Journal of Probability and Statistical Science
24(2), 1-15Aug. 2026

Fixed Points for Fitted Lines

David L. Farnsworth
School of Mathematics and Statistics
Rochester Institute of Technology, USA

ABSTRACT
For many fitting methods for lines in the plane, each observation possesses an associated point in the plane, or at infinity for particular observations. These associated points are called "fixed points," because properly chosen replacement points for the originally selected observation create other fitted lines that contain the same fixed point. Simple linear regression and orthogonal least-squares regression are investigated as examples. The phenomenon has been previously noted for simple regression, where the properly selected replacement points are related to regression diagnostics and the determination of outliers. Orthogonal regression lines are fundamental to principal component analysis. Mechanical devices for simple linear regression and for orthogonal regression are presented. The devices show the dependence of the fitted lines, which contain the fixed point, on the new replacement points.

Keywords: Fixed point; Mechanical device; Linear regression; Orthogonal regression

1. Introduction

Consider the set of $n-1$ paired observations

$$S_{(n-1)}: \{(x_i, y_i) : i = 1, 2, \dots, n-1\} \tag{1}$$

with $n > 3$. Subscripts in parentheses, such as $n-1$ in equation (1), indicate sample size. A fitted line to $S_{(n-1)}$ is

$$L_{(n-1)}: y = \bar{y}_{(n-1)} + b_{(n-1)}\left(x - \bar{x}_{(n-1)}\right). \tag{2}$$

The criteria for the fit are left unspecified in this and the next section to allow more general considerations at the outset. Specific fitting criteria are introduced in Sections 3 and 4. It is assumed that all fitted lines contain the points of means for the respective sets, because the fitting methods that are addressed have that property, as they are based upon least-squares. For $S_{(n-1)}$, the point of means is

$$M_{(n-1)}: \left(\bar{x}_{(n-1)}, \bar{y}_{(n-1)}\right).$$

The set $S_{(n-1)}$ is foundational and unchanging.

An n^{th} observation

$$R: (x_n, y_n) \tag{3}$$

is introduced, yielding the new set, point of means, and fitted line

$$S_{(n)}(x_n, y_n): \{(x_i, y_i) : i = 1, 2, \dots, n\},$$
$$M_{(n)}(x_n, y_n): \left(\bar{x}_{(n)}, \bar{y}_{(n)}\right).$$

and

$$L_{(n)}(x_n, y_n): y = \bar{y}_{(n)} + b_{(n)}(x_n, y_n)\left(x - \bar{x}_{(n)}\right), \tag{4}$$

where the same criteria as is used for $L_{(n-1)}$ is used for the fit of $L_{(n)}(x_n, y_n)$. In the set $S_{(n)}(x_n, y_n)$, the point of means $M_{(n)}(x_n, y_n)$, the line $L_{(n)}(x_n, y_n)$, and the slope $b_{(n)}(x_n, y_n)$, parenthetically identifying the coordinates of the n^{th} observation helps to distinguish these

- Received May 2, 2026, in final form August 2026.
- David L. Farnsworth (corresponding author) is affiliated with School of Mathematics and Statistics, Rochester Institute of Technology, USA
  Email: dlfsma@rit.edu

quantities from the set $S_{(n)}(s,t)$, point of means $M_{(n)}(s,t)$, line $L_{(n)}(s,t)$, and slope $b_{(n)}(s,t)$, in which the artificial point $T\colon(s,t)$ is used below as the n^{th} point, instead of the observation $R\colon(x_n,y_n)$.

If a set of n points had been obtained, then the observations can be ordered, so that the data point of interest, perhaps for examining its influence on the slope of the line or its being an outlier, is designated R and is in the n^{th} position.

Two fitting procedures for the lines are explored in detail. The topic of Section 3 is simple linear regression, where the x-variable is without error and the sum of the squares of the observations' distances from the fitted line, which are measured parallel to the y-axis, is minimized. Simple regression formulas with the distances measured parallel to the horizontal axis can be obtained from that one with changes in the names of variables. The topic of Section 4 is orthogonal least-squares fitting, i.e., orthogonal regression, where the distances are measured perpendicular to the fitted line and the sum of their squares is minimized. Orthogonal regression identifies the axis of the scatter ellipse, which is the main direction in principal component analysis [6, 7, 11].

Physical realizations in the form of mechanical devices are presented in Section 5. They show how certain locations of the n^{th} observation produce fitted lines that contain a unique fixed point that is associated with the data set's n^{th} observation with respect to the underlying $n-1$ observations. Concluding remarks are in Section 6.

2. Fixed points

In Lemma 1, the coordinates of the point $P\colon(p,q)$, which is the intersection of the line $L_{(n-1)}$ in (2) and the line $L_{(n)}(x_n,y_n)$ in (4) are found. It is shown below that $P\colon(p,q)$ is the "fixed point" that is mentioned in this paper's title. Families of lines that contain point P is the topic that is explored. See Figure 1, where the fixed point P is determined.

Lemma 1. The intersection of the line $L_{(n-1)}$ in (2) and the line $L_{(n)}(x_n,y_n)$ in (4), which are assumed to be distinct and non-parallel, is the point P with coordinates

$$(p,q)=\left(\bar{x}_{(n-1)}+\frac{(y_n-\bar{y}_{(n-1)})-b_{(n)}(x_n,y_n)(x_n-\bar{x}_{(n-1)})}{n\left(b_{(n-1)}-b_{(n)}(x_n,y_n)\right)},\right.$$
$$\left.\bar{y}_{(n-1)}+b_{(n-1)}\frac{(y_n-\bar{y}_{(n-1)})-b_{(n)}(x_n,y_n)(x_n-\bar{x}_{(n-1)})}{n\left(b_{(n-1)}-b_{(n)}(x_n,y_n)\right)}\right). \quad (5)$$

Proof. In (2) and (4), substitute p and q for x and y, respectively. Using the identities

$$\bar{x}_{(n)}=\bar{x}_{(n-1)}+\frac{1}{n}\left(x_n-\bar{x}_{(n-1)}\right) \text{ and } \bar{y}_{(n)}=\bar{y}_{(n-1)}+\frac{1}{n}\left(y_n-\bar{y}_{(n-1)}\right), \quad (6)$$

obtain

$$q-\bar{y}_{(n-1)}=b_{(n-1)}\left(p-\bar{x}_{(n-1)}\right) \quad (7)$$

and

$$q-\bar{y}_{(n-1)}=\frac{1}{n}\left(y_n-\bar{y}_{(n-1)}\right)+b_{(n)}(x_n,y_n)\left(p-\bar{x}_{(n-1)}-\frac{1}{n}\left(x_n-\bar{x}_{(n-1)}\right)\right).$$

