The Effects of Prior Hypothesis Testing on the Sampling Properties of Estimators and Tests: An Overview

David E. Giles

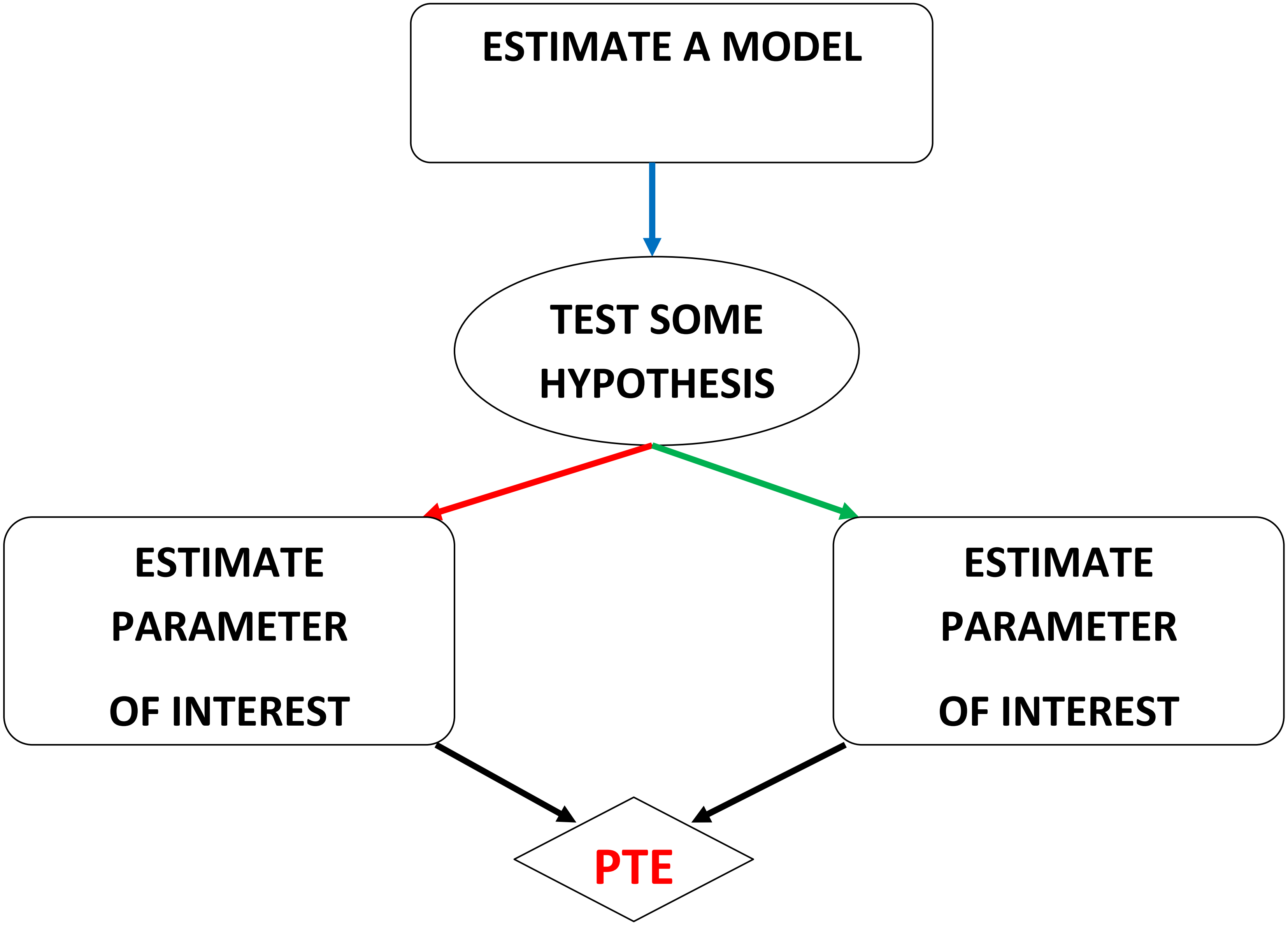
Dept. of Economics, UVic

dgiles@uvic.ca

Overview

- Motivation
- Historical overview
- Some general results
- Example 1: Testing the equality of two variances
- Example 2: Testing restrictions on regression coefficients
- Further examples
- Pre-test testing
- Extensions
- Summary

Motivation


ESTIMATE A MODEL
TEST SOME HYPOTHESIS
ESTIMATE PARAMETER OF INTEREST
ESTIMATE PARAMETER OF INTEREST
PTE

That is……….

- Test some hypothesis of interest
- If we Reject H_0, then estimate parameters using one estimator
- If we Do Not Reject H_0, estimate parameters using a different estimator
- Result is a "Pre-Test Estimator"
- $\hat{\theta} = I_{(S\in R)}\tilde{\theta} + I_{(S\in \bar{R})}\theta^*$
- $\hat{\theta}$ is a weighted sum of its 2 "component estimators", with *random weights*
- What are the sampling properties of $\hat{\theta}$?
- Similar situation if we pre-test, and then test again, based on either $\tilde{\theta}$ or θ^*
- What are the properties of the second test?

