

Weak Identification in Nonlinear Econometric Models

Lee C. Adkins

June 22, 2017

Abstract

The use of the Belsley, Kuh, and Welch Belsley et al. (1980) diagnostics in nonlinear models is explored via simulation. The two models examined are ordered probit and endogenous sample selection. The performance of the diagnostics are examined under several parameterizations and under different degrees of collinearity. The results suggest that the identification of the model's parameters is indeed related to collinearity of the data and by the parametric functional form of the model itself. The BKW diagnostics appear to be somewhat useful in the sense that serious problems with identification can be detected quite easily using the built in `vif` function of `gretl`.

Introduction

Adkins et al. (2015) wrote `gretl` functions to compute the Belsley et al. (1980) collinearity diagnostics that could be used to detect weak identification in linear and nonlinear models. In this companion, I explore how well the BKW measures perform in nonlinear contexts. In linear models the rules-of-thumb suggested by BKW indicate how severe the collinearity problem is and can tell the user whether the variance of a particular coefficient estimator is adversely affected by collinear data. The diagnostics are now available in the base versions of GRETL, but the performance of the rules of thumb have yet to be studied in any systematic way when used in this newer context.

In nonlinear models the BKW diagnostics are computed based on a rescaled variance covariance matrix, and are comparable to what is currently used in linear models estimated by least squares. In nonlinear models the covariance matrix depends on parameters as well as data and the interaction of the two can affect the identification of parameters.

In nonlinear models weak identification can lead to excessive numbers of iterations, to non-convergence of the algorithm, or can cause the covariance estimate (e.g., negative of the inverse Hessian) to be negative definite; if this happens the numerical algorithm fails to locate the optimum. When any of these happens, one would like to know whether the problem happens due to data or because of its interaction with parameters. A corrective strategy may materialize; get better data, adopt a simpler model, or impose additional constraints on the model.

The generalizability of the diagnostics to nonlinear models is studied using simulations of two nonlinear models. The average condition numbers and variance decompositions for each model are presented. The output also includes standard deviations from the Monte Carlo for each scenario. In this way, the relative variability due to parameter estimation can be measured. If the estimates play a fairly modest role, then much of the weak identification of the nonlinear estimator can be attributed to the data, which is similar to the situation in linear models. The goal is to form a foundation for the expanded use of the BKW condition number and variance decomposition analysis to the nonlinear world.

The paper is organized as follows. First, I review collinearity and the BKW diagnostics in linear models. In section 3 the extension of nonlinear models is reconsidered. In section 4 I examine two examples, ordered probit and sample selection. Within this section two sets of simulations are conducted and the performance of the BKW diagnostics are studied. I show that the diagnostics can be quite useful in detecting problems of identification in these models due to either collinearity of the data or to parameterization of the model.

1 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 , $i = 1, \dots, K$, are linearly dependent. That is, if there is at least one relation of the form $c_1x_1 + c_2x_2 + \dots + c_Kx_k = 0$, 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 $X^T X\beta = X^T y$ 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 $c_1x_1 + c_2x_2 + \dots + c_kx_k \approx 0$. We now examine the consequences of such near exact linear dependencies. The collinearity problem can more broadly be viewed as an identification problem. If collinearity is exact, then identification fails. If collinearity is strong, then the parameters are identified, but they may be estimated imprecisely given the data on hand.

1.1 Diagnosing Collinearity using the Eigenvalues and Eigenvectors of $X^T X$

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

$$C^T X^T X C = \Lambda \quad (1)$$

where Λ is a diagonal matrix with the real values $\lambda_1, \lambda_2, \dots, \lambda_k$ on the diagonal. An orthonormal matrix, sometimes also called an orthogonal matrix, has the property that $C^T = C^{-1}$, so that $C^T C = C C^T = I_k$, 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, $\lambda_1 \geq \lambda_2 \geq \dots, \geq \lambda_k$.

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 $Z = X C$ 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 $z_i^T z_i = \lambda_i$. If the characteristic root $\lambda_i = 0$, then $z_i = X c_i = 0$; we have an exact linear relation among the columns of X , and thus exact collinearity. If $\text{rank}(X) = \ell < k$, 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 $X^T X$ 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

$$\lambda_i = z_i^T z_i = c_i^T X^T X c_i \approx 0$$

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 $U^T U = V^T V = I_k$ and D is a diagonal matrix with non-negative diagonal values $\mu_1, \mu_2, \dots, \mu_k$, 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 $X^T X$, 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.

1.2 Collinearity and the Least Squares Estimator

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

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

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

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

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 one 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 $X^T X$ 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?”

1.3 Variance Decomposition of Belsley et al. (1980)

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

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

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. BKW's experiments lead them to the general guidelines that indices in the range

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

Table 1: Matrix of Variance Proportions

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 This depends on the number of large condition numbers identified in Step 1

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.

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.

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.

2 Identification in Nonlinear Models

Assessing the severity and consequences of nearly singular Hessians in nonlinear models is more complicated than in linear models, since the invertibility of the Hessian can be due to things other than collinearity of the variables of the model. The difficulties caused by the extension to nonlinear models can be illustrated using nonlinear least squares model and then extended to the context of maximum likelihood estimation, generalized linear models and other models that require nonlinear estimators. The basic BKW variance decomposition analysis extends easily to these situation.

2.1 Nonlinear Least Squares

Consider the nonlinear model

$$y = f(X, \beta) + e \tag{4}$$

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

$$Z(\beta)^T [y - f(X, \beta)] = 0 \quad (5)$$

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

$$\bar{y}(\beta_1) = Z(\beta_1)\beta + e \quad (6)$$

where $\bar{y}(\beta_1) = y - f(x, \beta_1) + Z(\beta_1)\beta_1$. Notice that the dependent 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 (6). The iterations continue until $\beta_{n=1} \approx \beta_n$.

It can be shown that asymptotically

$$Z(\beta)^T Z(\beta) / 2T \doteq H(\beta) / T. \quad (7)$$

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 among the columns of $Z(\beta)$ since $[Z(\beta_n)^T Z(\beta_n)]$ may become 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 regardless 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 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 (7) 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.

2.2 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-likelihood 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.

2.3 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

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

linear models the information matrix associated with the log-likelihood function can be expressed generally as

$$I(\beta) = X^T W X \quad (8)$$

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 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.

2.4 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.

Even though the BKW 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

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

inverse of the variance-covariance matrix. This is convenient in gretl since the variance-covariance of an 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. In recent releases of gretl, the `vif` command summons the BKW diagnostic matrix after estimation of any model. Refer to Adkins et al. (2015) for computational details based on covariance.

3 Examples

Two models are considered in the simulations that follow. The ordered probit is easily estimated in gretl and it contains additional parameters that are related to the number of categories, or bins, for the dependent variable. These “cutoff” parameters determine the points at which the observation moves from one category to the adjacent one when an independent variable changes.

The second model is that of sample selection as in (Greene, 2012, pp. 874-876). This model can be estimated in two-steps (Heckman (1979)) or by maximum likelihood and is based on two equations that may contain highly collinear variables both within each equation and across equations. In one extreme, the variables of the selection and regression equations are highly collinear within the equation and the variables of each equation are the same. As is shown below, this can create problems for estimating the “selection” parameter.

3.1 Ordered Probit

Following (Greene, 2012, p. 787-788) the ordered probit model is treated as a latent variable, y_i^* that depends on a linear index, $x_i^T \beta$.

$$y_i^* = x_i^T \beta + e_i$$

The latent variable y^* is unobserved. Instead we observe integers, y , such that

$$\begin{aligned} y=0 & \text{ if } y^* \leq 0 \\ y=1 & \text{ if } 0 \leq y^* \leq \tau_1 \\ y=2 & \text{ if } \tau_1 \leq y^* \leq \tau_2 \\ & \vdots \\ y=J & \text{ if } \tau_{J-1} \leq y^* \end{aligned}$$

The β and the τ s are unknown parameters. If e is assumed to be normally distributed across observations. To identify β , it is conventional to let $\sigma = 1$ in order to identify β . Then,

$$\begin{aligned} \text{Prob}(y = 0|x) &= \Phi(-x^T\beta) \\ \text{Prob}(y = 1|x) &= \Phi(\tau_1 - x^T\beta) - \Phi(-x^T\beta) \\ \text{Prob}(y = 2|x) &= \Phi(\tau_2 - x^T\beta) - \Phi(\tau_1 - x^T\beta) \\ &\vdots \\ \text{Prob}(y = J|x) &= 1 - \Phi(\tau_{J-1} - x^T\beta) \end{aligned}$$

with $0 < \tau_1 < \tau_2 < \dots < \tau_{J-1}$ and $\Phi(t)$ is the standard normal cdf evaluated at t .

In the following empirical example I model the probability of having children less than six years of age using the data are from Mroz. The regressors include a constant, mother's education, mother's experience, and mother's age. The dependent variable, `kidls6`, takes integer values 0, 1, 2, and 3. Identification is about linear relationships among the regressors as well as the parameterization of the model itself, which includes three regression coefficients, β s, and three cutoff parameters, τ s.

The results:

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

	Coefficient	Std. Error	z	p-value
<code>educ</code>	0.0437362	0.0265277	1.6487	0.0992
<code>exper</code>	-0.0282272	0.0101281	-2.7870	0.0053
<code>age</code>	-0.102429	0.00984913	-10.3999	0.0000
<code>cut1</code>	-2.91206	0.516607	-5.6369	0.0000
<code>cut2</code>	-1.74208	0.512837	-3.3970	0.0007
<code>cut3</code>	-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 results suggest that `educ` is not significant at 5% and neither is the right-most cutoff parameter. The collinearity diagnostics could shed light as to why.

The `vif` command in GRETTL produces the BKW table as follows:

		--- variance proportions ---					
lambda	cond	educ	exper	age	cut1	cut2	cut3
3.460	1.000	0.002	0.022	0.002	0.001	0.000	0.000
1.360	1.595	0.000	0.001	0.000	0.001	0.013	0.033
0.888	1.974	0.000	0.002	0.000	0.002	0.005	0.110
0.256	3.673	0.010	0.963	0.007	0.002	0.002	0.002
0.029	10.833	0.505	0.002	0.385	0.000	0.002	0.005
0.006	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 columns 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. The largest condition number plays a significant role in estimation of all three cutoff parameters, which all have variance proportions greater than 85%. This suggests that these parameters are relatively weakly identified by the model.

With one fairly high condition number there appears to be one moderately collinear relationship that involves education and age, and made worse by the existence of the cutoff parameters. Surprisingly, of the four variables only education falls victim to weakness of the data. Experience does not appear to be collinear with any of the other variables. The model seems to be fairly well identified despite these issues.

Simulation

There are two sources of trouble with the identification of the ordered probit model. 1) the data could be collinear and 2) the likelihood function as parameterized could be relatively flat. The simulation explores both scenarios.

Collinear regressors There are three regressors in the model. A constant is not needed since it is unidentified. However, in generation of the regressors for the simulations, one

is included so that a comparison with linear regression diagnostics can be made. For the simulation the regressor matrix

$$X = \{\text{const, age, educ, exper}\}. \quad (9)$$

is decomposed using SVD. $X = UDV$, where D is a diagonal matrix containing the eigenvalues of the original data. These are replaced using the desired ones Λ as in $X = U\Lambda V$.

