Correlation and Heterogeneity Robust Inference using Conservativeness of Test Statistic

Rustam Ibragimov
Harvard University

Ulrich K. Müller
Princeton University

London School of Economics
November 24, 2011
Introduction

• Most economic data is observational
  ⇒ data is plausibly not i.i.d.: think of countries, firms, or time series data

• Correlations and heterogeneity do not necessarily invalidate standard estimators, but standard errors and inference become a challenge
Consistent Variance Estimators

- White (1980) for heterogeneous and independent disturbances
- Rogers (1993) and Arellano (1987) for clustered and panel data
- Conley (1999) for spatially correlated data

⇒ based on a Law of Large Numbers, and thus require ”an infinite amount of independence”

⇒ 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:

⇒ We add to this literature and develop a general approach to robust inference when little is known about correlations and heterogeneity
Our Approach

- **Tests** on parameter $\beta$: $t$–**statistic** in **group** estimators

- Assume data can be classified in a finite number $q$ of groups that allow asymptotically independent normal inference about the (scalar) parameter of interest $\beta$, so that $\hat{\beta}_j \sim idN(\beta, v_j^2)$ for $j = 1, \cdots, q$. Time series example: Divide data into $q = 4$ consecutive blocks, and estimate the model 4 times.

- Treat $\hat{\beta}_j$ as observations for the usual $t$-statistic, and reject a 5% level test if $t$-statistic is larger than usual critical value for $q - 1$ degrees of freedom. Results in valid inference by small sample result.

- Exploits information $\hat{\beta}_j \sim idN(\beta, v_j^2)$ in an efficient way.

- Does not rely on single asymptotic model of sampling variability for estimated standard deviation

- Important precursor: Fama-MacBeth (1973) method
Our Approach

- **Tests** on equality of parameters $\beta_1 = \beta_2$ : Behrens-Fisher statistic in **group** estimators

- Partition data into $q_1 + q_2$ groups; estimate $\beta_1$ and $\beta_2$ with $q_1$ and $q_2$ groups: $\hat{\beta}_{ij} \sim \text{idN}(\beta_i, \nu_{ij}^2)$ for $j = 1, \ldots, q_i$, $i = 1, 2$.

- Treat $\hat{\beta}_{ij}$ as observations for the usual Behrens-Fisher (BF) statistic, and reject a 1% or 5% level test if BF statistic is larger than critical value of Student-t with $\min(q_1, q_2) - 1$ degrees of freedom. Valid inference by new small sample result.

- Exploits information $\hat{\beta}_{ij} \sim \text{idN}(\beta_i, \nu_{ij}^2)$ in an efficient way.

- Applications: robust analysis of **treatment effects**, **structural breaks**, inference under **heavy tails**, etc.
Plan of Talk

• Introduction

• The Small Sample t- and Behrens-Fisher Tests
  - Conservativeness for Heterogenous Variances
  - Optimality

• t-Statistic & Behrens-Fisher Statistic Based Large Sample Robust Inference
  - Basic Idea and Properties
  - Comparison with Standard Inference and Known Asymptotic Variance

• Applications: Panel data, Time Series, Inequality Measures (on-going work with Paul Kattuman, Univ. of Cambridge)

• Conclusions
t—statistic based robust inference: small samples

- Bakirov and Székely (2005), Ibragimov and Müller (2006): Usual small sample $t$—test of level $\alpha \leq 5\%$ : conservative for independent heterogeneous Gaussian observations (not $\alpha = 10\%$)

- $X_j \sim N(\mu, \sigma_j^2), \ j = 1, \cdots, q : H_0 : \mu = 0$ against $H_1 : \mu \neq 0$

  **t-statistic** $t = \sqrt{q} \frac{\bar{X}}{s_X}$

  $\bar{X} = q^{-1} \sum_{j=1}^{q} X_j, s_X^2 = (q - 1)^{-1} \sum_{j=1}^{q} (X_j - \bar{X})^2$

  $cv_q(\alpha) =$ critical value of $T_{q-1} : P(|T_{q-1}| > cv_q(\alpha)) = \alpha$

- $P(|t| > cv(\alpha)|H_0) \leq P(|t| > cv(\alpha)|H_0, \sigma_1^2 = \cdots = \sigma_q^2) = P(|T_{q-1}| > cv(\alpha)) = \alpha$

- Holds under **heavy tails**: mixtures of normals (stable, Student-$t$)
Conservativeness

Theorem (Bakirov and Székely 2005): Let $cv_q(\alpha)$ be the critical value of the usual two-sided t-test based on $|t|$ of level $\alpha$, i.e. $P(|T_q-1| > cv_q(\alpha)) = \alpha$, where $T_k$ is student-t distributed with $k$ degrees of freedom, and let $\Phi$ denote the cumulative density function of a standard normal random variable.

