Robust Generalized Empirical Likelihood for Heavy Tailed Autoregressions with Conditionally Heteroscedastic Errors

Jonathan B. Hill*
University of North Carolina – Chapel Hill

December 8, 2014

Abstract

We present a robust Generalized Empirical Likelihood estimator and confidence region for the parameters of an autoregression that may have a heavy tailed heteroscedastic error. The estimator exploits two transformations for heavy tail robustness: a redescending transformation of the error that robustifies against innovation outliers, and weighted least squares instruments that ensure robustness against heavy tailed regressors. Our estimator is consistent for the true parameter and asymptotically normally distributed irrespective of heavy tails.

1 Introduction

We present a robust Generalized Empirical Likelihood estimator and confidence region for an autoregression that may have a heavy tailed error. The setting is motivated first by the substantial evidence for heavy tails in macroeconomics, finance and insurance markets (e.g. Embrechts et al., 1997). Second, extreme values are frequently interpreted as outliers or one-off events (see, e.g., Jureckova and Sen, 1996). Third, the recent intense interest in information theoretic methods (Imbens, 1997; Kitamura, 1997; Kitamura and Stutzer, 1997; Smith, 1997; Antoine et al., 2007), in particular the recognition of higher order improvements afforded by GEL estimators (Newey and Smith, 2004), and use of the empirical likelihood method to sidestep complicated or case-dependent covariance matrix estimation (Owen, 1990).

The time series of interest is a stationary, ergodic autoregression $y_t$, defined by:

$$y_t = \xi + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t(\theta) = \theta' x_t + \epsilon_t(\theta) \text{ where } x_t = [1, y_{t-1}, ..., y_{t-p}]' \text{ and } \theta = [\xi, \phi']' \in \mathbb{R}^{p+1}. \quad (1)$$

*Dept. of Economics, University of North Carolina, Chapel Hill; http://www.unc.edu/~jbhill; jbhill@email.unc.edu.

Key words and phrases: Empirical Likelihood, heavy tails, autoregression, redescending transformation, tail trimming, robust estimation.

AMS subject classifications : 62F10, 62F35.
We assume there exists a unique point $\theta_0 = [\xi_0, \phi_0']'$ in a compact subset $\Theta \subset \mathbb{R}^{p+1}$ such that the roots of $1 - \sum_{i=1}^{p} \phi_{0,i} z^i$ lie outside the unit circle, and $\epsilon_t \equiv \epsilon_t(\theta_0)$ is a stationary martingale difference with respect to increasing $\sigma$-fields $\mathcal{F}_t \equiv \sigma(y_{\tau} : \tau \leq t)$.

We also assume $\epsilon_t$ is heteroscedastic with the form

$$\epsilon_t = \sigma_t u_t \quad \text{where } u_t \text{ is i.i.d.}$$

(2)

In this paper $\sigma_t$ does not need to be known, provided it satisfies a mixing condition. The time series $y_t$ may therefore be heavy tailed $E[|y_t|^2] = \infty$ for one of two reasons. First, $\epsilon_t$ may be i.i.d. and $E[\epsilon_t^2] = \infty$; or second, $\epsilon_t$ may be non-i.i.d. and exhibit stochastic recurrence type feedback, covering standard and asymmetric GARCH processes (e.g. Mikosch and Stáricá, 2000; Liu, 2006).

If $\epsilon_t$ is i.i.d. then we take $\sigma_t = 1$ for all $t$. Otherwise $E[u_t] = 0$ and $E[u_t^2] = 1$, and the random volatility component $\sigma_t$ is measurable with respect to $\mathcal{F}_{t-1}$, governed by a non-degenerate distribution, and bounded from below $\sigma_t \geq \omega$ a.s. for some constant $\omega > 0$. We assume the i.i.d. error $u_t$ has a distribution symmetric about zero in order to ensure our bounded transformed estimating equations identify $\theta_0$. This is clearly a shortcoming since conventional estimators for models with GARCH-like errors do not require the i.i.d. component $u_t$ to have a symmetric distribution, including QML, GMM and GEL (Francq and Zakoan, 2004; Chan and Ling, 2006; Skoglund, 2010). However, the model error $\epsilon_t = \sigma_t u_t$ itself may be asymmetrically distributed due to volatility $\sigma_t^2$. Moreover, it is well known that robustness by a bounded transform either ensures identification by assumption of distribution symmetry of an error term (e.g. Sakata and White, 1998; Cizek, 2008; Hill, 2013); or incurs bias that is removed either by simulation based methods or by modeling the bias, in both cases based on assuming an underlying distribution (e.g. Ronchetti and Trojani, 2001; Mancini et al., 2005). Model (1)-(2) therefore covers AR processes with GARCH or asymmetric GARCH errors, and it is only a matter of expanding the definition of $\mathcal{F}_t$ to allow for stochastic volatility $\sigma_t^2$. Other regressors, as well as nonlinear structure in the conditional mean response function, can be easily added, although the cost is cumbersome notation. See Hill (2013, Section 5) for discussion.

We deliver a robust GEL estimator for $\theta_0$ by using a redescending transform of the error $\epsilon_t$ and weighted least squares instruments. The estimating equations are then imbedded in a smooth criterion function that covers at least Empirical Likelihood, Continuously Updated Estimator (Hansen et al., 1996) and Exponential Tilting (Kitamura and Stutzer, 1997), and in general Minimum Discrepancy estimators in the Cressie and Read (1984) class. See Newey and Smith (2004).

In practice over identifying moment conditions may be desired, for example if the order $p$ is unknown. Let $z_t = [z_{i,t}]_{i=1}^{q}$ be the instruments that include AR regressors $x_t$, hence $q \geq p + 1$. Assume $z_t$ is $\mathcal{F}_{t-1}$-measurable for simplicity, but as above $\mathcal{F}_t$ may be expanded for more general cases. Let $W_t = [W_{i,t}]_{i=1}^{q} \in \mathbb{R}$ be any stationary, $\mathcal{F}_{t-1}$-measurable vector weight that satisfies:

$$\|W_t z_t\| \leq K \text{ a.s. and } E[|W_t z_{i,t} y_{t-j}|] \leq K \text{ for each } 1 \leq i \leq q \text{ and } 1 \leq j \leq p,$$

(3)

and let $Z_t = [Z_{i,t}]_{i=1}^{q} \equiv [W_{i,t} z_{i,t}]_{i=1}^{q}$ be the weighted instruments. Note that $\| \cdot \|$ is the spectral norm:
see notation conventions at the end of this section. Thus by (3) it follows $\|Z_t\| \leq K$ a.s. and the gradient is bounded $\|E[Z_t x_t^*]\| \leq K$. Our estimating equations are defined as

$$m_t(\theta) \equiv \left[ (y_t - \theta' x_t) \times W_{i,t} z_{i,t} \right]_{i=1}^q = \left[ \epsilon_t(\theta) \times Z_{i,t} \right]_{i=1}^q = \epsilon_t(\theta) \times Z_t.$$  

(4)

Under exact identification $z_t = x_t$ our estimator identically minimizes the weighted least squares criterion $\sum_{t=1}^n W_t (y_t - \theta' x_t)^2$.

Although $Z_t$ is bounded, this equation type is not robust to heavy tailed errors. However, use of a so-called redescending transformation $\psi(\epsilon_t(\theta), c)$ for some threshold parameter $c > 0$ together with bounded weighted instruments $Z_t$ ensure identification of $\theta_0$ and asymptotic normality. Redescending functions are usually constructed with a fixed threshold $c$ for estimation in the presence of contaminated data, one-off events and additive outliers. See Hampel et al. (1986), Jureckova and Sen (1996), Cizek (2008) and their references.

The transform $\psi(\epsilon, c)$ satisfies $\psi(\epsilon, c) \rightarrow 0$ as $|\epsilon|$ increases above $c$, and $\psi(\epsilon, c) \rightarrow \epsilon$ as $c \rightarrow \infty$. We use a central order statistic of $|\epsilon_t(\theta)|$ for $c$, hence our threshold $c \rightarrow (0, \infty)$ as the sample size $n \rightarrow \infty$. We specifically assume $\psi(\epsilon, c) = 0$ if $|\epsilon| > c$ and $|\psi(\epsilon, c)| \leq c$ to keep mathematical arguments short. Classic examples include simple trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$ where $I(\cdot)$ is the indicator function, cf. Huber (1964), Tukey’s bisquare $\epsilon(1 - (\epsilon/c)^2)^2 I(|\epsilon| \leq c)$, and an exponential variant $\epsilon \exp\{-\epsilon/c\} I(|\epsilon| \leq c)$, thus $\psi(\epsilon, c)$ operates like a smoothed trimmed version of $\epsilon$. In general truncation $\psi(\epsilon, c) = \text{sign}(\epsilon) \times \min\{|\epsilon|, c\}$ results in estimator bias by our method.

Examples of weights $W_t$ satisfying (3) are $1/\{\sum_{i<j}^q (1 + |z_{i,t}|)(1 + |z_{j,t}|) + \sum_{i=1}^p y_{i-1}^2\}$ and $1/\{\prod_{i=1}^q (1 + z_{i,t}^2)^{1/2} + \sum_{i=1}^p y_{i-1}^2\}$, and if $E[y_t] < \infty$ then also $1/\{\prod_{i=1}^q (1 + z_{i,t}^2)^{1/2}\}$. Ling (2005) weighs an LAD criterion for heavy tail robust AR model estimation with an i.i.d. error. We allow for non-i.i.d. errors and over identification, of course. In order to compare our weights, for simplicity drop the constant term and let $p = 1$, hence $x_t = y_{t-1}$. Under exact identification we minimize $\sum_{t=1}^n W_t (y_t - \theta y_{t-1})^2$ where $|W_t y_{t-1}| \leq K$ a.s. and $E[W_t y_{t-1}^2] \leq K$. Conversely, Ling (2005) minimizes $\sum_{t=1}^n \tilde{w}_t (y_t - \theta' x_t)$ where $\tilde{w}_t$ is positive and $\sigma(y_{t-1})$-measurable, but also $E[\tilde{w}_t (y_t - \theta' x_t)] < \infty$ to ensure asymptotic normality, hence $E[\tilde{w}_t (y_t - \theta_0 y_{t-1}^3)^3] < \infty$ must hold. Thus Ling (2005) does not impose boundedness $|W_t y_{t-1}| \leq K$ a.s. but requires a higher moment bound $E[\tilde{w}_t (y_t - \theta_0 y_{t-1}^3)^3] < \infty$, hence in principle we could lead to a less efficient estimator.