Equating these expressions for $q-\bar{y}_{(n-1)}$ gives

$$b_{(n-1)}\left(p-\bar{x}_{(n-1)}\right)=\frac{1}{n}\left(y_n-\bar{y}_{(n-1)}\right)+b_{(n)}(x_n,y_n)\left(p-\bar{x}_{(n-1)}-\frac{1}{n}\left(x_n-\bar{x}_{(n-1)}\right)\right),$$

so that

$$p = \bar{x}_{(n-1)} + \frac{(y_n - \bar{y}_{(n-1)}) - b_{(n)}(x_n,y_n)(x_n - \bar{x}_{(n-1)})}{n\left(b_{(n-1)} - b_{(n)}(x_n,y_n)\right)}. \quad (8)$$

Substitution of (8) into equation (2) for $L_{(n-1)}$ gives the second coordinate of P as

$$q = \bar{y}_{(n-1)} + b_{(n-1)}(p - \bar{x}_{(n-1)}) = \bar{y}_{(n-1)} + b_{(n-1)} \frac{(y_n - \bar{y}_{(n-1)}) - b_{(n)}(x_n,y_n)(x_n - \bar{x}_{(n-1)})}{n\left(b_{(n-1)} - b_{(n)}(x_n,y_n)\right)}. \quad (9)$$

■

The special case of parallel lines having a common point at infinity is clear from the denominators of (8) and (9), where the $b_{(n-1)} - b_{(n)}(x_n, y_n)$ would be zero. This is further addressed in Sections 3 and 4 for specific methods for fitting the lines.

Introduce the variable point

$$T\colon (s,t) \quad (10)$$

and the set

$$S_{(n)}(s,t) = S_{(n-1)} \cup \{T\}. \quad (11)$$

By inspection and by introducing the notation $b_{(n)}(s,t)$ for the slope, as in (4) but with the coordinates of the data point $R\colon (x_n, y_n)$ replaced by the coordinates of the moveable point $T\colon (s,t)$, there is Lemma 2, which contains the equation of the fitted line.

Lemma 2. The fitted line for the set (11) is

$$L_{(n)}(s,t)\colon\ y - \bar{y}_{(n-1)} - \frac{1}{n}(t - \bar{y}_{(n-1)}) = b_{(n)}(s,t)\left(x - \bar{x}_{(n-1)} - \frac{1}{n}(s - \bar{x}_{(n-1)})\right). \quad (12)$$

The point of means for the set (11) is

$$M_{(n)}(s,t)\colon \left(\bar{x}_{(n-1)} + \frac{1}{n}(s - \bar{x}_{(n-1)}),\ \bar{y}_{(n-1)} + \frac{1}{n}(t - \bar{y}_{(n-1)})\right),$$

which is on the line (12).

Lemma 3 says that the three points $M_{(n-1)}, T\colon (s,t)$, and $M_{(n)}(s,t)$ are collinear, if they are distinct. Designate that line by $L(s,t)$. This collinearity creates a linkage that is a major component in the influence that an observation has over a fitted line, as shown in Sections 3 and 4 and illustrated by the mechanical devices that are described and presented in Section 5. As point T traverses a curve, line $L(s,t)$ rotates about the stationary point $M_{(n-1)}$ as a hub, tracing the point of means $M_{(n)}(s,t)$, which is on the fitted line $L_{(n)}(s,t)$.

Lemma 3. The point of means $M_{(n-1)}$ of the set (1), the varying point $T\colon (s,t)$ in (10), and the point of means $M_{(n)}(s,t)$ of the set (11) are collinear, if they are distinct.

Proof. The slopes of the lines $M_{(n-1)}T$ and $M_{(n)}(s,t)T$, which both contain the point T, are equal to $(t - \bar{y}_{(n-1)})/(s - \bar{x}_{(n-1)})$. If $s = \bar{x}_{(n-1)}$ and $t \neq \bar{y}_{(n-1)}$, then the line is vertical; if $s \neq \bar{x}_{(n-1)}$ and $t = \bar{y}_{(n-1)}$, then the line is horizontal; and, if $s = \bar{x}_{(n-1)}$ and $t = \bar{y}_{(n-1)}$, then the three points coincide. ■

Lemma 4. For minimum sum-of-squares fitting, the fitted line is not changed by introducing an additional point that is on the fitted line for the original points.

Proof. The sum of the squared distances is not changed by the additional point, so the fit remains the same. Any change to the fitted line would cause an increase in the sum of the original $n - 1$ points' squared distances and also the addition of the squared distance to the additional point, so the total of the squared distances would increase. ■

Because of Lemma 4, it is assumed that the n^{th} observation is not on the fitted line for the remaining observations, i.e., neither the observation $R\colon (x_n, y_n)$ in (3) nor the varying point $T\colon (s,t)$ in (10) are on the line $L_{(n-1)}$ in (2). Any n^{th}point that is not on the line $L_{(n-1)}$ yields a line other than $L_{(n-1)}$, which can be seen because neither of the points of means of $S_{(n)}(x_n, y_n)$ or $S_{(n)}(s,t)$ is on line $L_{(n-1)}$, so those sets' fitted lines are not $L_{(n-1)}$.

Lemma 5. An equation for the coordinates s and t of a variable point $T\colon (s,t)$, such that the fitted line $L_{(n)}(s,t)$ in (12) for the set $S_{(n)}(s,t)$ in (11) contains the point $P\colon (p,q)$ in (5), is

$$\begin{aligned}&\big(b_{(n)}(s,t) - b_{(n-1)}\big)\Big(\big(y_n - \bar{y}_{(n-1)}\big) - b_{(n)}(x_n, y_n)\big(x_n - \bar{x}_{(n-1)}\big)\Big)\\ &\quad - \big(b_{(n)}(x_n, y_n) - b_{(n-1)}\big)\Big(\big(t - \bar{y}_{(n-1)}\big) - b_{(n)}(s,t)\big(s - \bar{x}_{(n-1)}\big)\Big) = 0. \qquad (13)\end{aligned}$$

Proof. The family of fitted lines in (12) for the set (11) that contains the point $P\colon (p,q)$ gives

$$q - \bar{y}_{(n-1)} - \frac{1}{n}\big(t - \bar{y}_{(n-1)}\big) = b_{(n)}(s,t)\left(p - \bar{x}_{(n-1)} - \frac{1}{n}\big(s - \bar{x}_{(n-1)}\big)\right), \qquad (14)$$

which is parameterized by s and t. Assume that the lines $L_{(n-1)}$ and $L_{(n)}(x_n, y_n)$ have different slopes, so that they have a point of intersection. Substituting from (8) and (9) for p and q, equation (14) is

$$\begin{aligned}&b_{(n-1)}\frac{\big(y_n - \bar{y}_{(n-1)}\big) - b_{(n)}(x_n,y_n)\big(x_n - \bar{x}_{(n-1)}\big)}{n\big(b_{(n-1)} - b_{(n)}(x_n,y_n)\big)} - \frac{1}{n}\big(t - \bar{y}_{(n-1)}\big)\\ &\quad = b_{(n)}(s,t)\left(\frac{\big(y_n - \bar{y}_{(n-1)}\big) - b_{(n)}(x_n,y_n)\big(x_n - \bar{x}_{(n-1)}\big)}{n\big(b_{(n-1)} - b_{(n)}(x_n,y_n)\big)} - \frac{1}{n}\big(s - \bar{x}_{(n-1)}\big)\right), \qquad (15)\end{aligned}$$

Obtain (13) by multiplying (15) by $n\left(b_{(n-1)} - b_{(n)}(x_n, y_n)\right)$, which is nonzero, and arranging the terms. ■

Because the center of attention is on a set of observations that is considered to be $n-1$ pairs augmented with an n^{th} pair, additional formulas that are similar to the identities (6), which gives the mean of n univariate observations in terms of the mean of $n-1$ of them and of the n^{th} observation, are very useful and are in Lemma 6. The derivations of the formulas can be accomplished with algebra by squaring, applying (6), and distributing the summation signs. When the n^{th} point is $T\colon (s,t)$, these formulas can be used by substituting (s,t) for (x_n, y_n).

