Bebonchu Atems and Jason Bergtold*

Revisiting the statistical specification of near-multicollinearity in the logistic regression model

Abstract: This paper revisits the statistical specification of near-multicollinearity in the logistic regression model. We argue that the *ceteris paribus* clause, which assumes that the maximum likelihood estimator of β remains constant as the correlation (ρ) between the regressors increases, invoked under the traditional account of near-multicollinearity is rather misleading. We derive the parameters of the logistic regression model and show that they are functions of ρ, indicating that the *ceteris paribus* clause is unattainable. Monte Carlo simulations confirm these findings and further show that: coefficient estimates and related statistics fluctuate in a non-symmetric, non-monotonic way as $|\rho| \rightarrow 1$; that the impact of near-multicollinearity is centered on the estimates of β; and that the impact on substantive inferences does not necessarily follow what the traditional account implies.

Keywords: logistic regression; model diagnostics; near-multicollinearity.

JEL Classification: C1; C2; C4.

1 Introduction

In econometrics, statistics and related disciplines, near-multicollinearity is seen as a pervasive problem that can have significant consequences for estimation and the reliability of substantive inference. Traditionally, when regressors in a regression model are perfectly or nearly perfectly correlated, estimates of the individual regression coefficients become unstable and the inferences based on the model may be misleading. This condition is known as multicollinearity (Mason, Webster, and Gunst 1975). Multicollinearity occurs when variables in the model are correlated to an extent that individual regression coefficient estimates become unreliable. When the regressors have an exact linear relationship, they are said to be perfectly collinear. When the relationship between the predictor variables is almost linear (but not exact), this results in the phenomenon known as *near-multicollinearity*, the problem specifically addressed in this paper and frequently encountered in the applied literature.

The problems traditionally associated with near-multicollineairty in the logistic regression model are similar to those traditionally found for the linear regression model (Menard 2002). This argument is based on the premise that the logistic regression model can be expressed linearly in terms of the log-odds as $ln(p/1-p) = \beta' \mathbf{X}_i + e_i$, where p is the mean of the binary dependent variable (Y_i), β is a vector of parameters, $\mathbf{X}_i$ is a ($k \times 1$) vector of explanatory variables, and e_i is a zero mean IID random error term. Therefore, similar to the linear regression model, near-multicollinearity in the logistic regression model may lead to near-singularity among the columns of $\mathbf{X}$, resulting in numerical instability, inflated variances of the maximum likelihood coefficient estimators, and misleading inferences. Many of the diagnostics suggested in the literature use the

*Corresponding author: Jason Bergtold, Kansas State University, Agricultural Economics, 342 Waters Hall Ag Economics Kansas State University, Manhattan, KS 66506-4011, USA, e-mail: bergtold@ksu.edu
Bebonchu Atems: Economics and Financial Studies, School of Business, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699, USA

linear form of the logistic regression for detecting problems with near-multicollinearity (e.g. variance infla-
tion and tolerance factors) (Hosmer and Lemeshow 2000; Menard 2010).

Spanos and McGuirk (2002) challenge the traditional account of near-multicollinearity for the linear
regression model. The traditional account relies upon a *ceteris paribus* clause that assumes that the estima-
tors of β and the R^2 remain constant as $|\rho| \to 1$, where ρ represents the correlation between two regressors.
However, the estimators of β and R^2 are in fact functions of ρ and change as $|\rho| \to 1$. This implies that changes
in associated statistics, such as the variances $[Var(\beta)]$ and t-ratios $[\tau(\beta)]$ of the estimators will be different
than previously thought. In fact, Spanos and McGuirk (2002) conclude that "the changes in the relevant
statistics $[Var(\beta), \tau(\beta)]$ are rather different than the usual account suggests; the *ceteris paribus* clause is unat-
tainable and the traditional account needs to be reconsidered (p: 369)." It is the unattainable *ceteris paribus*
clause that is also invoked when examining multicollinearity in the logistic regression model that is of major
concern in this paper. It is the link between the linear regression model and linearized form of the logistic
regression model in terms of the log odds that makes the findings from Spanos and McGuirk (2002) relevant
in re-evaluating the statistical specification of near-multicollinearity in the logistic regression model. Deriva-
tions by Scrucca and Weisberg (2004) show that the parameters of the logistic regression model may be func-
tions of ρ. Thus, the *ceteris paribus* clause is unlikely to hold for the logistic regression model.

The purpose of this paper is to revisit the statistical specification of near-multicollinearity in the logis-
tic regression model under the assumption of linearity of the predictor or index function. Generalizing the
derivations from Scrucca and Weisberg (2004), we derive the parameters of the logistic regression model in
terms of the correlation coefficient, ρ, between the regressors in the model. This parameterization allows for
an investigation of near-multicollinearity in the logistic regression model and why the *ceteris paribus* clause
is not appropriate. Evidence from simulations confirm these findings, making the impact of near-multicollin-
earity different than previously thought. The results in this paper extend what Spanos and McGuirk (2002)
find with regards to near-multicollinearity in the linear regression model to the logistic regression model,
but differ in that the the regression function is nonlinear and the impacts of near-multicollinearity are more
related to the changes in the parameter estimates of the regressors.

The paper proceeds as follows. Section 2 reviews the specification of near-multicollinearity for the linear
regression model. In Section 3, we examine the problem of near-multicollinearity in the logistic regression
modeling framework. Section 4 simulates a simple logistic model to illustrate the impacts of near-multicol-
linearity on parameter estimates, associated standard errors and asymptotic t-ratios. Section 5 concludes and
provides some general findings.

2 Background: Multicollinearity in the linear regression model

In the linear regression model $y = \mathbf{X}\beta + e$, perfect multicollinearity occurs when at least one of the columns of
$\mathbf{X}$ is a linear transformation of the others. This situation occurs when the correlation (ρ) between the regres-
sors is equal to one in absolute value. Perfect multicollinearity results in parameter identification problems
since the $(\mathbf{X}'\mathbf{X})$ matrix is singular. Thus, the OLS estimators, $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}y$, and their associated variances/
covariances, $\widehat{cov}(\hat{\beta}) = \hat{s}^2(\mathbf{X}'\mathbf{X})^{-1}$ (where $\hat{s}^2 = \dfrac{\hat{e}'e}{T-K}$), cannot be estimated. On the other hand, if $det(\mathbf{X}'\mathbf{X}) \approx 0$ but
$(\mathbf{X}'\mathbf{X})$ is still nondegenerate, then we have the associated problem of near-multicollinearity. Simply put, near-
multicollinearity occurs when $|\rho| \approx 1$. When the regressors are highly correlated in this way, the data matrix is
ill-conditioned, even though the OLS estimators still exist. However, high correlation among the regressors
may lead to numerical instability in the precision and significance of the estimates of β.[1]

Spanos and McGuirk (2002) explore in detail the problem of near-multicollinearity in the linear regres-
sion model. They consider the following linear model:

$$y_t = \beta_0 + \beta x_{1t} + \beta_2 x_{2t} + e_t, \ e_t \sim N(0, \sigma^2)$$

---

[1] See Greene (2011), page 256 for a more detailed discussion of the symptoms and consequences of near-multicollinearity.

