LECTURE 14: HC VARIANCE ESTIMATION

Ivan A. Canay
Northwestern University
**Past & Future**

**PART II: TOPICS**
- Non-parametric Regression
- RDD and Matching
- CART and Random Forest
- Binary Choice
- LASSO

**PART III: INFERENCE**
- HC Standard Errors
- HAC Standard Errors
- CR Standard Errors
- Bootstrap & Subsampling
- Randomization Tests
**Linear Model Setup**

- Let \((Y, X, U)\) be such that \(Y\) and \(U\) take values in \(\mathbb{R}\) and \(X\) takes values in \(\mathbb{R}^{k+1}\).

- The first component of \(X\) is a constant equal to one.

- Let \(\beta \in \mathbb{R}^{k+1}\) be such that
  \[
  Y = X' \beta + U.  
  \] (1)

  Suppose that 1. \(E[XU] = 0\), 2. that there is no perfect collinearity in \(X\), that 3. \(E[XX'] < \infty\), and that 4. \(\text{Var}[XU] < \infty\).

- Let \(P\) be the distribution of \((Y, X)\) and let \((Y_1, X_1), \ldots, (Y_n, X_n)\) be an i.i.d. sample from \(P\).

- Under these assumptions, we established the **asymptotic normality** of the OLS estimator, \(\hat{\beta}_n\):
  \[
  \sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \mathbb{V})
  \]
  for
  \[
  \mathbb{V} = E[XX']^{-1}E[XX'U^2]E[XX']^{-1}. 
  \]
**Testing Problem**

- We wish to test
  
  \[ H_0 : \beta \in \mathbf{B}_0 \quad \text{versus} \quad H_1 : \beta \in \mathbf{B}_1 \]

  where \( \mathbf{B}_0 \) and \( \mathbf{B}_1 \) form a partition of \( \mathbb{R}^{k+1} \). Particular attention to hypotheses for one component of \( \beta \).

- **WLOG**: assume we are interested in the first slope component of \( \beta \) so that,
  
  \[ H_0 : \beta_1 = c \quad \text{versus} \quad H_1 : \beta_1 \neq c \]

  (2)

  The CMT implies that
  
  \[ \sqrt{n}(\hat{\beta}_{1,n} - \beta_1) \xrightarrow{d} N(0, V_1) \]

  as \( n \to \infty \) where \( V_1 = \mathbf{V}_{[2,2]} \) is the element of \( \mathbf{V} \) corresponding to \( \beta_1 \).

- A natural choice of test statistic for this problem is the **absolute value of the t-statistic**, 
  
  \[ t_{\text{stat}} = \frac{\sqrt{n}(\hat{\beta}_{1,n} - c)}{\sqrt{\hat{V}_{1,n}}} , \]

  so that \( T_n = |t_{\text{stat}}| \). Required: a consistent estimator \( \hat{\mathbf{V}}_n \) of the limiting variance \( \mathbf{V} \).
Part III of this course covers consistent estimators of $V$ under different assumptions on the dependence and heterogeneity in the data.

We will, however, start with the usual i.i.d. setting, where one of such estimators is

$$
\hat{V}_n = \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X_i' \hat{U}_i^2 \right) \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X_i' \right)^{-1},
$$

where

$$
\hat{U}_i = Y_i - X_i' \hat{\beta}_n.
$$

This is the most widely used form of the robust, heteroskedasticity-consistent standard errors and it is associated with the work of White (1980) (see also Eicker, 1967; Huber, 1967).

We will refer to these as robust EHW (or HC) standard errors.
Consistency of HC standard errors

- **Wish to prove:** $\hat{V}_n \xrightarrow{P} V$.

- **Main difficulty:** showing that

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_iX'_i \hat{U}_i^2 \xrightarrow{P} \text{Var}[XU] \quad \text{as} \quad n \to \infty.$$

- **Note that**

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_iX'_i \hat{U}_i^2 = \frac{1}{n} \sum_{1 \leq i \leq n} X_iX'_i U_i^2 + \frac{1}{n} \sum_{1 \leq i \leq n} X_iX'_i(\hat{U}_i^2 - U_i^2).$$

- **First term:** under the assumption that $\text{Var}[XU] < \infty$, it converges in probability to $\text{Var}[XU]$.

