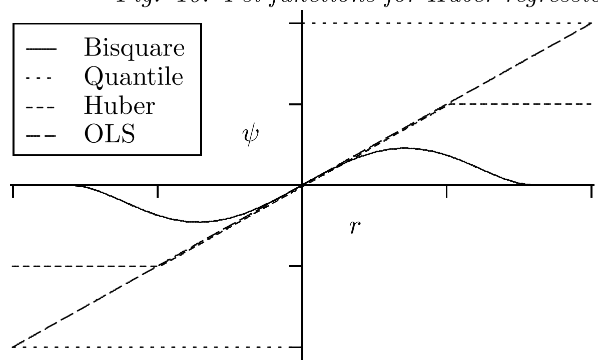
4. Quantile Regression Justification
   a. L^1 regression is MLE for Laplace error method.
   b. Has advantage that fit is less driven by outliers.
      i. Ex. L^2 will move line more towards outlier than will L^1.

MPV: 15.1.2

5. Robust estimation via Huber's method.
   a. Model: $Y_i = \boldsymbol{\beta}^\top \boldsymbol{x}_i + \epsilon_i$, ϵ_i iid, $\mathrm{Var}\left[\epsilon_i\right] = \sigma^2$ for σ^2 known.
   b. Recall: Least squares estimator minimizes $\sum_{i=1}^n (Y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i)^2$
      i. Solves: Least squares estimator $\sum_{i=1}^n 2(Y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i)\boldsymbol{x}_i = 0$
   c. Control the effect of residuals:
      i. Minimize $\sum_{i=1}^n \rho((Y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i)/\sigma)\sigma \boldsymbol{x}_i = 0$
      ii. Set $\sum_{i=1}^n \psi((Y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i)/\sigma)\sigma \boldsymbol{x}_i = 0$
   d. Can express as: $\sum_{i=1}^n w_i(Y_i - \boldsymbol{\beta}^\top \boldsymbol{x}_i)\boldsymbol{x}_i = 0$ for $w_i = w(r) = \psi(r)/r$, r is standardized residual.
6. Common choices for $w(r)$
   a. OLS: $\psi(r) = r$ and $w(r) = 1$.
   b. Quantile Regression: $\psi(r) = \mathrm{sgn}(r)$ and $w(r) = \begin{cases} 0 & \text{if } r = 0 \\ 1/\left|r\right| & \text{if } r \neq 0 \end{cases}$.
   c. Huber: $\psi(r) = \begin{cases} r & \text{if } |r| < c \\ c\,\mathrm{sgn}(r) & \text{if } |r| \geq c \end{cases}$ and $w(r) = \begin{cases} 1 & \text{if } |r| < c \\ c/\left|r\right| & \text{if } |r| \geq c \end{cases}$.

   d. Tukey's Bisquare: $w(r) = \begin{cases} (1-(r/c)^2)^2 & \text{if } |r| < c \\ 0 & \text{if } |r| \geq c \end{cases}$ and $\psi(r) = rw(r)$.
   e. It is useful to compare these functions.
      i. See Fig 10.

Fig. 10: Psi functions for Huber regression



7. To Estimate Parameters:
   a. Iteratively Re-Weighted Least Squares:
      i. Pick initial choice of weights.
      ii. Estimate linear and dispersion parameters
      iii. Recalculate weights.
      iv. Return to estimation step.
   b. In more realistic setting with $\mathrm{Var}\left[\epsilon_i\right]$ to be estimated, estimating equations are modified a bit to avoid opportunity to get better estimate by shrinking σ to zero.

MPV: 15.3

VII. Bioassay
  A. Preliminary problem:
    1. Definitions:
       a. $Y_j = \zeta + \epsilon_j$,
       b. $W_j = \mu + \delta_j$,
       c. $(\epsilon_j, \delta_j) \sim \mathcal{N}\left(\mathbf{o}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}\right)$.
    2. Estimate $\xi = \zeta/\mu$
       a. Consider estimator $\bar{Y}/\bar{W}$
       b. Problem: random variable in denomonator.
          i. Unfortunately distribution is non-standard
          ii. If $\rho = 0$ (easy case) expectation is

$$n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\frac{(w-\mu)^2}{2\tau^2/n} - \frac{(y-\zeta)^2}{2\sigma^2/n})}{2\pi\tau\sigma} \frac{y}{w}\, dw\, dy.$$

          iii. Can perform inner integration:

$$\sqrt{n} \int_{-\infty}^{\infty} \frac{\exp(-\frac{(w-\mu)^2}{2\tau^2/n})}{(2\pi)^{1/2}\tau} \frac{\zeta}{w}\, dw.$$

