Inference with Few Heterogeneous Clusters

Rustam Ibragimov — Imperial College

Ulrich K. Müller — Princeton University

November 2014

Introduction

- Most economic data is observational

  ⇒ data is plausibly not i.i.d.: countries, firms, or time series data

- Necessitates appropriate correction for standard errors
  - Newey and West (1987), Andrews (1991), etc. for time series disturbances
  - Rogers (1993) and Arellano (1987) for clustered and panel data
  - Conley (1999) for spatially correlated data

- Corrections based on law of large numbers

  ⇒ Poor small sample properties in many instances of interest

Inconsistent Variance Estimators

- Part of the problem is that sample variability of “consistent” variance estimators is neglected

- Approaches that account for sample variability of variance estimator:
  - time series: Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2002, 2005), Gonçalves and Vogelsang (2006), Müller (2007), Sun, Phillips and Jin (2008), Sun (2013)
  - panel data: Donald and Lang (2004), Hansen (2007)

- Add to this literature and suggest a general approach to robust inference when little is known about correlations and heterogeneity

Ibragimov and Müller (JBES, 2010)

- Assume data can be classified in a finite number q of groups that allow asymptotically independent normal inference about the (scalar) parameter of interest β, so that $\hat{\beta}_j \stackrel{a}{\sim} id\mathcal{N}(\beta, \omega^2)$ for $j = 1, \ldots, q$. Example: In a cross country regression, divide world into $q = 6$ regions, and estimate the model in each region.

- Treat $\hat{\beta}_j$, $j = 1, \ldots, q$ as q observations, and reject a 5% level test if t-statistic larger than usual critical value for $q - 1$ degrees of freedom.

- Potential concern: requires q groups to be homogeneous in terms of the precision of the independent information about β. Not plausible in regions example.

- Then $\hat{\beta}_j \stackrel{a}{\sim} id\mathcal{N}(\beta, \omega_j^2)$, with $\omega_j^2 \neq \omega_i^2$ for $i \neq j$. Properties of small sample t-test under heterogeneous variances?

Small Sample Result

- **Theorem (Bakirov and Székely, 2005):** If one applies the usual small sample t-test to independent Gaussian observations of possible heterogenous variance, then two-sided tests of level 5% or lower are conservative (i.e. rejection probability under the null hypothesis becomes smaller for unequal variances).

- Result not true at the 10% level for $q > 14$ observations

- Proof requires many clever steps, most of them in Bakirov (1989). Very special to t-statistic.

Large Sample t-statistic Based Inference

- Asymptotically valid 5% test of $H_0 : \beta = \beta_0$ under $n(\hat{\beta}_j - \beta) \Rightarrow id\mathcal{N}(0, \sigma_j^2)$ obtained by rejecting if $|t^{IM}| > \text{cv}(0.05, q-1)$, where

$$t^{IM} = \sqrt{q}\frac{\overline{\hat{\beta}} - \beta_0}{S},$$

  $\overline{\hat{\beta}} = q^{-1}\sum_{j=1}^{q}\hat{\beta}_j$, $S^2 = (q-1)^{-1}\sum_{j=1}^{q}(\hat{\beta}_j - \overline{\hat{\beta}})^2$ and $\text{cv}(\alpha, m)$ is $1 - \alpha/2$ quantile of student$-t_m$ distribution.

- Works with $q \geq 2$ clusters or groups

  $\Rightarrow$ At the cost of lower power when q is small

  $\Rightarrow$ But validity of bootstrap (for instance, Cameron, Gelbach and Miller (2008)) and consistent estimators require $q \to \infty$

- Allows for heterogeneity across groups

  $\Rightarrow$ in contrast to Bester, Conley and Hansen (2011)

Application to Spatial Correlations

- Obstfeld, Shambaugh and Taylor (2010) use study determinants of central bank reserve holdings with a cross country regression involving an unbalanced panel of 26 years and 134 countries, for a total of 2671 observations
- OST focus on four variables: "financial openness", dummies for a "Peg" or "Soft Peg", and "ln(M2/GDP)"
- OST assess the significance of these four variables in a horserace against other factors, clustering standard errors by country

Obstfeld, Shambaugh and Taylor (2010)

Region	fin. o. $\hat{\beta}_j$	Peg $\hat{\beta}_j$	Soft Peg $\hat{\beta}_j$	ln(M2/GDP) $\hat{\beta}_j$
Asia/Pacific	1.110	0.035	-0.060	0.627
W Europe	0.805	0.089	0.069	1.041
E. Europe	0.423	0.317	0.281	0.633
Africa	0.508	0.413	0.318	-0.019
Middle East	1.665	-0.236	-0.056	0.511
S. America	0.770	-0.279	-0.067	-0.201
p-values of $H_0 : \beta = 0$				
t^{IM}	0.51%	>10%	>10%	6.7%
OST	0.0%	24.2%	0.6%	0.1%

This Paper: Two Contributions

1. Inference about difference between two population parameters (for example: treated and control)

   ⇒ Rely on approach analogous to IM (2005)

   ⇒ New small sample result for two-sample t-test under variance heterogeneity

2. Testing the level of clustering

   ⇒ Example: In a cross country regression, can countries be treated as independent, or do only (few) regions provide independent information?

   ⇒ Formal test of corresponding null hypothesis

Plan of Talk

1. Introduction
2. Validity of IM and other approaches
3. Generalization to two sample problem
   (a) New small sample result
   (b) Large sample robust inference
4. Testing the level of clustering
5. Conclusions

Inference in Clustered Linear Regression

- Assume observations $i = 1, \ldots, n_j$ from cluster $j = 1, \ldots, q$ satisfy

$$y_{j,i} = X'_{j,i}\theta + \varepsilon_{j,i}$$

where (conditional on $\{X_{i,j}\}$) $\varepsilon_{j,i}$ is mean zero normal, $E[\varepsilon_{j,i}\varepsilon_{l,k}] = 0$ for $j \neq l$, but $E[\varepsilon_{j,i}\varepsilon_{j,k}] \neq 0$ in general.

- Usual OLS estimator $\hat{\theta}$ can be written as

$$\hat{\theta} = \theta + \left(\sum_{j=1}^{q} \Gamma_j \right)^{-1} \sum_{j=1}^{q} Z_j$$

where $\Gamma_j = \sum_{i=1}^{n_j} X_{j,i}X'_{j,i}$ and $Z_j = \sum_{i=1}^{n_j} X_{j,i}\varepsilon_{j,i} \sim id\mathcal{N}(0, \Psi_j)$ with $\Psi_j = \text{Var}[\sum_{i=1}^{n_j} X_{j,i}\varepsilon_{j,i}]$.

