Mean-Centering Does Nothing for Moderated Multiple Regression

Raj Echambadi
Associate Professor
Dept. of Marketing
College of Business Administration
University of Central Florida
P.O. Box 161400
Orlando, FL 32816-1400
Email: rechambadi@bus.ucf.edu

James D. Hess
Bauer Professor of Marketing Science
Dept. of Marketing and Entrepreneurship
C.T. Bauer College of Business
375H Melcher Hall
University of Houston
Houston, TX 77204
Email: jhess@uh.edu

Submitted to the *Journal of Marketing Research*
December 2004

The names of the authors are listed alphabetically. This is a fully collaborative work. We thank Edward A. Blair for his many helpful suggestions in crafting this article.

Mean-Centering Does Nothing for Moderated Multiple Regression

Abstract

The cross-product term in moderated regression may be collinear with its constituent parts, making it difficult to detect main and interaction effects. The commonplace response is to mean-center. However, we prove that mean-centering neither changes the computational precision of parameters, the sampling accuracy of main, simple, and interaction effects, nor R^2.

Moderated multiple regression models are widely used in marketing and have been the subject of much scholarly discussion (Irwin and McClelland 2001; Sharma, Durand, and Gur-Arie 1981). The interaction (or moderator) effect in a moderated regression model is estimated by including a cross-product term as an additional exogenous variable as in

(1) $$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \alpha_0 + \alpha_c x_c + \varepsilon,$$

where x_c plays the role of other covariates that are not part of the moderated element. This x_1x_2 cross-product is likely to be correlated with the term x_1 since we can think of x_2 as a non-constant multiplier coefficient of x_1. This has been interpreted as a form of multicollinearity, and collinearity makes it difficult to distinguish the separate effects of x_1x_2 and x_1 (and/or x_2).

In response to this problem, various researchers including Aiken and West (1991) and Jaccard, Wan, and Turrisi (1990) recommend mean-centering the variables x_1 and x_2 as an approach to alleviating collinearity related concerns. If the variables x_1 and x_2 are mean-centered, then the equation will be of the form

(2) $$y = \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2) + \beta_3(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + \beta_0 + \beta_c x_c + \upsilon.$$

In comparison, the interaction term in (2) involving $(x_1-\bar{x}_1)(x_2-\bar{x}_2)$ will have relatively smaller covariance with the term $x_1-\bar{x}_1$ because the multiplier coefficient, $x_2-\bar{x}_2$, is zero on average.

This practice of mean-centering has become commonplace throughout the social sciences; for example, the Social Science Citation Index shows 2,501 cites for Aiken and West (1991). A review of the influential marketing journals over the past decade reveals that mean-centering has become the standard method by which marketing researchers deal with collinearity concerns in moderated regression models, including eight citations in the *Journal of Marketing Research*, 15 in the *Journal of Marketing*, and four in the *Journal of Consumer Research*. A typical statement taken from Rokkan, Heide, and Wathne (2003, p. 219), which typifies the standard usage of mean-centering, is, "To mitigate the potential threat of multicollinearity, we mean-centered all independent variables that constituted an interaction term (Aiken and West 1991)."

Can such a simple shift in the location of the origin really help us see the pattern between variables? We use a hypothetical example to answer this question. Let the true model for this simulated data be: $y = x_1+½\, x_1x_2+\varepsilon$ where $\varepsilon \sim N(0,0.1)$. In Figure 1a, we graph the relationship between y and uncentered (x_1, x_2). In Figure 1b, we see the relationship between y and mean-centered (x_1, x_2). Obviously the same pattern of data is seen in both the graphs, since shifting the origin of the exogenous variables x_1 and x_2 does not change the relative position of any of the data points. Intuitive geometric sense tells us that looking for statistical patterns in the mean-centered data will neither be easier nor harder than looking for statistical patterns in the uncentered data.

------- Figure 1 about here -------

In fact, Aiken and West (1991, p. 182) made their recommendation not because of better statistical properties, but because of computational reasons stating "As was shown in chapter 4, centering versus not centering has no effect on the highest order interaction term in multiple regression with product variables. However, centering *may be* useful in avoiding computational difficulties." (emphasis added).