Spanos and Mcguirk (2002) show that the OLS estimators of β, σ^2 and R^2 are:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2, \ \hat{\beta}_1 = \frac{\hat{\sigma}_{33}\hat{\sigma}_{12} - \hat{\rho}\hat{\sigma}_{13}\sqrt{\hat{\sigma}_{22}\hat{\sigma}_{33}}}{(1-\hat{\rho}^2)\hat{\sigma}_{22}\hat{\sigma}_{33}}, \ \hat{\beta}_2 = \frac{\hat{\sigma}_{22}\hat{\sigma}_{13} - \hat{\rho}\hat{\sigma}_{12}\sqrt{\hat{\sigma}_{22}\hat{\sigma}_{33}}}{(1-\hat{\rho}^2)\hat{\sigma}_{22}\hat{\sigma}_{33}}$$

$$\hat{\sigma}^2 = \hat{\sigma}_{11} - \frac{\hat{\sigma}_{33}\hat{\sigma}_{12} - 2\hat{\rho}\hat{\sigma}_{12}\hat{\sigma}_{13}\sqrt{\hat{\sigma}_{22}\hat{\sigma}_{33}} + \hat{\sigma}_{22}\hat{\sigma}_{13}}{(1-\hat{\rho}^2)\hat{\sigma}_{22}\hat{\sigma}_{33}}; \ R^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_{11}^2}; \ \hat{\rho}^2 = \frac{\hat{\sigma}_{23}^2}{\hat{\sigma}_{22}\hat{\sigma}_{33}}, \ \text{where}$$

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}(z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j), \ i, j = 1, \ 2, \ 3; \ z_{1t} = y_t; \ z_{2t} = x_{1t}; \ z_{3t} = x_{2t}.$$

They further show that the estimated variances of the above estimators for β take the form:

$$Var(\hat{\beta}_1) = \frac{Ts^2}{(1-\hat{\rho}^2)\hat{\sigma}_{22}}, \ Var(\hat{\beta}_2) = \frac{Ts^2}{(1-\hat{\rho}^2)\hat{\sigma}_{33}}, \ \text{where} \ s^2 = \frac{\sum_{t=1}^{T}\hat{e}_t^2}{T-3};$$

and the corresponding *t-statistics*, $\tau(\beta_i)$ are:

$$\tau(\hat{\beta}_1) = \frac{\hat{\beta}_1}{s\sqrt{T}}\sqrt{(1-\hat{\rho}^2)\hat{\sigma}_{22}}, \ \tau(\hat{\beta}_2) = \frac{\hat{\beta}_2}{s\sqrt{T}}\sqrt{(1-\hat{\rho}^2)\hat{\sigma}_{33}}$$

The conventional account of near-multicollinearity in the linear regression model is associated with the effects on $\tau(\hat{\beta}_1)$, $\tau(\hat{\beta}_2)$, $Var(\hat{\beta}_1)$, and $Var(\hat{\beta}_2)$ as $|\hat{\rho}| \to 1$, ceteris paribus. The *ceteris paribus* clause assumes that $\hat{\beta}$, $\hat{\sigma}^2$, and R^2 are held constant, when examining the impact as $|\hat{\rho}| \to 1$ on the statistics $\tau(\hat{\beta}_1)$, $\tau(\hat{\beta}_2)$, $Var(\hat{\beta}_1)$, and $Var(\hat{\beta}_2)$. However, because $\hat{\beta}$, $\hat{\sigma}^2$, and R^2 are all functions of $\hat{\rho}$, the changes in the associated varainces and *t*-ratios will be different than the traditional account implies. Therefore the ceteris paribus assumption fails to hold.

The problems traditionally associated with near-multicollineairty in the logistic regression model are similar to those traditionally found for the linear regression model (Menard 2002). The analysis, however, is not directly applicable to the logistic regression model since the variance-covariance structure of the estimated parameters for the logistic regression model is different than that of the linear regression model and the regression is nonlinear. In the next section, the problem of near-multicollinearity in the logistic regression model is examined using the probabilistic reduction approach (Bergtold, Spanos, and Onukwugha 2010).

3 Multicollinearity in the logistic regression model

Let Y_i represent a Bernoulli random variable with mean parameter p and $\mathbf{X}_i$ be a $(k \times 1)$ vector of regressors. The logistic regression model takes the form $Y_i = E(Y_i|\mathbf{X}_i = \mathbf{x}_i) + e_i = [1 + exp(-\beta'\mathbf{X}_i)]^{-1} + e_i$. Bergtold, Spanos, and Onukwugha (2010) show that a sufficient condition for the existence of the model is the compatability between the conditional distribution $f_{Y|\mathbf{X}}(Y_i|\mathbf{X}_i;\beta)$ underlying the model and its associated inverse conditional distribution $f_{\mathbf{X}|Y=j}(\mathbf{X}_i;\theta_j)$. This condition is usually examined in logistic discriminant analysis, and has been explored by, among others Kay and Little (1987) and Cox and Snell (1989). The conditional distribution is usually modeled in logistic regression, whereas the inverse conditional distribution is modeled in discriminant analysis. While discriminant analysis is not commonly utilized in econometrics, the link between logistic regression and disciminant analysis provides a strong framework upon which to examine near-multicollinearity in the logistic regression model. The compatability condition provides the basis for a systematic approach to model specification and ensures that the underlying joint distribution of Y_i and $\mathbf{X}_i$ exists, which is a requirement for the existence of a model (Bergtold, Spanos, and Onukwugha 2010). Thus, the functional specification of the

logistic regression is intertwined with the inverse conditional distribution that captures the needed probability information in the explanatory variables, which is often overlooked (e.g. see Kay and Little 1987; Arnold, Castillo, and Sarabria 1999).

The model specification approach [the probabilistic reduction (PR) approach] adopted in this paper provides a more robust specification for examining near-multicollinearity. Traditional methods of simulating conditional binary discrete choice processes (e.g. Ryan 1997; Hosmer and Lemeshow 2000) may not capture or produce the true underlying parameters of the model or may impose restrictions upon the model, thereby limiting its flexibility. For example, the predictor or index function of the logisitc regression model may be nonlinear in the variables. If the inverse conditional distribution is multivariate normal with heterogenous covariance matrix, then the predictor or index function will be a quadratic function of the explanatory variables (McFadden 1976).