Four sets of eigenvalues for the SVD are considered. The first makes the regressors mutually orthogonal. Others impart various degrees and types of collinearity. The sets considered are:

Eigenvalues	Collinearity	
$\Lambda = \{1, 1, 1, 1\}$	None	
$\Lambda = \{10, 7, 4, 1\}$	Moderate	
$\Lambda = \{10, 1, 1, 0.1\}$	Moderate	(10)
$\Lambda = \{10, 10, 0.1, 0.1\}$	Moderately Severe	
$\Lambda = \{10, 0.1, 0.05, 0.05\}$	Severe	

To keep the overall variability of the data constant, the eigenvalues are rescaled to have equal length that is similar to the original Mroz data.

Parameters Gretl estimates the model using three cutoff parameters and no constant, which is not identified. Because the identification of the model could be affected by the cut-off parameters, several sets are used. The cutoff parameters are set by taking the mean of $X\beta$ and adding c such that $P(X < c) = p$ with the parameter p chosen from

Cumulative Prob	Bin Size	
$p = \{0.1, 0.5, 0.9\}$	Wide	
$p = \{0.1, 0.3, 0.5\}$	Moderate	(11)
$p = \{0.1, 0.2, 0.3\}$	Narrow	

The parameters from the regression are set to β is set to $\{-0.1, 0.05, -0.02\}$, which mimic the values of the parameters when the original data are used.

Table 2 shows the average condition numbers for each of the collinearity designs.

Table 2: Average condition numbers for the various collinearity designs. Large bins, $p = \{0.1, 0.5, 0.9\}$.

Cond	Collinearity among variables				
	$\{1,1,1,1\}$	$\{10,1,1,0.1\}$	$\{10,0,0.1,0.1\}$	$\{10,.1,.05,.05\}$	
η_1	1.000	1.000	1.000	1.000	1.000
η_2	1.101	1.260	1.568	1.331	1.601
η_3	1.138	1.351	1.831	1.517	1.899
η_4	1.326	1.573	5.270	1.669	17.320
η_5	1.972	2.624	8.134	15.470	62.650
η_l	14.280	18.530	24.280	75.860	171.300

Notice that even when the variables are mutually orthogonal (column 1) there is still evidence of weakening identification. This has to be due to the functional form for ordered probit. As collinearity worsens, the largest condition number increases in magnitude. By the time collinearity is very bad (rightmost column), there are several condition numbers that lie above the BKW suggested threshold of 30.

Bin size also affects the condition numbers. This is shown in Table 3. The collinearity regime for these results is $\Lambda = \{10, 7, 4, 1\}$, which is very mild. The first panel has a small bin at each extreme ($p < 0.1$ and $p > 0.9$), but wide bins in the middle of the distribution. The second panel has one very wide bin at the upper extreme and moderately narrow ones at the lowest. The bottom panel has very narrow bins at the lower end and 70% probability of landing in the last one. One can see that as the bins get narrower at the bottom end, conditioning worsens as η_i increases from 18.53 to 29.41. You also see that as the condition number gets larger, the standard errors of the cutoff parameters gets larger. This is indicated by the large and increasing variance proportions in each of the last three columns of the BKW tables.

Specific subsets of the simulation results appear in Tables 4-7 below. In Table 4 the best case scenario for ordered probit is found. Here, all of the regressors are orthogonal to one another and the bins for the ordered response are centered around 0.5 and relatively wide (at least in the center of the distribution). The summary statistics show that the MLE has low bias (perhaps unbiased) for this design. The overall variability of the estimates is very small. The t-ratio measures the relative precision of the MLE (not its unbiasedness).

The variance proportion table shows evidence of some ill-conditioning that is independent of the data, which are orthogonal. The first six condition numbers increase steadily from one to two and the last is 14.27. This falls within the moderate level for the BKW diagnostics in linear models.

Further, the largest condition number affects each of the parameters, though mostly x_1 and the three cutoff parameters.

The standard errors of the condition numbers and the variance decompositions appear at the bottom panel of each table. Since the regressors are not changing in the simulation, all of the variance is due to the parameter estimates. The relative precision of the parameter estimates is evidenced by the small standard deviations in this panel for all of the statistics computed. It is easy to conclude that parameter variation due to their estimation is not contributing to any significant degree the relative magnitudes of the diagnostics themselves. For instance, comparing the bottom panels of tables 4 and 7 (no collinearity and high collinearity, respectively) more collinear regressors are causing more variance in the condition numbers, but not in the variance proportions. These appear to be fairly invariant to collinearity among regressors. The implication is that the diagnostics themselves change in predictable ways as collinearity changes, but the variation of the measures is fairly constant in each design.

In Table 5 the degree of collinearity in the variables is introduced. The overall level is kept moderately low. The standard errors of the estimated cutoff parameters increases relative to the baseline.

The average largest condition number is 24.31 indicating that identification is becoming more challenging. The high variance proportions for the cutoff parameters is consistent with the higher standard errors found in the upper panel. The collinearity of the variables themselves does not appear to be much of a problem, except possibly for x_1 which shares a large variance proportion with the cutoffs.

Tables 6 and 7 show the effects of increasing collinearity among the regressors. In the first case, only two of the regressors are highly collinear and in the second, all three are collinear with each other. In Table 6 there is a single large condition number that affects estimation of the β s, but not the cutoffs. Relative to the β s, the cutoffs are not affected to a large extent by collinearity of the variables. So, as collinearity of the variables gets worse, identification of their parameters weakens while that of the cutoffs improves, at least relatively.

In Table 7 there are two large condition numbers, with β_1 and β_2 being weakly identified. The other parameter, β_3 , appears to be fairly well identified. In all of these cases, the variation caused by parameter estimation is low.

Table 3: Various bin sizes for a mildly collinear data in Ordered Probit. $\Lambda = 10, 7, 4, 1$. There is only 1 relatively large condition number in this scenario.

Bin Size: $p = 0.1, 0.5, 0.9$						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-3.366	-2.084	-0.803
Mean	-0.099	0.051	-0.020	-3.340	-2.062	-0.769
Std Err	0.012	0.009	0.003	0.377	0.365	0.354
η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}
18.530	0.978	0.634	0.448	0.977	0.989	0.968
Bin Size: $p = 0.1, 0.3, 0.5$						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-3.366	-2.609	-2.084
Mean	-0.099	0.052	-0.020	-3.322	-2.558	-2.043
Std Err	0.013	0.010	0.003	0.392	0.388	0.381
η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}
23.220	0.979	0.619	0.472	0.984	0.995	0.992
Bin Size: $p = 0.1, 0.2, 0.3$						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-3.366	-2.926	-2.609
Mean	-0.099	0.052	-0.020	-3.326	-2.880	-2.562
Std Err	0.014	0.010	0.004	0.429	0.424	0.423
η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}
29.410	0.980	0.594	0.500	0.991	0.997	0.995

Monte Carlo Summary Statistics						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-2.353	-1.071	0.2104
Mean	-0.101	0.051	-0.020	-2.361	-1.073	0.2263
Std Err	0.018	0.008	0.004	0.289	0.281	0.2793
t-ratio	-0.040	0.120	0.014	-0.030	-0.008	0.0569

Averages of the Variance Proportion Table						
η_l	β_1	β_2	β_3	τ_1	τ_2	τ_3
1.0000	0.0058	0.0200	0.0031	0.0027	0.0004	0.0070
1.1010	0.0020	0.0322	0.0043	0.0132	0.0009	0.0037
1.1380	0.0017	0.0014	0.0019	0.0020	0.0153	0.0061
1.3260	0.0013	0.0003	0.3843	0.0001	0.0000	0.0001
1.9690	0.0056	0.1656	0.0034	0.0191	0.0001	0.0257
14.2700	0.9836	0.7805	0.6029	0.9629	0.9832	0.9574

Standard Errors of the Variance Proportion Table						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
-	0.0003	0.0021	0.0005	0.0006	0.0003	0.0006
0.0082	0.0006	0.0037	0.0021	0.0017	0.0014	0.0018
0.0128	0.0010	0.0019	0.0022	0.0016	0.0018	0.0014
0.0107	0.0002	0.0004	0.0048	0.0001	0.0001	0.0001
0.0584	0.0006	0.0101	0.0041	0.0022	0.0001	0.0028
0.1252	0.0006	0.0098	0.0035	0.0029	0.0008	0.0036

Table 4: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with Wide Bins: $p = \{0.1, 0.5, 0.9\}$. Results based on 100 Monte Carlo samples.

Monte Carlo Summary Statistics						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-5.197	-3.916	-2.634
Mean	-0.101	0.049	-0.019	-5.252	-3.943	-2.667
Std Err	0.007	0.014	0.010	0.380	0.364	0.348
t-ratio	-0.119	-0.086	0.126	-0.145	-0.077	-0.096

Averages of the Variance Proportion Table						
η_l	β_1	β_2	β_3	τ_1	τ_2	τ_3
1.0000	0.0016	0.0043	0.0111	0.0006	0.0003	0.0005
1.5680	0.0000	0.0000	0.0001	0.0035	0.0080	0.0047
1.8320	0.0000	0.0009	0.0006	0.0125	0.0001	0.0144
5.2570	0.0035	0.0885	0.7976	0.0111	0.0044	0.0008
8.1320	0.1624	0.4825	0.0942	0.0022	0.0092	0.0208
24.3100	0.8325	0.4238	0.0964	0.9701	0.9780	0.9587

Standard Errors of the Variance Proportion Table						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
-	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
0.0149	0.0000	0.0000	0.0000	0.0006	0.0003	0.0006
0.0045	0.0000	0.0001	0.0002	0.0012	0.0001	0.0015
0.0691	0.0007	0.0057	0.0225	0.0010	0.0004	0.0002
0.1256	0.0072	0.0208	0.0162	0.0003	0.0008	0.0022
0.2551	0.0073	0.0159	0.0095	0.0015	0.0012	0.0032

Table 5: Moderate Collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with Wide Bins: $p=\{0.1, 0.5, 0.9\}$. Results based on 100 Monte Carlo samples.

Monte Carlo Summary Statistics						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-3.650	-2.368	-1.086
Mean	-0.103	0.054	-0.020	-3.706	-2.410	-1.124
Std Err	0.044	0.179	0.004	0.285	0.284	0.287
t-ratio	-0.062	0.023	-0.108	-0.200	-0.147	-0.131

Averages of the Variance Proportion Table						
η_l	β_1	β_2	β_3	τ_1	τ_2	τ_3
1.0000	0.0002	0.0001	0.0000	0.0008	0.0007	0.0007
1.3310	0.0000	0.0000	0.0003	0.0053	0.0064	0.0057
1.5160	0.0000	0.0000	0.0612	0.0068	0.0001	0.0068
1.6720	0.0000	0.0000	0.0612	0.0083	0.0002	0.0118
15.4900	0.0185	0.0031	0.1978	0.6253	0.6094	0.5738
75.8800	0.9813	0.9968	0.6794	0.3534	0.3833	0.4012

Standard Errors of the Variance Proportion Table						
	β_1	β_2	β_3	τ_1	τ_2	τ_3
-	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
0.0107	0.0000	0.0000	0.0003	0.0006	0.0002	0.0007
0.0154	0.0000	0.0000	0.0032	0.0010	0.0001	0.0011
0.0235	0.0000	0.0000	0.0035	0.0013	0.0001	0.0016
0.1385	0.0004	0.0001	0.0044	0.0091	0.0064	0.0038
0.2485	0.0004	0.0001	0.0050	0.0092	0.0065	0.0032

Table 6: Severe Collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with Wide Bins: $p = \{0.1, 0.5, 0.9\}$. Results based on 100 Monte Carlo samples.