Historical overview

- June 1944 issue of *AMS* had papers by Halmos, Hurwicz, Robbins, Scheffé, Tukey, Wald, Wolfowitz; .... and by Ted Bancroft (1907 - 1986)
- "On biases in estimation due to the use of preliminary tests of significance"
- Work was motivated by Berkson (*JASA*, 1942), "Tests of significance considered as evidence"
- Subsequently many papers by Bancroft & his students (*e.g.*, Han)
- Work by Dudley Wallace & students (*e.g.*, Toro-Vizcarrondo, Toyoda, Brook)
- Work by George Judge, Mary-Ellen Bock, Tom Yancey, students

Some general results

- Measure finite-sample performance in terms of estimators' risks
- $L\left(\hat{\theta}(y), \theta\right) \geq 0$; $L = 0$ iff $\hat{\theta}(y) = \theta$
- $r\left(\hat{\theta}\right) = E_y[L(\hat{\theta}(y)\ , \theta)]$
- *Quadratic loss* (i) Scalar θ : $r\left(\hat{\theta}\right) = MSE(\hat{\theta}\ )$

  (ii) Vector θ : $r\left(\hat{\theta}\right) = tr[MMSE(\hat{\theta}\ )]$

  $= tr[V(\hat{\theta}\ ) + \text{Bias}(\hat{\theta}\ )\text{Bias}(\hat{\theta}\ )']$
- PTE's are *inadmissible* under quadratic (& many other loss functions),

  because they're discontinuous functions of the sample data (Cohen, 1965)

Testing the equality of two variances

- Bancroft's first problem
- Simple random sampling, independently from 2 Normal populations
- $x_{1i} \sim N[\mu, \sigma_1^2]$; $i = 1, 2, ..., n_1$; $v_1 = (n_1 - 1)$
- $x_{2i} \sim N[\mu, \sigma_2^2]$; $i = 1, 2, ..., n_2$; $v_2 = (n_2 - 1)$
- H_0: $\sigma_1^2 = \sigma_2^2$ *vs.* H_1: $\sigma_1^2 > \sigma_2^2$.
- F-test is UMPI. $f = (s_1^2 / s_2^2)$, where $s_j^2 = \frac{1}{(v_j)} \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2$; $j = 1, 2$
- $f > c_{(\alpha)}$: Reject H_0. Use $s_1^2 = \frac{1}{(v_1)} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$
- $f \leq c_{(\alpha)}$: Do Not Reject H_0. Use $s^2 = \frac{1}{(v_1 + v_2)} [\, v_1 s_1^2 + v_2 s_2^2 \,]$

- The pre-test estimator of σ_1^2 is: $\hat{\sigma}_1^2 = I_{(f>c)} s_1^2 + I_{(f\leq c)} s^2$

- Note: $$E[\hat{\sigma}_1^2] = E\left[\left(1 - I_{(f\leq c)}\right)s_1^2\right] + E\left[I_{(f\leq c)} s^2\right]$$

$$= \sigma_1^2 + E\left[\left(s^2 - s_1^2\right) I_{(f\leq c)}\right]$$

- Going to be difficult to evaluate!

- Bancroft determined the Bias and the Variance of this pre-test estimator

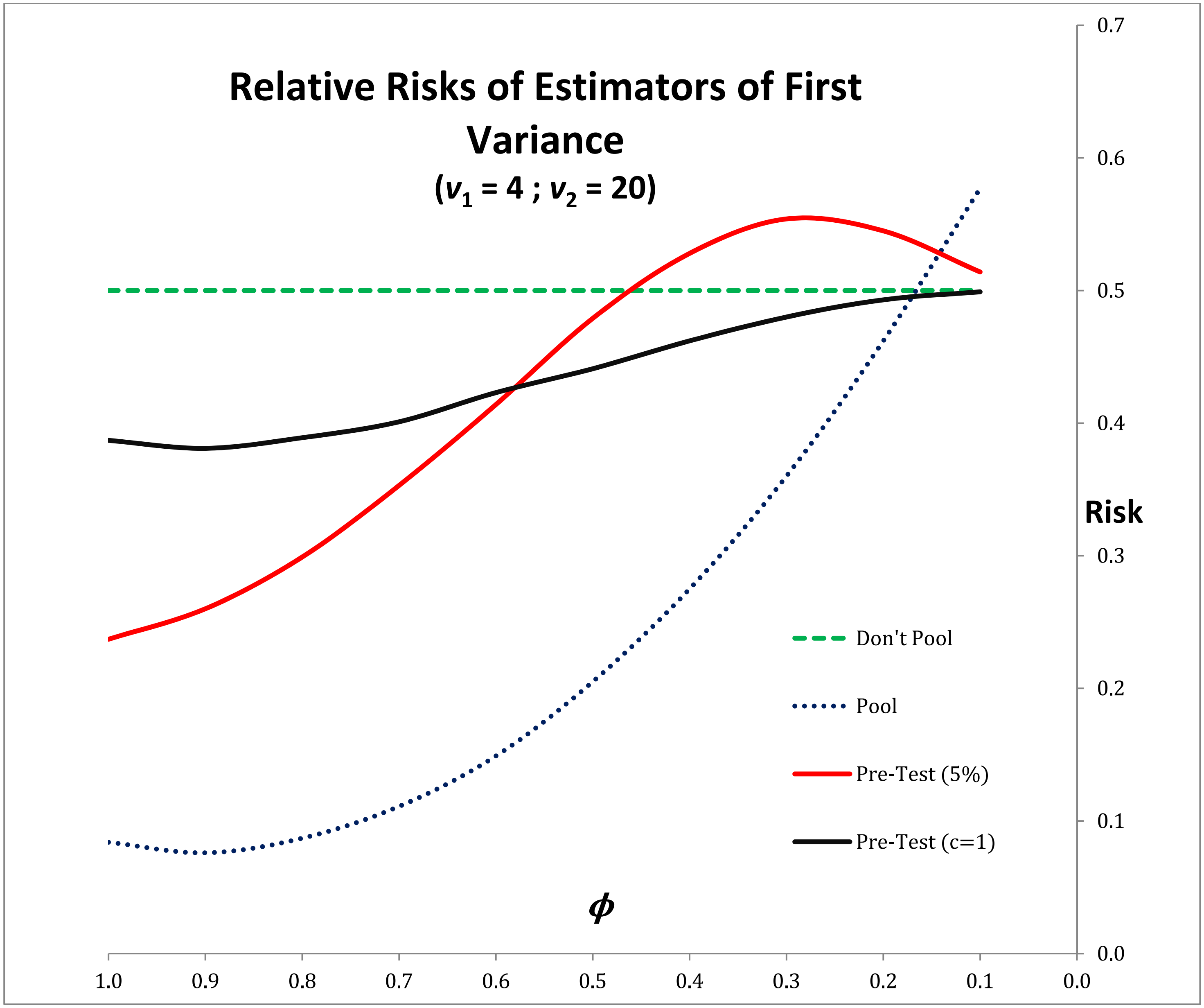
- Bias $(\hat{\sigma}_1^2) = \frac{\sigma_1^2 v_1}{(v_1+v_2)} [B_q\left(\frac{v_1}{2}; \frac{v_2}{2} + 1\right)\phi - B_q(\frac{v_1}{2} + 1; \frac{v_2}{2})]$

