Reversals of Least-Squares Estimates

Brian Knaeble
University of Utah
June 2013

Abstract
An adjusted estimate may be opposite the original estimate. This paper presents necessary and sufficient conditions for such a reversal, in the context of linear modeling, where adjustment is obtained through considering additional explanatory data associated with lurking variables.

Keywords: Yule–Simpson effect, Thales’ theorem, Multiple regression, Confounding

1 Introduction
The nature of an association between real-valued random variables $X_1$ and $Y$, can be crudely estimated through observing vectors of data, $x_1$ and $y$, and then computing a correlation coefficient $r(x_1, y)$. However, in general, such an estimate is susceptible to confounding: additional data, $x_2$, may lead to an adjusted estimate that is opposite the original estimate.

In this paper we study such reversals in the context of least-squares fitting of linear models. Given $x_1$, $y$, and the fitted coefficient $\hat{\beta}_1 \neq 0$, we seek necessary and sufficient conditions on $x_2$, so that the adjusted estimate, denoted with $\hat{\beta}_1$, is opposite the original estimate $\hat{\beta}_1$, which we henceforth denote with $\hat{\beta}_1$. Throughout this paper, left subscripts indicate those explanatory variables used to fit a model. Note also that we assume always that all vectors of data are of equal length.

Visual intuition suggests that a reversal occurs if and only if the vector $x_2$ lies “between” the vectors $x_1$ and $y$. Geometric analysis in higher dimensions then leads to the conclusion that this region is bounded by a right circular, conical surface of two nappes (this geometric object will be described in detail in Section 2). Such an understanding leads quickly to useful results.

The results are stated using “sign”, where sign of a positive real number is one, sign of a negative real number is negative one, and sign of zero remains undefined; the truth of $\text{sign}(2, \hat{\beta}_1) \neq \text{sign}(1, \hat{\beta}_1)$ entails that both quantities are defined. Note also that $r(y, x_1 + y) = r(x_1, x_1 + y)$.

Theorem 1.1. When $r(y, x_1) \neq 0$, a reversal, $\text{sign}(2, \hat{\beta}_1) \neq \text{sign}(1, \hat{\beta}_1)$, can occur only if $|r(x_2, y)|$ and $|r(x_2, x_1)|$ are both greater than $|r(x_1, y)|$. 
Theorem 1.2. When \( r(y, x_1) \neq 0 \), a reversal, \( \text{sign}(2, \hat{\beta}_1) \neq \text{sign}(1, \hat{\beta}_1) \), occurs if and only if \( |r(x_2, x_1 + y)| > |r(y, x_1 + y)| \).

These results can be generalized so as to regard the robustness of an (already) adjusted estimate, against confounding due to multiple lurking variables. See Theorems 5.1 and 5.2, in Section 5. In particular, a special case of Theorem 5.1 gives necessary and sufficient conditions for the Yule–Simpson effect. See [14], [5] and [7] (Section 2) for an overview of the Yule–Simpson effect.

Mathematics relating to reversals appears in literature from the 1950s regarding smoking and lung cancer (see [2]), and there remains potential for application of such mathematics within epidemiology today: see [10] and [11] for educational articles, and any of [3], [6], [12], [8] or [1] for scientific papers. For related mathematical papers see [13], [9] or [4].

2 Geometry

This section presents the geometric foundation for the proofs of Theorems 1–4. Notation has been chosen so as to best convey the ideas behind the proofs. Readily-verified lemmas are stated without proof.

2.1 Angles and Objects

Definition 2.1. Let \( u_1 \) and \( u_2 \) be two nonzero vectors within an inner product space. Define the angle between them, \( \theta \), to be

\[
\theta(u_1, u_2) = \cos^{-1}\left(\frac{(u_1, u_2)}{|u_1||u_2|}\right).
\]

Definition 2.2. Let \( u \in \mathbb{R}^p \) be a generic vector with coordinates \((u_1, u_2, \ldots, u_p)\).

