bkw_table(cond_n,VD, vn)
    return bkw
end function

#This computes the critical value for the s/n test.
# g = gamma (desired degree of evidence
# a = alpha, test size
# J = number of hypotheses, numerator df
# df = denominator degrees of freedom
function matrix signal_noise(scalar g, scalar a, scalar J, scalar df)
    matrix parms = g~a~J~df
    scalar c = critical(X,J,1-g)
    crit=invcdf(ncf,J,df,c,1-a)
    matrix xx = parms~crit
    colnames(xx,"gamma alpha J n-k critical")
    printf "The signal-to-noise parameters and critical value are: \n%12.3f\n", xx
    return xx
end function

# This computes t-ratios.  Redundant.
# This could be added as a row to BKW table if desired
function void t2(string vn, matrix t, matrix se)
    matrix phi_2 = ((t./se).^2)'
    vname = strsub(vn, ",", " ")
    rownames(phi_2,"t-squared")
    colnames(phi_2, " @vname ")
    printf "The signal-to-noise statistics are: \n\n%10.3f\n", phi_2
end function

open KLEING.gdt
list xlist = const W P A
ols C xlist
vif
scalar g = .90
scalar a = .05
```

```
matrix sn = signal_noise(g,a,1,$df)
string vn = varname(xlist)
matrix co = $vcv
matrix bb = bkw(co, vn)

t2(varname(xlist),$coeff,$stderr)
matrix sn = signal_noise(g,a,1,$df)

omit P A --test-only
matrix sn = signal_noise(g,a,2,$df)
string vn = varname(xlist)
matrix co = $vcv
matrix bb = bkw(co, vn)

open longley
list xlist = const 2 3 4 5 6 7
ols employ xlist
string vn = varname(xlist)
matrix cc = $vcv
matrix bb = bkw(cc, vn)
t2(vn,$coeff,$stderr)
scalar g = .90
scalar a = .05
matrix sn = signal_noise(g,a,1,$df)

open mroz
list xlist = const educ exper age
probit kidsl6 xlist
v=$coeff
matrix co = $vcv

string vn = "educ exper age cut1 cut2 cut3"
matrix bb = bkw(co, vn)

t2(vn,$coeff,$stderr)
scalar g = .90
scalar a = .05
matrix sn = signal_noise(g,a,1,$df)

open mroz
logs wage
list xlist = const educ exper age
```

```
list zlist = const exper age mothereduc fathereduc
tsls l_wage xlist; zlist
string vn = varname(xlist)
matrix co = $vcv
matrix bb = bkw(co, vn)
t2(vn,$coeff,$stderr)
scalar g = .90
scalar a = .05
matrix sn = signal_noise(g,a,1,$df)

ols l_wage xlist
string vn = varname(xlist)
matrix co = $vcv
matrix bb = bkw(co, vn)
t2(vn,$coeff,$stderr)
scalar g = .90
scalar a = .05
matrix sn = signal_noise(g,a,1,$df)
```

References

Belsley, D. A. (1982), 'Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise', *Journal of Econometrics* **20**, 211–253.

Belsley, D. A. (1991), *Collinearity Diagnostics: Collinearity and Weak Data in Regression*, Wiley, New York.

Belsley, David A. (1991), *Collinearity Diagnostics: Collinearity and Weak Data in Regression*, John Wiley & Sons, New York.

Belsley, David A., E. Kuh and R. E. Welsch (1980), *Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity*, John Wiley & Sons, New York.

Erkoç, Ali, Müjgân Tex and Kadri Ulaş Akay (2010), 'On multicollinearity in nonlinear regression models', *Selçuk Journal of Applies Mathematics* pp. 65–72.

Friendly, Michael and Ernest Kwan (2009), 'Wheres waldo: Visualizing collinearity diagnostics', *The American Statistician* **63**, 56–65.

Hill, R. Carter and Lee C. Adkins (2001), Collinearity, *in* B.Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Publishers Ltd, Malden Massachusettes, pp. 256–278.

Hillebrand, Eric and Tae-Hwy Lee (2012), 'Stein-rule estimation and generalized shrinkage methods for forecasting using many predictors', *Advances in Econometrics* **30**, 171–196.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl and Tsoung-Chao Lee (1988), *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York.

Lee, Kyung Yul and L. A. Weissfeld (1996), 'A multicollinearity diagnostic for the cox model with time dependent covariates', *Communications in Statistics - Simulation* **25**, 41– 60.

Lesaffre, E. and B. D. Marx (1993), 'Collinearity in generalized linear regression', *Communications in Statistics - Theory and Methods* **22**, 1933–1952.

Mackinnon, M. J. and M. L. Puterman (1989), 'Collinearity in generalized linear models', *Communications in Statistics - Theory and Methods* **18**, 3463–3472.

Neter, J., W. Wasserman and M. H. Kutner (1990), *Applied Linear Statistical Models*, Irwin, Boston.

Segerstedt, B. and H. Nyquist (1992), 'On the conditioning problem in generalized linear models', *Journal of Applied Statistics* **19**, 513–522.

Silvey, S. (1969), 'Multicollinearity and imprecise estimation', *Journal of the Royal Statistical Society, B* **31**, 539–552.

Weissfeld, L. A. and S. M. Sereika (1991), 'A multicollinearity diagnostic for generalized linear models', *Communication in Statistics, A* **20**, 1183–1198.