Lemma 6. For the set $S_{(n)}(x_n, y_n)$ of n pairs of observations,

$$\begin{aligned}S_{XX(n)} &= \textstyle\sum_{i=1}^{n}\big(x_i - \bar{x}_{(n)}\big)^2 = \sum_{i=1}^{n-1}\big(x_i - \bar{x}_{(n-1)}\big)^2 + \frac{n-1}{n}\big(x_n - \bar{x}_{(n-1)}\big)^2\\ &= S_{XX(n-1)} + \frac{n-1}{n}\big(x_n - \bar{x}_{(n-1)}\big)^2, \qquad (16)\\ S_{YY(n)} &= \textstyle\sum_{i=1}^{n}\big(y_i - \bar{y}_{(n)}\big)^2 = \sum_{i=1}^{n-1}\big(y_i - \bar{y}_{(n-1)}\big)^2 + \frac{n-1}{n}\big(y_n - \bar{y}_{(n-1)}\big)^2\\ &= S_{YY(n-1)} + \frac{n-1}{n}\big(y_n - \bar{y}_{(n-1)}\big)^2, \qquad (17)\end{aligned}$$

and

$$\begin{aligned}S_{XY(n)} &= \textstyle\sum_{i=1}^{n}\big(x_i - \bar{x}_{(n)}\big)\big(y_i - \bar{y}_{(n)}\big)\\ &= \textstyle\sum_{i=1}^{n-1}\big(x_i - \bar{x}_{(n-1)}\big)\big(y_i - \bar{y}_{(n-1)}\big) + \frac{n-1}{n}\big(x_n - \bar{x}_{(n-1)}\big)\big(y_n - \bar{y}_{(n-1)}\big)\\ &= S_{XY(n-1)} + \frac{n-1}{n}\big(x_n - \bar{x}_{(n-1)}\big)\big(y_n - \bar{y}_{(n-1)}\big). \qquad (18)\end{aligned}$$

For each fitting procedure of interest, such as linear regression in Section 3 and orthogonal least-squares fitting of lines in Section 4, the objective is to solve equation (13) in Lemma 5 for the relationship between the coordinates s and t of the variable n^{th} point $T\colon (s,t)$, such that the

fitted lines for the set $S_{(n)}(s,t)$ contain the point $P\colon(p,q)$, which is determined by the set $S_{(n)}(x_n,y_n)$, where $R\colon(x_n,y_n)$ is the selected point. It is required that the n^{th} data point R is one realization of the variable point T. Figure 2 illustrates this in a schematic fashion, where the curve $f(x,y)=0$ contains the points such as R and T.

Definition 1. When point $P\colon(p,q)$ plays the role of the common point in a family of lines, as described in the paragraph above, the point P is called a *fixed point*.

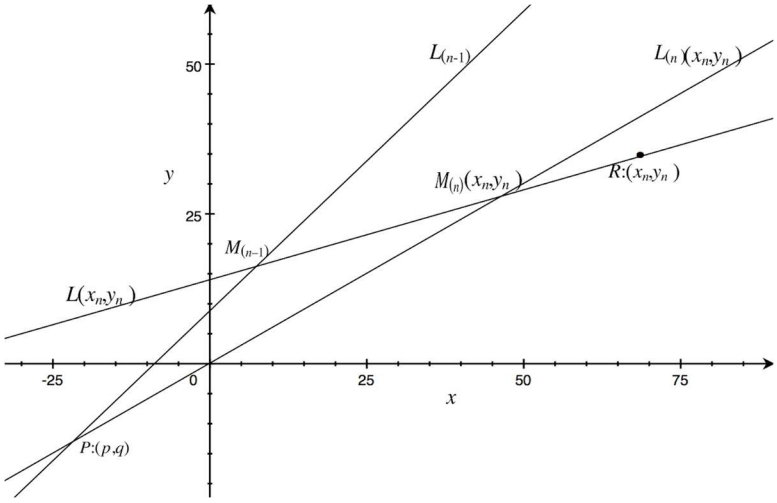


Figure 1: The point $P\colon(p,q)$ is determined by $S_{(n)}(x_n,y_n)$

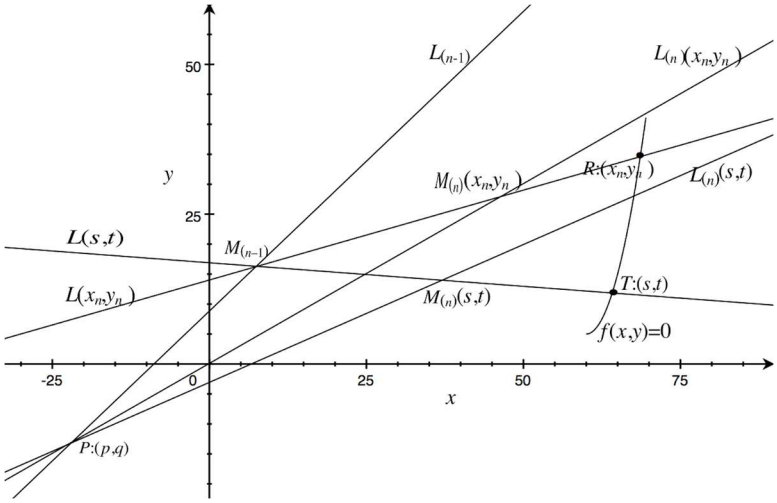


Figure 2: The curve $f(x,y)=0$ contains the n^{th} points, such as R and T, that create fitted lines containing the fixed point P

In order to have sufficient structure, it is assumed that R is not on the line $L_{(n-1)}$, as discussed following Lemma 4. The assumption that $S_{XY(n-1)} \neq 0$ in Sections 3 and 4 is reasonable, because the linear correlation coefficient for $S_{(n-1)}$ would be zero if $S_{XY(n-1)}$ were zero.

