Correlation and Heterogeneity Robust Inference using Conservativeness of Test Statistic

Rustam Ibragimov
Harvard University

Ulrich K. Müller
Princeton University

Econometrics Seminar, Boston University
November 9, 2011

Introduction

- Most economic data is observational

  $\Rightarrow$ 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
- Newey and West (1987), Andrews (1991), Davidson and de Jong (2000) and others for time series disturbances
- Rogers (1993) and Arellano (1987) for clustered and panel data
- Conley (1999) for spatially correlated data

$\Rightarrow$ based on a Law of Large Numbers, and thus require "an infinite amount of independence"

$\Rightarrow$ 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), Müller (2006)
  - panel data: Donald and Lang (2004), Hansen (2005)

⇒ 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 β : $t-$**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 β, so that $\hat{\beta}_j \overset{a}{\sim} id\mathcal{N}(\beta, v_j^2)$ 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 $\hat{\beta}_j \overset{a}{\sim} id\mathcal{N}(\beta, v_j^2)$ 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 β_1 and β_2 with q_1 and q_2 groups: $\hat{\beta}_{ij} \overset{a}{\sim} id\mathcal{N}(\beta_i, v_{ij}^2)$ for $j = 1, \cdots, q_i,\ i = 1, 2$.

- 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 $\hat{\beta}_{ij} \overset{a}{\sim} id\mathcal{N}(\beta_i, v_{ij}^2)$ in an efficient way.

Plan of Talk

1. Introduction
2. The Small Sample t-Test
   (a) Conservativeness for Heterogenous Variances
   (b) Optimality
3. t-Statistic Based Large Sample Robust Inference
   (a) Basic Idea and Properties
   (b) Comparison with Standard Inference and Known Asymptotic Variance
4. Applications: Panel data, Time series, Spatial correlation, Risk, Inequality & Concentration Measures
5. 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** $\alpha = 10\%$)

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

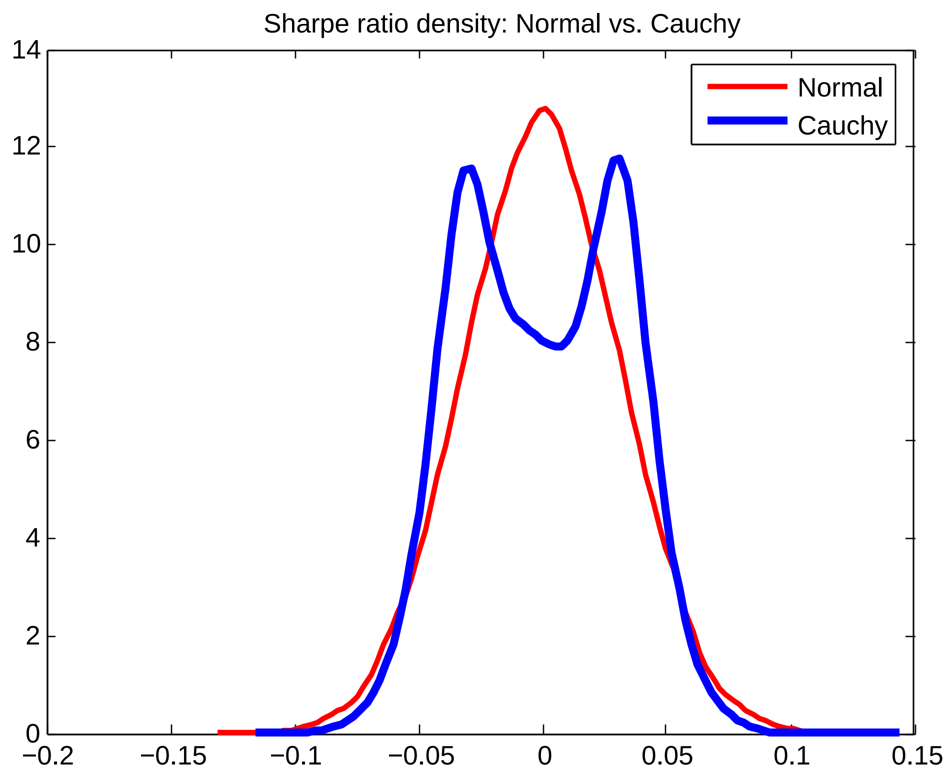
  t-statistic $t = \sqrt{q}\frac{\bar{X}}{s_X}$