            • Integral doesn't converge absolutely.
            • Similar to log odds ratio case
    3. Two Approximations to the mean ratio distribution.
       a. Delta method: $\boldsymbol{U}$ a random vector, $\boldsymbol{V} = g(\boldsymbol{U})$ for some known function g.
          i. Know g, $\mathrm{E}\left[\boldsymbol{U}\right]$, $\mathrm{Var}\left[\boldsymbol{U}\right]$.
          ii. Want $\mathrm{Var}\left[\boldsymbol{V}\right]$
       b. Construct a Taylor approximation for $g(\boldsymbol{U})$ about $\mathrm{E}\left[\boldsymbol{U}\right]$.

          i. $\boldsymbol{V} = g(\boldsymbol{U}) \approx g(\mathrm{E}\left[\boldsymbol{U}\right]) + g'(\mathrm{E}\left[\boldsymbol{U}\right])(\boldsymbol{U} - \mathrm{E}\left[\boldsymbol{U}\right])$.
          ii. $\mathrm{Var}\left[\boldsymbol{V}\right] \approx g'(\mathrm{E}\left[\boldsymbol{U}\right])^\top \mathrm{Var}\left[\boldsymbol{V}\right] g'(\mathrm{E}\left[\boldsymbol{U}\right])$
       c. Using delta method, mean and variance of approximating distribution are $\zeta/\mu = \xi$ and $\begin{pmatrix} \frac{1}{\mu} & -\frac{\xi}{\mu^2} \end{pmatrix} \begin{pmatrix} \sigma^2/n & \rho\sigma\tau/n \\ \rho\sigma\tau/n & \tau^2/n \end{pmatrix} \begin{pmatrix} \frac{1}{\mu} \\ -\frac{\xi}{\mu^2} \end{pmatrix} = \mu^{-2}n^{-1}(\sigma^2 - 2\,\rho\,\sigma\,\tau\,\xi + \tau^2\,\xi^2)$.
       d. Exact distribution:
          i. Let $U = \frac{\bar{W} - t\bar{Y} + t\mu - \zeta}{\sqrt{\tau^2/n + t^2\sigma^2/n - 2t\rho\sigma\tau/n}}$ and $V = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$
          ii. Let $u = \frac{\sqrt{n}(t\mu - \zeta)}{\sqrt{\tau^2 + t^2\sigma^2 - 2t\rho\sigma\tau}}$ and $v = \frac{-\mu}{\sigma/\sqrt{n}}$.
          iii. $\mathrm{P}\left[\bar{W}/\bar{Y} \leq t\right]$ is

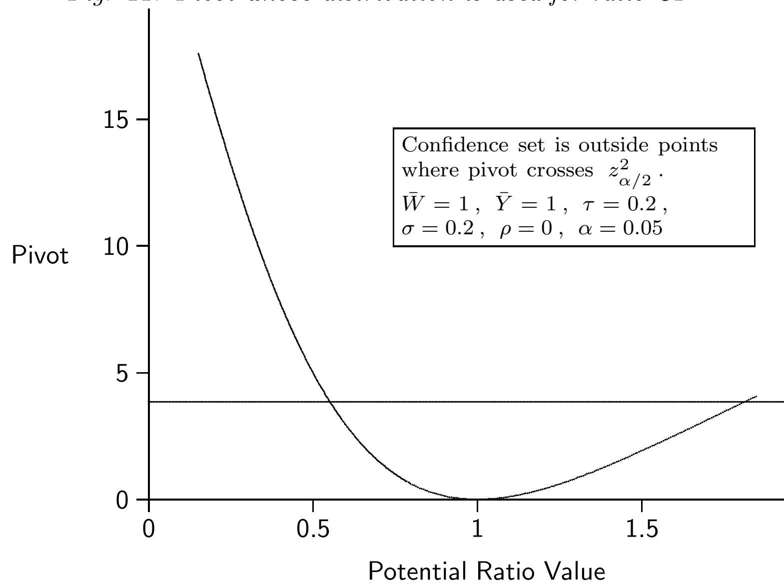
$$= \mathrm{P}\left[\bar{W} - t\bar{Y} \leq 0 \& \bar{Y} > 0\right] + \mathrm{P}\left[\bar{W} - t\bar{Y} \geq 0 \& \bar{Y} < 0\right]$$
$$= \mathrm{P}\left[U \leq u \& V > v\right] + \mathrm{P}\left[U \geq u \& V \leq v\right]$$
$$= \mathrm{P}\left[U \leq u\right] - \mathrm{P}\left[U \leq u \& V \leq v\right] + \mathrm{P}\left[U \geq u \& V \leq v\right]$$
$$= \Phi\left(\frac{\sqrt{n}(t-\xi)}{\sqrt{\tau^2 + t^2\sigma^2 - 2t\rho\sigma\tau}/\mu}\right) + R$$