- $\phi = \frac{\sigma_2^2}{\sigma_1^2}$; $q = (v_1 \phi c)/(v_2 + v_1 \phi c)$

- $B_z(a\, ; b)$ is incomplete Beta function

- Variance – very messy.

- Giles (1992) derived the exact sampling distribution of $\hat{\sigma}_1^2$, and used it to demonstrate effect of pre-testing on coverage probabilities of confidence intervals
- Let's compare the "never pool", "always pool" and "pre-test" estimators in terms of risk under quadratic loss - *i.e.*, MSE


Relative Risks of Estimators of First Variance
(v_1 = 4 ; v_2 = 20)
0.7
0.6
0.5
0.4
Risk
0.3
0.2
0.1
0.0
Don't Pool
Pool
Pre-Test (5%)
Pre-Test (c=1)
ϕ
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

- Always a region of parameter space where "never pool" is worst
- Always a region of parameter space where "always pool" is worst
- Always a region of parameter space where "pre-test" is worst
- Unless $c = 1$:

  (i) Always a region of parameter space where "never pool" is best

  (ii) Always a region of parameter space where "never pool" is best

  (iii) ***Never*** a region of parameter space where "pre-test" is best

- In general, the risk of the PTE depends on v_1, v_2, $\boldsymbol{\phi}$, and α

Testing restrictions on regression coefficients

- Bancroft's second problem
- $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$; $\varepsilon_i \sim iid\ N[0, \sigma^2]$
- H_0: $\beta_2 = 0$ *vs*. H_1: $\beta_2 \neq 0$
- $|t| > c_{(\alpha)}$: Reject H_0. Use $\tilde{\beta}_{1.2}$ (OLS = MLE)
- $|t| \leq c_{(\alpha)}$: Do Not Reject H_0. Use $\tilde{\beta}_1$ (RLS = RMLE)
- PTE: $\hat{\beta}_1 = \tilde{\beta}_{1.2}\, I_{(|t| > c_{(\alpha)})} + \tilde{\beta}_1\, I_{(|t| \leq c_{(\alpha)})}$
- Bancroft evaluated only the ***bias*** of the PTE
- Variance subsequently evaluated by Toro-Vizcarrondo (1968)

- Exact sampling distribution of PTE derived by Srivastava & Giles (1992)
- Problem generalizes to pre-testing the validity of m exact linear restrictions on coefficient vector for multiple linear regression model:
- $y = X\beta + \varepsilon$; $\varepsilon \sim N[0\,, \sigma^2 I_n]$; $v = (n-k)$
- $H_0: R\beta = r$ vs. $H_0: R\beta \neq r$; let $\delta = (R\beta - r)$
- F-test is UMPI . Test statistic is n.c. F, with n.c.p. $\lambda = (\delta'\delta)/(2\sigma^2)$
- $f > c_{(\alpha)}$: Reject H_0. Use $\tilde{\beta} = S^{-1}X'y$; $S = (X'X)$
- $f \leq c_{(\alpha)}$: Do Not Reject H_0. Use $\beta^* = \tilde{\beta} + S^{-1}R'[RS^{-1}R']^{-1}(r - R\tilde{\beta})$
- PTE: : $\hat{\beta} = \tilde{\beta} I_{(f > c_{(\alpha)})} + \beta^* I_{(f \leq c_{(\alpha)})}$
- Risk under quadratic loss – Brook (1972, 1976)