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

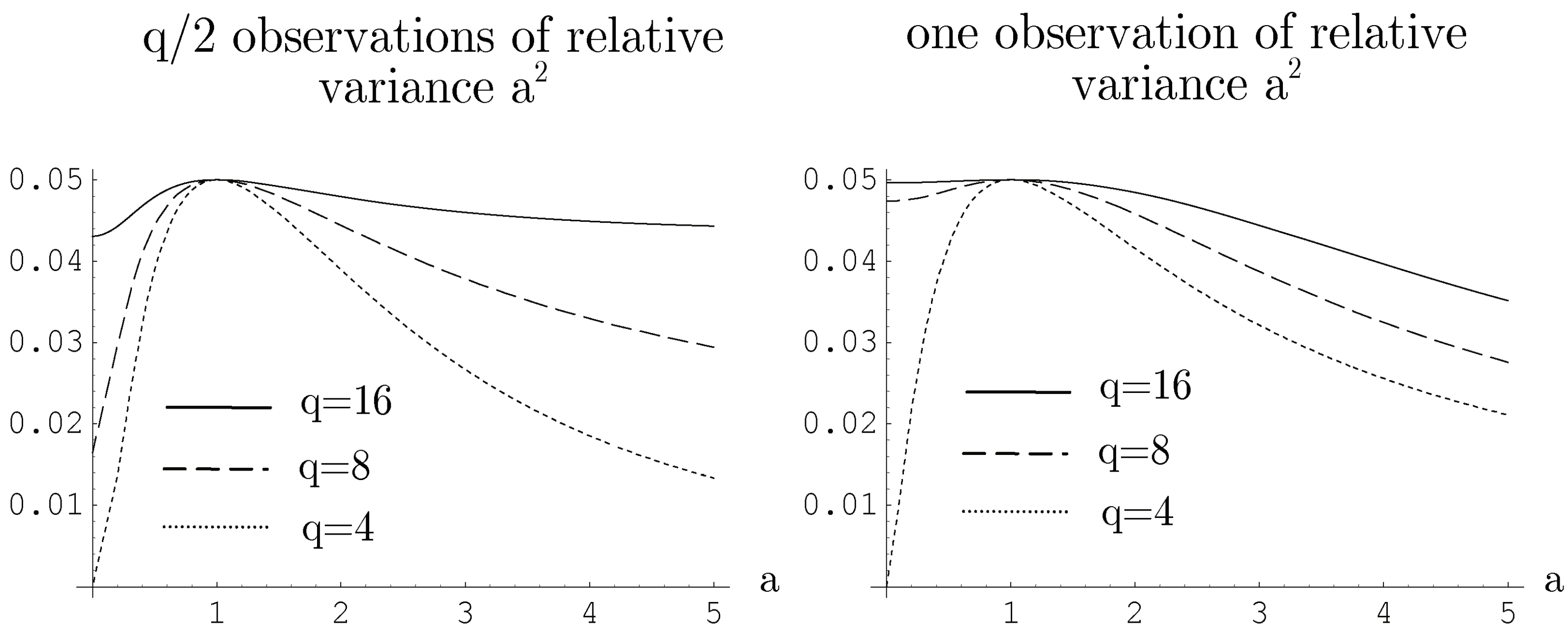
  $cv_q(\alpha) =$ critical value of $T_{q-1} : P(|T_{q-1}| > \mathrm{cv}_q(\alpha)) = \alpha$

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

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


Sharpe ratio density: Normal vs. Cauchy
Normal
Cauchy
14
12
10
8
6
4
2
0
−0.2
−0.15
−0.1
−0.05
0
0.05
0.1
0.15

Effective Rejection Probabilities



Optimality of t-Statistic

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

$$\begin{aligned} H_0 &: \mu = 0 \text{ and } \{\sigma_j^2\}_{j=1}^q \text{ arbitrary} \\ H_1 &: \mu \neq 0 \text{ and } \sigma_j^2 = \sigma^2 \text{ for all } j \end{aligned}$$

