Collinearity Diagnostics in gretl

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

June 4, 2015

Basic Outline

- Explore some of the tools gretl has to analyze collinearity
- BKW: condition numbers, variance decompositions and signal-to-noise tests in linear models
- Application of these diagnostics to nonlinear models
- Examples

Consequences of Collinearity

- Imprecise parameter estimates
- Weak hypothesis tests
- Poor predictions if collinearity structure changes out-of-sample
- Otherwise, no problem. OLS of the CLRM parameters is BLUE. Tests based on these are valid. :-)

Linear Model

$$y = X\beta + u \tag{1}$$

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 linear model it is a data problem.

Perfect	$c_1x_1 + c_2x_2 + \cdots + c_Kx_k = 0$
Nearly Perfect	$c_1x_1 + c_2x_2 + \cdots + c_Kx_k \approx 0$
- Perfect, collinearity exists when the columns of X, denoted x_i, $i = 1, \ldots, k$, are linearly dependent. The parameters are not identified.
- It's nearly perfect is the linear combination is close to zero. Here, the parameters are weakly identified.

Simple diagnostics

- Determinant of X^TX. Useful since $\det(X^TX) = \prod_{i=1}^{k} \lambda_i$ where the λ_i are the eignenvalues of X^TX and a small product indicates low data variation in at least one direction of the data space. But the magnitude of eigenvalues changes with the variables scaling. The columns of X should be first scaled to unit length.
- VIF. These are auxiliary regressions of each explanatory as a linear function of all of the others. Large R^2 make for large VIF which indicate multiple linear associations among variables. Gretl produces these and provides advice on the cutoff (>30 is high collinearity). These are not generalizable to nonlinear regressions.

Simple diagnostics continued

- Relative condition number.

$$r_c = \frac{1}{\|A\|_1 \|A^{-1}\|_1} \tag{2}$$

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. Gretl computes this for $A = \mathrm{Cov}(\hat{\beta})$ in the `vif` command, but the covariance isn't scaled and the measure is not as informative for diagnosing relative collinearity as it could be.

Diagnostics based on Eigenvalues and vectors

Silvey (1969) popularized the use of eigenvalues to diagnose collinearity, and Belsley et al. (1980) [hereinafter BKW] refined, and improved, the analysis. For symmetric matrices A their exists an orthonormal $k \times k$ matrix C such that

$$C^T AC = \Lambda \tag{3}$$

where Λ is a diagonal matrix with the real values $\lambda_1, \lambda_2, \ldots, \lambda_k$ on the diagonal. Note that covariance matrices, their inverses, and the regressor cross products matrix $X^T X$ are symmetric.

Diagnostics based on Eigenanalysis, cont.

- Note, $CC^T = I_k$.
- 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).
- Since the eigenvalues depend on the magnitude of the elements of X, Belsley suggests that the columns of X be scaled to unit length.
- Including the scale parameter as in $(X^T X)/\sigma^2$ has no net effect on the magnitudes of the eigenvalues. The decomposition can be performed on the scaled inverse covariance with identical results.

Using equation (3) and the properties of the matrix of eigenvectors C, we can write $X^TX = C\Lambda C^T$, and therefore

$$(X^TX)^{-1} = C\Lambda^{-1}C^T = \sum_{i=1}^{k} \lambda_i^{-1} c_i c_i^T \tag{4}$$

The covariance matrix of the least squares estimator b is $cov(b) = \sigma^2(X^TX)^{-1}$, and using equation (4) the variance of b_j is

$$var(b_j) = \sigma^2 \left(\frac{c_{j1}^2}{\lambda_1} + \frac{c_{j2}^2}{\lambda_2} + \ldots + \frac{c_{jk}^2}{\lambda_k} \right) \tag{5}$$

Variance Decomposition

Define $\phi_{jk} = \frac{c_{jk}^2}{\lambda_k}$, 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 $\frac{\phi_{jk}}{\phi_j}$.

Condition Index

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. The larger the condition index the worse the collinearity.

BKW Table

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: Matrix of Variance Proportions

Table 1 summarizes much of what we can learn about collinearity in data.

Recipe

- Indices 0-10 indicates weak near dependencies, 10-30 indicate moderately strong near dependencies, 30-100 is a large condition index, a strong near dependency, and indices in excess of 100 are very strong.
- One large η_k: 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.

More steps

- If $J \geq 2$ large η_j of roughly equal size. Sum the variance proportions for the coefficients across the J large condition number rows in Table 1. If the sum of the variance proportions exceeds 50%, then those variables are affected by strong collinearity. The variance proportions in a single row do not identify specific linear dependencies, as they did when there was just one large condition number.
- If $J \geq 2$ large condition numbers with 1 extremely large. η_k very large can "mask" the variables involved in other near exact linear dependencies. 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%.

Steps continued

- Determine if any coefficient is NOT affected by collinearity. A single large condition number with variance proportions less than 50% in the last row of Table 1 are not adversely affected. For $J \geq 2$ large condition numbers, sum the last J rows of the variance proportions. Sums less than 50% are not adversely affected by the collinear relationships.
- If key estimate adversely affected, further diagnostic steps may be taken.

Signal-to-noise

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

$$\tau_k \equiv \beta_k / \sigma_{b_k} \tag{6}$$

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.

