CESifo WORKING PAPERS

7459 2019

January 2019

Multicollinearity in the Presence of Errors-in-Variables Can Increase the Probability of Type-I Error

John Komlos

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7459*
Category 12: Empirical and Theoretical Methods

Multicollinearity in the Presence of Errors-in-Variables Can Increase the Probability of Type-I Error

Abstract

Multicollinearity, especially in combination with errors-in-variables, can increase the likelihood of a Type-I error by inflating the value of the estimated coefficients by more than it magnifies their standard errors, thereby increasing the likelihood of obtaining statistically significant results. This anomalous result may be due to an interaction effect between errors-in-variables and multicollinearity.

JEL-Codes: C010.

Keywords: multicollinearity, Type I error, errors-in-variables.

John Komlos
Professor Emeritus
University of Munich / Germany
John.Komlos@gmail.com

Comments on an earlier version of this paper by Scott Carson, Kam Hon Chu, Mark Gilthorpe, Volker Grzimek, Carl Mela, Edward Leamer, Hans Schneeweiss, Russel Warne, and Christopher Winship as well as those of Steven Marron and the participants of his statistical seminar at the University of North Carolina at Chapel Hill, are greatly appreciated. Any possible errors are, of course, the sole responsibility of the author. I am also grateful for the assistance of Michael Conroy for carrying out the computations.

It is common knowledge that in the context of regression analysis collinearity (or multicollinearity) can have severe consequences that include "unreliable" estimates (Draper and Smith, 1998); put another way, that it "leads to imprecise estimates of β" (Faraway, 2015, p. 106). In addition to "unstable" estimates, the most frequently expressed concern in textbooks is that it inflates the standard errors of the estimated coefficients and thereby increases the likelihood of a Type-II error (Kennedy, 1998, 184, 190; Asteriou and Hall, 2011, p. 174). However, there are a number of recent iconoclastic studies which emphasize just the opposite, namely, that it can also increase the probability of a Type-I error (false positives) under certain circumstances (Spanos and McGuirk, 2002; Chatelain, 2010; Chatelain and Ralf, 2014; Atems and Bergtold, 2016).[1] In fact, Kalnins goes as far as to conclude that "multicollinearity causes Type-I errors as well as Type-II errors" (Kalnins, 2018). This note reports such a case in the presence of errors-in-variables.

The substantive issues associated with the case at hand can be found elsewhere, and need not be reiterated here (Komlos and A'Hearn, 2016; 2019). Instead, we focus on the technical aspects of estimation in the presence of misspecification associated with multicollinearity in the presence of errors-in-variables.[2] Suffice it to say merely that the original analysis pertains to estimating the trend in the height of soldiers who enlisted during the American Civil War. The baseline model used most often in the literature is:

$$h_{bei} = \alpha + \sum_{b=1832}^{1842} \beta_b \cdot d_{bi} + \sum_{e=1862}^{1865} \gamma_e \cdot d_{ei} + \varepsilon_{bei} \qquad (1)$$

where h_{bei} is the height of soldier i, born in year b, and enlisting in year e, d_{bi} is a dummy variable for birth years, d_{ei} is a dummy variable for the years of enlistment that allow the level of heights to differ by enlistment year, α is a constant, β_b, γ_e are coefficients to be estimated, and ε_{bei} is a random disturbance term.[3] The reason it is important to control for the enlistment year is that the height of recruits decreased over the course of the conflict, as its bloody and prolonged nature became increasingly obvious, poorer, hence shorter, men were more likely to enter the military.

However, Bodenhorn, Guinnane, and Mroz recently questioned the usefulness of the above model arguing that selection on unobservables might bias the results (2017, pp. 177, 194). In order to test that hypothesis, BGM introduced 24 interaction dummy variables between birth-years and enlistment-years to the basic model[4] (BGM, 2017, Table 5, Model 4):

$$h_{bei} = \alpha + \sum_{b=1832}^{1842} \beta_b \cdot d_{bi} + \sum_{e=1862}^{1865} \gamma_e \cdot d_{ei} + \sum_b \sum_e \delta_{be} \cdot d_{bi} \cdot d_{ei} + \varepsilon_{bei} \qquad (2)$$

where the $(d_{bi} \cdot d_{ei})$ interactions allow height trends to differ by enlistment year. In the absence of varying selection effects, they argue, the interaction between enlistment-year and birth-year should not be statistically significant because only adults are included in the sample (between the ages of 23 and 30) and their height should not vary after controlling for both birth year and enlistment year insofar as humans stop growing in adulthood.[5]