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

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

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_β is the usual t-statistic

$$t_\beta = \sqrt{q} \frac{\overline{\hat{\beta}} - \beta_0}{s_{\hat{\beta}}}$$

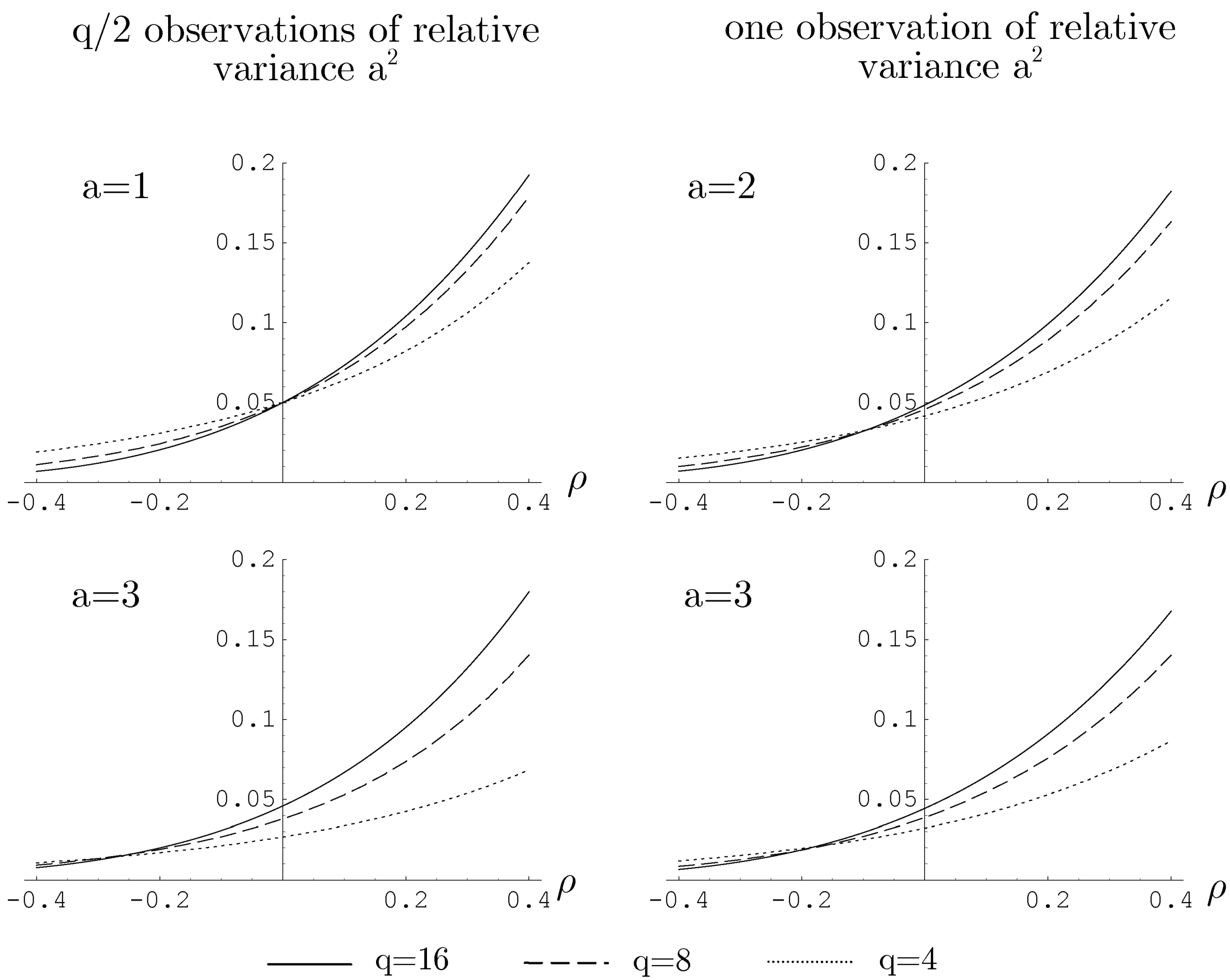
with $\overline{\hat{\beta}} = q^{-1} \sum_{j=1}^{q} \hat{\beta}_j$ and $s^2_{\hat{\beta}} = (q-1)^{-1} \sum_{j=1}^{q} (\hat{\beta}_j - \overline{\hat{\beta}})^2$, is asymptotically valid inference by small sample t-test result and the Continuous Mapping Theorem.

- 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.

Size Control under AR(1) Correlation



Comparison with Inference with Known Asymptotic Variance

- Typically,

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

  is weaker than what is required to consistently estimate the asymptotic variance.

- 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. We are interested in conducting inference about the first element β of the GMM vector parameter θ.

General Comparison

- Numerator of t-statistic is average of q GMM estimators, each using data from group j only. Assume that GMM estimators of different groups are asymptotically Gaussian and independent.

- Compare to inference based on usual full sample GMM estimator with asymptotic variance known. Not efficient under group heterogeneity, but generally efficient GMM estimator requires knowledge of the asymptotic variances in each group, which is assumed difficult to estimate.

- Both tests are consistent against fixed alternatives and have power against the same local alternatives.

- (Almost) nothing else can be said in general.

