

# The Exact Distribution of the t-Ratio with Robust and Clustered Standard Errors

by

**Bruce E. Hansen**

Department of Economics  
**University of Wisconsin**

June 2017

# t-ratio

- Gosset (1908)
  - ▶ The t-ratio of the sample mean has the exact  $t_{n-1}$  distribution
  - ▶ A fundamental intellectual achievement
- Linear regression
  - ▶ Gosset's result extends to classical  $t$ -ratios (classical standard errors)
  - ▶ Classical  $t$ -ratios have  $t_{n-k}$  distribution

## But...

- Classical standard errors are no longer used in economic research
- Papers use either
  - ▶ Heteroskedasticity-consistent (HC)
  - ▶ Cluster-robust (CR)
  - ▶ Heteroskedasticity-and-autocorrelation-consistent (HAC)
  - ▶ Justification is asymptotic
- Most assess significance (testing and confidence intervals) using finite sample distribution:
  - ▶  $t_{n-k}$  distribution (HC)
  - ▶  $t_{G-1}$  distribution (CR)
  - ▶ THIS IS WRONG!!!

## “reg y x, cluster(id” )

- Regression:

- ▶ Uses HC1 variance estimator
  - ★ White estimator scaled by  $n/(n - k)$
- ▶ Uses  $t_{n-k}$  distribution for p-values and confidence intervals
  - ★ UNJUSTIFIED!

- Clustered:

- ▶ Uses CR1 variance estimator
  - ★ Described later, ad hoc
- ▶ Uses  $t_{G-1}$  distribution for p-values and confidence intervals
  - ★ No finite sample justification

# This paper

- Provides an exact theory of inference
  - ▶ Linear regression with robust standard errors
  - ▶ Linear regression with clustered standard errors
- Exact distribution of HC and CR t-ratios under i.i.d. normality
  - ▶ Computable

# Linear Regression with Heteroskedasticity

- $y_i = x_i' \beta + e_i$
- $E(e_i | x_i) = 0$
- $E(e_i^2 | x_i) = \sigma_i^2$
- $n$  observations
- $k$  regressors
- Core model in applied econometrics

# Heteroskedastic (HC) Variance Estimation: Some History

- Eicker (1963): HC0
- Horn, Horn and Duncan (1975): HC2
- Hinkley (1977): HC1
- White (1980): HC0 for econometrics
- MacKinnon and White (1985): HC3
- Chesher and Jewitt (1987): Bias can be large
- Bera, Suprayitno and Premaratne (2002): Unbiased estimator
- Bell-McCaffrey (2002): Distributional approximation
- Cribari-Neto (2004): HC4
- Cribari-Neto, Souza and Vasconcellos (2007): HC5
- Cattaneo, Jansson and Newey (2017): Many regressors

# HC Variance Estimation

- OLS:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

- Residuals:

$$\hat{e}_i = y_i - x_i'\hat{\beta}$$

- HC0

$$\hat{V}_0 = (X'X)^{-1} \left( \sum_{i=1}^n x_i x_i' \hat{e}_i^2 \right) (X'X)^{-1}$$

- HC1

$$\hat{V}_1 = \frac{n}{n-k} (X'X)^{-1} \left( \sum_{i=1}^n x_i x_i' \hat{e}_i^2 \right) (X'X)^{-1}$$

- ▶ robust covariance matrix in Stata



- HC2:

$$\widehat{V}_2 = (X'X)^{-1} \left( \sum_{i=1}^n x_i x_i' \widehat{e}_i^2 (1 - h_i)^{-1} \right) (X'X)^{-1}$$

- ▶  $h_i = x_i' (X'X)^{-1} x_i$
- ▶ Unbiased under homoskedasticity

- HC3:

$$\widehat{V}_3 = (X'X)^{-1} \left( \sum_{i=1}^n x_i x_i' \widehat{e}_i^2 (1 - h_i)^{-2} \right) (X'X)^{-1}$$

- ▶ jackknife

## HC3 (jackknife) is a conservative estimator

**Theorem.** In the linear regression model,

$$E\left(\widehat{V}_3 \mid X\right) \geq V = E\left(\left(\widehat{\beta} - \beta\right)\left(\widehat{\beta} - \beta\right)' \mid X\right)$$

(However, **inference** using HC3 is not necessarily conservative.)

