# The Distribution of the Ratio of Normal Random Variables and the Ellipticity of the Earth

Gregory Kordas[1] and  George Petrakos[2]

**Abstract**

In this paper we consider inference regarding the ratio of two normal and the ratio of two t-distributed random variables, using both the popular Fieller interval, as well as, exact distributions. We apply these methods to a historical dataset regarding the shape of the Earth, and estimate the Earth's flatness coefficient as a ratio of regression coefficients. We demonstrate the equivalence of the inference using the exact density of this ratio with that using the Fieller interval.

## 1   Introduction

This paper grew out of our attempt to use a historical dataset to motivate least squares estimation and inference in our graduate econometrics course. The dataset in question concerns the shape of the Earth and consists of five observations of the length of $1°$ of latitude first analyzed by Roger Boscovich in the 1750's. The dataset and its story are discussed in length in Stephen Stigler's classic book *The History of Statistics* (Stigler, 1986: 39-50).

There are several reasons that make the analysis of this particular dataset interesting. First, from a physical perspective it provided an early support to Newton's theory of gravitation. Second, from a statistical standpoint, it illustrates the early history of various attempts to combine noisy observations in order to obtain less noisy estimates. Third, the parameters in question are known today, so the student can judge for him- or her-self the effectiveness of the "scientific method": careful collection of observations coupled with sound statistical analysis can reveal the "truth".

The most challenging part of the analysis is to compute the distribution and obtain confidence intervals for the Earth's flatness coefficient $f$ that is defined as the ratio of the two coefficients (slope/intercept) of a simple regression model. To obtain confidence intervals for $f$ we use Fieller's theorem, and to compute the entire distribution of $f$ we use the the ratio of normals distribution derived by Marsaglia (1965), as well as, the ratio of $t$-distributed random variables derived by Press (1969). We show that the two methods, i.e. Fieller intervals and exact distribution computation, yield identical results, that is, we show that the probability under the ratio density function between the Fieller bounds is indeed equal to the specified confidence level. Given that the two literatures, that of Fieller

[1] Department of Public Administration, Panteion University, Athens, Greece; gkordas@panteion.gr
[2] Department of Public Administration, Panteion University, Athens, Greece; petrakos@panteion.gr

intervals and that of exact distributions of normal ratios, have grown almost independently of each other, it is, to the best of our knowledge, the first time that the two approaches are used simultaneously and are shown to yield identical results.

Section 2 of this paper presents briefly the historical background of the question of the shape of the Earth, and discusses the dataset used by Boscovich. Section 3, presents the OLS estimates along with Fieller intervals for the Earth's flatness coefficient under alternative assumptions. Section 4, takes up the question of computing the exact distribution of $f$ and establishes the equivalence of the inference based on exact distributions with that based on Fieller intervals. The source R code used in the empirical application is provided in an Appendix.

## 2   The Ellipticity of the Earth

One of the implications of Isaac Newton's *Theory of Universal Gravitation*, introduced in his epoch-making book *Philosophia Naturalis Principia Mathematica*, often referred to as simply the *Principia*, was that, due to gravity, the rotation of the Earth around its axis would cause the Earth to bulge at the equator and flatten at the poles. More precisely, Newton proved that a rotating self-gravitating fluid body in equilibrium takes the form of an oblate ellipsoid of revolution (a spheroid). The exact amount of flattening depends on the body's density and the balance between the resulting gravitational and centrifugal forces. If gravity is therefore operational, it is unlikely that the Earth is a perfect sphere, but it should be an *oblate* spheroid, much like an orange.

Newton's theory was not the only theory around purporting to explain the motion of the celestial spheres. The French mathematician and physicist René Descartes had proposed the competing *Theory of Vortices*. According to Descartes' theory, space is filled with an invisible substance called the *ether*, that, much like water, creates vortices that sweep the planets into their apparent orbits. Now, plastic spherical objects inside a water vortex tend to flatten at the equator and bulge at the poles, so if the theory of Vortices was correct, the Earth should be a *prolate* spheroid, i.e., more like an egg or a lemon instead of an orange.

The two competing theories led to a prolonged controversy, much along nationalistic lines, between the English and their Continental rivals. Voltaire, who happened to be visiting London when Newton died in 1727, was greatly impressed by the State funeral and the honors bestowed on the great scientist, and, on the subject of the controversy, commented that:

> A Frenchman arriving in London finds things very different. [...] For us it is the pressure of the Moon that causes the tides of the sea; for the English it is the sea that gravitates towards the Moon. [...] In Paris you see the Earth shaped like a melon, in London it is flattened out on two sides.

The equation of a 3D ellipsoid centered at the origin with semi-axes $a$, $b$ and $c$ aligned along the coordinate axes is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$
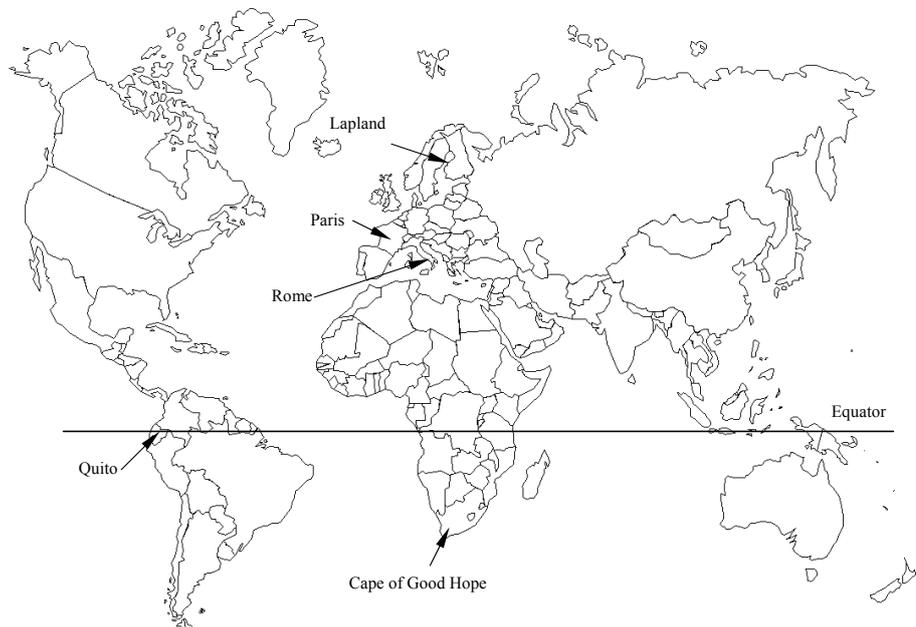
**Figure 1:** Location of available measurements on the World Map