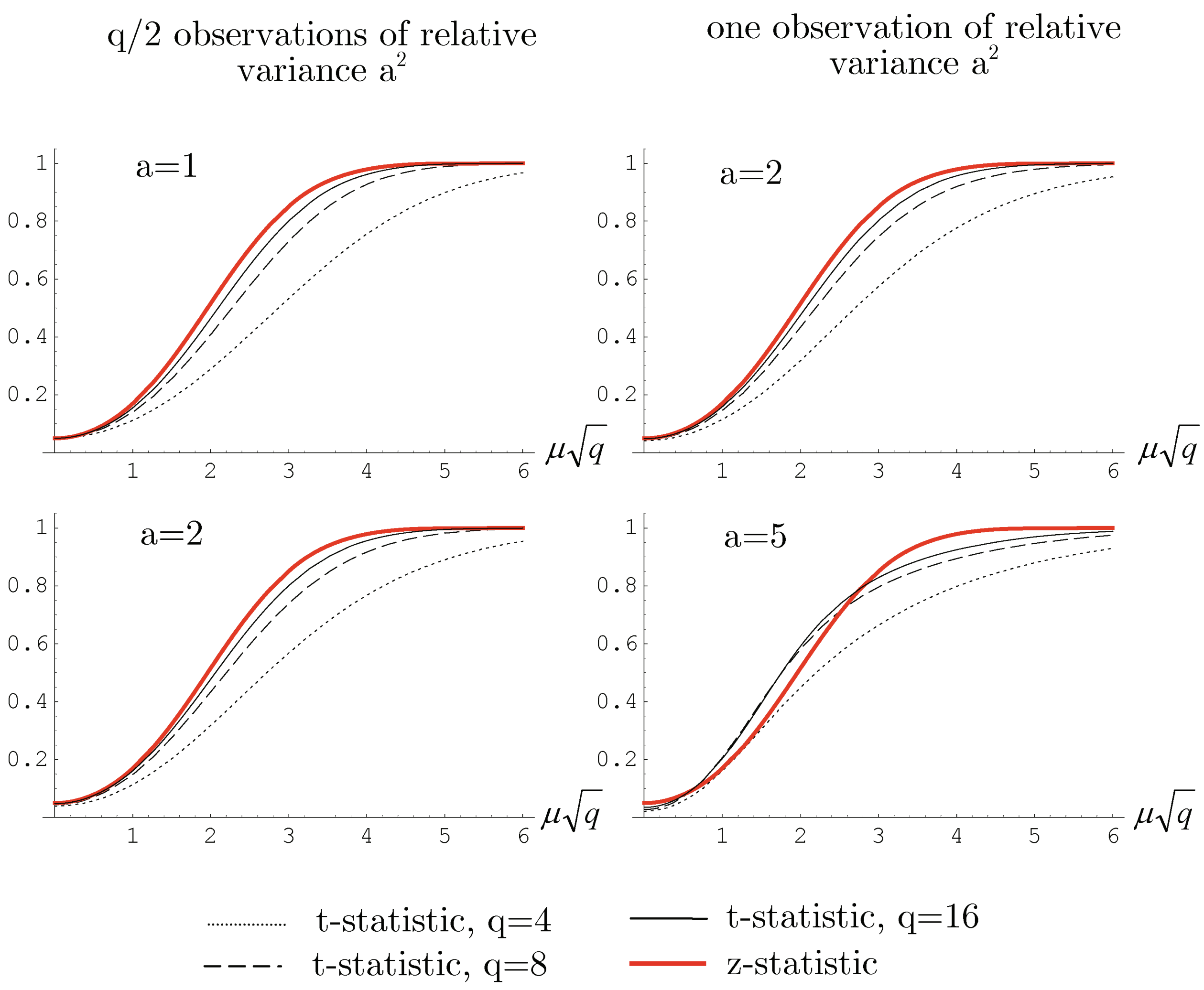
Special Case

- Consider the special case where the derivative of the moment condition averages to the same matrix in all groups. 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 full sample GMM estimator and group average estimator become asymptotically equivalent to order $\sqrt{n}$.

- The asymptotic local power comparison between tests based on t_β and tests based on full sample GMM estimator with known asymptotic variance simply reduces to the small sample power comparison between a t-statistic and a z-statistic with potentially heterogenous variances.

Numerical Comparison



Applications

- t-statistic approach requires

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

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

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

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 Γ_i and Ω_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": $u_{i,t} = \xi_t + \eta_{i,t}$, $\eta_{i,t} = \rho\eta_{i,t-1} + \varepsilon_{i,t}$

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

	Individual Persistence			Common Persistence		
ρ	0	0.7	0.9	0	0.7	0.9
			Size			
t-statistic $q = 2$	4.9	5.0	6.0	4.9	5.3	6.3
t-statistic $q = 4$	4.9	5.4	9.8	4.1	5.3	10.4
t-statistic $q = 8$	4.6	6.4	17.1	3.9	7.1	16.8
FM with Newey-West	12.6	19.8	34.8	11.4	14.2	23.4
cluster by i and t	9.3	8.8	7.0	10.2	29.9	49.5
cluster by i and t + cp	16.3	14.9	12.1	17.0	26.4	38.3
			Size Adjusted Power			
t-statistic $q = 2$	12.9	16.2	14.9	20.3	18.1	20.8
t-statistic $q = 4$	30.5	45.5	45.3	58.4	58.4	60.6
t-statistic $q = 8$	50.9	67.6	61.3	59.5	67.3	68.2
FM with Newey-West	100	91.6	58.9	57.3	47.4	47.1
cluster by i and t	46.8	67.6	74.8	86.3	66.2	70.2
cluster by i and t + cp	31.7	52.7	69.8	69.6	53.6	60.1

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

$$\sup_{0\leq r\leq 1} ||T^{-1}\sum_{t=1}^{\lfloor rT\rfloor} \frac{\partial g(a,y_t)}{\partial a}|_{a=\hat{\theta}} - \int_0^r \Gamma(\lambda)d\lambda|| \xrightarrow{p} 0 \qquad (1)$$

$$T^{-1/2}\sum_{t=1}^{\lfloor \cdot T\rfloor} g(\theta,y_t) \Rightarrow \int_0^{\cdot} h(\lambda)dW(\lambda) \qquad (2)$$

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

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

- 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 $\mathrm{Var}[\int_0^1 h(\lambda) dW(\lambda)]$ 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.