3. Linear regression

For simple linear regression, it is shown in [3] and [9] that a fixed point corresponds to each observation in the following sense. Designating the selected observation in $S_{(n)}(x_n, y_n)$ by $R\colon (x_n, y_n)$, it is shown there that any single replacement point on the vertical line $x = x_n$, along with the remainder of the original data set, produces a regression line that contains a fixed point that is associated with R. Frames from an animation showing numerous lines with one fixed point for a well-known data set is in [9]. Because one such replacement point is the intersection of $x = x_n$ and $L_{(n-1)}$, Lemma 4 says that the produced regression line is $L_{(n-1)}$. Thus, the fixed point is contained in both the line $L_{(n-1)}$ and the line $L_{(n)}(x_n, y_n)$ and its coordinates are given in Lemma 1.

Many diagnostic tools in linear regression for examining the influence and leverage of one of the data points, represented by $R\colon (x_n, y_n)$ here, involve "moving" that data point on the vertical line $x = x_n$ [5, 8, pp. 384, 420, 10, p. 321]. Fixed points might be considered to be intrinsic to those studies.

The special case in which the intersection point $P\colon (p, q)$ of $L_{(n-1)}$ and $L_{(n)}(x_n, y_n)$ in Lemma 1 is at infinity occurs when the first coordinate of $R\colon (x_n, y_n)$ is $\bar{x}_{(n-1)}$. Theorem 1 addresses that occurrence. Theorem 2 identifies all n^{th} points $T\colon (s, t)$ such that the regression line for the set $S_{(n)}(s, t)$ contains point P, including the case in which P is the appropriate point at infinity.

Theorem 1. In simple linear regression with $S_{XX(n-1)}$, $S_{XY(n-1)}$, and $b_{(n-1)}$ non zero, if the x-coordinate of the selected observation $R\colon (x_n, y_n)$ is the mean of the other observations, i.e., $x_n = \bar{x}_{(n-1)}$, and if $y_n \neq \bar{y}_{(n-1)}$, then the lines $L_{(n-1)}$ and $L_{(n)}(x_n, y_n)$ are parallel, and, thus, their point of intersection is a point at infinity. These are the only n^{th} points that yield parallel lines.

Proof. Consider the expression

$$b_{(n)}(x_n, y_n) = \frac{S_{XY(n)}(x_n, y_n)}{S_{XX(n)}(x_n, y_n)} = \frac{S_{XY(n-1)} + \frac{n-1}{n}(x_n - \bar{x}_{(n-1)})(y_n - \bar{y}_{(n-1)})}{S_{XX(n-1)} + \frac{n-1}{n}(x_n - \bar{x}_{(n-1)})^2} \tag{19}$$

for the slope of the regression line $L_{(n)}(x_n, y_n)$, where Lemma 6 has been employed. Setting this slope equal to

$$b_{(n-1)} = \frac{S_{XY(n-1)}}{S_{XX(n-1)}}, \tag{20}$$

which is the slope of regression line for the set $S_{(n-1)}$, yields

$$x_n = \bar{x}_{(n-1)}$$

or

$$b_{(n-1)} = \frac{S_{XY(n-1)}}{S_{XX(n-1)}} = \frac{y_n - \bar{y}_{(n-1)}}{x_n - \bar{x}_{(n-1)}}.$$

The second outcome says that $R\colon (x_n, y_n)$ is on the line $L_{(n-1)}$, which is disallowed. For the first outcome, since $y_n \neq \bar{y}_{(n-1)}$, the two lines are seen not to coincide because of their differing points of means with identical x-coordinates. They are parallel. ■

Theorem 2. For simple linear regression with $S_{XX(n-1)}$, $S_{XY(n-1)}$, and $b_{(n-1)}$ non zero, replacing the n^{th} data point $R\colon (x_n, y_n)$, which is in the set $S_{(n)}(x_n, y_n)$ and is not on the line $L_{(n-1)}$, with any point $T\colon (s, t)$ on the vertical line $x = x_n$ or on the regression line $L_{(n-1)}$ yields a regression line, which contains the fixed point P that is associated with point R with respect to the set $S_{(n)}(x_n, y_n)$. Points T on those two lines are the only replacement points for $R\colon (x_n, y_n)$ that yield a regression line containing that fixed point P. See Figure 3,

Proof. If $x_n = \bar{x}_{(n-1)}$, then P is a point at infinity by Theorem 1, and the conclusion to the theorem is true. In order to avoid zero denominators and so that in the remainder of this proof point P is a point in the plane, assume that $x_n \neq \bar{x}_{(n-1)}$ and $s \neq \bar{x}_{(n-1)}$.

Lemma 4 implies that any point on $L_{(n-1)}$ satisfies this theorem, because, from Lemma 1, $L_{(n-1)}$ contains the fixed point.

The coordinates of $P\colon (p, q)$ in (5) do not depend upon the y-coordinate of R [3, 9]. Indeed, from (19) and (20),

$$\begin{aligned}
&\frac{\left(y_n - \bar{y}_{(n-1)}\right) - b_{(n)}(x_n,y_n)\left(x_n - \bar{x}_{(n-1)}\right)}{n\left(b_{(n-1)} - b_{(n)}(x_n,y_n)\right)} \\
&\quad = \frac{\left(y_n - \bar{y}_{(n-1)}\right) - \frac{S_{XY(n-1)} + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)\left(y_n - \bar{y}_{(n-1)}\right)}{S_{XX(n-1)} + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)^2}\left(x_n - \bar{x}_{(n-1)}\right)}{n\left(\frac{S_{XY(n-1)}}{S_{XX(n-1)}} - \frac{S_{XY(n-1)} + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)\left(y_n - \bar{y}_{(n-1)}\right)}{S_{XX(n-1)} + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)^2}\right)} \\
&\quad = \frac{S_{XX(n-1)}\left(S_{XX(n-1)}\left(y_n - \bar{y}_{(n-1)}\right) - S_{XY(n-1)}\left(x_n - \bar{x}_{(n-1)}\right)\right)}{n\left(S_{XY(n-1)}\frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)^2 - S_{XX(n-1)}\frac{n-1}{n}\left(y_n - \bar{y}_{(n-1)}\right)\left(x_n - \bar{x}_{(n-1)}\right)\right)} \\
&\quad = -\frac{S_{XX(n-1)}}{(n-1)\left(x_n - \bar{x}_{(n-1)}\right)},
\end{aligned}$$

which is free of y_n. Therefore, any point on the vertical line $x = x_n$ yields the same fixed point and satisfies the theorem. For regression, the curve $f(x, y) = 0$ in Figure 2 is $f(x, y) = x - x_n = 0$.