# HC t-ratios

- t-ratio for  $R'\beta$ :

$$t = \frac{R'(\hat{\beta} - \beta)}{\sqrt{R'\widehat{V}R}}$$

- Distribution theory

- ▶ Asymptotic:  $t \rightarrow_d N(0, 1)$
- ▶ This is what we (typically) teach

- Distribution used in practical applications

- ▶ Finite Sample:  $t \sim t_{n-k}$
- ▶ This is what most applied papers use
- ▶ Incorrect

# Clustered Samples

- Observations are  $(y_{ig}, \mathbf{x}_{ig})$ 
  - ▶  $g = 1, \dots, G$  indexes cluster (group)
  - ▶  $i = 1, \dots, n_n$  indexes observation within  $g^{th}$  cluster
- Clusters are mutually independent
- Observations within a cluster have unknown dependence
- In panels,  $(y_{ig}, \mathbf{x}_{ig})$  could be demeaned observations
  - ▶ Assumptions fully allow for this
- Number of observations  $n_g$  per cluster may vary across cluster
- Total number of observations  $n = \sum_{g=1}^G n_g$

# Cluster Regression

- $\mathbf{y}_g = (y_{1g}, \dots, y_{n_g g})'$  is  $n_g \times 1$  vector of dependent variables
- $\mathbf{X}_g = (\mathbf{x}_{1g}, \dots, \mathbf{x}_{n_g g})'$  is  $n_g \times K$  regressor matrix for  $g^{th}$  cluster.
- Linear regression model
  - ▶  $\mathbf{y}_g = \mathbf{X}_g \boldsymbol{\beta} + \mathbf{e}_g$
  - ▶  $E(\mathbf{e}_g | \mathbf{X}_g) = \mathbf{0}$
  - ▶  $E(\mathbf{e}_g \mathbf{e}_g' | \mathbf{X}_g) = \mathbf{S}_g$

# Cluster-Robust (CR) Variance Estimation

- OLS:

$$\hat{\beta} = \left( \sum_{g=1}^G \mathbf{x}'_g \mathbf{x}_g \right)^{-1} \left( \sum_{g=1}^G \mathbf{x}'_g \mathbf{y}_g \right)$$

- Residual:

$$\hat{\mathbf{e}}_g = \mathbf{y}_g - \mathbf{x}_g \hat{\beta}$$

- Variance estimator

$$\hat{V}_0 = \left( \sum_{g=1}^G \mathbf{x}'_g \mathbf{x}_g \right)^{-1} \left( \sum_{g=1}^G \mathbf{x}'_g \hat{\mathbf{e}}_g \hat{\mathbf{e}}'_g \mathbf{x}_g \right) \left( \sum_{g=1}^G \mathbf{x}'_g \mathbf{x}_g \right)^{-1}$$

# Adjustments

- Chris Hansen (2007) adjustment

$$\hat{V} = \left( \frac{G}{G-1} \right) \hat{V}_0$$

Justified in “Large homogenous clusters” framework

- Stata adjustment

$$\hat{V}_1 = \left( \frac{n-1}{n-k} \right) \left( \frac{G}{G-1} \right) \hat{V}_0$$

No justification

# Other covariance matrix estimators

- CRV2

- ▶ Replace OLS residual  $\hat{\mathbf{e}}_g$  with  $\bar{\mathbf{e}}_g = (\mathbf{I} - \mathbf{H}_g)^{-1/2} \hat{\mathbf{e}}_g$
- ▶  $\mathbf{H}_g = \mathbf{X}_g \left( \sum_{g=1}^G \mathbf{X}'_g \mathbf{X}_g \right)^{-1} \mathbf{X}'_g$
- ▶ CRV2 is unbiased under i.i.d. dependence
- ▶ Recommended by Imbens-Kolesar (2016)