(i) If $\alpha \leq 2\Phi(-\sqrt{3}) = 0.08326...$, then for all $q \geq 2$,

$$\sup_{\{\sigma_1^2,\ldots,\sigma_q^2\}} P(|t| > cv_q(\alpha)|H_0) = P(|T_q-1| > cv_q(\alpha)) = \alpha. \quad (1)$$

(ii) Equation (1) also holds true for $2 \leq q \leq 14$ if $\alpha \leq \alpha_1 = 0.1$, and for $q \in \{2, 3\}$ if $\alpha \leq \alpha_2 = 0.2$. Moreover, define $\tilde{cv}_q(\alpha_i) = \sqrt{k_i(q-1)cv_{k_i}(\alpha_i)^2}/\sqrt{q(k_i-1) + (q-k_i)cv_{k_i}(\alpha_i)^2}$, $i \in \{1, 2\}$, where $k_1 = 14$ and $k_2 = 3$. Then for $q \geq k_i + 1$,

$$\sup_{\{\sigma_1^2,\ldots,\sigma_q^2\}} P(|t| > \tilde{cv}_q(\alpha_i)|H_0) = \alpha_i.$$
Sketch of Proof I

• Rewrite

\[ t^2 = \frac{(q - 1)(\sum_{j=1}^{q} X_j)^2}{q \sum_{j=1}^{q} (X_j - \bar{X})^2} > cv^2 \]

as \( X'AX > 0 \), where \( A = (1 + \frac{cv^2}{q-1})ee' - \frac{qcv^2}{q-1}I_q \), and \( e = (1, \cdots, 1)' \).

• Let \( D = \text{diag}(\sigma_1, \cdots, \sigma_q) \). Then \( X'AX \sim Z'BZ \), where \( B = DAD \) and \( Z \sim \mathcal{N}(0, I_q) \).

• \( Z'BZ \sim \sum_{j=1}^{q} \lambda_j Z_j^2 \), where \( \lambda_j \) are the eigenvalues of \( B \). Observation: One eigenvalue (say, \( \lambda_q \)) of \( B \) is positive, and \( q-1 \) are negative. Thus,

\[ P(t^2 > cv^2) = P(Z_q^2 > \sum_{j=1}^{q-1} c_j Z_j^2) \]

with \( c_j = -\lambda_j/\lambda_q \). But: eigenvalues of \( B \) are a very messy function of \( \{\sigma_1^2, \cdots, \sigma_q^2\} \).
Sketch of Proof II

- Fact, already exploited in Bakirov (1989): for $c_j > 0$, $j = 1, \cdots, q - 1$

$$P(Z_q^2 > \sum_{j=1}^{q-1} c_j Z_j^2) = \int_0^1 \frac{u^{(q-1)/2}(1 - u)^{-1/2}}{\pi \sqrt{\prod_{j=1}^{q-1}(u + c_j)}} du$$

- $\prod_{j=1}^{q-1}(u + c_j)$ is a polynomial in $u$ with roots $-c_j = \lambda_j / \lambda_q$. Eigenvalues $\lambda_j$ are roots of $\det(uI_q - B)$, and so after a scale normalization of $Z'BZ > 0$ to ensure $\lambda_q = 1$, we have

$$\prod_{j=1}^{q-1}(u + c_j) = \frac{\det(uI_q - B)}{u - 1}$$

- Obtain expression for $P(t^2 > cv^2)$ as an integral with an integrand that is a reasonably tractable function of $\{\sigma_1^2, \cdots, \sigma_q^2\}$. Argue why it cannot be maximized for positive but unequal variances...
Sharpe ratio density: Normal vs. Cauchy
Behrens-Fisher statistic robust inference: small sample results

- $X_{ij} \sim \mathcal{N}(\mu_i, \sigma_j^2)$, $j = 1, ..., q_i$, $i = 1, 2$

- $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$

- **Behrens-Fisher** statistic $BF = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

  $$\bar{X}_i = q_i^{-1} \sum_{j=1}^{q_i} X_{i,j}$$

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

- $P(|BF| \geq cv_{m-1}(\alpha)) \leq P(|T_{m-1}| \geq cv_{m-1}(\alpha)) = \alpha$, $m = \min(q_1, q_2)$

  for $2 \leq q_1, q_2 \leq 50$ and $\alpha \in \{0.1\%, 1\%, 5\\%\}$
Effective Rejection Probabilities

$q/2$ observations of relative variance $a^2$

one observation of relative variance $a^2$
Optimality of t-Statistic

- Let $X_j, j = 1, \cdots, q$ with $q \geq 2$, be distributed independent $\mathcal{N}(\mu, \sigma^2_j)$, and consider the hypothesis test

  $$H_0 : \mu = 0 \text{ and } \{\sigma^2_j\}_{j=1}^q \text{ arbitrary}$$

  $$H_1 : \mu \neq 0 \text{ and } \sigma^2_j = \sigma^2 \text{ for all } j$$

- Theorem: Let $cv$ be such that $P(|T_{q-1}| > cv) = \alpha \leq 0.05$. For any $q \geq 2$, a test that rejects the null hypothesis for $|t| > cv$ is the uniformly most powerful scale invariant level $\alpha$ test.

- Proof:
  - For equal variance case, t-test is UMP scale invariant.
  - Conservativeness of t-test implies that equal variance case under the null hypothesis is least favorable.
Asymptotic t-Statistic Based Inference

- Partition the data into $q \geq 2$ groups, with $n_j$ observations in group $j$, and $\sum_{j=1}^{q} n_j = n$.

- Denote by $\hat{\beta}_j$ the estimator of $\beta$ using observations in group $j$ only.

- Suppose the groups are chosen such that
  
  - $\sqrt{n}(\hat{\beta}_j - \beta) \Rightarrow \mathcal{N}(0, \sigma_j^2)$ for all $j$ (where $\max_{1 \leq j \leq q} \sigma_j^2 > 0$). Satisfied for many models as long as $\min_j n_j \to \infty$, linear or nonlinear.
  
  - $\sqrt{n}(\hat{\beta}_i - \beta)$ and $\sqrt{n}(\hat{\beta}_j - \beta)$ are asymptotically independent for $i \neq j$.