Inference in Clustered Linear Regression

- Parameter of interest is $\beta = \iota_1'\theta$. Clustered standard error of $\hat{\beta} = \iota_1'\hat{\theta}$ is $\hat{\sigma}_\beta$, where

$$\hat{\sigma}_\beta^2 = \frac{q}{q-1}\iota_1'\left(\sum_{j=1}^{q}\Gamma_j\right)^{-1}\left(\sum_{j=1}^{q}\hat{e}_j\hat{e}_j'\right)\left(\sum_{j=1}^{q}\Gamma_j\right)^{-1}\iota_1,$$

$\hat{e}_j = Z_j - \Gamma_j(\hat{\theta} - \theta)$, and corresponding t-statistic of $H_0 : \beta = \beta_0$ is $t^{\text{cluster}} = (\hat{\beta} - \beta_0)/\hat{\sigma}_\beta$.

- IM (05): Estimate $\hat{\theta}_j = \theta + \Gamma_j^{-1}Z_j$ in each group, so that $\hat{\beta}_j = \beta + \iota_1'\Gamma_j^{-1}Z_j \sim id\mathcal{N}(0, \iota_1'\Gamma_j^{-1}\Psi_j\Gamma_j^{-1}\iota_1)$, and use t-statistic

$$t^{IM} = \sqrt{q}\frac{\overline{\hat{\beta}} - \beta_0}{S}$$

and critical value cv$(\alpha, q-1)$, where $\overline{\hat{\beta}} = q^{-1}\sum_{j=1}^{q}\hat{\beta}_j$ and $S^2 = \frac{1}{q-1}\sum_{j=1}^{q}(\hat{\beta}_j - \overline{\hat{\beta}})^2$.

Inference in Clustered Linear Regression

- Cameron, Gelbach and Miller (2008): Bootstrap critical value for t^{cluster}

- Bester, Conley and Hansen (2011): Use $\text{cv}(\alpha, q-1)$ for t^{cluster}

  $\Rightarrow$ BCH approach valid if Γ_j homogenous across clusters, since then $\hat{\beta} = \overline{\hat{\beta}}$ and $t^{\text{cluster}} = t^{IM}$

- In general, null rejection probability of CGM and BCH with finite q unknown and functions of $\{\Psi_j\}_{j=1}^{q}$ and $\{\Gamma_j\}_{j=1}^{q}$

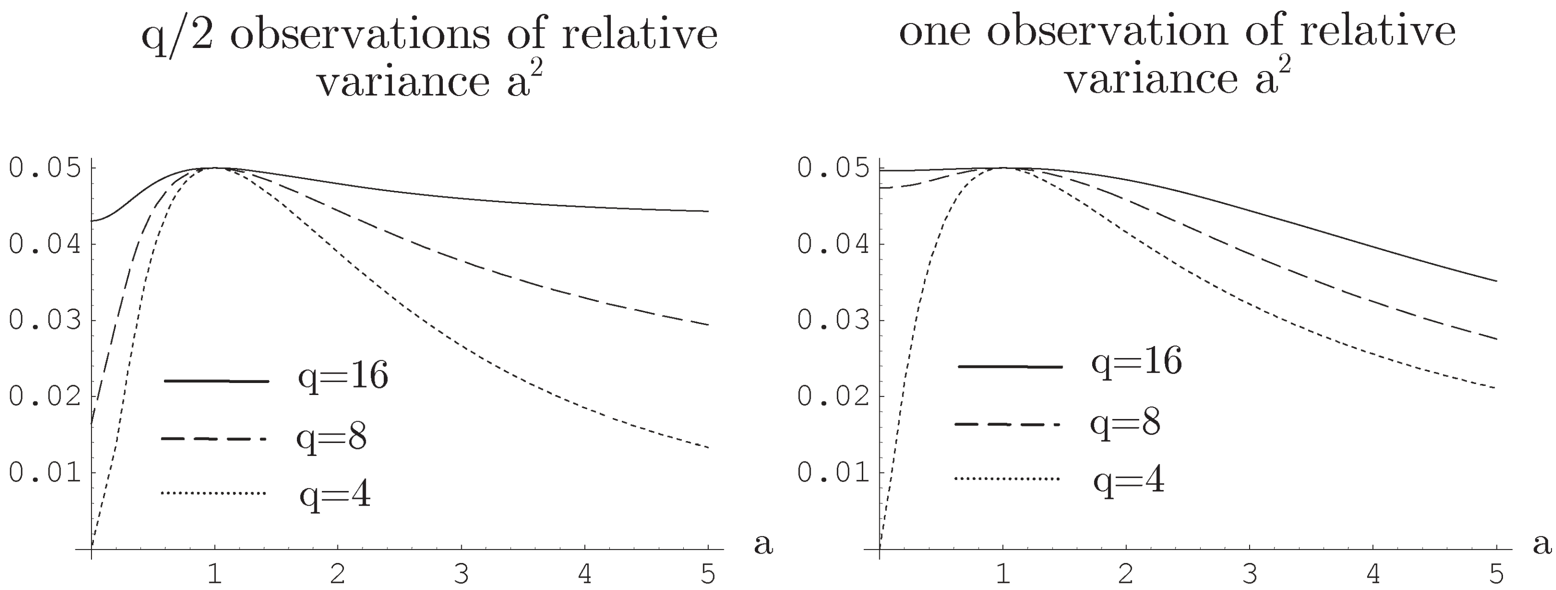
  $\Rightarrow$ to get some sense of size distortions, draw Γ_j $\sim$iid $\text{Wishart}_{2k}(I_k)$, and try to maximize null rejection probability over $\{\Psi_j\}_{j=1}^{q}$

Size Distortions of CGM and BCH

	CGM					BCH				
dim θ	min	Q_1	med	Q_3	max	min	Q_1	med	Q_3	max
					$q = 4$					
1	0.0	0.0	0.0	0.0	0.0	3.9	8.6	14.0	17.4	100.0
2	3.9	6.6	8.8	12.2	77.5	4.7	8.6	15.3	20.0	100.0
3	6.4	11.0	14.3	19.3	89.5	6.2	11.8	17.5	20.8	100.0
					$q = 8$					
1	4.3	4.4	4.5	4.8	4.9	6.6	9.7	11.8	16.7	27.3
2	4.8	9.1	11.1	13.9	33.2	6.3	9.3	11.6	14.9	27.4
3	6.3	10.8	14.0	17.8	35.8	6.0	9.3	11.6	14.7	26.2
					$q = 12$					
1	4.5	4.6	4.7	5.0	5.1	7.2	9.8	12.0	15.9	24.2
2	5.3	6.8	8.3	10.2	19.9	6.0	8.4	10.0	12.9	28.2
3	6.3	8.5	9.9	13.0	22.2	6.3	8.5	10.6	12.8	22.2