- **Second term:** we wish to show it converges in probability to zero.
Step 1: We argue this separately for each of the \((k + 1)^2\) terms in

\[
\frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i (\hat{U}_i^2 - U_i^2).
\]
Step 2: Intermediate lemma needed to show $\max_{1 \leq i \leq n} |\hat{U}_i^2 - U_i^2| = o_P(1)$.

**Lemma**

Let $Z_1, \ldots, Z_n$ be an i.i.d. sequence of random vectors such that $E[|Z_i|^r] < \infty$. Then

$$\max_{1 \leq i \leq n} |Z_i| = o_P\left(n^{\frac{1}{r}}\right) \quad \text{i.e.} \quad n^{-\frac{1}{r}} \max_{1 \leq i \leq n} |Z_i| \xrightarrow{P} 0.$$  

**Proof:**
**Proof**

**Step 3:** Show that \( \max_{1 \leq i \leq n} |\hat{U}_i^2 - U_i^2| = o_P(1) \) using \( E[|X|^2] < \infty \) and \( E[|UX|^2] < \infty \).
We just proved that
\[
\frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i \hat{U}_i^2 \overset{P}{\to} E[X_i X'_i U_i^2] \]

We also know that
\[
\frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i \overset{P}{\to} E[XX'] .
\]

By the CMT it then follows that
\[
\hat{V}_n = \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i \right)^{-1} \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i \hat{U}_i^2 \right) \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_i X'_i \right)^{-1} \overset{P}{\to} V .
\]
Back to the t-test

- Let $\hat{V}_{1,n}$ denote the $(2, 2)$-diagonal element of $\hat{\nabla}_n$ - i.e., the entry corresponding to $\beta_1$.

- The test that rejects $H_0 : \beta_1 = c$ when

$$T_n = |t_{\text{stat}}| = \left| \frac{\sqrt{n}(\hat{\beta}_{1,n} - c)}{\sqrt{\hat{V}_{1,n}}} \right|$$

exceeds $z_{1-\frac{\alpha}{2}}$, is **consistent in levels**.

- **Duality**: between hypothesis testing and the construction of confidence regions leads to

$$C_n = \left\{ c \in \mathbb{R} : \frac{\sqrt{n}(\hat{\beta}_{1,n} - c)}{\sqrt{\hat{V}_{1,n}}} \leq z_{1-\frac{\alpha}{2}} \right\} = \left\{ \hat{\beta}_{1,n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{V}_{1,n}}{n}}, \hat{\beta}_{1,n} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{V}_{1,n}}{n}} \right\}.$$ 

This confidence region satisfies

$$P[\beta_1 \in C_n] \to 1 - \alpha \quad n \to \infty.$$
QUESTIONS?
**Finite Sample Performance**

- **Stata**: does not compute $\hat{V}_n$ in the default “robust” option.

- It includes a finite sample adjustment to inflate the estimated residuals (known to be too small in finite samples).

- **HC1**: This version of the HC estimator is commonly known as HC1 and given by

$$
\hat{V}_{hc1,n} = \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_iX_i' \right)^{-1} \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_iX_i' \hat{U}_i^* \right) \left( \frac{1}{n} \sum_{1 \leq i \leq n} X_iX_i' \right)^{-1},
$$

where

$$
\hat{U}_i^* = \frac{n}{n - k - 1} \hat{U}_i^2.
$$

- **Obvious Result**: this estimator is also consistent for $V$ and are the ones used to compute “robust” confidence intervals in Stata.
An alternative to $\hat{V}_n$ and $\hat{V}_{hc1,n}$ is what MacKinnon and White (1985) call the HC2 variance estimator, here denoted by $\hat{V}_{hc2,n}$.

In order to define this estimator, we need additional notation. Let

$$P = X(X'X)^{-1}X'$$

be the $n \times n$ projection matrix, with $i$-th column denoted by

$$P_i = X(X'X)^{-1}X_i$$

and $(i, i)$-th element denoted by

$$P_{ii} = X'_i(X'X)^{-1}X_i.$$

Let $\Omega$ be the $n \times n$ diagonal matrix with $i$-th diagonal element equal to $\sigma^2(X_i) = \text{Var}[U_i|X_i]$.

Let $e_{n,i}$ be the $n$-vector with $i$-th element equal to one and all other elements equal to zero.