  $\Rightarrow$ this amounts to

  $\sqrt{n}(\hat{\beta}_1 - \beta, \cdots, \hat{\beta}_q - \beta)' \Rightarrow \mathcal{N}(0, \text{diag}(\sigma_1^2, \cdots, \sigma_q^2))$
Asymptotic t-Statistic Based Inference

• Rejection of $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ if $|t_\beta|$ exceeds the $(1 - \alpha/2)$ percentile of the student-t distribution with $q - 1$ degrees of freedom, where $t_\beta$ is the usual t-statistic

$$t_\beta = \frac{\bar{\beta} - \beta_0}{s_\beta}$$

with $\bar{\beta} = q^{-1} \sum_{j=1}^{q} \hat{\beta}_j$ and $s_\beta^2 = (q - 1)^{-1} \sum_{j=1}^{q} (\hat{\beta}_j - \bar{\beta})^2$, is asymptotically valid inference by small sample t-test result and the Continuous Mapping Theorem.

• This way of conducting inference efficiently exploits the information

$$\sqrt{n}(\hat{\beta}_1 - \beta, \cdots, \hat{\beta}_q - \beta)' \Rightarrow \mathcal{N}(0, \text{diag}(\sigma_1^2, \cdots, \sigma_q^2))$$

as it maximizes asymptotic local power uniformly against all local alternatives where $\beta = \beta_n = \beta_0 + c/\sqrt{n}$ for some $c \neq 0$ and $\sigma_i^2 = \sigma_j^2$ for all $i, j$, among all scale invariant tests by optimality of t-statistic.
Asymptotic Behrens-Fisher Statistic Based Inference

- **Robust tests** of $H_0 : \beta_1 = \beta_2$ against $H_1 : \beta_1 \neq \beta_2$

- Partition the data into $q_1 + q_2$ groups

- $\hat{\beta}_{ij}, j = 1, ..., q_i$ : estimators of $\beta_i, i = 1, 2$, using observations in group $j$ only

- **Behrens-Fisher** statistic $BF = \frac{\overline{\beta}_1 - \overline{\beta}_2}{\sqrt{\frac{s_1^2}{q_1} + \frac{s_2^2}{q_2}}}$

  \[ \overline{\beta}_i = q_i^{-1} \sum_{j=1}^{q_i} \hat{\beta}_{i,j} \]

  \[ S_i^2 = (q_i - 1)^{-1} \sum_{j=1}^{q_i} (\hat{\beta}_{i,j} - \overline{\beta}_i)^2, \quad i = 1, 2 \]

- Reject $H_0$ if $|BF|$ is larger than the critical value of Student-t distribution with $\min(q_1, q_2) - 1$ degrees of freedom
Size Control under AR(1) Correlation

$q/2$ observations of relative variance $a^2$

$a=1$

$a=2$

$a=3$

$q=16$  
$q=8$  
$q=4$
Comparison with Inference with Known Asymptotic Variance

• Typically,
\[ \sqrt{n}(\hat{\beta}_1 - \beta, \ldots, \hat{\beta}_q - \beta) \rightarrow N(0, \text{diag}(\sigma_1^2, \ldots, \sigma_q^2)) \]

is weaker than what is required to consistently estimate the asymptotic variance. Consistent estimation not only requires more assumptions on correlation structure, but typically also higher moments.

• What are the efficiency cost of this additional robustness, i.e. how does the t-statistic approach compare to an approach based on consistent variance estimation (if the variance can indeed be consistently estimated)?

• Consider question in exactly identified GMM framework with \( k \times 1 \) parameter \( \theta \) and moment condition \( E[g(\theta, y_i)] = 0 \). We are interested in conducting inference about the first element of \( \theta \), denoted by \( \beta \).
Properties of Group Estimators

- Let $G_j$ the set of indices of group $j$ observations. Suppose the estimator $\hat{\theta}_j$ based on group $j = 1, \cdots, q$ data satisfies

$$
\sqrt{n}(\hat{\theta}_j - \theta) = \Gamma_j^{-1}Q_j + o_p(1) \Rightarrow \mathcal{N}(0, \Gamma_j^{-1}\Omega_j\Gamma_j'^{-1})
$$

where

$$
n^{-1} \sum_{i \in G_j} \frac{\partial g(a,y_i)}{\partial a} \bigg|_{a=\hat{\theta}_j} \xrightarrow{p} \Gamma_j \quad \text{and} \quad Q_j = n^{-1/2} \sum_{i \in G_j} g(\theta, y_i) \Rightarrow \mathcal{N}(0, \Omega_j), \quad \text{and} \quad (Q'_1, \cdots, Q'_q)' \Rightarrow \mathcal{N}(0, \text{diag}(\Omega_1, \cdots, \Omega_q)).
$$

- The simple average of the group estimators $\bar{\hat{\theta}}$ then satisfies

$$
\sqrt{n}(\bar{\hat{\theta}} - \theta) = q^{-1} \sum_{j=1}^{q} \Gamma_j^{-1}Q_j + o_p(1) \Rightarrow \mathcal{N}(0, \bar{\Sigma}_q)
$$

where $\bar{\Sigma}_q = q^{-2} \sum_{j=1}^{q} \Gamma_j^{-1}\Omega_j\Gamma_j'^{-1}$.
**Full Sample Estimator**

- In contrast, full sample GMM estimator $\hat{\theta}$ (with first element $\hat{\beta}$) that solves $n^{-1} \sum_{i=1}^{n} g(\hat{\theta}, y_i)'g(\hat{\theta}, y_i) = 0$ satisfies

\[
\sqrt{n}(\hat{\theta} - \theta) = \left( \sum_{j=1}^{q} \Gamma_j \right)^{-1} \sum_{j=1}^{q} Q_j + o_p(1) \Rightarrow \mathcal{N}(0, \Sigma_q)
\]

where $\Sigma_q = \left( \sum_{j=1}^{q} \Gamma_j \right)^{-1} \left( \sum_{j=1}^{q} \Omega_j \right) \left( \sum_{j=1}^{q} \Gamma_j' \right)^{-1}$.