	Monte Carlo Summary Statistics					
	β_1	β_2	β_3	τ_1	τ_2	τ_3
True	-0.100	0.050	-0.020	-5.974	-4.692	-3.411
Mean	-0.098	0.027	-0.009	-6.069	-4.771	-3.472
Std Err	0.050	0.191	0.078	0.356	0.344	0.337
t-ratio	0.041	-0.123	0.144	-0.267	-0.228	-0.181

	Averages of the Variance Proportion Table					
η_l	β_1	β_2	β_3	τ_1	τ_2	τ_3
1.0000	0.0000	0.0000	0.0001	0.0005	0.0003	0.0004
1.6050	0.0000	0.0000	0.0000	0.0037	0.0074	0.0049
1.9000	0.0000	0.0000	0.0000	0.0122	0.0001	0.0164
17.3200	0.0010	0.0003	0.0153	0.5755	0.5384	0.4904
62.7300	0.0660	0.0355	0.9750	0.1401	0.1510	0.1537
171.5000	0.9330	0.9643	0.0096	0.2680	0.3028	0.3342

	Standard Errors of the Variance Proportion Table					
-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0142	0.0000	0.0000	0.0000	0.0006	0.0003	0.0007
0.0069	0.0000	0.0000	0.0000	0.0014	0.0001	0.0019
0.2352	0.0001	0.0000	0.0005	0.0167	0.0132	0.0113
0.3206	0.0010	0.0006	0.0013	0.0128	0.0112	0.0099
0.6394	0.0010	0.0006	0.0012	0.0160	0.0134	0.0105

Table 7: Severe Collinearity: $\Lambda = \{10, 0.1, 0.05, 0.05\}$, with Wide Bins: $p = \{0.1, 0.5, 0.9\}$. Results based on 100 Monte Carlo samples.

The variance decomposition diagnostics clearly show that the smallest eigenvalue is wreaking havoc on estimation of the β s, but not so much for the cutoffs which had very substantial average t-ratios due to smallish standard errors. I would tentatively conclude that the ordered probit functional form is fairly robust with respect to identification of parameters. Of course, other designs may reveal otherwise.

3.2 Sample Selectivity

The sample selectivity model has been studied within the context of collinearity between the variable of the regression equation and that of the selection equation (e.g., Nawata and Nagase (1996)) and some have commented that having different regressors in the selection and regression equations improves identification (or that the normality assumption for the likelihood provides weak identification if regressors are not somehow different. Hill et al. (2003) found that using the same variables in both equations distorted the sampling distribution of the two-step estimator. A more specific examination of the effects of collinearity on the MLE, and the corresponding diagnostics, follows in this section of the paper.

The Model

Following (Greene, 2012, pp. 874-876) consider a model consisting of two equations. The first is the “selection equation,” defined

$$z_i^* = w_i' \gamma + u_i, \quad i = 1, \dots, N \quad (12)$$

where z_i^* is a latent variable, γ is a $K \times 1$ vector of parameters, w_i' is a $1 \times K$ row vector of observations on K exogenous variables, and u_i is a random disturbance. The latent variable is unobservable, but we do observe the dichotomous variable

$$z_i = \begin{cases} 1 & z_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

The second equation is the linear model of interest. It is

$$y_i = x_i' \beta + e_i, \quad i = 1, \dots, n, \quad N > n \quad (14)$$

where y_i is an observable random variable, β is an $M \times 1$ vector of parameters, x_i' is a $1 \times M$ vector of exogenous variables, and e_i is a random disturbance. It is assumed that the random disturbances of the two equations are distributed as

$$\begin{bmatrix} u_i \\ e_i \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma_e^2 \end{pmatrix} \right] \quad (15)$$

A selectivity problem arises when y_i is observed only when $z_i = 1$ and $\rho \neq 0$. In this case the ordinary least squares estimator of β in (14) is biased and inconsistent. A consistent estimator has been suggested by Heckman (1979) and is commonly referred to as Heckman's two-step estimator, or more simply, *Heckit*. As computational power has improved, many now use maximum likelihood to estimate the parameters. Since both Heckit and the MLE are derived based on the joint normality of the errors the choice is up to the user, though in finite samples the properties are likely to differ.

The conditional mean of y_i given that it is observed is

$$E[y_i | z_i > 0] = x_i' \beta + \beta_\lambda \lambda_i \quad (16)$$

where $\lambda_i = \phi(w_i' \gamma) / \Phi(w_i' \gamma)$ is the *inverse Mill's ratio*, $\phi(\cdot)$ is the standard normal probability density function evaluated at the argument, and $\Phi(\cdot)$ is the cumulative density function of the standard normal random variable evaluated at the argument.

The log-likelihood (Greene, 2012, p. 878) is

$$\ln L = \sum_{z=1} \left[\frac{\exp(-(1/2)e_i^2/\sigma_e^2)}{\sigma_e \sqrt{2\pi}} \Phi \left(\frac{\rho e_i / \sigma_e + w_i' \gamma}{\sqrt{1 - \rho^2}} \right) \right] + \sum_{z=0} [1 - \ln \Phi(w_i' \gamma)] \quad (17)$$

where $e_i = y_i - x_i' \beta$.

3.3 Example

In this example, the Mroz data are again used to explore the return to schooling for women. A woman may or may not participate in the labor force and work for wages. This leads to a selection event. If in the labor force, the woman's (log) wage can be modeled as a function of education and some controls. Thus, the selection equation includes education, experience, and the number of children less than six years of age. The regression includes only education and experience.

The MLE reveals:

Model 6: ML Heckit, using observations 1–753
 Dependent variable: l_wage
 Selection variable: lfp Standard errors based on Hessian

	Coefficient	Std. Error	z	p-value
const	-0.219618	0.425389	-0.5163	0.6057
educ	0.103330	0.0192341	5.3722	0.0000
exper	0.0126179	0.00758164	1.6643	0.0961
lambda	-0.106730	0.224406	-0.4756	0.6344

Selection equation

const	-1.75508	0.284841	-6.1616	0.0000
educ	0.121601	0.0226111	5.3779	0.0000
exper	0.0546770	0.00671292	8.1450	0.0000
kidsl6	-0.488228	0.0990169	-4.9308	0.0000

Mean dependent var	1.190173	S.D. dependent var	0.723198
$\hat{\sigma}$	0.670926	$\hat{\rho}$	-0.159079
Log-likelihood	-876.3777	Akaike criterion	1760.755
Schwarz criterion	1776.992	Hannan-Quinn	1767.168

Total observations: 753

Censored observations: 325 (43.2%)

In the regression, education is significantly positive at 5% but experience is not. A test of whether the sample selection is endogenous is not significant. The BKW table reveals:

--- variance proportions ---								
cond	const	educ	exper	lambda	const	educ	exper	kidsl6
1.000	0.000	0.001	0.004	0.003	0.000	0.000	0.003	0.002
1.128	0.000	0.000	0.001	0.000	0.003	0.003	0.027	0.025
2.119	0.000	0.000	0.000	0.000	0.001	0.000	0.049	0.840
2.907	0.000	0.000	0.095	0.051	0.000	0.000	0.005	0.004
3.872	0.000	0.000	0.001	0.001	0.019	0.022	0.900	0.114
7.215	0.002	0.127	0.243	0.164	0.001	0.000	0.002	0.000
15.413	0.000	0.003	0.002	0.003	0.976	0.972	0.013	0.013
30.041	0.998	0.869	0.653	0.778	0.001	0.001	0.000	0.000

There is one very large condition number that affects the parameters of the regression equation (but not the selection). In fact the strength of the selection equation is evident. Each variable, other than `educ`, has a single condition number associated with a large variance proportion: `exper` with λ_5 and `kidsl6` with λ_3 ; the constant and `educ` have large variance proportions shared by λ_7 . The insignificance of β_λ may be because it is weakly identified in this specification.

Simulation

There are several things that may affect identification in this model. The usual culprits 1) collinearity among independent variables and 2) the interaction of parameter estimates and independent variables. There are other sources as well: 3) the choice of variables to include in the selection and regression equations—collinearity between x and w and 4) the fact that the independent variables are different for the two equations since the regression observations are only a subset of the selection ones. The difference changes at each iteration of the simulation, which adds another source of variation to the BKW statistics.

Collinear regressors In this example we again consider three regressors in the model plus a constant that will be included in both regression and selection equations. For the simulation the regressor matrix

$$X = \{\text{const, educ, exper, kidsl6}\}. \quad (18)$$

is decomposed using SVD. $X = UDV$, where D is a diagonal matrix containing the eigenvalues of the original data. These are replaced using the desired ones Λ as in $X = U\Lambda V$.

Four sets of eigenvalues for the SVD are considered. The first makes the regressors mutually orthogonal. Others impart various degrees and types of collinearity. The sets considered are:

Eigenvalues	Collinearity
$\Lambda = \{1, 1, 1, 1\}$	None
$\Lambda = \{10, 1, 1, 0.1\}$	Moderate
$\Lambda = \{10, 10, 0.1, 0.1\}$	High
$\Lambda = \{10, 0.1, 0.1, 0.05\}$	Severe

(19)

It is important to know that in this sample selection model, it is extremely unlikely that no collinearity in the population does not imply that subsample regressors used in the regression equation are uncorrelated. The variables in the selection equation may be uncorrelated, while those of the regression equation are probably not. This adds to the magnitudes of the condition numbers and to the variability observed in their computation. The size and characteristics of the subsample change with each iteration of the simulation, which leads to another source of variation in the BKW diagnostics.

Parameters With two equations we have two sets of regressors which may or may not be collinear with one another. So that each equation may have variables within it that are collinear with one another, a constant and two regressors are included in x and in w . This produces a model that will have six regression parameters plus β_λ , which appears in the regression equation to account for the sample selection problem.

To parameterize the regression coefficients, the model is estimated using the original Mroz data. The coefficients from this initial estimation, $\hat{\beta}$ and $\hat{\gamma}$ are used to generate the simulated samples.³

The other parameters in the model include ρ , which determines the level of censoring and σ_e , which determines both the variance of the regression as well as the value of β_λ , which theoretically is equal to $\rho\sigma_e$.

The selection degree is chosen from

$$\rho = \{0, -0.5, -0.99\}$$

and the variance from

$$\sigma_e = \{1, 2\}.$$

Note, the increase in variance in the regression equation has no effect on the collinearity diagnostic results. The main effect is to double the estimated value of λ as well as its standard error. It also doubles the measured standard errors of the regression equation, having no effect on the selection equation. This is verified in the simulations.

Two sets of regressors are chosen based on the collinear induced X obtained above.

These are

$$x_i = 1, \text{educ}_i, \text{exper}_i$$

$$w_i = 1, \text{educ}_i, \text{exper}_i$$

and

$$x_i = 1, \text{educ}_i, \text{exper}_i$$

$$w_i = 1, \text{educ}_i, \text{kidsl6}_i$$

.

This results in 48 possible designs. A subset of these is presented below. The gretl code appears in the appendix.