Following Bergtold, Spanos, and Onukwugha (2010), the logistic regression model can be expressed as:

$$Y_i = [1 + exp\{-\eta(\mathbf{X}_i; \beta)\}]^{-1} + e_i \tag{1}$$

where $\eta(\mathbf{X}_i; \beta) = ln\left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)}\right) + \kappa$, $\kappa = ln\left(\frac{p}{1-p}\right)$, and $p = Prob(Y_i = 1)$. Equation (1) is usually reparametized such that $\beta = \beta(\theta_j)$, $j = 0, 1$. To date, the literature examining near-multicollinearity in the logistic regression model has not captured the fact that $\beta = \beta(\rho)$, where ρ is the correlation between the explanatory variables when $f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta_j)$ is multivariate normal. Thus, one can show that the traditional account of near-multicollinearity and the invocation of the *ceteris paribus* clause are unattainable, as in the case for the linear model.

Following the commonly used specification in the literature of linearity in the variables and parameters of the function $\eta(\cdot)$, assume that $f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta_j)$ is multivariate normal with a homoegenous covariance matrix. That is:

$$f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta_j) = (2\pi)^{-\frac{k}{2}} |\mathbf{V}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\mathbf{X}_i - \mu_j)' \mathbf{V}^{-1}(\mathbf{X}_i - \mu_j)'\right\} \tag{2}$$

where μ_j is the mean vector for $\mathbf{X}_i$ conditional on $Y_i = j$, and $\mathbf{V}$ is the covariance matrix for $\mathbf{X}_i$.[2] Following Day and Kerridge (1967), a more general distributional functional form $f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta) = \alpha_j exp\left\{-\frac{1}{2}(\mathbf{X}_i - \mu_j)' \mathbf{V}^{-1}(\mathbf{X}_i - \mu_j)'\right\} \delta(\mathbf{X}_i)$ can be expressed, as well. When $\delta(\mathbf{X}_i) = 1$, $f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta_j)$ is distributed multivariate normal. However, if $\delta(\mathbf{X}_i) \neq 1$, this functional form can represent a wider range of alternatives, including skewed distributions (Byth and McLachlan 1980). Thus, the results from this paper should hold for more general specifications of the logistic regression model than when the inverse conditional model is multivariate normal. This flexibility may be particularly important in economics when considering cases with a skewed distribution.

With the above assumptions, the $ln\left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)}\right) = [(\mu_1 - \mu_0)' \mathbf{V}^{-1}] \mathbf{X}_i + \left[\frac{1}{2}\mu_0' \mathbf{V}^{-1}\mu_0 - \frac{1}{2}\mu_1' \mathbf{V}^{-1}\mu_1\right]$, giving:

$$\eta(\mathbf{X}_i; \beta) = \beta_0 + \beta' \mathbf{X}_i, \tag{3}$$

where $\beta_0 = ln\left(\frac{p}{1-p}\right) + \left[\frac{1}{2}\mu_0' \mathbf{V}^{-1}\mu_0 - \frac{1}{2}\mu_1' \mathbf{V}^{-1}\mu_1\right]$; and $\beta = [(\mu_1 - \mu_0)' \mathbf{V}^{-1}]$.

Without loss of generality, standardize $\mathbf{X}_i$ by dividing each element by its respective standard deviation, so that $Var(X_{m,i}) = 1$ for $m = 1, \ldots, k$. Then repartition the covariance matrix, $\mathbf{V}$ so that $\mathbf{V} = \begin{pmatrix} 1 & \rho_m' \\ \rho_m & P_{-m} \end{pmatrix}$, where $cov(X_m, X_s) = \rho_{m,s}$ is the correlation between X_m and X_s; ρ_m is the $(k-1 \times 1)$ vector of correlation coefficients

---

2 Arnold, Castillo, and Sarabria (1999) note that linearity in the parameters and variables for the predictor or index function requires severe restrictions and that many of the applied logistic regression models in the literature may be questionable in light of this information. Furthermore, note that the covariance matrix $\mathbf{V}$ is not conditional on Y_i for this case.

between X_m and all the other explanatory variables; and P_{-m} is the remainder of the covariance matrix **V**, with 1 s along the diagonal and correlation coefficients on the off-diagonal elements between the remaining regressors. The repartition of **V** allows us to isolate the parameterization of the coefficient β_m for $m=1, \ldots, k$. Invoking the Schur lemma following Spanos and McGuirk (2002):

$$\mathbf{V}^{-1}=\begin{pmatrix} \gamma^{-1} & -\gamma^{-1}\rho'_m P_{-m}^{-1} \\ -P_{-m}^{-1}\rho_m\gamma^{-1} & P_{-m}^{-1}\rho_m\gamma^{-1}\rho'_m P_{-m}^{-1}+P_{-m}^{-1} \end{pmatrix} \tag{4}$$

where $\gamma=1-\rho'_m P_{-m}^{-1}\rho_m$. Plugging equation (4) into the formula for β in equation (3), gives:

$$\beta=\begin{pmatrix} \beta_m \\ \beta_{-m} \end{pmatrix}=\begin{pmatrix} \gamma^{-1}(\mu_{1,m}-\mu_{0,m})-\gamma^{-1}\rho'_m P_{-m}^{-1}(\mu_{1,-m}-\mu_{0,-m}) \\ -P_{-m}^{-1}\rho_m\gamma^{-1}(\mu_{1,m}-\mu_{0,m})+[P_{-m}^{-1}\rho_m\gamma^{-1}\rho'_m P_{-m}^{-1}+P_{-m}^{-1}](\mu_{1,-m}-\mu_{0,-m}) \end{pmatrix} \tag{5}$$

where $\mu_{j,m}$ is the mth element of μ_j for $j=0, 1$; and $\mu_{j,-m}$ represents the $(k-1\times1)$ vector of μ_j without $\mu_{j,m}$ for $j=0, 1$. Equation (5) illustrates that β_m for $m=1, \ldots, k$ are functions of ρ_m. Furthermore, β_m is affected by the correlation between other explanatory variables via P_{-m}^{-1}. Thus, the *ceteris paribus* clause is unattainable in the logistic regression model because β is a function of ρ.

Following Schaefer, Roi, and Wolfe (1984) and Gourieroux (2000), let $\widehat{Prob}(y_i=1|\mathbf{X}_i;\beta)=F(\beta'_{MLE}X_i)=F_i$ so that the asymptotic variance estimator of β can be written as $Var(\beta_{MLE})=(\mathbf{X}'\Omega\mathbf{X})^{-1}$, where $\Omega=diag[F_i(1-F_i)]_{i=1}^n$ and **X** is the $(n\times k)$ dimensional matrix of regressors. Then, $Var(\beta_{MLE})$ is a function of ρ through F_i. Consider the case where $|\rho_{m,s}|\to1$ for some $s=1, \ldots, k$. If $|\beta_m|\to\infty$, then F_i will aproach 0 or 1, depending on the sign of β_m. This implies that $F_i(1-F_i)\to0$ as $|\beta_m|\to\infty$, suggesting that the $Var(\beta_{MLE})$ may follow the shifts in β_m as $|\rho_{m,s}|\to1$ for some $s=1, \ldots, k$. While changes in β with respect to ρ_m can be plotted using equation (5), $Var(\beta_{MLE})$ and asymptotic t-ratios $[\tau(\beta_{MLE})]$ need to be examined using simulation methods as no closed form expressions exist for these statistics (Gourieroux 2000).[3]

4 Simulation

This section simulates a simple logistic regression model with two covariates to illustrate the impacts of near multicollinearity on parameter estimates, associated standard errors, and asymptotic t-ratios.

