Journal of Econometrics 172 (2013) 255–274

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Moment condition tests for heavy tailed time series

Jonathan B. Hill[a,*], Mike Aguilar[b]

[a] Department of Economics, CB 3305, University of North Carolina, Chapel Hill, NC 27599, United States
[b] Department of Economics, University of North Carolina, Chapel Hill, United States

ARTICLE INFO

Article history:
Available online 23 August 2012

JEL classification:
C13
C20
C22

Keywords:
Moment condition test
Heavy tails
Tail trimming
Robust inference

ABSTRACT

We develop an asymptotically chi-squared statistic for testing moment conditions $E[m_t(\theta^0)] = 0$, where $m_t(\theta^0)$ may be weakly dependent, scalar components of $m_t(\theta^0)$ may have an infinite variance, and $E[m_t(\theta)]$ need not exist for any θ under the alternative. Score tests are a natural application, and in general a variety of tests can be heavy-tail robustified by our method, including white noise, GARCH affects, omitted variables, distribution, functional form, causation, volatility spillover and over-identification. The test statistic is derived from a tail-trimmed sample version of the moments evaluated at a consistent plug-in $\hat{\theta}_T$ for θ^0. Depending on the test in question and heaviness of tails, $\hat{\theta}_T$ may be *any* consistent estimator including sub-$T^{1/2}$-convergent and/or asymptotically non-Gaussian ones, since $\hat{\theta}_T$ can be assured not to affect the test statistic asymptotically. We adapt bootstrap, *p*-value occupation time, and covariance determinant methods for selecting the trimming fractile in any sample, and apply our statistic to tests of white noise, omitted variables and volatility spillover. We find it obtains sharp empirical size and strong power, while conventional tests exhibit size distortions.

1. Introduction

We propose an asymptotically chi-squared statistic for testing moment conditions in the presence of heavy tails.

Casting inference within a moment condition framework covers a broad array of tests. Moment condition [MC] tests, as well as those with MC interpretations (like score tests), include tests of omitted or instrumental variables, functional form, distribution, conditional heteroscedasticity, over-identification, causation, volatility spillover, structural change, order selection and encompassing tests. Notable theory contributions for iid data include Hausman (1978), White (1981), Hansen (1982), Newey (1985), Bierens (1990), Wooldridge (1990), Imbens et al. (1998) and Kitamura et al. (2004) to name very few. Moment equality and inequality tests for time series are developed in de Jong (1996), Florens et al. (1998), Bai (2003), Ghysels and Guay (2003), Hill (2008) and Bontemps and Meddahi (in press), amongst others.

Applying moment based inference to heavy tailed data is particularly relevant since evidence for heavy tails across disciplines is substantial, ranging from financial, macroeconomic, auction, actuarial, and meteorological to network telecommunication data. The literature is vast, but notable surveys include Embrechts et al. (1997) and Finkenstädt and Rootzén (2003). See also Hill and Shneyerov (2010). In general, by their sample moment form, MC tests require the existence of higher moments to ensure standard asymptotics. As a consequence, this forces possibly very severe moment restrictions on an underlying process that may fail for even mildly volatile data. One example is estimating equations for GARCH models as the basis of a test of over-identifying restrictions or model mis-specification. A finite variance may require the GARCH process to have a finite eighth moment, or an error to have a finite fourth moment, depending on the equation form (Hill and Renault, 2010). The problem of requiring "more than just the hypotheses of interest" was noted in Wooldridge (1990, p. 18) concerning a variety of contexts, although not heavy tails.

Let $m_t : \Theta \rightarrow \mathbb{R}^q$ be parametric test equations where Θ is a compact subset of $\mathbb{R}^r$, and $q, r \geq 1$, and for simplicity of exposition assume $m_t(\theta)$ is continuous and differentiable. The null hypothesis is

$$H_0 : E\left[m_t(\theta^0)\right] = 0 \quad \text{for unique } \theta^0 \in \Theta \tag{1}$$

with a general alternative

$$H_1 : \text{the null is false.} \tag{2}$$

We allow for heavy tails such that $E[m_{i,t}^2(\theta^0)] = \infty$ and do not require the moment $E[m_t(\theta)]$ to exist under H_1 for any θ. If the test equation $m_t(\theta)$ is integrable uniformly on Θ, then the alternative becomes $H_1 : E[m_t(\theta^0)] \neq 0$. In general θ may represent a subset of parameters, such as when testing for the autoregression order in

* Corresponding author. Tel.: +1 919 966 5343; fax: +1 919 966 4986.
E-mail addresses: jbhill@email.unc.edu (J.B. Hill), maguilar@email.unc.edu (M. Aguilar).
URL: http://www.unc.edu/~jbhill (J.B. Hill).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.08.013

an AR-GARCH. Alternatively, the equations may be parameter free $m_t(\theta) = m_t$, such as in a test of white noise on an observable time series. See Section 3 for examples.