- CRV3:

- ▶ Replace  $\hat{\mathbf{e}}_g$  with  $\tilde{\mathbf{e}}_g = (\mathbf{I} - \mathbf{H}_g)^{-1} \hat{\mathbf{e}}_g$
- ▶ **Theorem:** CRV3 conservative under clustered dependence:

$$E \left( \hat{V}_3 \mid X \right) \geq V$$



# Cluster-Robust (CR) Variance Estimation: Some History

- **Methods:** Moulton (1986, 1990), Arellano (1987)
- **Popularization:** Rogers (1993), Bertrand, Duflo and Mullainathan (2004)
- **Large  $G$  asymptotics:** White (1984), C. Hansen (2007), Carter, Schnepel and Steigerwald (2017)
- **Fixed  $G$  asymptotics:** C. Hansen (2007), Bester, Conley and C. Hansen (2011), Conley and Taber (2011), Ibragimov and Mueller (2010, 2016)
- **Small Sample:** Donald and Lang (2007), Imbens and Kolesar (2016), Young (2017), Canay, Romano, and Shaikh (2017)
- **Bootstrap:** Cameron, Gelbach and Miller (2008), MacKinnon and Webb (2017)

# Illustration: Heteroskedastic Dummy Variable Regression

- Dummy variable model
  - ▶ Angrist and Pinchke (2009)
  - ▶ Imbens and Kolesar (2016)
- $y_i = \beta_0 + \beta_1 x_i + e_i$
- $\sum_{i=1}^n x_i = 3$
- $e_i \sim N(0, 1)$
- Coefficient of interest:  $\beta_1$
- Simulation with 100,000 replications

# Large Size Distortion with HC Standard Errors

Rejection Probability of Nominal 5% Tests  
Using  $t_{n-k}$  Critical Values

	$n = 30$
HC0	0.18
HC1	0.17
HC2	0.14
HC3	0.10

Notice that even conservative HC3 t-ratio over-rejects.  
That is because the  $t_{n-k}$  distribution is incorrect.

## Distortion increases with Sample size!

Rejection Probability of Nominal 5% Tests  
Using  $t_{n-k}$  Critical Values

	$n = 30$	$n = 100$	$n = 500$
HC0	0.18	0.23	0.24
HC1	0.17	0.22	0.24
HC2	0.14	0.17	0.18
HC3	0.10	0.13	0.14

**Reason: Highly Leveraged Design Matrix**

# Simulation Results

- All procedures over-reject
- HC1 correction doesn't help
- Unbiased estimator HC2 over-rejects
- Conservative estimator HC3 over-rejects
- $t_{n-k}$  vs  $N(0, 1)$  ineffective
- Conclusion: **Distributional** approximation needs improvement

# Exact Distribution of White t-ratio

**Assumption:** Observations are i.i.d.,  $e_i|x_i \sim N(0, \sigma^2)$

- Step 1: t-ratio is ratio of normal to weighted sum of chi-squares

$$t \sim \frac{Z}{\sqrt{Q}}$$
$$Q = \sum_{i=1}^K w_i Q_i$$

where  $Z \sim N(0, 1)$ ,  $Q_1 \sim \chi_1^2$ , ...,  $Q_K \sim \chi_1^2$

- Step 2: The exact distribution of  $Q$  is a chi-square mixture
- Step 3: The exact distribution of  $t$  is a student t mixture

# Step 1

- $R'(\hat{\beta} - \beta) = \left(\sigma^2 R'(X'X)^{-1}R\right)^{1/2} Z$  where  $Z \sim N(0, 1)$
- $d_i = R'(X'X)^{-1}x_i$ ,  $D = \text{diag}\{d_1^2, \dots, d_n^2\}$ ,  $M = I - X(X'X)^{-1}X'$ ,  
 $B = MDM$ ,  $Q_i$  iid  $\chi_1^2$
- $\lambda_1, \dots, \lambda_K$  are the non-zero eigenvalues of  $B$ .
- Then

$$R'\hat{V}R = \sum_{i=1}^n d_i^2 \hat{e}_i^2 = \hat{e}'D\hat{e} = e'Be = \sigma^2 \sum_{i=1}^K \lambda_i Q_i$$