The standard errors of the condition numbers and the variance decompositions appear at the bottom panel of the tables. Since the regressors are not changing in the simulation, all of the variance is due to parameter estimates. The relative precision of the parameter estimates is evidenced by the small standard deviations in this panel for all of the statistics computed. It is easy to conclude that parameter variation due to their estimation is not contributing to any significant degree the relative magnitudes of the diagnostics themselves.

³The parameters are rounded to two decimal places for neatness

For instance, comparing the bottom panels of tables 9 and 12 (some selection, with no collinearity and high collinearity, respectively) more collinearity causes more variation in the condition numbers, but not in the variance proportions. These appear to be fairly invariant to changes in collinearity among regressors. The implication is that the diagnostics themselves change in predictable ways as collinearity changes, and the quality of the information they reveal does not degrade as collinearity increases.

In general, the diagnostics show that as the degree of selection increases (ρ), its parameter β_λ is better identified. This can be seen by comparing the standard error of $\hat{\beta}_\lambda$ in tables 17, 19 (where $\rho = 0$) tables 9, 11 (where $\rho = -0.5$) and tables 13, 15 (where $\rho = -0.99$). No matter what the largest condition number is, the variance proportion associated with $\hat{\beta}_\lambda$ falls substantially and is no longer associated with estimation of the other slopes of the regression equation as ρ increases in magnitude. Its estimation also benefits from increasing collinearity.

The proportion of samples that cause singularity of the Hessian or fail to converge increase as the condition numbers increase.

In tables 21 - 32 the regressors are the same in the regression and selection equations. Note, since $n < N$ there are some sample differences. The most notable effect on the results is that the MLE of β_λ is now badly biased. When $\rho = -0.5$ the average estimate is roughly half of what it should be. However, as ρ increases, the bias declines. This is summarized in table 8. At $\rho = -0.75$ bias has fallen to about 20%. At $\rho = -0.99$ the bias is almost gone. The variance proportion diagnostic is very good at detecting this. At $\rho = -0.5$ the proportion is almost 98% of the variation associated with the smallest eigenvalue. By the time ρ reaches -0.99 it has fallen to 0.088.

In Table 33 the effects of using different orthogonal variables in regression and selection equations are given. Each regressor is orthogonal to the rest. The panels are arranged from the strongest conditioning to the weakest, which is due only to the choice of variables.⁴ The largest condition number in the first panel is 2.705. In this case the variables in selection and regression equations (aside from the constant) are completely different. The selection equation contains more variables, however. The most interesting feature of this comparison is the fact that the presence of *cexper* in the selection equation helps to identify *cexper* in the regression. That can be seen by comparing the variance proportion of *cexper* in the regression in panels 2, 3, 4 and 5.

When $x=\{1, \text{ceduc}\}$ and $w=\{1, \text{ceduc}, \text{ckidsl6}\}$, $w=\{1, \text{ckidsl6}\}$, or $w=\{1, \text{cexper}\}$ the maximum condition numbers fall in the 2.5 – 2.9 range. Having variables other than

⁴Recall, even though the variables are orthogonal there is some variation in the c_{ij} of the eigenvectors. It is for this reason that the data were based on actual observations from Mroz and converted to orthogonality through the eigenvalues. The eigenvectors from the original data remain intact.

education in the selection equation is good for conditioning in this case.

In the second panel, the largest condition number $c_l = 4.769$ and identification worsens. Surprisingly, the variables in selection and regression are the same and include *ceduc*. However, the coefficient for education and the selectivity parameter β_λ are problematic. High proportions of variance are due to the largest condition number. This does not affect estimation of the selection parameters, however. As the other simulations show, estimation of β_λ is badly biased in this scenario when ρ diverges from zero.

Conditionings in the third and fourth panels are quite similar, and only marginally worse than the second best case above. The addition of *cexper* to the regression (x) increases the diagnostic substantially. The *ceduc* is switched with *cexper* in w , both of which are included in x . This causes a substantial weakening of the identification of the regression parameters. This last case is equivalent to that found in the simulations and presents the worst conditioning of the lot. So, the identification of the model's parameters is sensitive to the orientation of the variables in the model, even if those variables are mutually uncorrelated. This is a surprising finding and points to the possible utility of using the BKW diagnostics in this context.

3.4 Conclusion

1. As in Ordered Probit, orthogonal variables does not mean that all the parameters will be strongly identified. There is interaction with the degree of selection bias and with the choice of variables that can have dramatic effects on condition numbers and the variance decompositions.
2. Identification in the sample selection model depends on collinear variables as well as functional form and parameters. The more collinear the variables in each equation the weaker the identification, other things being equal.
3. As the degree of selection bias increases, β_λ becomes better identified. Its estimation becomes increasingly isolated from the collinear variables in the model.
4. Perfect collinearity between x and w severely biases estimation of $\beta_\lambda \neq 0$. The bias shrinks as $\rho \rightarrow 1$, but is substantial even when $\rho = 0.75$. Otherwise, the effects of $x = w$ in terms of bias and precision are nothing special. The BKW diagnostics can tell you which parameters are likely measured imprecisely.
5. BKW diagnostics useful in diagnosing problems, both as collinearity worsens and as bins size shrinks.
6. As in the Ordered Probit case, the proportion of BKW diagnostics variability due to parameter estimation is minimal.

4 Final Thoughts

To sum things up, the BKW diagnostics appear to be quite useful in identifying problems with nonlinear estimation. Large condition numbers that have large effects on the variances of two or more coefficients can signal issues with the data or the model. Also, higher collinearity among variables may actually be useful if interest in on other parameters in the model. This is counterintuitive and needs some additional investigation.

Table 8: When the regressors are the same in each equation, the estimator of β_λ is biased. As ρ increases in magnitude, the bias diminishes. The diagnostic seems to anticipate this. No collinearity, $x = w$, $\sigma_e = 1$.

ρ		β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
0.000	True	-0.170	0.330	0.270	0.000	-0.050	0.010	0.080
	Mean	-0.170	0.330	0.270	0.003	-0.050	0.009	0.081
	Std Err	0.029	0.051	0.012	0.459	0.009	0.008	0.009
	η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}	ϕ_{l7}
	11.220	0.867	0.977	0.575	0.985	0.004	0.003	0.007
ρ		β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-0.250	True	-0.170	0.330	0.270	-0.250	-0.050	0.010	0.080
	Mean	-0.179	0.314	0.267	-0.105	-0.050	0.009	0.081
	Std Err	0.029	0.051	0.012	0.459	0.009	0.008	0.009
	η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}	ϕ_{l7}
	11.200	0.861	0.976	0.567	0.984	0.004	0.003	0.007
ρ		β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-0.500	True	-0.170	0.330	0.270	-0.500	-0.050	0.010	0.080
	Mean	-0.184	0.305	0.264	-0.263	-0.050	0.009	0.081
	Std Err	0.028	0.050	0.012	0.458	0.009	0.008	0.009
	η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}	ϕ_{l7}
	10.250	0.826	0.968	0.508	0.978	0.004	0.004	0.008
ρ		β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-0.750	True	-0.170	0.330	0.270	-0.750	-0.050	0.010	0.080
	Mean	-0.181	0.311	0.266	-0.571	-0.050	0.009	0.081
	Std Err	0.026	0.045	0.011	0.415	0.009	0.008	0.009
	η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}	ϕ_{l7}
	6.614	0.674	0.928	0.283	0.943	0.004	0.016	0.019
ρ		β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-0.990	True	-0.170	0.330	0.270	-0.990	-0.050	0.010	0.080
	Mean	-0.174	0.323	0.268	-0.924	-0.050	0.009	0.081
	Std Err	0.019	0.034	0.009	0.305	0.008	0.008	0.008
	η	ϕ_{l1}	ϕ_{l2}	ϕ_{l3}	ϕ_{l4}	ϕ_{l5}	ϕ_{l6}	ϕ_{l7}
	4.754	0.501	0.634	0.067	0.326	0.252	0.461	0.114

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	0.280	26.820	17.250	-0.500	-7.650	1.270	-6.660
Mean	-0.086	26.318	16.895	-0.466	-7.762	1.139	-6.830
Std Err	3.344	5.249	3.204	0.290	1.351	1.300	1.350
t-ratio	-0.109	-0.096	-0.111	0.117	-0.083	-0.101	-0.126

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.001	0.013	0.009	0.012	0.004	0.040	0.000
1.233	0.084	0.003	0.004	0.000	0.244	0.010	0.013
1.370	0.005	0.004	0.091	0.002	0.026	0.237	0.127
1.430	0.005	0.001	0.024	0.000	0.026	0.053	0.703
1.619	0.012	0.016	0.039	0.004	0.153	0.517	0.057
1.681	0.103	0.009	0.007	0.000	0.546	0.140	0.046
9.573	0.790	0.955	0.826	0.981	0.003	0.003	0.054

Standard Errors of the Variance Proportion Table							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-	0.001	0.005	0.003	0.005	0.003	0.020	0.001
0.036	0.026	0.002	0.005	0.000	0.050	0.014	0.023
0.024	0.012	0.002	0.066	0.001	0.061	0.084	0.108
0.029	0.004	0.002	0.022	0.000	0.078	0.124	0.242
0.142	0.011	0.012	0.018	0.005	0.115	0.176	0.091
0.174	0.074	0.007	0.014	0.001	0.158	0.086	0.100
3.098	0.092	0.025	0.078	0.011	0.004	0.005	0.061

Table 9: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-16.590	9.280	-3.120	-0.500	-76.420	7.480	-6.310
Mean	-22.166	9.412	-2.957	-0.437	-76.782	7.503	-6.421
Std Err	28.667	1.274	1.307	0.311	13.780	1.120	1.325
t-ratio	-0.195	0.104	0.125	0.201	-0.026	0.021	-0.084

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.001	0.001	0.004	0.000	0.000	0.000
1.541	0.000	0.000	0.000	0.002	0.007	0.007	0.024
2.036	0.000	0.000	0.000	0.001	0.001	0.001	0.927
3.959	0.000	0.005	0.016	0.220	0.000	0.000	0.013
15.450	0.014	0.220	0.197	0.014	0.512	0.513	0.001
17.730	0.052	0.426	0.423	0.166	0.356	0.360	0.011
28.090	0.934	0.348	0.363	0.595	0.124	0.119	0.024

Standard Errors of the Variance Proportion Table							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
-	0.000	0.000	0.000	0.002	0.000	0.000	0.000
0.123	0.000	0.000	0.000	0.002	0.000	0.000	0.012
0.044	0.000	0.000	0.000	0.001	0.000	0.000	0.062
0.253	0.000	0.001	0.004	0.100	0.000	0.000	0.018
0.288	0.033	0.094	0.184	0.038	0.215	0.215	0.004
0.903	0.077	0.251	0.210	0.147	0.195	0.196	0.017
6.681	0.092	0.197	0.258	0.256	0.156	0.151	0.038