Because planets revolve around their north-south axis, physical considerations dictate that planets are *solids of revolution*, i.e., solids that can be obtained by revolving a 2D curve around the $z$ (North-South) axis to obtain a 3D body. This means that for planets $a = b$, so the equation becomes

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1.$$

If $c < a$ the ellipsoid is called *oblate* (orange-like), while if $c > a$ the ellipsoid is *prolate* (lemon-like). Of course, if $c = a$ we get a perfect sphere. Our interest is in the quantity of polar *flattening* or *ellipticity* $f$ given by

$$f = \frac{a - c}{a} = 1 - \frac{c}{a}.$$

Letting $a = R_E$ be the Earth's equatorial radius and $c = R_P$ be its polar radius, we can write

$$f = 1 - \frac{R_P}{R_E}.$$

The Earth is oblate if $f > 0$, prolate if $f < 0$, and spherical if $f = 0$.

To settle the matter once and for all, in 1735 the *Académie des Sciences Française* send expeditions to Ecuador, Lapland, and South Africa to measure meridians at widely separated latitudes. Along with the pre-existing measurements from Paris and Rome, the *Académie* managed to collect the five data points given in Table 1 (Stigler, 1986).

The length of $1°$ of latitude $\ell$ at the various locations was measured in toise, a popular measure of the time. To get a better idea, the table also presents these lengths in kilometers. A simple inspection of the table makes it apparent that the length of $1°$ of latitude

**Table 1:** Data on the length of $1°$ of latitude at various locations

| Location | Latitude $\theta$ | $\sin^2(\theta)$ | Length of $1°$ $\ell$ [*toises*][a] | Length of $1°$ $\ell$ [*km*] |
|---|---|---|---|---|
| Quito, Ecuador | $0°0'$ | 0.0000 | 56,751 | 110.551 |
| COGH,[b] S. Africa | $33°18'$ | 0.2987 | 57,037 | 111.108 |
| Rome, Italy | $42°59'$ | 0.4648 | 56,979 | 110.995 |
| Paris, France | $49°23'$ | 0.5762 | 57,074 | 111.180 |
| Lapland, Finland | $66°19'$ | 0.8386 | 57,422 | 111.858 |

[a] 1 toise = 1948 meters
[b] Cape of Good Hope

grows as we move from the equator ($\theta = 0°$) to the poles ($\theta = 90°$). At Quito, which is on the equator, the length of a degree was measured to be 110.551 km, while at Lapland, which is the closest people of the time could get to the north pole on account of the cold, the length of the degree was measured to be more than a kilometer longer. Clearly these data favor Newton's prediction that the Earth flattens at the poles. There are, however, discrepancies too: at the Cape of Good Hope the length of a degree is longer than that at Rome, despite the fact that Rome has a larger (north) latitude than the (south) latitude of the Cape of Good Hope. Graphing these measurements we see that, with the exception of the (Cape of Good Hope - Rome) pair, there seems to be a consistent tendency for the length to grow as we move away from the equator.

For short arcs, the approximation (see Stigler, 1986)

$$\ell = \beta_0 + \beta_1(3\sin^2\theta) + \text{higher-order terms},$$

where $\ell$ is the length of $1°$ of latitude and $\theta$ is the angle of the latitude, was known to be satisfactory. The parameters $\beta_0$ and $\beta_1$ can be interpreted as the length of $1°$ of latitude at the equator and the excess in length of $1°$ at the poles over its value at the equator, respectively. Ellipticity is therefore given by $f = \beta_1/\beta_0$.

These parameters are, of course, "known" today. Table 2 (from Abramowitz and Stegun, 1972, p. 8.) presents the geodetic constants for the *International Hayford Spheroid* (IHS). We see that the Earth flattens at the poles with an average oblate ellipticity of $f = 1/297$. The length of $1°$ of latitude at the equator is $60 \times 1,842.925 = 110,576$ m, while that at the poles is $60 \times 1,861.666 = 111,700$ m. This means that the circumference of the Earth around the equator is approximately[2] $C_E = 39,807$ km, while that around the poles is approximately $C_P = 39,807 \times (1 - 1/297) = 39,673$ km, a mere 134 km less than that around the equator.

In Book III of the *Principia*, Newton himself predicted that:

> [...] the diameter of the Earth at the equator is to its diameter from pole to pole as 230 to 229. – *Principia*, Book III, Proposition XIX, Problem III.

---

[2]That the circumference of the earth in km's is almost a round number (40,000 km) is *not* a coincidence. The *meter* was originally defined as the one ten-millionth (1/10,000,000) of the distance between a pole and the equator along a great circle over water. Since to go around the earth one has to travel 4 times this distance, the circumference of the earth is 40 million meters.
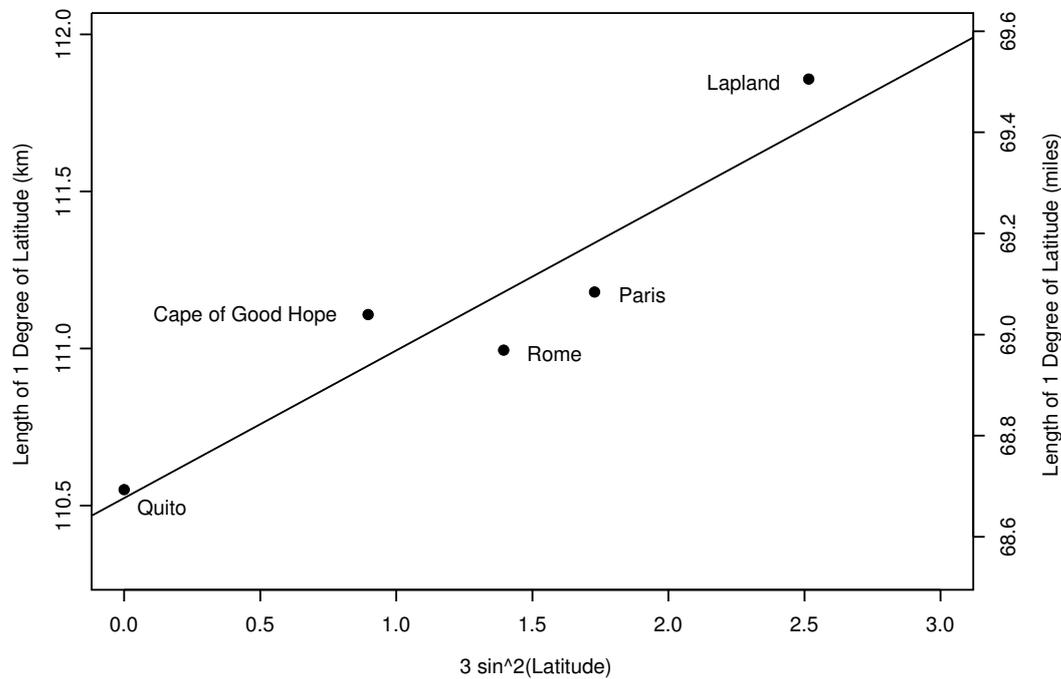
**Figure 2:** Graph of $\ell$ against $3\sin^2\theta$, along with the least squares line