Null Rejection Probability of t^{IM}

- t^{IM} inference valid by Bakirov and Székely (2005) result
- But null rejection probability is smaller than α under variance heterogeneity



Power of t^{IM}

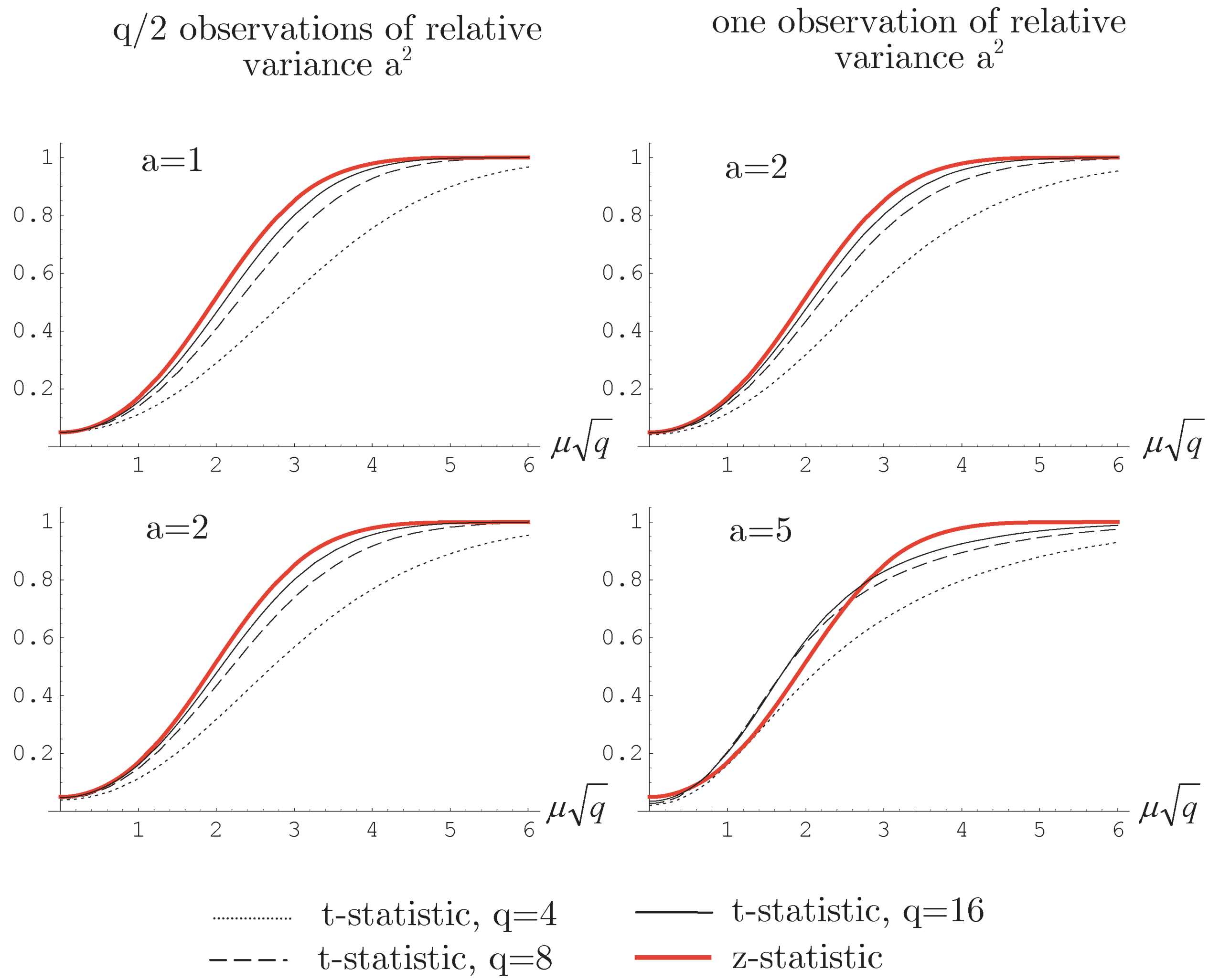
- Compare power of $t^{IM} = \sqrt{q}(\overline{\hat{\beta}} - \beta_0)/S$ to (infeasible) inference based on $z^{\text{cluster}} = (\hat{\beta} - \beta_0)/\sigma_{\beta}$.

- General comparison difficult, since $\hat{\beta}$ and $\overline{\hat{\beta}}$ are not identical unless Γ_j are homogeneous ($\Gamma_j = \Gamma_l$ for $j, l = 1, \ldots, q$).

- Theorem: $\text{Var}[\hat{\beta}]/\text{Var}[\overline{\hat{\beta}}]$ can take on (almost) any value as long as Γ_j are not homogenous.

- For homogeneous Γ_j, power difference reduces to difference between power of t and z test (under variance heterogeneity).