- Estimator $\hat{\theta}$ not efficient under group heterogeneity. Efficient estimator would exploit $q \times k$ moment conditions $E[g(\theta, y_i)] = 0$ for $i \in G_j, j = 1, \ldots, q$.

- Optimal weighting matrix depends on $\{\Omega_j\}_{j=1}^{q}$, which is assumed difficult to estimate. We thus focus on comparison of t-statistic approach with $\hat{\beta}$ in the numerator with inference based on $\hat{\beta}$ and known $\sigma$ (the (1,1) element of $\Sigma_q$).
General Comparison I

- In general, $\widehat{\Sigma}_q$ and $\Sigma_q$ are not identical. t-statistic approach and inference based on $\hat{\beta}$ with $\sigma^2$ known differ not only in the denominator, but also in the numerator.

- Both tests are consistent against fixed alternatives and have power against the same local alternatives $\beta = \beta_0 + c/\sqrt{n}$. 
General Comparison II

**Theorem:** (i) Let \( \iota \) be the \( k \times 1 \) vector with a one in the first row and zeros elsewhere. Then

\[
\inf_{\{\Gamma_i\}_{i=1}^q, \{\Omega_i\}_{i=1}^q} \frac{\iota' \Sigma_{ql}}{\iota' \overline{\Sigma}_{ql}} = 0 \quad \text{and} \quad \inf_{\{\Gamma_i\}_{i=1}^q, \{\Omega_i\}_{i=1}^q} \frac{\iota' \overline{\Sigma}_{ql}}{\iota' \Sigma_{ql}} = \begin{cases} 1/q^2 & \text{if } k = 1 \\ 0 & \text{if } k \geq 2 \end{cases}
\]

(ii) For any sequence of full rank matrices \( \{\Gamma_i\}_{i=1}^q \) there exists a positive definite sequence \( \{\overline{\Omega}_j\}_{j=1}^q \) so that \( \Sigma_q - \overline{\Sigma}_q \) is positive semidefinite for \( \{\Omega_j\}_{j=1}^q = \{\overline{\Omega}_j\}_{j=1}^q \), and for any sequence of symmetric positive definite matrices \( \{\Gamma_i\}_{i=1}^q \) there exists a positive definite sequence \( \{\Omega_j\}_{j=1}^q \) so that \( \Sigma_q - \overline{\Sigma}_q \) is negative semidefinite for \( \{\Omega_j\}_{j=1}^q = \{\Omega_j\}_{j=1}^q \).

(iii) If \( \Gamma_i = \Gamma \) for \( i = 1, \ldots, q \), then \( \overline{\Sigma}_q = \Sigma_q \) for all \( \{\Omega_j\}_{j=1}^q \).
Special Case

- Consider the special case $\Gamma_j = \Gamma$ for $j = 1, \cdots, q$. This naturally arises when the groups have an equal number of observations $n/q$, and the average of the derivative of the moment condition is homogenous across groups (leading example: i.i.d. data). Then

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\hat{\theta} - \theta) + o_p(1)$$

and $\hat{\beta}$ and $\bar{\beta}$ are asymptotically equivalent to order $\sqrt{n}$.

- The asymptotic local power of tests based on $t_\beta$ and $\hat{\beta}$ with $\sigma^2$ known simply reduces to the small sample power of the small sample t-statistic and $z$-statistic of $H_0 : \mu = 0$ when $X_i \sim \mathcal{N}(\mu, \sigma_i^2)$, that is

$$z = \frac{\sum_{i=1}^{q} X_i}{\sqrt{\sum_{i=1}^{q} \sigma_i^2}}$$

where $\sigma_i^2$ is the $(1,1)$ element of $\Gamma^{-1}\Omega_i\Gamma'^{-1}$. 
Numerical Comparison

\[ \frac{q}{2} \text{ observations of relative variance } a^2 \]

\[ \text{a=1} \]

\[ \text{a=2} \]

\[ \text{a=2} \]

\[ \text{a=5} \]

\[ \frac{\mu \sqrt{q}}{1}  \]

\[ \text{t-statistic, } q=4 \]

\[ \text{t-statistic, } q=8 \]

\[ \text{t-statistic, } q=16 \]

\[ z\text{-statistic} \]
Applications

- t-statistic approach requires

$$\sqrt{n}(\hat{\beta}_1 - \beta, \ldots, \hat{\beta}_q - \beta)' \Rightarrow \mathcal{N}(0, \text{diag}(\sigma^2_1, \ldots, \sigma^2_q)).$$

1. **Panel Data**

2. **Time Series** Data

3. **Inequality, poverty, risk & concentration** measures
   - Small sample results
   - Large sample inference
Panel Data

- Panel with potential time series correlation and few individuals $N$

$$y_{i,t} = x_{i,t}'\theta + u_{i,t}, \text{ } i = 1, \cdots, N, \text{ } t = 1, \cdots, T$$

where $\{x_{i,t}, u_{i,t}\}_{t=1}^{T}$ are independent across $i$ and $E[x_{i,t}u_{i,t}] = 0$ for all $i, t$.