Table 10: Moderate Collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

True	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
Mean	-26.750	64.390	-53.110	-0.500	-164.900	15.680	-66.660
Std Err	-37.092	62.656	-50.865	-0.442	-165.721	15.766	-68.315
t-ratio	55.201	13.760	15.126	0.288	28.175	2.282	13.244
	-0.187	-0.126	0.148	0.202	-0.029	0.038	-0.125
Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.003	0.000	0.000	0.005
1.387	0.000	0.000	0.000	0.002	0.001	0.001	0.047
3.448	0.000	0.000	0.000	0.078	0.002	0.002	0.634
3.815	0.000	0.000	0.000	0.150	0.001	0.001	0.265
31.970	0.056	0.001	0.000	0.049	0.785	0.793	0.019
43.720	0.539	0.018	0.005	0.147	0.207	0.198	0.003
233.400	0.405	0.981	0.995	0.570	0.004	0.005	0.027
Standard Errors of the Variance Proportion Table							
-	0.000	0.000	0.000	0.001	0.000	0.000	0.002
0.130	0.000	0.000	0.000	0.002	0.000	0.000	0.004
0.101	0.000	0.000	0.000	0.104	0.001	0.001	0.288
0.196	0.000	0.000	0.000	0.070	0.001	0.001	0.275
1.206	0.058	0.001	0.001	0.045	0.168	0.163	0.010
2.497	0.132	0.004	0.002	0.039	0.169	0.163	0.004
44.250	0.176	0.004	0.003	0.144	0.007	0.008	0.037

Table 11: Severe Collinearity: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0.02. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-11.910	3.730	1.800	-0.500	-75.310	8.950	-105.200
Mean	-15.091	3.918	1.793	-0.464	-75.946	9.023	-106.530
Std Err	23.231	1.418	0.120	0.233	13.742	1.195	9.902
t-ratio	-0.137	0.133	-0.059	0.152	-0.046	0.061	-0.134

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.001	0.004	0.001	0.001	0.007
1.304	0.001	0.001	0.016	0.009	0.001	0.001	0.013
1.885	0.000	0.000	0.648	0.001	0.000	0.000	0.000
4.583	0.002	0.006	0.075	0.238	0.011	0.004	0.341
5.178	0.001	0.009	0.224	0.178	0.014	0.003	0.372
22.330	0.060	0.044	0.010	0.091	0.834	0.868	0.255
32.080	0.937	0.940	0.026	0.480	0.140	0.123	0.013

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.001
0.144	0.000	0.000	0.013	0.005	0.000	0.000	0.002
0.087	0.000	0.000	0.069	0.000	0.000	0.000	0.000
0.256	0.001	0.005	0.078	0.202	0.010	0.003	0.295
0.374	0.001	0.008	0.104	0.146	0.010	0.003	0.285
1.162	0.080	0.064	0.010	0.083	0.168	0.154	0.027
4.592	0.081	0.064	0.050	0.194	0.168	0.154	0.019

Table 12: Severe Collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	0.280	26.820	17.250	-0.990	-7.650	1.270	-6.660
Mean	0.338	26.981	17.258	-0.995	-7.708	1.157	-6.755
Std Err	1.296	1.352	0.776	0.061	1.235	1.256	0.759
t-ratio	0.044	0.119	0.010	-0.074	-0.047	-0.090	-0.125

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.030	0.011	0.006	0.038	0.033	0.013	0.006
1.110	0.015	0.044	0.011	0.018	0.016	0.051	0.013
1.388	0.007	0.003	0.244	0.075	0.008	0.006	0.242
1.574	0.003	0.004	0.421	0.011	0.003	0.004	0.409
2.352	0.015	0.010	0.240	0.480	0.054	0.160	0.245
3.824	0.368	0.345	0.011	0.089	0.547	0.332	0.025
4.699	0.562	0.583	0.066	0.290	0.338	0.435	0.061

Standard Errors of the Variance Proportion Table							
0.000	0.011	0.009	0.007	0.016	0.012	0.010	0.007
0.067	0.014	0.015	0.017	0.024	0.016	0.017	0.017
0.073	0.006	0.004	0.150	0.038	0.008	0.007	0.121
0.112	0.004	0.006	0.187	0.023	0.005	0.005	0.198
0.218	0.008	0.018	0.082	0.108	0.032	0.128	0.115
0.599	0.218	0.230	0.024	0.089	0.213	0.204	0.027
2.608	0.227	0.233	0.079	0.133	0.234	0.219	0.067

Table 13: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0.005. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-16.590	9.280	-3.120	-0.990	-76.420	7.480	-6.310
Mean	-15.896	9.321	-3.198	-0.994	-76.327	7.467	-6.349
Std Err	12.529	0.968	0.479	0.052	12.318	0.993	0.796
t-ratio	0.055	0.043	-0.163	-0.085	0.008	-0.013	-0.049

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.002	0.001	0.000	0.000	0.002
1.913	0.000	0.000	0.000	0.259	0.000	0.000	0.284
2.352	0.001	0.000	0.001	0.206	0.001	0.001	0.525
3.401	0.001	0.003	0.018	0.222	0.008	0.005	0.118
11.290	0.110	0.014	0.524	0.222	0.020	0.037	0.042
14.420	0.037	0.147	0.444	0.084	0.093	0.106	0.026
38.100	0.851	0.835	0.011	0.005	0.878	0.850	0.002

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001
0.079	0.000	0.000	0.000	0.063	0.000	0.000	0.077
0.122	0.000	0.000	0.001	0.066	0.001	0.001	0.100
0.393	0.000	0.000	0.006	0.087	0.002	0.001	0.069
1.372	0.035	0.025	0.242	0.093	0.018	0.026	0.040
2.090	0.063	0.041	0.241	0.093	0.039	0.048	0.035
5.137	0.053	0.032	0.042	0.021	0.033	0.036	0.032

Table 14: Moderate Collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0013. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-26.750	64.390	-53.110	-0.990	-164.900	15.680	-66.660
Mean	-26.128	65.021	-53.729	-0.992	-164.312	15.640	-67.467
Std Err	24.560	4.967	4.629	0.046	24.614	1.979	7.760
t-ratio	0.025	0.127	-0.134	-0.033	0.024	-0.020	-0.104

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.001	0.000	0.000	0.009
2.150	0.000	0.000	0.000	0.455	0.000	0.000	0.013
2.690	0.000	0.000	0.000	0.072	0.000	0.000	0.320
4.333	0.000	0.000	0.000	0.169	0.003	0.002	0.557
25.760	0.139	0.005	0.005	0.260	0.083	0.106	0.089
71.730	0.834	0.018	0.059	0.009	0.867	0.841	0.002
148.000	0.026	0.977	0.936	0.034	0.047	0.049	0.010

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001
0.056	0.000	0.000	0.000	0.089	0.000	0.000	0.022
0.226	0.000	0.000	0.000	0.089	0.000	0.000	0.078
0.421	0.000	0.000	0.000	0.057	0.000	0.000	0.106
3.735	0.040	0.003	0.001	0.070	0.023	0.025	0.067
9.002	0.039	0.003	0.027	0.019	0.039	0.039	0.010
14.440	0.023	0.004	0.028	0.029	0.041	0.042	0.022

Table 15: Severe Collinearity: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0.008. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-11.910	3.730	1.800	-0.990	-75.310	8.950	-105.200
Mean	-11.873	3.742	1.795	-0.991	-75.620	8.990	-106.032
Std Err	11.638	0.858	0.091	0.043	11.783	0.977	7.230
t-ratio	0.003	0.014	-0.050	-0.023	-0.026	0.041	-0.115

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.001	0.004	0.000	0.000	0.003
2.005	0.000	0.000	0.330	0.075	0.000	0.000	0.004
2.319	0.000	0.000	0.103	0.379	0.000	0.000	0.003
3.531	0.004	0.005	0.105	0.128	0.001	0.001	0.030
7.704	0.000	0.003	0.373	0.096	0.023	0.003	0.571
20.180	0.142	0.114	0.008	0.317	0.074	0.154	0.329
52.860	0.854	0.878	0.080	0.001	0.901	0.841	0.059

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.001
0.098	0.000	0.000	0.097	0.050	0.000	0.000	0.002
0.129	0.000	0.000	0.079	0.078	0.000	0.000	0.003
0.375	0.000	0.000	0.041	0.090	0.000	0.000	0.010
0.936	0.001	0.001	0.066	0.050	0.008	0.001	0.084
2.348	0.033	0.030	0.015	0.081	0.025	0.024	0.082
6.891	0.033	0.030	0.021	0.002	0.025	0.025	0.022

Table 16: High Collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0.01. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	0.280	26.820	17.250	0.000	-7.650	1.270	-6.660
Mean	0.304	26.938	17.292	-0.004	-7.770	1.147	-6.765
Std Err	4.277	7.032	4.173	0.390	1.350	1.298	1.380
t-ratio	0.006	0.017	0.010	-0.010	-0.089	-0.095	-0.076

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.001	0.010	0.009	0.010	0.002	0.021	0.000
1.266	0.081	0.002	0.005	0.000	0.219	0.008	0.021
1.359	0.010	0.003	0.040	0.001	0.068	0.289	0.211
1.408	0.005	0.002	0.027	0.000	0.069	0.153	0.500
1.509	0.009	0.008	0.034	0.001	0.193	0.364	0.116
1.547	0.049	0.006	0.013	0.000	0.447	0.164	0.116
10.820	0.845	0.970	0.871	0.987	0.002	0.001	0.036

Standard Errors of the Variance Proportion Table							
0.000	0.001	0.004	0.003	0.004	0.002	0.020	0.000
0.037	0.028	0.002	0.006	0.000	0.075	0.020	0.032
0.023	0.017	0.002	0.045	0.001	0.096	0.092	0.140
0.027	0.005	0.002	0.023	0.000	0.112	0.203	0.309
0.117	0.010	0.007	0.022	0.002	0.141	0.204	0.137
0.142	0.050	0.005	0.019	0.000	0.178	0.099	0.171
2.581	0.062	0.015	0.051	0.006	0.004	0.003	0.064

Table 17: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with no selectivity: $\rho = 0.0$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-16.590	9.280	-3.120	0.000	-76.420	7.480	-6.310
Mean	-18.021	9.345	-3.097	0.014	-76.870	7.510	-6.362
Std Err	35.173	1.413	1.573	0.395	13.835	1.125	1.361
t-ratio	-0.041	0.046	0.014	0.034	-0.033	0.027	-0.038
Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.001	0.001	0.003	0.000	0.000	0.000
1.453	0.000	0.000	0.000	0.001	0.007	0.007	0.018
2.003	0.000	0.000	0.000	0.000	0.001	0.000	0.944
4.031	0.000	0.005	0.015	0.161	0.000	0.000	0.007
15.510	0.005	0.171	0.114	0.004	0.659	0.660	0.001
17.430	0.017	0.578	0.346	0.078	0.291	0.292	0.006
31.520	0.978	0.245	0.524	0.754	0.042	0.041	0.025
Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.102	0.000	0.000	0.000	0.001	0.000	0.000	0.007
0.040	0.000	0.000	0.000	0.001	0.000	0.000	0.068
0.235	0.000	0.001	0.004	0.071	0.000	0.000	0.012
0.217	0.010	0.113	0.115	0.009	0.224	0.224	0.001
0.914	0.037	0.230	0.166	0.100	0.208	0.208	0.013
11.350	0.042	0.148	0.211	0.168	0.086	0.083	0.055

Table 18: Moderate Collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-26.750	64.390	-53.110	0.000	-164.900	15.680	-66.660
Mean	-30.241	64.056	-52.574	0.016	-166.163	15.796	-67.756
Std Err	66.241	15.856	17.715	0.355	28.216	2.286	13.453
t-ratio	-0.053	-0.021	0.030	0.046	-0.045	0.051	-0.081