For a fixed \( \theta \), with \( 0 < \theta < \pi/2 \), define the following geometric objects:

\[
V_\theta := \left\{ u \in \mathbb{R}^p : \frac{\sqrt{u_1^2 + u_2^2 + \ldots + u_{p-1}^2}}{|u_p|} \leq \tan(\theta/2) \right\},
\]

\[
B_\theta := \left\{ u \in \mathbb{R}^p : u_p = \pm \cos(\theta/2), u_1^2 + u_2^2 + \ldots + u_{p-1}^2 \leq \sin^2(\theta/2) \right\}.
\]

The boundary of \( V_\theta \) is a right circular, conical surface of two nappes, and \( B_\theta \) is the intersection of \( V_\theta \) with the union of two affine planes: with \( A^+ = \{(u_1, u_2, \ldots, u_p) \in \mathbb{R}^n : u_k = \cos(\theta/2)\} \cong \mathbb{R}^{p-1} \) and \( A^- = \{(u_1, u_2, \ldots, u_p) \in \mathbb{R}^n : u_p = -\cos(\theta/2)\} \cong \mathbb{R}^{p-1} \) define \( B_\theta^+ = V_\theta \cap A^+ \) and \( B_\theta^- = V_\theta \cap A^- \) so that \( B_\theta = B_\theta^+ \cup B_\theta^- \). Similarly, \( V_\theta^+ \) and \( V_\theta^- \) to represent the upper \((x_p > 0)\) and lower \((x_p < 0)\) portions of \( V_\theta \).

We use “\( \partial \)” to signal the boundary of a set and “\( \circ \)” to signal the interior of a set. Expressions such as \( \partial V_\theta \) and \( V_\theta^\circ \) have meanings that are readily understood. For \( B_\theta^- \), however, since it is embedded within \( A^+ \), which itself is embedded within \( \mathbb{R}^p \), we should specify that the topological operations are carried out
within $A^+$, and likewise for $B^- \subset A^-$ and $B^- \subset A$. Thus, $\partial_B \theta = A \cap \partial V_\theta$ and $B^\theta_\theta = A \cap V_\theta^\circ$, etc.

The boundary $\partial V_\theta$ separates $\mathbb{R}^p$ into the complement of $V_\theta$, namely $V_\theta^\circ$, and the interior of $V_\theta$, $V_\theta^\circ$. Furthermore, the interior $V_\theta^\circ$ is separated by the origin into what we will call $V_\theta^\circ^+$ and $V_\theta^\circ^-$. In summary, the complement of the boundary of $V_\theta$ consists of three separate regions: $\partial V_\theta = V_\theta^\circ \cup V_\theta^\circ^+ \cup V_\theta^\circ^-$.  

2.2 Projections

**Definition 2.3.** Given an affine subspace $L \subseteq \mathbb{R}^d$ consider $u \in \mathbb{R}^d$, $u \notin L$. The projection of $u$ onto $L$ is

$$p_L(u) = \arg\min_{l \in L} \|u - l\|.$$  

**Lemma 2.1.** Given an affine subspace $L \subseteq \mathbb{R}^d$ consider $u \in \mathbb{R}^d$, $u \notin L$. A point $p \in L$ is $p_L(u)$ if and only if for all $l \in L$ we have $(1 - p) \perp (u - p)$.

**Definition 2.4.** Given a set of vectors, $\{x_k\}_{k \in K}$, define the vector subspace

$$K_L = \text{span}\{\{x_k\}_{k \in K}\}.$$  

**Lemma 2.2.** Consider a vector subspace $K_L \subseteq \mathbb{R}^d$ and a vector $y \in \mathbb{R}^d$, $y \notin K_L$. If $q \in \mathbb{R}^d$ and $q \notin K_L$ then $|y - p_{q,K_L}(y)| < |y - p_{K_L}(y)|$.

**Proposition 2.1.** Set $e_d = (0, 0, ..., 1) \in \mathbb{R}^d$. Consider $y \in \mathbb{R}^d$ and $x_1 \in \mathbb{R}^d$ with $\langle y, e_d \rangle > 0$ and $\langle x_1, e_d \rangle > 0$. If a nonzero $x_2 \in \mathbb{R}^d$ satisfies $x_2 \perp e_d$ then $\langle p_{2,1,L}(y), e_d \rangle \neq 0$.

**Proof.** Define $E_d = \{u \in \mathbb{R}^d : u \perp e_d\}$ so that $y \notin E_d$, $x_1 \notin E_d$ and $x_2 \in E_d$. Observe that $\{\text{span}\{x_1, x_2\} \cap E_d\}$ is one-dimensional, and thus by Lemma 2.1, $\langle p_{2,1,L}(y), e_d \rangle = 0 \Rightarrow p_{2,1,L}(y) = p_{2,L}(y)$. This however contradicts the strict inequality in Lemma 2.2, so we reject $\langle p_{2,1,L}(y), e_d \rangle = 0$ and conclude $\langle p_{2,1,L}(y), e_d \rangle \neq 0$.  