for $|R| \leq \Phi(-\sqrt{n}\mu/\sigma)$.
    4. Confidence intervals
       a. $\xi\bar{W} - \bar{Y} \sim \mathcal{N}(0, \xi^2\tau^2/n + \sigma^2/n - 2\rho\sigma\tau\xi/n)$
       b. $\mathrm{P}\left[\frac{(\xi\bar{W}-\bar{Y})^2}{\xi^2\tau^2/n + \sigma^2/n - 2\rho\sigma\tau\xi/n} \leq z^2_{\alpha/2}\right] = 1 - \alpha$.
       c. Set of ξ satisfying statement inside probability is CI (sort of)
          i. Restriction is quadratic inequality.
       d. If $\rho = 0$, then

i. There's no need to consider the pairing
ii. You need not require equal number of contributors to each mean.
iii. Suppose $\bar{W}$ and $\bar{Y}$ are the means of m and n i.i.d. observations resp..
iv. Defining analog is $\mathrm{P}\left[\frac{(\xi\bar{W}-\bar{Y})^2}{\xi^2\tau^2/m+\sigma^2/n}\le z^2_{\alpha/2}\right]=1-\alpha$.

5. Solution at equality gives confidence set endpoints.
   a. End points are
      $\frac{R-\rho\,\sigma\,\tau\,Z^2\pm Z\sqrt{R^2\,\tau^2-2\,R\,\rho\,\sigma\,\tau+\sigma^2\,(1-(1-\rho^2)\,\tau^2\,Z^2)}}{(1-\tau^2\,Z^2)}$
      for $Z=z_{\alpha/2}/(\bar{W}\sqrt{n})$
      i. If $\rho=0$, end points are
         $\frac{R\pm Z\sqrt{R^2\,\tau^2+(n/m)\sigma^2\,(1-\tau^2\,Z^2)}}{(1-\tau^2\,Z^2)}$
   b. Method generally called *Fieller's method*.
      i. See Fig 11.
6. Possible deviant behavior
   a. Value of squared normal deviate for very large $|\xi|$ is $\bar{W}n/\tau^2$
   b. Hence confidence interval is outside of end points.
   c. If and only if denominator of endpoints is negative.
      i. See Fig. 12.
   d. More extreme case: quadratic equation has no roots.
      i. See Fig 13.
7. Special Cases
   a. $\tau=0\Rightarrow$ End points are $R\pm Z\sigma$, the usual CI with denominator known
   b. For large n,

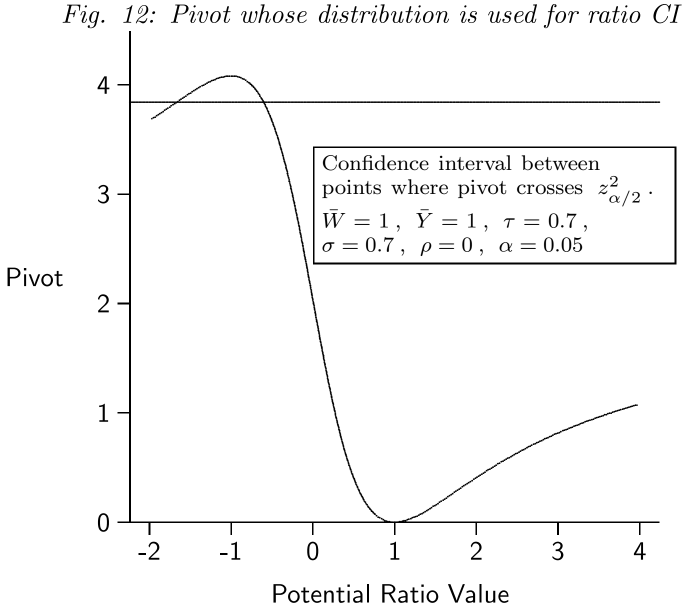
Fig. 11: Pivot whose distribution is used for ratio CI