4.1 Logisitc regression model simulated

Assume that $\mathbf{X}_i=(x_{1,i}, x_{2,i})'$ conditional on $Y_i=j$ is bivariate normal with homogenous covariance matrix. That is:

$$\begin{pmatrix} x_{1,i} \\ x_{2,i} \end{pmatrix}|Y_i=j \sim N\left[\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]$$

where $\rho=Corr(x_{1,i}, x_{2,i})$. The inverse conditional distribution function can be expressed as (see Spanos 1986: p 119–121 and Bergtold, Spanos, and Onukwugha 2010):

$$f_{\mathbf{X}_i|Y=j}(x_1, x_2;\theta_j)=\frac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi\sigma_1\sigma_2}\cdot exp\left\{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_{1,j}}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_{2,j}}{\sigma_2}\right)^2-2\rho\left(\frac{x_1-\mu_{1,j}}{\sigma_1}\right)\left(\frac{x_2-\mu_{2,j}}{\sigma_2}\right)\right]\right\} \tag{6}$$

---

3 We examine asymptotic t-ratios in this paper, however, similar results hold for other asymptotic test statisitcs, such as the likelihood ratio or Wald test statistics.

Let $\vartheta_j = -\dfrac{1}{2}\dfrac{1}{(1-\rho^2)}\left[\left(\dfrac{x_1-\mu_{1,j}}{\sigma_1}\right)^2 + \left(\dfrac{x_2-\mu_{2,j}}{\sigma_2}\right)^2 - 2\rho\left(\dfrac{x_1-\mu_{1,j}}{\sigma_1}\right)\left(\dfrac{x_2-\mu_{2,j}}{\sigma_2}\right)\right]$, so that equation (6) can be rewritten as:

$$f_{\mathbf{X}_i|Y=j}(x_1,x_2;\theta_j)=\dfrac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi\sigma_1\sigma_2}\cdot exp\{\vartheta_j\} \tag{7}$$

Then:

$$ln\left(\dfrac{f_{\mathbf{X}_i|Y=1}(x_1,x_2;\theta_1)}{f_{\mathbf{X}_i|Y=0}(x_1,x_2;\theta_0)}\right)=ln\dfrac{exp\{\vartheta_1\}}{exp\{\vartheta_0\}}=\vartheta_1-\vartheta_0 \tag{8}$$

Combining equations (6) and (8) for $j=0$, 1 simplifying, and rearranging terms gives:

$$\begin{aligned}
\vartheta_1-\vartheta_0 = &-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{\mu_{1,1}^2-\mu_{1,0}^2}{\sigma_1^2}-\dfrac{2\rho(\mu_{1,1}\mu_{2,1}-\mu_{1,0}\mu_{2,0})}{\sigma_1\sigma_2}+\dfrac{\mu_{2,1}^2-\mu_{2,0}^2}{\sigma_2^2}\right)+\left(\dfrac{2\rho(\mu_{2,1}-\mu_{2,0})}{\sigma_1\sigma_2}-\dfrac{2(\mu_{1,1}-\mu_{1,0})}{\sigma_1^2}\right)x_1\right.\\
&\left.+\left(\dfrac{2\rho(\mu_{1,1}-\mu_{1,0})}{\sigma_1\sigma_2}-\dfrac{2(\mu_{2,1}-\mu_{2,0})}{\sigma_2^2}\right)x_2\right]
\end{aligned} \tag{9}$$

Using the relationships in equations (1), (6), (8) and (9), the corresponding logistic regression model can be written as:

$$Y_i=[1+exp\{-\beta_0-\beta_1 x_1-\beta_2 x_2\}]^{-1}+u_i. \tag{10}$$

From equations (3) and (9), it follows that:

$$\beta_0=ln\left(\dfrac{p}{1-p}\right)-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{\mu_{1,1}^2-\mu_{1,0}^2}{\sigma_1^2}-\dfrac{2\rho(\mu_{1,1}\mu_{2,1}-\mu_{1,0}\mu_{2,0})}{\sigma_1\sigma_2}+\dfrac{\mu_{2,1}^2-\mu_{2,0}^2}{\sigma_2^2}\right)\right] \tag{11}$$

$$\beta_1=\dfrac{1}{2(1-\rho^2)}\left[\dfrac{2(\mu_{1,1}-\mu_{1,0})}{\sigma_1^2}-\left(\dfrac{2\rho(\mu_{2,1}-\mu_{2,0})}{\sigma_1\sigma_2}\right)\right] \tag{12}$$

$$\beta_2=\dfrac{1}{2(1-\rho^2)}\left[\dfrac{2(\mu_{2,1}-\mu_{2,0})}{\sigma_2^2}-\left(\dfrac{2\rho(\mu_{1,1}-\mu_{1,0})}{\sigma_1\sigma_2}\right)\right] \tag{13}$$

4.2 Simplifications

The simplifications below reduce the dimensionality of the problem being simulated without necessarily limiting the generalizability of the analysis, only affecting scaling of the results.

4.2.1 Admissable parameter values

In light of the statistical reparametizations shown above, it is important to determine the range of admissible values for which the model exists. The importance of this is emphasized by Spanos and McGuirk (2002). Let Σ

denote the the variance-covariance matrix of the inverse conditional distribution, then $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Following this parametization, the above parameters $\{\beta_0, \beta_1, \beta_2\}$ are valid when (Spanos 1986):

$$det(\Sigma) = \sigma_1^2\sigma_2^2(1-\rho^2) > 0 \tag{14}$$

Condition (14) implies that the inverse conditional distribution given by equation (6) will not exist when $\Sigma \not\geq 0$. Thus, during simulations if $det(\Sigma) = 0$ then zeros are assigned for all the parameter values and associated statistics for that simulation run.

