Size-Dependent Probability Bounds for t-Tests

Alessandro Palandri*

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
G. Parenti, Università degli Studi di Firenze

DCU Business School, Dublin City University

This Version: February 29, 2020

Abstract

The paper extends Chebyshev's inequality to incorporate moments' convergence in t-tests of model parameters. Size-dependent probability bounds are derived from one conditional higher-order moment of the distribution of the test statistic. Monte Carlo simulations attest that, in the cases of heteroskedastic and autocorrelated observations, the proposed bounds over-reject less than the asymptotic approximation and bootstrap methods. Therefore, when asymptotic critical values are suspected to lead to the over-rejection of the null hypothesis, the proposed inequalities may be used in conjunction to bootstrap methods to reduce the number of instances in which multiple re-samplings and associated estimations have to be performed.

Keywords: Chebyshev's Inequality, Asymptotic Approximation, Over Rejection, Bootstrap, Wild Bootstrap.

JEL classification: C01, C12, C15.

---

*Address correspondence to: Alessandro Palandri, DCU Business School, Dublin City University, Dublin 9, Ireland, Email: `alessandro.palandri@gmail.com`

1 Introduction

When the number of observations is not large enough (relative to the data characteristics) for the asymptotic distribution of a test statistic to be a good approximation to its unknown sample distribution, asymptotic inference will lead to size distortions. In this situation, bootstrap methods present a valid alternative: artificial datasets, generated by resampling the data, are aggregated into statistics which may be used to construct confidence intervals for the statistic calculated on real data. When applied to approximate the distribution of model parameters, bootstrap methods require parameter estimates for each of the artificial datasets. This otherwise time consuming aspect, depending on the bootstrap resampling scheme, becomes computationally demanding for parameter estimates that are not available in closed form as each bootstrap replication will require the numerical optimization of the estimator's objective function.

In principle, another approach to testing could rely on the inequalities of Chebyshev and Cantelli which provide bounding probabilities of c^{-2} and $(1+c^2)^{-1}$ to outcomes that are c standard deviations *below and above* and *below or above* the mean, respectively. Numerous extensions and refinements of these inequalities exist which focus on bounding the probabilities of the $\{z_i\}_{i=1}^N$ draws of random variables: $\mathbb{P}\left(\bigcap_{i=1}^2 |z_i| < c\right)$ due to Berge (1938), $\mathbb{P}\left(\bigcap_{i=1}^I |z_i| < c_i\right)$ due to Olkin and Pratt (1958) and $\mathbb{P}\left(\sum_{i=1}^I z_i^2 \geq c_i^2\right)$ due to Birnbaum *et al.* (1947), just to name a few. Isii (1959) gives bounds for $\mathbb{P}\left(\bigcap_{i=1}^2 c_1 < x_i < c_2\right)$ for *non*-standardized random variables while Kotz *et al.* (2000) refine the bounds of Olkin and Pratt (1958) for independent random variables. Further refinements exist for random variables that are non-negative, bounded, have known distributional characteristics and higher-order moments, as in Zelen (1954) and Bhattacharyya (1987). However, when applied to the testing of model parameters, none of these inequalities takes into account the sample-size effects, induced by the Central Limit Theorem (CLT) and moments' convergence, on the shape of the distribution. As a result, the ensuing loose[1] probability bounds are hardly ever used in hypotheses testing.

This paper extends Chebyshev's inequality to incorporate sample-size effects into the probability bounds. In particular, since the general concern behind size distor-

[1]Chebyshev's (Cantelli's) are sharp only for the worst case distributions of Section 3.1. For example, if the true distribution is Gaussian, the probabilities of observing values that are more than 2, 3, and 4 standard deviations from the mean are 4.55%, 0.27% and 0.006%, respectively. Chebyshev's bounds, on the other hand, are 25%, 11% and 6.25%, respectively.

tions is the over-rejection of null hypotheses, the proposed probability bounds are explicitly derived for t-tests with sample distributions that are leptokurtic, relative to their asymptotic limit. Size-dependence is attained by considering one higher-order moment of the estimator's distribution and the rate at which it converges to its asymptotic value. Furthermore, distinctively from other Chebyshev's inequalities, those put forth in this paper are explicitly aimed at estimates of models' parameters and associated t-tests. Monte Carlo Simulations show that the proposed bounds may be used effectively in hypotheses testing and that they are a robust and fast alternative to bootstrap methods. In particular, whenever the proposed bounds reject the null hypothesis, simulation studies confirm that it is safe to forgo lengthier bootstrap procedures on the grounds that they will also reject. Besides robustness, the main advantage of the proposed approach to testing is simplicity: its implementation only requires the calculation of an arithmetic average from the building blocks of the test statistic.

The paper is organized as follows. Section 2 introduces size-dependence from the convergence of higher-order moments. Properties of the components of the proposed inequalities are studied in Section 3. Implementation of the size-dependent bounds is discussed in Section 4 followed by the simulations' setup and results of Section 5. Section 6 concludes.

2 Higher-Order Moments Bounds

Let $\{z_i\}_{i=1}^N$ be a sample of independent and identically distributed standardized random variables and $\widetilde{z}_N = N^{-1/2}\sum_{i=1}^N z_i$ be the corresponding standardized mean estimator. Chebyshev's inequality is derived from the variance of $\widetilde{z}_N$ as follows:

$$1 = \mathbb{E}\left(\widetilde{z}_N^2\right) \geq \mathbb{E}\left(\widetilde{z}_N^2 \cdot \mathbb{1}[|\widetilde{z}_N| > c\,]\right) \geq c^2 \cdot \mathbb{P}\left(|\widetilde{z}_N| > c\right) \tag{1}$$

where $\mathbb{1}$ is the indicator function. Although the bounds could be tightened by eliminating the first inequality in (1):

$$\mathbb{E}\left(\widetilde{z}_N^2 \cdot \mathbb{1}[|\widetilde{z}_N| > c\,]\right) \geq c^2 \cdot \mathbb{P}\left(|\widetilde{z}_N| > c\right) \tag{2}$$

it is not obvious how to estimate the tail variance of $\widetilde{z}_N$. Furthermore, the inequality in (2) doesn't account for the asymptotic effects of CLT and moments' convergence.