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.003	0.000	0.000	0.003
1.273	0.000	0.000	0.000	0.001	0.001	0.001	0.046
3.403	0.000	0.000	0.000	0.021	0.003	0.003	0.813
3.780	0.000	0.000	0.000	0.146	0.000	0.000	0.101
31.400	0.017	0.000	0.000	0.016	0.911	0.916	0.015
42.640	0.458	0.017	0.003	0.150	0.081	0.077	0.002
257.500	0.525	0.982	0.997	0.664	0.003	0.004	0.021

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.002
0.101	0.000	0.000	0.000	0.001	0.000	0.000	0.003
0.082	0.000	0.000	0.000	0.048	0.001	0.001	0.195
0.231	0.000	0.000	0.000	0.058	0.001	0.001	0.178
0.992	0.028	0.001	0.000	0.024	0.105	0.101	0.007
2.747	0.130	0.004	0.001	0.042	0.105	0.100	0.004
54.230	0.145	0.004	0.002	0.112	0.005	0.006	0.038

Table 19: Severe Collinearity: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0.025. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	-11.910	3.730	1.800	0.000	-75.310	8.950	-105.200
Mean	-13.120	3.805	1.797	0.013	-75.907	9.017	-106.396
Std Err	28.156	1.666	0.130	0.304	13.892	1.209	9.995
t-ratio	-0.043	0.045	-0.024	0.043	-0.043	0.055	-0.120

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.002	0.004	0.001	0.001	0.007
1.137	0.000	0.000	0.006	0.005	0.001	0.001	0.012
1.779	0.000	0.000	0.650	0.001	0.000	0.000	0.000
4.468	0.001	0.003	0.046	0.091	0.018	0.006	0.553
4.993	0.001	0.013	0.259	0.247	0.006	0.002	0.171
21.090	0.009	0.006	0.006	0.021	0.947	0.968	0.252
32.860	0.988	0.977	0.032	0.633	0.027	0.023	0.005

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.001
0.113	0.000	0.000	0.005	0.002	0.000	0.000	0.002
0.077	0.000	0.000	0.080	0.000	0.000	0.000	0.000
0.220	0.001	0.004	0.082	0.147	0.009	0.003	0.271
0.455	0.001	0.008	0.111	0.135	0.009	0.003	0.260
1.160	0.028	0.021	0.008	0.038	0.063	0.056	0.020
5.753	0.029	0.023	0.062	0.120	0.063	0.056	0.013

Table 20: High Collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} \neq w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	5.710	36.600	18.920	-0.500	-7.750	1.520	12.610
Mean	3.270	32.160	17.890	-0.248	-7.808	1.410	12.807
Std Err	4.631	8.429	2.020	0.464	1.392	1.276	1.488
t-ratio	-0.527	-0.527	-0.510	0.543	-0.042	-0.086	0.132

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.002	0.011	0.011	0.012	0.006	0.039	0.005
1.231	0.056	0.005	0.013	0.000	0.197	0.020	0.066
1.309	0.022	0.000	0.128	0.000	0.066	0.025	0.301
1.537	0.009	0.002	0.028	0.007	0.108	0.625	0.114
1.642	0.042	0.011	0.130	0.000	0.269	0.268	0.081
1.693	0.043	0.001	0.155	0.001	0.350	0.019	0.425
10.530	0.826	0.970	0.536	0.979	0.004	0.004	0.008

Standard Errors of the Variance Proportion Table							
0.000	0.002	0.006	0.004	0.007	0.005	0.021	0.003
0.037	0.022	0.004	0.012	0.000	0.046	0.013	0.026
0.024	0.013	0.001	0.068	0.000	0.086	0.021	0.069
0.136	0.007	0.003	0.022	0.009	0.046	0.077	0.051
0.171	0.038	0.010	0.109	0.001	0.140	0.074	0.110
0.172	0.065	0.002	0.094	0.001	0.144	0.025	0.103
4.606	0.110	0.024	0.206	0.018	0.006	0.006	0.012

Table 21: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	30.650	10.920	-6.980	-0.500	-81.540	1.030	6.370
Mean	9.715	10.295	-5.397	-0.256	-81.960	0.911	6.516
Std Err	39.886	1.864	3.154	0.454	14.520	1.360	1.040
t-ratio	-0.525	-0.335	0.502	0.538	-0.029	-0.087	0.140

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
0.998	0.000	0.001	0.000	0.002	0.001	0.000	0.001
1.336	0.000	0.000	0.000	0.002	0.003	0.002	0.004
3.320	0.000	0.001	0.003	0.112	0.000	0.000	0.000
11.850	0.003	0.001	0.000	0.019	0.444	0.004	0.641
16.080	0.048	0.159	0.008	0.000	0.249	0.508	0.205
20.980	0.203	0.415	0.049	0.040	0.278	0.432	0.119
44.200	0.743	0.421	0.937	0.823	0.023	0.050	0.028

Standard Errors of the Variance Proportion Table							
0.050	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.175	0.000	0.000	0.000	0.002	0.000	0.000	0.000
0.239	0.000	0.000	0.001	0.073	0.000	0.000	0.000
0.994	0.006	0.002	0.001	0.029	0.044	0.004	0.071
0.873	0.034	0.070	0.011	0.002	0.127	0.181	0.068
2.583	0.215	0.170	0.131	0.118	0.113	0.187	0.054
17.800	0.247	0.203	0.146	0.212	0.067	0.139	0.058

Table 22: Orthogonal Regressors: $\Lambda = \{10, 1, 1, 0.1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	66.550	103.710	-95.090	-0.500	-161.720	-44.790	54.450
Mean	24.974	88.014	-78.089	-0.259	-163.389	-45.970	55.667
Std Err	77.910	31.639	33.474	0.435	28.399	10.279	9.629
t-ratio	-0.534	-0.496	0.508	0.553	-0.059	-0.115	0.126

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000
1.323	0.000	0.000	0.000	0.002	0.001	0.000	0.000
3.294	0.000	0.000	0.000	0.114	0.000	0.000	0.000
26.410	0.022	0.000	0.000	0.012	0.815	0.010	0.009
38.600	0.310	0.005	0.001	0.030	0.179	0.001	0.003
148.400	0.017	0.002	0.003	0.025	0.001	0.958	0.959
460.800	0.650	0.993	0.995	0.815	0.004	0.031	0.029

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.159	0.000	0.000	0.000	0.002	0.000	0.000	0.000
0.178	0.000	0.000	0.000	0.075	0.000	0.000	0.000
0.956	0.025	0.000	0.000	0.016	0.145	0.002	0.002
2.809	0.161	0.003	0.001	0.014	0.145	0.001	0.002
9.389	0.021	0.006	0.009	0.037	0.001	0.054	0.052
170.000	0.202	0.008	0.010	0.137	0.006	0.054	0.052

Table 23: Orthogonal Regressors: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	29.470	1.510	1.780	-0.500	-78.700	6.010	1.570
Mean	8.735	2.618	1.758	-0.254	-79.339	6.047	1.591
Std Err	39.406	2.172	0.119	0.458	13.878	1.048	0.164
t-ratio	-0.526	0.510	-0.186	0.538	-0.046	0.035	0.130

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.003	0.004	0.000	0.000	0.000
1.391	0.000	0.000	0.047	0.002	0.003	0.003	0.033
1.649	0.000	0.000	0.229	0.002	0.001	0.000	0.292
2.064	0.000	0.000	0.451	0.000	0.000	0.000	0.506
5.100	0.002	0.014	0.207	0.270	0.000	0.000	0.002
20.670	0.031	0.020	0.008	0.063	0.893	0.894	0.158
41.840	0.967	0.965	0.056	0.659	0.103	0.102	0.009

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000
0.126	0.000	0.000	0.069	0.002	0.001	0.001	0.061
0.041	0.000	0.000	0.101	0.004	0.001	0.001	0.112
0.231	0.000	0.000	0.100	0.001	0.000	0.000	0.066
0.519	0.001	0.006	0.063	0.146	0.000	0.000	0.005
0.994	0.053	0.039	0.008	0.076	0.145	0.144	0.022
15.720	0.054	0.041	0.087	0.223	0.145	0.144	0.015

Table 24: High collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with moderate selectivity: $\rho = -0.5$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	5.710	36.600	18.920	-0.990	-7.750	1.520	12.610
Mean	4.838	35.037	18.501	-0.896	-7.797	1.422	12.730
Std Err	3.528	6.288	1.659	0.346	1.226	1.231	1.266
t-ratio	-0.247	-0.248	-0.253	0.272	-0.038	-0.080	0.095

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.013	0.017	0.019	0.032	0.014	0.020	0.018
1.119	0.038	0.013	0.018	0.026	0.051	0.015	0.022
1.294	0.011	0.032	0.077	0.009	0.011	0.039	0.073
2.071	0.023	0.002	0.059	0.507	0.055	0.141	0.013
3.279	0.088	0.108	0.566	0.025	0.128	0.148	0.531
3.670	0.321	0.193	0.175	0.062	0.488	0.194	0.224
4.847	0.505	0.635	0.087	0.340	0.251	0.443	0.118

Standard Errors of the Variance Proportion Table							
0.000	0.011	0.009	0.008	0.021	0.012	0.011	0.009
0.071	0.024	0.016	0.020	0.030	0.056	0.018	0.026
0.131	0.013	0.016	0.045	0.019	0.015	0.015	0.074
0.261	0.016	0.007	0.030	0.177	0.031	0.153	0.031
0.678	0.190	0.131	0.286	0.034	0.232	0.131	0.309
0.813	0.215	0.184	0.225	0.064	0.235	0.174	0.230
1.216	0.224	0.218	0.135	0.223	0.216	0.233	0.129

Table 25: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.05. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	30.650	10.920	-6.980	-0.990	-81.540	1.030	6.370
Mean	21.137	10.617	-6.252	-0.869	-82.455	0.985	6.480
Std Err	34.476	1.525	2.669	0.381	12.394	1.207	0.881
t-ratio	-0.276	-0.198	0.273	0.318	-0.074	-0.037	0.125

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001
2.070	0.001	0.000	0.000	0.302	0.001	0.001	0.001
3.125	0.001	0.002	0.005	0.178	0.004	0.002	0.002
11.200	0.042	0.013	0.042	0.280	0.097	0.016	0.166
13.860	0.050	0.079	0.034	0.113	0.086	0.121	0.085
27.780	0.408	0.059	0.589	0.040	0.339	0.056	0.500
42.200	0.498	0.846	0.329	0.086	0.473	0.804	0.245

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.282	0.000	0.000	0.000	0.124	0.001	0.001	0.001
0.351	0.000	0.000	0.001	0.074	0.001	0.001	0.001
1.247	0.028	0.018	0.028	0.159	0.117	0.023	0.170
2.220	0.034	0.056	0.030	0.145	0.079	0.143	0.082
4.673	0.083	0.153	0.197	0.038	0.055	0.139	0.144
6.675	0.092	0.194	0.219	0.246	0.163	0.268	0.078

Table 26: Moderate collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.07. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	66.550	103.710	-95.090	-0.990	-161.720	-44.790	54.450
Mean	52.814	98.225	-89.210	-0.902	-163.173	-45.402	55.127
Std Err	61.773	24.289	25.729	0.329	25.083	9.368	8.741
t-ratio	-0.222	-0.226	0.229	0.268	-0.058	-0.065	0.077