That is, Newton gave $f = 1/230$. Fitzpatrick (2009: Section 2.12), presents a theoretical model of rotational flattening that, under simplifying homogeneity assumptions about the rotating body, predicts $f = 1/233$. He says that this is (essentially) the model that Newton used to make his prediction, and comments that "the discrepancy [with the actual $f = 1/297$ value] is due to the fact that the Earth is strongly inhomogeneous, being much denser at its core than in its outer regions".

In terms of our model, $\beta_0 = 110.576$ km, while from $f$ we obtain a polar exceedance of $\beta_1 = 110.576/297 = 0.3723$ km per $1°$ of latitude. In what follows, we will estimate $\beta_0$, $\beta_1$, and $f$ from the data in Table 1 and see how close the scientists of the time came to discovering the "truth".

# 3   Estimating the Ellipticity of the Earth

Using the data in Table 1 we estimate by least squares the regression model

$$\ell = \beta_0 + \beta_1(3\sin^2\theta) + u.$$

According to the Earth parameters in Table 2, the "true" equation is

$$\ell = 110.576 + 0.3723\ (3\sin^2\theta)$$

**Table 2:** Geodetic Constants – Hayford International Spheroid[a]

$$a = 6,378,388\,m;\ c = 6,356,912\,m;\ f = 1/297$$

| Latitude [°] | Length of 1′ of longitude [m] | Length of 1′ of latitude [m] | Acceleration of gravity [m/s²] |
|---|---|---|---|
| 0 | 1855.398 | 1842.925 | 9.780 350 |
| 15 | 1792.580 | 1844.170 | 9.783 800 |
| 30 | 1608.174 | 1847.580 | 9.793 238 |
| 45 | 1314.175 | 1852.256 | 9.806 154 |
| 60 | 930.047 | 1856.951 | 9.819 099 |
| 75 | 481.725 | 1860.401 | 9.828 593 |
| 90 | 0.000 | 1861.666 | 9.832 072 |

[a] From Abramowitz and Stegun (1972), p. 8

so that,

$$f = 0.3723/110.576 = 1/297\ ;$$
$$C_E = 110.576 \times 360 = 39,807\,\text{km}.$$

The OLS estimates (along with their s.e.'s in parentheses below them) are

$$\hat{\ell} = \underset{(0.158)}{110.525} + \underset{(0.1014)}{0.4697}\ (3\sin^2\theta),\quad \begin{aligned} R^2 &= 0.8773 \\ \hat{\sigma}_u &= 0.1903 \end{aligned}$$

so that,

$$\hat{f} = 0.4697/110.525 = 1/235.3\ ;$$
$$\hat{C}_E = 110.525 \times 360 = 39,789\,\text{km}.$$

The variance-covariance matrix for $\hat{\beta}$ is given by

$$V(\hat{\beta}) = \hat{\sigma}_u^2 (X'X)^{-1} = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{pmatrix} = \begin{pmatrix} 0.02481 & -0.01344 \\ -0.01344 & 0.01028 \end{pmatrix}.$$

Given the smallness of the sample size, the only way to perform inference is to assume that the neoclassical (normal and spherical errors) model applies. Assuming normal and spherical errors, we see that $\hat{\beta}_0$ is very statistically significant with a t-statistic of 702, while $\hat{\beta}_1$ is less significant, but still significant at the 5% level, with a t-statistic of 4.63 (*p*-value of 0.019 on 3 d.f.). Compared to the now known quantities, we see that $\beta_0$ and $\beta_1$, and therefore $C_E$ and $f$ also, were estimated quite accurately by these data, and that the data clearly support Newton's prediction that the Earth bulges at the equator and flattens at the poles. Figure 3 presents individual 95% CI's for $\beta_0$ and $\beta_1$, as well as, a joint 95% confidence region for the two parameters. The joint confidence region is given by the set of values for which the quadratic form

$$\frac{1}{2} \begin{pmatrix} 110.525 - \beta_0 \\ 0.4697 - \beta_1 \end{pmatrix}' \begin{pmatrix} 0.02481 & -0.01344 \\ -0.01344 & 0.01028 \end{pmatrix}^{-1} \begin{pmatrix} 110.525 - \beta_0 \\ 0.4697 - \beta_1 \end{pmatrix}$$
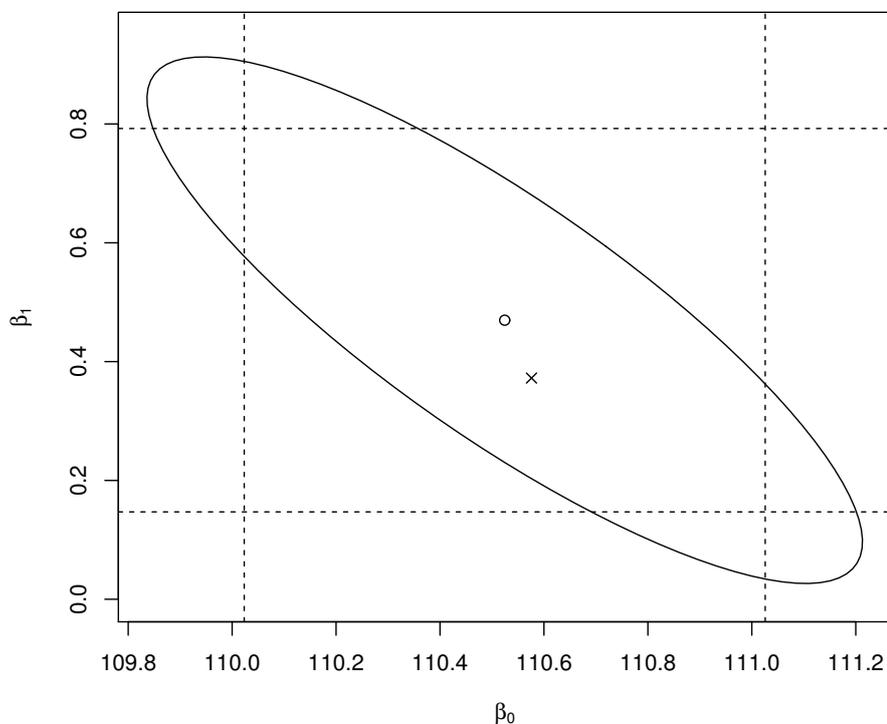
**Figure 3:** The 95% joint confidence region for $\beta_0$ and $\beta_1$ along with the marginal 95% CI's. The marginal 95% CI for $\beta_0$ is $[110.023, 111.026]$, while that for $\beta_1$ is $[0.1470, 0.7924]$. The now known true parameter values of $\beta_0 = 110.576$ and $\beta_1 = 0.3723$ are also shown as point $\times$.

is less than or equal to the critical $F_{1-\alpha,2,n-k} = F_{.95,2,3} = 9.552$ value.