- “Modern” derivation of risk of PTE for this problem

- $\hat{\beta} = \tilde{\beta} I_{(f > c_{(\alpha)})} + \beta^* I_{(f \leq c_{(\alpha)})}$

- $I_{(f \leq c_{(\alpha)})} \times I_{(f > c_{(\alpha)})} = 0 \quad ; \quad I_{(f > c_{(\alpha)})} = 1 - I_{(f \leq c_{(\alpha)})}$

- $r(\hat{\beta}) = E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right]$

$$= r(\tilde{\beta}) - E\left[I_{(f \leq c_{(\alpha)})}(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)\right] + \delta'\delta E\left[I_{(f \leq c_{(\alpha)})}\right]$$

- **Th.1**: If $w \sim MVN[\theta, I_J]$, and A is p.d.s., then for any measurable fctn., $\boldsymbol{\phi}$,

$$E[\phi(w'Aw)w'Aw] = E[\phi\left(\chi^{2'}_{\left(J+2;\frac{\theta'\theta}{2}\right)}\right)]tr.(A) + E[\phi(\chi^{2'}_{(J+4;\theta'\theta/2)})]\theta'A\theta$$

- **Th. 2**: If $w \sim MVN[\theta \,, I_J]$, and A is p.d.s., then for any measurable fctn., $\boldsymbol{\phi}$,

$$E[\phi(w'Aw)w] = \theta E[\phi(\chi^{2'}_{\left(J+2;\frac{\theta'\theta}{2}\right)})]$$

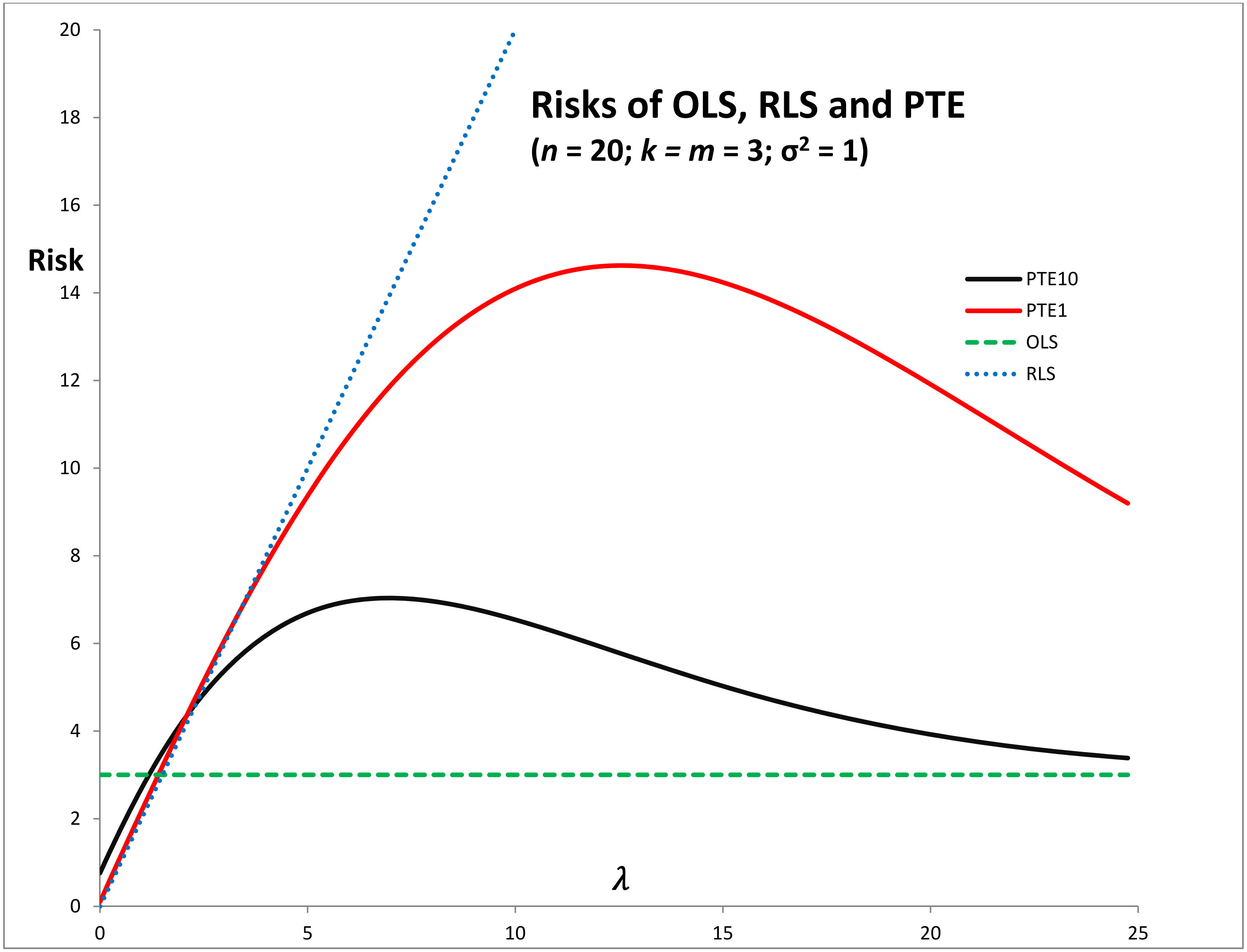
- Using these results, we can show that

$$\boldsymbol{r(\hat{\beta}) = \sigma^2[k + (4\lambda - m)P_{20} - 2\lambda P_{40}]}$$

where

$$P_{ij} = Pr.[F'_{(m+i,v+j;\lambda)} \leq (cm(v+j))/(v(m+i))] \;;\; i, j, = 0, 1, 2, \ldots.$$

and $v = (n - k)$



Risks of OLS, RLS and PTE
(n = 20; k = m = 3; σ² = 1)
Risk
λ
PTE10
PTE1
OLS
RLS
0
2
4
6
8
10
12
14
16
18
20
0
5
10
15
20
25