4.2.2 Simplification 1

Standardize the variables x_1 and x_2 by dividing them by σ_m for $m=1, 2$, so that $\sigma_1 = \sigma_2 = 1$. It follows that equation (14) becomes $det(\Sigma) = (1-\rho^2) > 0$. Thus, the admissible parameter range for ρ is $\rho \epsilon (-1, 1)$. Following this simplification, the parameters $\{\beta_0, \beta_1, \beta_2\}$ from equations (11) to (13) reduce to:

$$\beta_0 = ln\left(\frac{p}{1-p}\right) - \frac{1}{2(1-\rho^2)}[\mu_{1,1}^2 - \mu_{1,0}^2 - 2\rho(\mu_{1,1}\mu_{2,1} - \mu_{1,0}\mu_{2,0}) + \mu_{2,1}^2 - \mu_{2,0}^2]$$

$$\beta_1 = \frac{1}{2(1-\rho^2)}[2(\mu_{1,1} - \mu_{1,0}) - 2\rho(\mu_{2,1} - \mu_{2,0})] \tag{15}$$

$$\beta_2 = \frac{1}{2(1-\rho^2)}[2(\mu_{2,1} - \mu_{2,0}) - 2\rho(\mu_{1,1} - \mu_{1,0})]$$

4.2.3 Simplification 2

Further, consider the following mean deviation forms of $x_{1,i}$ and $x_{2,i}$: $\bar{x}_1 = x_1 - \mu_{1,0}$ and $\bar{x}_2 = x_2 - \mu_{2,0}$, where $\mu_{m,0} = E(x_m|Y=0)$ for $m=1, 2$. Then $\bar{\mu}_{m,0} = E(\bar{x}_m, |Y=0) = 0$ for $m=1, 2$, making $\bar{\mu}_{m,1} = E(\bar{x}_m, |Y=1) = \mu_{m,1} - \mu_{m,0}$ for $m=1, 2$. This further reduces the dimensionality of the parameters $\{\beta_0, \beta_1, \beta_2\}$ from equations (15) to:

$$\beta_0 = ln\left(\frac{p}{1-p}\right) - \frac{1}{2(1-\rho^2)}[\bar{\mu}_{1,1}^2 - 2\rho(\bar{\mu}_{1,1}\bar{\mu}_{2,1}) + \bar{\mu}_{2,1}^2]$$

$$\beta_1 = \frac{1}{2(1-\rho^2)}[2\bar{\mu}_{1,1} - 2\rho\bar{\mu}_{2,1}] \tag{16}$$

$$\beta_2 = \frac{1}{2(1-\rho^2)}[2\bar{\mu}_{2,1} - 2\rho\bar{\mu}_{1,1}]$$

4.3 Monte Carlo simulation procedure

To generate binary choice data, we use a two stage process following Bergtold, Spanos, and Onukwugha (2010) and Scrucca and Weisberg (2004). First, using a binomial random number generator, realizations of the vector stochastic process $\{Y_i, i=1, \ldots, n\}$ are generated. Second, using Y_i as the conditioning variable, the vector stochastic process of predictors $\{\mathbf{X}_i, i=1, \ldots, n\}$ is generated using the bivariate normal inverse conditional distribution $f_{\mathbf{X}|Y=j}(\mathbf{X}_i; \theta_j)$ using appropriate random number generators. Given these simplifications, Monte Carlo simulations are conducted for different mean pairs of $(\bar{\mu}_{11}, \bar{\mu}_{21})$.[4] Monte Carlo simulations involved conducting 1000 runs for all mean pair combinations for each value of ρ examined, where ρ was varied between

---

4 Monte Carlo simulations were conducted for a large number of different mean pairs. All simulations and graphics were conducted in MATLAB. These results are available upon request from the authors.

−1 and 1 by increments of 0.005 for the plots in Figure 1, with specific values reported in Tables 1 and 2. The means of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, their associated standard errors $[se(\hat{\beta}_0), se(\hat{\beta}_1), se(\hat{\beta}_2)]$, and asymptotic t-ratios $[\tau(\hat{\beta}_0), \tau(\hat{\beta}_1), \tau(\hat{\beta}_2)]$ were calculated across runs.