In order to test their hypothesis, BGM ran the regression specified in Equation 2 on a data set of 7,458 records originally collected by Robert Fogel (Fogel et al., 1990). These include all the native-born white soldiers within the above age range in the original sample. BGM find that the interaction variables' X^2 (Chi-square) statistic's p-value is 0.17, i.e., they are jointly not statistically significant (BGM, 2017, p. 196). In other words, they do not find the hypothetical selection effects on unobservables and therefore should have rejected their hypothesis (BGM, 2017, Table 5, Model 4, Column 1). However, they did not stop there.

Instead, they repeated the regression on a subset of this data set with N=3,245. (We shall refer these as the large and the small data sets.) While the large data set uses the age recorded at the time of enlistment (age1), the small data set uses an alternative age variable that was recorded decades later, derived from pension applications. It is noteworthy that the small data set is not a random sample from the large data set because the small data set is comprised of those records on which there is another age variable available on the soldiers (age2). However, age2 is less accurate than age1, because the former was recorded decades after the war, when the veteran applied for a pension and the petitioner had an incentive to misreport his age in order to qualify. In addition, the smaller sample also suffers from survival bias insofar as it pertains only to those veterans who not only survived the war but survived into the 1890s as well (Costa, 1995, p. 301).[6] Thus, the small data set is not a random sample of the soldiers and is also plagued by errors-in-variables. Nonetheless, they do run the regression on the small data set and report jointly statistically significant interaction effects of Equation 2 (BGM, 2017, Table 5, Model 4, Column 2).

However, they fail to address the obvious inconsistency between the two results; these are anomalous, because with a smaller number of observations combined with errors-in-variables one would expect that the estimates would be more likely to be less significant rather than more so (Goodhue, Lewis and Thompson, 2018). Hence, this anomaly is sufficiently puzzling to warrant further exploration. Upon some analysis of this mystery we find that their results reported in their Table 5, row 4, bring to light an unusual and often overlooked aspect of multicollinearity. As it turns out, the interaction terms of Equation 2 introduce severe multicollinearity into the regressions that was exacerbated by the use of the less accurate age2 variable.

A common way to measure multicollinearity is with the variance inflation factor (calculated for each of the independent variables). It is defined as $VIF_i = \frac{1}{1-R_i^2}$, where R_i^2 is the coefficient of determination obtained from the auxiliary regression of each of the x_i on all the other regressors.[7] The VIFs multiply the standard estimator of the variance of the least-squares coefficients. Clearly, VIF_i increases as $R_i^2 \to 1$ thereby increasing the variance of the estimated coefficient of the i'th variable. VIF values less than 5 are considered within the normal range while those above 5 is taken as evidence of some degree of multicollinearity. VIF values above 10 are considered indicative of extreme multicollinearity.

We first report the VIF values of the enlistment year variables. Those of the baseline model (Eq. 1) are all well below 5, indicating that multicollinearity is not present (Table 1, column 2). Their average value is just 1.59. In contrast, the VIF values of the same enlistment year variables in Equation 2 (with the interaction terms) are extremely large for both small and large data sets with values as high as 56.1 (Table 1, Columns 5 and 6). Their average value is much greater than 10: (VIFs $\approx$ 28) indicating the presence of severe multicollinearity. Thus, there is a substantial difference between the VIF values of Equation 1 and Equation 2.[8]

Table 1. A Comparison of Three Sets of Estimates of the Enlistment-Year Coefficents (γ), Union Army Soldiers									
	___No Interaction__		__________With Interaction___________				____________Magnification_____________		
	Coefficient	VIF	Coefficient (γ)		VIF				
	γ		N= Small	N=Large	N= Small	N=Large	N= Small	N=Large	Ratio (7/8)
	1	2	3	4	5	6	7	8	9
1862	0.01	1.66	-1.22	-0.90	56.13	52.64	-174.8	-128.7	1.36
1863	-0.46	1.23	-1.61	-1.20	13.45	15.16	3.50	2.62	1.34
1864	-0.50	1.84	-2.46	-1.32	36.01	34.58	4.90	2.62	1.87
1865	-0.62	1.64	-1.93	-1.26	8.62	8.56	3.13	2.05	1.53
Mean	-0.39	1.59	-1.81	-1.17	28.55	27.74	4.60	2.98	1.52

Note: "No interaction" refers to Equation 1 whereas "With Interaction" refers to Equation 2. Calculations are based on data in Fogel et al., (1990).