Finally, we provide a consistent estimator of the asymptotic variance based on a Jacobian estimator for non-smooth equations in Pakes and Pollard (1989), amongst others. However, since Wilks’ Theorem holds in general for our transformed version of $m_t(\theta)$, the empirical likelihood method leads to confidence regions for $\theta_0$ (cf. Owen, 1990). The well known benefit is that we by-pass Jacobian and covariance matrix estimation required for inference.

Our focus is robustness against heavy tailed errors, or so-called innovation outliers as in Hill (2013) and Hill and Aguilar (2013). The allowance of additive outliers in the sense of contaminated data in the conditional mean and volatility structures is straightforward, but outliers in the volatility process complicates demonstrating estimator consistency in view of the nonlinear propagation of the outliers (see Muler and Yohai, 2008). Computational methods, however, are presented in Mancini
et al. (2005). Nevertheless, our estimator is based on an asymptotically bounded version of $m_t(\theta)$, hence the influence function of our estimator is bounded, although we do not show it.\footnote{Similar derivations can be found in Cantoni and Ronchetti (2001) and Ronchetti and Trojani (2001).} This implies infinitesimal robustness in the sense of robustness to single point contamination (Hampel, 1974; Hampel et al., 1986).

In a recent paper, Kitamura et al. (2013) show that Beran (1977)’s pioneering Minimum Hellinger Distance Estimator is a GEL estimator in the sense of being a Minimum Discrepancy estimator within the Cressie-Read class (see, e.g., Newey and Smith, 2004, Theorem 2.2). They also show that when the estimating equations have bounded support, or unbounded support and are trimmed with a threshold that increases with the sample size, this GEL estimator has the smallest bias and mean-squared-error [mse] when the model assumptions are not true. Both bias and mse are computed using measures in a $n^{1/2}$-Hellinger neighborhood of the true distribution. Their trimmed GEL estimator, however, is not robust to heavy tails since the supremum of their estimating equations on a neighborhood containing the true parameter must have a fourth moment, while this may imply very stringent moment conditions on the underlying process depending on how the equations are constructed. Moreover, they do not show that their estimator is asymptotically normal and unbiased in its limit distribution, in the sense $n^{1/2}(\hat{\theta}_n - \theta_0 - B_n) \xrightarrow{d} N(0, V)$ where $B_n$ is a bias term that vanishes at rate $n^{1/2}$. $n^{1/2}||B_n|| \to 0$. They only show that it is close to a GEL estimator computed with a distribution near the true one, and that estimator may be biased in its limit distribution. Thus, in what sense their trimmed estimator is robust to heavy tails is unknown, and since the threshold increases with the sample size, the estimator need not be infinitesimally robust.

By contrast, our GEL estimator is based on asymptotically bounded estimating equations and is therefore infinitesimally robust and heavy tail robust, and asymptotically unbiased in the limit distribution (since $B_n = 0$), albeit with a cost of assuming the i.i.d. term $u_t$ in the conditionally heteroscedastic error $\epsilon_t = \sigma_t u_t$ has a symmetric distribution. Moreover, since we work in the broad GEL class, covering Minimum Discrepancy criteria in the Cressie-Read class, Kitamura et al. (2013)’s criteria is allowed here. Hence, when we use their criteria, our GEL estimator is infinitesimally and heavy tail robust, asymptotically normal and unbiased in the limit distribution, and has the smallest bias and mse as defined in Kitamura et al. (2013, Theorems 3.1 and 3.2). Further, as discussed above, asymptotic unbiasedness in the limit distribution when a bounded transform is used requires either distribution symmetry (e.g. Sakata and White, 1998; Cizek, 2008; Hill, 2013); simulation based indirect inference in order to sidestep bias, which requires knowledge of an error distribution (e.g. Ronchetti and Trojani, 2001; Ortelli and Trojani, 2005); or a parametric model of the bias which, again, requires knowledge of the error distribution (e.g. Mancini et al., 2005). In the present paper, we only use symmetry of the i.i.d. term $u_t$, allowing for an asymmetrically distributed conditionally heteroscedastic error $\epsilon_t = \sigma_t u_t$.

In a similar vein as Kitamura et al. (2013), Broniatowski and Keziou (2012) present a dual class of minimum divergent estimators that, in special cases, are GEL estimators. Toma (2013) shows Broniatowski and Keziou (2012)’s class of estimators are infinitesimally robust. However,
Broniatowski and Keziou (2012, Assumption 1) require higher moments of their estimating equations, hence their estimator is not heavy tail robust. Further, it is known that Empirical Likelihood obtains the smallest higher order bias relative to GMM and GEL estimators (Newey and Smith, 2004): it is unknown whether the non-GEL members of Broniatowski and Keziou (2012)’s class have such an optimality property.


Section 2 contains the main results and in Section 3 we discuss inference. A simulation study is presented in Section 4, and Section 5 contains parting comments.

Let $[z]$ be the integer part of $z$; $K$ is a positive constant, the value of which may change from line to line; $|A|$ and $||A||$ are respectively $L_1$ and spectral norms for matrix $A$, i.e. $||A|| = (\lambda_{\text{max}}(A^TA))^{1/2}$ where $\lambda_{\text{max}}(\cdot)$ is the maximum eigenvalue. A function $A(\theta)$ is written $A = A(\theta_0)$ and if it is differentiable then $(\partial/\partial \theta)A = (\partial/\partial \theta)A(\theta)|_{\theta_0}$.

## 2 GEL with Equation Tail-Trimming

We first build the equations and estimator, and then state the assumptions and main results. Recall $\mathbb{S}_t \equiv \sigma(y_t : \tau \leq t)$.

### 2.1 Transformed Equations

We transform the errors based on how they relate to a stochastic threshold derived from the order statistics $\epsilon_{(1)}(\theta) \geq \cdots \geq \epsilon_{(n)}(\theta) \geq 0$ of the two-tailed error $\epsilon_i(\theta) \equiv |\epsilon_i(\theta)|$. Let $\{k_n(\xi)\}$ be a central order sequence

$$k_n(\xi) = [\xi n] \quad \text{for some } \xi \in (0, 1).$$

In general we do not show dependence on $\xi$, e.g. we write $k_n = k_n(\xi)$. The threshold used is the central order statistic $\epsilon_{(k_n)}^{\text{(a)}}(\theta)$, a common boundary for selecting extreme observations in the robustness literature (cf. Huber, 1964; Hampel et al., 1986; Jureckova and Sen, 1996; Cizek, 2008).

The relationship between $\epsilon_i(\theta)$ and $\epsilon_{(k_n)}^{\text{(a)}}(\theta)$ for robust estimation of $\theta_0$ is based on a mapping $\psi: \mathbb{R} \times [0, \infty) \rightarrow \mathbb{R}$ with the following descending, smoothness and boundedness properties.

**Assumption A (transform).**

1. $\psi(\epsilon, c) = \varphi(\epsilon, c)I(\mid \epsilon \mid \leq c)$ where $\varphi(\cdot, c)$ is for each $c$ Borel measurable; 2. the set $\{\epsilon \in \mathbb{R} : \varphi(\epsilon, c) = 0\}$ for any fixed $c \in (0, \infty)$ contains $c = 0$ and is countable; 3. $\varphi(\epsilon, c)$ is twice continuously differentiable in $\epsilon$ and $c$; 4. for $i = 0, 1, 2$ let $|((\partial/\partial c)^i \varphi(\epsilon, c))| \leq K |\epsilon|^s$ for some finite $s > 0$ and $|((\partial/\partial c)^i \varphi(\epsilon, c))| \leq K \min\{|\epsilon|, c\}$; 5. $\varphi(-\epsilon, c) = -\varphi(\epsilon, c)$; 6. $\varphi(\epsilon, c) \rightarrow \epsilon$ as $c \rightarrow \infty$ for each $\epsilon \in \mathbb{R}$; 7. $\varphi(uc, c) = \varphi(u, c/\sigma) \times \sigma$ for any $u \in \mathbb{R}$ and $\sigma > 0$. 


Remark 1  We use A.2 in order to ensure that a key covariance term is positive definite. In particular, it implies that a Riemann-Stieltjes integral of \( \varphi(\epsilon, c)^2I(\mid \epsilon \mid \leq c) \) is positive for a fixed \( c > 0 \) since \( \varphi(\epsilon, c) = 0 \) for only countably many \( \epsilon \). See Footnote 2 below.

Remark 2  \( \psi(\epsilon, c) = \varphi(\epsilon, c)I(\mid \epsilon \mid \leq c) \) operates like a smoothed \( \epsilon I(\mid \epsilon \mid \leq c) \). Simple trimming \( \varphi(\epsilon, c) = \epsilon, \) Tukey’s bisquare \( \varphi(\epsilon, c) = \epsilon(1 - (\epsilon/c)^2)^2, \) and the exponential \( \varphi(\epsilon, c) = \epsilon \exp\{-(\epsilon/c)^2\} \) are all valid. Notice A.2 applies in these three cases: \( \varphi(\epsilon, c) = 0 \) if and only if \( \epsilon = 0 \) in simple trimming and exponential cases, and for Tukey’s bisquare if and only if \( \epsilon \in \{0, c\}. \)

Now use \( \psi(\epsilon_t(\theta), \epsilon_{(kn)}(\theta)) \), and the instruments \( Z_{i,t} \equiv W_{i,t}z_{i,t} \) in (4), to construct estimating equations:

\[
\hat{m}_{n,t}^*(\theta) = \left[ \hat{m}_{i,n,t}^*(\theta) \right]_{i=1}^q = \left[ \psi \left( \epsilon_t(\theta), \epsilon_{(kn)}(\theta) \right) \times Z_{i,t} \right]_{i=1}^q = \psi \left( \epsilon_t(\theta), \epsilon_{(kn)}(\theta) \right) \times Z_t.
\]

2.2 GEL Estimation

Let \( \rho : \mathcal{D} \to \mathbb{R} \) be a concave function with domain \( \mathcal{D} \) containing zero, and three times continuously differentiable in a neighborhood of zero. Write \( \rho^{(i)}(u) = (\partial/\partial u)^i \rho(u), i = 0, 1, 2, \) and assume the normalizations \( \rho(0) = 0 \) and \( \rho^{(1)} = \rho^{(2)} = -1. \) Ignoring normalizations, examples include CUE-GMM or Euclidean Likelihood \( \rho(u) = -u^2/2 - u \) (Hansen et al., 1996; Antoine et al., 2007); empirical likelihood \( \rho(u) = \ln(1 - u) \) for \( u < 1 \) (Owen, 1990); and exponential tilting \( \rho(u) = -\exp\{u\} \) (Kitamura and Stutzer, 1997).