To show that those are the only points that yield the desired line, take advantage of the symmetry in (13) in Lemma 5, where the second term is the same as the first term after interchanging the roles of (s, t) and (x_n, y_n). Also, the number of variables can be reduced by one by employing (20) to replace $S_{XY(n-1)}$ with $S_{XX(n-1)}b_{(n-1)}$. The identity

$$b_{(n)}(s, t) = \frac{S_{XY(n)}(s,t)}{S_{XX(n)}(s,t)} = \frac{S_{XY(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)\left(t - \bar{y}_{(n-1)}\right)}{S_{XX(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)^2}, \tag{21}$$

implies that

$$\begin{aligned}
&b_{(n)}(s, t) - b_{(n-1)} \\
&\quad = \frac{S_{XY(n)}(s,t)}{S_{XX(n)}(s,t)} - b_{(n-1)} = \frac{S_{XY(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)\left(t - \bar{y}_{(n-1)}\right)}{S_{XX(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)^2} - b_{(n-1)} \\
&\quad = \frac{S_{XX(n-1)}b_{(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)\left(t - \bar{y}_{(n-1)}\right)}{S_{XX(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)^2} - b_{(n-1)} \\
&\quad = \frac{S_{XX(n-1)}b_{(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)\left(t - \bar{y}_{(n-1)}\right) - S_{XX(n-1)}b_{(n-1)} - \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)^2 b_{(n-1)}}{S_{XX(n-1)} + \frac{n-1}{n}\left(s - \bar{x}_{(n-1)}\right)^2}
\end{aligned}$$

$$= \frac{\frac{n-1}{n}(s-\bar{x}_{(n-1)})\left((t-\bar{y}_{(n-1)})-b_{(n-1)}(s-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)}+\frac{n-1}{n}(s-\bar{x}_{(n-1)})^2}. \quad (22)$$

Applying the interchange of roles to (22) gives

$$b_{(n)}(x_n, y_n) - b_{(n-1)} = \frac{\frac{n-1}{n}(x_n-\bar{x}_{(n-1)})\left((y_n-\bar{y}_{(n-1)})-b_{(n-1)}(x_n-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)}+\frac{n-1}{n}(x_n-\bar{x}_{(n-1)})^2}. \quad (23)$$

Also,

$$\begin{aligned}
&\left(t - \bar{y}_{(n-1)}\right) - b_{(n)}(s,t)\left(s - \bar{x}_{(n-1)}\right) \\
&\quad = \left(t - \bar{y}_{(n-1)}\right) - \frac{S_{XX(n-1)}b_{(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})(t-\bar{y}_{(n-1)})}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2}\left(s - \bar{x}_{(n-1)}\right) \\
&\quad = \frac{S_{XX(n-1)}\,(t-\bar{y}_{(n-1)}) + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2(t-\bar{y}_{(n-1)})}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2} \\
&- \frac{S_{XX(n-1)}b_{(n-1)}(s-\bar{x}_{(n-1)}) + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2(t-\bar{y}_{(n-1)})}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2} \\
&\quad = \frac{S_{XX(n-1)}\,(t-\bar{y}_{(n-1)}) - S_{XX(n-1)}b_{(n-1)}(s-\bar{x}_{(n-1)})}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2} \\
&\quad = \frac{S_{XX(n-1)}\left((t-\bar{y}_{(n-1)})-b_{(n-1)}(s-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2}. \quad (24)
\end{aligned}$$

Applying the interchange of roles to (24) gives

$$\left(y_n - \bar{y}_{(n-1)}\right) - b_{(n)}(x_n, y_n)\left(x_n - \bar{x}_{(n-1)}\right) = \frac{S_{XX(n-1)}\left((y_n-\bar{y}_{(n-1)})-b_{(n-1)}(x_n-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)} + \frac{n-1}{n}(x_n-\bar{x}_{(n-1)})^2}. \quad (25)$$

Substituting (22)–(25) into (13) gives

$$\begin{aligned}
&\frac{\frac{n-1}{n}(s-\bar{x}_{(n-1)})\left((t-\bar{y}_{(n-1)})-b_{(n-1)}(s-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)}+\frac{n-1}{n}(s-\bar{x}_{(n-1)})^2}\,\frac{S_{XX(n-1)}\left((y_n-\bar{y}_{(n-1)})-b_{(n-1)}(x_n-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)} + \frac{n-1}{n}(x_n-\bar{x}_{(n-1)})^2} \\
&- \frac{\frac{n-1}{n}(x_n-\bar{x}_{(n-1)})\left((y_n-\bar{y}_{(n-1)})-b_{(n-1)}(x_n-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)}+\frac{n-1}{n}(x_n-\bar{x}_{(n-1)})^2}\,\frac{S_{XX(n-1)}\left((t-\bar{y}_{(n-1)})-b_{(n-1)}(s-\bar{x}_{(n-1)})\right)}{S_{XX(n-1)} + \frac{n-1}{n}(s-\bar{x}_{(n-1)})^2} = 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left(\frac{\frac{n-1}{n}S_{XX(n-1)}}{\left(S_{XX(n-1)}+\frac{n-1}{n}(s-\bar{x}_{(n-1)})^2\right)\left(S_{XX(n-1)}+\frac{n-1}{n}(x_n-\bar{x}_{(n-1)})^2\right)}\right)\left(\left(s-\bar{x}_{(n-1)}\right)-\left(x_n-\bar{x}_{(n-1)}\right)\right) \\
&\left(\left(t-\bar{y}_{(n-1)}\right)-b_{(n-1)}\left(s-\bar{x}_{(n-1)}\right)\right)\left(\left(y_n-\bar{y}_{(n-1)}\right)-b_{(n-1)}\left(x_n-\bar{x}_{(n-1)}\right)\right) = 0. \quad (26)
\end{aligned}$$

The first factor in (26) is nonzero. The fourth factor is not zero, because R: (x_n, y_n) is not on the line $L_{(n-1)}$. The second and third factors show again that replacement points on the vertical line $s = x_n$, i.e., $x = x_n$, and on the regression line $L_{(n-1)}$ give regression lines containing point P: (p, q). There are no other factors, so those are the only points. ■

4. Orthogonal regression

Analogously to equations (19)–(21) for simple linear regression, for orthogonal regression, the slopes are

$$b_{(n)}(x_n, y_n) = \frac{S_{YY(n)}(x_n,y_n) - S_{XX(n)}(x_n,y_n) + \sqrt{\left(S_{YY(n)}(x_n,y_n) - S_{XX(n)}(x_n,y_n)\right)^2 + 4\left(S_{XY(n)}(x_n,y_n)\right)^2}}{2S_{XY(n)}(x_n,y_n)}, \quad (27)$$

$$b_{(n-1)} = \frac{S_{YY(n-1)} - S_{XX(n-1)} + \sqrt{\left(S_{YY(n-1)} - S_{XX(n-1)}\right)^2 + 4\left(S_{XY(n-1)}\right)^2}}{2S_{XY(n-1)}}, \quad (28)$$

and

$$b_{(n)}(s,t) = \frac{S_{YY(n)}(s,t) - S_{XX(n)}(s,t) + \sqrt{\left(S_{YY(n)}(s,t) - S_{XX(n)}(s,t)\right)^2 + 4\left(S_{XY(n)}(s,t)\right)^2}}{2S_{XY(n)}(s,t)} \quad (29)$$

[12].

Notice that $\bar{x}_{(n-1)}$ or $\bar{y}_{(n-1)}$ is subtracted on each occasion of the appearance of a data value. (This is true for linear regression in Section 3, as well.) Thus, without loss of generality, the centering

$$\bar{x}_{(n-1)} = \bar{y}_{(n-1)} = 0$$

of $S_{(n-1)}$ is assumed in the proofs in this section. Translations of the coordinates can be employed to accomplish this, and, at any juncture, the corresponding inverse transformations can be used to reintroduce non-zero means.