**Theorem 2.1** (Thales’ theorem). Consider the geometric sphere:

$$S_r^m = \{(u_1, u_2, ..., u_m) \in \mathbb{R}^m : u_1^2 + u_2^2 + u_3^2 = r^2 > 0\}.$$  

For three points on $S_r^m$, namely $t$, its opposite $-t$, and another point $s$, the following holds:

$$(s - t) \perp (s - (-t)).$$

Thales’ theorem is typically proved for $m = 2$ using elementary geometry of angles and triangles; the higher-dimensional version that we have stated follows since $\text{span}\{(t, -t, v)\} \cong \mathbb{R}^2$ and $\text{span}\{(t, -t, s)\} \cap S_r^m \cong S_s^2$. Note that the theorem also holds for lower dimensional ($q < m$), embedded spheres ($S_2^q \hookrightarrow \mathbb{R}^m$), that are not necessarily centered at the origin. For example, the theorem holds when applied on $\partial B_\theta$.  

3
3 Zeros

This section builds on the development of the previous section. Fixing $y$ and $x_1$ within $\mathbb{R}^n$, we characterize the level set $\{x_2 \in \mathbb{R}^n : 2,1\beta_1 = 0\}$, as a step toward identification of those $x_2$ that are associated with reversals. The analysis proceeds cleanly if we assume centered (mean zero) vectors of data. Tildes indicate centered data.

3.1 Space of Centered Vectors

Assume $\tilde{y}$, $\tilde{x}_1$ and $\tilde{x}_2$, each within $\mathbb{R}^n$, with $0 < \theta(\tilde{y}, \tilde{x}_1) < \pi/2$. Being centered, each vector is orthogonal to $e = (1, 1, ..., 1)$, and thus $\text{span}(\{\tilde{y}, \tilde{x}_1, \tilde{x}_2\})$ is contained within an $(n-1)$-dimensional, vector subspace $E$.

Place orthonormal coordinates on $E$ so that $\tilde{y}/|\tilde{y}|+\tilde{x}_1/|\tilde{x}_1|$ can be expressed as $(0,0,...,2\cos(\theta(\tilde{y},\tilde{x}_1)/2))$. Set $p = n-1$ so as to write $E \cong \mathbb{R}^p$, suggesting and allowing application of the remaining terminology from Section 2. Note that $\tilde{y}/|\tilde{y}|$ and $\tilde{x}_1/|\tilde{x}_1|$ are opposites on $\partial B^+$, and $\tilde{x}_2$ is somewhere in $E\setminus 0$.

Because $0 < \theta(\tilde{y}, \tilde{x}_1) < \pi/2$, $2,1\tilde{\beta}_1 > 0$, but depending on the positioning of $\tilde{x}_2$, $2,1\tilde{\beta}_1$ could be positive, negative, or zero.\footnote{In general, $\kappa \tilde{\beta}_k$ denotes the $k$th coefficient, when the projection of $y$ onto the span of explanatory vectors indexed by $K$, is expressed in terms of such basis vectors.} We specify the “sign” function to be “1” for positives, “−1” for negatives, and undefined for zero, so that $\text{sign}(2,1\tilde{\beta}_1) \neq \text{sign}(1,\tilde{\beta}_1)$ signals a change from a positive to a negative or vice versa: i.e. $\text{sign}(2,1\tilde{\beta}_1) \neq \text{sign}(1,\tilde{\beta}_1) \iff \text{sign}(2,1\tilde{\beta}_1) + \text{sign}(1,\tilde{\beta}_1) = 0$.

In case $\tilde{x}_2/|\tilde{x}_2| = \pm \tilde{x}_1/|\tilde{x}_1|$ think of $\tilde{x}_2$ as dominating, and define $2,1\tilde{\beta}_2$ to be $2,1\tilde{\beta}_2$ while $2,1\tilde{\beta}_1 = 0$. For the unrealistic case $\tilde{x}_2 = 0$, think of arrival along $\text{span}(\{\tilde{y}\})$, so that by continuity it makes sense to formally define $2,1\tilde{\beta}_1 = 0$.