The GEL criterion is \( 1/n \sum_{t=1}^n \rho(\lambda^t \hat{m}_{n,t}^*(\theta)) \) where the multiplier \( \lambda \in \mathbb{R}^q \) satisfies \( \lambda^t \hat{m}_{n,t}^*(\theta) \in \mathcal{D}. \) The estimator is then framed as a saddle-point optimization problem (see Newey and Smith, 2004):

\[
\hat{\theta}_n = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \widehat{\Lambda}_n(\theta)} \left\{ \frac{1}{n} \sum_{t=1}^n \rho(\lambda^t \hat{m}_{n,t}^*(\theta)) \right\} \quad \text{and} \quad \hat{\lambda}_n = \arg \sup_{\lambda \in \widehat{\Lambda}_n(\theta)} \left\{ \frac{1}{n} \sum_{t=1}^n \rho(\lambda^t \hat{m}_{n,t}^*(\theta)) \right\},
\]

where \( \Theta \) is a compact subset of \( \mathbb{R}^{p+1} \), and \( \widehat{\Lambda}_n(\theta) \) contains those \( \lambda \) such that \( \lambda^t \hat{m}_{n,t}^*(\theta) \in \mathcal{D} \) with probability one for each \( t = 1, ..., n \). Computation of \( \hat{\theta}_n, \hat{\lambda}_n \) is straightforward since standard iterative methods for non-smooth criteria apply. See, e.g., Pakes and Pollard (1989).

2.3 Equations for Asymptotic Representations

In order to characterize the limit distribution of \( \hat{\theta}_n, \hat{\lambda}_n \) we require a non-random quantile sequence \( \{c_n(\theta)\} \) associated with the sample order statistic \( \epsilon_{(kn)}(\theta) \): \( c_n(\theta) \) satisfies

\[
P \left( \mid \epsilon_t(\theta) \mid \geq c_n(\theta) \right) = \frac{k_n}{n} = \frac{[\xi n]}{n} \sim \xi.
\]

(6)
We hide that $c_n(\theta)$ depends on $\xi$. Distribution continuity ensures $\{c_n(\theta)\}$ exists for any $\theta$, and $\epsilon_{(k_n)}(\theta)$ estimates $c_n(\theta)$. See below for distribution assumptions. The equations in this case are

$$m_{n,t}^*(\theta) = [m_{i,n,t}^*(\theta)]_{i=1}^q = [\psi(\epsilon_t(\theta), c_n(\theta)) \times \mathcal{Z}_{i,t}]_{i=1}^q = \psi(\epsilon_t(\theta), c_n(\theta)) \times \mathcal{Z}_t.$$ 

In the simple trimming case $\sup_{\theta \in \Theta} \|1/n^{1/2} \sum_{t=1}^n [\hat{m}_{n,t}(\theta) - m_{n,t}^*(\theta)]\| \to 0$ hence we can work with $m_{n,t}^*(\theta)$ in order to characterize the limit law of $\hat{\theta}_n$. In other error transform cases the relationship between $\hat{m}_{n,t}(\theta)$ and $m_{n,t}^*(\theta)$ is complicated by the appearance of $\epsilon_{(k_n)}(\theta)$ in the smooth functional $\varphi(\epsilon_t(\theta), \epsilon_{(k_n)}(\theta))$. See Lemma A.3 in Appendix A.1.

The equations $m_{n,t}^*(\theta)$ are particularly useful since they employ non-stochastic trimming thresholds, and they identify $\theta_0$ since $m_{n,t}^*$ is a martingale difference with respect to $\mathcal{F}_t$. This follows by noting $\psi(-\epsilon, c) = -\psi(\epsilon, c)$ and $\psi(\sigma, c) = \psi(\sigma/c) \times \psi$ under Assumptions A.5 and A.7, while $\sigma_t \geq \omega > 0$ a.s. and $u_t$ is symmetrically distributed about zero. Now let $F$ be the distribution function of $u_t$. Then by independence of $u_t$ and $\mathcal{F}_{t-1}$-measurability of $\sigma_t$:

$$E[\sigma_t \times E\left[\psi\left(u_t, \frac{c_n}{\sigma_t}\right) | \mathcal{F}_{t-1}\right]] = \sigma_t \times \int_0^\infty \psi\left(u_t, \frac{c_n}{\sigma_t}\right) dF(u) + \sigma_t \times \int_{-\infty}^0 \psi\left(-u_t, \frac{c_n}{\sigma_t}\right) dF(-u) = \sigma_t \times \int_0^\infty \psi\left(u_t, \frac{c_n}{\sigma_t}\right) dF(u) - \sigma_t \times \int_0^\infty \psi\left(-u_t, \frac{c_n}{\sigma_t}\right) dF(u) = 0 \text{ a.s.}$$

Further, by continuity $E|m_{n,t}^*(\theta)| \neq 0$ for $\theta \neq \theta_0$, each $i = 1, \ldots, q$, and all $n \geq N$ for some $N \geq 1$. For small enough $n$ it is possible that trimming forces $E[m_{n,t}^*(\theta)] = 0$, but as $n$ increases eventually $E[m_{n,t}^*(\theta)] \neq 0$ unless $\theta = \theta_0$. See the proof of consistency Theorem 2.1 in Appendix A.2.

2.4 Main Results

Define the moment supremum $\kappa \equiv \sup\{\alpha > 0 : E|\epsilon_t|^\alpha < \infty\}$. If over-identifying restrictions exist such that $q > p + 1$ then the instrument set is $z_t = [z_{i,t}]_{i=1}^q = [z_{i}, w_{i}']$ where $w_{i} \in \mathbb{R}^{q-(p+1)}$ are the additional instruments. The following assumption formally defines the data generating processes for $\{y_t, \epsilon_t, w_t\}$ and the instrument weights $\{W_t\}$.

**ASSUMPTION B (DGP).**

1. All random variables exist on the same complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and all functions of $\theta$ and transformations on $\Theta$ satisfy Pollard (1984, Appendix C)'s permissibility criteria.

2. There exists a unique point $\theta_0$ in the interior of compact $\Theta \subset \mathbb{R}^{p+1}$ such that the roots of $1 - \sum_{i=1}^p \phi_{0,i} z^i$ lie outside the unit circle, and $\epsilon_t$ is a stationary martingale difference with respect to $\mathcal{F}_t \equiv \sigma(y_t : t \leq t)$.

3. $\epsilon_t = \sigma_t u_t$, where $u_t$ is i.i.d. and has a distribution symmetric about zero. $\epsilon_t$ has an absolutely continuous and non-degenerate distribution on $\mathbb{R}$, with a density $(\partial / \partial \alpha) P(\epsilon_t \leq a) \text{ that is bounded}$.
sup_{a \in \mathbb{R}} (\partial/\partial a)P(\epsilon_t \leq a) \leq K, and positive on \mathbb{R}. Specifically, if \epsilon_t is i.i.d. then \epsilon_t = u_t, and \(E|u_t|^t < \infty\) for some tiny \(t > 0\). Otherwise \(E[u_t] = 0\) and \(E[u_t^2] = 1\), and \(\sigma_t\) is \(\mathcal{F}_{t-1}\)-measurable, governed by a non-degenerate, absolutely continuous, bounded and stationary distribution that is positive on \(\mathbb{R}\), \(\sigma_t \geq \omega\) a.s. for some constant \(\omega > 0\), and \(E[\sigma_t^2] < \infty\) for some tiny \(t > 0\).

4. If \(q > p + 1\) then the additional instruments \(w_t \in \mathbb{R}^{q-(p+1)}\) are strictly stationary, \(\mathcal{F}_{t-1}\)-measurable, geometrically \(\beta\)-mixing, governed by a non-degenerate, absolutely continuous distribution on \(\mathbb{R}\). Further, \(w_t\) does not contain linear combinations of \(x_t\).

5. The instrument weights \(W_t\) are stationary, \(\mathcal{F}_{t-1}\)-measurable and satisfy \(W_t \neq 0\) a.s., while \(Z_t = W_tz_t\) satisfies \(|Z_t| \leq K\) a.s. and \(E[Z_t y_{t-j}] \leq K\) for each \(t, j\). Further \(|((\partial/\partial \theta)E[\psi(\epsilon_t(\theta), c(\theta))Z_t]|_{\theta_0}| > 0\) for any continuous bounded \(c : \Theta \rightarrow (0, \infty)\).

6. \(\{y_t, \sigma_t\}\) are geometrically \(\beta\)-mixing.

**Remark 3** B.1 is invoked since we require probabilities and expectations of majorants of functions of the data. Pollard (1984, Appendix C)'s criteria ensures majorants of functions of the data on \((\Omega, \mathcal{F}, \mathcal{P})\) are \(\mathcal{F}\)-measurable. In particular, probabilities and expectations of majorants are outer probabilities and expectations (cf. Dudley, 1978; Hoffman-Jørgensen, 1984; Pollard, 1984). As an example, a well known problem is the supremum: we require ULLN Lemma A.1 of Appendix A.1, which shows \(\sup_{\theta \in \Theta} \|1/n \sum_{t=1}^n \{g_n(\epsilon_t(\theta), z_t) - E[g_n(\epsilon_t(\theta), z_t)]\} \|_p \rightarrow 0\) for some Borel measurable mapping \(g_n : \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}\) for each \(n \in \mathbb{N}\); and UCLT Lemma A.6, which shows \(E[\sup_{\theta \in \Theta} |n^{-1/2} \sum_{t=1}^n \zeta'(m_{n,t}(\theta)) - E[m_{n,t}(\theta)]|] = O(1)\) where \(\zeta\zeta' = 1\). Thus, \(\sup_{\theta \in \Theta} \|1/n \sum_{t=1}^n \{g_n(\epsilon_t(\theta), z_t) - E[g_n(\epsilon_t(\theta), z_t)]\} \|\) and \(\sup_{\theta \in \Theta} |n^{-1/2} \sum_{t=1}^n \zeta'(m_{n,t}(\theta)) - E[m_{n,t}(\theta)]|\) must be \(\mathcal{F}\)-measurable. Pollard (1984)'s framework ensures such measurability.

**Remark 4** In B.3 we assume the distribution of \(\epsilon_t\) has support \(\mathbb{R}\) to allow for heavy tails. We assume \((\partial/\partial a)P(\epsilon_t \leq a)\) is positive on \(\mathbb{R}\) in order to obtain an expansion for the central order statistic \(\epsilon_{\frac{a}{(a)(k_n)}}\), used for trimming, where the expansion links \(\epsilon_{\frac{a}{(a)(k_n)}}\) asymptotically to a partial sum of indicator functions for which a limit theory exists. See Lemma A.2 in Appendix A.1. Finally, we ensure the trimmed estimating equations identify \(\theta_0\) by assuming the i.i.d. term \(u_t\) in the error \(\epsilon_t = \sigma_t u_t\) has a symmetric distribution.