We next consider the effect of multicollinearity on the size of the estimated enlistment-year coefficients themselves ($\hat{\gamma}_e$). Those from the regression Equation (1) *without* interaction terms are of reasonable size (Table 1, Column 1). (It is common knowledge that as the war progressed, poorer, less well-nourished, and hence somewhat shorter men were recruited into the Union Army.) However, the inclusion of the interaction terms (Equation 2) multicollinearity magnifies these coefficients. Their average values are between -1.17 and -1.81 compared to -0.39 without the interaction terms (Table 1, Columns 1, 3, and 4).[9] From a Bayesian perspective, the enlistment-year coefficients of Equation 2 are well outside of the realistic range (Leamer, 1994). Such large coefficient estimates are not unheard of in case of multicollinearity. As Greene suggests, in the presence of multicollinearity "coefficients may have… implausible magnitudes" (Greene, 2003, 57). In fact, the coefficients are magnified relative to the baseline values in Column 1 on average by a factor of 2.98 (= -1.17/-0.39) and 4.6 (= -1.81/-0.39) (Table 1, Columns 7 and 8).[10] Thus, the coefficients ($\hat{\gamma}_e$) from the small data set are magnified on average by a factor of 1.52 (= 4.6/2.98) times as much as those of the large data set (Table 1, Column 9).

While the enlistment-year coefficients ($\hat{\gamma}_e$) are magnified greatly in the negative direction, the estimated interaction coefficients ($\hat{\delta}_{be}$) are similarly magnified but with the opposite sign (Table 2, Columns 1 and 4). The 24 estimated interaction coefficients are reported in Table A1 and their mean values by enlistment year are reported in Table 2 (Columns 1 and 4). Note that the mean of the $\hat{\delta}_{be}$'s are all large and positive whereas those of the $\hat{\gamma}_e$'s are all large (in absolute value) and negative (Table 2, Columns 2 and 5). Hence, the δ_{be}'s are also unrealistically large especially those of the small data set: 19 out of the 24 coefficients exceed 1 inch (Table A1, Column 1). It is unreasonable to suppose that soldiers born in 1832 and recruited in 1862 were 1.8 inches taller than those born and recruited a year earlier after controlling for enlistment year effects. Hence, because of multicollinearity, both the $\hat{\delta}_{be}$'s and the $\hat{\gamma}_e$'s are unreasonably large (in absolute value), but they offset each other remarkably. Their sums (Table 2, Columns 3 and 6) are practically identical to the coefficients obtained with the baseline regression without the interaction terms (Table 2, Columns 7, 8 and 9). They differ from the baseline on average merely by between -0.16 and +0.09 inches (Table 2, Columns 8 and 9).

Table 2. Interaction Coefficient Estimates (δ), Summary Statistics

	Regressions with Interactions Eq. 2							Baseline (Eq. 1)		
	N=Small				N=Large			Regression	Difference	
	Mean of	Enlistment	Sum (δ+γ)		Mean of	Enlistment	Sum (δ+γ)	Enlistment	N=Small	N=Large
	δ's	γ	1+2		δ's	γ	4+5	γ	7-3	7-6
	1	2	3		4	5	6	7	8	9
1862	1.35	-1.22	0.13		0.90	-0.90	0.00	0.01	-0.12	0.01
1863	1.32	-1.61	-0.29		0.59	-1.20	-0.61	-0.46	-0.17	0.15
1864	1.89	-2.46	-0.57		0.59	-1.32	-0.73	-0.50	0.07	0.22
1865	1.72	-1.93	-0.21		0.68	-1.26	-0.58	-0.62	-0.41	-0.03
Grand Mean	1.57	-1.81	-0.24		0.69	-1.17	-0.48	-0.39	-0.16	0.09

Note: The average values of the δ's are from Table A1.