4.4 Simulation results and discusssion

Simulation results for the mean pairs $(\bar{\mu}_{11}, \bar{\mu}_{21})=(−1, −0.5)$, $(−1, 1)$, $(−0.5, −0.25)$, $(0.25, 0.5)$, $(0.5, 1)$, and $(0.5, 2)$ for 15 different values of ρ ranging from $−0.99$ to 0.99 for $\hat{\beta}_i$ and $se(\hat{\beta}_i)$ for $i=1, 2$ are provided in
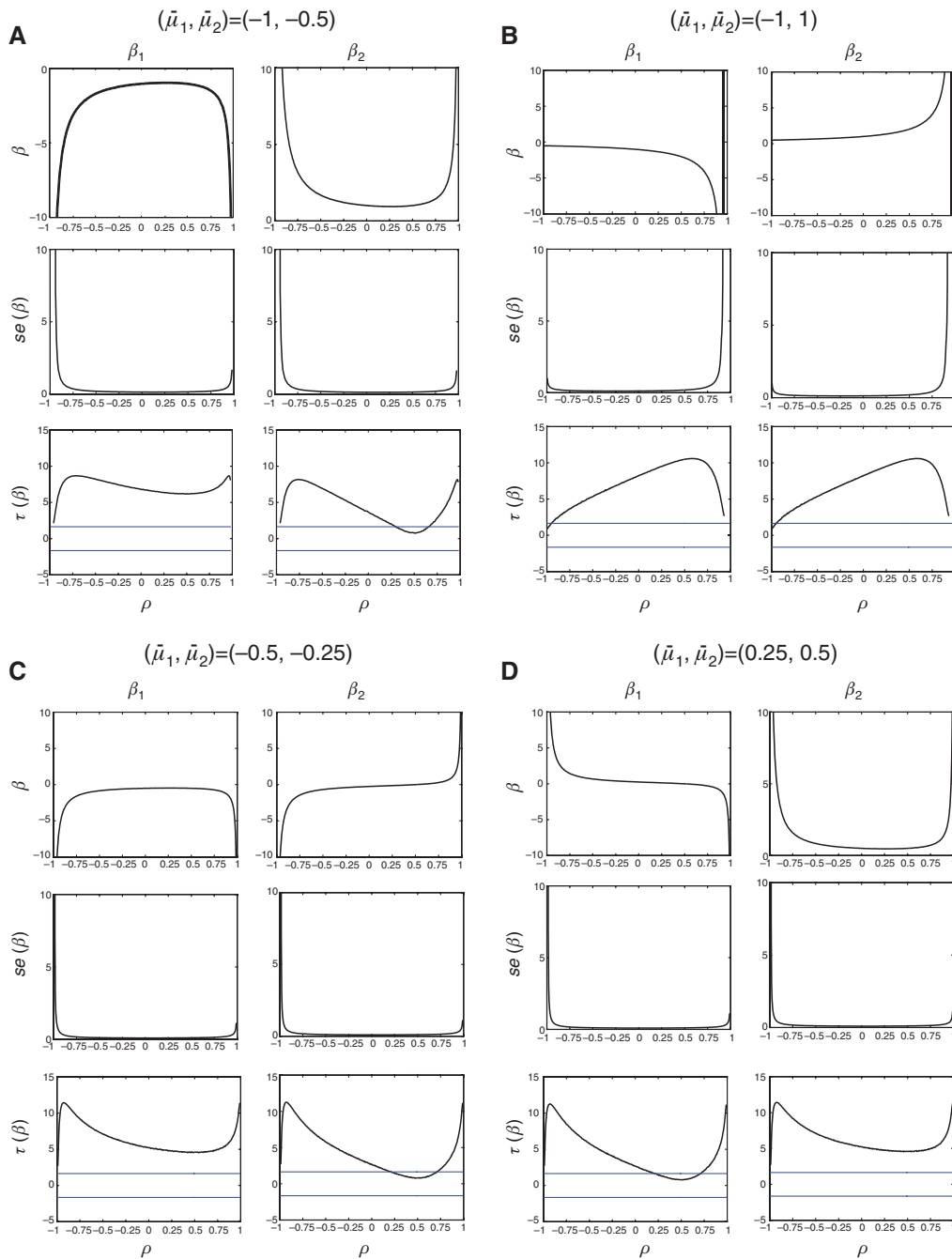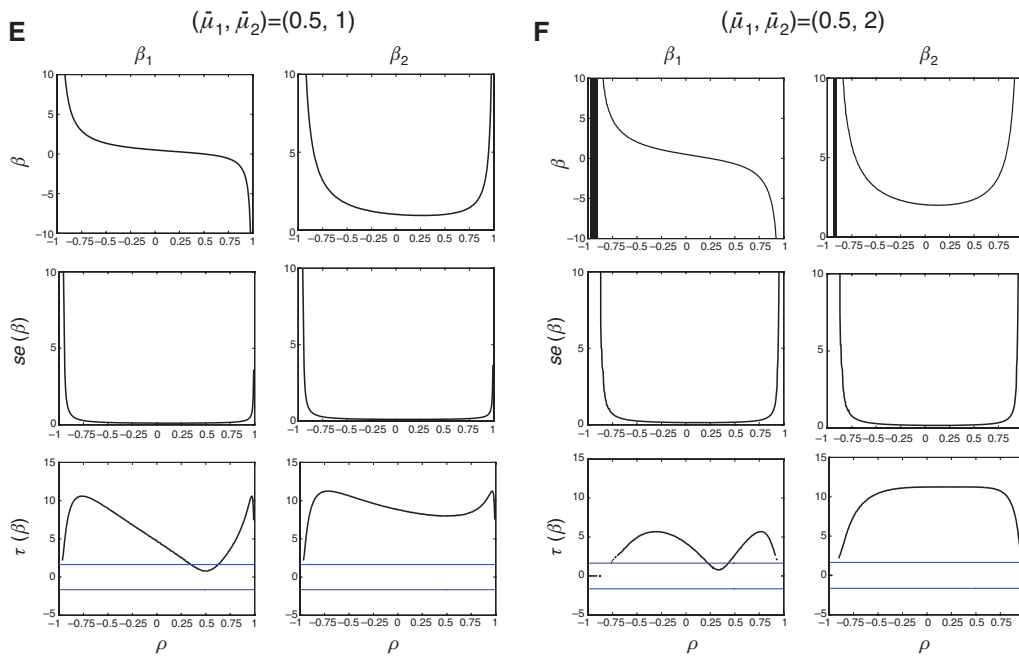

Figure 1 (continued)

Figure 1: Plots of $\hat{\beta}$, $se(\hat{\beta})$, and $\tau(\hat{\beta})$ for alternative mean pairs $(\bar{\mu}_1, \bar{\mu}_2)$ as ρ changes.

Tables 1 and 2. Figure 1 provides plots of the mean estimates of $\hat{\beta}_1$, $\hat{\beta}_2$, $se(\hat{\beta}_1)$, $se(\hat{\beta}_2)$, $\tau(\hat{\beta}_1)$, and $\tau(\hat{\beta}_2)$ as $|\rho| \to 1$ for the mean pairs mentioned above. The mean pairs presented here were selected as representative of the general findings of the large number of runs conducted. Many of the results from the simulation for alternative mean pairs not shown follow the general patterns presented.

In general, contrary to the traditional account, the plots of $\hat{\beta}_1$, $\hat{\beta}_2$ (and $\hat{\beta}_0$) for all the mean pairs simulated were not constant as $|\rho| \to 1$. Furthermore, as illustrated by the standard error estimates in Figure 1, $se(\hat{\beta})$ did not always substatially inflate as $|\rho| \to 1$. Simulation studies by Ryan (1997) support this finding. If $se(\hat{\beta})$ did become inflated, it was not necessarily distinguishable if it would increase as $\rho \to -1$, $\rho \to 1$ or both. For example, $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$ for the mean pair $(\bar{\mu}_1, \bar{\mu}_2)=(0.5, 2)$ in Figure 1(F) inflate as $|\rho| \to 1$ as the traditional account suggests. However, for the mean pair $(\bar{\mu}_1, \bar{\mu}_2)=(-1, -0.5)$ in Figure 1(A), $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$ increases only as $\rho \to -1$; and for the mean pair $(\bar{\mu}_1, \bar{\mu}_2)=(-1, 1)$ in Figure 1(B), $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$ only increase as $\rho \to 1$. These results are supported by examining the corresponding changes in the the standard error estimates in Tables 1 and 2 for the associated mean pairs, as well. This is contrary to the traditional account of near-multicollinearity identified by Schaefer, Roi, and Wolfe (1984). A possible explanation for the difference is that traditionally, it is assumed β remains constant as $|\rho| \to 1$ (i.e. the *ceteris paribus* clause) when in fact this assumption is not attainable since β is a function of ρ. Of particular interest is that in general, the numerical results in Tables 1 and 2, as well as the plots in Figure 1 provide evidence that the estimates of $se(\hat{\beta})$ tend to follow the changes in $\hat{\beta}$. This is further corroborated by examining changes in $\tau(\hat{\beta})=\hat{\beta}/se(\hat{\beta})$ as $|\rho| \to 1$. In each of the plots in Figure 1, when $|\hat{\beta}|$ increases, the $|se(\hat{\beta})|$ increases; and for most of these cases, $\tau(\hat{\beta})$ was still statistically significant. Thus, $\tau(\hat{\beta})>1$ in many cases, providing evidence that $\hat{\beta}$ may be increasing at a faster rate than the $se(\hat{\beta})$ as $|\rho| \to 1$. The implications of multicollinearity on statisitcal significance are unclear *a priori*.

The *ceteris paribus* clause has a strong impact on the implications from near-multicollinearity for the logistic regression model. When it is shown that β and other related statistics are functions of ρ, the results seem to indicate that the traditional implications may not hold. Figure 1 provides evidence that $se(\hat{\beta})$ may not become inflated and if they do it may not have an appreciable impact on statistical significance. For many of the plots examined using different mean pairs, $\hat{\beta}$ remained statistically significant at the 5% level as $|\rho| \to 1$, while the magnitude of the test statistic did fluctuate as $|\rho| \to 1$. We expect the same results would

Table 1: Simulation results for $\hat{\beta}_1$ and $se(\hat{\beta}_1)$ for different mean pairs $(\bar{\mu}_{11}, \bar{\mu}_{21})$ as ρ changes.