2.1 Convergence and Size-Dependence

Assume $\widetilde{z}_N$ is leptokurtic and that although CLT applies it does not yield good finite sample approximations. Furthermore, assume the m-th moment of $\widetilde{z}_N$ exists and converges asymptotically to that of a Gaussian random variable. Let $M_L(N, m, c)$ be the absolute m-th moment of $\widetilde{z}_N$ conditional on $\widetilde{z}_N < -c$ for some tail value $c > 0$:

$$M_L(N, m, c) \equiv \mathbb{E}\left(|\widetilde{z}_N|^m \cdot \mathbb{1}[\widetilde{z}_N < -c\,]\right) \geq c^m \cdot \mathbb{P}(\widetilde{z}_N < -c) \tag{3}$$

As N diverges, the m-th moment converges to that of a Gaussian random variable at the rate $\pi_{N,m}$. Since $M_L(N, m, c)$ is a component of the m-th moment, its rate of convergence is at least $\pi_{N,m}$:

$$M_L(N, m, c) = (1 - \pi_{N,m})\, M_L(\infty, m, c) + \pi_{N,m} M_L(1, m, c) \tag{4}$$

Substituting equation (4) in (3) yields:

$$\mathbb{P}_m(\widetilde{z}_N < -c) \leq c^{-m} \left\{(1 - \pi_{N,m})\, M_L(\infty, m, c) + \pi_{N,m} M_L(1, m, c)\right\}$$

and *mutatis mutandis*:

$$\mathbb{P}_m(\widetilde{z}_N > \quad c) \leq c^{-m} \left\{(1 - \pi_{N,m})\, M_L(\infty, m, c) + \pi_{N,m} M_H(1, m, c)\right\}$$

where $M_H(N, m, c)$ is the absolute m-th moment of $\widetilde{z}_N$ conditional on $\widetilde{z}_N > c$ and $M_H(\infty, m, c) = M_L(\infty, m, c)$, from asymptotic symmetry. The resulting two-sided probability is:

$$\mathbb{P}_m(|\widetilde{z}_N| > c) \leq c^{-m} \left\{2\,(1 - \pi_{N,m})\, M_L(\infty, m, c) + \pi_{N,m} M_{L+H}(1, m, c)\right\}$$

where $M_{L+H}(1, m, c) = M_L(1, m, c) + M_H(1, m, c)$. For third moment bounds (Skew) $\pi_{N,3} = N^{-1/2}$ and $M_L(\infty, 3, c) = (c^2+2)\phi(c)$ while for fourth moment bounds (Kurt) $\pi_{N,4} = N^{-1}$ and $M_L(\infty, 4, c) = (c^3 + 2c)\phi(c) + 3[1 - \Phi(c)]$.

3 Properties

The probability bounds of Section 2 are convex combinations of the $N = 1$ components $M_L(1, m, c)$, $M_H(1, m, c)$ and the $N = \infty$ component $M_L(\infty, m, c)$. What follows outlines their respective properties to shed light on the behavior of Skew and Kurt bounds both in small samples and asymptotically.

3.1 $N = 1$ Components

Since $M_L(1, m, c)$ and $M_H(1, m, c)$ may be consistently estimated from the data, in this section they will be treated as known. Let the standardized random variable z_i be distributed so that Chebyshev is sharp:

$$z_i = \begin{cases} -(2p)^{-1/2} & \text{w.p.} \quad p \\ 0 & \text{w.p.} \quad 1-2p \\ (2p)^{-1/2} & \text{w.p.} \quad p \end{cases} \quad \Rightarrow \quad \mathbb{P}_{\textsc{True}}(|z_i| \geq c) = \begin{cases} 2p & \text{for} \quad c \leq (2p)^{-1/2} \\ 0 & \text{for} \quad c > (2p)^{-1/2} \end{cases}$$

with $p < 0.5$. The probability bound from Chebyshev's inequality is $\mathbb{P}_{\textsc{Cheby}}(|z_i| \geq c) \leq c^{-2}$ which is sharp for $c = (2p)^{-1/2}$. Skewness and kurtosis bounds are:

$$\mathbb{P}_{\textsc{Skew,Kurt}}(|z_i| \geq c) \leq \begin{cases} (2p)^{1-m/2}c^{-3} & \text{for} \quad c \leq (2p)^{-1/2} \\ 0 & \text{for} \quad c > (2p)^{-1/2} \end{cases}$$

with $m = 3$ for Skew and $m = 4$ for Kurt. Hence, for $c < (2p)^{-1/2}$ Cheby gives the sharpest bounds, for $c = (2p)^{-1/2}$ all approaches provide sharp bounds of $2p$, while for $c > (2p)^{-1/2}$ Skew and Kurt are sharper than Chebyshev[2]. Since statistical tests are generally conducted for threshold values of two standard deviations or more, compared to Chebyshev and Cantelli, the use of higher-order moments doesn't necessarily result in losses of sharpness.

3.2 $N = \infty$ Components

The limiting bounding probabilities of the proposed approaches are:

$$\lim_{N\to\infty} \mathbb{P}_{\textsc{Skew}}(|\widetilde{z}_N| > c) \leq 2M_L(\infty, 3, c) = 2(c^{-1} + 2c^{-3})\phi(c)$$
$$\lim_{N\to\infty} \mathbb{P}_{\textsc{Kurt}}(|\widetilde{z}_N| > c) \leq 2M_L(\infty, 4, c) = 2(c^{-1} + 3c^{-3})\phi(c) + 6c^{-4}[1 - \Phi(c)]$$