- Together

$$t = \frac{R'(\hat{\beta} - \beta)}{\sqrt{R'\hat{V}_1R}} = \frac{Z}{\sqrt{\sum_{i=1}^K w_i Q_i}}$$

where  $w_i = \lambda_i / R'(X'X)^{-1}R$

# Ratio of normal to weighted sum of chi-squares

- Under normality

$$t = \frac{R'(\hat{\beta} - \beta)}{\sqrt{R'\hat{V}_1R}} = \frac{Z}{\sqrt{\sum_{i=1}^K w_i Q_i}}$$

- This representation holds for HC0, HC1, HC2, HC3, HC4 heteroskedasticity-robust t-ratios
  - ▶ The weights  $w_i$  depend on the specific estimator
- This representation holds for CRV0, CRV1, CRV2, CRV3 cluster-robust t-ratios
  - ▶ The weights  $w_i$  depend on the specific estimator



## Step 2: Exact Distribution of $Q$

- **Weighted sum of chi-square random variables**

For  $Q_1 \sim \chi_{k_1}^2, \dots, Q_N \sim \chi_{k_N}^2$  mutually independent,  $w_i > 0, k_i > 0$

$$Q = \sum_{i=1}^N w_i Q_i$$

- We write its distribution as

$$G(u | w_1, \dots, w_N; k_1, \dots, k_N) = P(Q \leq u).$$

- Conventional chi-square when  $w_1 = \dots = w_N$
- Distribution function  $G$  unknown
- Classic problem in statistical theory
- Approximation methods dominate
- We now provide the exact distribution

# Theorem 1: Distribution of Q

$$G(u|w_1, \dots, w_N; k_1, \dots, k_N) = \sum_{m=0}^{\infty} b_m G_{K+2m} \left( \frac{u}{\delta} \right)$$

where  $G_r(u)$  is the  $\chi_r^2$  distribution,

$$K = \sum_{i=1}^N k_i$$

$$\delta = \min_m w_m$$

$$b_0 = \prod_{i=1}^N \left( \frac{\delta}{w_i} \right)^{k_i/2}$$

$$b_m = \frac{1}{m} \sum_{\ell=1}^m b_{m-\ell} a_{\ell}, \quad m \geq 1$$

$$a_m = \sum_{i=1}^N \frac{k_i}{2} \left( 1 - \frac{\delta}{w_i} \right)^m$$

# Comments

- Theorem 1 shows that the distribution of  $Q$  can be written as an infinite mixture of chi-square distributions
- The weights are non-negative, sum to one
- Weights are determined by a simple recursion in known parameters
- Theorem 1 is a refinement of Castano and Lopez (2005).
  - ▶ Obtained by inversion of transformed MGF
  - ▶ Uses theory of MVUE of Gamma distributions
  - ▶ Written in terms of Laguerre polynomials
  - ▶ Their result is written as a function of two tuning parameters.
  - ▶ Theorem 1 is obtained as a limiting case (taking the limit as one tuning parameter limits to zero and the other is set at its boundary).
  - ▶ Theorem 1 is a simpler, more convenient, and numerically accurate.

## Step 3: Exact Distribution of t-ratio

- **Generalized T distribution**

For  $Z \sim N(0, 1)$ ,  $Q_1 \sim \chi_{k_1}^2$ , ...,  $Q_N \sim \chi_{k_N}^2$ , mutually independent,  
 $w_i > 0$ ,  $k_i > 0$

$$T = \frac{Z}{\sqrt{\sum_{i=1}^N w_i Q_i}}$$

- We write its distribution as

$$F(u | w_1, \dots, w_N; k_1, \dots, k_N) = P(T \leq u)$$

- If  $k_1 = \dots = k_N = 1$  we write the distribution as  $F(u | w_1, \dots, w_N)$ .
- Conventional student  $t$  when  $w_1 = \dots = w_N$
- Step 1 showed that HC t-ratios are distributed generalized  $T$