While the VIF values are not very different for the small and for the large data sets (9.9 vs. 9.4), the smaller data set has the additional problem of using the less accurate age variable. Presumably this is why the statistical tests are less reliable with the small data set. It is known that measurement errors increase the problems associated with multicollinearity: "the fit becomes very sensitive to measurement errors where small changes in *y* can lead to large changes in $\hat{\beta}$" (Faraway, 2015, p. 106). So, the less accurate age variable in the small regression compounds the problem of multicollinearity and thus has a larger impact on the estimates. Such results are possible in the presence of multicollinearity with misspecified models: "some estimates may be nonsensical… [and] may be highly sensitive to the model specification… [and] can create problems…when the model is misspecified… [and] can enormously magnify the effects of model misspecification" (Winship and Western, 2016, pp. 628-629). Equation 2 with the interaction terms correspond to such a misspecification and the less accurate age variable exacerbates the problem of multicollinearity.

Hence, the joint significance of the X^2 statistic of the interaction terms in the smaller data set is due to misspecification of the model caused by multicollinearity. "[A] serious consequence of multicollinearity is that a slight modification of the data might induce substantial changes in the results of the regression analysis…" (Tu and Gilthorpe, 2016, 81). Another reason for the above result might be that standard errors can actually decrease because of multicollinearity: "collinearity can reduce parameter variance estimates" (Mela and Kopalle, 2002).

To be sure, the standard error does not decrease in this case. However, multicollinearity in the small sample increases the coefficient estimate by more than it increases its estimated standard errors. The coefficients $\widehat{\delta}'s$ are on average 2.23 (=1.54/0.69) times larger in the small data set than in the large one, but their standard errors are only 1.52 (=0.78/0.51) times as large (Table A1 columns 7 and 8). This implies that the Wald z-tests (coefficient/se) in Table A1 are on average 1.50 (=2.08/1.37) times as large in the small sample than in the large one (Table A1 Column 9). No wonder that the X^2 statistic is significant in the small subsample even though it is insignificant in the large one. The less accurate age variable magnifies the instability due to multicollinearity. Johnston actually mentions such a possibility: "It is also possible to find… highly significant t values on individual coefficients, even though multicollinearity is serious. This can arise if individual coefficients happen to be numerically well in excess of the true value, so that the effect still shows up in spite of the inflated standard error" (Johnston, 1984, 249).[11] Kalnins also reports that multicollinearity can inflate a coefficient by more than it increases its standard error (Kalnins, 2018).

In fact, three experiments support the view that the less accurate age variable magnifies the impact of multicollinearity. First we substituted the more accurate age value (age1) into the small data set (with 3,245 observations) instead of the less accurate age value (age2). We found that the X^2 (Chi-square) statistic's p-value becomes 0.8, i.e. the interaction terms are no longer jointly significant although the average of the VIF values is still large (=9.4). This points to the variable age2 causing the interaction terms to become significant.

Next, we drew 20 random samples of 3,245 observations each from the large data set (N=7,458) and used the more accurate age variable (age1) in these regressions. We obtained just one case in which the interaction terms were jointly statistically significant at the 5% level.[12] In other words, we obtained false positives in just 5% of the cases, as we would expect with the significance level set at 5%. The statistical significance is not enhanced due to the small size of the data set.

Finally, we ran 100 regressions with the large data set (N=7,458) but this time used the less accurate age variable (age2) if it was available (in roughly 43% of the cases) but used the more accurate age1 variable for the missing cases (in roughly 57% of the cases). The results indicate that in six out of the 100 cases the interaction terms were jointly statistically significant. This is one more than the number one would expect when testing 100 hypotheses at the 5% significance level. Hence, the use of the less accurate age variable in about 43% of the cases did not suffice to increase the probability of Type-I error excessively but it did increase it slightly. In sum, errors-in-variables do magnify the likelihood of a Type-1 error in regressions plagued with multicollinearity.