In order to obtain a first approximate confidence interval for $\psi = 1/f$ we resort to a trick: we divide both sides by $\hat{\beta}_0$ and estimate the OLS regression

$$\frac{\ell}{\hat{\beta}_0} = 1 + \frac{1}{\psi}(3\sin^2\theta) + \frac{u}{\hat{\beta}_0}.$$

This regression has the same $R^2$ as the original regression and the $t$-statistic for $1/\psi$ is the same as the $t$-statistic for $\beta_1$. We get $1/\hat{\psi} = .0042498$ with an s.e. of $.0009174$, so, using the $t_{.975,3} = 3.182$ critical value, we obtain the 95% CI for $1/\psi$ as $[.001330, .007169]$. Upon inverting we get $\hat{\psi} = 235.3$ as above, and the 95% CI for $\psi$ is given by

$$CI_{naive}(\psi; .95) = [139.48, 751.77].$$

The CI is quite wide, but it importantly contains only positive values for $\psi$ that correspond to a prolate Earth.

This CI for $\psi$ treats $\beta_0$ as known and equal to the estimated value without uncertainty, i.e., it does not take into account the variability in $\hat{\beta}_0$. Since, in our application, $\beta_0$ is estimated very accurately (i.e., it's s.e. is very small relative to its magnitude) this omission

should not matter a lot. In any case, the correct $95\%$ CI should be wider[3].

To obtain the correct CI we use Fieller's (1944) theorem. The following Aside presents the method as it is adapted to the General Linear Model by Zerbe (1978).

**Aside** (Fieller's Theorem) (Zerbe, 1978) Let

$$\psi = K\beta/L\beta,$$

where $K$ and $L$ are $1 \times k$ vectors of known constants, be the ratio of two linear combinations of a $k \times 1$ parameter vector $\beta$. If an estimator $\hat{\beta}$ is distributed as $\hat{\beta} \sim N(\beta, \Sigma)$, we have that, for a $\sqrt{n}$-estimator $\hat{\Sigma}$ of $\Sigma$,

$$T = \frac{K\hat{\beta} - \psi L\hat{\beta}}{\left[K\hat{\Sigma}K' - 2\psi K\hat{\Sigma}L' + \psi^2 L\hat{\Sigma}L'\right]^{1/2}} \sim t_{n-k}.$$

Letting $t = t_{1-\alpha/2, n-k}$ be the critical value from the $t_{n-k}$ distribution, we have

$$1 - \alpha = \Pr\{-t \le T \le t\} = \Pr\{T^2 - t^2 \le 0\} = \Pr\{a\psi^2 + b\psi + c \le 0\},$$

where,

$$a = (L\hat{\beta})^2 - t^2 L\hat{\Sigma}L',$$
$$b = 2\left[t^2 K\hat{\Sigma}L' - (K\hat{\beta})(L\hat{\beta})\right],$$
$$c = (K\hat{\beta})^2 - t^2 K\hat{\Sigma}K'.$$

The last expression says that the interval containing the required $1 - \alpha$ probability is characterized by the values for which the binomial $a\psi^2 + b\psi + c$ is *negative*. If $a$ is positive, the function is convex and takes negative values. If, furthermore, the discriminant $b^2 - 4ac$ is also positive, the binomial has two distinct real roots that define the required CI. Thus, the $100(1 - \alpha)\%$ CI for $\psi$ is given by

$$\left[\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right],$$

provided that $a > 0$ and $b^2 - 4ac > 0$. For the pathological $a < 0$ and/or $b^2 - 4ac < 0$ cases, as well as, for a nice geometrical interpretation of Fieller's theorem, see Luxburg and Franz (2009).

In our application, $\psi = \beta_0/\beta_1$, and $\hat{\beta}$ is the OLS coefficient that, under the normal and spherical errors assumption, is distributed as $N(\beta, \sigma_u^2 (X'X)^{-1})$. Letting $K = (1, 0)$ and $L = (0, 1)$, we can write $\psi = K\beta/L\beta$ as required. Then, using $t = t_{.975,3} = 3.182$, we compute

$$a = 0.1165 > 0, \quad b = -104.1, \quad c = 12,215.4$$

---

[3]Note that we can obtain a correct CI for $\psi$ by evaluating the Sum of Square Residuals on a grid of values for $\beta_0$ to obtain the profile likelihood for $\psi$. Then, upon inverting the LR-test statistic we can obtain an asymptotically correct CI. To avoid confusing the discussion, we do not pursue this alternative here.
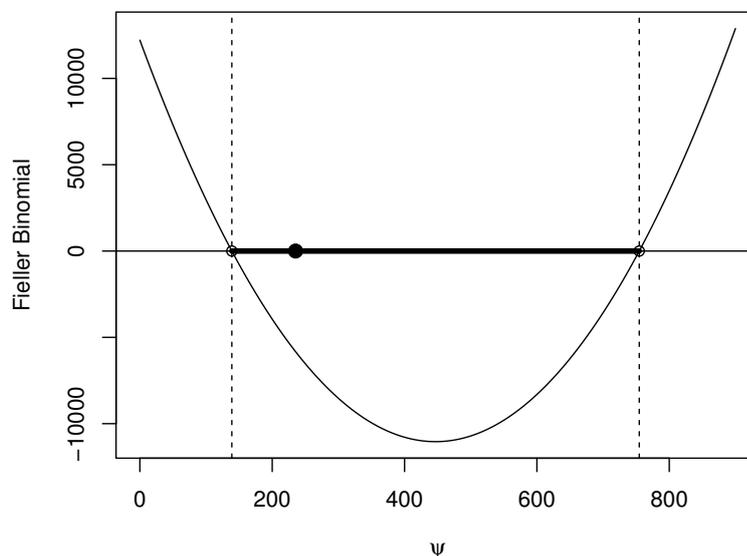
**Figure 4:** The Fieller binomial $a\psi^2 + b\psi + c$ (thin line) that, when negative, defines the Fieller 95% CI for $\psi$ (bold line). The point estimate $\hat{\psi} = 235.3$ is also shown.