Let $I$ be the $n \times n$ identity matrix and $M = I - P$ be the residual maker matrix.
Residuals: $\hat{U}_i = Y_i - X_i^T \hat{\beta}_n$ can be written as
\[
\hat{U}_i = e_{n,i}^T MU, \quad \text{or, in vector form,} \quad \hat{U} = MU .
\]

The (conditional) expected value of the square of the residual is
\[
E[\hat{U}_i^2 | X_1, \ldots, X_n] = E[(e_{n,i}^T MU)^2 | X_1, \ldots, X_n]
= (e_{n,i} - P_i)^T \Omega (e_{n,i} - P_i).
\]

If we further assume homoskedasticity (i.e., $\text{Var}[U | X] = \sigma^2$), the last expression reduces to
\[
E[\hat{U}_i^2 | X_1, \ldots, X_n] = \sigma^2 (1 - P_{ii}),
\]
by exploiting that $P$ is an idempotent matrix.

Take away: even when the error term $U$ is homoskedastic, the LS residual $\hat{U}$ is heteroskedastic (due to the presence of $P_{ii}$).

Downward Bias: Since it can be shown that $\frac{1}{n} \leq P_{ii} \leq 1$, it follows that $\text{Var}[\hat{U}_i]$ underestimates $\sigma^2$ under homoskedasticity.
Natural Correction: it makes sense to consider
\[ \tilde{U}_i^2 \equiv \frac{\hat{U}_i^2}{1 - P_{ii}}, \]
as the squared residual to use in variance estimation.

It follows that \( \tilde{U}_i^2 \) is unbiased for \( E[U_i^2 | X_1, \ldots, X_n] \) under homoskedasticity.

HC2: this is the motivation for the variance estimator known as HC2,
\[ \hat{V}_{hc2,n} = \left( \frac{1}{n} \sum_{i=1}^{n} X_iX'_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} X_iX'_i \tilde{U}_i^2 \right) \left( \frac{1}{n} \sum_{i=1}^{n} X_iX'_i \right)^{-1}. \]

Under heteroskedasticity this estimator is unbiased only in some simple examples (e.g., The Behrens-Fisher problem), but it is biased in general.

However, it is expected to have lower bias relative to HC/HC1 - a statement supported by simulations.
There are other finite sample adjustments that give place to HC3, HC4, and even HC5.....

**HC3** is equivalent to HC2 with

\[ U_{i}^{*2} \equiv \frac{\hat{U}_{i}^{2}}{(1 - P_{ii})^{2}}, \]

replacing \( \hat{U}_{i}^{2} \), and its justification is related to the Jackknife estimator of the variance of \( \hat{\beta}_n \).

We will not consider these in class as these adjustments do not deliver noticeable additional benefits relative to HC2 (at least for the purpose of this class).

It is worth noting that HC2 and HC3 are available as an option in Stata.
QUESTIONS?
The Behrens-Fisher Problem

- **Behrens-Fisher**: compare means of two normals when variances are unknown:

\[
Y(0) \sim \mathcal{N}(\mu_0, \sigma^2(0)) \quad \text{and} \quad Y(1) \sim \mathcal{N}(\mu_1, \sigma^2(1)).
\]  

(4)

- Special case of linear regression with a binary regressor, i.e. \(X = (1, D)\) and \(D \in \{0, 1\}\).

- The coefficient on \(D\) identifies the average treatment effect: \(\mu_1 - \mu_0\).

- To be specific, consider the linear model

\[
Y = \beta_0 + \beta_1 D + U \quad \text{and} \quad Y = Y(1)D + (1 - D)Y(0)
\]

with \(U|D\) assumed to be normally distributed with zero conditional mean and

\[
\text{Var}[U|D = d] = \sigma^2(d) \quad \text{for } d \in \{0, 1\}.
\]

- We are interested in

\[
\beta_1 = \frac{\text{Cov}(Y, D)}{\text{Var}(D)} = E[Y|D = 1] - E[Y|D = 0],
\]

which can be estimated as

\[
\hat{\beta}_{1,n} = \bar{Y}_1 - \bar{Y}_0 \quad \text{where} \quad \bar{Y}_d = \frac{1}{n_d} \sum_{i=1}^{n} Y_i I[D_i = d] \quad \text{and} \quad n_d = \sum_{i=1}^{n} I[D_i = d].
\]
The Behrens-Fisher Problem

