



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, WZB, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPSTEIN, Columbia University
ALLAN GREGORY, Queen’s University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit,
London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
JEROEN WEESIE, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

DAVID CULWELL and DEIRDRE SKAGGS

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$ 98	1-year subscription	\$138
2-year subscription	\$165	2-year subscription	\$245
3-year subscription	\$225	3-year subscription	\$345
1-year student subscription	\$ 75	1-year student subscription	\$ 99
1-year institutional subscription	\$245	1-year institutional subscription	\$285
2-year institutional subscription	\$445	2-year institutional subscription	\$525
3-year institutional subscription	\$645	3-year institutional subscription	\$765
Electronic only		Electronic only	
1-year subscription	\$ 75	1-year subscription	\$ 75
2-year subscription	\$125	2-year subscription	\$125
3-year subscription	\$165	3-year subscription	\$165
1-year student subscription	\$ 45	1-year student subscription	\$ 45

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.



Copyright © 2014 by StataCorp LP

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, fileservers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **STATA**, Stata Press, Mata, **MATA**, and NetCourse are registered trademarks of StataCorp LP.

Stata tip 118: Orthogonalizing powered and product terms using residual centering

Carsten Sauer
Bielefeld University
Bielefeld, Germany
carsten.sauer@uni-bielefeld.de

In multiple regressions, powered variables are commonly included to represent higher-order nonlinear effects. Interaction effects are often represented by the product of two variables. In many cases, including powered or product terms leads to increasing correlation or collinearity. This correlation between the powered or the product term and the first-order predictor variable can lead to unstable estimates and bouncing beta weights. To tackle this problem, one can center the predictor variables before the estimation at the means (Cohen 1978). An alternative approach to mean centering is residual centering. Unlike mean centering, residual centering ensures orthogonality between the powered or the product term and the first-order predictor variables. The theoretical background of residual centering in comparison with mean centering is described in Little, Bovaird, and Widaman (2006). Technically, residual centering is a two-stage ordinary least-squares regression procedure introduced by Lance (1988). In the first step, the powered term or the product term is regressed on the first-order predictor variables. The residuals of this regression are then used to represent the powered or the product term. The variance of the powered or the product term obtained by this regression is independent of the variance of the first-order predictor variable.

The following examples show the residual centering method for powered and product terms. First, we run a regression with one linear effect; then we include the powered term.

```
. sysuse auto
(1978 Automobile Data)
. quietly regress price weight
. estimates store M1
. generate weight2 = weight^2
. quietly regress price weight weight2
. estimates store M2
```

```
. estat vif
```

Variable	VIF	1/VIF
weight	58.95	0.016963
weight2	58.95	0.016963
Mean VIF	58.95	

The variance inflation factor is 58.95, which indicates multicollinearity. Now we calculate the residual-centered powered term and rerun the regression model.

```
. quietly regress weight2 weight
. predict weight2_rc, res
. quietly regress price weight weight2_rc
. estimates store M3
. estat vif
```

Variable	VIF	1/VIF
weight	1.00	1.000000
weight2_rc	1.00	1.000000
Mean VIF	1.00	

The test for collinearity indicates that perfect orthogonality could be achieved via residual centering. The following table shows the coefficients of the models with only the linear effect (M1), the quadratic effects without residual centering (M2), and the quadratic effects with residual centering (M3). The results show that including the interaction term in the third model does not change the coefficients of the first-order variable and the constant term. Note that residual centering does not affect the proportion of explained variance (R^2).

```
. estimates table M1 M2 M3, b(%8.3f) stats(r2 N) star
```

Variable	M1	M2	M3
weight	2.044***	-7.273**	2.044***
weight2		0.002***	
weight2_rc			0.002***
_cons	-6.707	1.3e+04**	-6.707
r2	0.290	0.394	0.394
N	74	74	74

legend: * p<0.05; ** p<0.01; *** p<0.001

The second example shows the residual centering for the interaction terms of two continuous variables, weight and length.

```
. quietly regress price weight length
. estimates store M4
. generate interact = weight*length
. quietly regress price weight length interact
. estimates store M5
. estat vif
```

Variable	VIF	1/VIF
interact	155.59	0.006427
weight	91.62	0.010915
length	21.84	0.045795
Mean VIF	89.68	

```
. quietly regress interact weight length
. predict interact_rc, res
. quietly regress price weight length interact_rc
. estimates store M6
```

The test for collinearity indicates results similar to those shown above. The variance inflation factor is very high for the interaction model without centering and remarkably lower for the interaction model with residual centering.

```
. estat vif
```

Variable	VIF	1/VIF
length	9.52	0.105068
weight	9.52	0.105068
interact_rc	1.00	1.000000
Mean VIF	6.68	

The following table shows the coefficients for the two first-order predictor variables (M4), the interaction effect without residual centering (M5), and the interaction effect with residual centering (M6). Again including the orthogonalized product term does not change the coefficients for the first-order predictor variables and the constant term.

```
. estimates table M4 M5 M6, b(%8.3f) stats(r2 N) star
```

Variable	M4	M5	M6
weight	4.699***	-0.967	4.699***
length	-97.960*	-174.570**	-97.960*
interact		0.029	
interact_rc			0.029
_cons	1.0e+04*	2.5e+04**	1.0e+04*
r2	0.348	0.375	0.375
N	74	74	74

legend: * p<0.05; ** p<0.01; *** p<0.001

Residual centering is an easy two-step procedure that makes it possible to obtain unbiased powered or interaction effects and main effects. While it was originally used for multiple regression models, Little, Bovaird, and Widaman (2006) extend the residual centering procedure to represent latent variable interactions in structural equation modeling. Geldhof et al. (2012) provide an overview about applications and caveats for latent interactions.

References

- Cohen, J. 1978. Partialled products are interactions; partialled powers are curve components. *Psychological Bulletin* 85: 858–866.
- Geldhof, G. J., S. Pornprasertmanit, A. M. Schoemann, and T. D. Little. 2012. Orthogonalizing through residual centering: Extended applications and caveats. *Educational and Psychological Measurement* 73: 27–46.
- Lance, C. E. 1988. Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Applied Psychological Measurement* 12: 163–175.
- Little, T. D., J. A. Bovaird, and K. F. Widaman. 2006. On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal* 13: 497–519.