and

$$b^2 - 4ac = 5,144.7 > 0,$$

and the Fieller (correct) 95% CI for $\psi$ is given by

$$CI_{Fieller,t}(\psi, .95) = [138.95, 754.66]. \tag{3.1}$$

As expected, since $\beta_0$ is estimated quite accurately here, this correct CI for $\psi$ (that takes into account the variability in both $\hat{\beta}_0$ and $\hat{\beta}_1$) is only marginally wider than the naive CI we computed above. Note that both CI's are asymmetric around the point estimate $\hat{\psi} = 235.3$ (see Figure 4), with the longer tail towards large $\psi$'s that correspond to a less oblate and more spherical Earth.

For comparison purposes, we can also treat the means, variances and correlation as known, so that the ratio $f = \beta_0/\beta_1$ is a ratio of normals standardized by their *true* variances and correlation (see Section 4). This simply amounts to using $t = z_{\alpha/2} = 1.96$, so that

$$a = 0.1811 > 0, \quad b = -103.9, \quad c = 12,215.6$$

and

$$b^2 - 4ac = 1,951.4 > 0,$$

which yields the much shorter interval

$$CI_{Fieller,normal}(\psi, .95) = [164.96, 408.84]. \tag{3.2}$$

Yet a fourth, albeit only asymptotically valid, method of obtaining a CI for $\psi$ is the $\delta$-method, which produces $s.e.(\hat{\psi}) = 51.08$, so the 95% CI is given by

$$CI_{\delta-method}(\psi, .95) = 235.3 \pm t_{.975,3}51.08 = [72.75, 397.86].$$

This CI is symmetric around the point estimate, but it disagrees with the correct Fieller-t CI enormously because it disregards the extreme smallness of the sample size.

Compared to the now known quantities, the estimates obtained from the data in Table 1 were indeed quite accurate. When published in the mid 1700's, these data lent considerable support to Newton's *Theory of Gravitation*. One could indeed compare this verification of Newton's predictions to the experimental verifications of Einstein's *General Theory of Relativity* in the 20th century, that was found to correctly account for many anomalies that were left unexplained by Newtonian mechanics, the most famous of which are the advance of Mercury's perihelion and the bending of light rays passing through the gravitational field of the Sun.

# 4   Ratios of Normal and t Random Variables

In the previous section we obtained several CI's for $\psi$ under alternative assumptions. In this section we wish to compute the entire distribution of $\hat{\psi}$ under the same assumptions and verify that the two approaches yield identical results. We start with the case in which the parameters (means, variances and correlation) of the joint normal distribution are known, and then discuss how the analysis should be modified if these parameters are treated only as estimates from a small sample, as it is the case here.

The distribution of the ratio of normal random variables has been derived by several authors, including Marsaglia (1965), Hinkley (1969), and Cedilnik, Košmelj, and Blejec (2004). In our developments below we use the form given by Marsaglia (1965) because his result was extended by Press (1969) to ratios of $t$-distributed random variables that we will also need.

Marsaglia (1965) considered the distribution of the ratio

$$R = \frac{Z + a}{W + b}, \quad a \geq 0, b \geq 0, \tag{4.1}$$

where $a$, $b$ are nonnegative real constants and $Z$ and $W$ are independent standard normal random variables. The following lemma gives the distribution of $R$.

**Lemma 1** (Marsaglia, 1965) *If $Z$ and $W$ are independent normal random variables, the probability density function of the ratio $R$ in (4.1) is given by*

$$f_R(t) = \frac{e^{-(a^2 + b^2)/2}}{\pi(1 + t^2)} \left\{ 1 + \frac{q[2\Phi(q) - 1]}{2\phi(q)} \right\}, \quad -\infty < t < \infty, \tag{4.2}$$

*where,*

$$q = \frac{at + b}{\sqrt{1 + t^2}},$$

*and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively.*

Press (1969) considered again the ratio in (4.1), only this time he assumed that $Z$ and $W$ are independent $t_\nu$ random variables.

**Lemma 2** (Press, 1969) *If $Z$ and $W$ are independent $t_\nu$ random variables, the probability density function of the ratio $R$ in (4.1) is given by*

$$g_{R,\nu}(t) = \frac{k_1}{1+t^2} \left\{ 1 + \frac{k_2 q}{q^{*\nu+1}} \left[ 2F_{\nu+1}\left( \frac{q\sqrt{\nu+1}}{q^*} \right) - 1 \right] \right\}, \qquad -\infty < t < \infty,$$

*where $F_n$ is the c.d.f of the Student t-distribution with $n$ degrees of freedom, $q$ is as in Lemma 1, $q^* = \sqrt{a^2 + b^2 + \nu - q^2}$ and*

$$k_1 = \frac{1}{\pi} \left( 1 + \frac{a^2+b^2}{\nu} \right)^{-\nu/2}, \quad k_2 = \frac{\sqrt{\pi}\nu^{\nu+2}\,\Gamma\left(\dfrac{\nu+1}{2}\right)}{2\Gamma\left(\dfrac{\nu+2}{2}\right)\left(1 + \dfrac{a^2+b^2}{\nu}\right)^{-\nu/2}}.$$

Press (1969) shows that as $\nu \to \infty$, $g_{R,\nu}(t)$ converges to $f_R(t)$. Both results in Lemmas 1 and 2 provide the distribution for uncorrelated standardized variables, with only their means $a$ and $b$ being nonzero. To use these distributions in applications we need a way of transforming arbitrary jointly normal or jointly $t$-distributed random variables to the form considered in (4.1). Marsaglia (1965) states that there are constants $c_1$ and $c_2$ for which the ratio of arbitrary jointly normal variables $X$ and $Y$, can be written as $X/Y = c_1 + c_2(Z+a)/(W+b)$, i.e., that the distribution of $X/Y$ is a translation of that of $R$. After we had derived the transformation ourselves we found the paper by Marsaglia (2006) that gives exactly the same formulas. In the hope that our derivation of the formulas is more transparent than that of Marsaglia, we include here a brief account of our derivation of the constants $c_1, c_2, a$ and $b$.