- t-statistic approach asymptotically valid if $T^{-1}\sum_{t=1}^{T} x_{i,t}x_{i,t}' \xrightarrow{p} \Gamma_i$ and $T^{-1/2}\sum_{t=1}^{T} x_{i,t}u_{i,t} \Rightarrow \mathcal{N}(0, \Omega_i)$ for all $i$ as $T \to \infty$ and $N$ fixed for some full rank matrices $\Gamma_i$ and $\Omega_i$.

- Hansen (2005) shows that usual t-statistic with Rogers (1993) standard errors converges under the null to a scaled t-statistic with $q-1$ degrees of freedom under 'asymptotic homogeneity', i.e. when $\Gamma_i = \Gamma$ and $\Omega_i = \Omega$ for all $i$. 
Panel Data II

• For applications in Finance, concern about cross section correlation. Our results justify Fama–MacBeth method where regression is run cross sectionally for each time period, and inference is based on $t$-statistic of resulting estimators $\hat{\beta}_j$, $j = 1, \cdots, T$, even for small $T$ and potential heterogeneity of variances as long as no correlation across $t$.

• For corporate Finance applications, uncorrelatedness in time is often implausible. Rather than to try to consistently estimate the long-run variance with few observations, the approach taken here suggests forming groups of more than one unit in time to achieve approximate independence.

• Alternatively, assume independence in cross section dimension, say, across industries, as in Froot (1989). Combinations are possible.

• Same possibilities for long-run event studies, country panel data, city panel data and so forth.
Monte Carlo Results

Same design as in Thompson (2006): Linear Regression, one nonconstant regressor, \(N = 50\), \(T = 25\).

"individual persistence": \(u_{i,t} = \xi_t + \eta_{i,t}, \quad \eta_{i,t} = \rho \eta_{i,t-1} + \varepsilon_{i,t}\)

"common persistence": factor structure \(u_{i,t} = h_{i} f_{t} + \varepsilon_{i,t}, \quad f_{t} = \rho f_{t-1} + \xi_t, \quad h_{i} \sim N(1, 0.25)\)

<table>
<thead>
<tr>
<th>(\rho)</th>
<th>Individual Persistence</th>
<th>Common Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic (q = 2)</td>
<td>4.9</td>
<td>5.0</td>
</tr>
<tr>
<td>t-statistic (q = 4)</td>
<td>4.9</td>
<td>5.4</td>
</tr>
<tr>
<td>t-statistic (q = 8)</td>
<td>4.6</td>
<td>6.4</td>
</tr>
<tr>
<td>FM with Newey-West</td>
<td>12.6</td>
<td>19.8</td>
</tr>
<tr>
<td>cluster by (i) and (t)</td>
<td>9.3</td>
<td>8.8</td>
</tr>
<tr>
<td>cluster by (i) and (t + cp)</td>
<td>16.3</td>
<td>14.9</td>
</tr>
<tr>
<td>Size Adjusted Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic (q = 2)</td>
<td>12.9</td>
<td>16.2</td>
</tr>
<tr>
<td>t-statistic (q = 4)</td>
<td>30.5</td>
<td>45.5</td>
</tr>
<tr>
<td>t-statistic (q = 8)</td>
<td>50.9</td>
<td>67.6</td>
</tr>
<tr>
<td>FM with Newey-West</td>
<td>100</td>
<td>91.6</td>
</tr>
<tr>
<td>cluster by (i) and (t)</td>
<td>46.8</td>
<td>67.6</td>
</tr>
<tr>
<td>cluster by (i) and (t + cp)</td>
<td>31.7</td>
<td>52.7</td>
</tr>
</tbody>
</table>
Time Series Data

- In absence of more specific knowledge, exploit the default assumption that correlations between observations become weaker the further apart in time they are: Divide the sample of size $T$ into $q$ (approximately) equal sized groups of consecutive observations.

- Under a wide range of assumptions on the underlying model and observations, exactly identified GMM inference satisfies

$$\sup_{0 \leq r \leq 1} \left\| T^{-1} \sum_{t=1}^{[rT]} \frac{\partial g(a, y_t)}{\partial a} \right\|_{a=\hat{\theta}} - \int_0^r \Gamma(\lambda)d\lambda \| \overset{p}{\to} 0 \quad (1)$$

$$T^{-1/2} \sum_{t=1}^{[\cdot T]} g(\theta, y_t) \Rightarrow \int_0^r h(\lambda)dW(\lambda) \quad (2)$$

where $\Gamma(\cdot)$ is a positive definite $k \times k$ matrix function and $h(\cdot)$ is nonzero.
Time Series Data

- With that convergence, t-statistic approach is asymptotically valid, since

\[
\sqrt{T} \begin{pmatrix}
\hat{\theta}_1 - \theta \\
\hat{\theta}_2 - \theta \\
\vdots \\
\hat{\theta}_q - \theta 
\end{pmatrix} \Rightarrow \begin{pmatrix}
\left( \int_0^{1/q} \Gamma(\lambda)d\lambda \right)^{-1} \int_0^{1/q} h(\lambda)dW(\lambda) \\
\left( \int_1^{2/q} \Gamma(\lambda)d\lambda \right)^{-1} \int_1^{2/q} h(\lambda)dW(\lambda) \\
\vdots \\
\left( \int_{(q-1)/q}^{1} \Gamma(\lambda)d\lambda \right)^{-1} \int_{(q-1)/q}^{1} h(\lambda)dW(\lambda)
\end{pmatrix}
\]

- In contrast, no other known way of conducting asymptotically valid inference under (1) and (2):
  
  - Kiefer and Vogelsang (2002, 2005) approach requires \(\Gamma(\cdot)\) and \(h(\cdot)\) to be constant
  