	t-statistic (q)			$\hat{\omega}^2_{QA}$	$\hat{\omega}^2_{PW}$	$\hat{\omega}^2_{BT}(b)$			
	2	4	8			0.05	0.1	0.3	1
ρ				Size					
0	4.9	4.7	4.6	7.1	8.1	6.7	6.6	6.0	6.2
0.5	4.8	4.6	4.6	10.4	9.9	9.4	8.4	7.5	7.0
0.8	4.8	4.9	5.4	19.1	17.3	18.6	15.6	12.8	11.9
0.9	4.9	5.1	6.1	28.9	25.4	29.9	24.9	20.5	18.8
ρ				Size Adjusted Power					
0	15.1	38.4	53.7	62.7	60.6	60.7	58.6	51.9	47.2
0.5	14.5	38.2	55.9	57.0	56.2	56.0	53.5	48.4	44.2
0.8	15.4	45.1	66.0	52.9	51.7	54.0	52.6	46.9	42.4
0.9	17.2	56.7	77.6	57.5	54.6	58.7	57.5	51.4	46.6

Heavy-tailed inference

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

$S_t \sim$ 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

$y_t = \beta + u_t$, S_t : symm. stable with index $\alpha \in (0, 2)$

AR: $u_t = 0.5u_{t-1} + S_t$,

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

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

				$t-$stat. (q)			subs. $t-$stat. (b)		
				2	4	8	2	4	8
DGP	α	T	β	Size					
MA	1.8	100	0	4.9	5.0	6.7	0.0	0.1	1.6
MA	1.8	1000	0	4.9	4.7	5.0	0.1	0.1	0.2
AR	1.8	100	0	4.7	4.4	5.3	0.0	1.1	6.5
AR	1.8	1000	0	4.7	4.6	4.8	0.1	0.3	1.9
DGP	α	T	β	Non-Size Adjusted Power					
MA	1.8	100	0.4	14.2	38.4	58.9	0.0	2.9	20.6
MA	1.8	1000	0.2	19.0	56.8	76.0	0.1	0.1	15.8
AR	1.8	100	0.8	14.2	37.7	55.4	0.1	24.1	51.2
AR	1.8	1000	0.4	18.7	57.2	75.6	0.1	34.6	70.0

The subsampled $t-$statistic rejects if full sample OLS $t-$statistic falls outside 2.5% and 97.5% quantiles of the empirical cdf of $t-$statistics for subsamples of size b.

Heterogeneity, dependence & heavy tails

- Data in many **economic**, **financial** and **insurance** markets:
  - **Heterogeneous**, **dependent** & **heavy tailed**
  - **Financial returns**, **exchange rates**: Conditional & time-varying **volatility**, **heavy tails**
  - **Income** & **wealth**: **heterogeneity** in subgroups & countries
- Many **risk, inequality, poverty** & **concentration** measures: very **sensitive** to **extremes**, **heavy tails** & **dependence**
  - Value at risk (VaR), expected shortfall, Sharpe ratio (SR)
  - Gini, generalized entropy (GE), Theil, MLD, poverty gap & squares
  - Herfindahl-Hirschman index (HHI)

Heavy tails: empirics

- **Heavy-tails** & **power** laws: $P(X > x) \sim x^{-\zeta}$
- Financial **returns**: $2 < \zeta < 4$: **finite** σ^2 & **infinite** fourth moments
- $\zeta \approx 3$ (Gabaix *et al.*, 2003)
  - **Sizes** of **largest** market participants (**mutual funds**): $\zeta \approx 1$
  - **Trading volume** V : $\zeta \approx 1.5$
  - **Price impact** of trade: $r = \Delta p \approx CV^{1/2}$
- $\zeta \approx 1$ (Zipf's law): **firm sizes** & **city sizes** (Gabaix, 1999, Axtell, 2001)
- **Income**: $1.5 < \zeta < 3$ **Wealth**: $\zeta \approx 1.5$

Heavy tails in exchange rates: Preliminary results

Currency	N	$\hat{\zeta}$	S.e.$=\frac{2\hat{\zeta}}{\sqrt{n}}$	95% CI
Euro	2820	4.64	0.39	(3.87, 5.40)
Swiss Frank	2829	4.60	0.39	(3.84, 5.36)
GBP	2826	4.47	0.38	(3.73, 5.20)
Norwegian Kroner	2833	3.53	0.30	(2.95, 4.12)
Singapore Dollar	1824	3.37	0.35	(2.68, 4.06)
Japanese Yen	2828	3.23	0.27	(2.69, 3.76)
Canadian Dollar	1945	2.96	0.30	(2.37, 3.55)
Australian Dollar	1946	2.61	0.26	(2.09, 3.13)
Indian Rupee	2644	2.57	0.22	(2.13, 3.01)
Hong Kong Dollar	1425	2.56	0.30	(1.97, 3.16)
Malaysian Ringgit	1479	2.44	0.28	(1.89, 3.00)
Taiwan Dollar	1497	2.34	0.27	(1.81, 2.86)
Thai Baht	1644	2.14	0.24	(1.68, 2.60)
Russian Ruble	2263	1.90	0.18	(1.55, 2.25)
Philippine Peso	1466	1.62	0.12	(1.38, 1.86)
Korean Won	1544	1.63	0.19	(1.27, 2.00)
Uzbek Soum: Official	696	1.20	0.20	(0.80, 1.60)
Uzbek Soum: Unofficial $(n = 0.2N)$	109	1.93	0.58	(0.79, 3.07)

