Collinearity Diagnostics in gretl

Lee C. Adkins and Melissa S. Waters

February 27, 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 documented in Hill and Adkins (2001) and in particular, we introduce hansel routines to perform the variance decomposition of Belsely, Kuh, and Welch (1980) for both linear and nonlinear models. The latter has been a feature of SAS, but is generally not available in other software of which we are aware. We also provide a function to identify collinearity-influential observations as suggested by Belsley (1991) and explore the usefulness of the Belsley (1982) signal-to-noise test statistic. Other diagnostics will be provided, time permitting.

1 Collinearity in Nonlinear Models

Assessing the severity and consequences of collinearity in nonlinear models is a bit 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.

Consider the nonlinear model

$$y = f(X, \beta) + e \tag{1}$$

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 $S(\beta) = e'e$. The least squares solution is

$$Z(\beta)'[y - f(X, \beta)] = 0 \tag{2}$$

where $Z(\beta) = \partial f(X, \beta)/\partial\beta$. The matrix of second derivatives is referred to as the Hessian and is $H(\beta) = \partial^2 f(X, \beta)/\partial\beta\beta'$. 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

$$\bar{y}(\beta) = Z(\beta_1)\beta + e \tag{3}$$

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, $\beta_2 = [Z(\beta_1)'Z(\beta_1)]^{-1}Z(\beta_1)'\bar{y}(\beta_1)$, on equation (3). The iterations continue until $\beta_{n=1} \approx \beta_n$.

It can be shown that asymptotically

$$Z(\beta)'Z(\beta)/2T \doteq H(\beta)/T. \tag{4}$$

Therefore, if H is nearly singular, then $Z(\beta)'Z(\beta)$ 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 $[Z(\beta_n)'Z(\beta_n)]$ 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 equation (4) 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.

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.

This basic approach has been used in other contexts. Weissfeld and Sereika (1991) explore

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

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

$$I(\beta) = X'WX \tag{5}$$

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 diagnostics 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.

References

Belsley, David 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.

[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.

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

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.

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.

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

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