Conditional on $D^{(n)} = (D_1, \ldots, D_n)$, the exact finite sample variance of $\hat{\beta}_{1,n}$ is

$$V_1^* = \text{Var} \left( \hat{\beta}_{1,n} \mid D^{(n)} \right) = \frac{\sigma^2(0)}{n_0} + \frac{\sigma^2(1)}{n_1},$$

so that, under normality, it follows that

$$\hat{\beta}_{1,n} \mid D^{(n)} \sim N \left( \beta_1, \frac{\sigma^2(0)}{n_0} + \frac{\sigma^2(1)}{n_1} \right).$$

Question: is there a $\kappa \in \mathbb{R}$ such that for some estimator $\hat{V}^*_{1,n}$ we get

$$\frac{\hat{\beta}_{1,n} - \beta_1}{\sqrt{\hat{V}^*_{1,n}}} \sim t(\kappa),$$

(5)

where $t(\kappa)$ denotes a $t$-distribution with $\kappa$ degrees of freedom (dof)?

**Word on Notation**

Today we talk about the “actual” conditional variance of $\hat{\beta}_{1,n}$ as opposed to the asymptotic variance. Thus, the estimator $\hat{V}^*_{1,n}$ above is an estimator of such variance (also explains why there is no $\sqrt{n}$ in (5)). Of course, if $\hat{V}_{1,n}$ is a consistent estimator of the asymptotic variance of $\hat{\beta}_{1,n}$, then $\hat{V}^*_{1,n} = \frac{1}{n} \hat{V}_{1,n}$ is an estimator of the variance of $\hat{\beta}_{1,n}$. We use $*$ to denote finite sample variances.
Assumption: $\sigma^2 = \sigma^2(0) = \sigma^2(1)$ so that the exact conditional variance of $\hat{\beta}_{1,n}$ is

$$V^*_1 = \sigma^2 \left( \frac{1}{n_0} + \frac{1}{n_1} \right).$$

We can estimate $\sigma^2$ by

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - X'_i \hat{\beta}_n)^2$$

and let

$$\hat{V}^*_{1,ho} = \hat{\sigma}^2 \left( \frac{1}{n_0} + \frac{1}{n_1} \right),$$

be the estimator of $V^*_1$. This estimator has two important features.

(A) **Unbiased.** Since $\hat{\sigma}^2$ is unbiased for $\sigma^2$, it follows that $\hat{V}^*_{1,ho}$ is unbiased for the true variance $V^*_1$.

(B) **Chi-square.** Under normality of $U$ given $D$, the scaled distribution of $\hat{V}^*_{1,ho}$ is chi-square with $n - 2$ dof,

$$(n - 2) \frac{\hat{V}^*_{1,ho}}{V^*_1} \sim \chi^2(n - 2).$$

Under normality, the $t$-stat has an exact $t$-distribution under the null

$$t_{ho} = \frac{\hat{\beta}_{1,n} - c}{\sqrt{\hat{V}^*_{1,ho}}} \sim t(n - 2).$$  (6)
The robust EHW variance estimator

- **BF Example:** the component of the EHW variance estimator \( \frac{1}{n} \hat{\sigma}^2_n \) corresponding to \( \beta_1 \) simplifies to

\[
\hat{V}_{1, hc}^* = \frac{\hat{\sigma}^2(0)}{n_0} + \frac{\hat{\sigma}^2(1)}{n_1}
\]

where \( \hat{\sigma}^2(d) = \frac{1}{n_d} \sum_{i=1}^{n} (Y_i - \bar{Y}_d)^2 I\{D_i = d\} \) for \( d \in \{0, 1\} \).

- **No assumptions** under which there exists a value of \( \kappa \) such that (5) holds, even when \( U \) is normally distributed conditional on \( D \).

- In small samples: \( \hat{V}_{1, hc}^* \) is **biased downward**, i.e.,

\[
E[\hat{V}_{1, hc}^*] = \frac{n_0 - 1}{n_0} \frac{\sigma^2(0)}{n_0} + \frac{n_1 - 1}{n_1} \frac{\sigma^2(1)}{n_1} < V_{1}^*,
\]

and confidence intervals based off these have **coverage substantially below** \( 1 - \alpha \).