In order to visually simplify the equations, designate

$$S_{YY(n-1)} - S_{XX(n-1)} = \frac{n-1}{n}A,\ S_{XY(n-1)} = \frac{n-1}{n}B, \text{ and } b_{(n-1)} = b.$$

Those quantities are functions only of elements of the foundational set $S_{(n-1)}$. Using Lemma 6 and this notation, equations (27)–(29) can be expressed

$$b_{(n)}(x_n, y_n) = \frac{A + y_n^2 - x_n^2 + \sqrt{\left(A + y_n^2 - x_n^2\right)^2 + 4(B + x_n y_n)^2}}{2(B + x_n y_n)}, \quad (30)$$

$$b = \frac{A + \sqrt{A^2 + 4B^2}}{2B}, \quad (31)$$

and

$$b_{(n)}(s,t) = \frac{A + t^2 - s^2 + \sqrt{(A + t^2 - s^2)^2 + 4(B + st)^2}}{2(B + st)}. \quad (32)$$

Data sets that produce zero denominators in these equations create isolated points in the loci of interest. Those values would correspond to linear correlation of value zero in the observations. Take $b \neq 0$. Equation (31) shows that the sign of b is the same as the sign of B, i.e., $bB > 0$, and that

$$A = \frac{b^2 - 1}{b}B, \quad (33)$$

since $b \neq 0$.

Theorem 3 contains the result about fixed points at infinity for orthogonal linear regression that corresponds to Theorem 1 for simple linear regression.

Theorem 3. For orthogonal linear regression, assume that $S_{XY(n-1)}$ and $b_{(n-1)}$ are nonzero, i.e., $B \neq 0$ and $b \neq 0$. If the selected observation $R\colon (x_n, y_n)$ is on the line that is perpendicular to $L_{(n-1)}$ and containing the point of means $M_{(n-1)} = \left(\bar{x}_{(n-1)}, \bar{y}_{(n-1)}\right)$ of the other $n-1$ observations, but R is not the point $M_{(n-1)}$, then the orthogonal regression lines $L_{(n-1)}$ and

$L_{(n)}(x_n, y_n)$ are parallel, and, thus, their point of intersection is a point at infinity. These are the only n^{th} points that yield parallel lines.

Proof. Consider the variable point $T\colon (s,t)$. Equation (32) can be written

$$(B+st)\left(b_{(n)}(s,t)\right)^2 - (A+t^2-s^2)b_{(n)}(s,t) - (B+st) = 0.$$

Equation (32) is the immediate result of this equation in the minimization procedure. Taking $b_{(n)}(s,t) = b$, this is

$$(B+st)(b)^2 - (A+t^2-s^2)b - (B+st) = 0,$$

or

$$bt^2 - (b^2-1)st - bs^2 - (Bb^2 - Ab - B) = 0.$$

The constant term is zero from equation (33), and the remaining terms factor to
$(t-bs)(bt+s) = 0$.

The first factor yields the line $L_{(n-1)}\colon y = bx$. The second factor gives

$$t = -\frac{1}{b}s. \tag{34}$$

Equation (34) represents the line perpendicular to the line $L_{(n-1)}$ at $M_{(n-1)}$, i.e.,

$$y = -\frac{1}{b}x. \tag{35}$$

The lines $L_{(n-1)}$ and $L_{(n)}(x_n, y_n)$ are distinct because their points of means differ. ■

Theorem 4 supplies the locations of replacement points that produce fitted lines that contain the fixed point.

Theorem 4. For orthogonal regression, assume that $S_{XY(n-1)}$ and $b_{(n-1)}$ are nonzero, i.e., $B \neq 0$ and $b \neq 0$, and that $\bar{x}_{(n-1)} = \bar{y}_{(n-1)} = 0$. Replacing the n^{th} data point $R\colon (x_n, y_n)$, which is in the set $S_{(n)}(x_n, y_n)$ and is not on the line $L_{(n-1)}$, with any point $T\colon (s,t)$ on the line $L_{(n-1)}$ or on the circle

$$C\colon \left(x - \frac{n^2p^2+A-bB}{2n}\right)^2 + \left(y - \frac{bn^2p^2+bA-b^2B}{2np}\right)^2 = \frac{1}{4n^2p^2}\,(8bn^2p^2B - 4n^2p^2A + (b^2+1)(n^2p^2+A-bB)^2), \tag{36}$$

omitting the intersection points of C and $y = -(1/b)x$, yields an orthogonal regression line, which contains the fixed point P associated with point R with respect to the set $S_{(n)}(x_n, y_n)$. Also, points on circle C and line $L_{(n-1)}$ are the only such replacement points for point R that yield an orthogonal regression line containing that fixed point P. See Figure 4,

Proof. As in the proof of Theorem 2, Lemma 4 implies that any point on $L_{(n-1)}$ satisfies this theorem, because, from Lemma 1, $L_{(n-1)}$ contains the fixed point.

If the coordinates of the points $R\colon (x_n, y_n)$ and $T\colon (s,t)$ satisfy equation (35), then P is a point at infinity by Theorem 3 and the conclusion to this theorem is true with the circle C in (36) replaced by the line (35). In order to avoid that case and so that P is a point in the plane, assume in the remainder of this proof that the coordinates of the points $R\colon (x_n, y_n)$ and $T\colon (s,t)$ do not satisfy equation (35).

Instead of using equation (13), which as in the proof of Theorem 2, a more direct approach is taken here to find the coordinates of T. The expressions that are derived for the coordinates of the point T contain point P's x-coordinate p, which can be computed with equation (5) using the elements of the known set $S_{(n)}(x_n, y_n)$ and equations (27) and (28). From equation (7),

$$q = bp,$$

and then, from equation (14),

$$b_{(n)}(s,t) = \frac{t-b}{s-np}, \tag{37}$$

which is the equation to be solved for the coordinates $x = s$ and $y = t$ of T. The point $(s,t) = (np, bnp)$ is excluded from the analysis and is an isolated point of C. From equation (32), equation (37) becomes

$$\frac{A+t^2-s^2+\sqrt{(A+t^2-s^2)^2+4(B+st)^2}}{2(B+st)} = \frac{t-bnp}{s-np},$$

or

$$\sqrt{(A+t^2-s^2)^2+4(B+st)^2} = 2\frac{(B+st)(t-bnp)}{s-np} - (A+t^2-s^2).$$

Squaring and arranging terms yield

$$(B+st)((t-bnp)^2-(s-np)^2) - (s-np)(A+t^2-s^2)(t-bnp) = 0, \tag{38}$$

which is a third-degree polynomial in t. Because any point on the line $L_{(n-1)}$ satisfies this equation, one factor of the left-hand side of (38) must be $t - bs$. By expressing t as $t - bs + bs$, equation (38) becomes

$$\Big(B + s\big((t-bs)+bs\big)\Big)\Big(\big((t-bs)+bs-bnp\big)^2 - (s-np)^2\Big)$$
$$-(s-np)\Big(A + \big((t-bs)+bs\big)^2 - s^2\Big)\big((t-bs)+bs-bnp\big) = 0,$$

or

$$\big(s(t-bs)+(bs^2+B)\big)((t-bs)^2+2b(s-np)(t-bs)+(b^2-1)(s-np)^2)$$
$$-(s-np)((t-bs)^2+2bs(t-bs)+(b^2-1)s^2+A)\big((t-bs)+b(s-np)\big) = 0. \tag{39}$$