In this paper, we will demonstrate that geometric intuition is correct: mean-centering in moderated regression does not help. Specifically, we show the following: 1) in contrast to Aiken and West's (1991) suggestion, mean-centering does not improve the accuracy of numerical computation of statistical parameters, 2) it does not change the sampling accuracy of main effects, simple effects, and/or interaction effects (point estimates and standard errors are identical with or without mean-centering), and 3) it does not change overall measures of fit such as R^2 and adjusted-R^2. It does not hurt, but it does not help, not one iota.

1. Mean Centering Neither Helps Nor Hurts

Straight-forward algebra of equation (1) shows that it is equivalent to:

(3)
$$y = (\alpha_1 + \alpha_3\bar{x}_2)(x_1 - \bar{x}_1) + (\alpha_2 + \alpha_3\bar{x}_1)(x_2 - \bar{x}_2) + \alpha_3(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) \\ + \alpha_0 + \alpha_1\bar{x}_1 + \alpha_2\bar{x}_2 + \alpha_3\bar{x}_1\bar{x}_2 + \alpha_c x_c + \varepsilon.$$

Comparing (2) and (3), there is a linear relationship between the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameter vectors:

(4)
$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_0 \\ \beta_c \end{bmatrix} = \begin{bmatrix} 1 & 0 & \bar{x}_2 & 0 & 0 \\ 0 & 1 & \bar{x}_1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \bar{x}_1 & \bar{x}_2 & \bar{x}_1\bar{x}_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_0 \\ \alpha_c \end{bmatrix} = W\boldsymbol{\alpha}.$$

The inverse of W is easily computed as:

(5)
$$W^{-1} = \begin{bmatrix} 1 & 0 & -\bar{x}_2 & 0 & 0 \\ 0 & 1 & -\bar{x}_1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -\bar{x}_1 & -\bar{x}_2 & \bar{x}_1\bar{x}_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Notice that the determinants of both W and W^{-1} equal 1.

Suppose a data set consists of an $n\times5$ matrix of explanatory variable values $X \equiv [X_1 \vdots X_2 \vdots X_1{*}X_2 \vdots \mathbf{1} \vdots X_c]$, where X_j is a column n-vector of observations of the jth variable, $X_1{*}X_2$ is an n-vector whose typical component is $X_{i1}X_{i2}$, and $\mathbf{1}$ is a vector of ones. The empirical version of (1) is therefore $Y=X\boldsymbol{\alpha}+\varepsilon$. This is equivalent to $Y=XW^{-1}W\boldsymbol{\alpha}+\varepsilon=XW^{-1}\boldsymbol{\beta}+\varepsilon$. It is easily seen that $XW^{-1} \equiv [X_1-\bar{x}_1\mathbf{1} \vdots X_2-\bar{x}_2\mathbf{1} \vdots (X_1-\bar{x}_1\mathbf{1}){*}(X_2-\bar{x}_2\mathbf{1}) \vdots \mathbf{1} \vdots X_c]$, the mean-centered version of the data.

An immediate conclusion is that ordinary least squares (OLS) estimates of (1) and (2) produce identical estimated residuals e, and because the residuals are identical, the R^2 for both formulations are identical. OLS estimators $\mathbf{a}=(X'X)^{-1}X'Y$ and $\mathbf{b}=((XW^{-1})'(XW^{-1}))^{-1}(XW^{-1})'Y$ are related to each other by $\mathbf{b}=W\mathbf{a}$. Finally, the variance-covariance of the uncentered and mean-centered OLS estimators are $S_{\mathbf{a}}=s^2(X'X)^{-1}$ and $S_{\mathbf{b}}=s^2(W'^{-1}X'XW^{-1})^{-1}=s^2W(X'X)^{-1}W'$, where the estimator of σ^2 is $s^2=e'e/(n-5)$.

As noted earlier, Aiken and West (1991) recommend mean-centering because it may help avoid computational problems. What are these unspecified computational problems? Rounding errors may be large when computing the inverse of X'X using finite precision digital calculations. When the determinant of X'X is near zero as it might be with collinear data, the computation of $(X'X)^{-1}$ will eventually lead to division by almost zero (recall $A^{-1}=adj(A)/|A|$ for a square matrix A), which produces rounding error that might make estimates computationally unstable. Each computation done at double-precision on a modern computer will be accurate to at least 15 digits of accuracy, but repeated computations can cause the errors to accumulate. However, since the mid-1980s, major statistical software packages have inverted matrices with singular value decomposition algorithms, which have been shown to dramatically reduce this accumulation compared to Gaussian elimination (Hammarling 1985). McCullough (1999) demonstrates that while cumulative computational errors are indeed possible in statistical software such as SAS and SPSS, for even complex linear regression problems we will get 7 to 10 digits of computational accuracy. This is more than enough computational accuracy for typical purposes, especially given that raw data may come from a survey with one significant digit of accuracy (say, using seven-point scales).