Power Comparison under Homogeneous Γ_j



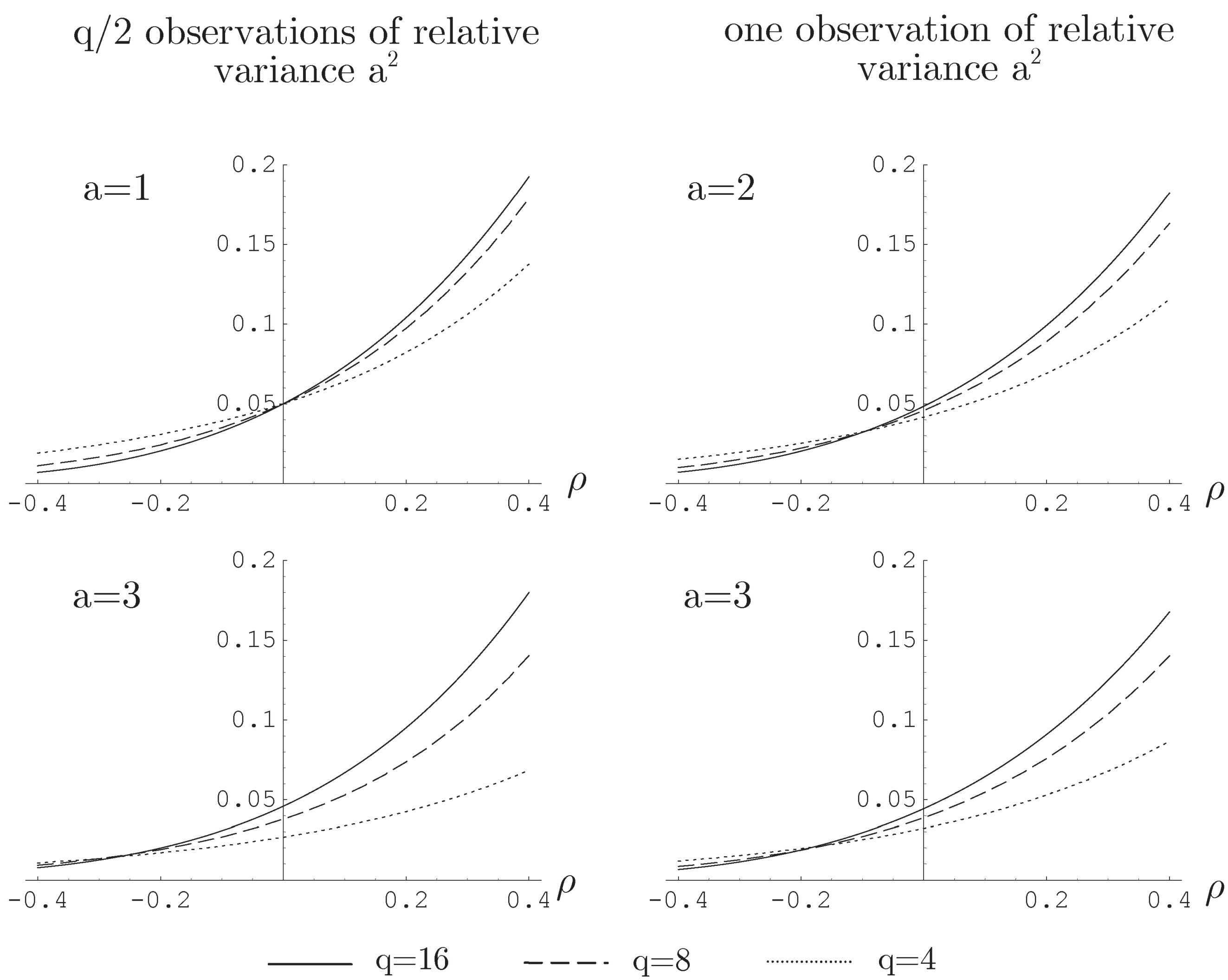
General Clustering Problem

- Above observations also relevant for GMM problems, since in large samples, Γ_j (=derivative of moment condition at true parameter) converges to constant by LLN, and group estimators are approximately independent and Gaussian by CLT.

- IM validity requires independence of $\hat{\beta}_j$ across groups

  $\Rightarrow$ typically implied by usual weak dependence assumptions in time series and spatial econometrics

Effect of AR(1) Correlation in $\hat{\beta}_j$



Additional (cheap) Generalization

- Suppose the variances of $\hat{\beta}_j$ are random: $\hat{\beta}_j | \{V_j\}_{j=1}^q \overset{a}{\sim} id\mathcal{N}(\beta, V_j)$, where V_j is potentially correlated across j.

- t^{IM} remains valid, since null rejection probability is below α conditional on $\{V_j\}_{j=1}^q$, and thus also unconditionally

  $\Rightarrow$ t^{IM} remains valid if $\hat{\beta}_j$ follows any any scale mixture of normals centered at β

  $\Rightarrow$ Applicable to time series with stochastic volatility

Time Series Application

- Keim (1983) provides evidence that the size anomaly of stock returns is, to a substantial degree, due to very high excess returns in January.

- Study differences between daily CRSP excess returns of portfolios constructed from firms in the top and bottom decile of equity market value, for each January of the 17 years 1963 to 1979.

- Overall significance of January difference assessed by OLS standard errors (treating trading days as uncorrelated).

Keim (1983)

Table 2

Average differences (*t*-statistics) between daily (CRSP) excess returns (in percent) of portfolios constructed from firms in the top and bottom decile of size (measured by market value of equity) on the NYSE and AMEX over the period 1963–1979.

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Mean daily return over all months
1963	0.309	0.093	−0.085	−0.045	0.172	0.056	−0.026	0.000	0.040	−0.021	0.011	−0.123	0.032
	(2.26)	(1.23)	(−0.81)	(−0.55)	(2.56)	(0.73)	(−0.39)	(0.00)	(0.51)	(−0.26)	(0.08)	(−1.38)	(1.16)
1964	0.170	0.105	0.097	0.007	−0.084	(−0.037	0.121	0.104	0.077	0.136	−0.019	−0.111	0.048
	(1.52)	(1.30)	(1.37)	(0.12)	(−1.34)	(−0.56)	(1.66)	(1.42)	(1.08)	(2.03)	(−0.23)	(−1.64)	(2.14)
1965	0.288	0.228	0.204	0.133	0.025	−0.212	0.070	0.104	−0.023	0.314	0.343	0.202	0.137
	(2.34)	(2.68)	(2.63)	(2.25)	(0.38)	(−2.41)	(1.05)	(1.34)	(−0.44)	(4.16)	(3.74)	(1.79)	(5.29)
1966	0.388	0.448	0.183	0.192	−0.278	0.017	−0.009	−0.177	−0.025	−0.423	0.138	0.001	0.033
	(4.63)	(4.44)	(1.52)	(2.11)	(−1.82)	(0.24)	(−0.08)	(−1.78)	(−0.23)	(−2.31)	(1.47)	(0.01)	(0.92)
1967	0.765	0.413	0.142	0.149	0.240	0.599	0.403	0.235	0.512	0.268	−0.120	0.427	0.336
	(4.59)	(5.04)	(2.35)	(1.54)	(2.56)	(3.87)	(4.15)	(3.09)	(5.59)	(2.70)	(−0.64)	(4.40)	(9.34)
1968	0.834	−0.197	−0.079	0.427	0.727	0.096	0.222	0.348	0.345	−0.002	0.091	0.434	0.285
	(6.20)	(−1.25)	(−0.52)	(2.70)	(6.52)	(1.11)	(1.30)	(3.37)	(4.15)	(−0.03)	(0.88)	(4.12)	(6.78)
1969	0.128	−0.253	−0.059	−0.139	0.082	−0.265	−0.241	−0.073	−0.006	0.247	−0.148	−0.349	−0.085
	(1.00)	(−2.49)	(−0.62)	(−1.94)	(1.19)	(−2.47)	(−1.88)	(−0.61)	(−0.07)	(2.48)	(−2.16)	(−4.06)	(−2.76)
1970	0.612	−0.257	0.033	−0.213	−0.016	−0.082	−0.164	−0.074	0.315	−0.019	−0.434	0.136	−0.011
	(2.49)	(−1.58)	(0.31)	(−1.39)	(−0.07)	(−0.47)	(−0.80)	(−0.48)	(2.91)	(−0.16)	(−4.44)	(1.08)	(−0.22)