It is a well-known fact that if $U_1$ and $U_2$ are independent $N(0,1)$ random variables and $\rho \in R$ such that $|\rho| < 1$, then

$$Z_1 = U_1 \quad \text{and} \quad Z_2 = \rho U_1 + \sqrt{1-\rho^2}U_2,$$

are $BVN(0,0,1,1,\rho)$. Also, for $\mu_1, \mu_2 \in R$ and $\sigma_1 > 0, \sigma_2 > 0$, $X_1 = \mu_1 + \sigma_1 Z_1$ and $X_2 = \mu_2 + \sigma_2 Z_2$ are $BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Now let $Z$ and $W$ be independent $N(0,1)$ random variables and consider the rotations

$$X = \sigma_x \rho(W+b) + \sigma_x\sqrt{1-\rho^2}(Z+a) \quad \text{and} \quad Y = \sigma_y(W+b),$$

where $a$ and $b$ are nonnegative constants and $|\rho| < 1$. Then, $X$ and $Y$ are jointly normal with variances $\sigma_x^2$ and $\sigma_y^2$ and correlation $\rho$. Their ratio is

$$\frac{X}{Y} = \frac{\sigma_x \rho}{\sigma_y} + \frac{\sigma_x\sqrt{1-\rho^2}}{\sigma_y}\left( \frac{Z+a}{W+b} \right),$$

as required, with

$$c_1 = \frac{\sigma_x}{\sigma_y}\rho \quad \text{and} \quad c_2 = \frac{\sigma_x}{\sigma_y}\sqrt{1-\rho^2}. \tag{4.3}$$

The means of $X$ and $Y$ are $\mu_x = b\sigma_x\rho + a\sigma_x\sqrt{1-\rho^2}$ and $\mu_y = \sigma_y b$, respectively, from which we obtain,

$$a = \frac{\mu_x/\sigma_x - \rho\mu_y/\sigma_y}{\sqrt{1-\rho^2}} \quad \text{and} \quad b = \frac{\mu_y}{\sigma_y}. \tag{4.4}$$

Since the distributions of $Z$ and $-Z$ and $W$ and $-W$ are the same, if $a$ and $b$ have the same sign, $c_1$ and $c_2$ can be taken as given in (4.3). If, however, only one of the constants $a$ or $b$ is positive, we take $c_2 = -\sigma_x\sqrt{1 - \rho^2}/\sigma_y$.

We have the following lemma.

**Lemma 3** *Let $X$ and $Y$ be jointly normal random variables with means $\mu_x$, $\mu_y$, variances $\sigma_x$, $\sigma_y$ and correlation $\rho$, and let $R' = X/Y$ be their ratio.*

1. *The distribution of $R'$ is unaffected if $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ are all rescaled by a common positive factor.*

2. *If $f_R(t)$ is the density of $R$ in (4.1) where $a$ and $b$ are as given in (4.4) then the density of $R'$ is given by*

$$f_{R'}(t) = \frac{1}{c_2} f_R\left(\frac{t - c_1}{c_2}\right), \qquad -\infty < t < \infty,$$

*where $c_1$ and $c_2$ are given in (4.3).*

Part 1. of the lemma says that, as far as ratios are concerned, of the five parameters in a bivariate normal distribution, only four are *free*: the ratios $R_1 = X_1/Y_1$ from $(X_1, Y_1) \sim BVN(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ and $R_2 = X_2/Y_2$ from $(X_2, Y_2) \sim BVN(\lambda\mu_x, \lambda\mu_y, \lambda\sigma_x, \lambda\sigma_y, \rho)$, where $\lambda > 0$, have the same distribution. This answers Hinkley's (1969) objection that in the general problem there are five parameters but Marsaglia (1965) uses only four, namely $(a, b, c_1, c_2)$.

Part 2. of the lemma gives the relation between the two sets of parameters and yields some interesting insights into the problem. The location and scale constants $c_1$ and $c_2$ depend only on the parameters of the covariance matrix of $X$ and $Y$, and although necessary to pin down the exact distribution of $R$ for each set of parameters, they are not very interesting in terms of the various shapes that this distribution assumes. As Marsaglia (1965) asserted, the shape is indeed determined by the constants $a$ and $b$, which we will call "the numerator and denominator standardized means", respectively: $b$ is indeed the standardized mean $\mu_y/\sigma_y$ of the denominator variable $Y$, while $a$ is the standardized mean of the numerator variable $X$, albeit adjusted for correlation if $\rho \neq 0$. For example, since, for $\mu_y \neq 0$, $a = 0$ if and only if $(\mu_x/\sigma_x)/(\mu_y/\sigma_y) = \rho$, and since $c_1$ and $c_2$ do not depend on $\mu_x$ and $\mu_y$, the ratio $R_1$ from $(X_1, Y_1) \sim BVN[\mu_x = m, \mu_y \neq 0, \sigma_x, \sigma_y, \rho = 0]$ has the same distribution as the ratio $R_2$ from $(X_2, Y_2) \sim BVN[\mu_x = m/\sigma_x - \rho\mu_y\sigma_x/\sigma_y, \mu_y \neq 0, \sigma_x, \sigma_y, \rho]$. So, although the statement "$b$ is small, large, or zero" is equivalent to "$\mu_y/\sigma_y$ is small, large, or zero, respectively", the statement "$a$ is small, large or zero" should be interpreted as: "there is an *equivalent model* (shifted by $c_1$ and scaled by $c_2$) with $\rho = 0$ for which $a = \mu_x/\sigma_x$ is small, large or zero, respectively".

Recall that $U$ follows a Cauchy $C(\xi, \nu)$ distribution if

$$g_U(u) = \left\{\pi\nu\left[1 + \left(\frac{u - \xi}{\nu}\right)^2\right]\right\}^{-1} \quad \text{and} \quad G_U(u) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{u - \xi}{\nu}\right), \quad (4.5)$$

for $-\infty < u < \infty$. The following corollary characterizes the Cauchy densities that are included in $f_R(t)$.

**Corollary 1** *Taken together, Lemmas 1 and 3 yield that:*

1. *The ratio $R$ of two centered jointly normal random variables $(X, Y) \sim BVN(\mu_x = 0, \mu_y = 0, \sigma_x, \sigma_y, \rho)$ is distributed as $C(\xi = c_1, \nu = c_2)$.*

2. *The distribution of the ratio $R$ of two jointly normal random variables for which either $\mu_x \neq 0$ and/or $\mu_y \neq 0$, does not belong to the Cauchy $C(\nu, \xi)$ family.*

The familiar result that the ratio of two independent standard normal variables is distributed as standard Cauchy $C(0, 1)$ is a special case of (i). Note also that by (i), $C(0, 1)$ is also the distribution of the ratio of any independent centered jointly normal random variables with equal variances. Furthermore, the general variances and correlation with zero means, completely exhaust the Cauchy location-scale family: the negative result in part (ii) means that the densities graphed by Marsaglia and other authors for $a \neq 0$ and/or $b \neq 0$, although Cauchy-like in some cases, do not belong to the $C(\xi, \nu)$ location-scale family.