## Derivation

The distribution of  $T$  is

$$P(T \leq u) = P(Z \leq \sqrt{Q}u) = E(\Phi(\sqrt{Q}u))$$

Its density is

$$E(\phi(\sqrt{Q}u) \sqrt{Q}) = \int_0^\infty \phi(\sqrt{q}u) \sqrt{q}g(q) dq$$

where  $g$  is the density of  $Q$

Applying Theorem 1, this equals

$$\begin{aligned} & \sum_{m=0}^{\infty} \frac{b_m}{\delta} \int_0^\infty \phi(\sqrt{q}u) \sqrt{q}g_{K+2m}(q/\delta) dq \\ &= \sum_{m=0}^{\infty} b_m (\delta(K+2m))^{1/2} f_{K+2m}\left(u\sqrt{\delta(K+2m)}\right) \end{aligned}$$

where  $f_{K+2m}$  is the student  $t$  density

## Theorem 3: Distribution of $T$

$$F(u|w_1, \dots, w_N; k_1, \dots, k_N) = \sum_{m=0}^{\infty} b_m F_{K+2m} \left( u \sqrt{(K+2m)\delta} \right)$$

where  $F_r$  is the student distribution

### Comments:

- Exact distribution is an infinite mixture of student  $t$  distributions
- Specializes to conventional student  $t$  when  $w_i$  are all equal

## Theorem 4: Alternative expression

$$F(u|w_1, \dots, w_N; k_1, \dots, k_N) \\ = F_K(u\sqrt{K\delta}) + u\sqrt{\delta} \sum_{m=1}^{\infty} b_m^* \frac{f_{K+2m-2}(u\sqrt{(K+2m-2)\delta})}{\sqrt{K+2m-2}}$$

where

$$b_m^* = 1 - \sum_{j=0}^{m-1} b_j$$

### Comments:

- Obtained by applying sequential integration by parts
- Preferable computational form
  - ▶ Only one distribution evaluation

## Theorem 5: Exact Distribution of White t-ratio

$$t \sim F(u | w_1, \dots, w_N)$$

where

- $w_i = \lambda_i / R' (X'X)^{-1} R$
- $\lambda_1, \dots, \lambda_K$  are the non-zero eigenvalues of  $B = D^{1/2} M D^{1/2}$
- $d_i = R' (X'X)^{-1} x_i$
- $D = \text{diag} \{ d_1^2, \dots, d_n^2 \}$
- $M = I - X(X'X)^{-1} X'$



# Finite Sample Distribution

- This is the **exact** finite sample distribution of the White HC t-ratio under normality.
- The distribution is determined by the design matrix  $X'X$
- This is entirely new
- The exact distribution is not student  $t$ . It is a mixture of student  $t$  distributions.
- The difference can be large when the design matrix is highly leveraged.

# Computation Issue 1

- Computation of eigenvalues of  $B = D^{1/2}MD^{1/2}$ 
  - ▶  $n \times n$  matrix
  - ▶ Unreasonable to compute  $B$  for very large  $n$
  - ▶ Eigenvalue calculation reasonable for  $n \leq 1000$ .
    - ★ Unreasonable for  $n \geq 5000$
- Solution for  $n > 1000$ :
  - ▶ Use algorithm which uses function  $a(x) = Bx$  instead of matrix  $B$  itself
  - ▶ Only calculate largest, say  $L = 10$ , eigenvalues
  - ▶ Matlab “eigs” function very fast, even for  $n = 1,000,000$
- When only  $L$  eigenvalues calculated
  - ▶  $\sum_{i=1}^N w_i = \text{tr}(B) = \sum d_i^2 - \text{tr}((X'X)^{-1}(X'DX))$
  - ▶  $\lambda_{L+1}^* = \sum_{i=L+1}^N w_i = \text{tr}(B) - \sum_{i=1}^L w_i$
  - ▶  $w_{L+1}^* = \lambda_{L+1}^* / (n - k - L)$
  - ▶ Approximate  $\sum_{i=1}^N w_i Q_i \simeq \sum_{i=1}^L w_i Q_i + w_{L+1}^* \chi_{n-k-L}^2$