Regardless, Aiken and West (1991) seem to suggest that mean-centering reduces the covariance between the linear and interaction terms, thereby increasing the determinant of X'X and mitigating the roundoff errors in inverting the product matrix. Is this true? In the uncentered data, we must invert $X'X$ and in the centered data we must invert $W'^{-1}X'XW^{-1}$. Intuitively, reducing the collinearity between X_1, X_2, and X_1*X_2 should reduce computational errors. However, mean-centering not only reduces the off-diagonal elements (such as $X_1'X_1*X_2$), but it also reduces the elements on the main diagonal (such as $X_1*X_2'X_1*X_2$). Furthermore, mean-centering has no effect whatsoever on the determinant.

Theorem 1: The determinant of the uncentered data product matrix $X'X$ equals the determinant of the centered data product matrix $W'^{-1}X'XW^{-1}$.

(Proofs of all theorems are relegated to the appendix). Because the source of computational problems in inverting these matrices is a small determinant, the same computational problems exist for mean-centered data as for uncentered data.

Also, assuming that the random variable ε is normally distributed, the OLS $\mathbf{a}$ is normally distributed with a mean $\boldsymbol{\alpha}$, and variance-covariance matrix $\sigma^2 (X'X)^{-1}$. Because $\mathbf{b}$ is a linear combination of these, $W\mathbf{a}$, $\mathbf{b}$ must be normal with mean $W\boldsymbol{\alpha}$, and an estimated variance-covariance matrix $WS_{\mathbf{a}}W'$. As Aiken and West (1991) have shown, estimation of the interaction term is identical for uncentered and centered data; we repeat this for sake of completeness.

Theorem 2: The OLS estimates of the interaction terms α_3 and β_3, a_3 for (1) and b_3 for (2), have identical point estimates and standard errors.

This result generalizes to all other effects as seen in the next three theorems.

Theorem 3: The main effect of x_1 (β_1 from equation (2) or $\alpha_1+\alpha_3\bar{x}_2$ from equation (3)) as measured by the OLS estimate b_1, or by the OLS estimate $a_1+a_3\bar{x}_2$, have identical point estimates and standard errors.

Note that the coefficient α_1 in equation (1) is not the main effect of x_1; the "main effect" means the "average effect" of x_1, namely $\alpha_1+\alpha_3\bar{x}_2$. Instead, the coefficient α_1 is the simple effect of x_1 when $x_2=0$. Algebraic rearrangement of (4) states that this simple effect can also be measured from the main effects found in the mean-centered equation (2) because $a_1=b_1-\bar{x}_2b_3$.

Theorem 4: The simple effect of x_1 when x_2=0 is either α_1 in equation (1), or $\beta_1-\bar{x}_2\beta_3$ from equation (2), and the OLS estimates of each of these (a_1 for (1) and $b_1-\bar{x}_2b_3$ for (2)) have identical point estimates and standard errors.

Theorem 5: The simple effect of x_1 when x_2=1 is either $\alpha_1+\alpha_3$ in equation (1), or $\beta_1-(1-\bar{x}_2)\beta_3$ from equation (2) and the OLS estimates of each of these (a_1+a_3 for (1) and $b_1-(1-\bar{x}_2)$b3 for (2)) have identical point estimates and standard errors.

In summary, although some researchers may believe that mean-centering variables in moderated regression will reduce collinearity between the interaction term and linear terms and will therefore miraculously improve their computational or statistical conclusions, this is not so. We have demonstrated that mean-centering does not improve computational accuracy nor does it change the ability to detect relationships between variables in moderated regression.

2. Comments

Why do so many researchers mean-center their moderated variables? Clearly they do so to counter the fear that by including a term x_1x_2 in the regressors, they will create collinearity with the main regressor, such as x_1, so that it will become difficult to distinguish the separate effects of x_1 and x_1x_2 on y. If we make x_2, the multiplier of x_1 in the interaction term, closer to zero on average, then we can reduce the covariance and correlation. One simple way to do this is to replace the multiplier x_2 by $x_2-\bar{x}_2$. By subtracting the mean, the typical value of the multiplier is zero and hence the covariance between the regressor and the interaction terms is smaller. This appears to reduce the "potential threat of multicollinearity" and hopefully improves our ability to distinguish the effect of changes in x_1 from changes in x_1x_2.