Inference on tail index ζ

- **Log-log rank-size** regression
  - $\log(t - 1/2) = \hat{a} - \hat{\zeta} r_t$ (Gabaix & Ibragimov, 2007)

    $r_1 \geq r_2 \geq ... \geq r_n$, $n = 0.2N$
  - **Shift** $Rank - 1/2$: optimal
  - Correct 95% **CI's**: $\zeta \in (\hat{\zeta} - 1.96 \times \sqrt{\frac{\mathbf{2}}{n}}\hat{\zeta}, \hat{\zeta} + 1.96 \times \sqrt{\frac{\mathbf{2}}{n}}\hat{\zeta})$

Heavy tails & (non-)robustness

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

- **Income, wealth, firm sizes**: **heavy-tailed** (Singh-Maddala, Pareto, log-normal) $\Rightarrow$ **poor performance**
  - **Asymptotic**, standard & non-standard (moon) **bootstrap**
  - **Semiparametric** bootstrap & asymptotic methods $\Rightarrow$ Improvement
  - **Computationally expensive** (bootstrap), problematic **s.e.'s** (asymptotic)

Heavy tails & (non-)robustness

- HHI & $\zeta < 2$: **Inconsistent**, **random** limits (Mandelbrot)

- **Similar** conclusions: Sharpe ratio, coefficient of variation & self-normalized sums

- SR & CV: **problematic** estimation for $2 < \zeta < 4$

  Infinite fourth moments $\Rightarrow$ **S.e.** of **sample variance**=?

Sharpe ratio & self-normalized measures

- $r_j,\ j = 1, \cdots, q$: **excess returns**
  - **Heterogeneous** $r_j \sim \mathcal{N}(\mu, \sigma_j^2)$
  - **Heavy-tailed** mixtures (Student-t : $\zeta = 4$; Cauchy, $\zeta = 1$)
- **Sharpe ratio** $SR = \frac{\overline{r}}{s_r}$

  $P(SR > z) \leq P\Big(SR > z | \sigma_1^2 = ... = \sigma_q^2\Big)$, $z \geq cv(0.05) \times \sqrt{q}$

  $P(SR > z | X_j \sim$ **Cauchy**$) \leq P\Big(SR > z | X_j \sim$ **N(0, 1)**$\Big)$
- **Less volatility** $\Rightarrow$ **Higher SR** far out in **tails = large** z (Hedge funds?)
- **Not so in general** (smaller z : $z > cv(0.1) \times \sqrt{q}$)

Coefficient of variation

- **Heterogeneous** $X_j \sim \mathcal{N}(\mu, \sigma_j^2)$; **Heavy-tailed** mixtures (Student-t : $\zeta = 4$; Cauchy, $\zeta = 1$)

- **Coefficient of variation** $CV = \frac{s_X}{\overline{X}}$

  $$P(0 < CV < z) \leq P\Big(0 < CV < z | \sigma_1^2 = ... = \sigma_q^2\Big)$$

  $$z \leq 1/(cv(0.05) \times \sqrt{q})$$

  $$P(0 < CV < z | X_j \sim \textbf{Cauchy}) \leq P\Big(0 < CV < z | X_j \sim \ \textbf{N(0, 1)}\Big)$$

- **Not so in general:** $z > 1/(cv(0.1) \times \sqrt{q})$

- **Homogeneity & light tails** $\Rightarrow$ **reduction** in **inequality & disparity only for small** values, **not in general**

Herfindahl-Hirschman index of concentration

- **HHI** $HHI = \sum_{j=1}^{q} X_j^2 / \left(\sum_{j=1}^{q} X_j\right)^2$

  $P(0 < HHI < z) \leq P\left(0 < HHI < z | \sigma_1^2 = ... = \sigma_q^2\right),$

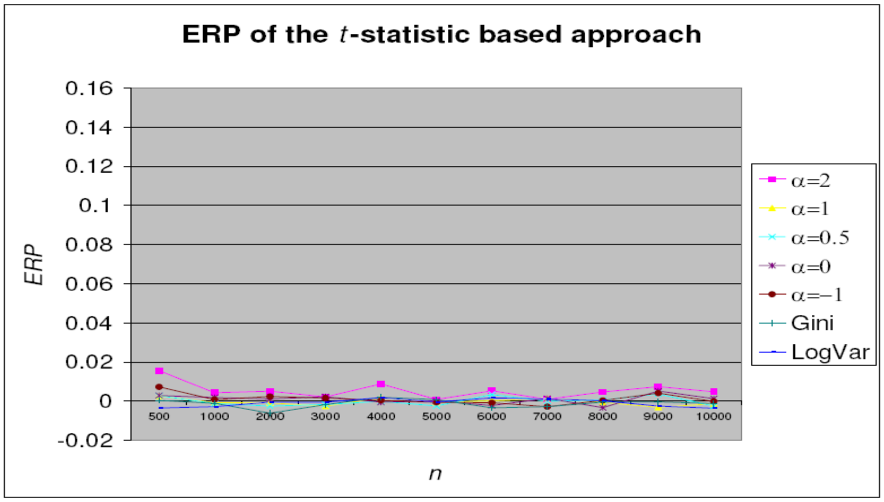
  $z \leq \sqrt{\frac{qcv^2(0.05)}{cv^2(0.05)+q-1}}$