Since they do not converge to Gaussian probabilities, it is worth investigating their limits. To begin, there exist c^* such that for every $c > c^*$ Skew and Kurt are sharper than Cantelli (Cant) and Chebyshev (Cheby):

$$\begin{array}{lcll} \lim_{N\to\infty} \mathbb{P}_{\textsc{Skew}}(\ \widetilde{z}_N \ > c) & < & (1 + c^2)^{-1} & ; \quad \forall c > 1.171274 \\ \lim_{N\to\infty} \mathbb{P}_{\textsc{Skew}}(|\widetilde{z}_N| > c) & < & c^{-2} & ; \quad \forall c > 1.279774 \\ \lim_{N\to\infty} \mathbb{P}_{\textsc{Kurt}}(\ \widetilde{z}_N \ > c) & < & (1 + c^2)^{-1} & ; \quad \forall c > 1.394795 \\ \lim_{N\to\infty} \mathbb{P}_{\textsc{Kurt}}(|\widetilde{z}_N| > c) & < & c^{-2} & ; \quad \forall c > 1.543475 \end{array}$$

[2] Analogous results may be derived for Cantelli (Cant) by setting $p = 0.5$. For $c < 1$ Cant gives the sharpest bounds, for $c = 1$ all approaches provide sharp bounds of 0.5, while for $c > 1$ Skew and Kurt are sharper than Cantelli.

For $N \to \infty$ and $c \to \infty$ the ratios of the Gaussian probability to Cant, Cheby, Skew and Kurt are:

$$\lim_{c\to\infty}\left(\mathbb{P}_{\text{Gauss}}/\mathbb{P}_{\text{Cant}}\right) = 0 \quad ; \quad \lim_{c\to\infty}\left(\mathbb{P}_{\text{Gauss}}/\mathbb{P}_{\text{Cheby}}\right) = 0$$
$$\lim_{c\to\infty}\left(\mathbb{P}_{\text{Gauss}}/\mathbb{P}_{\text{Skew}}\right) = 1 \quad ; \quad \lim_{c\to\infty}\left(\mathbb{P}_{\text{Gauss}}/\mathbb{P}_{\text{Kurt}}\right) = 1$$

Since, for diverging threshold values, Skew and Kurt probabilities converge to Gauss, their asymptotic tail-behavior is superior to both Cant and Cheby. For nominal sizes of (0.01, 0.05, 0.10), asymptotic significance levels[3] are: (0.0000, 0.0000, 0.0016) for Cheby, (0.0070, 0.0301, 0.0542) for Skew and (0.0062, 0.0254, 0.0445) for Kurt.

4 Extension to Model Parameters

Implementation of the proposed approaches to the class of M-estimators is straightforward. To begin, recall that, given the objective function ψ, standard asymptotic confidence intervals for the estimates $\widehat{\beta}$ are obtained starting from the mean-value-theorem expansion around the vector of true parameter values β_0:

$$\widehat{\beta} = \beta_0 - \left[\frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial^2\psi_i}{\partial\beta\partial\beta'}\right|_{\beta=\beta^*}\right]^{-1} \cdot \left[\frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial\psi_i}{\partial\beta}\right|_{\beta=\beta_0}\right]$$

where ψ_i is the objective function evaluated at the i-th observation and β^* is a convex combination of β_0 and $\widehat{\beta}$. Now consider the same mean-value-theorem expansion but around the vector of parameter estimates $\widehat{\beta}$:

$$\widehat{\beta} = \widehat{\beta} - \left[\frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial^2\psi_i}{\partial\beta\partial\beta'}\right|_{\beta=\widehat{\beta}}\right]^{-1} \cdot \left[\frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial\psi_i}{\partial\beta}\right|_{\beta=\widehat{\beta}}\right]$$

From the above *tautology* it is possible to define the pseudo-observations $\{\widehat{\beta}_i\}_{i=1}^N$ with average $\widehat{\beta}$:

$$\widehat{\beta}_i \equiv \widehat{\beta} - \left[\frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial^2\psi_i}{\partial\beta\partial\beta'}\right|_{\beta=\widehat{\beta}}\right]^{-1} \cdot \left[\left.\frac{\partial\psi_i}{\partial\beta}\right|_{\beta=\widehat{\beta}}\right]$$

The $\{\widehat{\beta}_i\}_{i=1}^N$, which are readily available as a by-product of the calculation of standard t-tests, may be used to calculate p-values according to the proposed bounds. It is

[3] Let $g_i(c) \equiv \mathbb{P}_i\left(|\widetilde{z}_N| > c\right)$ be the monotonically non-decreasing p-value function. The actual significance level is $q_i \equiv \mathbb{P}\left(g_i(c) \leq p\right) = \mathbb{P}\left(c \geq g_i^{-1}(p)\right)$, where p is the nominal level. Since under the null, c converges to a standardized Gaussian random variable, it follows that $q_i = 2\left[1 - \Phi\left(g_i^{-1}(p)\right)\right]$.

important to emphasize that, since estimating the variance of the k-th parameter $\widehat{\beta}_k$ from the pseudo-observations $\{\widehat{\beta}_{k,i}\}_{i=1}^N$ produces White's (1980) Heteroskedasticity-Consistent (HC) variance estimates, the following approach to the calculation of probability bounds is by construction robust to heteroskedasticity.

Estimation of the $N = 1$ components, needed for Skew and Kurt bounds, is particularly undemanding. From the pseudo-observations of the k-th parameter $\{\widehat{\beta}_{k,i}\}_{i=1}^N$:

$$\widetilde{\beta}_{k,i} = \frac{\widehat{\beta}_{k,i} - \beta_{0,k}}{\widehat{\sigma}_k} \quad ; \quad \widehat{\sigma}_k^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\beta}_{k,i} - \widehat{\beta}_k\right)^2 \quad ; \quad c_k = \frac{\sqrt{N}\left|\widehat{\beta}_k - \beta_{0,k}\right|}{\widehat{\sigma}_k} \tag{5}$$

where $\beta_{0,k}$ is the parameter value under the null, $\widehat{\beta}_k$ is the parameter estimate and c_k is the realization of the standard t-test statistic. Then, the $N = 1$ components may be estimated consistently from:

$$\widehat{M}_L(1, m, c) = \frac{1}{N}\sum_{\widetilde{\beta}_{k,i} < -c_k} |\widetilde{\beta}_{k,i}|^m \quad \text{and} \quad \widehat{M}_H(1, m, c) = \frac{1}{N}\sum_{\widetilde{\beta}_{k,i} > c_k} |\widetilde{\beta}_{k,i}|^m$$