ρ	$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-3, -1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-1, -0.5)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-1, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-0.5, -0.25)$	
	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$
−0.99	−17.07	1.32e6	−41.63	1601.12	−0.50	0.51	−51.23	20.84
−0.95	−15.96	2299.90	−17.41	3.68	−0.51	0.23	−7.62	0.49
−0.90	−16.66	2.37e5	−7.84	0.67	−0.53	0.17	−3.83	0.24
−0.80	−31.89	1.37e10	−3.92	0.26	−0.56	0.12	−1.95	0.14
−0.65	−7.51	1.88	−2.31	0.15	−0.60	0.10	−1.15	0.10
−0.50	−4.93	0.63	−1.68	0.11	−0.67	0.089	−0.84	0.084
−0.25	−3.54	0.32	−1.21	0.090	−0.81	0.084	−0.61	0.073
0	−3.05	0.25	−1.001	0.082	−1.01	0.088	−0.50	0.069
0.25	−2.97	0.23	−0.94	0.082	−1.34	0.10	−0.47	0.071
0.50	−3.38	0.26	−1.00	0.090	−2.01	0.14	−0.50	0.079
0.65	−4.15	0.34	−1.18	0.10	−2.89	0.20	−0.59	0.090
0.80	−6.30	0.65	−1.69	0.14	−5.07	0.41	−0.84	0.12
0.90	−13.45	3.48	−2.91	0.21	−10.94	1.74	−1.45	0.16
0.95	−82.44	1.03e14	−5.44	0.35	−50.11	5.32e10	−2.72	0.24
0.99	−31.86	1.01e4	−26.16	2.48	−32.10	9507.56	−12.76	0.80

ρ	$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-0.1, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.25, 0.5)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.5, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.5, 2)$	
−0.99	−0.032	0.46	52.00	21.83	41.55	1650.52	26.16	3.43e4
−0.95	−0.055	0.21	7.51	0.49	17.18	3.47	34.32	5.91e9
−0.90	−0.045	0.15	3.73	0.24	7.54	0.65	17.51	10.35
−0.80	−0.054	0.11	1.81	0.14	3.64	0.25	6.11	0.74
−0.65	−0.056	0.085	1.00	0.099	2.00	0.14	3.16	0.28
−0.50	−0.066	0.075	0.67	0.083	1.34	0.11	2.02	0.17
−0.25	−0.078	0.067	0.41	0.071	0.80	0.084	1.07	0.12
0	−0.10	0.65	0.25	0.068	0.50	0.076	0.51	0.10
0.25	−0.14	0.67	0.13	0.069	0.27	0.076	−0.0012	0.10
0.50	−0.20	0.075	0.0015	0.077	5.6e−3	0.084	−0.68	0.12
0.65	−0.29	0.086	−0.13	0.089	−0.26	0.096	−1.40	0.15
0.80	−0.50	0.11	−0.42	0.11	−0.84	0.13	−3.09	0.26
0.90	−1.01	0.16	−1.06	0.16	−2.11	0.20	−7.02	0.70
0.95	−2.01	0.23	−2.31	0.24	−4.64	0.33	−17.39	4.54
0.99	−10.05	0.67	−12.39	0.79	−25.21	2.40	−40.75	1942.99

hold for other asymptotic test statistics, as well. The impact of the results was dependent upon the mean pair combination simulated and may be somewhat uncertain for a given dataset *a priori,* even if high correlation among the regressors is expected.

We conjecture from these findings that the systematic implications from near-multicollinearity may result from changes in $\hat{\beta}$ as $|\rho| \to 1$ in the logistic regression model. For many of the simulations of different mean pairs, as $|\rho| \to 1$, $|\hat{\beta}| \to \infty$, or got very large in magnitude. As seen in Tables 1 and 2, the inflation of $\hat{\beta}$ tended to be more severe as the absolute difference between $\bar{\mu}_{11}$ and $\bar{\mu}_{21}$ increased, indicating the scale of the data plays a role on the impact of near multicollinearity. Scaling tends to be more of a numerical issue and did give rise to problems during estimation when the absolute difference between $\bar{\mu}_{11}$ and $\bar{\mu}_{21}$ was larger [e.g. the simulation results were more erratic for mean pair $(\bar{\mu}_{11}, \bar{\mu}_{21})=(-3, -1)$ in Tables 1 and 2 as $|\rho| \to 1$]. Given equation (10), the inflation in $\hat{\beta}$ is expected as the denominator is a function of $(1-\rho^2)$, while the numerator is only a function of ρ. The findings here are different than the findings for the linear regression model presented by Spanos and McGuirk (2002) in that the implications of near-multicollinearity for the logistic regression model seem to be tied up with $\hat{\beta}$. Near-multicollinearity affects both the standard errors of $\hat{\beta}$ and the associated *t*-ratios as in the linear regression model, but changes in the standard errors seem to follow the changes in $\hat{\beta}$.

Table 2: Simulation results for $\hat{\beta}_2$ and $se(\hat{\beta}_2)$ for different mean pairs $(\bar{\mu}_{11}, \bar{\mu}_{21})$ as ρ changes.

ρ	$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-3, -1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-1, -0.5)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-1, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-0.5, -0.25)$	
	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$
−0.99	−16.97	2.01e5	−41.44	1656.25	0.51	0.51	−51.08	20.79
−0.95	−15.58	3172.13	−17.11	3.62	0.53	0.23	−7.49	0.49
−0.90	−15.81	3.06e4	−7.57	0.66	0.53	0.17	−3.70	0.24
−0.80	−28.53	1.31e10	−3.64	0.25	0.56	0.12	−1.81	0.14
−0.65	−6.06	1.58	−2.01	0.14	0.61	0.10	−1.00	0.099
−0.50	−3.52	0.50	−1.34	0.11	0.67	0.089	−0.67	0.083
−0.25	−1.91	0.24	−0.81	0.084	0.81	0.084	−0.40	0.071
0	−1.02	0.17	−0.50	0.076	1.00	0.087	−0.25	0.068
0.25	−0.27	0.15	−0.27	0.076	1.34	0.10	−0.13	0.069
0.50	0.68	0.18	−0.0038	0.084	2.01	0.14	−0.0023	0.077
0.65	1.67	0.24	0.26	0.096	2.89	0.20	0.13	0.088
0.80	4.01	0.51	0.85	0.13	5.07	0.41	0.42	0.11
0.90	10.89	2.93	2.12	0.20	10.94	1.74	1.05	0.16
0.95	74.06	9.30e13	4.66	0.33	50.16	5.22e10	2.33	0.24
0.99	31.21	1.42e4	25.38	2.43	32.08	9494.38	12.37	0.79