- Always a region of parameter space where OLS is worst
- Always a region of parameter space where RLS is worst
- Always a region of parameter space where PTE is worst
- Always a region of parameter space where OLS is best
- Always a region of parameter space where RLS is best
- ***Never*** a region of parameter space where PTE is best
- In general, risk of PTE depends on β, σ^2, R, r, n, k, m, α, X

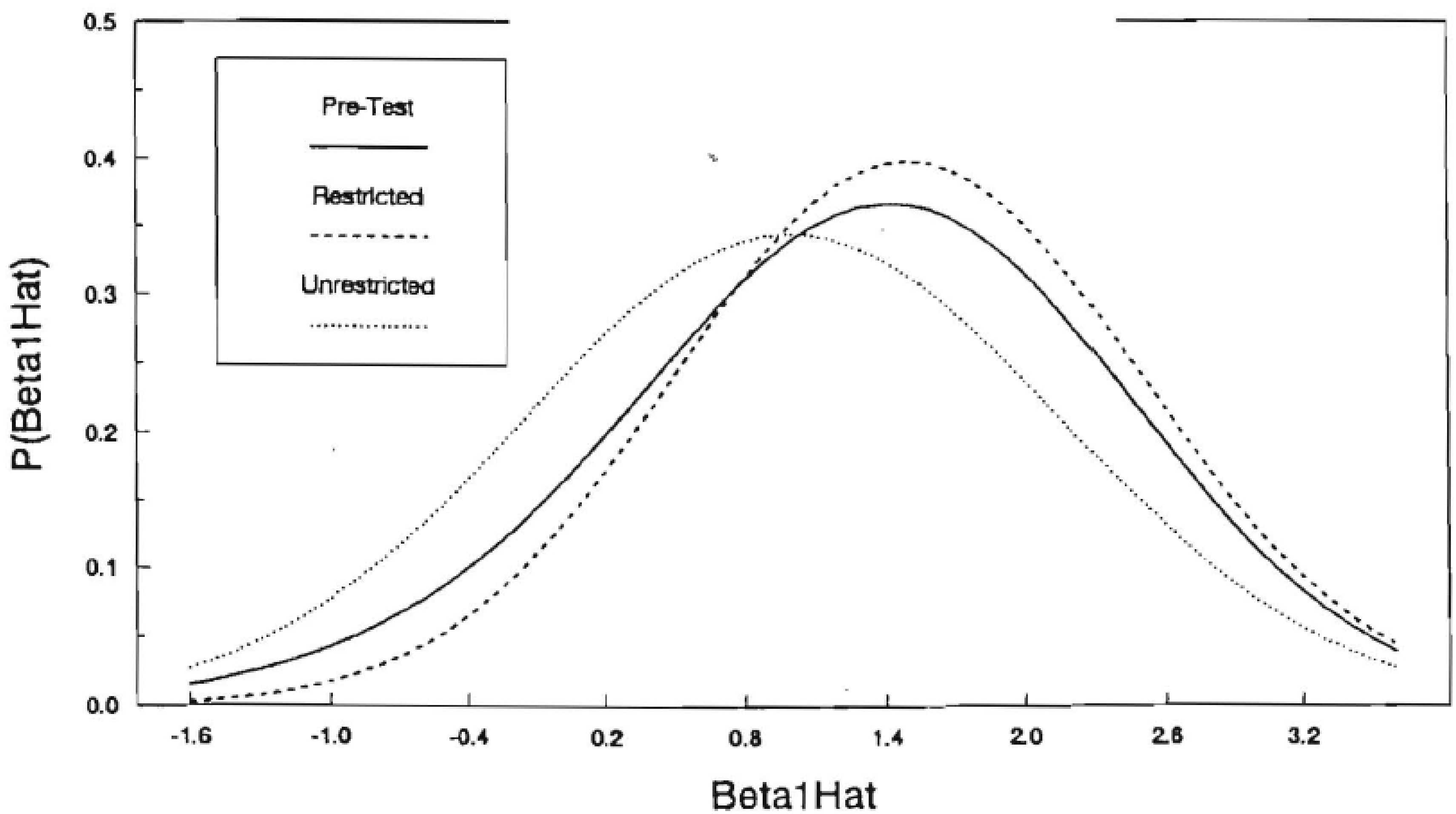


Fig. 2. Probability density function for the pre-test estimator ($\beta_1 = \beta_2 = \sigma = 1$).