In order to conquer the challenge of heavy tails, and arrive at a test statistic that is easy to compute and interpret due to a standard limit distribution, we trim a negligible sample fraction of those $m_{i,t}(\theta)$ that may be heavy tailed. Let $\{k_{1,i,T}, k_{2,i,T}\}$ be integer *fractile* sequences representing the number of trimmed left-tailed and right-tailed observations from each sample $\{m_{i,t}(\theta)\}_{t=1}^{T}$ with sample size T. We enforce negligible trimming by assuming $\{k_{1,i,T}, k_{2,i,T}\}$ are intermediate order sequences: $k_{j,i,T} \to \infty$ and $k_{j,i,T}/T \to 0$ (Leadbetter et al., 1983; Hahn et al., 1991). Define tail specific observations of $m_{i,t}(\theta)$ and their sample order statistics:

$$m_{i,t}^{(-)}(\theta) := m_{i,t}(\theta) \times I\left(m_{i,t}(\theta) < 0\right) \quad \text{and}$$
$$m_{i,(1)}^{(-)}(\theta) \le \cdots \le m_{i,(T)}^{(-)}(\theta) \le 0$$
$$m_{i,t}^{(+)}(\theta) := m_{i,t}(\theta) \times I\left(m_{i,t}(\theta) > 0\right) \quad \text{and}$$
$$m_{i,(1)}^{(+)}(\theta) \ge \cdots \ge m_{i,(T)}^{(+)}(\theta) \ge 0.$$

If an equation $m_{i,t}(\theta^0)$ has an infinite variance, or its higher moments are unknown, we trim $m_{i,t}(\theta)$ such that it lies between its lower $k_{1,i,T}/T$th and upper $k_{2,i,T}/T$th sample quantiles:

$$\begin{aligned}\hat{m}_{T,i,t}^{*}(\theta) &= m_{i,t}(\theta) \\ &\quad \times I\left(m_{i,(k_{1,i,T})}^{(-)}(\theta) \le m_{i,t}(\theta) \le m_{i,(k_{2,i,T})}^{(+)}(\theta)\right) \\ &= m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta) \end{aligned} \qquad (3)$$
$$\hat{m}_{T,t}^{*}(\theta) = \left[m_{i,t}(\theta) \times \hat{I}_{i,T,t}(\theta)\right]_{i=1}^{q}$$

where $\hat{I}_{i,T,t}(\theta) = 1$ if equation i is not trimmed.

Note $I(A) = 1$ is A is true, and 0 otherwise. If $m_{i,t}(\theta^0)$ is symmetric then use[1] $I(|m_{i,t}(\theta)| \le m_{i,(k_{i,T})}^{(a)}(\theta))$ where $m_{i,t}^{(a)}(\theta) := |m_{i,t}(\theta)|$, $k_{i,T} \to \infty$ and $k_{i,T}/T \to 0$.

Now let $\hat{\theta}_T$ be any consistent estimator of θ^0. The proposed Tail-Trimmed Moment Condition [TTMC] test statistic has a quadratic form

$$\hat{W}_T = \left(\sum_{t=1}^{T} \hat{m}_{T,t}^{*}(\hat{\theta}_T)\right)' \hat{S}_T^{-1}(\hat{\theta}_T) \left(\sum_{t=1}^{T} \hat{m}_{T,t}^{*}(\hat{\theta}_T)\right) \qquad (4)$$

where $\hat{S}_T(\theta)$ is a kernel covariance estimator

$$\hat{S}_T(\theta) := \sum_{s,t=1}^{T} k\left((s-t)/\gamma_T\right) \times \left\{\hat{m}_{T,s}^{*}(\theta) - \hat{m}_{T}^{*}(\theta)\right\}\left\{\hat{m}_{T,t}^{*}(\theta) - \hat{m}_{T}^{*}(\theta)\right\}',$$

and $\hat{m}_T^{*}(\theta) := 1/T\sum_{t=1}^{T}\hat{m}_{T,t}^{*}(\theta)$, $k(\cdot)$ is a kernel function and $\gamma_T \to \infty$ is a bandwidth parameter. Trimming introduces spurious dependence unless $m_{i,t}(\theta^0)$ is iid, so an HAC is in general preferred. See Section 2.3.

Our framework is built on the principles of Generalized of Method of Tail-Trimmed Moments by Hill and Renault (2010) and denoted HR (2010). Indeed, $\sum_{t=1}^{T}\hat{m}_{T,t}^{*}(\theta)' \times \hat{S}_T^{-1}(\hat{\theta}_T) \times \sum_{t=1}^{T}\hat{m}_{T,t}^{*}(\theta)$ is simply the efficiently weighted GMTTM criterion with consistent plug-in $\hat{\theta}_T$, provided there are at least as many equations as parameters $q \ge r$, and (1) holds. The primary contribution of this paper is to provide an accompanying theory of heavy tail robust inference for GMM, with an arbitrary plug-in $\hat{\theta}_T$ that may not have a Gaussian limit nor be $T^{1/2}$-convergent.