Since in finite samples there is a non-zero probability that $-c < \widehat{z}_{\min}$ and $\widehat{z}_{\max} < c$, for which $\widehat{M}_L(1, m, c) = 0$ and $\widehat{M}_H(1, m, c) = 0$, the resulting p-values would be downward biased and lead to over-rejections. A simple solution to this problem is to use Cantelli/Chebyshev to replace the $N = 1$ components whenever the conditioning sets are empty[4]. Specifically, in the case of third moment bounds for the t-test statistic t_k associated to the k-th model parameter, if $\sum_{|\widetilde{\beta}_{k,i}| > c_k} 1 > 0$:

$$\mathbb{P}_{\text{Skew}}(|t_k| > c_k) \leq 2(1 - N^{-1/2})(c_k^{-1} + 2c_k^{-3})\phi(c_k) + c_k^{-3}N^{-3/2}\sum_{|\widetilde{\beta}_{k,i}| > c_k} |\widetilde{\beta}_{k,i}|^3$$

while in the case of an empty conditioning set:

$$\mathbb{P}_{\text{Skew}}(|t_k| > c_k) \leq 2(1 - N^{-1/2})(c_k^{-1} + 2c_k^{-3})\phi(c_k) + c_k^{-2}N^{-1/2}$$

With respect to Cantelli and Chebyshev, which require estimation of the parameters' standard errors $\widehat{\sigma}_k$, the proposed bounds require the estimation of one additional higher-order moment. Using consistent moment estimates produces the same probability bounds as if the higher-order moment was known, asymptotically. In finite samples, however, moment estimates, which contain estimation error, are likely to have a non-negligible impact on the resulting probability bounds. The simulation studies allow to evaluate, among others, the effects that estimation errors in moments have on the overall goodness of the bounds.

[4] Using the bound of equation (3), $\mathbb{P}(z_i < -c) = c^{-m}M_L(1, m, c)$. When the moment's conditioning set is empty, it may substituted with $\mathbb{P}(z_i < -c) \leq (1+c^2)^{-1}$. Similarly for two-sided probability bounds where $c^{-m}\left[M_L(1, m, c) + M_H(1, m, c)\right]$ is substituted with $\mathbb{P}(|z_i| > c) \leq c^{-2}$.

4.1 Extension to Autocorrelated Observations

Since time series observations may not be assumed to be serially uncorrelated, the proposed probability bounds are not immediately applicable. The standard approach is to define the number of lags L after which the correlations either vanish or become negligible. In Newey and West (1987) L corresponds to the truncation lag while in block-bootstrap[5] it corresponds to the block-size. Following the same principle, let $\{\widehat{\beta}_{k,\tau}\}_{\tau=1}^{T/\text{L}}$ be the L-aggregated pseudo-observations. Then, the setup of equation (5) becomes:

$$\widetilde{\beta}_{k,\tau} = \frac{\widehat{\beta}_{k,\tau} - \text{L} \cdot \beta_{0,k}}{\sqrt{\text{L}} \cdot \widetilde{\sigma}_k} \quad \text{with} \quad \widetilde{\sigma}_k^2 = T \cdot \mathbb{VAR}_{HAC}\left[\widehat{\beta}_k\right] \quad \text{and} \quad c_k = \frac{\sqrt{T}\left|\widehat{\beta}_k - \beta_{0,k}\right|}{\widetilde{\sigma}_k}$$

where $\mathbb{VAR}_{HAC}\left[\widehat{\beta}_k\right]$ is a Heteroskedasticity-Autocorrelation-Consistent (HAC) estimator of the variance of $\widehat{\beta}_k$.

5 Simulations Set Up and Results

In the following experiments Monte Carlo simulations are set at $S = 10000$ and bootstrap replications at $B = 2000$. Both *pairs*-bootstrap P-Boot and *residual*-bootstrap R-Boot are calculated and reported for comparisons. The *pairs*-bootstrap only assumes existence of the joint empirical distribution function, from which it samples. Hence, since it makes no underlying homoskedasticity assumptions, P-Boot is robust to heteroskedasticity by construction. On the other hand, *residual*-bootstrap requires that the parametric model correctly describes the conditional expectation of the dependent variable. *Residual*-bootstrap that accounts for heteroskedasticity is the *wild*-bootstrap developed in Liu (1988). Specific details of the *pairs*- and *residual*-bootstrap implementation are discussed in Sections 5.1 and 5.2.

In addition, the performance of the proposed Skew and Kurt bounds is benchmarked against that of Chebyshev's and Berry-Esseen's inequalities. Berry-Esseen's inequality (B&E), bounds the distance between the true distribution of the test statistic t_k and its Gaussian limit. In the context of this paper, it reads:

$$\sup_{c \in \mathbb{R}} \left|\mathbb{P}(t_k < c) - \Phi(c)\right| \leq \kappa N^{-1/2} \mathbb{E}\left(|\widetilde{\beta}_{k,i}^3|\right)$$

[5] See Künsch (1989), Liu and Singh (1992), Politis and Romano (1992,1994), Carlstein *et al.* (1998) and Paroditis and Politis (2001), among others, for some block-bootstrap methods and resampling procedures.

which leads to the following probability bounds:

$$\begin{aligned}
\mathbb{P}_{\text{B\&E}}(t_k < -c) &\leq 1 - \Phi(c) + \kappa \cdot N^{-1/2}\mathbb{E}\left(|\widetilde{\beta}^3_{k,i}|\right) \\
\mathbb{P}_{\text{B\&E}}(t_k > \quad c) &\leq 1 - \Phi(c) + \kappa \cdot N^{-1/2}\mathbb{E}\left(|\widetilde{\beta}^3_{k,i}|\right) \\
\mathbb{P}_{\text{B\&E}}(|t_k| > c) &\leq 2[1 - \Phi(c)] + 2\kappa \cdot N^{-1/2}\mathbb{E}\left(|\widetilde{\beta}^3_{k,i}|\right)
\end{aligned}$$

Currently, the best estimate of the constant is $\kappa = 0.4748$, due to Shevtsova (2011). Notice that, contrary to SKEW and KURT, B&E bounds converge to Gaussian probabilities. Due to its undisputed theoretical properties[6], B&E delivers *state of the art* probability bounds incorporating asymptotic convergence to Gaussian probabilities.