Keim (1983)

- Wanto to allow for arbitrary serial correlation between trading days within each January

  $\Rightarrow$ use t-statistic with 16 degrees of freedom

  $\Rightarrow$ still significant January effect at 0.1% level

Comparing Parameters between Two Populations

- Similar to above:
  - Two populations with parameters δ_1 and δ_2, interest in $\beta = \delta_1 - \delta_2$. Null hypothesis $H_0 : \beta = \beta_0$
  - Divide data into q_1 and q_2 groups, respectively, and estimate model on each of the $q_1 + q_2$ groups
  - Groups chosen such that $\hat{\delta}_{1,j} \overset{a}{\sim} \mathcal{N}(\delta_1, \omega^2_{1,j})$, $j = 1, \ldots, q_1$, $\hat{\delta}_{2,j} \overset{a}{\sim} \mathcal{N}(\delta_2, \omega^2_{2,j})$, $j = 1, \ldots, q_2$, and $\hat{\delta}_{i,j}$ (approximately) independent

- Properties of two sample t-test

$$t^{IM} = \frac{\bar{\hat{\delta}}_1 - \bar{\hat{\delta}}_2 - \beta_0}{\sqrt{\frac{S_1^2}{q_1} + \frac{S_2^2}{q_2}}}$$

where $\bar{\hat{\delta}}_i = q_i^{-1} \sum_{j=1}^{q_i} \hat{\delta}_{i,j}$ and $S_i^2 = (q_i - 1)^{-1} \sum_{j=1}^{q_i} (\hat{\delta}_{i,j} - \bar{\hat{\delta}}_i)^2$ for $i = 1, 2$?

Small Sample Analogue

- $X_{i,j}$ independent RVs with $X_{i,j} \sim \mathcal{N}(\mu_i, \sigma^2_{i,j})$, $j = 1, \ldots, q_i$, $i = 1, 2$, where $q_i \geq 2$.

- Use t-test for $H_0 : \beta = \mu_1 - \mu_2 = \beta_0$ against $H_a : \beta \neq \beta_0$

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \beta_0}{\sqrt{\frac{s_1^2}{q_1} + \frac{s_2^2}{q_2}}}$$

  where $\bar{X}_i = q_i^{-1} \sum_{i=1}^{q_i} X_{i,j}$ and $s_i^2 = (q_i - 1)^{-1} \sum_{i=1}^{q_i} (X_{i,j} - \bar{X}_i)^2$ for $i = 1, 2$.

- Generalization of Behrens-Fisher problem, where $\sigma^2_{i,j} = \sigma^2_i$, but σ^2_1 potentially different from σ^2_2.

New Small Sample Result

- **Theorem:** Under the null hypothesis

$$\sup_{\{\sigma_{1,j}\}_{j=1}^{q_1}, \{\sigma_{2,j}\}_{j=1}^{q_2}} P(|t| > \text{cv}(\alpha, \min(q_1, q_2) - 1)) = \alpha$$

for $2 \leq q_1, q_2 \leq 50$ and $\alpha \in \{0.001, 0.002, \ldots, 0.083\}$, and also for $\alpha = \{0.084, 0.085, \ldots, 0.1\}$ if $2 \leq q_1, q_2 \leq 14$.

- Proof relies on Bakirov (1998), Bakirov and Székely (2005) and novel arguments, and requires computations for each α.

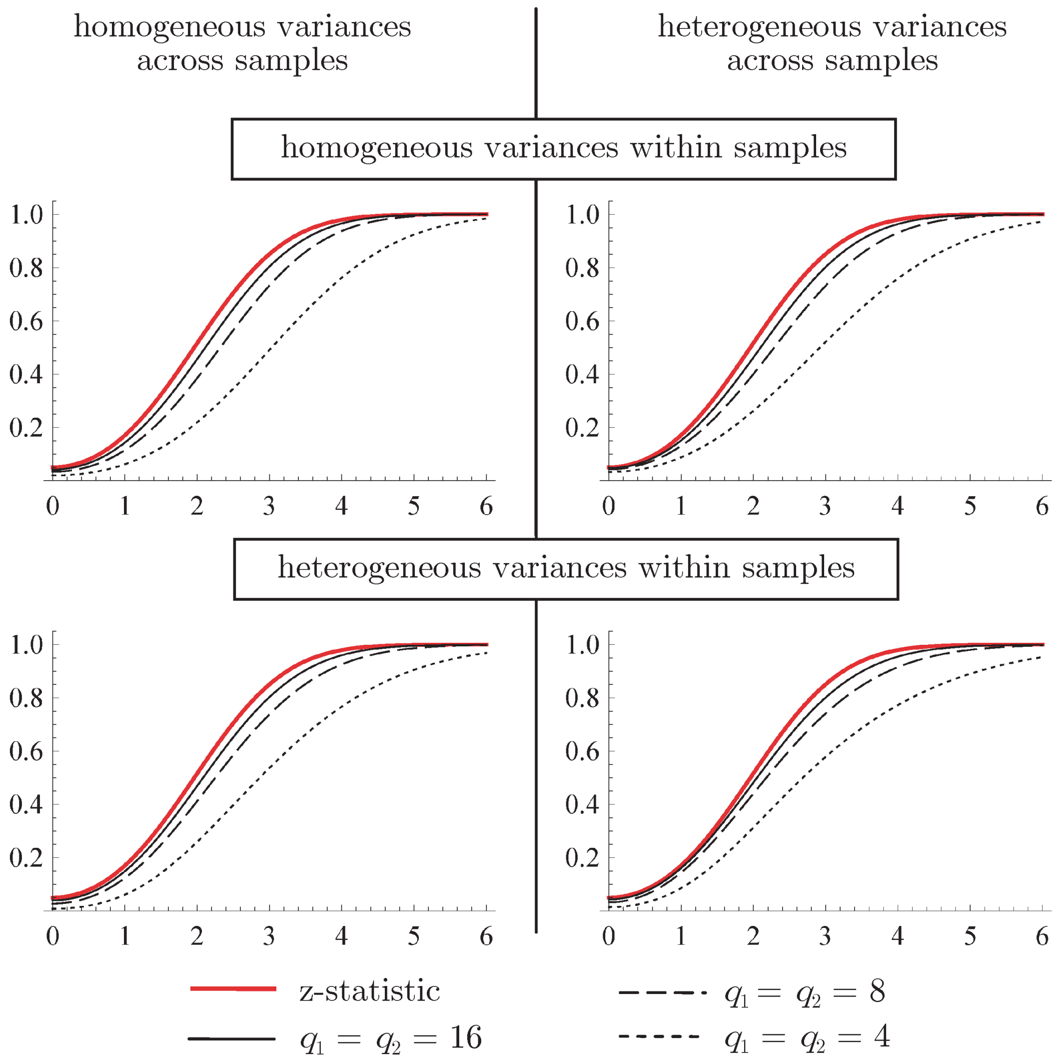
- Justifies asymptotically valid inference about β_0 whenever

$$\sqrt{n}(\hat{\delta}_{i,j} - \delta_i) \Rightarrow id\mathcal{N}(0, \sigma_{i,j}^2).$$