Carl Mela, professor of marketing at Duke University, suggested that one can gain some intuition why multicollinearity can inflate the coefficient estimates with opposite signs by considering the following example. Suppose the correct model is:

$$y = \beta_0 + \beta_3 x_1 + \varepsilon \qquad (3)$$

but instead we estimate a misspecified model (Asteriou and Hall, 2011, p. 173):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \qquad (4)$$

and suppose furthermore that x_1 and x_2 are perfectly collinear so that $x_1 = x_2$. Then

$$y = \beta_0 + (\beta_1 + \beta_2)x_1 + \varepsilon \qquad (5)$$

and thus $\beta_3 = \hat{\beta}_1 + \hat{\beta}_2$. Of course, the coefficient β_3 remains constant for infinitely many estimated combinations of $\hat{\beta}_1$ and $\hat{\beta}_2$. So, if $\hat{\beta}_1$ increases with sampling variation, $\hat{\beta}_2$ must decrease by the same amount so that their sums remain constant and equal to the true value of β_3. Thus, the changes in $\hat{\beta}_1$ and $\hat{\beta}_2$ will be of similar magnitude (in absolute value) but have opposite signs (See also Schneeweiss, 134).[13] Schneeweiss also shows that a positive correlation between x_1 and x_2 leads to a covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ of the same magnitude (in absolute value) but with the opposite sign[14] (Schneeweiss, 1990, 139). This is why both the enlistment year parameters and the interaction-term variables in the above example are magnified almost identically in absolute value but with the opposite sign so that they move in the opposite direction and essentially offset each other. Their combined effect is within -0.16 to +0.09 inches of the enlistment-year coefficients estimated with the baseline model without the interaction terms (Table 2, Columns 8 and 9). In other words, the two collinear variables repel each other as the same poles of two magnets.[15]

Hence, we came across an oft-overlooked property of multicollinearity, namely, that it can actually induce some variables to become statistically significant by inflating the estimated coefficients by more than it magnifies their standard errors. In such cases multicollinearity "may be associated with large and false t-statistics" (Kalnins, 2018), or "excessive false positives" especially in the presence of errors-in-variables (Goodhue, Lewis and Thompson, 2018), thereby revealing a "statistical blind spot" in multiple regression analysis (Goodhue, Lewis, and Thompson, 2017, 668). Thus, multicollinearity is a more serious issue then is currently commonly acknowledged: "Because of a lack of awareness about the M+ME [multicollinearity + measurement error] blind spot, it appears that too often we may have mislead ourselves into believing we had support for hypotheses, when actually we did not" (Goodhue, Lewis and Thompson, 2017, 682).

References

Asteriou, Dimitrios and Stephen G. Hall, 2011. *Applied Econometrics*, 2nd ed., (New York: Palgrave Macmillan).

Atems, Bebonchu and Jason Bergtold, 2016. "Revisiting the statistical specification nof ner-multicollinearity in the logistic regression model," Studies in Nonlinear Dynamics and Econometrics 20, 2: 199-210.

Baltagi, Badi, H. 2008. *Econometrics* (4th ed. Berlin: Springer).

Bodenhorn, Howard, Timothy Guinnane, and Thomas Mroz. 2017. "Sample-Selection Biases and the "Industrialization Puzzle," *The Journal of Economic History* 77, 1: 171-207.

Chatelain, Jean-Bernard and Ralf, Kirsten. 2014. “Spurious regressions and near-multicollinearity, with an application to aid, policies and growth,” *Journal of macroeconomics* 39, Iss. A: 85-96.

Chatelain, Jean-Bernard, 2010. “Can Statistics Do without Artefacts?” Unpublished Manuscript, https://hal.archives-ouvertes.fr/hal-00750495/document

Costa, Dora, L. “Pensions and Retirement: Evidence from Union Army Veterans.” *Quarterly Journal of Economics* 110, no. 2 (1995): 297-319.

Draper, Norman R. and Harry Smith, 1998. *Applied Regression Analysis* (3rd ed., John Wiley & Sons).

Faraway, Julian J. 2015. *Linear Models with R* (2nd ed. Boca Raton: Chapman &Hall, CRC Press).

Fogel, Robert W., Stanley L. Engerman, Clayne Pope, et al. “Union Army Recruits in White Regimens in the United States, 1861-1865.” [Computer file]. (ICPSR 9425-v2). Chicago: University of Chicago, Center for Population Economics [producer], Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1990.

Goldberger, Arthur, 1991. *A course in Econometrics* (Cambridge, MA: Harvard University Press).

Goodhue, Dale, Will Lewis and Ron Thompson, 2017. “A Multicollinearity and Measurement Error Statistical Blind Spot: Correcting False Positives in Regression and PLS,” *MIS Quarterly*, 41, 3: 667-684.