5.1 Heteroskedastic OLS

Simulated data is generated from the linear regression model:

$$y_i = 0 + 1 \cdot x_{1,i} + 2 \cdot x_{2,i} + |x_{1,i} \cdot x_{2,i}|^d \cdot \epsilon_i$$

where $x_{1,i}, x_{2,i}$ and ϵ_i are uncorrelated zero mean random variables with distributions $N(0,1)$, $\chi^2_{(1)} - 1$, $F_{(1,10)} - 5/4$ and $F_{(1,5)} - 5/3$. $d = 1$ except for the Gaussian control case with $d = 0$. The sample sizes considered are $N = \{100, 200, 1000\}$.

R-BOOT is the *wild*-bootstrap of Davidson and Flachaire (2008) which, as shown by their simulation studies, provides the most accurate rejection rates in the heteroskedastic case. Specifically, regression residuals $\widehat{\epsilon}_i$ are rescaled by $(1 - h_i)$, with $h_i = x_i'(X'X)^{-1}x_i$, and their sign changed[7] with probability 1/2.

Simulation results of Table 1 show that in the homoskedastic Gaussian case the sizes of GAUSS, P-BOOT and the *wild*-bootstrap R-BOOT are very accurate. B&E bounds perform the worst with no observed rejections of the null hypothesis. CHEBY does slightly better but still under-rejects severely. Also the proposed bounds SKEW and KURT exhibit substantial under-rejections: due to the lack of tail observations, in most Monte Carlo simulations, infeasible estimates of the $N = 1$ components are replaced with Chebyshev's bounds, hence the similar, although less severe, under-rejections.

Results for the heteroskedastic case with Chi-square innovations are reported in the second panel of Table 1. Despite the use of White's (1980) HC variance-covariance

[6] See van der Vaart and Wellner (1996).

[7] Davidson and Flachaire (2008) show that residuals' sign changes following the Radamacher distribution give better results than changes following the distribution of Mammen (1993).

estimates, GAUSS over-rejects severely. Over-rejection rates similar to those of the asymptotic test are reported for P-BOOT. On the other hand, *wild*-bootstrap R-BOOT displays substantially lower size distortions in small samples. In fact, for sample sizes of $N = 1000$, the best performing *wild bootstrap* R-BOOT exhibits rejection rates similar to those of the asymptotic GAUSS. At the other hand of the spectrum: CHEBY and B&E under-reject, with the latter displaying rejection rates that are *one-tenth* less than nominal. SKEW and KURT have rejection rates that are closest to nominal for all sample sizes considered with the former slightly more conservative than the latter.

In Table 2, with F-distributed innovations, GAUSS exhibits the worst over-rejections. P-BOOT does substantially better than HC asymptotic tests but still over-rejects. Interestingly, with F-distributed shocks, R-BOOT does only marginally better than P-BOOT. The proposed bounds display better rejection rates with SKEW and KURT performing always and nearly always better than R-BOOT, respectively. In this setting, the generally severe under-rejections of CHEBY come in handy to deliver rejection rates that are coincidentally closest to nominal. B&E still under-rejects severely: *one-fiftieth* less than nominal.

5.2 Heteroskedastic and Autocorrelated OLS

The T dependent variables are generated from:

$$y_t = 0 + 1 \cdot x_{1,t} + 2 \cdot x_{2,t} + u_t \qquad \text{with} \qquad u_t = 0.7 \cdot u_{t-1} + |x_{1,t} \cdot x_{2,t}| \cdot \epsilon_t$$

where $x_{1,t}, x_{2,t}$ and ϵ_t are zero mean uncorrelated random variables with distributions $N(0,1)$, $\chi^2_{(1)} - 1$, $F_{(1,10)} - 5/4$ and $F_{(1,5)} - 5/3$. Following Andrews (1991), the number of lags L in the HAC estimator (GAUSS and CHEBY), is automatically selected by a quadratic spectral kernel with $\alpha(2) = 4\rho^2(1-\rho^2)^{-4}$. For comparison purposes, the same L is used as block-size (P-BOOT and R-BOOT) and to aggregate the pseudo-parameters' estimates[8] (SKEW, KURT and B&E). In particular, while P-BOOT is the standard *pairs-block*-bootstrap, R-BOOT is the *wild*-bootstrap with block-sampling of size L. The latter is presented purely for completeness since the sign changes of the scaled resampled residuals are expected to disrupt the autocorrelation structure despite block-resampling.

[8] The rates of convergence of the moments in equation (4) and that of the Berry-Esseen inequality are derived under the assumption of no serial correlation.

In the Gaussian case of Table 3 Gauss over-rejects in small samples. P-Boot and R-Boot do better and worse than asymptotic test, respectively. Cheby severely under-rejects while B&E never rejects. Severe under-rejections are observed also for Skew and Kurt: as for the homoskedastic Gaussian case of Table 1, the lack of tail observations causes infeasible estimates of the $N = 1$ components to be replaced by Chebyshev's bounds, causing under-rejections.

Chi-square innovations, in the second panel of Table 3 produce substantial over-rejections in Gauss. P-Boot does better than asymptotic tests in small samples while it is indistinguishable for large N. Despite block-sampling, R-Boot does worse than Gauss due to the sign changes of the *wild*-bootstrap that destroy persistence. Skew and Kurt display rejection rates that are closest to nominal, with the former slightly more conservative than the latter. With the exception of very small samples, Cheby severely under-rejects while B&E barely ever rejects.