**Remark 5** If \(\epsilon_t\) is i.i.d. then we only require a trivial moment condition \(E|\epsilon_t|^t < \infty\) for tiny \(t > 0\). Conversely, in the non-i.i.d. case we impose standard GARCH-like moments \(E[u_t] = 0\) and \(E[u_t^2] = 1\), allowing \(\epsilon_t\) to be very heavy tailed through the conditional variance term \(\sigma_t^2\) (cf. Basrak et al., 2002; Liu, 2006).

**Remark 6** The B.5 Jacobian bound \(|((\partial/\partial \theta)E[\psi(\epsilon_t(\theta), c(\theta))Z_t]|_{\theta_0}| > 0\) rules out degenerate cases, it is standard in the robust estimation literature, and ensures our estimator’s asymptotic variance is positive definite in the case of simple trimming.

**Remark 7** Geometric \(\beta\)-mixing B.6 applies to AR processes \(\{y_t\}\) with standard or asymmetric GARCH errors \(\{\epsilon_t\}\) that have smooth and bounded distributions (eg. Meitz and Saikkonen, 2008). Note \(\beta\)-mixing implies mixing (in the ergodic sense), hence ergodicity: see Petersen (1983).
We now state the main results. See Appendix A.2 for proofs of the main theorems.

**Theorem 2.1** Under Assumptions A and B, \( \hat{\theta}_n = \theta_0 + O_p(1/n^{1/2}) \) and \( \hat{\lambda}_n = O_p(1/n^{1/2}) \).

**Remark 8** In general we require all conditions under Assumption B to prove just consistency \( \hat{\theta}_n \overset{p}{\rightarrow} \theta_0 \) due to the presence of parametric order statistics \( \epsilon_{(k_0)}^{(a)}(\theta) \) in the non-smooth criterion.

Next, \( \hat{\theta}_n \) and \( \hat{\lambda}_n \) are jointly asymptotically normally distributed. In order to characterize the asymptotic variance we need standard GEL matrix components. Let \( r(\theta, \xi) \) be the \( \xi \)-upper tail quantile of \( |\epsilon_t(\theta)| \):

\[
P(|\epsilon_t(\theta)| \geq r(\theta, \xi)) = \xi
\]

and notice \( r(\theta, \xi) \in (0, \infty) \) by distribution positiveness on \( \mathbb{R} \) and \( \xi \in (0, 1) \). Define covariance and scale matrices:

\[
S(\xi) = E\left[\psi(\epsilon_t, r(\xi))Z_tZ_t'\right] \in \mathbb{R}^{q \times q} \quad \text{and} \quad J(\xi) = \frac{\partial}{\partial \theta} E[\psi(\epsilon_t(\theta), r(\theta, \xi)) \times Z_t]\big|_{\theta_0} \in \mathbb{R}^{q \times (p+1)}
\]

\[
V(\xi) = (J(\xi)'S(\xi)^{-1}J(\xi))^{-1} \in \mathbb{R}^{(p+1) \times (p+1)}
\]

\[
P(\xi) = S(\xi)^{-1} - S(\xi)^{-1}J(\xi)(J(\xi)'S(\xi)^{-1}J(\xi))^{-1}J(\xi)'S(\xi)^{-1} \in \mathbb{R}^{q \times q}
\]

\[
A(\xi) = \begin{bmatrix} V(\xi) & 0 \\ 0 & P(\xi) \end{bmatrix} \in \mathbb{R}^{(p+1+q) \times (p+1+q)}.
\]

The matrix \( V(\xi) \) is the usual scale in the GMM and GEL literature, cf. Hansen (1982) and Newey and Smith (2004). Assumptions A.1 and B.5 imply the covariance \( S(\xi) \) is finite and the Jacobian \( J(\xi) \) is finite and has full row rank, hence as long as \( S(\xi) \) is positive definite then \( V(\xi) \) is finite and positive definite. It is easily verified that \( S(\xi) \) is positive definite on \( \xi \in (0, 1) \) since there is conditional variance non-degeneracy for the transformed errors \( E[\psi(\epsilon_t, r(\xi))^2|\mathcal{Z}_{t-1}] > 0 \) a.s., and each weighted instrument \( Z_{i,t} \) provides unique information in the sense of linear independence \( \sum_{i=1}^q \lambda_i Z_{i,t} \neq 0 \) a.s. \( \forall \lambda \in \mathbb{R}^q/0 \). Both properties follow from transform properties A.2, and A.5 and A.7, and data generating properties B.2-B.5.

Notice the Jacobian need not have a simple form due to the non-smooth transform. In the simple trimming case \( \psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c) \), for example, we have

\[
J(\xi) = \frac{\partial}{\partial \theta} E[\epsilon_t(\theta)I(|\epsilon_t(\theta)| \leq r(\theta, \xi)) Z_t]\big|_{\theta_0} = -E \left[ Z_t x_t' I(|\epsilon_t| \leq r(\xi)) \right] + \frac{\partial}{\partial \theta} E[\epsilon_t Z_t I(|\epsilon_t(\theta)| \leq r(\theta, \xi))]
\]

Define \( f(u) \equiv (\partial/\partial u)P(u \leq u) \) the symmetric probability density function of i.i.d. \( u_t \). Then for \( \xi \in (0, 1) \):

\[
E[\psi(\epsilon_t, r(\xi))^2|\mathcal{Z}_{t-1}] = \sigma^2 \times E \left[ \varphi(u_t, r(\xi)/\sigma_t)^2 I(|u_t| \leq r(\xi)/\sigma_t) \right] \geq 2\omega^2 \int_0^{r(\xi)/\sigma_t} \varphi(u, r(\xi)/\sigma_t)^2 f(u)du > 0 \text{ a.s.}
\]

The first inequality is a result of \( \sigma_t \geq \omega > 0 \) a.s., and \( \varphi(-\epsilon, c) = -\varphi(\epsilon, c) \) and \( \varphi(u\sigma, c) = \varphi(u/c/\sigma) \times \sigma \) under A.5 and A.7. The second inequality follows from three properties. First, \( r(\xi)/\sigma_t > 0 \) a.s., since \( r(\xi) \in (0, \infty) \) and \( \sigma_t < \infty \) a.s. by B.3. Second, \( f(u) > 0 \) on \( \mathbb{R} \) by B.3. Third, under A.2 and B.3 \( \varphi(u, r(\xi)/\sigma_t)^2 > 0 \) a.s. for all \( u \in [0, \infty)/\mathcal{U} \) where \( \mathcal{U} \) is a countable set. Next, \( \sum_{i=1}^q \lambda_i Z_{i,t} = \mathcal{W}_t \sum_{i=1}^q \lambda_i z_{i,t} \neq 0 \) a.s. since \( \mathcal{W}_t \neq 0 \) a.s. and \( \sum_{i=1}^q \lambda_i z_{i,t} \neq 0 \) a.s. \( \forall \lambda \in \mathbb{R}^q/0 \) follow directly from B.2-B.5.
In general \( J(\xi) \) does not reduce to \(-E[Z_t x'_t I(\|\xi_t\| \leq r(\xi))]\). Although \( J(\xi) \) is straightforward to estimate as in Pakes and Pollard (1989), the empirical likelihood method allows us to by-pass estimating \( \mathcal{V}(\xi) \) altogether. See Section 3 below.

In non-simple trimming cases \( \psi(\epsilon, c) = \varphi(\epsilon, c)I(\|\epsilon\| \leq c) \) where \((\partial/\partial c)\varphi(\epsilon, c) \neq 0\) as in Tukey’s bisquare, the order statistic \( \epsilon^{(a)}_{(k_n)} \) that appears in \( \varphi(\epsilon_t, \epsilon^{(a)}_{(k_n)}) \) impacts the limit distribution of our estimator. In this case \( \mathcal{V}(\xi) \) comprises only part of the asymptotic variance of \( \hat{\theta}_n \). We show in Appendix A.1 that the limit distribution of \( n^{1/2}(\epsilon^{(a)}_{(k_n)} - c_n) \) is equivalent to the limit for a scaled and standardized indicator sum \( 1/n^{1/2} \sum_{i=1}^{n} \{I(\|\epsilon_i\| > c_n) - E[I(\|\epsilon_i\| > c_n)]\} \). We therefore need the following structures associated with this sum to characterize the asymptotic variance of \( \hat{\theta}_n \).

Define a derivative of \( \varphi(\epsilon, c) \) and the error density:

\[
\delta(\epsilon, c) = \frac{\partial}{\partial c} \varphi(\epsilon, c) \text{ and } f(r) = \frac{\partial}{\partial r} P(\epsilon \leq r).
\]

Now define:

\[
B(\xi) \equiv \frac{\xi}{f(r(\xi)) + f(-r(\xi))} E[\delta(\epsilon_t, r(\xi))Z_t] \in \mathbb{R}^q
\]

\[
\Upsilon(\xi) \equiv \lim\limits_{n \to \infty} E \left( \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left\{ I(\|\epsilon_i\| > r(\xi)) - E[I(\|\epsilon_i\| > r(\xi))] \right\} \right)^2
\]

\[
R(\xi) \equiv B(\xi)\Upsilon(\xi)B(\xi)' + \sum_{i=1}^{\infty} E \left[ \varphi(\epsilon_1, r(\xi)) I(\|\epsilon_i\| \leq r(\xi)) \{B(\xi)Z'_i + Z_iB(\xi)\}' I(\|\epsilon_{i+1}\| > r(\xi)) \right]
\]

\[
\mathcal{H}(\xi) \equiv \left( J(\xi)'S(\xi)^{-1}J(\xi) \right)^{-1} \text{ and } \hat{\mathcal{H}}(\xi) \equiv \left[ \begin{array}{c} \mathcal{H}(\xi) \\ \mathcal{P}(\xi) \end{array} \right].
\]

Notice \(|\delta(\epsilon, c)| \times I(\|\epsilon\| \leq c) \leq Kc \) by Assumption A; while \( f(r(\xi)) + f(-r(\xi)) > 0 \) follows from \( f(a) > 0 \) for \( a \in \mathbb{R} \) under Assumption B.3, and \( \xi \in (0, 1) \). Thus, \(|B(\xi)| < \infty \). Further, \(|\Upsilon(\xi) < \infty \) and \(|R(\xi) < \infty \) since \( Z_t \) and \( I(\|\epsilon_i\| > r(\xi)) \) are bounded, stationary and geometrically \( \beta \)-mixing (cf. Ibragimov, 1962: Lemma 1.2 or Theorem 1.6).