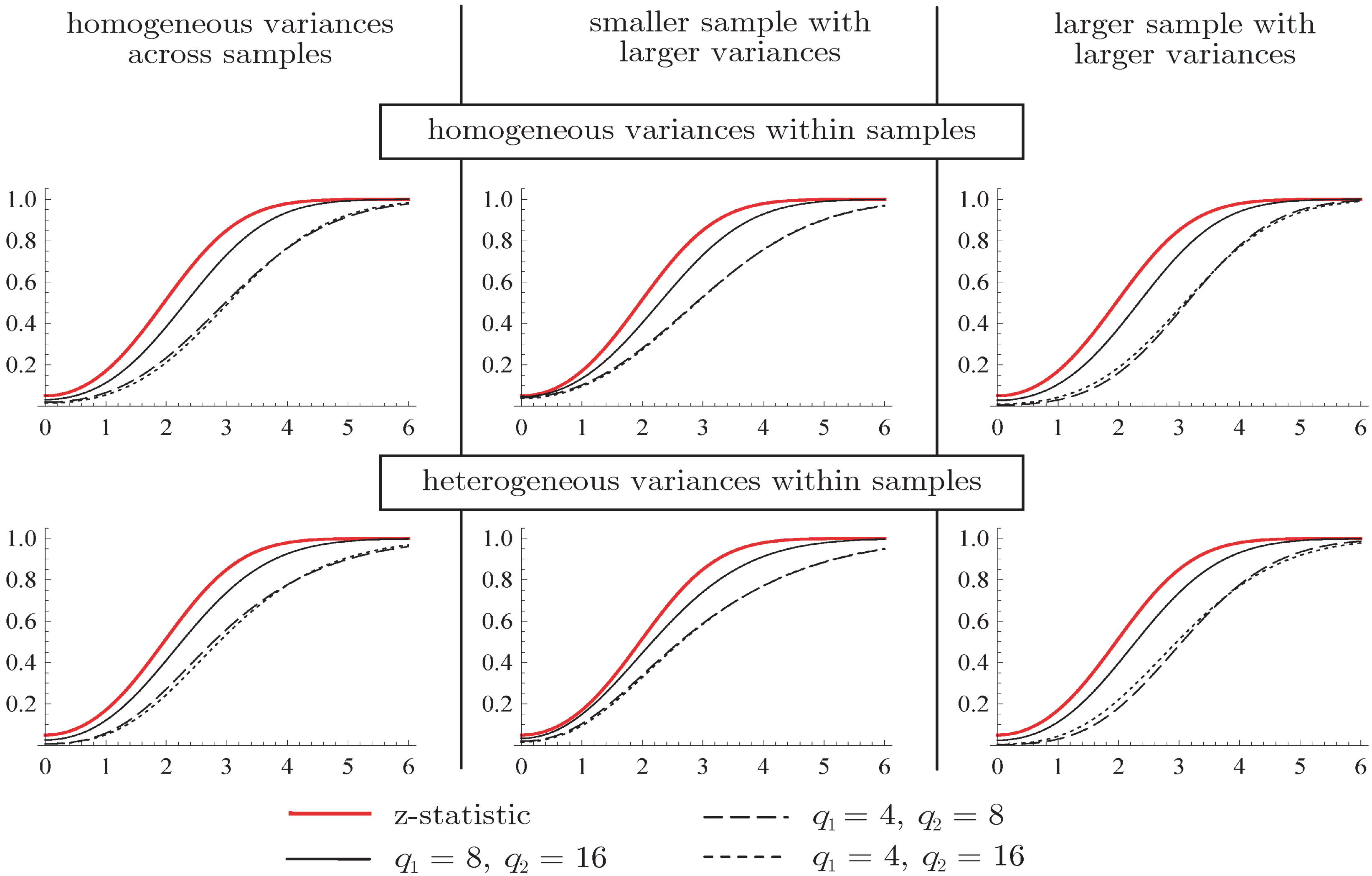
Size Distortions of CGM and BCH in Two-Sample Linear Regression Design

	CGM					BCH				
q_1	min	Q_1	med	Q_3	max	min	Q_1	med	Q_3	max
					$q_1 + q_2 = 4$					
2	4.6	10.2	12.2	14.7	26.3	18.3	22.6	35.6	100.0	100.0
					$q_1 + q_2 = 8$					
2	19.6	38.9	42.3	46.1	100.0	20.9	28.9	32.0	100.0	100.0
3	10.8	17.0	20.3	25.9	47.8	11.7	21.7	27.5	32.4	100.0
4	7.5	13.0	15.8	19.4	38.9	13.4	22.8	27.3	30.6	100.0
					$q_1 + q_2 = 12$					
2	36.1	44.2	46.0	47.8	100.0	21.0	32.8	34.5	100.0	100.0
3	16.9	20.6	22.6	28.5	100.0	11.3	19.3	24.2	31.8	100.0
4	10.2	13.1	16.2	21.0	41.3	10.1	16.6	22.2	26.8	100.0
5	8.2	10.8	12.3	16.1	100.0	11.2	16.4	21.7	25.2	100.0
6	7.7	12.1	14.9	18.4	36.1	13.4	20.2	24.1	27.8	100.0

Power of Two-Sample t-Test under $\Gamma_{i,j} = \Gamma$



Power of Two-Sample t-Test under $\Gamma_{i,j} = \Gamma$



Dal Bó and Fréchette (2011)

- Dal Bó and Fréchette (2011) experimentally study degree of cooperation in infinitely repeated games as function of the probability of continuation p, and the pay-off of cooperation R.

- Six treatments (pairs of values (p, R)), with three sessions each.