Table 4 reports simulation results in the case of F-distributed shocks. Although Gauss over-rejects severely, R-Boot does worse. P-Boot does substantially better than the asymptotic test although the over-rejection is still severe. Like in the Chi-square case, both Skew and Kurt exhibit better size properties than bootstrap with the former, again, mildly more conservative than the latter. With the exception of large samples $N = 1000$, Cheby under-rejects with $F(1, 10)$ shocks but over-rejects with $F(1, 5)$ shocks. One more time, B&E barely ever rejects.

6 Conclusions

In finite samples, higher-order moments bounds have better size properties than the Berry-Esseen's inequality which exhibits under-rejection rates[9] that are more severe than Chebyshev's. Most interestingly, the rejection rates of the proposed bounds are found to be lower than those of bootstrap, regardless of sample size and degree of leptokurtosis of the test statistic. This finding suggests that the proposed bounds may be used as a robust and fast alternative to bootstrap in suspected cases of over-rejection of the null hypothesis. In fact, should either Skew or Kurt reject the null, performing bootstrap would not produce a different outcome. The advantage of calculating the proposed bounds first is that, in case of rejection of the null, the more

[9]The actual significance level of $B\&E$ is $q = p - 2\kappa N^{-1/2}\mathbb{E}(|z_i|^3)$. For Gaussian random variables, a no more than 10% discrepancy between actual and nominal size requires $N = \{23104, 92416, 2310400\}$ observations for $p = \{0.1, 0.05, 0.01\}$, respectively. For leptokurtic random variables, the same level of discrepancy is attained by significantly larger sample sizes.

computationally demanding bootstrap may be avoided altogether. On the other hand, should SKEW or KURT fail to reject, it is always possible to implement bootstrap methods, should they be deemed more accurate in the given setting. Refinements of the proposed bounds in the case of empty conditioning sets and extensions to unit-root testing are left as areas for future work.

References

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, **59**(3) 817-858.

Berge, P. O. (1938), "A note on a form of Tchebysheff's theorem for two variables", *Biometrika*, **29** 405-406.

Bhattacharyya, B. B. (1987), "One-sided chebyshev inequality when the first four moments are known", *Communications in Statistics - Theory and Methods*, **16**(9) 2789-2791.

Birnbaum, Z. W., J. Raymond and H. S. Zuckerman (1947), "A Generalization of Tshebyshev's Inequality to Two Dimensions", *The Annals of Mathematical Statistics*, **18**(1) 70-79.

Carlstein, E., K. A. Do, P. Hall, T. Hesterberg and H. R. Künsch (1998), "Matched-block bootstrap for dependent data", *Bernoulli*, **4** 305-328.

Davidson, R. and E. Flachaire (2008), "The wild bootstrap, tamed at last", *Journal of Econometrics*, **146**, 162-169.

Isii, K. (1959), "On a method for generalizations of Tchebycheff's inequality", *Annals of the Institute of Statistical Mathematics*, **10**(2) 65-88.

Kotz, S., N. Balakrishan and N. L. Johnson (2000), Continuous Multivariate Distributions, Volume 1, Models and Applications, (2nd ed.), *Wiley Series in Probability and Statistics.*

Künsch, H. R. (1989), "The jacknife and the bootstrap for general stationary observations", *The Annals of Statistics*, **17**(3) 1217-1241.

Liu, R. Y. (1988), "Bootstrap procedure under some non-i.i.d. models", *Annals of Statistics*, **16** 1696-1708.

Liu, R. Y. and K. Singh (1992), "Moving blocks jacknife and bootsratp capture weak dependence", In: LePage, R. and L. Billard (Eds.), Exploring the Limits of Bootstrap, *Wiley Series in Probability and Statistics.*

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models", *Annals of Statistics*, **21**, 255-285.

Newey, W. K. and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, **55**(3) 703-708.

Olkin, I. and J. W. Pratt (1958), "A Multivariate Tchebycheff Inequality", *The Annuals of Mathematical Statistics*, **29**(1) 226-234.

Paroditis, E. and D. N. Politis (2001), "The tapered block bootstrap", *Biometrika*, **88** 1105-1119.

Politis, D. N. and J. P. Romano (1992), "A circular block resampling procedure for stationary data", In: LePage, R. and L. Billard (Eds.), Exploring the Limits of Bootstrap, *Wiley Series in Probability and Statistics.*

Politis, D. N. and J. P. Romano (1994), "The stationary bootstrap", *Journal of the American Statistical Association*, **89** 1303-1313.

Shevtsova, I. (2011), "On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands", *Working Paper.*

van der Vaart, A. and J. A. Wellner (1996), "Weak Convergence and Empirical Processes (With Applications to Statistics", Springer, New York.

White, H. (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, **48** 817-838.

Zelen M. (1954), "Bounds on a distribution function that are functions of moments to order four", *Journal of Research of the National Bureau of Standards*, **53**(6) 377-381.

Table 1: Tests of OLS Paramter with Gaussian Homoskedastic and Centered $\chi^2_{(1)}$ Heteroskedastic Variables

The table reports probabilities of a Type I error. In the Nom column are reported the nominal values of the test. Empirical rejection rates are reported for the bounds based on Gaussian (Gauss), Pairs-Bootstrap (P-Boot), Residual-Bootstrap (R-Boot), Chebyshev (Cheby), Berry-Esseen (B&E), Skewness (Skew) and Kurtosis (Kurt). N indicates the sample size of the tests. The number of Monte Carlo simulations is 10000 and the number of bootstrap replications 2000.