In the form given in (4.2), the density of $R$ is numerically unstable for large $a$ and $b$. To see this, note that when $a$ and/or $b$ is large $e^{-(a^2+b^2)/2}$ becomes very small, while $1/\phi(q)$ in the expression inside the brackets becomes very large, leading to $0 \times$ infinity computations in terms of floating-point computer accuracy. Also, as given, the normal approximation in eq. (4.2) is not apparent. To make the density numerically stable and reveal the nature of the normal approximation let

$$h = \frac{bt - a}{\sqrt{1 + t^2}},$$

and use

$$2\pi\phi(h)\phi(q) = e^{-(a^2+b^2)/2},$$

to rewrite the density in (4.2) as

$$f_R(t) = \frac{q}{1 + t^2}\phi(h)\{2\Phi(q) - 1\} + \frac{e^{-(a^2+b^2)/2}}{\pi(1 + t^2)} \tag{4.6}$$

$$= \phi(h)\left[\frac{q\{2\Phi(q) - 1\} + 2\phi(q)}{1 + t^2}\right]. \tag{4.7}$$

Either of the above representations are completely general and have the added advantage that they can be easily evaluated without numerical problems for any $a$ and $b$ no matter how large or small they happen to be.

To see what happens as $b$ gets large, or as $a$ *and* $b$ get small, let

$$f_R^*(t) = \frac{q}{1 + t^2}\phi(h) = \frac{dh}{dt}\phi(h) = \frac{d}{dt}\Phi(h),$$

and write

$$f_R(t) = f_R^*(t) \times \delta(t),$$

where

$$\delta(t) = \{2\Phi(q) - 1\} + \frac{2\phi(q)}{q}.$$
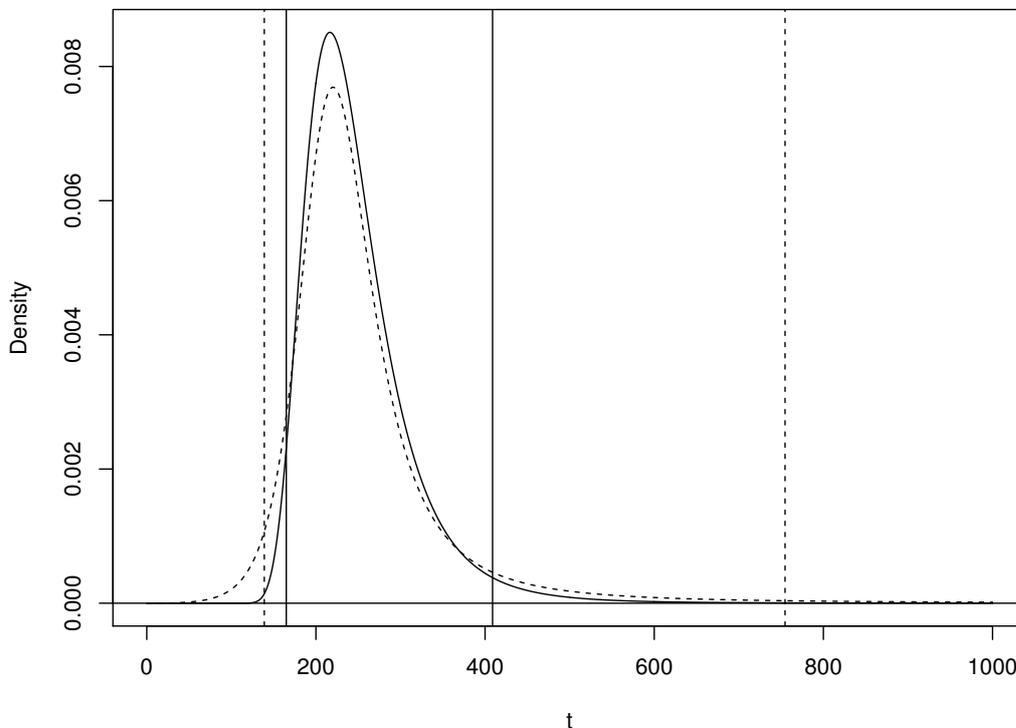
**Figure 5:** The distribution of $\hat{\psi}$ under bivariate normality of $\hat{\beta}_0$ and $\hat{\beta}_1$ (solid line) and under bivariate $t$-distribution with 3 degrees of freedom (dotted line). Also shown are the corresponding 95% CI's in each of the two cases, that is, [164.96, 408.84] under normality, and [138.95, 754.66] under $t$-distribution.

As $b$ becomes large, the first term of $\delta(t)$ goes to one and the second term goes to zero so, for large $b$, $f_R(t)$ is approximated by the normal density $f_R^*(t)$ in (16), and $f_{R'}(t)$ by the shifted and rescaled according to Lemma 3 normal density $f_{R'}^*(t)$. Note that this approximation is valid irrespective of the value of $a$. On the other hand, as *both* $a$ and $b$ go to zero, $f_R(t)$ goes to the standard Cauchy density and $f_{R'}(t)$ goes to the $C(\xi, \nu)$ density in (4.5) with $\xi = c_1$ and $\nu = c_2$. This is obvious from either (4.2) or (4.6) and Lemma 3.