This logic seems plausible, but it is incomplete. Mean-centering not only reduces the covariance between x_1 and x_1x_2, which is "good," but it also reduces the variance of the exogenous variable x_1x_2, which is "bad." For accurate measurement of the slope of the relationship, we need the exogenous variables to sweep out a large set of values; however, mean-centered $(x_1-\bar{x}_1)(x_2-\bar{x}_2)$ has a smaller spread than x_1x_2. When both the improvement in collinearity and the deterioration of exogenous variable spread are considered, mean-centering provides no change in the accuracy with which the regression coefficients are estimated. The complete analysis of mean-centering shows that mean-centering neither helps nor hurts moderated regression.

A point that may confuse some researchers in this regard is that t-statistics for individual regressors may change when data are mean-centered. This does not occur for the x_1x_2 term. As noted by Aiken and West (1991) and shown here, the coefficient and the standard error for the interaction (highest order) term, and hence the significance of this term, will be identical with or without mean-centering. However, t-statistics may change for x_1 or x_2 terms as a result of shifting the interpretation of the effect. In a regression without mean-centering, the coefficients represent *simple* effects of the exogenous variables, i.e., the effects of each variable when the other variables are at zero. When data are mean-centered, the coefficients represent *main* effects of these variables, i.e., the effects of each variable when the other variables are at their mean values. When there is a meaningful interaction between x_1 and x_2, the main effect will not equal the simple effect, and may have a significant t-statistic where the simple effect does not.

We illustrate this point that results of linear effects may change across the uncentered and mean-centered models by running separate regressions on the synthetic data used earlier. Suppose that, as above, the true model is: $y = x_1+½\, x_1x_2+\varepsilon$ where $\varepsilon \sim N(0,0.1)$, and the mean of x_1 is equal to 1.5, and the mean of x_2 is equal to 1.0. Table 1 shows the results of both the uncentered and mean-centered regressions from a simulated sample of n = 121 observations.

------- Table 1 about here -------

Table 1a shows the results of the mean-centered regression model and Table 1b shows the results of the uncentered model. Both the mean-centered and the uncentered models provided an identical fit to the data, and yielded the same model R^2. As expected, the coefficients of the interaction, the standard errors, and the t-statistics obtained from both the models are identical.

An examination of the linear effects from Tables 1a and 1b reveals a different story. The linear effect of x_1 is significant in both the uncentered and mean-centered models, whereas x_2 is significant only in the mean-centered model. As discussed earlier, the significant result for x_2 in the mean-centered model should not be taken to imply that the mean-centering approach is superior in alleviating collinearity concerns. The effects tested in these two models are vastly different (simple effects from the uncentered models vis-à-vis main effects from the centered models), and hence, direct comparisons of the corresponding effects are inappropriate. The infamous "comparison of apples and oranges" metaphor is appropriate. Using equation (1) and $\mathbf{b}=W\mathbf{a}$, we can recover an equally accurate measure of the main effect from the uncentered data, and using equation (2) and $\mathbf{a}=W^{-1}\mathbf{b}$, we can recover an equally accurate measure of the simple effect from the centered data. This equivalence between uncentered and mean-centered models may be viewed as an extension of Irwin and McClleland (2001).

In such a circumstance, should a researcher mean-center or not? One might argue that the main effects are the more meaningful term because they better characterize the overall relationships, so the data should be mean-centered. However, one might also argue that the simple effects are preferable because they provide a more fine-grained understanding of the patterns, so the data should be uncentered. Both arguments may be persuasive, but the choice should be made independent of the spurious rationale pertaining to multicollinearity since the information can be recovered from either approach. Of course, recovery of the proper standard errors requires computing the diagonal elements of matrices such as $WS_{\mathbf{a}}W'$ in the former or $W^{-1}S_{\mathbf{b}}W^{-1}{}'$ in the latter, and this may be more easily accomplished by reversing the data-centering decision. However, mean-centering does not hurt, so there is no need to re-evaluate the conclusions of the many published papers that have used mean-centering as long as the researchers are clear about the proper interpretation of the linear terms.