Goodhue, Dale, Will Lewis and Ron Thompson, 2018. “What Do You Mean, My Results are Incorrect? The Impact of Multicollinearity and Measurement Error in Tests of Statistical Significance,” advances in Management Information Systems Research, August.

Greene, William H., 2003. *Econometric Analysis* (5th ed. Upper Saddle River, NJ: Prentice Hall).

Johnston, John, 1984. *Econometric Methods* (3rd edition, New York: McGraw-Hill).

Kalnins, Arturs, 2018. “Multicollinearity: How common factors cause Type 1 errors in multivariate regression,” *Strategic Management Journal* 22, issue 10, May.

Kennedy, Peter, 1998. *A Guide to Econometrics* (Cambridge, MA: The MIT Press).

Komlos, John and Brian A'Hearn. 2016. The Decline in the Nutritional Status of the U.S. Antebellum Population at the Onset of Modern Economic Growth. Cambridge, MA. NBER Working Paper 21845, January.

Komlos, John and Brian A'Hearn. 2019. "Clarifications of a Puzzle: the Decline in Nutritional Status at the Onset of Modern Economic Growth in the U.S.A." with Brian A'Hearn, *Journal of Economic History*, forthcoming, September.

Leamer, Edward E. 1994. *Sturdy Econometrics* (Hants, England: Edward Elgar).

Mela, Carl F. and Praveen K. Kopalle. 2002. "The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations." Applied Economics 34: 667-677.

Schneeweiss, Hans. 1990. *Ökonometrie* (Heidelberg: Physica-Verlag), 4th edition.

Tu, Yu-Kang and Mark S. Gilthorpe, 2016. *Statistical Thinking in Epidemiology* (Boca Raton: CRC Press)

Winship, Christopher and Bruce Western. 2016. "Multicollinearity and Model Misspecification," *Sociological Science* 3: 627-649.

Appendix

Table A1. The 24 Estimated Interaction Coefficients (δ) of Equation 2										
Enlistment	Birth	N=Small			N=large			Magnification		
Year		δ	s.e.	Wald z	δ	se	Wald z	δ	s.e.	Wald z
		1	2	3	4	5	6	7	8	9
1862	1832	1.82	0.79	**2.31**	0.91	0.50	1.8	2.0	1.6	1.3
	1833	0.90	0.76	1.19	0.96	0.50	**1.95**	0.9	1.5	0.6
	1834	1.03	0.74	1.38	0.88	0.49	1.81	1.2	1.5	0.8
	1835	1.57	0.74	**2.12**	0.70	0.48	1.46	2.2	1.5	1.5
	1836	1.75	0.72	**2.43**	1.18	0.47	**2.49**	1.5	1.5	1.0
	1837	1.43	0.71	**2.01**	0.80	0.47	1.72	1.8	1.5	1.2
	1838	0.93	0.66	1.41	0.85	0.42	**2.01**	1.1	1.5	0.7
	Mean	1.35			0.90					
1863	1833	0.72	1.17	0.61	0.93	0.67	1.39	0.8	1.7	0.4
	1834	2.09	1.02	**2.04**	0.70	0.68	1.04	3.0	1.5	2.0
	1835	2.08	0.89	**2.33**	0.40	0.60	0.67	5.2	1.5	3.5
	1836	0.72	1.01	0.72	1.10	0.62	1.78	0.7	1.6	0.4
	1837	1.47	0.89	1.65	0.96	0.60	1.6	1.5	1.5	1.0
	1838	0.95	0.88	1.08	0.27	0.62	0.44	3.5	1.4	2.5
	1839	1.22	0.93	1.31	-0.21	0.59	-0.36	5.7	1.6	-3.6
	Mean	1.32			0.59					
1864	1834	2.05	0.72	**2.85**	0.95	0.49	**1.94**	2.2	1.5	1.5
	1835	1.85	0.73	**2.54**	0.83	0.50	1.68	2.2	1.5	1.5
	1836	2.49	0.73	**3.42**	0.97	0.48	**2.03**	2.6	1.5	1.7
	1837	2.08	0.70	**2.97**	0.79	0.47	1.67	2.6	1.5	1.8
	1838	2.16	0.65	**3.34**	1.09	0.44	**2.5**	2.0	1.5	1.3
	1839	1.22	0.67	1.82	-0.44	0.44	-1.01	2.8	1.5	-1.8
	1840	1.35	0.59	**2.27**	-0.06	0.42	-0.14	23.5	1.4	-16.2
	Mean	1.89			0.59					
1865	1835	1.61	0.73	**2.21**	0.85	0.46	**1.85**	1.9	1.6	1.2
	1836	1.90	0.65	**2.94**	1.36	0.45	**3.04**	1.4	1.4	1.0
	1837	1.65	0.66	**2.50**	-0.16	0.44	-0.38	10.0	1.5	-6.6
	Mean	1.72			0.68					
Grand Mean		1.54	0.78	**2.06**	0.69	0.51	1.37	2.23	1.52	1.50