Equation (39) becomes

$$(t-bs)\big((np)(t-bs)^2+(2bnps-bn^2p^2+B)(t-bs)$$
$$+(s-np)(b^2nps+nps+2bB-A)\big) = 0. \tag{40}$$

The first factor in equation (40) confirms that the points on the line $L_{(n-1)}$: $y = bx$ satisfy the hypothesis. The conic section in the second factor of (40) is the circle

$$nps^2+npt^2-(n^2p^2-bB+A)s-(bn^2p^2-b^2B+bA)t+npA-2bnpB = 0,$$

or

$$C: npx^2+npy^2-(n^2p^2-bB+A)x-(bn^2p^2-b^2B+bA)y+npA-2bnpB = 0,$$

which yields (36). For orthogonal regression, the curve $f(x,y) = 0$ in Figure 2 is the circle C in equation (36). ■

Equation (36) shows that the center of C is on line $L_{(n-1)}$. It may be somewhat surprising that the curve is a circle. The curve's symmetry might be somewhat expected, because the model uses distances perpendicular to a line. Also, the variables are treated equally.

Sometimes data are standardized, i.e., replaced with z-scores [1, 11]. If the x-coordinates and the y-coordinates of the foundational set $S_{(n-1)}$ are separately standardized, then $A = 0$, and, from (31), $b = \pm 1$. The circle in (36) becomes

$$\Big(x - \frac{n^2p^2-b}{2np}\Big)^2 + \Big(y - \frac{bn^2p^2-B}{2np}\Big)^2 = \frac{1}{2}\Big(\frac{bn^2p^2+B}{np}\Big)^2. \tag{41}$$

The points of intersection of the circle (41) and the line $L_{(n-1)}$ are (np, bnp) and $(-bB/(np), -B/(np))$.

5. Mechanical devices

Figures 3 and 4 show physical realizations of the dependence on the varying replacement n^{th} observation for fitting the regression lines and orthogonal regression lines in Theorems 2 and 4, respectively. As the location of the brown replacement point $T\colon(s,t)$ is changed on its gray track, the red linkage rod $L(s,t)$ rotates, like a spoke of a wheel, about the stationary green hub at $M_{(n-1)}$. The blue point of means $M_{(n)}(s,t)$ slides on the red rod $L(s,t)$, and on its gray track. The point $M_{(n)}(s,t)$ reacts to the movement of the point $T\colon(s,t)$, because of the linkage $L(s,t)$, which is the red rod. The tan rod is the fitted line $L_{(n)}(s,t)$, which contains the orange fixed point $P\colon(p,q)$, which is also a hub, and the blue point of means $M_{(n)}(s,t)$. In both figures, the original n^{th}-observation $R\colon(x_n,y_n)$ is at one of the points on the gray track for $T\colon(s,t)$. There are two hubs, which are located at $P\colon(p,q)$ and $M_{(n-1)}$. The line $L_{(n-1)}$, which fits the elements of the set $S_{(n-1)}$, is not shown in either figure. It would be a stationary rod connecting the orange $P\colon(p,q)$ and the green $M_{(n-1)}$ hubs. It is a component of the underlying structure for the mechanisms but not a moving part.

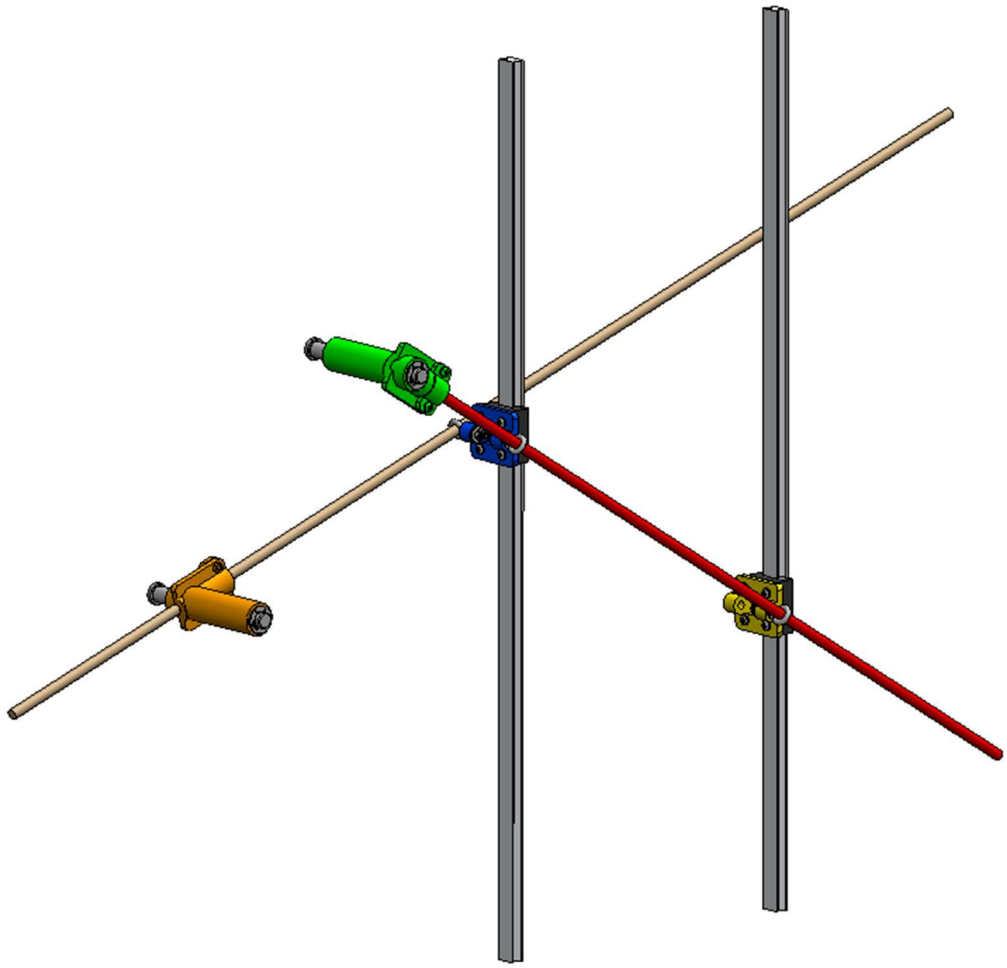


FIGURE 3: Mechanism for Theorem 2

The legend for Figure 3 is:

The brown slider is the potential replacement point T: (s, t).
The right-hand gray vertical track is the location of replacement points, such as T: (s, t).
The blue slider is the point of means $M_{(n)}(s, t)$.
The left-hand gray vertical track is the location of the points of means, such as $M_{(n)}(s, t)$.
The green fixture is a pivot at a center of rotation and is $M_{(n-1)}$.
The red rod $L(s, t)$ is the linear linkage that is assured by Lemma 3.
The orange fixture is a pivot at a center of rotation and is the fixed point P: (p, q).
The tan rod is the regression line $L_{(n)}(s, t)$ for the set $S_{(n)}(s, t)$.