Denote the total set of parameters and their estimated counterparts: \( \beta^0 = [\theta_0', 0]' \in \mathbb{R}^{p+1+q} \) and \( \hat{\beta}_n \equiv [\hat{\theta}_n', \hat{\lambda}_n']' \).

**Theorem 2.2** Under Assumptions A and B \( n^{1/2}(\hat{\beta}_n - \beta^0) \overset{d}{\to} N(0, A(\xi) + \hat{\mathcal{H}}(\xi)R(\xi)\hat{\mathcal{H}}(\xi)'), \) in particular \( n^{1/2}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, \mathcal{V}(\xi) + \mathcal{H}(\xi)R(\xi)\mathcal{H}(\xi)'), \)

**Remark 9** The asymptotic variance \( \mathcal{V}(\xi) + \mathcal{H}(\xi)R(\xi)\mathcal{H}(\xi)' \) for \( \hat{\theta}_n \) reflects dispersion in the transformed equation sum \( 1/n^{1/2} \sum_{t=1}^{n} m^*_{n,t} \) in the usual GEL form \( \mathcal{V}(\xi) \), as well as dispersion in the tail indicator sum \( 1/n^{1/2} \sum_{t=1}^{n} \{I(\|\epsilon_t\| > c_n) - E[I(\|\epsilon_t\| > c_n)]\} \) which is the driving mechanism for \( \epsilon^{(a)}_{(k_n)} \) in \( \varphi(\epsilon_t, \epsilon^{(a)}_{(k_n)}) \), and the covariance between the two.

**Remark 10** In the simple trimming case, the threshold \( \epsilon^{(a)}_{(k_n)} \) itself does not impact \( \hat{\beta}_n \) asymptotically. Simply note \( \varphi(\epsilon_t, c) = \epsilon_t \) hence \( \delta(\epsilon_t, r(\xi)) = 0 \) a.s. thus \( \mathcal{R}(\xi) = 0 \). This implies \( n^{1/2}(\hat{\beta}_n - \beta^0) \overset{d}{\to} \)
Theorem 3.2

Under Assumptions A and B, point framework is \( \lambda \) computed with trimming parameter likelihood function (cf. Owen, 1990). By the Lagrange multiplier problem in (5) and \( \rho \)

Theorem 3.1


Inference

In practice \( \mathcal{V}(\xi) \) can be estimated since the Jacobian \( J(\xi) \) can be estimated indirectly. Let \( e_j \) be a \( (p + 1) \times 1 \) zero vector with 1 in the \( j^{th} \) position, e.g. \( e_2 = [0, 1, 0, ..., 0]' \). Let \( \hat{\theta}_n(\xi) \) express \( \hat{\theta}_n \) computed with trimming parameter \( \xi \), define profile functions

\[
\hat{\pi}_{n,t}^*(\hat{\theta}_n(\xi)) \quad \text{where} \quad \hat{\pi}_{n,t}^*(\theta) = \frac{\rho^{(1)}(\lambda'\hat{m}_{n,t}^*(\theta))}{\sum_{t=1}^{n} \rho^{(1)}(\lambda'\hat{m}_{n,t}^*(\theta))},
\]

and define estimators \( \hat{S}_n(\xi) \equiv \sum_{t=1}^{n} \hat{\pi}_{n,t}^*(\hat{\theta}_n(\xi)) \times \psi(\epsilon_t(\hat{\theta}_n(\xi)), \epsilon_{(k_n(\xi))}^{(a)})(\hat{\theta}_n(\xi)))^2 \times Z_t Z_t' \) and

\[
\hat{J}_n(\xi) \equiv \left[ \frac{1}{2\epsilon_n} \sum_{t=1}^{n} \hat{\pi}_{n,t}^*(\hat{\theta}_n(\xi)) \times \left\{ \hat{m}_{n,t}^*(\hat{\theta}_n(\xi) + \epsilon_j\epsilon_n) - \hat{m}_{n,t}^*(\hat{\theta}_n(\xi) - \epsilon_j\epsilon_n) \right\} \right]_{j=1}^{p+1},
\]

where \( \{\epsilon_n\} \) is any sequence of constants that satisfies \( \epsilon_n \to 0 \) and \( \epsilon_n n^{1/2} \to \infty \) (cf. Pakes and Pollard, 1989). See also Back and Brown (1993), Kitamura and Stutzer (1997) and Newey and Smith (2004) for theory details on the higher order efficiency improvements of profile weighted moment estimators.

Theorem 3.1 Let \( \hat{V}_n(\xi) \equiv (\hat{J}_n(\xi)'\hat{S}^{-1}_n(\xi)\hat{J}_n(\xi))^{-1} \). Under Assumptions A and B, and \( \epsilon_n \to 0 \) and \( \epsilon_n n^{1/2} \to \infty \), \( \hat{V}_n(\xi) \overset{p}{\to} \mathcal{V}(\xi) \).

In order to sidestep estimating \( \mathcal{V}(\xi) \), confidence regions may be computed by inverting the log-likelihood function (cf. Owen, 1990). By the Lagrange multiplier problem in (5) and \( \rho(0) = 0 \) we obtain the log-GEL representation

\[
l_n(\theta) = 2 \sum_{t=1}^{n} \rho(\lambda'\hat{m}_{n,t}^*(\theta))
\]

where the multiplier in the saddle point framework is \( \lambda = \lambda(\theta) \).

Theorem 3.2 Under Assumptions A and B \( \hat{l}_n(\theta_0) \overset{d}{\to} \chi^2(q) \) and \( \hat{l}_n(\hat{\theta}_n) \overset{d}{\to} \chi^2(q - (p + 1)) \).

Remark 11 The result \( \hat{l}_n(\theta_0) \overset{d}{\to} \chi^2(q) \) can be used for confidence region computation. Let \( \chi_{q,\alpha}^2 \) satisfy \( \Pr(\chi_q^2 > \chi_{q,\alpha}^2) = \alpha \) where \( \chi_q^2 \) is a \( \chi^2(q) \) random variable, and let \( l_n(a; b) \) be the likelihood function evaluated at \( \theta = [a', b']' \) for some partition \( \{a, b\} \). The \( 1 - \alpha \) confidence region for some component \( \theta_{0,i} \) of the vector \( \theta_0 = [\theta_{0,j}]_{j=1}^{p+1} \) with significance level \( \alpha \in (0, 1) \) is

\[
\mathcal{I}_{1-\alpha} = \left\{ \hat{\theta}_{0,i} : l_n(\hat{\theta}_{0,i}; [\theta_{0,j}]_{j \neq i}) \leq \chi_{q,\alpha}^2 \right\}.
\]

The second result \( \hat{l}_n(\hat{\theta}_n) \overset{d}{\to} \chi^2(q - (p + 1)) \) for non-robust estimation problems was suggested by Smith (1997) as a way to test for over-identification as in Hansen (1982)'s GMM setting.
4 Simulation Study

We now perform a Monte Carlo experiment. Let $P_\kappa$ denote a power-law distribution. If $\epsilon_t$ is distributed $P(\epsilon_t \leq -a) = P(\epsilon_t \geq a) = .5(1 + a)^{-\kappa_\epsilon}$ for $a \geq 0$. If $\kappa_\epsilon > 2$ then we standardize $\epsilon_t$ such that $E[\epsilon_t^2] = 1$.

We estimate $\rho$ using EL and CUE criterion functions. In the case of no trimming, the bias, rmse and the confidence regions are generally larger, in particular when $\epsilon$ is similar, hence we relegate them to the supplemental material Hill (2014, Tables T.2-T.6). In the case of error $\epsilon$ for each estimator and sample size. The remaining estimation results are qualitatively similar, and we control for how many trimming fractiles are used, and heaviness of tails.

4.1 Experiment 1: Small Subset of $\xi$

We estimate $\theta_0$ for each model using two instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1/\prod_{i=1}^{2}(1 + y_{t-2i})^{1/2}$, hence there is over-identification. It is easily verified that Assumption B.5 holds. We use EL and CUE criterion functions $\rho(u) = \ln(1 - u)$ for $u < 1$ and $\rho(u) = -u^2/2 - u$, and one of three redescending transforms $\psi(\epsilon, c)$: simple trimming $\epsilon I(|\epsilon| \leq c)$, Tukey’s bisquare $(1 - (\epsilon/c)^2)^2 I(|\epsilon| \leq c)$ and exponential $\epsilon \exp\{-((\epsilon/c)^2)\} I(|\epsilon| \leq c)$. The fractile is $k_n = \max\{1, (\xi_n)^\prime\}$ for $\xi \in \{0, .01, .05, 10, .20, .30\}$, where $\xi = 0$ corresponds to the case of no trimming. We explore a larger and finer grid of $\xi$ below. We compute empirical likelihood 95% regions $\mathcal{I}_{.95}$ based on formula (10) for each estimator by centering at $\theta = \hat{\theta}_n$ and increasing and decreasing $\theta$ by .001, and we compute coverage probabilities based on the average region over all simulated samples.

Table 1 summarizes the results for EL in the simple trimming case by reporting the simulation bias, median, average 95% confidence regions and their coverage probabilities, and the root-mean-squared-error [rmse] for each estimator and sample size. The remaining estimation results are qualitatively similar, hence we relegate them to the supplemental material Hill (2014, Tables T.2-T.6). In the case of no trimming, the bias, rmse and the confidence regions are generally larger, in particular when $\epsilon_t$ is heavy tailed.

---

4Starting values are $y_1 = \epsilon_1$ and $\sigma_1^2 = 1$. We generate 1000 observations and retain the last n.