  $P(0 < HHI < z | X_j \sim \textbf{Cauchy}) \leq P\left(0 < HHI < z | X_j \sim \textbf{N(0, 1)}\right)$

- **Not so in general:** $z > \sqrt{\frac{qcv^2(0.1)}{cv^2(0.1)+q-1}}$

- **Homogeneity & light tails** $\Rightarrow$ **reduction** in **concentration only for small** values, **not in general**

Robust inference for risk, inequality, poverty & concentration: *t*-statistic based approach vs. alternative procedures



F.A. Cowell, E. Flachaire / Journal of Econometrics 141 (2007) 1044–1072

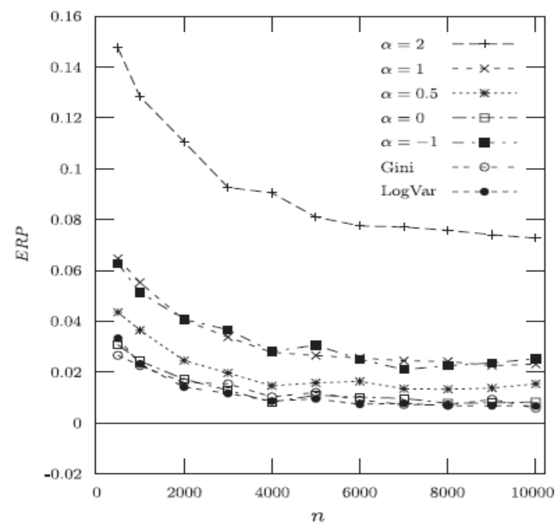


Fig. 7. ERP of asymptotic tests.

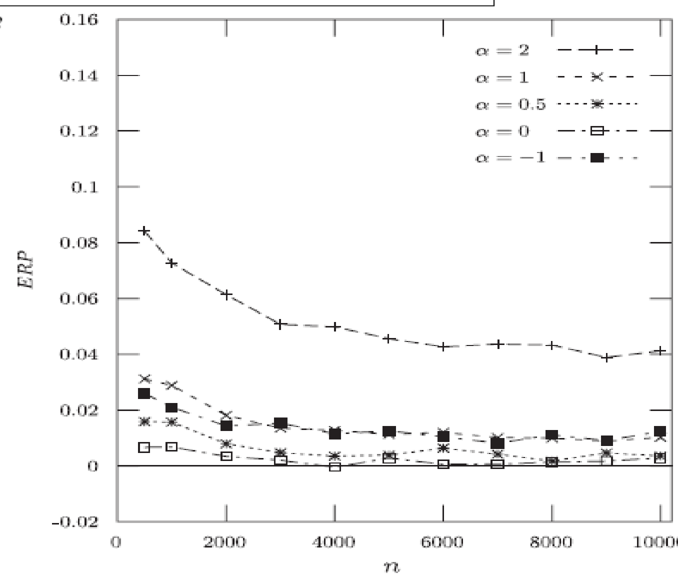


Fig. 8. ERP of bootstrap tests.

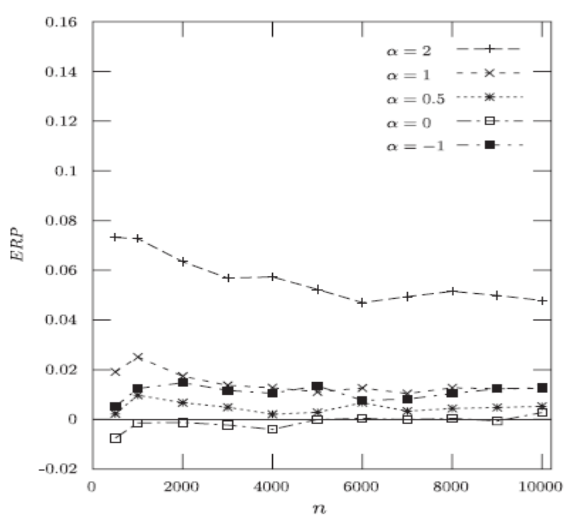


Fig. 9. *Moon* bootstrap.