  Sessions have heterogenous number of individuals (12-20)

- Propensity to cooperate measured by probit coefficient.

- Dal Bó and Fréchette (2011) conduct 7 pairwise comparisons with standard errors clustered by session.

Dal Bó and Fréchette (2011)

Probit coefficients $\hat{\delta}_{i,j}$ in session j of treatment i

i	(p, R)	$\hat{\delta}_{i,1}$	$\hat{\delta}_{i,2}$	$\hat{\delta}_{i,3}$
1	$(\frac{1}{2}, 32)$	-1.538	-0.963	-1.698
2	$(\frac{1}{2}, 40)$	-1.052	-0.813	-0.878
3	$(\frac{1}{2}, 48)$	-0.262	-0.261	-0.684
4	$(\frac{3}{4}, 32)$	-0.833	-0.698	-0.974
5	$(\frac{3}{4}, 40)$	0.176	0.905	-0.200
6	$(\frac{3}{4}, 48)$	0.458	1.037	0.674

Dal Bó and Fréchette (2011)

	H_0 is equal cooperation under (p_1, R_1) and (p_2, R_2)						
(p_1, R_1)	$(\frac{1}{2}, 32)$	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 32)$	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 48)$	$(\frac{3}{4}, 32)$	$(\frac{3}{4}, 40)$
(p_2, R_2)	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 48)$	$(\frac{3}{4}, 32)$	$(\frac{3}{4}, 40)$	$(\frac{3}{4}, 48)$	$(\frac{3}{4}, 40)$	$(\frac{3}{4}, 48)$
	p-values of Tests of H_0						
DBF	8.6%	0.0%	3.9%	0.0%	0.0%	0.0%	11.5%
t^{IM}	$>$ 10%	8.4%	$>$ 10%	6.8%	3.6%	7.7%	$>$ 10%

Diff-in-Diff

- Bloom et al. (2013) conduct field experiment on randomly selected firms in the textile industry in India to determine the importance of management practices on productivity

- Fourteen treatment plants and 6 control plants

- Let $\hat{\delta}_{1,j}$ be the the difference in average productivity between post- and pre-treatment periods for the control plants $j = 1, \ldots, 6$, and $\hat{\delta}_{2,j}$, $j = 1, \ldots, 14$ analogously for the treatment plants

  $\Rightarrow$ any plant fixed effects cancel in these differences

- Note that in $\bar{\hat{\delta}}_1 - \bar{\hat{\delta}}_2 = q_1^{-1} \sum_{j=1}^{q_1} \hat{\delta}_{1,j} - q_2^{-1} \sum_{j=1}^{q_2} \hat{\delta}_{2,j}$, as well as in $S_i = (q_j - 1)^{-1} \sum_{j=1}^{q_j} (\hat{\delta}_{i,j} - \bar{\hat{\delta}}_i)^2$, also any time series fixed effects cancel

  $\Rightarrow$ two sample t-statistic approach numerically invariant to presence of both time and plant fixed effects

- Bloom et al. report results based on t^{IM} with 5 degrees of freedom

Testing the Level of Clustering

- Consider initially one-sample set-up: $\hat{\beta}_j \overset{a}{\sim} id\mathcal{N}(\beta, \omega_j^2)$, $j = 1, \ldots, q$, where $\hat{\beta}_j$ is based on data from coarse cluster j (region)

- Want to test the null hypothesis that fine clustering (countries) would also be valid, maintaining validity of coarse clustering under alternative

- Under null hypothesis, can estimate ω_j well by clustered standard error $\hat{\omega}_j$ (use ", cluster" when estimating $\hat{\beta}_j$, $j = 1, \ldots, q$)

Obstfeld, Shambaugh and Taylor (2010)

Estimators $\hat{\beta}_j$ and estimated standard errors $\hat{\omega}_j$ (clustered by country)

	ln(M2/GDP)	
Region	$\hat{\beta}_j$	$\hat{\omega}_j$
Asia/Pacific	0.627	0.164
W Europe	1.041	0.319
E. Europe	0.633	0.144
Africa	-0.019	0.179
Middle East	0.511	0.152
S. America	-0.201	0.196

Testing the Level of Clustering, ctd

- Variance of $\overline{\hat{\beta}} = q^{-1}\sum_{j=1}^{q}\hat{\beta}_j$ is $q^{-2}\sum_{j=1}^{q}\omega_j^2$. Idea of test: Compare the two estimators

$$q^{-2}\sum_{j=1}^{q}\hat{\omega}_j^2 \quad \text{and} \quad S^2 = q^{-1}(q-1)^{-1}\sum_{j=1}^{q}(\hat{\beta}_j - \overline{\hat{\beta}})^2$$

- Under asymptotics where fine clustering has more and more clusters in each coarse cluster, sampling variability of S^2 dominates under null hypothesis

  $\Rightarrow$ Under null hypothesis, distribution of S^2 well approximated by distribution of $S_X^2 = q^{-1}(q-1)^{-1}\sum_{i=1}^{q}(X_j - \bar{X})^2$, where X_j are $X_j \sim id\mathcal{N}(0, \hat{\omega}_j^2)$ (conditional on $\{\hat{\omega}_j^2\}$)

  $\Rightarrow$ Under alternative, clustered standard error ω_j are incorrect, and $S^2 \nsim S_X^2$ (typically, $\hat{\omega}_j < \omega_j$, so distribution of S^2 stochastically dominates distribution of S_X^2)

  $\Rightarrow$ reject if S^2 is larger than 95% quantile of distribution of S_X^2

- Analogous test for comparisons between two populations

Small Sample Performance

- Panel data $y_{j,t}$, $j = 1, \ldots, q$, $t = 1, \ldots, T$. Under H_0, data is i.i.d. in time, under alternative, time series autocorrelation with coefficient 0.5. Coarse cluster by j, fine cluster by j, t (=OLS standard error)

- Under variance homogeneity, q cross-sectional units are i.i.d., under heterogeneity, $q/2$ have twofold standard deviation

	homogenous σ_i			heterogeneous σ_i		
$q \backslash T$	5	10	20	5	10	20
	Null hypothesis					
4	7.2	6.6	5.2	8.3	7.2	5.5
8	6.0	5.4	5.0	7.0	5.9	5.5
16	5.1	5.4	5.0	5.9	5.6	5.1
	Alternative hypothesis					
4	45.7	47.1	46.2	42.8	42.7	41.6
8	65.7	69.1	69.7	58.6	61.8	60.8
16	87.2	89.7	90.0	79.6	82.7	83.1

Obstfeld, Shambaugh and Taylor (2010)