## Computation Issue 2

- Coefficient recursion  $b_m = \frac{1}{m} \sum_{\ell=1}^m b_{m-\ell} a_\ell$
- Fast for  $m \leq 1000$ . Slow for large  $m$
- Convergence when  $\sum_{m=0}^M b_m \simeq 1$ 
  - ▶ Requires large  $M$  when weights are highly unbalanced
- In such cases, we may need to make a computational approximation
  - ▶ Under investigation

# Computation Issue 3

- Distribution function evaluation

- $F_K \left( u\sqrt{K\delta} \right) + u\sqrt{\delta} \sum_{m=1}^{\infty} b_m^* \frac{f_{K+2m-2} \left( u\sqrt{(K+2m-2)\delta} \right)}{\sqrt{K+2m-2}}$

- Computation using this formula is fast

# Exact Distribution

- Advantages
  - ▶ Computable exact distribution under normality
  - ▶ Improved accuracy when regressor matrix is highly leveraged
- Disadvantages
  - ▶ Increased computation cost relative to classical methods
  - ▶ Reliable algorithm in development
- Limitations
  - ▶ Assumes homoskedasticity
  - ▶ Assumes normality
  - ▶ Linear parameters

# Alternative

- Bell-McCaffrey (2002)
  - ▶ Satterthwaite (1946) approximation for  $Q$  is  $\alpha\chi_K^2$  where  $\alpha$  and  $K$  match first two moments of  $Q$
  - ▶ Approximate distribution of  $t$  by  $t_K$
- Endorsed by Imbens-Kolesar (2016)
- An “approximation” but no formal theory

# Simulation Experiment

- Dummy variable model
  - ▶ Angrist and Pischke (2009)
  - ▶ Imbens and Kolesar (2016)
- $y_i = \beta_0 + \beta_1 x_i + e_i$
- $\sum_{i=1}^n x_i = 3$
- Coefficient of interest:  $\beta_1$
- $n = 50, 100, 500$
- Compare:
  - ▶ HC1, HC2, HC3
  - ▶  $t_{n-k}$ , Bell-McCaffrey, and  $T$  distributions
- Size and median length of confidence regions
- $e_i \sim N(0, 1)$ , Heteroskedastic, and student-t errors
- 100,000 replications

# Design Matrix is Highly Leveraged

- $n = 50$ 
  - ▶ HC1 weights  $w_j = \{0.33, 0.33, 0.0013, 0.0013, \dots\}$
  - ▶ HC2 weights  $w_j = \{0.47, 0.47, 0.0013, 0.0013, \dots\}$
  - ▶ HC3 weights  $w_j = \{0.70, 0.70, 0.0013, 0.0013, \dots\}$
- $n = 100$ 
  - ▶ HC1 weights  $w_j = \{0.33, 0.33, 0.0003, 0.0003, \dots\}$
  - ▶ HC2 weights  $w_j = \{0.48, 0.48, 0.0003, 0.0003, \dots\}$
  - ▶ HC3 weights  $w_j = \{0.73, 0.73, 0.0003, 0.0003, \dots\}$
- Highly unequal, contrast increases with sample size
- Due to high leverage



Rejection Probability of Nominal 5% Tests  
 Median Length of 95% Confidence Intervals  
 Normal Homoskedastic Errors

		$t_{n-k}$	Bell-McCaffrey		Exact $T$	
			size	Length	size	Length
$n = 50$	HC1	0.174	0.032	3.5	0.053	3.0
	HC2	0.139	0.033	3.7	0.052	3.2
	HC3	0.101	0.035	3.9	0.052	3.3
$n = 100$	HC1	0.224	0.036	3.9	0.052	3.4
	HC2	0.173	0.040	4.0	0.051	3.6
	HC3	0.126	0.042	4.0	0.051	3.7
$n = 500$	HC1	0.240	0.046	4.1	0.051	3.9
	HC2	0.183	0.047	4.1	0.051	3.9
	HC3	0.137	0.049	4.1	0.051	4.0