- **Ad-hoc correction:** A common “correction” is to replace \( z_{1 - \frac{\alpha}{2}} \) with \( t_{n-2, 1 - \frac{\alpha}{2}} \) - the quantile of a t-distribution with \( n - 2 \) dof.

- Such a correction is often **ineffective**.
**Alternative:** the HC2 variance estimator, here denoted by \( \frac{1}{n} \hat{V}_{hc2,n} \).

This estimator is unbiased under homoskedasticity but, in general, it removes only part of the bias under heteroskedasticity.

**BF problem:** in this case the HC2 correction removes the entire bias.

Its form in this case is

\[
\hat{V}_{1, hc2}^* = \frac{\tilde{\sigma}^2(0)}{n_0} + \frac{\tilde{\sigma}^2(1)}{n_1},
\]

where

\[
\tilde{\sigma}^2(d) = \frac{1}{n_d - 1} \sum_{i=1}^{n} (Y_i - \bar{Y}_d)^2 I\{D_i = d\}.
\]

These conditional variance estimators differ from \( \hat{\sigma}^2(d) \) by a factor \( n_d / (n_d - 1) \).

The estimator \( \hat{V}_{1, hc2}^* \) is unbiased for \( V_1^* \), but it does not satisfy the chi-square property in (b) above. As a result, the associated confidence interval based off a normal critical value is still not exact.

**No assumptions** under which there exists a value of \( \kappa \) such that (5) holds, even when \( U \) is normally distributed conditional on \( D \). In fact, in small samples these standard errors do not work very well.
Simulations

- **Simple simulation.** From Imbens and Kolesar (2016) and MHE:

\[ U_i|D_i \sim N(0, \sigma^2(D_i)) , \]

with \( n_1 = 3, n_0 = 27, \sigma^2(1) = 1, \sigma^2(0) \in \{0, 1, 2\}, \) and \( 1 - \alpha = 0.95. \)

<table>
<thead>
<tr>
<th></th>
<th>dof</th>
<th>( \sigma^2(0) = 0 )</th>
<th>( \sigma^2(0) = 1 )</th>
<th>( \sigma^2(0) = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\vartheta}_{ho} )</td>
<td>( n - 2 )</td>
<td>72.5</td>
<td>94.0</td>
<td>99.8</td>
</tr>
<tr>
<td></td>
<td>( \infty )</td>
<td>74.5</td>
<td>95.0</td>
<td>99.8</td>
</tr>
<tr>
<td>( \hat{\vartheta}_{1, hc} )</td>
<td>( n - 2 )</td>
<td>76.8</td>
<td>80.5</td>
<td>86.6</td>
</tr>
<tr>
<td></td>
<td>( \infty )</td>
<td>78.3</td>
<td>82.0</td>
<td>88.1</td>
</tr>
<tr>
<td>( \hat{\vartheta}_{1, hc2} )</td>
<td>( n - 2 )</td>
<td>82.5</td>
<td>85.2</td>
<td>89.8</td>
</tr>
<tr>
<td></td>
<td>( \infty )</td>
<td>83.8</td>
<td>86.5</td>
<td>91.0</td>
</tr>
</tbody>
</table>

**Table:** Angrist-Pischke design. \( n_1 = 3, n_0 = 27. \)

- **DOF:** \( n - 2 \) may be a **poor choice** for dof. Suppose \( n_1 = 3 \) and \( n_0 = 1,000,000. \) Here \( E[Y_i|D_i = 0] \) is precisely estimated with variance \( \sigma^2(0)/n_0 \approx 0. \) Heuristically then,

\[
    t_{\text{stat}} \approx \frac{\bar{Y}_1 - E[Y_i|D_i = 1]}{\sqrt{\hat{\sigma}^2(1)/n_1}}.
\]

Under normality this has an exact \( t \)-distribution with dof equal to \( n_1 - 1 = 2 << n - 2 \approx \infty. \)
One of the most attractive proposals for the Behrens-Fisher problem is due to Welch (1951).

Welch suggests approximating the distribution of the $t$-statistic based on HC2 by a $t$-distribution.

Suggests using moments of the variance estimator $\frac{1}{n} \hat{V}_{hc2}$ to determine the most appropriate value for the degrees of freedom.