Endnotes

[1] Atems and Bergtold conclude in the context of logistic regression that in the presence of near-multicollinearity "the parameters and their associated variances and t-ratios may be different than the traditional account implies" (Atems and Bergtold, 2016, 210). Similarly, Spanos and McGuirk "demonstrate that increasing the correlation among the regressors does not necessarily… worsen the significance of the coefficients" (Spanos and McGuirk, 2002, 366).

[2] A "consequence of multicollinearity is that it can easily lead to specification errors" (Kennedy, 1998, 185). Others warn about spurious inference in the presence of multicollinearity and suppressor variables (Chatelain and Ralf, 2014). (Suppressor variables are correlated with the independent variable(s) but not with the dependent variable, yet they improve the R^2 of the regression.)

[3] Because of minimum height requirements, the model is estimated by reduced-sample maximum likelihood truncated regression using STATA (Komlos, and A'Hearn, 2016).

[4] There are 40 possible interaction terms between birth-years and enlistment-years observed in the sample (5 enlistment years times 8 ages from 23 to 30). However, Equation 2 has eleven birth-years (*b*) and four enlistment-years (*e*) with the 1831 birth year and the 1861 enlistment year omitted as the reference groups. Including the constant, that makes sixteen parameters to be estimated in addition to the interaction terms. This implies that only 24 of the possible 40 (*b* x *e*) interaction effects can be estimated and 16 have to be omitted.

[5] In other words, after controlling for enlistment-year effects, a 23-year-old born in 1840 and enlisting in 1863 should be as tall on average as a 24-year-old born in 1840 and enlisting in 1964.

[6] Dora Costa, who oversees the Union Army Data website, confirmed that age1 is more accurate than age2.

[7] The VIF's are a property of the independent variable and are therefore not affected by the fact that the dependent variable (height) is truncated.

[8] Similarly, the average VIFs for all the variables for the baseline regression (including the birth cohorts) are 4.39 and 4.96, whereas the introduction of the interaction terms increases the average VIF values for all the variables to 9.4 for the large data set and 9.9 for the small one.

[9] However, the birth-year effects estimated with Equation 2 (that include the interaction effects) are identical to the ones estimated with Equation 1. There is no difference between them to two decimal places.

[10] Note that the magnification values for enlistment year 1862 are extremely large on account of the tiny baseline estimate.

[11] Goldberger quotes this statement from Johnston and Baltagi and confirms that the t-statistic of an estimated coefficient can be significant in the presence of multicollinearity (Goldberger, 1991, 247; Baltagi, 2008, 76).

[12] The mean of the VIF values of the 20 draws was 9.5 with a range between 8.6 and 11.0. The VIF value of the regression with the significant interaction terms were close to the mean values.

[13] To be sure, this does not explain why the estimated coefficients are inflated by more than their standard errors.

[14] Similarly, Johnston writes, "a positive covariance for the X's gives a negative covariance for the b's, and vice versa… if b_2 is below β_2, b_3 is most likely to exceed b_2, β_3, and vice versa (provided the X's are positively correlated)" (Johnston, 1984, 240). This property explains the findings reported in Table 2.

[15] Schneeweiss states that while it is true that the estimated coefficients remain unbiased with multicollinearity, "this does not preclude that in particular cases the estimated value can differ considerably from the true parameter value. This is especially true under multicollinearity. There the parameter estimate may hugely deviate in either direction, and it is of no help to know that the estimate is unbiased…. So, the problem of multicollinearity is not bias but the extreme variance of the parameter estimates, making them completely unreliable (personal communication).