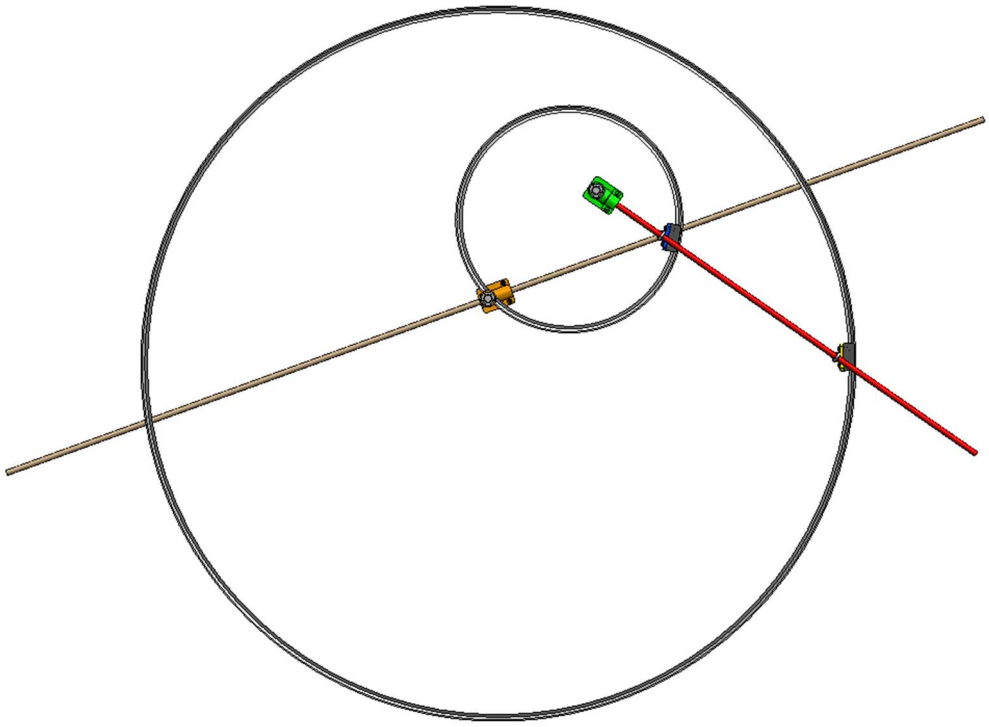


FIGURE 4: Mechanism for Theorem 4

The legend for Figure 4 is:

The brown slider is the potential replacement point T: (s, t).
The outer gray circular track is C, which is the location of replacement points, such as T: (s, t).
The blue slider is the point of means $M_{(n)}(s, t)$.
The inner gray circular track is the location of the points of means, such as $M_{(n)}(s, t)$.
The green fixture is a pivot at a center of rotation and is $M_{(n-1)}$.
The red rod $L(s, t)$ is the linear linkage that is assured by Lemma 3.
The orange fixture is a pivot at a center of rotation and is the fixed point P: (p, q).
The tan rod is the orthogonal regression line $L_{(n)}(s, t)$ for the set $S_{(n)}(s, t)$.

6. Concluding comments

The goal here is to show that there are some interesting points and curves associated with certain models for bivariate observations. Simple linear regression and orthogonal regression are used as examples. Another choice for a linear model is Deming regression, where a different constant angle replaces the 90° degree angle that appears in orthogonal regression [2, pp. 128–147, 12]. Since the line contains the point of means, the lemmas and ideas in Sections 1 and 2 can be applied to it. One approach to Deming regression is with errors in measurement in the data for each coordinate. For orthogonal regression, it is assumed, or believed, that all measurements have equal errors. That might occur when the same gauge or dial is read for both variables, and, thus, the gauge's pointer or dial's markings produce the same uncertainty for both variables. For Deming regression, the errors in the two variables may have different variances, whose constant ratio appears in the model [2, 12].

A well-known problem with orthogonal regression is that a coordinate transformation or scaling alters the identity of the fit, which cannot be undone with the inverse transformation of the fit. There is still debate about the suitability of transformations in some circumstances [1, 4, pp. 93–94, 6, pp. 64–65, 7]. Centering does not affect perpendicularity, for example, so that centering with $\bar{x}_{(n-1)} = \bar{y}_{(n-1)} = 0$ in Section 4 is not problematical. However, transforming the observations to z-scores is not advised by some authors and practitioners, because that irreversibly alters the fit of the observations.

Acknowledgement

Figures 3 and 4 were created by Hilary L. Farnsworth, Managing Partner, Instill Tech, using a 3D modeling tool. The author wishes to express his appreciation and thankfulness for her contributions to this paper, especially the figures.

Conflict of interest and funding

The author declares that there is no conflict of interest for this paper. No funds were received by the author regarding this paper.

References

[1]. Abdi, H., and Williams, L. J. (2010). Principal Component Analysis. *WIREs Computational Statistics*, 2(4, July/August), 433–459.

[2]. Deming, W. E. (1964). *Statistical Adjustment of Data*. New York, NY, USA: Dover Publications.

[3]. Farnsworth, D. L. (2025). Fixed Points in Linear Regression. *Open Journal of Statistics*, 15(3, June), 251–271.

[4]. Gwet, K. L. (2020). *Beginner's Guide to Principal Components*. Gaithersburg, MD, USA: AgreeStat Analytics.

[5]. Hoaglin, D. C. (1988). Using Leverage and Influence to Introduce Regression Diagnostics. *The College Mathematics Journal*, 19(5, November), 387–401.

[6]. Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York, NY, USA: Springer-Verlag.

[7]. Jolliffe, I. T., and Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(April)(No. 2065), 1–16.

[8]. Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). New York, NY, USA: McGraw-Hill/Irwin.

[9]. Lutzer, C. V., and Farnsworth, D. L. (2021). Pivot Points in Bivariate Linear Regression. *Open Journal of Statistics*, 11(3, June), 393–399.

[10]. Montgomery, D. C., Runger, G. C. and Hubele, N. F. (2011). *Engineering Statistics* (5th ed.). New York, NY, USA: Wiley.

[11]. Tofallis, C. (2023). Fitting an Equation to Data Impartially. *Mathematics*, 11(article 3957).

[12]. Xu, S. (2014). A Property of Geometric Mean Regression. *The American Statistician*, 68(4, November), 277–281.