3.2 Geometric Results

Proposition 3.1. Let $\tilde{y}$ and $\tilde{x}_1$ be such that $0 < \theta(\tilde{y}, \tilde{x}_1) < \pi/2$. For any $\tilde{x}_2$,

$$2,1\tilde{\beta}_1 = 0 \iff \tilde{x}_2 \in \partial V_{\theta(\tilde{y}, \tilde{x}_1)}.$$  

Proof. First observe that $\tilde{x}_2 \in \text{span}(\{\tilde{x}_1\}) \cup \text{span}(\{\tilde{y}\}) \implies 2,1\tilde{\beta}_1 = 0$. Also, by Proposition 2.1, $2,1\tilde{\beta}_1 \neq 0$ for any $\tilde{x}_2 \neq 0$ such that $\langle \tilde{x}_2, (0,0,...,1) \rangle = 0$. Furthermore, for $c_1 \neq 0$, the scaling, $\tilde{x}_2 \mapsto c_1\tilde{x}_2$ leaves $2,1\tilde{\beta}_1$ unchanged. Thus, we assume $\tilde{x}_2 \in A^+ \setminus \{\tilde{y}, \tilde{x}_1\}$.

Additionally, note that for $c_2 > 0$, the scaling, $\tilde{p} \mapsto c_2\tilde{p}$ leaves $\text{sign}(2,1\tilde{\beta}_1)$ unchanged, so we may consider $2,1\tilde{\beta}_1^*$ (in place of $2,1\tilde{\beta}_1$), where $2,1\tilde{\beta}_1^*$ is the expression of $\tilde{p}^*$ in the basis $\{\tilde{x}_1, \tilde{x}_2\}$. 


Lemma 4.1. 

Corollary 3.1. 

Finally, thinking of $l$ as a function of $\tilde{x}_2 \in A^+ \setminus \{\tilde{y}, \tilde{x}_1\}$, we can conclude that $\text{sign}(\tilde{l}) = 0$ $\iff$ $\tilde{x}_2 \in \partial B^+$ by appealing to Thales’ theorem. 

**Proposition 3.2.** Let $\tilde{y}$ and $\tilde{x}_1$ be such that $0 < \theta(\tilde{y}, \tilde{x}_1) < \pi/2$. For any $\tilde{x}_2$,

$$2,1\hat{\beta}_1 < 0 \iff \tilde{x}_2 \in V_{\theta(\tilde{y}, \tilde{x}_1)}. $$

*Proof.* Given our setup, $2,1\hat{\beta}_1$ is a function, $f$, of $\tilde{x}_2$, and, by Proposition 3.1, the zero set of $f$ is $\partial V_{\theta(\tilde{y}, \tilde{x}_1)}$. Also, as mentioned in Section 2.1, $\partial V_\theta = V_\theta^- \cup V_\theta^+$ and $\partial V_\theta = \partial V_\theta^- \cup V_\theta^-$, since, $f$ is continuous we need only sample one representative point from each of $V_\theta^-, V_\theta^+$ and $V_\theta^-$, in order to determine the sign of $2,1\hat{\beta}_1$ throughout each region. The following points are conveniently selected to avoid projection, and direct computation reveals that $\tilde{x}_2 = \pm \tilde{y} + \tilde{x}_1 \implies 2,1\hat{\beta}_1 < 0$ and $\tilde{x}_2 = \tilde{x}_1 - \tilde{y} \implies 2,1\hat{\beta}_1 > 0$. 

In case $\pi/2 < \theta(\tilde{y}, \tilde{x}_1) < \pi$, replace $\tilde{x}_1$ with $\tilde{x}_1$, so that Proposition 3.2 still applies, effectively extended the range of application, resulting in the following corollary.

**Corollary 3.1.** Let $\tilde{y}$ and $\tilde{x}_1$ be such that $\theta(\tilde{y}, \tilde{x}_1) \in \{(0, \pi/2) \cup (\pi/2, \pi)\}$. For any $\tilde{x}_2$,

$$2,1\hat{\beta}_1 \neq 1\hat{\beta}_1 \iff \tilde{x}_2 \in V_{\theta(\tilde{y}, \tilde{x}_1)}. $$

4 Correlations

In this section we translate the results of Section 3 into the language of correlations, in order to prove Theorems 1.1 and 1.2. We posit the observation of $y$, $x_1$, and $x_2$, each within $\mathbb{R}^n$. In case $r(y, x_1) = \pm 1$, the truth of the theorems can be verified with the comments written at the end of Section 3.1, and $r(y, x_1) \neq 0$ is assumed in both theorems. Thus, henceforth, we tacitly assume $r(y, x_1) \in \{(-1, 0) \cup (0, 1)\}$, ensuring applicability of Corollary 3.1.