4In the GARCH case $P(|\epsilon_t| > c) = dc^{-\kappa_\epsilon}(1 + o(1))$ where the tail index $\kappa_\epsilon$ satisfies $E[.3u_t^2 + .406]^2 = 1$ (Mikosch and Stărică, 2000). We draw $N = 100,000$ i.i.d. $u_t$ from $P_{2.5}$ or $P_{4.5}$ and compute $\kappa = \arg\min_{\kappa \in \mathcal{K}} |1/N \sum_{t=1}^N |.3\epsilon_t^2 + .406|^{1/2} - 1|$ where $\mathcal{K} = \{.001, .002, ..., 10\}$. We repeat this 10,000 times giving median values $\kappa_\epsilon \in \{2.13, 2.24\}$ for $\kappa_u \in \{2.5, 4.5\}$. See Mikosch and Stărică (2000).
In general, bias, median, rmse and the confidence region width increase when the trimming fractile is large $\xi > .10$. Bias is also smaller and the confidence regions are tighter for larger $n$, as expected, and bias and the confidence regions tend to be larger when the errors are conditionally heteroscedastic. Further, bias exhibits a nonlinearity: in general it is negative for small $\xi \in \{0,.05,.10\}$ and becomes positive for larger $\xi \in \{.20,.30\}$, where the value $\xi$ that aligns with roughly zero bias appears to be sensitive to heaviness of tails, conditional heteroscedasticity, and sample size. The untrimmed estimator is generally suboptimal to the trimmed estimator with $\xi \in \{.05,.10\}$ in the heaviest tailed cases, specifically i.i.d. $\epsilon_t$ distributed $P_{1.5}$, or GARCH $\epsilon_t$ with i.i.d. $u_t$ distributed $P_{2.5}$. Simple trimming, Tukey’s bisquare, and the exponential transform each lead to qualitatively similar results. Overall EL and CUE perform roughly the same.

4.2 Experiment 2: Large Grid of $\xi$

As a separate experiment, we compute the EL estimator with simple trimming, instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1/\prod_{i=1}^2(1 + y_{t-i}^2)^{1/2}$, over the grid $\xi \in \{0,.01,.02,...,.40\}$. In Figure 1 we plot the simulation median and average 95% confidence regions over $\xi$ for each AR model when $n = 100$. In Figure 2 we plot simulation bias, median, rmse, and 95% coverage probabilities over $\xi$ when $n = 100$. See Hill (2014, Figures F.1-F.4) for all figures with both $n \in \{100,500\}$. Similar to the results in Table 1, bias, median, rmse and the confidence region width increase when the trimming fractile is large $\xi > .10$, while values $\xi \in [.01,.10]$ lead to the best results in the heavy tailed case, and generally in thin tailed cases. Figure 1 shows monotonically expanding confidence regions once $\xi > .10$ when $\epsilon_t$ is i.i.d. $P_{1.5}$, but in all other cases a noticeable expansion starts closer to $\xi > .05$.

Further, we now clearly see that absolute bias, median and the confidence region width increase monotonically as $\xi$ increases roughly above .10 in each case. Conversely, in all cases except the IGARCH error with a Gaussian shock, the rmse is inversely parabolic: it tends to be small for small and large $\xi$. Since bias eventually increases monotonically, this arises due to a strong monotonic drop in dispersion for large $\xi$. The exception is IGARCH: rmse is monotonically increasing due to a monotonic increase in bias and dispersion.

Both Figures 1 and 2, and Table 1, reveal that for values $\xi \geq .15$ our estimator behaves starkly different, and indeed suboptimal in terms of bias, rmse and confidence region width. Thus, in summary of the results so far, in general for the sharpest confidence regions and median, and the smallest bias and rmse, values $\xi \in [.01,.05]$ seem best, but any $\xi \in [.01,.10]$ generally works very well.

Coverage probability is based on the empirical likelihood method of inverting the likelihood function to obtain confidence regions based on the asymptotic chi-square distribution, and computing the percent occurrence of the true $\theta_0 = .90$ in the region. Hence, the results are slightly different than those for the bias, median and rmse of the EL estimator. Interestingly, despite the strong monotonic increase in bias and the median in the IGARCH case, the coverage probability is fairly close to 95% for nearly all $\xi$. In Figure 2 coverage probability is closer to 95% in all other cases when $\xi$ is small, except when $\epsilon$ is i.i.d. $P_{1.5}$, in which case coverage is closest to the nominal level for large $\xi > .15$. The reason is revealed to some degree by the confidence region plots in Figure 1. We see that of the
five cases, when \( \epsilon \) is i.i.d. \( P_{1.5} \) the region width remains relatively constant until \( \xi > .15 \), when it begins to expand to a width that, apparently, approximately covers 95%. This reveals that basing the empirical likelihood method on the asymptotic chi-square distribution when tails are heavy works only under a greater degree of trimming, and strongly fails without trimming.

### 4.3 Experiment 3: Very Heavy Tails

In our last experiment, we use heavier tailed data and a smaller weight. We simulate the AR process

\[
y_t = \theta_0 y_{t-1} + \epsilon_t \quad \text{with} \quad \theta_0 = .9, \quad \text{where} \quad \epsilon_t \text{ is i.i.d. and } P_{\kappa_\epsilon} \text{ distributed with } \kappa_\epsilon \in \{.75, 1.5, 2.5\}.\]

It follows \( E[\epsilon_t] = 0 \) when \( \kappa_\epsilon > 1 \) since \( P_{\kappa_\epsilon} \) is symmetric at zero. The instruments are \( z_t = [y_{t-1}, y_{t-2}]' \), and in this case the weight is \( W_t = 1/ \prod_{i=1}^{2} (1 + y_{t-i}^2) \), so that again Assumption B.5 holds. The results for EL with simple trimming and \( \xi \in \{0, .01, .10, .20, .30\} \) are reported in Table 2, and Figures 3 and 4 contain corresponding plots of confidence regions, bias, median, rmse and coverage probability over \( \xi \in \{.01, .02, \ldots, .40\} \) when \( n = 100 \). See Hill (2014, Tables T.7-T.9, Figures F.5-F.8) for all tables and plots based on EL with Tukey’s bisquare and exponential transforms, and each \( n \in \{100, 500\} \).

In the very heavy tailed case \( \kappa_\epsilon = .75 \) the untrimmed estimator is highly biased. Further, relative to thinner tail cases, when tails are very heavy we may trim more and achieve improvements in bias, rmse, and confidence region width. Indeed, bias, rmse and the confidence region width decrease monotonically over \( \xi \in \{0, \ldots, .40\} \). The median is relatively flat and near the true \( \theta_0 = .90 \) for all \( \xi \in \{.05, \ldots, .35\} \), with deviation from the true value for small \( \xi < .05 \) and large \( \xi > .35 \). When \( \kappa_\epsilon > 1 \) then the optimal range of \( \xi \) falls, as above, to \( [.01, .10] \). The confidence regions for the trimmed estimator are noticeably broader when \( \kappa_\epsilon = .75 \), but otherwise the estimator works well.

If \( n = 500 \) then across all experiments the only difference worth noting occurs with the coverage probability in the very heavy tailed case \( \kappa = .75 \) (Hill, 2014, Figure F.8). If \( n = 100 \) then coverage tends to be too small except when \( \kappa = 2.5 \). However, at the larger sample size \( n = 500 \) coverage improves substantially, and is now accurate even in the very heavy tailed case \( \kappa = .75 \), provided \( \xi \in [.03, .07] \). Recall that we are using a smaller weight \( W_t \) than in the previous experiments. Hence, the results for \( \kappa \in \{1.5, 2.5\} \) are not surprisingly slightly different than those discussed in Sections 4.1 and 4.2.

### 5 Concluding Remarks

We develop a heavy tail robust GEL estimator and confidence regions for autoregressions by transforming weighted least squares estimating equations. The model error may be heteroscedastic of unknown form as long as measurability and mixing conditions hold, covering at least AR models with symmetric or asymmetric GARCH errors. The transformations are based on a class of redescending functions, allowing for heavy tailed errors, while the weight allows for heavy tailed regressors. A simulation study shows our estimator with EL and CUE criterion functions works well for AR models with i.i.d. or GARCH errors, resulting in comparatively low bias, sharp median, small mean-squared-
error and tight confidence regions in most cases studied. The most challenging case involves a process
with an i.i.d. error that has an infinite mean, but even here our estimator works well, and is far
better than the untrimmed estimator.

Acknowledgements

We would like to thank three referees for constructed comments that lead to a much improved
manuscript.

A Appendix: Proofs of Main Theorems

Recall \( \rho(0) = 0 \) and \( \rho(1) = \rho(2) = -1 \). Throughout \( o_p(1) \) terms do not depend on \( \theta \) and \( \lambda \).
We drop the trimming quantile \( \xi \) from all matrix arguments for notational ease. Let \( w.p.a.1 \) denote "with probability approaching one".

We require the following definitions:

\[
\hat{Q}_n(\theta, \lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} \rho(\lambda^t \hat{m}^*_n t(\theta)) \quad \text{and} \quad \Lambda_n \equiv \left\{ \lambda : \|\lambda\| \leq Kn^{-1/2} \right\}
\]

\[
m^*_n(\theta) \equiv \frac{1}{n} \sum_{t=1}^{n} m^*_n t(\theta) \quad \text{and} \quad \hat{m}^*_n(\theta) \equiv \frac{1}{n} \sum_{t=1}^{n} \hat{m}^*_n t(\theta)
\]

\[
S_n \equiv E \left[ m^*_n t m^*_n t \right], \quad S = E \left[ \psi(\epsilon_t, r)^2 Z_t Z'_t \right], \quad \mathcal{J}_n = \frac{\partial}{\partial \theta} E \left[ m^*_n t \right] \quad \text{and} \quad \mathcal{J} = \lim_{n \to \infty} \mathcal{J}_n.
\]

Let \( \Rightarrow^* \) denote weak convergence on a Polish space as in Hoffman-Jörjensen (1984). See also Dudley (1978) and Doukhan et al. (1995) for background details and deep results.

A.1 Preliminary Results

The proofs of the main theorems utilize the following results. We present primitive results first that
follow from the assumptions. In all cases Assumptions A and B hold. We begin with a required
general ULLN for tail-trimmed sequences, and an order statistic expansion. Recall the instrument
set is \( z_t = [x'_t, w'_t] \).

Lemma A.1 (ULLN) Let \( g_n(\epsilon, z) \) be for each \( n \in \mathbb{N} \) a Borel measurable mapping from \( \mathbb{R} \times \mathbb{R}^q \) to \( \mathbb{R} \)
that satisfies \( \sup_{\theta \in \Theta} \left| g_n(\epsilon_t(\theta), z_t) \right| \leq K \) a.s. Then \( \sup_{\theta \in \Theta} |1/n \sum_{t=1}^{n} \{ g_n(\epsilon_t(\theta), z_t) - E[g_n(\epsilon_t(\theta), z_t)] \} | \nabla 0.