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad ; \quad \varepsilon_i \sim iid\ N[0\,,\sigma^2]$$

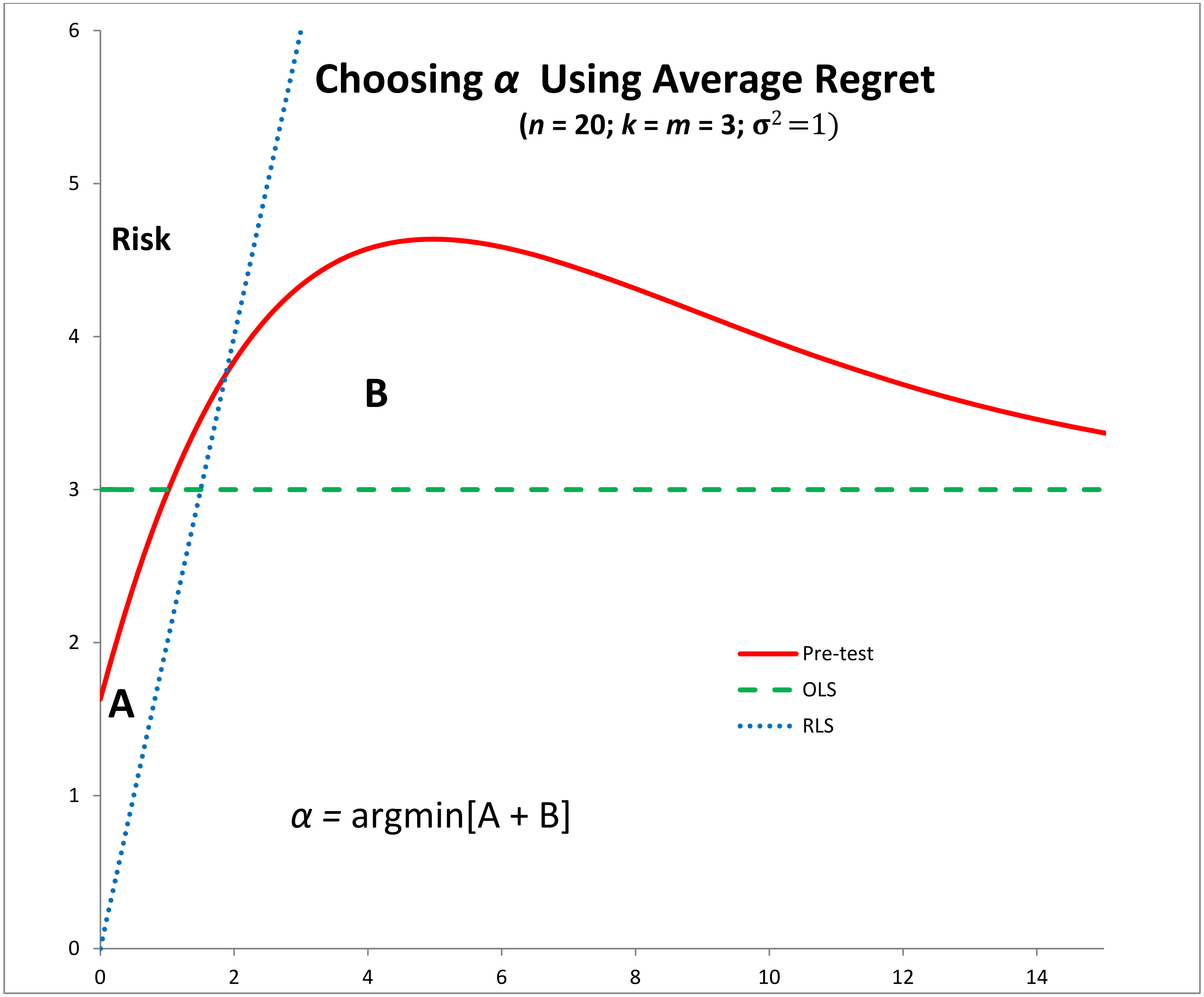
- Dependence of results on choice of significance level for pre-test suggests the question: "Is there an *optimal* choice of α ?"
- Addressed by several authors: Sawa & Hirumatsu (1973), Brook (1972, 1976),Toyoda & Wallace (1975, 1977), Ohtani & Toyoda (1980), Brook & Fletcher (1981), Bancroft & Han (1983), J. Giles & Lieberman (1992), Giles *et al*. (1992)
- Several ways of defining "optimal". For example:

  (i) Minimax regret

  (ii) Minimum average regret

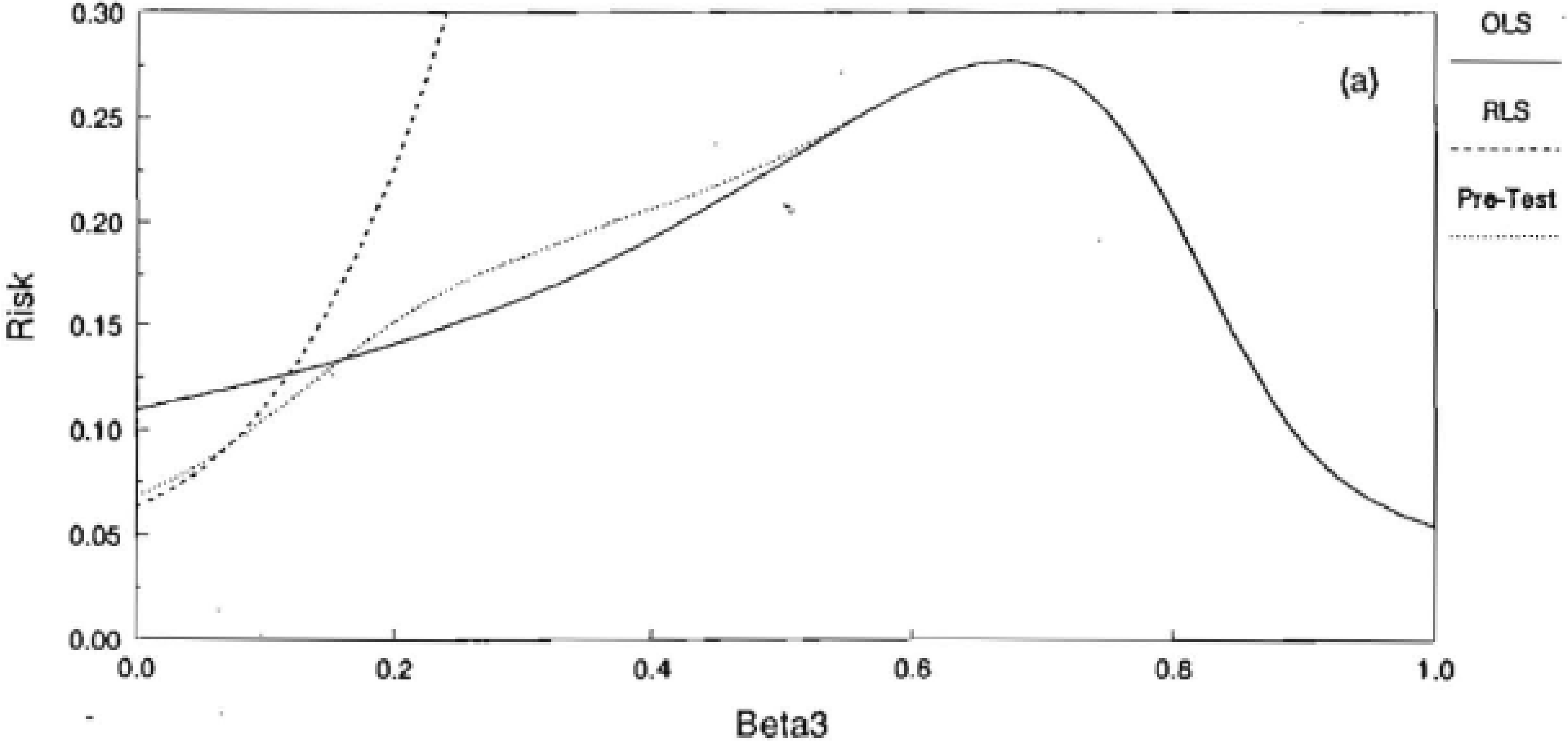
  (iii) Pseudo-Bayesian approach

Choosing α Using Mini-max Regret
(n = 20; k = m = 3 ; $\sigma^2 = 1$; α = 35.7%)
Risk
λ
Pre-test (c = 1.39)
OLS
RLS
0
1
2
3
4
5
6
0
2
4
6
8
10
12
14


Choosing α Using Average Regret
(n = 20; k = m = 3; σ² = 1)
Risk
A
B
α = argmin[A + B]
Pre-test
OLS
RLS
0
1
2
3
4
5
6
0
2
4
6
8
10
12
14