Due to the fact that mean-centering does not mitigate multicollinearity in moderated regression, one might ask, "What else can be done?" One alternative is to use the residual-centering method proposed by Lance (1986), but this is a distinctly bad idea. Echambadi, Arroniz, Reinartz, and Lee (2004) show that residual-centering biases the x_1 and x_2 effects, which is undesirable. Because collinearity problems cannot be remedied after the data has been collected in most cases, we recommend that researchers carefully design their research studies prior to collecting their data. If feasible, one can address it by using a data collection scheme that isolates the interaction effect (for example, a factorial design). Likewise, if feasible, one can address the loss of power associated with multicollinearity by increasing the sample size; in this regard, Woolridge (2001) notes that the effects of multicollinearity are indistinguishable from the effects of micronumerosity, or small sample sizes.

<u>Summary</u>: Whether we estimate uncentered moderated regression equation (1) or the mean-centered equation (2), all the point estimates, standard errors and t-statistics of the main effects, simple effects, and interaction effects are identical, and will be computed with the same accuracy by modern double-precision statistical packages. This is also true of the overall measures of accuracy such as R^2 and adjusted-R^2.

Table 1
Results from Regression Analysis Utilizing Uncentered and Mean-Centered Terms [a]

True Model: $Y = 1X_1 + 0\,X_2 + ½\,X_1 \times X_2 + \varepsilon$, where $\varepsilon \sim N(0,0.1)$ and $\overline{X}_1 = 1.5$, $\overline{X}_2 = 1.0$

a. Mean-centered model: OLS regression coefficients for main effects

Dependent variable: Y

Variables	Unstandardized Coefficients	t-statistic	Interpretation
Constant	2.237* (0.009)	241.767	
$X_1 - \overline{X}_1$	1.505* (0.029)	51.423	Main effect of X_1 at mean levels of X_2
$X_2 - \overline{X}_2$	0.736* (0.029)	25.157	Main effect of X_2 at mean levels of X_1
$(X_1 - \overline{X}_1)\times(X_2 - \overline{X}_2)$	0.386* (0.093)	4.173	Interaction
	R^2	0.966	
N=121	Adjusted R^2	0.965	

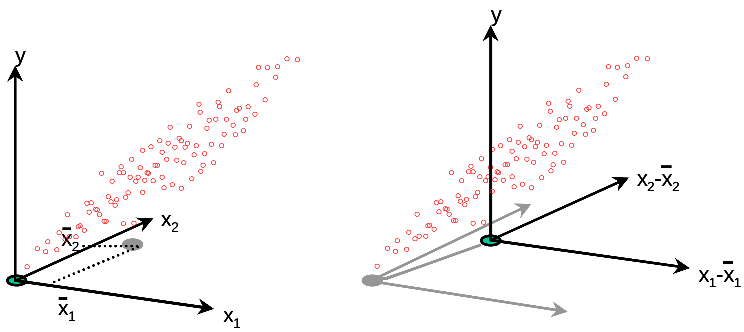
b. Uncentered model: OLS regression coefficients for simple effects

Dependent variable: Y

Variables	Unstandardized Coefficients	t-statistic	Interpretation
Constant	-0.177 (0.149)	-1.189	
X_1	1.119* (0.097)	11.526	Simple effect of X_1 for $X_2 = 0$
X_2	0.157 (0.142)	1.106	Simple effect of X_2 for $X_1 = 0$
$X_1 \times X_2$	0.386* (0.093)	4.173	Interaction
	R^2	0.966	
N=121	Adjusted R^2	0.965	

* significant at 0.01 level;
[a] standard errors are given in parentheses

Figure 1
Graphical Representation of Uncentered and Mean-centered Data in 3D Variable Space



a. Uncentered Data b. Mean-centered Data

References

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage Publications.

Echambadi, Raj., Inigo Arroniz, Werner Reinartz, and Junsoo Lee (2004), "A Critical Analysis of the Residual-Centering Method," Working paper, University of Central Florida.

Hammarling, S. (1985), "The Singular Value Decomposition in Multivariate Statistics," *ACM Special Interest Group in Numerical Mathematics*, 20, 2-25.

Irwin, Julie R. and Gary H. McClelland (2001), "Misleading Heuristics and Moderated Multiple Regression Models," *Journal of Marketing Research*, 38(1), 100-09.

Jaccard, James R., Robert Turrisi, and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Newbury Park, CA: Sage Publications.