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
2.117	0.000	0.000	0.000	0.315	0.000	0.000	0.000
3.162	0.001	0.000	0.000	0.215	0.001	0.000	0.000
23.150	0.077	0.001	0.001	0.134	0.172	0.002	0.002
66.840	0.844	0.012	0.008	0.008	0.822	0.008	0.012
131.800	0.036	0.075	0.093	0.239	0.004	0.204	0.209
360.300	0.043	0.911	0.898	0.088	0.001	0.785	0.777

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.256	0.000	0.000	0.000	0.117	0.000	0.000	0.000
0.368	0.000	0.000	0.000	0.077	0.000	0.000	0.000
3.039	0.023	0.000	0.000	0.059	0.194	0.002	0.002
12.350	0.123	0.002	0.002	0.014	0.194	0.002	0.003
20.580	0.022	0.027	0.031	0.092	0.014	0.211	0.209
59.290	0.140	0.028	0.033	0.196	0.002	0.211	0.208

Table 27: Severe collinearity: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.105. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	29.470	1.510	1.780	-0.990	-78.700	6.010	1.570
Mean	20.727	1.977	1.771	-0.881	-79.029	6.024	1.584
Std Err	31.753	1.779	0.118	0.360	12.291	0.922	0.140
t-ratio	-0.275	0.263	-0.074	0.304	-0.027	0.015	0.101

Averages of the Variance Proportion Table							
η_l	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.001	0.008	0.000	0.000	0.001
1.452	0.000	0.000	0.075	0.017	0.000	0.000	0.067
2.161	0.000	0.000	0.054	0.411	0.001	0.001	0.035
3.482	0.003	0.005	0.103	0.166	0.003	0.002	0.074
4.442	0.000	0.002	0.634	0.105	0.000	0.000	0.630
17.480	0.107	0.086	0.006	0.238	0.171	0.185	0.105
48.440	0.889	0.907	0.127	0.055	0.824	0.811	0.088

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.001
0.092	0.000	0.000	0.031	0.013	0.001	0.001	0.029
0.207	0.000	0.000	0.076	0.151	0.001	0.001	0.095
0.603	0.001	0.002	0.161	0.091	0.001	0.001	0.166
0.569	0.001	0.005	0.183	0.107	0.001	0.001	0.222
2.205	0.034	0.029	0.004	0.096	0.221	0.218	0.038
8.684	0.034	0.027	0.033	0.154	0.221	0.218	0.021

Table 28: High collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with high selectivity: $\rho = -0.99$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.032. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	5.710	36.600	18.920	0.000	-7.750	1.520	12.610
Mean	5.686	36.630	18.916	-0.001	-7.793	1.412	12.788
Std Err	4.677	8.413	2.113	0.459	1.396	1.270	1.488
t-ratio	-0.005	0.004	-0.002	-0.002	-0.030	-0.085	0.120

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.001	0.009	0.011	0.010	0.004	0.029	0.004
1.247	0.053	0.004	0.009	0.000	0.194	0.014	0.067
1.317	0.020	0.000	0.100	0.000	0.091	0.029	0.329
1.482	0.007	0.001	0.028	0.004	0.104	0.643	0.098
1.570	0.030	0.007	0.124	0.000	0.252	0.260	0.051
1.620	0.024	0.001	0.126	0.000	0.350	0.023	0.445
11.660	0.866	0.978	0.603	0.985	0.004	0.003	0.007

Standard Errors of the Variance Proportion Table							
0.000	0.001	0.005	0.004	0.005	0.004	0.022	0.003
0.037	0.021	0.003	0.010	0.000	0.059	0.013	0.029
0.022	0.012	0.000	0.068	0.000	0.120	0.023	0.069
0.121	0.006	0.002	0.025	0.006	0.055	0.095	0.055
0.156	0.034	0.007	0.097	0.000	0.144	0.097	0.084
0.157	0.042	0.001	0.090	0.001	0.132	0.030	0.079
4.653	0.084	0.017	0.178	0.012	0.009	0.005	0.012

Table 29: Orthogonal Regressors: $\Lambda = \{1, 1, 1, 1\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	30.650	10.920	-6.980	0.000	-81.540	1.030	6.370
Mean	30.885	10.970	-7.026	-0.008	-81.997	0.912	6.519
Std Err	40.453	1.879	3.170	0.456	14.015	1.354	0.982
t-ratio	0.006	0.027	-0.015	-0.017	-0.033	-0.087	0.151

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001
1.271	0.000	0.000	0.000	0.001	0.003	0.002	0.004
3.300	0.000	0.001	0.002	0.086	0.000	0.000	0.000
11.560	0.001	0.000	0.000	0.010	0.446	0.004	0.655
16.200	0.037	0.139	0.006	0.000	0.300	0.582	0.216
20.060	0.142	0.434	0.021	0.013	0.242	0.392	0.111
47.600	0.820	0.424	0.971	0.889	0.009	0.019	0.013

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.149	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.184	0.000	0.000	0.001	0.053	0.000	0.000	0.000
0.791	0.003	0.001	0.000	0.018	0.033	0.003	0.050
0.284	0.029	0.077	0.007	0.001	0.133	0.198	0.062
2.347	0.139	0.151	0.066	0.062	0.123	0.194	0.054
16.220	0.163	0.187	0.071	0.119	0.035	0.070	0.031

Table 30: Moderate collinearity: $\Lambda = \{10, 1, 1, 0.1\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	66.550	103.710	-95.090	0.000	-161.720	-44.790	54.450
Mean	67.217	104.690	-96.025	-0.010	-163.253	-46.064	55.743
Std Err	78.777	32.051	33.817	0.435	28.630	10.311	9.679
t-ratio	0.008	0.031	-0.028	-0.023	-0.054	-0.124	0.134

Averages of the Variance Proportion Table							
η_i	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
1.254	0.000	0.000	0.000	0.001	0.001	0.000	0.000
3.256	0.000	0.000	0.000	0.087	0.000	0.000	0.000
26.100	0.012	0.000	0.000	0.007	0.878	0.010	0.010
37.450	0.255	0.004	0.001	0.027	0.117	0.001	0.002
144.500	0.010	0.001	0.001	0.013	0.001	0.972	0.973
495.200	0.723	0.996	0.998	0.864	0.004	0.017	0.015

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.145	0.000	0.000	0.000	0.001	0.000	0.000	0.000
0.187	0.000	0.000	0.000	0.054	0.000	0.000	0.000
0.965	0.018	0.000	0.000	0.011	0.125	0.001	0.002
2.721	0.127	0.002	0.001	0.012	0.124	0.001	0.002
9.128	0.016	0.002	0.004	0.025	0.001	0.030	0.029
158.800	0.157	0.004	0.004	0.097	0.006	0.031	0.029

Table 31: Severe collinearity: $\Lambda = \{10, 0.1, 0.1, 0.05\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.089. Results based on 400 Monte Carlo samples.

Monte Carlo Summary Statistics							
	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
True	29.470	1.510	1.780	0.000	-78.700	6.010	1.570
Mean	29.430	1.524	1.782	-0.005	-79.192	6.036	1.588
Std Err	39.823	2.213	0.126	0.457	13.972	1.054	0.164
t-ratio	-0.001	0.006	0.016	-0.012	-0.035	0.025	0.111

Averages of the Variance Proportion Table							
η	β_1	β_2	β_3	β_λ	γ_1	γ_2	γ_3
1.000	0.000	0.000	0.003	0.003	0.000	0.000	0.000
1.337	0.000	0.000	0.026	0.001	0.003	0.003	0.017
1.648	0.000	0.000	0.262	0.001	0.000	0.000	0.332
1.966	0.000	0.000	0.427	0.000	0.000	0.000	0.482
5.185	0.001	0.013	0.222	0.219	0.000	0.000	0.001
20.390	0.013	0.008	0.007	0.035	0.946	0.946	0.162
43.510	0.985	0.978	0.052	0.740	0.050	0.050	0.005

Standard Errors of the Variance Proportion Table							
0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000
0.121	0.000	0.000	0.052	0.002	0.001	0.001	0.044
0.030	0.000	0.000	0.092	0.002	0.001	0.001	0.106
0.214	0.000	0.000	0.102	0.000	0.000	0.000	0.080
0.492	0.001	0.006	0.056	0.114	0.000	0.000	0.004
1.028	0.033	0.023	0.009	0.058	0.095	0.094	0.019
15.010	0.034	0.025	0.085	0.170	0.095	0.094	0.010

Table 32: High collinearity: $\Lambda = \{10, 10, 0.1, 0.1\}$, with no selectivity: $\rho = 0$. $\sigma_e = 1$ and $x_{i2} = w_{i2}$. Failure proportion=0.002. Results based on 400 Monte Carlo samples.

Table 33: Average condition numbers and variance proportions for models having no collinearity and no selection bias.

λ	cond	Regression				Selection			
		cconst	ceduc	cexper	β_λ	cconst	ceduc	cexper	ckidsl6
1.717	1.000	0.062	0.065		0.110	0.080		0.000	0.000
1.473	1.080	0.131	0.095		0.029	0.136		0.017	0.012
1.073	1.265	0.000	0.000		0.000	0.000		0.442	0.418
0.963	1.335	0.021	0.047		0.010	0.001		0.371	0.447
0.539	1.785	0.388	0.017		0.039	0.766		0.049	0.041
0.235	2.705	0.396	0.776		0.811	0.017		0.121	0.082
λ	cond	cconst	ceduc	cexper	β_λ	cconst	ceduc	cexper	ckidsl6
2.097	1.000	0.005	0.032	0.020	0.036	0.012	0.065	0.006	
1.578	1.153	0.108	0.015	0.006	0.000	0.174	0.033	0.018	
1.313	1.264	0.015	0.000	0.251	0.006	0.011	0.004	0.276	
0.748	1.675	0.001	0.008	0.121	0.031	0.071	0.607	0.113	
0.608	1.857	0.046	0.044	0.404	0.002	0.076	0.268	0.289	
0.564	1.928	0.216	0.001	0.120	0.006	0.651	0.001	0.172	
0.092	4.769	0.608	0.900	0.078	0.920	0.004	0.023	0.128	
λ	cond	cconst	ceduc	cexper	β_λ	cconst	ceduc	cexper	ckidsl6
1.899	1.000	0.000	0.023	0.035	0.023	0.000	0.001	0.001	0.000
1.165	1.277	0.144	0.008	0.001	0.002	0.149	0.001	0.060	0.010
1.087	1.321	0.049	0.001	0.000	0.001	0.331	0.107	0.131	0.132
1.054	1.342	0.007	0.000	0.014	0.000	0.003	0.182	0.409	0.296
0.976	1.395	0.001	0.000	0.011	0.000	0.002	0.620	0.003	0.371
0.919	1.438	0.001	0.011	0.556	0.002	0.044	0.053	0.001	0.012
0.856	1.489	0.003	0.000	0.039	0.000	0.469	0.035	0.395	0.178
0.043	6.630	0.794	0.957	0.344	0.971	0.001	0.001	0.001	0.001
λ	cond	cconst	ceduc	cexper	β_λ	cconst	ceduc	cexper	ckidsl6
1.904	1.000	0.000	0.022	0.034	0.022	0.000		0.001	0.000
1.163	1.280	0.139	0.006	0.002	0.002	0.170		0.071	0.010
1.073	1.332	0.066	0.001	0.002	0.002	0.294		0.313	0.039
1.033	1.358	0.000	0.000	0.010	0.000	0.032		0.221	0.687
0.925	1.435	0.001	0.009	0.550	0.001	0.070		0.003	0.063
0.862	1.486	0.004	0.000	0.066	0.000	0.434		0.390	0.200
0.041	6.796	0.790	0.961	0.336	0.973	0.001		0.001	0.001
λ	cond	cconst	ceduc	cexper	β_λ	cconst	ceduc	cexper	ckidsl6
2.004	1.000	0.000	0.005	0.007	0.005	0.000	0.000		0.000
1.157	1.316	0.059	0.005	0.002	0.000	0.018	0.000		0.002
1.071	1.368	0.000	0.000	0.000	0.000	0.315	0.232		0.371
0.977	1.432	0.001	0.000	0.000	0.000	0.192	0.749		0.070
0.949	1.453	0.000	0.000	0.000	0.000	0.468	0.016		0.557
0.832	1.552	0.016	0.010	0.043	0.000	0.002	0.002		0.000
0.009	14.665	0.923	0.980	0.948	0.995	0.005	0.000		0.000