Further examples

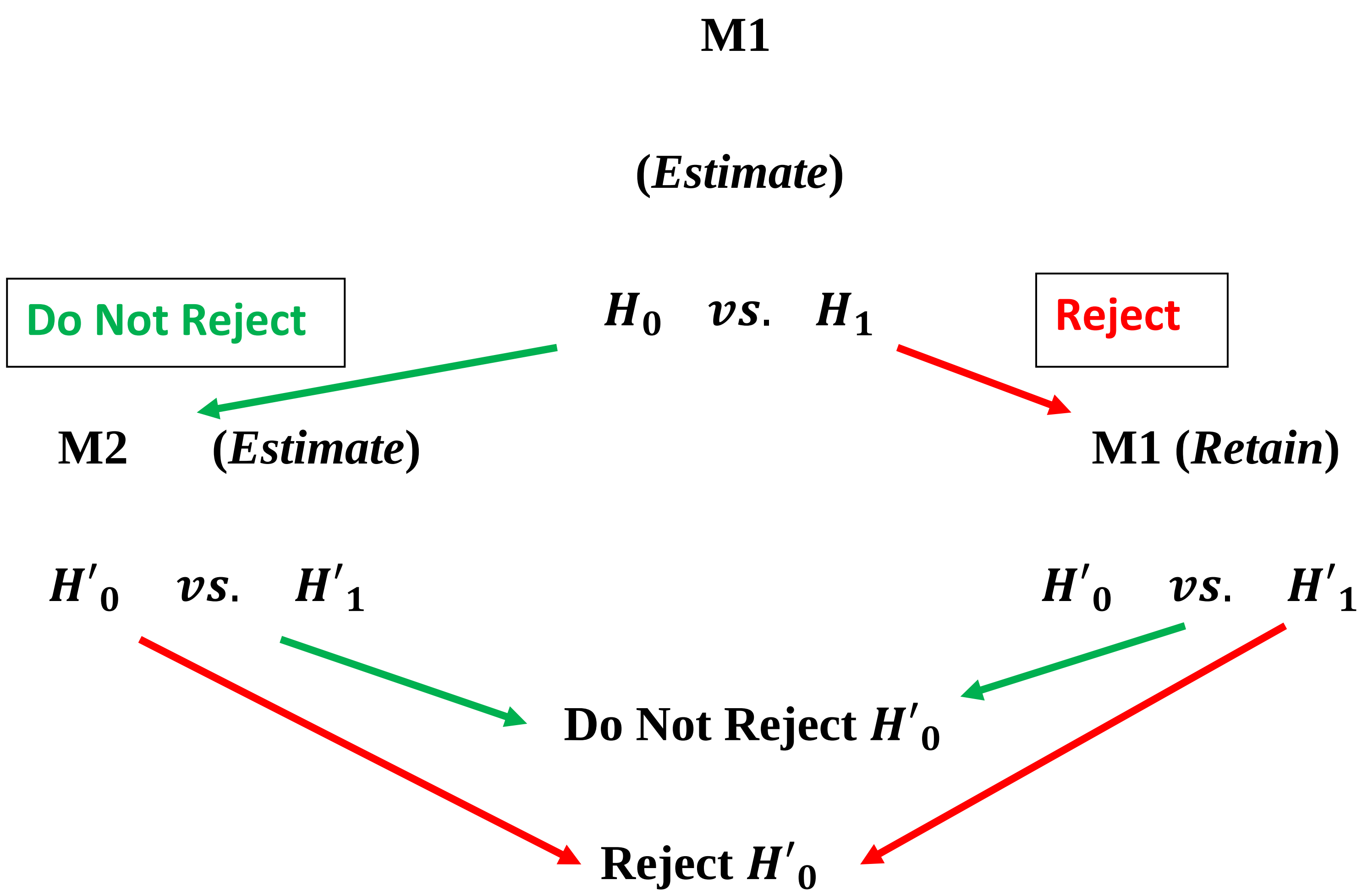
- Can pre-testing ever dominate ***both*** of the component estimators?
- **YES, in some situations – recall Bancroft's 1st problem**
- Ozcam *et al*. (1991) – SURE model
- J. Giles (1992) – variance estimation after pre-test of homogeneity in regression with multivariate Student-t errors: PTE dominates both component estimators in terms of risk under quadratic loss
- Giles & Cuneen (1994) – autoregressive models and pre-test of exact restrictions on coefficients
- Generally, ranges where PTE dominates are quite limited



$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t \quad ; \quad \varepsilon_t \sim iid\ N[0\,, \sigma^2]$$

$$H_0\colon \beta_3 = 0 \quad vs. \quad H_0\colon \beta_3 > 0$$

Pre-test testing



- Generally, when conducting a sequence of tests, the test statistics are ***not independent of each other***
- Size distortion and implications for power
- Example 1

M1 $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \;;\; \varepsilon_i \sim N[0, \sigma_\varepsilon^2]$

$H_0: \beta_2 = 0 \quad vs. \quad H_1: \beta_2 \neq 0$

M2 $y_i = \beta_1 x_{1i} + v_i \;;\; v_i \sim N[0, \sigma_v^2]$

H'_0: Errors serially independent $\quad vs. \quad H'_1$: Errors are AR(1)

- Giles & Lieberman (1992) – size of DW test is distorted upwards
- Recommend (nominal) significance level of up to 50% at 1st stage

- Power results mixed, but can have situations where pre-testing actually *increases* power of DW test (after controlling for size distortion)

- <u>Example 2</u>

M1 $$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad ; \quad u_t \sim N[0, \sigma_u^2] \quad ; \quad -1 < \rho < 1$$

$$H_0: \rho = 0 \quad vs. \quad H_1: \rho \neq 0$$

M2 $$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \; ; \; \varepsilon_t \sim N[0, \sigma_\varepsilon^2]$$

$$H'_0: \beta_2 = 0 \quad vs. \quad H'_1: \beta_2 \neq 0$$

- King & Giles (1984) – little size distortion/power loss if α = 50% at 1st stage

Extensions

- Are the results robust to choice of loss function?

  (i) Absolute error loss – Giles (1993)

  (ii) LINEX loss (asymmetric) – J. Giles & Giles (1993, 1996)

- Are the results robust to non-normality?

  J. Giles – various papers – Spherically symmetric disturbances

- Are the results robust to model mis-specification?

  Omitted regressors – J. Giles, Giles, various papers

- Multi-stage pre-test estimation – very little evidence available – lots of interesting problems here

- Some recent PTE developments include:

(i) Magnus & Durbin (1999) - model averaging

(ii) Danilov & Magnus (2004) - model averaging

(iii) Chmelarova & Hill (2010) – Hausman pre-test estimation

(iv) Guggenberger (2010) – Hausman pre-test testing

(v) De Luca & Magnus (2011) - model averaging

(vi) Llorente & Martín Apaolaza (2011) – symmetry model of categorical data

(vii) Baltagi *et al*. (2011, 2012) – panel data regression with spatial data

Summary

- Pre-testing is very common, but its consequences are often ignored
- Pre-test strategies are *inadmissible*
- Care needs to be paid to "size" of a pre-test
- Pre-testing alters the sampling distributions of subsequent estimators and tests, often in very complicated ways
- Several surveys of the "pre-testing" literature
- Bancroft and Han (1977)
- Han *et al*. (1988)
- J. Giles and Giles, *Journal of Economic Surveys*, 1993