Test and critical values

- The null hypothesis, A_0, is that the s/n is inadequate. The alternative, A_1, is that it is adequate.
- User decides what is 'adequate' through the choice of a parameter, γ. Larger values of γ increase the burden on the model/data to be adequate.
- The test is based on the usual t-ratio (squared) or a Wald statistic, ϕ_2 which has a noncentral F-distribution.
- The parameter γ determines the value of the noncentrality parameter, $({}_{\gamma}\chi^2_J)$, used in the test. $({}_{\gamma}\chi^2_J)$ is the γ level critical value from the χ^2_J.

Test and critical values, cont.

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

$$F_\alpha = {}_{1-\alpha}F_{J,n-k}({}_\gamma\chi^2_J) \qquad (7)$$

where J is the number of linear relationships to test and α the desired level of the test.

3. If t^2 or $\phi_2 > F_\alpha$ reject A_0 in favor of A_1.

Gretl can't natively compute critical values from noncentral distributions so this is done using R within a gretl script.

Rejection Means?

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

1. There is very little signal, e.g., $\beta_i \approx 0$.
2. The data are very noisy, i.e., in CLRM $\sigma 2 \gg 0$.
3. Very high 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.

Models

The BKW diagnostics can be used in any estimator whose covariance can be estimated. The s/n diagnostics can be used for estimators that are normally, or approximately normally distributed.

1. Nonlinear Least Squares
2. Maximum Likelihood
3. Generalized Linear Models
4. GMM

How?

Estimate the model, retrieve the estimated variance covariance matrix, invert it, and scale so that its diagonal elements are equal to one.

Let Σ be the inverse of the estimated covariance matrix and s_i being the i^{th} diagonal element. Define S to be a $k \times k$ diagonal matrix with the s_i, $i = 1, \ldots, k$ on the diagonal. Then, the inverse of the covariance is scaled

$$\Sigma_s = S^{-1/2} \Sigma S^{-1/2} \tag{8}$$

.

Why does it work?

- It works for any linear model producing exactly the same condition indices and variance decompositions as if done on the scaled X^TX matrix. The scaling removes any influence of σ^2.
- Variance decomposition is not unique to the linear model. Interpretation is less straightforward elements of the covariance are functions of variables and parameters. The cause of the collinearity cannot be directly isolated to one or the other (of course there are exceptions).
- Likewise, the s/n diagnostics should still work as well as long as the estimators are (asymptotically) normally distributed. Belsley's s/n analysis relies only on this normality.

Three examples from the paper:

1. Klein-Goldberger Consumption function from ITPE II
2. Longley: famously bad collinearity
3. Ordered Probit: Nonlinear example
4. Linear Instrumental Variables: Weak Instruments or Collinearity?

Klein-Goldberger Consumption

The first is the collinearity analysis of the Klein-Goldberger consumption function model from (Judge et al., 1988, Chapter 21).

$$C = \beta_1 + W\beta_2 + P\beta_3 + A\beta_4 + u \tag{9}$$

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.

Variance Decomposition Table

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

Signal-to-Noise Statistics

The signal-to-noise statistics are:

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

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

```
gamma     alpha       J       n-k   critical
0.900     0.050   1.000    16.000     13.275
```

VIFs from gretl

```
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
```

BKW variance decomposition

7 regressors: constant, GNP deflator, GNP, unemployment, size of the armed forces, population, and year.

cond	const	prdefl	gnp	unemp	armfrc	pop	year
1	0.0	0.000	0.000	0.000	0.000	0.000	0.0
9	0.0	0.000	0.000	0.014	0.092	0.000	0.0
12	0.0	0.000	0.000	0.001	0.064	0.000	0.0
25	0.0	0.000	0.001	0.065	0.427	0.000	0.0
230	0.0	0.457	0.016	0.006	0.115	0.010	0.0
1048	0.0	0.505	0.328	0.225	0.000	0.831	0.0
43275	1.0	0.038	0.655	0.689	0.302	0.160	1.0

Signal-to-Noise: $\beta_i = 0$

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

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
```

BKW variance decomposition

Data: Mroz, dependent variable = kidsl6 (0, 1, 2, 3), 6 parameters, 3 regressors

cond	educ	exper	age	cut1	cut2	cut3
1.000	0.002	0.022	0.002	0.001	0.000	0.000
1.595	0.000	0.001	0.000	0.001	0.013	0.033
1.974	0.000	0.002	0.000	0.002	0.005	0.110
3.673	0.010	0.963	0.007	0.002	0.002	0.002
10.833	0.505	0.002	0.385	0.000	0.002	0.005
23.627	0.482	0.012	0.605	0.994	0.977	0.850

Signal-to-Noise: $\beta_i = 0$

	t-squared	F=10.869 (a=.05,g=.9)
educ	2.718	low
exper	7.767	low
age	108.157	adequate
cut1	31.775	adequate
cut2	11.539	adequate
cut3	1.597	low

gamma	alpha	J	n-k	critical
0.900	0.050	1.000	747.000	10.869

BKW OLS

$$ln(\text{wage}) = \beta_1 + \text{educ}\beta_2 + \text{exper}\beta_3 + \text{age}\beta_4 + u \qquad (10)$$

First, the model is estimated (Mroz) 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

BKW TSLS

TSLS with educ endogenous and mothereduc and fathereduc are instruments. The instruments are strong and so is collinearity.

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

	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

Signal-to-noise

	OLS t-squared	TSLS t-squared
const	1.735	0.300
educ	59.211	4.180
exper	12.613	12.569
age	0.086	0.200

gamma	alpha	J	n-k	critical
0.900	0.050	1.000	424.000	10.904

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

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.

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