  - Müller (2004) shows that no long-run variance estimator can be consistent for \(\text{Var}[\int_0^1 h(\lambda)dW(\lambda)]\) for all processes that satisfy (2)
Monte Carlo Results

Same design as in Andrews (1991): Linear Regression, 5 regressors, 4 nonconstant regressors are independent draws from stationary Gaussian AR(1), as are the disturbances, + heteroskedasticity. $T = 128$, 5% level test about coefficient of one nonconstant regressor.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$t$-statistic $(q)$</th>
<th>$\hat{\omega}_Q^2$</th>
<th>$\hat{\omega}_P^2$</th>
<th>$\hat{\omega}_B^2(b)$</th>
<th>Size</th>
<th>Size Adjusted Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>0.05</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>0</td>
<td>4.9</td>
<td>4.7</td>
<td>4.6</td>
<td>7.1</td>
<td>8.1</td>
<td>6.7</td>
</tr>
<tr>
<td>0.5</td>
<td>4.8</td>
<td>4.6</td>
<td>4.6</td>
<td>10.4</td>
<td>9.9</td>
<td>9.4</td>
</tr>
<tr>
<td>0.8</td>
<td>4.8</td>
<td>4.9</td>
<td>5.4</td>
<td>19.1</td>
<td>17.3</td>
<td>18.6</td>
</tr>
<tr>
<td>0.9</td>
<td>4.9</td>
<td>5.1</td>
<td>6.1</td>
<td>28.9</td>
<td>25.4</td>
<td>29.9</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18
**Heavy-tailed inference**

\[ y_t = \beta + \sum_{j=-\infty}^{\infty} \psi_j S_{t-j}, \]

\( S_t \sim \text{i.i.d. stable domain of attraction}, \alpha \in (1, 2] \)

Subsampling: works under strong mixing for \( y_t \), extensions for GARCH (McElroy and Politis, 2002)

\( t \)-statistic approach: symmetric stable (scale mixtures of normals)
Differences:

**Subsampling**: first calculate block $t$–statistic, then approximate cdf of full sample $t$–statistic by empirical cdf’s

**Our approach**: first calculate estimators over blocks, then form their $t$–statistic to conduct asymptotically valid inference
Monte Carlo Results

\[ y_t = \beta + u_t, \quad S_t : \text{symm. stable with index } \alpha \in (0, 2) \]

AR: \[ u_t = 0.5 u_{t-1} + S_t, \]

MA: \[ u_t = \sum_{j=0}^{10} \psi_j S_{t-j}, \]

\{\psi_j\}_{j=0}^{10} = \{.03, .05, .07, .1, .15, .2, .15, .1, .07, .05, .03\}
Robust inequality measurement

(on-going work with Paul Kattuman, University of Cambridge)

Economic & financial data: correlated, heterogeneous or heavy-tailed

- **Income, wealth & financial returns:** heterogeneity, dependence & outliers

- **Heterogeneous observations & outliers** ⇐ **heavy-tailed** power laws \( P(I > x) \sim \frac{C}{x^\zeta} \)

- **Tail index** \( \zeta \): heavy-tailedness, likelihood of extreme observations & outliers
  - **Income:** \( 1.5 < \zeta < 3 \); **wealth:** \( \zeta \approx 1.5 \)
  - **Infinite variance** ⇒ Failure of standard OLS!
  - **Problematic** inference on income & wealth inequality
  - **Tail index** \( \zeta \Rightarrow** inequality measures: **Gini** \( G = 1/(2\zeta - 1) \) (inequality in upper tails)
  - Similar relations: Singh-Maddala power law distributions, log-normal

- **Crises, emerging markets**: more volatile, more extreme shocks ⇒ more pronounced heterogeneity, outliers & heavy tails (e.g., IIK, 2009: income & wealth in CIS)

- **Financial returns**: \( 2 < \zeta < 4 \Rightarrow** finite variance, infinite fourth moment
A tale of two tails

Simulated data from Normal, Cauchy and Levy distributions, n=25

Normal
Cauchy $\alpha=1$
Levy $\alpha=1/2$

Figure: Heavy-tailed distributions: more extreme observations
Figure: Tails of Cauchy distributions are heavier than those of normal distributions. Tails of Lévy distributions are heavier than those of Cauchy or normal distributions.
Robust inequality measurement

- Many inequality measures: problematic under heterogeneity, heavy-tailedness & correlation

- Inequality, poverty & concentration measures: very sensitive to extremes & heavy tails (GE, Theil, MLD, Gini, HHI)


- Robustness of measure choice, estimation & inference
  - Asymptotic, standard & non-standard (moon) bootstrap
  - Semiparametric bootstrap & asymptotic methods ⇒ Improvement
  - Computationally expensive (bootstrap), problematic s.e.’s (asymptotic)

- Many studies: only point estimates of inequality
  - Conclusions may need to be modified: standard errors & statistical significance
\textbf{t-statistic based robust inference} \\

- Inference on income inequality measure $I$ (Gini or Theil)

- Data: heavy tails, heterogeneity or dependence

- $H_0 : I = I_0$ against the alternative $H_a : I \neq I_0$

- $t$–statistic based robust test of level $\alpha = 5\%$

  1. \textbf{Partition} sample $I_1, I_2, \ldots, I_n$ of observations on incomes into $q \geq 2$ groups

  2. \textbf{Estimate income inequality} measure $I$ for each group $\Rightarrow$ \textbf{empirical inequality} measures $\hat{I}_j, j = 1, \ldots, q$