Proof. Note \( \mathcal{Z}_{n,t}(\theta) \equiv g_n(\epsilon_t(\theta), z_t) - E[g_n(\epsilon_t(\theta), z_t)] \) is geometrically \( \beta \)-mixing by measurability and
B.4 and B.6, stationary over \( 1 \leq t \leq n \), and uniformly \( L_s \)-bounded for any \( s > 2 \). Therefore pointwise
\( 1/n \sum_{t=1}^{n} \mathcal{Z}_{n,t}(\theta) \nabla 0 \) by Theorem 2 and Example 4 in Andrews (1988). Next, by construction \( \mathcal{Z}_{n,t}(\theta) \)
is uniformly \( L_1 \)-bounded, hence it belongs to a separable Banach space. This implies the \( L_1 \)-bracketing
numbers satisfy \( N_{\mathcal{I}}(\|\cdot\|_1) \) < \( \infty \) (Dudley, 1999: Proposition 7.1.7). Combine the pointwise law
and \( N_{\mathcal{I}}(\|\cdot\|_1) \) < \( \infty \) to deduce \( \sup_{\theta \in \Theta} |1/n \sum_{t=1}^{n} \mathcal{Z}_{n,t}(\theta) | \nabla 0 \) by Theorem 7.1.5 of Dudley
(1999).
As usual, $\epsilon(a)_1(\theta) \geq \cdots \geq \epsilon(a)_n(\theta) \geq 0$ are order statistics of the two-tailed error $\epsilon(a)_t(\theta) \equiv |\epsilon_t(\theta)|$, and $\{k_n\}$ is a central order sequence.

**Lemma A.2 (order statistic expansion)** Define $f(a) \equiv (\partial/\partial a)P(\epsilon \leq a)$, and recall $r(\theta)$ satisfies $P(|\epsilon_t(\theta)| \geq r(\theta)) = \xi$, and $r = r(\theta_0)$. Then

$$n^{1/2} \left( \epsilon(a)_{(k_n)} - c_n \right) = \frac{\xi}{f(r) + f(-r)} \times \frac{1}{n^{1/2}} \sum_{t=1}^{n} \{I(|\epsilon_t| > c_n) - E[I(|\epsilon_t| > c_n)]\} + o_P(1).$$

**Proof.** Define $I_{n,t} \equiv I(|\epsilon_t| > c_n)$. We require one preliminary CLT.

**Lemma B.1 (indicator CLT).** Define $I_{n,t}(u) \equiv I(|\epsilon_t| > c_n + u)$ for any $u \in \mathbb{R}$, $Y_n(u) \equiv E(n^{-1/2} \sum_{t=1}^{n} \{I_{n,t}(u) - E[I_{n,t}(u)]\}^2)$ and $Y(u) \equiv \lim_{n \to \infty} Y_n(u)$. Then $n^{-1/2} \sum_{t=1}^{n} \{I_{n,t}(u/n^{1/2}) - E[I_{n,t}(u/n^{1/2})]\} \xrightarrow{d} N(0, Y(0))$ where $Y(0) < \infty$.

**Proof.** Let $U$ be an arbitrary compact subset of $\mathbb{R}$. $I_{n,t}(u)$ is stationary over $1 \leq t \leq n$, geometrically $\beta$-mixing by B.6, and $L_s$-bounded for any $s > 2$. Furthermore, distribution continuity and boundedness B.3 imply the Lipschitz bound $E[(I_{n,t}(u) - I_{n,t}(\bar{u}))^2] \leq K||u - \bar{u}||$ for any $u, \bar{u} \in U$. This implies a metric entropy with $L_s$-bracketing numbers $N_{||.||}(\epsilon, U, ||.||_2)$. See Pollard (1984). Therefore $n^{-1/2} \sum_{t=1}^{n} \{I_{n,t}(u) - E[I_{n,t}(u)]\}$ : $u \in U \implies \{3(u) : u \in U\}$, a Gaussian process with almost surely uniformly continuous sample paths and variance $Y(u) < \infty$, by Theorem 1 of Doukhan et al. (1995). See their (2.17) and Application 4. Hence $n^{-1/2} \sum_{t=1}^{n} \{I_{n,t}(u/n^{1/2}) - E[I_{n,t}(u/n^{1/2})]\} \xrightarrow{d} N(0, \lim_{n \to \infty} Y_n(u/n^{1/2}))$. Finally, $\lim_{n \to \infty} Y_n(u/n^{1/2}) = Y(0)$ in view of $Y_n(u) = O(1), c_n \to r$, and dominated convergence: $E[I_{n,t}(u/n^{1/2})] \to E[\lim_{n \to \infty} I_{n,t}(u/n^{1/2})] = E[I(|\epsilon_t| > r)]$. QED

Under B.3 $f(r) + f(-r) \in (0, \infty)$ since $r \in (0, \infty)$. Now observe that $n^{1/2}(\epsilon(a)_{(k_n)} - c_n) \leq u$ for $u \in \mathbb{R}$ if and only if $1/n \sum_{t=1}^{n} I_{n,t}(u/n^{1/2}) \leq 1$, if and only if

$$\frac{1}{n} \sum_{t=1}^{n} \{I_{n,t}(u/n^{1/2}) - E[I_{n,t}(u/n^{1/2})]\} \leq 1 - \frac{n}{k_n} P\left(|\epsilon_t| > c_n + u/n^{1/2}\right) = 1 - \frac{P(|\epsilon_t| > c_n + u/n^{1/2})}{P(|\epsilon_t| > c_n)}.$$

Expand $P(|\epsilon_t| > c_n + u/n^{1/2})$ around $u = 0$ to obtain for some $|u^*| \leq |u|$:

$$\frac{\xi}{f(r) + f(-r)} \frac{1}{n^{1/2}} \sum_{t=1}^{n} \{I_{n,t}(u/n^{1/2}) - E[I_{n,t}(u/n^{1/2})]\} \leq \frac{\xi}{f(r) + f(-r)} \frac{f(c_n + u^*/n^{1/2}) + f(-c_n + u^*/n^{1/2})}{P(|\epsilon_t| > c_n)} u = u \times (1 + o(1)),$$

where $o(1)$ is non-random. The last equality follows from $P(|\epsilon_t| > c_n) \to \xi$, and $f(\pm(c_n + u^*/n^{1/2})) \to f(\pm r)$ by continuity. Finally, $1/n^{1/2} \sum_{t=1}^{n} \{I_{n,t}(u/n^{1/2}) - E[I_{n,t}(u/n^{1/2})]\}$ has the same limit distribution as $1/n^{1/2} \sum_{t=1}^{n} \{I_{n,t} - E[I_{n,t}]\}$ by Lemma B.1 and the construction $I_{n,t} = I_{n,t}(0)$. This proves the claim. 

Next, we show $m^*_n(\theta)$ based on trimming with a stochastic threshold $\epsilon(a)_{(k_n)}(\theta)$, can be replaced with $m^*_n(\theta)$ based on a non-stochastic threshold $c_n(\theta)$.
Lemma A.3 (approximation) Define $\delta(\epsilon, c) \equiv (\partial/\partial c)\varphi(\epsilon, c)$ and let $r(\theta)$ satisfy $P(|\epsilon_t(\theta)| \geq r(\theta)) = \xi \in (0, 1)$. Then the following is $o_p(1)$:

$$
\sup_{\theta \in \Theta} \left\| \frac{1}{n^{1/2}} \sum_{t=1}^{n} \{ \hat{m}_{n,t}(\theta) - m_{n,t}(\theta) \} - E[\delta(\epsilon_t(\theta), r(\theta)) I(|\epsilon_t(\theta)| \leq r(\theta)) \mathcal{Z}_t] \times n^{1/2} \left\{ \epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta) \right\} \right\|
$$

**Proof.** Define $\hat{\mathcal{E}}_{n,t}(\theta) \equiv |\epsilon_t(\theta)| - \epsilon^{(a)}_{(k_n)}(\theta)$, $\mathcal{E}_{n,t}(\theta) \equiv |\epsilon_t(\theta)| - c_n(\theta)$, and $\mathcal{E}_t(\theta) \equiv |\epsilon_t(\theta)| - r(\theta)$ where $\sup_{\theta \in \Theta} |c_n(\theta) - r(\theta)| \to 0$, and write $I(u) = I(u \leq 0)$. Write $\hat{\varphi}_{n,t}(\theta) = \varphi(\epsilon_t(\theta), \epsilon^{(a)}_{(k_n)}(\theta))$, $\varphi_{n,t}(\theta) = \varphi(\epsilon_t(\theta), c_n(\theta))$, and define $h(\epsilon, c) \equiv (\partial/\partial c)^2 \varphi(\epsilon, c)$. We repeatedly use $\sup_{\theta \in \Theta} |\epsilon^{(a)}_{(k_n)}(\theta) - r(\theta)| = O_p(n^{-1/2})$ under B.1-B.4 and B.6 by uniform order statistic laws in Cizek (Appendix A 2008); and $\sup_{\theta \in \Theta} |c_n(\theta) - r(\theta)| \to 0$ and $\sup_{\theta \in \Theta} r(\theta) \leq K$ by construction of $\{c_n(\theta), r(\theta)\}$, $\xi \in (0, 1)$, and model linearity and distribution continuity.

Note $1/n^{1/2} \sum_{t=1}^{n} \{ \hat{m}_{n,t}(\theta) - m_{n,t}(\theta) \} = \mathfrak{A}_n(\theta) + \mathfrak{B}_n(\theta)$ where

$$
\mathfrak{A}_n(\theta) \equiv \frac{1}{n^{1/2}} \sum_{t=1}^{n} \hat{\varphi}_{n,t}(\theta) \left\{ I(\hat{\mathcal{E}}_{n,t}(\theta)) - I(\mathcal{E}_{n,t}(\theta)) \right\} \mathcal{Z}_t
$$

$$
\mathfrak{B}_n(\theta) \equiv \frac{1}{n^{1/2}} \sum_{t=1}^{n} \{ \hat{\varphi}_{n,t}(\theta) - \varphi_{n,t}(\theta) \} I(\mathcal{E}_{n,t}(\theta)) \mathcal{Z}_t.
$$

We need only show $\sup_{\theta \in \Theta} \|\mathfrak{A}_n(\theta)\| \overset{p}{\to} 0$ and $\sup_{\theta \in \Theta} \|\mathfrak{B}_n(\theta) - E[\delta(\epsilon_t(\theta), r(\theta)) I(\mathcal{E}_t(\theta)) \mathcal{Z}_t] \times n^{1/2} \left\{ \epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta) \right\}\| \overset{p}{\to} 0$.