Rejection Probability of Nominal 5% Tests  
 Median Length of 95% Confidence Intervals  
 Normal Heteroskedastic Errors  
 $\sigma^2(x) = 1(x = 1) + 0.5(x = 0)$

		$t_{n-k}$	Bell-McCaffrey		$T$	
			size	Length	size	Length
$n = 50$	HC1	0.201	0.053	3.3	0.079	2.8
	HC2	0.158	0.049	3.6	0.072	3.0
	HC3	0.115	0.046	3.8	0.065	3.3
$n = 100$	HC1	0.228	0.051	3.8	0.065	3.4
	HC2	0.175	0.050	4.0	0.061	3.6
	HC3	0.128	0.049	4.0	0.058	3.7
$n = 500$	HC1	0.259	0.052	4.0	0.057	3.8
	HC2	0.197	0.052	4.0	0.055	3.9
	HC3	0.144	0.052	4.0	0.054	3.9

Rejection Probability of Nominal 5% Tests  
 Median Length of 95% Confidence Intervals  
 Normal Heteroskedastic Errors  
 $\sigma^2(x) = 1(x = 1) + 2(x = 0)$

		$t_{n-k}$	Bell-McCaffrey		$T$	
			size	Length	size	Length
$n = 50$	HC1	0.112	0.003	4.5	0.017	3.8
	HC2	0.093	0.009	4.5	0.021	3.8
	HC3	0.068	0.013	4.5	0.024	3.8
$n = 100$	HC1	0.182	0.012	4.3	0.021	3.8
	HC2	0.140	0.017	4.3	0.027	3.9
	HC3	0.106	0.025	4.3	0.033	3.9
$n = 500$	HC1	0.231	0.034	4.2	0.039	4.0
	HC2	0.177	0.039	4.2	0.042	4.0
	HC3	0.132	0.042	4.2	0.044	4.1

Rejection Probability of Nominal 5% Tests  
 Median Length of 95% Confidence Intervals  
 $t_5$  Errors

		$t_{n-k}$	Bell-McCaffrey		$T$	
			size	Length	size	Length
$n = 50$	HC1	0.153	0.022	4.2	0.039	3.6
	HC2	0.122	0.023	4.4	0.039	3.7
	HC3	0.086	0.024	4.5	0.039	3.9
$n = 100$	HC1	0.182	0.012	4.6	0.038	4.0
	HC2	0.140	0.017	4.6	0.039	4.2
	HC3	0.106	0.025	4.7	0.040	4.3
$n = 500$	HC1	0.226	0.035	4.7	0.038	4.5
	HC2	0.166	0.036	4.7	0.039	4.5
	HC3	0.119	0.037	4.7	0.040	4.6

# Expanded Dummy Variable Design

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- $k = 5$
- Each dummy variable only equals 1 for 3 observations
- Each dummy variable overlaps with first regressor

## Normal Homoskedastic Errors

		$t_{n-k}$	Bell-McCaffrey		Exact $T$	
			size	Length	size	Length
$n = 50$	HC1	0.188	0.044	3.6	0.052	3.5
	HC2	0.113	0.042	3.7	0.051	3.5
	HC3	0.047	0.039	3.9	0.051	3.6
$n = 100$	HC1	0.219	0.044	3.6	0.051	3.5
	HC2	0.118	0.042	3.7	0.050	3.5
	HC3	0.151	0.040	3.9	0.050	3.5
$n = 500$	HC1	0.234	0.044	3.6	0.050	3.5
	HC2	0.125	0.042	3.8	0.050	3.5
	HC3	0.053	0.040	3.9	0.051	3.6