  3. Compute the \textbf{usual $t$–statistic} $t_I = \sqrt{q \frac{\bar{I} - I_0}{\hat{s}_I}}$ with $\bar{I} = q^{-1} \sum_{j=1}^{q} \hat{I}_j$ and $s_{\hat{I}}^2 = (q - 1)^{-1} \sum_{j=1}^{q} (\hat{I}_j - \bar{I})^2$ (t–statistic treating $q$ group estimators as observations)

  4. \textbf{Reject} $H_0$ if $t_R$ exceeds the 97.5–percentile of \textbf{Student–t distribution} $T_{q-1}$ with $q - 1$ degrees of freedom

  5. 95\% \textbf{confidence interval} for unknown $I : (\bar{I} - \tau_{0.05} s_{\hat{I}}, \bar{I} + \tau_{0.05} s_{\hat{I}})$ where $P(|T_{q-1}| > \tau_{0.05}) = 0.05$
Connection to inequality measures

- Ex.: coefficient of variation for logs $Y_j = \log X_j$ of $X_1, \ldots, X_q$

- $\tilde{Y}_1, \ldots, \tilde{Y}_q$: i.i.d. standard normal r.v.’s: $\tilde{X}_j \sim \mathcal{N}(0,1)$

- $P(|T_{q-1}| > \tau_{0.05}) = 0.05$

- Empirical coefficient of variation $\hat{CV}_X = \frac{s_X}{\bar{X}} = \sqrt{q}/t$
  - Sharpe ratio, Herfindahl-Hirschman Index

- Analogues for logs

- If $Y_j \sim \mathcal{N}(0, \sigma_j^2)$ or scale mixtures of normals (stable, etc.) $\Rightarrow$
  
  \[ P(0 < CV_Y < y) \leq P(0 < CV_{\tilde{Y}} < y) \] for $y < 1/(\tau_{0.05} \sqrt{q})$

- In general, does not hold for $y < 1/(\tau_{0.1} \sqrt{q})$

- Homogeneity and thin-tails (normality): Reduce the inequality in region of their small values

- Does not not hold $\Rightarrow$ poor measure of inequality for medium & high
$t$–statistic based robust inference

- Asymptotically valid & efficient if empirical inequality measures $\hat{\mathcal{I}}_j, j = 1, \ldots, q$:
  asymptotically independent, unbiased & Gaussian of possibly different variances

- Standard results on CLT for empirical income inequality measures (Gini, Theil, etc.):
  $\frac{1}{\sqrt{n}}(\hat{\mathcal{I}}_j - \mathcal{I}) \to_d \mathcal{N}(0, \sigma_j^2)$
  - **Same conditions** as for empirical inequality measure $\hat{\mathcal{I}}$ over full sample $I_1, I_2, \ldots, I_N$:
    $\frac{1}{\sqrt{N}}(\hat{\mathcal{I}} - \mathcal{I}) \to_d \mathcal{N}(0, \sigma^2)$

- Asymptotic validity: also holds when group estimators $\hat{\mathcal{I}}_j$ converge (at arbitrary rate) to mixed normal
  - Symmetric stable
  - **Conditionally normal** variates, dependence through second moments
  - **Heavy tails, extremes** and outliers
  - **Dependent** models with common shocks
$t-$statistic based robust inference

First applications in inequality measurement

Assumptions: Asymptotic normality & independence for group estimators

- **Asymptotic normality:** From usual CLT for (full) sample
- **Many applications:** Natural choice of groups for asymptotic independence
  - **Time series:** Divide into $q = 4$ consecutive blocks & estimate model 4 times
  - **Panel data, cross section correlation:** Justification of Fama-MacBeth
    * Regression is run cross sectionally for each time period & inference using $t-$statistic if no correlation across $t$
    * Other combinations
- **Income & wealth:** spatial dependence, dependence across states, common shocks, crises, etc.
  - Groups by geographical regions ⇒ asymptotically correlated estimators
- **Groups** to achieve asymptotic independence of group estimators?
- **Independence** condition: methodological problems for direct applications
Asymptotic independence through randomization

- **Solution:** randomization of initial samples $I_1, I_2, ..., I_N$ of income observations
  
  - Randomization stage: randomly assign each observation $I_i$, $i = 1, ..., N$, to one of $q$ groups $j = 1, ..., q$ with equal probability $1/q$
  
  - $t$–statistic stage: Apply $t$–statistic approach to $q$ groups of consecutive observations in randomized sample of incomes $\tilde{I}_1, \tilde{I}_2, ..., \tilde{I}_N$

  As in time series applications: observation $\tilde{I}_i$ in randomized sample is element of group $j$ if $(j - 1)N/q < i \leq jN/q$

- **Group estimators** for randomized sample: independent by construction

- **Both conditions** for validity of $t$–statistic inference: satisfied

- **Randomization-based** modification of $t$–statistic robust inference: may also prove to be useful in other problems with natural dependence for group estimators
### Risk, Inequality & Concentration: MC Results

Cowell & Flachaire (2007): Theil measure $\mathcal{I} = \left( \frac{E[Y^\alpha]}{(E[Y])^\alpha} - 1 \right) / (\alpha(\alpha - 1))$

Singh-Maddala: $F(y) = 1 - (1 + ay^b)^{-c}$, $a = 100$, $b = 2.8$, $c = 1.7 \Rightarrow$ tail index $\zeta = bc = 4.8$, $\mathcal{I} = 0.14$

Pareto: $F(y) = 1 - (y_l/y)^\zeta$, $y_l = 0.1$, $\zeta = 2.5 \Rightarrow \mathcal{I} = 1/(\zeta - 1) + \log((\zeta - 1)/\zeta) = 0.16$