We prove $\hat{W}_T$ is asymptotically chi-squared under the null (1) and suitable regularity conditions. Further, under these conditions $\hat{W}_T$ has non-negligible power against a sequence of local alternatives, which implies $\hat{W}_T \to \infty$ under (2) with probability one. In both cases $\hat{W}_T$ has the same limiting properties whether tails are heavy or not due to self-standardization and trimming negligibility. This is a major advantage over other methods (see below) since other than $E[m_t(\theta^0)] = 0$ under H_0 we never need to know if $m_t(\theta^0)$ lacks higher moments. Small sample power, however, can in principle be affected by trimming when tails are thin. See Theorem 2.2 for details, and see Section 5 for simulation evidence that trimming need not reduce power.

In the presence of heavy tails $\sum_{t=1}^{T}\hat{m}_{T,t}^{*}(\theta^0)$ and $\hat{\theta}_T$ typically have different rates of convergence. In many cases the test component $\sum_{t=1}^{T}\hat{m}_{T,t}^{*}(\theta^0)$ can be slowed simply by trimming more (i.e. faster $k_{j,i,T} \to \infty$), implying many plug-ins $\hat{\theta}_T$ will not affect $\hat{W}_T$, hence $\hat{\theta}_T$ need not be $T^{1/2}$-convergent *nor* possess a Gaussian limit. This is evidently the first study to allow sub-$T^{1/2}$ *and* super-$T^{1/2}$-convergent estimation, although Antoine and Renault (2012) tackle heterogeneous rates at or below $T^{1/2}$ for GMM. We further develop these ideas by example in Section 3 since deep results require a specification for $m_t(\theta)$. Depending on the form of $m_{i,t}(\theta^0)$ and tails, valid plug-ins include at least OLS, LAD, QML and GMM, and information theoretic estimators like Empirical Likelihood, each with non-Gaussian limits when tails are heavy. Similarly, we may use estimators robust to data contamination that are *not* robust to heavy tails, like Least Trimmed Squares and Quasi-Maximum Trimmed Likelihood (see Čižek (2008) and his references). Finally, heavy tailed robust estimators with Gaussian limits are valid, like Peng and Yao's (2003) Log-LAD, Ling's (2005, 2007), Least Weighted Absolute Deviations [LWAD] and Quasi-Maximum Weighted Likelihood [QMWL], and HR's (2010) Generalized Method of Tail-Trimmed Moments [GMTTM].

Further transformations of the equations may lead to robustness of $\hat{W}_T$ to *any* $\hat{\theta}_T$ that converges no slower than GMTTM, including orthogonal projections (Wooldridge, 1990; Bai, 2003; Bontemps and Meddahi, in press). In the literature robustness is evidently only ensured for $T^{1/2}$-convergent plug-ins.[2] We only briefly discuss orthogonal transformations in Section 2 due to space constraints.

In Section 4 we discuss data-driven methods for choosing the number of trimmed $m_t(\theta)'s$, including p-value occupation time, wild bootstrap, and a covariance determinant technique. We then investigate tests of white noise and omitted variables in a simulation study in Section 5, and we study tests of volatility spillover in Hill and Aguilar (2011), the technical appendix to this paper. Our simulations serve two purposes. First, they demonstrate heavy tails may distort empirical size of non-robust tests, adding to existing evidence (e.g. de Lima (1997) and Runde (1997)). Second, trimming remarkably few large $m_t(\theta)$ leads to sharp empirical size while still retaining power in most cases, and substantial power

[1] If $m_{i,t}(\theta^0)$ is symmetric, in *theory* it suffices to use $I(m_{i,(k_{1,i,T})}^{(-)}(\theta) \le m_{i,t}(\theta) \le m_{i,(k_{2,i,T})}^{(+)}(\theta))$ with $k_{1,i,T} = k_{2,i,T}\forall T$. In *practice*, however, $m_{i,(k_{1,i,T})}^{(-)}(\theta)$ and $m_{i,(k_{2,i,T})}^{(+)}(\theta)$ may be quite different, even for large T, and our simulation experiments show $I(|m_{i,t}(\theta)| \le m_{i,(k_{i,T})}^{(a)}(\theta))$ works exceptionally well because it forces the same left- and right-tail threshold.

[2] GMTTM for heavy tailed GARCH is $o_p(T^{1/2})$ and is guaranteed *not* to affect $\hat{W}_T$ for a test of volatility spillover couched in a GARCH framework (see HR (2010) and Hill and Aguilar (2011)). Sub-$T^{1/2}$-convergence also arises due to kernel smoothing in estimating equations, in-fill asymptotics, and nearly-weak GMM. See Antoine and Renault (2012) for examples and references.