**Lemma 4.1.** Let $u_1$ and $u_2$ be two vectors in $\mathbb{R}^n$, and let their centered versions be denoted with $\tilde{u}_1$ and $\tilde{u}_2$. Then, $\left\langle \frac{\tilde{u}_1}{|\tilde{u}_1|}, \frac{\tilde{u}_2}{|\tilde{u}_2|} \right\rangle = r(u_1, u_2)$.

Given $y$ and $x_1$ projection determines their centered versions, $\tilde{y}$ and $\tilde{x}_1$. Define $\tilde{h} = \tilde{y} + \tilde{x}_1$. Since projection and addition commute, we know that $\tilde{h}$ is the centered version of $h = y + x_1$. With this notation, Corollary 3.1 asserts that $\text{sign}(2,1\hat{\beta}_1) \neq \text{sign}(1\hat{\beta}_1) \iff \left\langle \frac{\tilde{x}_2}{|\tilde{x}_2|}, \frac{\tilde{h}}{|\tilde{h}|} \right\rangle > \left\langle \frac{\tilde{y}}{|\tilde{y}|}, \frac{\tilde{h}}{|\tilde{h}|} \right\rangle$, and by Lemma 4.1 this becomes $\text{sign}(2,1\hat{\beta}_1) \neq \text{sign}(1\hat{\beta}_1) \iff r(x_2, h) > r(y, h)$, thus demonstrating the truth of Theorem 1.2.
To prove Theorem 1.1, simply observe that within $A$ the intersection $F = \{ \tilde{r}_1 \in A : |\tilde{r}_1 - \tilde{y}| < \sin(\theta(\tilde{y}, \tilde{x}_1)/2) \} \cap \{ \tilde{r}_2 \in A : |\tilde{r}_2 - \tilde{x}_1| < \sin(\theta(\tilde{y}, \tilde{x}_1)/2) \}$ strictly contains $B$. Thus, by Corollary 3.1, the existence of a reversal implies that there exists a $c \neq 0$ such that $c \tilde{x}_2 \in B^\circ \subset F$. An interpretation of $c \tilde{x}_2 \in F$ in terms of correlation coefficients produces the conclusion of Theorem 1.1.

5 Generalizations

If we denote with $JR$ the positive square root of the coefficient of determination, where $J$ indexes a finite subset of the (perhaps very large) set of all conceivable, explanatory vectors of data, then the following generalization of Theorem 1.2 can be stated.

**Theorem 5.1.** When $r(y, x_1) \neq 0$, a reversal, $\text{sign}(J, 1\hat{\beta}_1) \neq \text{sign}(1\hat{\beta}_1)$, occurs if and only if $JR > |r(y, x_1 + y)|$.

Theorem 5.1 is related to Theorem 1.2 and Proposition 3.2 through the observation that a reversal occurs, in this more general setting, if and only if $p_{J, V}(\tilde{y}) \in V_{\theta(\tilde{y}, \tilde{x}_1)}^\circ$. Alternatively, one can think of reversals as occurring precisely when $\text{span}\{\tilde{x}_1, \{\tilde{x}_j\}_{j \in J}\}$ intersects $V_{\theta(\tilde{y}, \tilde{x}_1)}^\circ$.

We can generalize further. Let $I$ index a finite, orthogonal set of explanatory vectors, and let $J$ index a disjoint, finite set of additional explanatory vectors.

**Theorem 5.2.** For any $i \in I$, when $r(y, x_i) \neq 0$, a reversal, $\text{sign}(J, I, 1\hat{\beta}_i) \neq \text{sign}(1\hat{\beta}_i)$, occurs if and only if $JR > |r(y, x_i + y)|$.

References


†The assumption of orthogonality allows for disregard of the supporting vectors, since they lie orthogonal to the space of interest. For a specific example that shows the importance of the assumption of orthogonality, and for a discussion of the practical applications of Theorem 5.2, see [7] (Table 1 and Sections 1 and 3).


[8] Lignell et al. “Prenatal exposure to polychlorinated biphenvyls and polybriminated diphenyl ethers may influence birth weight among infants in a Swedish cohort with background exposure: a cross-sectional study.” *Environmental Health* (2013), 12:44 http://www.ehjournal.net/content/12/1/44