Lance, Charles E. (1988), "Residual Centering, Exploratory and Confirmatory Moderator Analysis, and Decomposition of Effects in Path Models Containing Interactions," *Applied Psychological Measurement*, 12 (June), 163-75.

McCullough, B. D. (1999), "Assessing the Reliability of Statistical Software: Part II," American Statistician, 53(2), 149-59.

Rokkan, Aksel I., Jan B. Heide, Kenneth H. Wathne (2003), "Specific Investments in Marketing Relationships: Expropriation and Bonding Effects," *Journal of Marketing Research* 40(2), 210-224.

Sharma, Subhash, Richard M. Durand, and Oded Gur-Arie (1981), "Identification and Analysis of Moderator Variables," *Journal of Marketing Research*, 18 (3), 291-300.

Woolridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Appendix

Proof of Theorem 1: Recall that the determinant of W^{-1} equals 1. So, $\det(W'^{-1}X'XW^{-1})=$

$\det(W'^{-1})\det(X'X)\det(W^{-1})=\det(W^{-1}{}')\det(X'X)\det(W^{-1})=\det(X'X)$. Q.E.D.

Proof of Theorem 2: From the third row of (4), $b_3=a_3$. In this appendix, we will denote $S_{\mathbf{a}}$ by S.

Using matrix multiplication of (4), the third column of SW' is

$$\begin{bmatrix} S_{31} \\ S_{32} \\ S_{33} \\ S_{30} \\ S_{3c} \end{bmatrix}.$$

The third row of W is [0 0 1 0 0], so the 3rd row×3rd column of WSW' is S_{33}. That is,

$SE(b_3)=SE(a_3)=\sqrt{S_{33}}$. Q.E.D.

Proof of Theorem 3: From the first row of (4), the point estimates are equal. The first column of SW' is

$$\begin{bmatrix} S_{11}+\bar{x}_2 S_{13} \\ S_{21}+\bar{x}_2 S_{23} \\ S_{31}+\bar{x}_2 S_{33} \\ S_{01}+\bar{x}_2 S_{03} \\ S_{c1}+\bar{x}_2 S_{c3} \end{bmatrix}.$$

The first row of W is $[1\ 0\ \bar{x}_2\ 0\ 0]$, so the variance of b_1 (the 1st row×1st column of WSW') is $S_{11}+2\bar{x}_2 S_{13}+\bar{x}_2^2 S_{33}$. The variance of $a_1+a_3\bar{x}_2$ is $\mathrm{var}(a_1)+2\bar{x}_2$

$cov(a_1,a_3)+\bar{x}_2^2 var(a_3)= S_{11}+2\bar{x}_2 S_{13}+\bar{x}_2^2 S_{33}$. That is,

$SE(b_1)=SE(a_1+a_3\bar{x}_2)=\sqrt{S_{11}+2\bar{x}_2 S_{13}+\bar{x}_2^2 S_{33}}$. Q.E.D.

Proof of Theorem 4: The variance of $b_1-\bar{x}_2 b_3$ equals $var(b_1)-2\bar{x}_2\, cov(b_1,b_3)+\bar{x}_2^2 var(b_3)$. From the proofs of Theorems 2 and 3 we know that $var(b_3)=S_{33}$ and $var(b_1)= S_{11}+2\bar{x}_2 S_{13}+\bar{x}_2^2 S_{33}$. The first column of SW' is

$$\begin{bmatrix} S_{11}+\bar{x}_2 S_{13} \\ S_{21}+\bar{x}_2 S_{23} \\ S_{31}+\bar{x}_2 S_{33} \\ S_{01}+\bar{x}_2 S_{03} \\ S_{c1}+\bar{x}_2 S_{c3} \end{bmatrix}$$

and the third row of W is [0 0 1 0 0], so the covariance of b_1 and b_3 (the 3rd row×1st column of WSW') is $S_{31}+\bar{x}_2 S_{33}$. Hence the variance of $b_1-\bar{x}_2 b_3$ equals $S_{11}+2\bar{x}_2 S_{13}+\bar{x}_2^2 S_{33} -2\bar{x}_2(S_{31}+\bar{x}_2 S_{33})+ \bar{x}_2^2 S_{33} = S_{11}$. That is, $SE(a_1)=SE(b_1-b_3\bar{x}_2)=\sqrt{S_{11}}$. Q.E.D.

Proof of Theorem 5: a variant of the above.