i. keeping terms that are constant or have a multiple of $1/\sqrt{n}$.
ii. deleting others,
iii. terms with Z^2 get deleted,
iv. the result is $R\pm Z\sqrt{\sigma^2-2\,R\,\rho\,\sigma\,\tau+R^2\,\tau^2}$
v. Same as delta method solution with R and $\bar{W}$ in place of ξ and μ

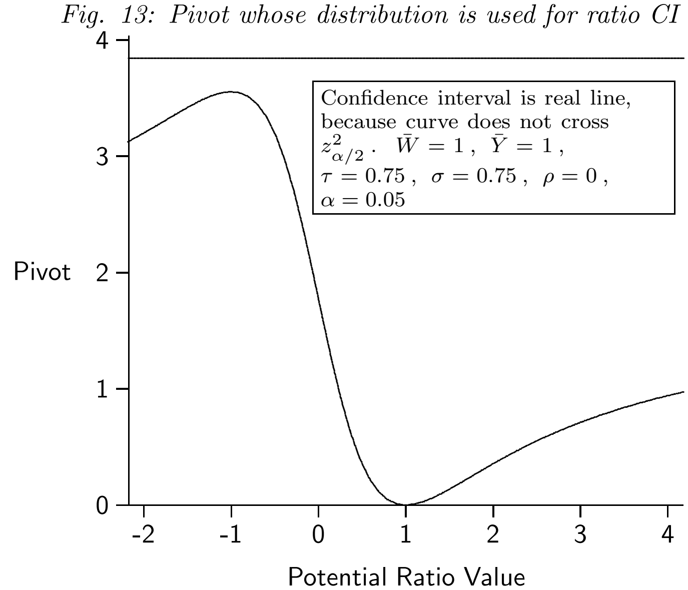
8. More realistic case: σ and τ unknown.

Fig. 12: Pivot whose distribution is used for ratio CI



   a. Assume $\sigma=\theta\tau$, for θ known, and ρ known.
   b. Estimate σ with quantity $\tilde{\sigma}$ so that $\tilde{\sigma}^2/\sigma^2\sim\chi^2_d$
   c. For example
      i. θW_j and $(Y_j-\rho\theta W_j)/\sqrt{1-\rho^2}$ are uncorrelated, mean zero, variance σ^2
      ii. $s^2=\frac{\theta^2\sum_j(W_j-\bar{W})^2+\sum_j(Y_j-\bar{Y}-\rho\theta(W_j-\bar{W}))^2/(1-\rho^2)}{n-1+n-1}$
         is unbiased estimator of σ^2, independent of $\bar{W}$ and

Fig. 13: Pivot whose distribution is used for ratio CI



         $\bar{Y}$, with a $\chi^2_{n-1+n-1}$ distribution
      iii. Hence squared deviate defining CI has a $t_{n-1+n-1}$ distribution before squaring
   d. Same CI except with t critical value.
9. Example: inference on ordinary regression inverse
   a. Want CI for x_0 satisfying $\beta_0+\beta_1x_0=y_0$.
   b. Let $\hat{x}_0=(y_0-\hat{\beta}_0)/\hat{\beta}_1$
   c. y_0 has no error, since we pick it.
   d. $(\hat{\beta}_0,\hat{\beta}_1)$ has a bivariate normal distribution with known

correlation.

e. CI is

$$\hat{x}_0 + \frac{(\hat{x}_0 - \bar{x})g \pm (t\hat{\sigma}/\hat{\beta}_1)\sqrt{(\hat{x}_0 - \bar{x})^2/S_{xx} + (1-g)(1+1/n)}}{1-g}$$

i. for $g = (t^2\hat{\sigma}^2/(\hat{\beta}_1^2 S_{xx})$ and t the critical value.

ii. This version of formula is from Greenwell and Kabban (2014), R investr package documentation.

This page intentionally left blank.

This page intentionally left blank.

This page intentionally left blank.