ρ	$(\bar{\mu}_{11}, \bar{\mu}_{21})=(-0.1, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.25, 0.5)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.5, 1)$		$(\bar{\mu}_{11}, \bar{\mu}_{21})=(0.5, 2)$	
−0.99	0.069	0.46	52.22	21.92	41.70	1593.70	26.32	2.71e4
−0.95	0.049	0.21	7.64	0.49	17.44	3.51	35.24	5.99e9
−0.90	0.064	0.15	3.86	0.25	7.81	0.67	18.69	11.02
−0.80	0.058	0.11	1.95	0.14	3.92	0.26	6.99	0.82
−0.65	0.067	0.085	1.15	0.10	2.31	0.15	4.09	0.32
−0.50	0.064	0.075	0.84	0.085	1.67	0.11	3.03	0.21
−0.25	0.080	0.067	0.60	0.073	1.21	0.090	2.28	0.15
0	0.10	0.065	0.50	0.069	1.01	0.082	2.01	0.13
0.25	0.13	0.067	0.47	0.071	0.94	0.081	2.02	0.13
0.50	0.20	0.075	0.50	0.079	1.01	0.090	2.36	0.15
0.65	0.29	0.086	0.59	0.090	1.18	0.10	2.92	0.19
0.80	0.51	0.11	0.84	0.12	1.68	0.14	4.50	0.31
0.90	1.01	0.16	1.45	0.16	2.91	0.21	8.36	0.78
0.95	2.02	0.23	2.70	0.24	5.41	0.35	18.96	4.89
0.99	10.06	0.67	12.78	0.80	25.97	2.45	41.44	1659.79

Furthermore, near-multicollinearity may have a direct effect on the magnitude of $\hat{\beta}$, related to the difference in scale of the explanatory variables involved. These differences highlight the fact that multicollineariy in nonlinear models, such as the logisitc regression model, may need to be further examined, as they may differ from the consequences of near-multicollinearity for the linear regression model.

In addition, as $|\rho| \rightarrow 1$, changes in $\hat{\beta}$ will have subsequent impacts on odds ratios and marginal effects. Odd ratios, which may be caluated as $exp(\hat{\beta})$ may grow exponentially large as $|\rho| \rightarrow 1$, because they are a direct transformation of $\hat{\beta}$. Marginal effects of the logistic regression model in equation (1) can be expressed as (see Bergtold, Spanos, and Onukwugha 2010):

$$\frac{\partial Y_i}{\partial \mathbf{X}_i} = F_i(1-F_i) \cdot \frac{\partial \eta(\mathbf{X}_i; \beta)}{\partial \mathbf{X}_i} \tag{17}$$

where all variables and functions are as previously defined. As seen in equation (17), the marginal effects are also functions of ρ, since they are functions of β through F_i and $\eta(\mathbf{X}_i; \beta)$. This implies that changes in the estimates of $\hat{\beta}$ from near-multicollinearity impact marginal effect estimates, as well.

5 Conclusion

In this paper, we derive and show that the parameters of the logistic regression model, where the predictor is linear in both the variables and parameters, are functions of ρ. Thus, our analysis calls into question the *ceteris paribus* clause in the traditional account of near-multicollinearity for the logistic regression model and shows that it is inappropriate. Monte Carlo simulations provide evidence that the parameters and their associated variances and *t*-ratios may be different than the traditional account implies. The implications of our findings are that $\hat{\beta}$, $se(\hat{\beta})$, and $\tau(\hat{\beta})$ can fluctuate in a non-systematic, non-monotonic way as $|\rho| \to 1$; that changes to relevant statistics are data specific and uncertain; that the impact of near-multicollinearity is likely centered on the estimates of β; and the impact on substantive inferences does not follow what the traditional account would imply. As the probit model may seen as a close approximation of the logistic regression model (Amemiya 1981), the results here may extend to that model, as well, though this requires further exploration. In addition, future research needs to extend the analysis to logistic regression models with predictors that are nonlinear in the regressors and linear predictors that incorporate binary covariates.

References

Amemiya, T. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19: 1483–1536.

Arnold, B. C., E. Castillo, and J. M. Sarabria. 1999. *Conditional Specifications of Statistical Models*. New York, NY: Springer.

Bergtold, J., A. Spanos, and E. Onukwugha. 2010. "Bernoulli Regression Models: Revisiting the Specification of Statistical Models with Binary Dependent Variables." *Journal of Choice Modelling* 2: 1–28.

Byth, K., and G. McLachlan. 1980. "Logistic Regression Compared to Normal Discrimination for Non-Normal Populations." *Australian and New Zealand Journal of Statistics* 22: 188–196.

Cox, D. R., and E. Snell. 1989. *Analysis of Binary Data*. 2nd ed. New York, NY: Chapman and Hall.

Day, N. E., and D. F. Kerridge. 1967. "A General Maximum Likelihood Discriminant." *Biometrics* 23: 313–323.

Gourieroux, C. 2000. *Econometrics of Qualitative Dependent Variables*. Cambridge, UK: Cambridge University Press.

Greene, W. H. 2011. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression*. 2nd ed. New York: John Wiley and Sons, Inc.

Kay, R., and S. Little. 1987. "Transformations of the Explanatory Variables in the Logistic Regression Model for Binary Data." *Biometrika* 74: 495–501.

Mason, R., R. F. Gunst, and J. T. Webster. 1975. " Regression Analysis and Problems of Multicollinearity." *Communications in Statistics* 4: 277–292.

McFadden, D. 1976. "A Comment on Discriminant Analysis 'Versus' Logit Analysis." *Annals of Economic and Social Measurement* 5/4: 511–523.

Menard, S., 2002. *Applied Logistic Regression Analysis*. Thousand Oaks: Sage.

Menard, S. 2010. *Logistic Regression: From Introductory to Advanced Concepts and Applications*. Los Angeles, CA: Sage Publications, Inc.

Ryan, T. P. 1997. *Modern Regression Methods*. New York: Wiley.

Schaefer, R. L., L. D. Roi, and R. A. Wolfe. 1984. "A Ridge Logistic Estimator." *Communications in Statistics – Theory and Methods* 99–113.

Scrucca, L., and S. Weisberg. 2004. "A Simulation Study to Investigate the Behavior of the Log-Density Ratio Under Normality." *Communications in Statistics: Simulation and Computation* 33: 159–178.

Spanos, A. 1986. *Statistical Foundations of Econometric Modelling*. Cambridge, UK: Cambridge University Press.

Spanos, A., and A. McGuirk. 2002. 'The Problem of Near-Multicollinearity Revisited: Erractic vs Systematic Volatility." *Journal of Econometrics* 108: 365–393.