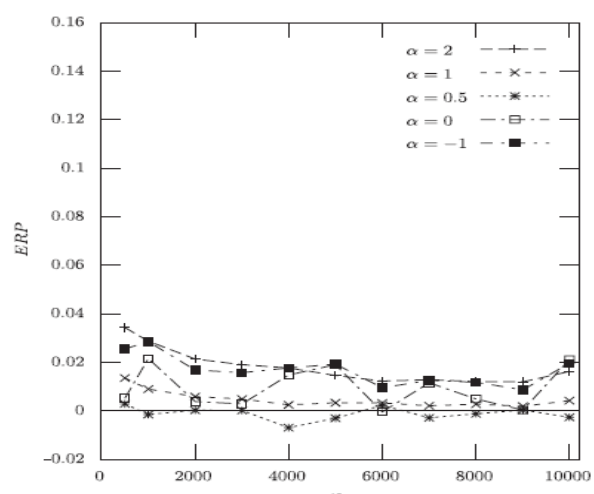


Fig. 10. *Semiparametric* bootstrap.

Singh-Maddala: $P(X.>y)=(1+ay^b)^{-c}$, a=100, b=2.8, c=1.7 $\Rightarrow$ $\xi = bc = 4.8$

Risk, Inequality & Concentration: MC Results

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

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

	Singh-Maddala		Pareto		Lognormal	
N	500	1000	500	1000	500	1000
			Size			
t-statistic $q = 2$	5.4	4.8	7.2	6.9	5.3	5.2
t-statistic $q = 4$	6.9	6.4	16.6	14.5	9.3	7.3
Asymptotic	11.5	10.5	29.4	25.9	16.1	14.1
Standard bootstrap	8.1	7.9	17.8	16.3	10.4	9.6
Moon bootstrap	6.9	7.5	6.9	10.0	8.2	8.1
Semiparametric bootstrap	6.4	5.9	10.6	9.9	4.8	4.8
Asymptotic test with semiparametric measure	6.6	6.3	11.6	10.4	2.9	2.8
% of cases semipar. measure cannot be computed	0	0	8.9%	6.3%	9.0%	3.3%

Risk, Inequality & Concentration: MC Results

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

	Singh-Maddala		Pareto		Lognormal	
N	500	1000	500	1000	500	1000
			Size			
t-statistic $q = 2$	5.3	5.3	6.2	5.9	5.4	5.0
t-statistic $q = 4$	6.3	5.5	10.2	8.0	6.1	5.5
Asymptotic	8.1	7.4	19.3	16.9	9.2	8.9
Standard bootstrap	5.7	5.7	11.7	10.5	6.3	6.6
Moon bootstrap	4.3	4.9	5.5	7.5	4.8	5.7
Semiparametric bootstrap	5.5	5.2	8.7	9.4	4.3	5.3
Asymptotic test with semiparametric measure	5.1	5.2	12.4	11.9	3.7	4.3
% of cases semipar. measure cannot be computed	0	0	8.9%	6.3%	9%	3.3%

Table 1: ERP of the $t-$statistic based tests on inequality measures

N	$\alpha=-1$	$\alpha=0$	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	Gini	LogVar
				$q=2$			
500	0.0073	0.0029	0.0028	0.0026	0.0156	0.0004	−0.0035
1000	0.0008	0.0019	−0.0012	−0.0008	0.0043	−0.0011	−0.0028
2000	0.0024	0.0011	−0.0025	−0.0009	0.0051	−0.0061	−0.0005
3000	0.0016	0.0023	−0.0017	−0.0025	0.0022	−0.0018	−0.0005
4000	0.0004	−0.0005	−0.0001	0.0017	0.0088	0.0020	0.0018
5000	−0.0006	0.0001	−0.0024	−0.0006	0.0009	0.0008	−0.0004
6000	−0.0008	−0.0022	0.0034	0.0002	0.0053	−0.0034	0.0017
7000	−0.0028	0.0014	0.000	0.0014	0.0010	−0.0028	0.0011
8000	0.0004	−0.0033	0.0006	−0.0006	0.0047	−0.0005	0.0002
9000	0.0041	0.0052	0.0037	−0.0031	0.0074	−0.0004	−0.0025
10000	−0.0001	0.0013	−0.0025	−0.0013	0.0048	−0.0012	−0.0036
500	0.0154	0.0093	0.0076	0.0182	0.0580	0.0080	0.0080
1000	0.0098	0.0050	0.0076	0.0048	0.0416	0.0032	0.0018
2000	0.0045	0.0024	0.0037	0.0093	0.0274	0.0033	0.0004
3000	0.0035	0.0017	0.0032	0.0060	0.0213	0.0060	0.0014
4000	0.0096	−0.0041	0.0043	−0.0009	0.0258	0.0003	0.0000
5000	0.0015	0.0016	0.0040	0.0011	0.0210	0.0010	0.0040
6000	−0.0010	0.0060	−0.0008	0.0046	0.0178	0.0011	0.0015
7000	0.0076	−0.0002	0.0036	0.0008	0.0187	0.0015	−0.0033
8000	0.0012	0.0005	0.0002	0.0024	0.0154	0.0016	0.0030
9000	0.0050	−0.0011	0.0001	0.0048	0.0191	0.0021	0.0023
10000	−0.0007	0.0003	-0.0038	0.0036	0.0094	0.0018	−0.0021

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.