Estimators $\hat{\beta}_j$ and estimated standard errors $\hat{\omega}_j$ (clustered by country)

	fin. openness		Peg		Soft Peg		ln(M2/GDP)	
Region	$\hat{\beta}_j$	$\hat{\omega}_j$	$\hat{\beta}_j$	$\hat{\omega}_j$	$\hat{\beta}_j$	$\hat{\omega}_j$	$\hat{\beta}_j$	$\hat{\omega}_j$
Asia/Pacific	1.110	0.221	0.035	0.113	-0.060	0.119	0.627	0.164
W Europe	0.805	0.430	0.089	0.179	0.069	0.147	1.041	0.319
E. Europe	0.423	0.353	0.317	0.168	0.281	0.111	0.633	0.144
Africa	0.508	0.433	0.413	0.151	0.318	0.101	-0.019	0.179
Middle East	1.665	0.438	-0.236	0.193	-0.056	0.153	0.511	0.152
S. America	0.770	0.309	-0.279	0.165	-0.067	0.146	-0.201	0.196
p-values of H_0 : Clustering by country is fine								
	19.3%		1.3%		10.7%		0.1%	

Keim (1983)

Table 2

Average differences (*t*-statistics) between daily (CRSP) excess returns (in percent) of portfolios constructed from firms in the top and bottom decile of size (measured by market value of equity) on the NYSE and AMEX over the period 1963–1979.

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Mean daily return over all months
1963	0.309	0.093	−0.085	−0.045	0.172	0.056	−0.026	0.000	0.040	−0.021	0.011	−0.123	0.032
	(2.26)	(1.23)	(−0.81)	(−0.55)	(2.56)	(0.73)	(−0.39)	(0.00)	(0.51)	(−0.26)	(0.08)	(−1.38)	(1.16)
1964	0.170	0.105	0.097	0.007	−0.084	(−0.037	0.121	0.104	0.077	0.136	−0.019	−0.111	0.048
	(1.52)	(1.30)	(1.37)	(0.12)	(−1.34)	(−0.56)	(1.66)	(1.42)	(1.08)	(2.03)	(−0.23)	(−1.64)	(2.14)
1965	0.288	0.228	0.204	0.133	0.025	−0.212	0.070	0.104	−0.023	0.314	0.343	0.202	0.137
	(2.34)	(2.68)	(2.63)	(2.25)	(0.38)	(−2.41)	(1.05)	(1.34)	(−0.44)	(4.16)	(3.74)	(1.79)	(5.29)
1966	0.388	0.448	0.183	0.192	−0.278	0.017	−0.009	−0.177	−0.025	−0.423	0.138	0.001	0.033
	(4.63)	(4.44)	(1.52)	(2.11)	(−1.82)	(0.24)	(−0.08)	(−1.78)	(−0.23)	(−2.31)	(1.47)	(0.01)	(0.92)
1967	0.765	0.413	0.142	0.149	0.240	0.599	0.403	0.235	0.512	0.268	−0.120	0.427	0.336
	(4.59)	(5.04)	(2.35)	(1.54)	(2.56)	(3.87)	(4.15)	(3.09)	(5.59)	(2.70)	(−0.64)	(4.40)	(9.34)
1968	0.834	−0.197	−0.079	0.427	0.727	0.096	0.222	0.348	0.345	−0.002	0.091	0.434	0.285
	(6.20)	(−1.25)	(−0.52)	(2.70)	(6.52)	(1.11)	(1.30)	(3.37)	(4.15)	(−0.03)	(0.88)	(4.12)	(6.78)
1969	0.128	−0.253	−0.059	−0.139	0.082	−0.265	−0.241	−0.073	−0.006	0.247	−0.148	−0.349	−0.085
	(1.00)	(−2.49)	(−0.62)	(−1.94)	(1.19)	(−2.47)	(−1.88)	(−0.61)	(−0.07)	(2.48)	(−2.16)	(−4.06)	(−2.76)
1970	0.612	−0.257	0.033	−0.213	−0.016	−0.082	−0.164	−0.074	0.315	−0.019	−0.434	0.136	−0.011
	(2.49)	(−1.58)	(0.31)	(−1.39)	(−0.07)	(−0.47)	(−0.80)	(−0.48)	(2.91)	(−0.16)	(−4.44)	(1.08)	(−0.22)

$\Rightarrow$ Test of level of clustering at day level against clustering at year level rejects at 0.1% level

Dal Bó and Fréchette (2011)

Probit coefficients $\hat{\delta}_{i,j}$ and estimated standard errors $\hat{\omega}_{i,j}$ (clustered by individual) in session j of treatment i

i	(δ, R)	$\hat{\delta}_{i,1}$	$\hat{\omega}_{i,1}$	$\hat{\delta}_{i,2}$	$\hat{\omega}_{i,2}$	$\hat{\delta}_{i,3}$	$\hat{\omega}_{i,3}$
1	$(\frac{1}{2}, 32)$	-1.538	0.163	-0.963	0.183	-1.698	0.216
2	$(\frac{1}{2}, 40)$	-1.052	0.147	-0.813	0.146	-0.878	0.148
3	$(\frac{1}{2}, 48)$	-0.262	0.185	-0.261	0.221	-0.684	0.179
4	$(\frac{3}{4}, 32)$	-0.833	0.142	-0.698	0.167	-0.974	0.198
5	$(\frac{3}{4}, 40)$	0.176	0.153	0.905	0.099	-0.200	0.205
6	$(\frac{3}{4}, 48)$	0.458	0.118	1.037	0.132	0.674	0.113

Dal Bó and Fréchette (2011)

Treatment pair under consideration							
(p_1, R_1)	$(\frac{1}{2}, 32)$	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 32)$	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 48)$	$(\frac{3}{4}, 32)$	$(\frac{3}{4}, 40)$
(p_2, R_2)	$(\frac{1}{2}, 40)$	$(\frac{1}{2}, 48)$	$(\frac{3}{4}, 32)$	$(\frac{3}{4}, 40)$	$(\frac{3}{4}, 48)$	$(\frac{3}{4}, 40)$	$(\frac{3}{4}, 48)$
p-value of Test of validity of clustering at level of individuals							
	2.5%	28.5%	3.6%	0.0%	3.7%	0.0%	0.0%

Conclusion

- General approach to correlation and heterogeneity robust inference
- Method imposes only a "finite amount of independence" through the assumption that estimators from different groups are (approximately) independent
- Straightforward to implement
- Transparency of clustering assumption in time series and spatial applications arguably advantage over kernel methods
- Difficult to generalize to F-tests