Nom	Gauss	P-Boot	R-Boot	Cheby	B&E	Skew	Kurt
Gaussian Homoskedastic							
	N= 100						
0.0010	0.0021	0.0012	0.0010	0.0000	0.0000	0.0000	0.0000
0.0100	0.0163	0.0128	0.0097	0.0000	0.0000	0.0000	0.0004
0.0500	0.0634	0.0580	0.0470	0.0000	0.0000	0.0005	0.0059
0.1000	0.1175	0.1080	0.0953	0.0030	0.0000	0.0023	0.0133
	N= 200						
0.0010	0.0017	0.0015	0.0010	0.0000	0.0000	0.0000	0.0000
0.0100	0.0114	0.0105	0.0079	0.0000	0.0000	0.0000	0.0006
0.0500	0.0532	0.0532	0.0482	0.0000	0.0000	0.0004	0.0076
0.1000	0.1052	0.1050	0.0969	0.0029	0.0000	0.0030	0.0186
	N=1000						
0.0010	0.0001	0.0008	0.0016	0.0000	0.0000	0.0000	0.0000
0.0100	0.0091	0.0104	0.0090	0.0000	0.0000	0.0000	0.0013
0.0500	0.0479	0.0524	0.0467	0.0000	0.0000	0.0008	0.0165
0.1000	0.0947	0.1014	0.0976	0.0006	0.0000	0.0080	0.0321
Centered $\chi^2_{(1)}$ Heteroskedastic							
	N= 100						
0.0010	0.0471	0.0167	0.0160	0.0000	0.0000	0.0000	0.0092
0.0100	0.1068	0.0654	0.0524	0.0000	0.0000	0.0198	0.0431
0.0500	0.2046	0.1514	0.1233	0.0137	0.0003	0.0667	0.0932
0.1000	0.2840	0.2203	0.1835	0.0537	0.0029	0.1068	0.1261
	N= 200						
0.0010	0.0286	0.0184	0.0151	0.0000	0.0000	0.0008	0.0069
0.0100	0.0742	0.0606	0.0503	0.0000	0.0000	0.0141	0.0338
0.0500	0.1578	0.1356	0.1182	0.0062	0.0004	0.0597	0.0810
0.1000	0.2263	0.2044	0.1790	0.0339	0.0035	0.0944	0.1121
	N=1000						
0.0010	0.0127	0.0140	0.0084	0.0000	0.0000	0.0004	0.0047
0.0100	0.0416	0.0461	0.0365	0.0000	0.0000	0.0102	0.0251
0.0500	0.1019	0.1114	0.0958	0.0016	0.0000	0.0482	0.0630
0.1000	0.1591	0.1739	0.1557	0.0152	0.0012	0.0764	0.0864

Table 2: Tests of OLS Paramter with Centered $F_{(1,10)}$ and $F_{(1,5)}$ Heteroskedastic Variables

The table reports probabilities of a Type I error. In the Nom column are reported the nominal values of the test. Empirical rejection rates are reported for the bounds based on Gaussian (Gauss), Pairs-Bootstrap (P-Boot), Residual-Bootstrap (R-Boot), Chebyshev (Cheby), Berry-Esseen (B&E), Skewness (Skew) and Kurtosis (Kurt). N indicates the sample size of the tests. The number of Monte Carlo simulations is 10000 and the number of bootstrap replications 2000.

Nom	Gauss	P-Boot	R-Boot	Cheby	B&E	Skew	Kurt
Centered $F_{(1,10)}$ Heteroskedastic							
	N= 100						
0.0010	0.1053	0.0247	0.0267	0.0003	0.0000	0.0010	0.0280
0.0100	0.1768	0.0812	0.0775	0.0049	0.0000	0.0471	0.0965
0.0500	0.2874	0.1745	0.1587	0.0490	0.0001	0.1384	0.1671
0.1000	0.3649	0.2497	0.2188	0.1157	0.0020	0.1863	0.2048
	N= 200						
0.0010	0.0656	0.0261	0.0280	0.0000	0.0000	0.0029	0.0218
0.0100	0.1304	0.0844	0.0763	0.0010	0.0000	0.0397	0.0767
0.0500	0.2354	0.1738	0.1556	0.0211	0.0004	0.1188	0.1416
0.1000	0.3099	0.2465	0.2202	0.0738	0.0022	0.1641	0.1790
	N=1000						
0.0010	0.0291	0.0234	0.0200	0.0000	0.0000	0.0011	0.0155
0.0100	0.0707	0.0666	0.0582	0.0002	0.0000	0.0285	0.0468
0.0500	0.1502	0.1448	0.1286	0.0061	0.0003	0.0818	0.0964
0.1000	0.2120	0.2081	0.1942	0.0343	0.0017	0.1162	0.1253
Centered $F_{(1,5)}$ Heteroskedastic							
	N= 100						
0.0010	0.2014	0.0325	0.0476	0.0072	0.0000	0.0020	0.0183
0.0100	0.2836	0.1067	0.1152	0.0347	0.0000	0.0276	0.0890
0.0500	0.3958	0.2195	0.2007	0.1247	0.0000	0.1322	0.2041
0.1000	0.4728	0.3016	0.2656	0.2112	0.0003	0.2123	0.2659
	N= 200						
0.0010	0.1663	0.0468	0.0498	0.0034	0.0000	0.0016	0.0188
0.0100	0.2508	0.1270	0.1142	0.0187	0.0000	0.0307	0.0999
0.0500	0.3566	0.2344	0.2043	0.0901	0.0000	0.1439	0.1988
0.1000	0.4319	0.3067	0.2697	0.1794	0.0005	0.2119	0.2554
	N=1000						
0.0010	0.1066	0.0445	0.0451	0.0012	0.0000	0.0015	0.0218
0.0100	0.1737	0.1158	0.1023	0.0089	0.0000	0.0375	0.0898
0.0500	0.2706	0.2151	0.1940	0.0501	0.0001	0.1284	0.1670
0.1000	0.3475	0.2827	0.2574	0.1165	0.0002	0.1839	0.2104

Table 3: Tests of OLS Paramter with Gaussian and Centered $\chi^2_{(1)}$ Autocorrelated and Heteroskedastic Variables

The table reports probabilities of a Type I error. In the Nom column are reported the nominal values of the test. Empirical rejection rates are reported for the bounds based on Gaussian (Gauss), Pairs-Bootstrap (P-Boot), Residual-Bootstrap (R-Boot), Chebyshev (Cheby), Berry-Esseen (B&E), Skewness (Skew) and Kurtosis (Kurt). N indicates the sample size of the tests. The number of Monte Carlo simulations is 10000 and the number of bootstrap replications 2000.