Lognormal: $\log(Y) \sim N(\mu, \sigma^2)$, $\mu = -2$, $\sigma = 1 \Rightarrow \mathcal{I} = \sigma^2/2 = 0.5$

<table>
<thead>
<tr>
<th>$N$</th>
<th>Singh-Maddala</th>
<th>Pareto</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>1000</td>
<td>500</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic $q = 2$</td>
<td>5.4</td>
<td>4.8</td>
<td>7.2</td>
</tr>
<tr>
<td>t-statistic $q = 4$</td>
<td>6.9</td>
<td>6.4</td>
<td>16.6</td>
</tr>
<tr>
<td>Asymptotic</td>
<td>11.5</td>
<td>10.5</td>
<td>29.4</td>
</tr>
<tr>
<td>Standard bootstrap</td>
<td>8.1</td>
<td>7.9</td>
<td>17.8</td>
</tr>
<tr>
<td>Moon bootstrap</td>
<td>6.9</td>
<td>7.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Semiparametric bootstrap</td>
<td>6.4</td>
<td>5.9</td>
<td>10.6</td>
</tr>
<tr>
<td>Asymptotic test with semiparametric measure</td>
<td>6.6</td>
<td>6.3</td>
<td>11.6</td>
</tr>
<tr>
<td>% of cases semipar. measure cannot be computed</td>
<td>0</td>
<td>0</td>
<td>8.9%</td>
</tr>
</tbody>
</table>
Cowell & Flachaire (2007): Mean logarithmic deviation $\mathcal{I} = \log(E[Y]) - E[\log(Y)]$

Singh-Maddala: $F(y) = 1 - (1 + ay^b)^{-c}$, $a = 100$, $b = 2.8$, $c = 1.7 \Rightarrow$ tail index $\zeta = bc = 4.8$, $\mathcal{I} = 0.15$

Pareto: $F(y) = 1 - (y_l/y)^\zeta$, $y_l = 0.1$, $\zeta = 2.5 \Rightarrow \mathcal{I} = -1/\zeta - \log((\zeta - 1)/\zeta) = 0.11$

Lognormal: $\log(Y) \sim N(\mu, \sigma^2)$, $\mu = -2$, $\sigma = 1 \Rightarrow \mathcal{I} = \sigma^2/2 = 0.5$

<table>
<thead>
<tr>
<th></th>
<th>Singh-Maddala</th>
<th>Pareto</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>500</td>
<td>1000</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic $q = 2$</td>
<td>5.3</td>
<td>5.3</td>
<td>6.2</td>
</tr>
<tr>
<td>t-statistic $q = 4$</td>
<td>6.3</td>
<td>5.5</td>
<td>10.2</td>
</tr>
<tr>
<td>Asymptotic</td>
<td>8.1</td>
<td>7.4</td>
<td>19.3</td>
</tr>
<tr>
<td>Standard bootstrap</td>
<td>5.7</td>
<td>5.7</td>
<td>11.7</td>
</tr>
<tr>
<td>Moon bootstrap</td>
<td>4.3</td>
<td>4.9</td>
<td>5.5</td>
</tr>
<tr>
<td>Semiparametric bootstrap</td>
<td>5.5</td>
<td>5.2</td>
<td>8.7</td>
</tr>
<tr>
<td>Asymptotic test with semiparametric measure</td>
<td>5.1</td>
<td>5.2</td>
<td>12.4</td>
</tr>
<tr>
<td>% of cases semipar. measure cannot be computed</td>
<td>0</td>
<td>0</td>
<td>8.9%</td>
</tr>
</tbody>
</table>
Robust inference for risk, inequality, poverty & concentration: $t$-statistic based approach vs. alternative procedures

Singh-Maddala: $P(X > y) = (1 + ay^b)^{-c}$, $a=100$, $b=2.8$, $c=1.7 \Rightarrow \xi = bc = 4.8$
Tail index estimation & inequality in upper tails

- **Inequality in upper tails**, applications of semiparametric asymptotic and bootstrap inequality inference: *Estimates of tail index* $\zeta$ in *power laws*
  - Hill’s estimator
  - OLS log-log rank-size: optimal shifts & correct s.e.’s, Gabaix & Ibragimov (11)

- $I_1 \geq I_2 \geq \ldots \geq I_n \geq I_{n+1}$: *power law* $P(I > x) \sim \frac{C}{x^\zeta}$

- **OLS approach**: regression $\log(t) = a - b \cdot \log(I(t)) \iff \log(\text{Rank}) = a - b \cdot \log(\text{Size})$
  - $b$ : estimate of *tail index*: Simple & robust
  - Biased in small samples

- One should use the $\text{Rank} - 1/2$, & run $\log(\text{Rank} - 1/2) = a - b \cdot \log(\text{Size})$
  - Shift of $1/2$: optimal, and reduces bias to a leading order
  - S.e. on $\zeta$: not OLS s.e., but is asymptotically $\sqrt{2\zeta}/\sqrt{n}$
  - **Numerical results**: advantage over standard OLS estimation procedures
  - Performs well under dependence & deviations from power laws
Conclusion

- New approach to correlation and heterogeneity robust inference in large samples.

- Method imposes only a 'finite amount of independence' through the assumption that estimators from different groups are independent.

- Approach exploits this assumption in an efficient way. Many potential applications, encouraging Monte Carlo results.

- Challenge to choose groups in practice. But inference requires some assumption on correlation structure, and other methods make more implicit and even less interpretable assumptions.