To use these results in our application, we compute the four parameters in (4.3) and (4.4) and obtain:

$$a = 1305.7, \quad b = 4.632, \quad c_1 = -1.307, \quad \text{and} \quad c_2 = 0.8395,$$

from which we obtain the distributions in Figure 5. Note that for these large values of $a$ and $b$, the density in (4.2) cannot be evaluated due to the above-mentioned numerical problems, but the alternative forms in (4.6) and (4.7) can be evaluated without any problems. The solid line in Figure 5 corresponds to the Marsaglia distribution and the solid vertical lines define the Fieller-normal 95% CI in (3.2). Using numerical integration in R, we compute the area under the curve between the bounds in (3.2) and find that the area is indeed 95%: Using the `marsaglia()` function in R given in the Appendix and the

OLS estimates from Section 3 to specify $par = (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$, we obtain

```
> par <- c(110.525,0.4697,0.15751,0.10140,-0.84139)
> mars <- function(x){marsaglia(x,par)}
> integrate(mars,164.96,408.84)
0.9499881 with absolute error < 6e-07
```

Similarly, the dotted line in Figure 5 corresponds to the Press distribution and the dotted vertical lines define the Fieller-t $95\%$ CI in (3.1). Using again numerical integration in R, we compute the area under the dotted curve between the bounds in (3.1) and find that the area is also $95\%$ as expected: Using the `press()` function given in the Appendix, the same *par* vector and *nu* $= 3$ d.f. we compute

```
> pres <- function(x){press(x,par,3)}
> integrate(pres,138.95,754.66)
0.9499949 with absolute error < 1.2e-07
```

We conclude that inference based on Fieller intervals and that based on exact distribution computations yield identical results.

# Appendix: Source R Code for Fieller Intervals and Ratio Densities

The following R functions compute the densities discussed in the main text. The function `marsaglia()` computes the ratio of normals density in (4.10), while the function `press()` computes the ratio of $t$'s density in Lemma 2. Their inputs are $z$, a vector of values over which the density is to be evaluated, and *par* a $5 \times 1$ vector containing $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$, in that order. Function `press()` also takes the single-value parameter *nu*, the degrees of freedom of the $t$ variables. The parameters $a, b, c_1, c_2$ in (4.3) and (4.4) are computed internally from *par*.

```
marsaglia <- function(z,par){
   mx <-par[1];my<-par[2];sx<-par[3];sy<-par[4];r<-par[5]
   vx <-sx^2;vy<-sy^2
   a <- mx/(sx*sqrt(1-r^2))-(r*my)/(sy*sqrt(1-r^2)); b<-my/sy
   c1 <- (sx*r)/sy; c2 <-(sx*sqrt(1-r^2))/sy
   zn <- (z-c1)/c2
   q <- (b+a*zn)/sqrt(1+zn^2)
   h <- (b*zn-a)/sqrt(1+zn^2)
   f1 <- b/sqrt(1+zn^2)*dnorm(h)
   f2 <- q*(2*pnorm(q)-1)/(b*sqrt(1+zn^2))
   f3 <- exp(-.5*(a^2+b^2))/(pi*(1+zn^2))
   prob <- (1/c2)*(f1*f2+f3)
   return(prob)
}


press <- function(z,par,nu){
   mx <-par[1];my<-par[2];sx<-par[3];sy<-par[4];r<-par[5]
```

```
    vx <-sx^2;vy<-sy^2
    a <- mx/(sx*sqrt(1-r^2))-(r*my)/(sy*sqrt(1-r^2)); b<-my/sy
    c1 <- (sx*r)/sy; c2 <-(sx*sqrt(1-r^2))/sy
    zn <- (z-c1)/c2
    k1 <- 1/(pi*(1+(a^2+b^2)/nu)^(nu/2))
    k2 <- (sqrt(pi)*nu^((nu+2)/2)*gamma((nu+1)/2))/
        (2*gamma((nu+2)/2)*(1+(a^2+b^2)/nu)^(-nu/2))
    q1 <- -(a*zn+b)/sqrt(1+zn^2)
    q2 <- sqrt(a^2+b^2+nu-q1^2)
    prob <- (1/c2)*((k1/(1+zn^2))*(1+(k2*q1/q2^(nu+1))*
            (2 * pt(q1*sqrt(nu+1)/q2,nu+1) -1)))
    return(prob)
}
```

The following R code computes the OLS estimates and implements Zerbe's (1978) algorithm to compute the $95\%$ Fieller-t interval for $\psi$, for the data in Table 1.

```
X <- cbind(c(1,1,1,1,1), 3*c(0,.2987,.4648,.5762,.8386))
y <- c(110.551,111.108,110.995,111.18,111.858)

bhat <- solve(t(X)%*%X)%*%t(X)%*%y
print(bhat)
           [,1]
[1,] 110.5245067
[2,]   0.4697037

yhat <- X%*%bhat; uhat<-y-yhat; shat<-sqrt(sum(uhat^2)/3)
Vhat <- shat^2 * (solve(t(X)%*%X))
print(Vhat)
            [,1]          [,2]
[1,]  0.02480802 -0.01343743
[2,] -0.01343743  0.01028128

psi <- bhat[1]/bhat[2]
print(psi)
[1] 235.3069

K <- matrix(c(1,0),1,2); L <- matrix(c(0,1),1,2)
a1 <- (L%*%bhat)^2 - qt(.975,3)^2 * (L%*%Vhat%*%t(L))
a2 <- 2*(qt(.975,3)^2*(K%*%Vhat%*%t(L))-(K%*%bhat)*(L%*%bhat))
a3 <- (K%*%bhat)^2-qt(.975,3)^2*(K%*%Vhat%*%t(K))

low <- (-a2-sqrt(a2^2-4*a1*a3))/(2*a1)
up  <- (-a2+sqrt(a2^2-4*a1*a3))/(2*a1)
print(c(low,up))
[1] 138.9486 754.6641
```

# Acknowledgment

# References

[1] Abramowitz M., and Stegun, I. A. (1972): *Handbook of Mathematical Tables*, New York: Dover.

[2] Cedilnik, A., Košmelj, K., and Blejec, A. (2004): The distribution of the ratio of jointly normal variables. *Metodološki zvezki*, **1**, 99–108.

[3] Fieller, E.C. (1944): A fundamental formula in the statistics of biology assay and some applications. *Quartely Journal of Pharmacy and Pharmacology*, **17**, 117–123.

[4] Fitzpatrick R. (2009): *Theoretical Fluid Mechanics*. Austin: University of Texas at Austin Press.

[5] Hinkley, D.V. (1969): On the ratio of two correlated normal random variables. *Biometrika*, **56**, 635–639.

[6] Luxburg, von U., and Franz, V.H. (2009): A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations and bootstrap. *Statistica Sinica*, **19**, 1095–1117.

[7] Marsaglia, G.M. (1965): Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, **60**, 193–204.

[8] Marsaglia, G.M. (2006): Ratios of normal variables. *Journal of Statistical Software*, **16**, 1–10.

[9] Press, S.J. (1969): The t ratio distribution. *Journal of the American Statistical Association*, **64**, 242–252.

[10] Stigler, S.M. (1986): *The History of Statistics*, Cambridge: Harvard University Press.

[11] Zerbe, G.O. (1978): On Fieller's theorem and the general linear model. *The American Statistician*, **32**, 103–105.