**Idea:** suppose (Assumption 1) there was a constant $\kappa$ such that

$$\kappa \frac{\hat{V}_{1, hc2}^*}{V_1^*} \sim \chi^2(\kappa) .$$

Recall that the mean and variance of a $\chi^2(\kappa)$ are $\kappa$ and $2\kappa$.

**Welch:** find $\kappa$ by matching the first two moments of a chi-square distribution. This is, find $\kappa$ such that

$$E \left[ \kappa \frac{\hat{V}_{1, hc2}^*}{V_1^*} \right] = \kappa \quad \text{and} \quad \text{Var} \left[ \kappa \frac{\hat{V}_{1, hc2}^*}{V_1^*} \right] = 2\kappa . \quad (7)$$

The first equality automatically holds if $E[\hat{V}_{1, hc2}^*] = V_1^*$ so the value of $\kappa$ is determined by the second equality if we assume (Assumption 2) $\hat{V}_{1, hc2}^*$ is unbiased.
To find the variance, Welch assumes (Assumption 3) normality.

Under normality we obtain that

$$
\hat{V}_{1, hc2}^* = \frac{\sigma^2(0)}{n_0(n_0 - 1)} \frac{(n_0 - 1)\bar{\sigma}^2(0)}{\sigma^2(0)} + \frac{\sigma^2(1)}{n_1(n_1 - 1)} \frac{(n_1 - 1)\bar{\sigma}^2(1)}{\sigma^2(1)},
$$

is a linear combination of two chi-squared random variables,

$$
\frac{(n_0 - 1)\bar{\sigma}^2(0)}{\sigma^2(0)} \sim \chi^2(n_0 - 1) \quad \text{and} \quad \frac{(n_1 - 1)\bar{\sigma}^2(1)}{\sigma^2(1)} \sim \chi^2(n_1 - 1),
$$

where $\bar{\sigma}^2(0)$ and $\bar{\sigma}^2(1)$ are independent of each other and of $(\hat{\beta}_{1,n} - c)$. It follows that,

$$
\text{Var}[\hat{V}_{1, hc2}^*] = \frac{2\sigma^4(0)}{(n_0 - 1)n_0^2} + \frac{2\sigma^4(1)}{(n_1 - 1)n_1^2}.
$$

**Welch’s DoF**

$$
\kappa_W = \frac{2V_{1}^{*2}}{\text{Var}[\hat{V}_{1, hc2}^*]} = \frac{2 \left( \frac{\sigma^2(0)}{n_0} + \frac{\sigma^2(1)}{n_1} \right)^2}{\frac{2\sigma^4(0)}{(n_0 - 1)n_0^2} + \frac{2\sigma^4(1)}{(n_1 - 1)n_1^2}}.
$$
**Extra: Test with DoF adjustments**

- **Simplification**: A slightly different degrees of freedom adjustment arises if we further assume (Assumption 4) homoskedasticity at the time of computing $\kappa$.

- $\kappa_w$ then simplifies to

$$\kappa_{bm} = \frac{2\left(\frac{\sigma^2}{n_0} + \frac{\sigma^2}{n_1}\right)^2}{\frac{2\sigma^4}{(n_0-1)n_0^2} + \frac{2\sigma^4}{(n_1-1)n_1^2}} = \frac{(n_0 + n_1)^2(n_0 - 1)(n_1 - 1)}{n_1^2(n_1 - 1) + n_0^2(n_0 - 1)}.$$  

- The associated $1 - \alpha$ confidence interval is now

$$CS_{bm}^{1-\alpha} = \left\{ \hat{\beta}_{1,n} - t^{\kappa_{bm}}_{1-\frac{\alpha}{2}} \sqrt{\hat{V}^*_{1, hc2}}, \hat{\beta}_{1,n} + t^{\kappa_{bm}}_{1-\frac{\alpha}{2}} \sqrt{\hat{V}^*_{1, hc2}} \right\}.$$  

- **Intuition**: note that

$$\kappa_{bm} \to \begin{cases} n_1 - 1 & \text{if } n_0 \to \infty, n_1 \text{ fixed} \\ n_0 - 1 & \text{if } n_1 \to \infty, n_0 \text{ fixed} \\ n - 2 & \text{if } n_0 = n_1 = \frac{n}{2} \end{cases},$$

so the DoF adapt to the example in our previous table.

- For further details, see Imbens and Kolesar (2016).