5 Sample Code for Ordered Probit Simulation

```
set echo off
set messages off
function matrix diagrv(matrix L)
    it = ones(cols(L),1)
    matrix m = it * L .* I(cols(L))
    return m
end function
function string lam_names (scalar k)
    string lall = null
    loop for i=1..k --quiet
        sprintf l "%d",i
        string ls = "@l "
        string lall += ls
    endloop
    return lall
end function

function matrix bkvar_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

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

function matrix cond_num (matrix vc)
    matrix lambda = svd(vc)
    scalar ll = max(lambda)
```

```

    matrix condnumber = sqrt(11./lambda)
    matrix rc = rcond(vc)
    return condnumber
end function

function scalar relative_cn (matrix vc)
    matrix scaled_vc = scale_cov(vc)
    scalar rc = rcond(vc)
    printf "Relative condition number based on scaled inverse covariance \n%10.4e\n\n"
    return rc
end function

function matrix bkw_table (matrix l,
    matrix p,
    string vn)
    string lnames = lam_names(nelem(l))
    string lname = strstr(lnames, ",", " ")
    string vname = strstr(vn, ",", " ")
    matrix p = l'~p
    colnames(p, " cond @vname " )
    rownames(p, " @lname ")
    # printf "\n\nThe BKW variance decomposition \n%10.3f\n\n ",p
    return p
end function

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

open mroz
probit kidsl6 const age educ exper --quiet
matrix beta = $coeff
list xlist = const age educ exper
matrix X = { const, age, educ, exper }
matrix U V

```

```

matrix L = svd(X,&U,&V)

# Pick one of these
matrix eigs = {1,1,1,1}
matrix eigs = {10,1,1,0.1}
matrix eigs = {10, .1, .05, .05}
matrix eigs = {10,10,0.1,0.1}
matrix eigs = { 10, 7, 4, 1}

#cut-offs
matrix p = {.1, .5, .9}
matrix p = {.4, .7, .9}
matrix p = {.1, .2, .3}

matrix eigs = (sumr(L)/sumr(eigs)).*eigs
printf "Eigenvalues %12.4f \n", eigs
matrix mm = diagrv(eigs)
matrix xs = U * mm * V
matrix xsx = xs'*xs

# put the regressors into a data set
scalar k = nelem(xlist)
list std_var = null
loop foreach i xlist --quiet
    series c$i = xs[,i]
endloop

# For reference use the data for OLS and OProbit
list xlist = const age educ exper
ols kidsl6 xlist
matrix co = $vcv
string vn = "const educ exper age"
matrix bb = bkw(co, vn)
bb

list xlist_collin = cconst cage ceduc cexper
ols kidsl6 xlist_collin
matrix co = $vcv
string vn = "const educ exper age"
matrix bb = bkw(co, vn)
bb

```

```

list xlist_collin = cage ceduc cexper
probit kidsl6 xlist_collin
matrix beta = $coeff
string vn = "const educ exper age"
matrix co = $vcv
matrix bb = bkw(co, vn)
bb

# verify that xtx is Identity when eigenvalues are equal
matrix X = { cage, ceduc, cexper }
# matrix XsX = X'*X
# XsX
# create the cut offs, taus
matrix b = { -.1 , .05 , -.02 }
matrix xb = X*b'
xb_bar = mean(xb)

scalar tau1 = xb_bar + invcdf(z,p[1])
scalar tau2 = xb_bar + invcdf(z,p[2])
scalar tau3 = xb_bar + invcdf(z,p[3])
matrix taus = tau1 ~ tau2 ~ tau3
taus

scalar NMC = 100
probit kidsl6 xlist_collin
scalar k=nelem($coeff)

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

matrix bkw_dat = zeros(NMC,k*(k+1))
matrix op_coeff = zeros(NMC,k)
matrix op_coef_ls = zeros(NMC,nelem(xlist_collin)+1)

loop i=1..NMC --quiet
    series ystar = xb + normal(0,1)
    series y1 = (ystar < tau1)*0
    series y2 = ((ystar>tau1)&&(ystar<tau2))*1
    series y3 = ((ystar>tau2)&&(ystar<tau3))*2
    series y4 = (ystar>tau3)*3

```

```

series y = y1 + y2 + y3 + y4
probit y xlist_collin --quiet
# save the estimated coefficients
matrix co = $vcv
matrix bk = bkw(co,vn)
matrix bkw_dat[i,] = (vec(bk))'
matrix op_coef[i,]=($coeff)'
endloop

stats=meanc(bkw_dat)
sd_stuff=sd_c(bkw_dat)

coef = meanc(op_coef)
sd_coef=sd_c(op_coef)
# printf "true values\n %12.5f\n", b~taus
# printf "average of coefficients\n %12.5f\n", coef
# printf "std deviation of coefficients\n %12.5g\n", sd_coef
# printf "t-ratio of coefficients\n %12.5g\n", coef./sd_coef
matrix result = (b~taus)|coef|sd_coef|(coef-(b~taus))./sd_coef
printf "Results\n %12.4f\n", result

xx=mshape(stats,k,k+1)
yy=mshape(sd_stuff,k,k+1)

# printf "average of bkw matrix\n %12.4g\n", xx
# printf "std dev of bkw matrix\n %12.4g\n", yy
printf "avg and std dev\n %12.4g\n", xx|yy

```

6 Sample Code for Sample Selection Simulation

```

open mroz
logs wage
scalar N=$nobs
# heckit l_wage const educ exper; lfp const educ kidsl6

list xlist = const educ exper kidsl6
matrix X = { const, educ, exper, kidsl6 }
matrix U V
matrix L = svd(X,&U,&V)

```

```

L
# Pick one of these
matrix eigs = {1,1,1,1}
matrix eigs = {10,1,1,0.1}
matrix eigs = {10, .1, .1, .05}
# matrix eigs = {10,10,0.1,0.1}
# matrix eigs = { 10, 7, 4, 1}
scalar rho = -0.5 # set r = (0.0, 0.5, or 0.99)
scalar p = 0.5 # set p = (0, or 0.5)
scalar sig_e = 1
# matrix eigs = (sumr(L)/sumr(eigs)).*eigs

printf "Eigenvalues %12.4f \n", eigs
matrix mm = diagrv(eigs)
matrix xs = U * mm * V
matrix xsx = xs'*xs

# put the regressors into a data set
scalar k = nelem(xlist)
list std_var = null
loop foreach i xlist --quiet
    series c$i = xs[,i]
endloop

list Xlist = cconst ceduc cexper
list Wlist = cconst ceduc cexper

matrix X = { Xlist }
matrix W = { Wlist }

heckit l_wage Xlist; lfp Wlist --quiet

matrix parms = $coeff
matrix parms = round(parms*100)/100
scalar k=nelem($coeff)
matrix co = $vcv

string vn = "const educ exper age"
matrix bb = bkw(co, vn)

matrix S = {1, rho; rho, 1}
matrix C = cholesky(S)

```

```

scalar nparms=nelem(parms)
matrix beta = parms[1:nelem(Xlist),]
matrix gam = parms[nelem(Xlist)+2:nparms,]

scalar NMC = 400
matrix co = $vcv
string vn = "const educ exper lambda const educ exper "
matrix bk = bkw(co,vn)

matrix bkw_dat = zeros(NMC,k*(k+1))
matrix op_coeff = zeros(NMC,k)
# matrix op_coef_ls = zeros(NMC,nelem(xlist_collin)+1)
err=0
er=0
j=1
loop while j <= NMC --quiet
    matrix errors = mnormal(N,2)*C'
    matrix vi = errors[,1]
    matrix ei = errors[,2].*(sig_e)
    matrix zi = W*gam+vi
    matrix z = (zi.>0)
    matrix yi = X*beta+ei
    series zs = z
    series ys = yi

    catch heckit ys Xlist;zs Wlist --quiet
    scalar err = $error
    if err
        scalar er = er + 1

    else
        # save the estimated coefficients
        matrix co = $vcv
        matrix bk = bkw(co,vn)
        matrix bkw_dat[j,] = (vec(bk))'
        matrix op_coeff[j,]=($coeff)'
        j=j+1
    endif

# save the estimated coefficients

```

```

#     matrix co = $vcv
#     matrix bk = bkw(co,vn)
#     matrix bkw_dat[i,] = (vec(bk))'
#     matrix op_coeff[i,]=($coeff)'

endloop
printf "The proportion of errors is %12.5f\n", er/(NMC+er)
matrix stats=meanc(bkw_dat)
matrix sd_stuff=sdc(bkw_dat)
parms[nelem(Xlist)+1]=rho*sig_e
matrix coef = meanc(op_coeff)
matrix sd_coef=sdc(op_coeff)
matrix mc_sd_coef=sdc(op_coeff)./sqrt(rows(op_coeff))
matrix des = eigs~rho~p~sig_e
# printf "various stats from the design \n %12.5f\n", des
# printf "true values\n %12.5f\n", b~taus
# printf "average of coefficients\n %12.5f\n", coef
# printf "std deviation of coefficients\n %12.5g\n", sd_coef
# printf "t-ratio of coefficients\n %12.5g\n", coef./sd_coef
matrix result = des|parms'|coef|sd_coef|(coef-(parms'))./sd_coef
printf "Results\n %12.3f\n", result

xx=mshape(stats,k,k+1)
yy=mshape(sd_stuff,k,k+1)
printf "avg and std dev\n %12.4g\n", xx|yy

```

References

- Adkins, Lee C., Melissa Waters and R. C. Hill (2015), 'Collinearity diagnostics in gretl', *Presented at 4th Gretl Conference, Berlin Germany*.
URL: http://learneconometrics.com/pdf/Collin/collin_gretl.pdf
- 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.
- Greene, W. (2012), *Econometric Analysis*, 7th edn, Prentice Hall, Upper Saddle River, NJ.

- Greene, William H. (2012), *Econometric Analysis*, 7th edn, Prentice-Hall, Upper Saddle River, NJ.
- Heckman, James J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**(1), 153–161.
- Hill, Carter, Lee Adkins and Keith Bender (2003), 'Test statistics and critical values in selectivity models', *Advances in Econometrics* **17**, 75–105.
- 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.
- McCullagh, P. and J.A. Nelder (1989), *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.
- Nawata, Kazumitsu and Nobuko Nagase (1996), 'Estimation of sample selection bias models', *Econometric Reviews* **15**(4), 387–400.
- 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.