Nom	Gauss	P-Boot	R-Boot	Cheby	B&E	Skew	Kurt
Gaussian Autocorrelated Heteroskedastic							
	N= 100, l= 6						
0.0010	0.0148	0.0019	0.0147	0.0001	0.0000	0.0000	0.0000
0.0100	0.0425	0.0188	0.0416	0.0010	0.0000	0.0002	0.0015
0.0500	0.1112	0.0732	0.1090	0.0044	0.0000	0.0053	0.0126
0.1000	0.1790	0.1297	0.1731	0.0166	0.0000	0.0132	0.0267
	N= 200, l= 7						
0.0010	0.0058	0.0017	0.0027	0.0000	0.0000	0.0000	0.0000
0.0100	0.0239	0.0172	0.0259	0.0000	0.0000	0.0001	0.0017
0.0500	0.0832	0.0701	0.1100	0.0009	0.0000	0.0057	0.0161
0.1000	0.1434	0.1253	0.1957	0.0072	0.0000	0.0155	0.0324
	N=1000, l=10						
0.0010	0.0011	0.0015	0.0070	0.0000	0.0000	0.0000	0.0001
0.0100	0.0119	0.0137	0.0488	0.0000	0.0000	0.0004	0.0040
0.0500	0.0546	0.0596	0.1417	0.0000	0.0000	0.0100	0.0202
0.1000	0.1054	0.1127	0.2199	0.0014	0.0000	0.0243	0.0393
Centered $\chi^2_{(1)}$ Autocorrelated Heteroskedastic							
	N= 100, l= 6						
0.0010	0.0633	0.0090	0.0083	0.0002	0.0000	0.0002	0.0005
0.0100	0.1259	0.0442	0.0559	0.0011	0.0000	0.0083	0.0407
0.0500	0.2258	0.1236	0.1959	0.0234	0.0000	0.0709	0.0994
0.1000	0.3046	0.1898	0.3207	0.0713	0.0002	0.1119	0.1329
	N= 200, l= 7						
0.0010	0.0349	0.0118	0.0169	0.0000	0.0000	0.0000	0.0005
0.0100	0.0865	0.0504	0.0853	0.0000	0.0000	0.0089	0.0375
0.0500	0.1690	0.1233	0.2414	0.0084	0.0000	0.0702	0.0899
0.1000	0.2424	0.1864	0.3748	0.0404	0.0004	0.1056	0.1194
	N=1000, l=10						
0.0010	0.0133	0.0113	0.0430	0.0000	0.0000	0.0000	0.0065
0.0100	0.0399	0.0404	0.1397	0.0000	0.0000	0.0156	0.0242
0.0500	0.1046	0.1053	0.3088	0.0009	0.0001	0.0514	0.0609
0.1000	0.1613	0.1645	0.4442	0.0163	0.0008	0.0790	0.0863

Table 4: Tests of OLS Paramter with Centered $F_{(1,10)}$ and $F_{(1,5)}$ Autocorrelated and Heteroskedastic Variables

The table reports probabilities of a Type I error. In the Nom column are reported the nominal values of the test. Empirical rejection rates are reported for the bounds based on Gaussian (Gauss), Pairs-Bootstrap (P-Boot), Residual-Bootstrap (R-Boot), Chebyshev (Cheby), Berry-Esseen (B&E), Skewness (Skew) and Kurtosis (Kurt). N indicates the sample size of the tests. The number of Monte Carlo simulations is 10000 and the number of bootstrap replications 2000.

Nom	Gauss	P-Boot	R-Boot	Cheby	B&E	Skew	Kurt
Centered $F_{(1,10)}$ Autocorrelated Heteroskedastic							
	N= 100, l= 6						
0.0010	0.1141	0.0099	0.0096	0.0005	0.0000	0.0010	0.0054
0.0100	0.1849	0.0522	0.0669	0.0063	0.0001	0.0204	0.0733
0.0500	0.2911	0.1420	0.2192	0.0576	0.0009	0.1156	0.1536
0.1000	0.3742	0.2126	0.3543	0.1234	0.0023	0.1710	0.1921
	N= 200, l= 7						
0.0010	0.0787	0.0180	0.0249	0.0000	0.0000	0.0001	0.0049
0.0100	0.1395	0.0599	0.1104	0.0020	0.0000	0.0231	0.0754
0.0500	0.2481	0.1524	0.2891	0.0284	0.0004	0.1158	0.1410
0.1000	0.3215	0.2273	0.4294	0.0871	0.0015	0.1657	0.1799
	N=1000, l=10						
0.0010	0.0297	0.0227	0.0795	0.0000	0.0000	0.0000	0.0153
0.0100	0.0766	0.0663	0.2019	0.0001	0.0000	0.0311	0.0520
0.0500	0.1559	0.1454	0.4087	0.0070	0.0001	0.0896	0.1019
0.1000	0.2240	0.2080	0.5519	0.0359	0.0013	0.1266	0.1339
Centered $F_{(1,5)}$ Autocorrelated Heteroskedastic							
	N= 100, l= 6						
0.0010	0.2021	0.0137	0.0101	0.0052	0.0001	0.0021	0.0068
0.0100	0.2804	0.0644	0.0709	0.0355	0.0004	0.0167	0.0572
0.0500	0.3875	0.1683	0.2374	0.1300	0.0023	0.0921	0.1587
0.1000	0.4577	0.2457	0.3844	0.2138	0.0041	0.1652	0.2206
	N= 200, l= 7						
0.0010	0.1710	0.0252	0.0348	0.0045	0.0000	0.0018	0.0086
0.0100	0.2481	0.0879	0.1358	0.0237	0.0001	0.0246	0.0737
0.0500	0.3556	0.1961	0.3309	0.1008	0.0011	0.1156	0.1738
0.1000	0.4292	0.2729	0.4767	0.1821	0.0021	0.1860	0.2293
	N=1000, l=10						
0.0010	0.1110	0.0444	0.1166	0.0023	0.0000	0.0003	0.0183
0.0100	0.1815	0.1116	0.2651	0.0078	0.0000	0.0343	0.0883
0.0500	0.2792	0.2107	0.4881	0.0494	0.0000	0.1310	0.1686
0.1000	0.3529	0.2827	0.6218	0.1205	0.0001	0.1845	0.2135