## Normal Heteroskedastic Errors

$$\sigma^2(x) = 1(x = 1) + 0.5(x = 0)$$

		$t_{n-k}$	Bell-McCaffrey		Exact $T$	
			size	Length	size	Length
$n = 50$	HC1	0.255	0.085	2.6	0.092	2.5
	HC2	0.166	0.076	2.8	0.088	2.6
	HC3	0.079	0.069	3.0	0.084	2.7
$n = 100$	HC1	0.288	0.082	2.6	0.091	2.5
	HC2	0.174	0.074	2.8	0.087	2.6
	HC3	0.084	0.068	3.0	0.082	2.8
$n = 500$	HC1	0.312	0.087	2.6	0.098	2.5
	HC2	0.187	0.080	2.8	0.092	2.6
	HC3	0.092	0.073	3.0	0.089	2.7

## Normal Heteroskedastic Errors

$$\sigma^2(x) = 1(x = 1) + 2(x = 0)$$

		$t_{n-k}$	Bell-McCaffrey		Exact $T$	
			size	Length	size	Length
$n = 50$	HC1	0.126	0.023	6.0	0.027	5.8
	HC2	0.072	0.023	6.1	0.028	5.7
	HC3	0.026	0.021	6.3	0.029	5.8
$n = 100$	HC1	0.152	0.024	6.1	0.027	5.8
	HC2	0.077	0.023	6.2	0.028	5.8
	HC3	0.030	0.022	6.3	0.030	5.8
$n = 500$	HC1	0.172	0.022	6.0	0.026	5.8
	HC2	0.079	0.021	6.2	0.028	5.8
	HC3	0.030	0.021	6.3	0.028	5.8



$t_5$  Errors

		$t_{n-k}$	Bell-McCaffrey		Exact $T$	
			size	Length	size	Length
$n = 50$	HC1	0.172	0.037	4.4	0.043	4.2
	HC2	0.099	0.035	4.6	0.042	4.3
	HC3	0.039	0.032	4.7	0.042	4.3
$n = 100$	HC1	0.226	0.040	4.4	0.046	4.2
	HC2	0.115	0.037	4.6	0.045	4.3
	HC3	0.046	0.035	4.8	0.045	4.4
$n = 500$	HC1	0.227	0.038	4.4	0.044	4.2
	HC2	0.115	0.035	4.5	0.043	4.3
	HC3	0.046	0.033	4.7	0.044	4.3

# Continuous Design

- $X \sim \log \text{Normal}$ , otherwise similar
  - ▶ Also creates highly leveraged samples
- Results very similar

# Simulation Summary

- $t_{n-k}$  criticals inappropriate
- Bell-McCaffrey can be quite conservative
- $T$  is precise under homoskedastic normality (as expected)
- Both Bell-McCaffrey and  $T$  sensitive to heteroskedasticity and non-normality
- HC3 appears least sensitive
- HC3 with  $T$  distribution reasonably reliable

# Clustered Samples

- Same analysis applies to clustered regression and CR standard errors
- Under i.i.d. normality, clustered t-ratios have exact  $T$  distributions
- Weights are determined by regressor matrix
- Distortions from normality when design matrix is highly leveraged
  - ▶ When clusters are heterogeneous
  - ▶ When only a few clusters are “treated”
- Accuracy of conventional distribution theory depends on the number of clusters  $G$  and degree of leverage
  - ▶ Conventional asymptotics requires a large  $G$ , not large  $n$
  - ▶ Many applied papers don't even report  $G$
  - ▶  $G$  should be reported, along with sample size!

# Conclusion

- In 1908, Gosset revolutionized statistical inference by providing the exact distribution of the classical t-ratio
- Applied econometrics relies on heteroskedasticity-robust and cluster-robust standard errors
- There is no finite sample theory for HC and CR t-ratios
- This paper provides the first exact distribution theory

# Findings

- HC and CR t-ratios are NOT student  $t_{n-k}$
- The deviation from  $t_{n-k}$  can be very substantial
- The exact distribution (under iid normality) is generalized  $T$
- Exact distribution depends on regressor matrix  $X$
- Correct finite sample p-values and confidence intervals can be reported