First, observe $I(u)$ can be approximated by a continuous, differentiable, uniformly bounded positive function $\mathcal{I}_N(u)$ that has a uniformly bounded derivative $D_N(u) \equiv (\partial/\partial u)\mathcal{I}_N(u)$, where $N \in \mathbb{N}$ guides the approximation: $\lim_{N \to \infty} \sup_{u \in \mathbb{R}} |\mathcal{I}_N(u) - I(u)| = 0$ and $\lim_{N \to \infty} \sup_{u \in \mathbb{R}} |D_N(u)| = 0$ for all $n$. See Hill (2013, proof of Lemma A.1), cf. Lighthill (1958, p. 22). In particular there exists a double array of positive non-random numbers $\{e^{(N)}_n : n, N \in \mathbb{N}\}$, $e^{(N)}_n \downarrow 0$ as $N \to \infty$ for each $n$, such that:

$$
\mathfrak{A}_n(\theta) = \frac{1}{n^{1/2}} \sum_{t=1}^{n} \hat{\varphi}_{n,t}(\theta) \left\{ \mathcal{I}_N(\hat{\mathcal{E}}_{n,t}(\theta)) - \mathcal{I}_N(\mathcal{E}_{n,t}(\theta)) \right\} \mathcal{Z}_t + e^{(N)}_n.
$$

By the mean value theorem for some $\tilde{c}_n(\theta)$, $|\epsilon^{(a)}_{(k_n)}(\theta) - \tilde{c}_n(\theta)| \leq |\epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta)|$:

$$
\|\mathfrak{A}_n(\theta)\| \leq \left\| \frac{1}{n^{1/2}} \sum_{t=1}^{n} \hat{\varphi}_{n,t}(\theta) D_N(\hat{\mathcal{E}}_{n,t}(\theta)) \mathcal{Z}_t \left( \epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta) \right) \right\| + e^{(N)}_n
$$

where $\hat{\mathcal{E}}_{n,t}(\theta) \equiv |\epsilon_t(\theta)| - \tilde{c}_n(\theta)$. The derivative can be made close to zero since $\lim_{N \to \infty} \sup_{u \in \mathbb{R}} |D_N(u)| = 0$ a.s. for all $n$. In particular we can always set $N \to \infty$ as $n \to \infty$ fast enough so ensure $\max_{1 \leq t \leq n} |\sup_{\theta \in \Theta} |\epsilon_t(\theta)|^s D_N(\hat{\mathcal{E}}_{n,t}(\theta))| \overset{p}{\to} 0$ for any $s > 0$ and $e^{(N)}_n \to 0$ as $n \to \infty$ (see the proof of Lemma A.1 in Hill, 2013). Now use $\sup_{\theta \in \Theta} |\epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta)| = O_p(n^{-1/2})$, $|\hat{\varphi}_{n,t}(\theta)| \leq K|\epsilon_t(\theta)|^s$ for some $s > 0$ by A.4, and $|\mathcal{Z}_t| \leq K$ a.s. by B.5 to obtain $\sup_{\theta \in \Theta} \|\mathfrak{A}_n(\theta)\| = O_p(1/n \sum_{t=1}^{n} \sup_{\theta \in \Theta} |\epsilon_t(\theta)|^s \times |D_N(\hat{\mathcal{E}}_{n,t}(\theta))|) + e^{(N)}_n = o_p(1)$.

For $\mathfrak{B}_n(\theta)$, by the mean-value theorem there exists $\epsilon^{(a)}_{(k_n)}(\theta) \leq |\epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta)|$:

$$
\mathfrak{B}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \delta(\epsilon_t(\theta), c_n(\theta)) I(\mathcal{E}_{n,t}(\theta)) \mathcal{Z}_t \times n^{1/2} \left( \epsilon^{(a)}_{(k_n)}(\theta) - c_n(\theta) \right)
$$

(A.1)
Since by A.4 and B.5 \( \sup_{\theta \in \Theta} |\delta(\epsilon_t, \theta) - c_n(\theta)| I(\mathcal{E}_{n,t}(\theta))| \leq \sup_{\theta \in \Theta} c_n(\theta) \leq K \) and \( \|Z_t\| \leq K \ a.s. \) we have by Lemma A.1 \( \sup_{\theta \in \Theta} |1/n \sum_{t=1}^n \delta(\epsilon_t(\theta), c_n(\theta)) I(\mathcal{E}_{n,t}(\theta)) Z_t - E[\delta(\epsilon_t(\theta), c_n(\theta)) I(\mathcal{E}_{n,t}(\theta)) Z_t]| \xrightarrow{p} 0. \)

Since \( \sup_{\theta \in \Theta} |\epsilon_{k_n}(\theta) - c_n(\theta)| = O_p(n^{-1/2}) \), the proof is complete if we show \( \sup_{\theta \in \Theta} |1/n \sum_{t=1}^n \delta(\epsilon_t(\theta), c_n(\theta)) \times I((\mathcal{E}_{n,t}(\theta)) Z_t| \xrightarrow{p} 0. \)

The mean-value-theorem implies for some \( c_n^*(\theta) \), \( |c_n^*(\theta) - c_n^*(\theta)| \leq |c_n^*(\theta) - c_n(\theta)| \):

\[
\frac{1}{n} \sum_{t=1}^n \{\delta(\epsilon_t(\theta), c_n^*(\theta)) - \delta(\epsilon_t(\theta), c_n(\theta))\} I(\mathcal{E}_{n,t}(\theta)) Z_t = \frac{1}{n} \sum_{t=1}^n h(\epsilon_t(\theta), c_n^*(\theta)) I(\mathcal{E}_{n,t}(\theta)) Z_t \times \{c_n^*(\theta) - c_n(\theta)\}. \tag{A.2}
\]

Write \( \mathcal{E}_{n,t}^* \equiv |\epsilon_t(\theta) - c_n^*(\theta)| \), and note \( \sup_{\theta \in \Theta} |c_n^*(\theta) - c_n(\theta)| \xrightarrow{p} 0 \) by construction. By the indicator smoothing argument above \( \sup_{\theta \in \Theta} |1/n \sum_{t=1}^n h(\epsilon_t(\theta), c_n^*(\theta)) I(\mathcal{E}_{n,t}(\theta)) - I(\mathcal{E}_{n,t}^*(\theta)) Z_t| \xrightarrow{p} 0 \) \( \sup_{\theta \in \Theta} r(\theta) \leq K \) to conclude \( \|1/n \sum_{t=1}^n \{\delta(\epsilon_t(\theta), c_n^*(\theta)) - \delta(\epsilon_t(\theta), c_n(\theta))\} I(\mathcal{E}_{n,t}(\theta)) Z_t\| = \|1/n \sum_{t=1}^n h(\epsilon_t(\theta), c_n^*(\theta)) I(\mathcal{E}_{n,t}^*(\theta)) Z_t \times \{c_n^*(\theta) - c_n(\theta)\} + o_p(1) \leq K \sup_{\theta \in \Theta} c_n^*(\theta) \times \sup_{\theta \in \Theta} |c_n^*(\theta) - c_n(\theta)| + o_p(1) \xrightarrow{p} 0 \) as required. \( \blacksquare \)

Lemmas A.2 and A.3 imply the following expansion for \( 1/n \sum_{t=1}^n \tilde{m}_{n,t}^* \):

**Corollary A.4** (expansion for \( \tilde{m}_{n,t}^* \)) Define \( \delta(\epsilon, c) \equiv (\partial/\partial c) \varphi(\epsilon, c) \) and \( f(a) \equiv (\partial/\partial a) P(\epsilon_t \leq a) \), and let \( r \) satisfy \( P(|\epsilon_t| \geq r) = \xi \in (0,1) \). Then

\[
\frac{1}{n} \sum_{t=1}^n \tilde{m}_{n,t}^* = \frac{1}{n} \sum_{t=1}^n m_{n,t}^* + \frac{\xi E[\delta(\epsilon_t, r(\xi)) I(|\epsilon_t| > r) Z_t]}{f(r) + f(-r)} \frac{1}{n} \sum_{t=1}^n \{I(|\epsilon_t| > c_n) - E[I(|\epsilon_t| > c_n)]\} + o_p(1/n^{1/2}).
\]

Recall \( S_n \equiv E[m_{n,t}^* m_{n,t}^*] \) and \( S \equiv E[\psi(\epsilon_t, r)^2 Z_t Z_t'] \).

**Lemma A.5** (covariance) Let \( \{\tilde{\theta}_n\}_{n \geq 1} \) be any stochastic sequence that satisfies \( \tilde{\theta}_n \xrightarrow{p} \theta_0 \) and define \( \tilde{S}_n(\tilde{\theta}_n) \equiv 1/n \sum_{t=1}^n \tilde{m}_{n,t}^*(\tilde{\theta}_n) \tilde{m}_{n,t}^*(\tilde{\theta}_n)' \). Then \( \|\tilde{S}_n(\tilde{\theta}_n) - S_n\| \xrightarrow{p} 0 \) and \( \|S_n - S\| \xrightarrow{p} 0 \) where \( S \) is positive definite.

**Proof.** Write \( \psi_{n,t}(\theta) \equiv \psi(\epsilon_t(\theta), c_n(\theta)) \), \( \varphi_{n,t}(\theta) \equiv \varphi(\epsilon_t(\theta), c_n(\theta)) \) and \( I_{n,t}(\theta) \equiv I(|\epsilon_t(\theta)| \leq c_n(\theta)) \). Drop \( \theta_0 \). By the argument used to prove Lemma A.3: \( \tilde{S}_n(\tilde{\theta}_n) = 1/n \sum_{t=1}^n \psi_{n,t}^2(\tilde{\theta}_n) Z_t Z_t' + o_p(1) \).

Trivially for each \( \theta \in \Theta: \)

\[
\psi_{n,t}^2(\theta) = \varphi_{n,t}^2(\theta) + \{\varphi_{n,t}^2(\theta) - \varphi_{n,t}^2(\theta)\} I_{n,t}(\theta) - I_{n,t}(\theta) = \varphi_{n,t}^2(\theta) + \Psi_{n,t}(\theta), \tag{A.4}
\]
say. The bound $|\varphi(\epsilon,c)I(|\epsilon| \leq c)| \leq c$ by A.4 and $\sup_{\theta \in \Theta} c_n(\theta) \leq K$ imply $\Psi_{n,t}(\theta)$ is uniformly bounded on $\Theta$. Therefore $\sup_{\theta \in \Theta} [1/n \sum_{t=1}^n \Psi_{n,t}(\theta) - E[\Psi_{n,t}(\theta)]] \overset{p}{\to} 0$ by Lemma A.1, hence $1/n \sum_{t=1}^n \Psi_{n,t}(\tilde{\theta}_n) - E[\Psi_{n,t}(\tilde{\theta}_n)] \overset{p}{\to} 0$. In view of continuity and boundedness of $E[\Psi_{n,t}(\theta)]$ on compact $\Theta$, $\Psi_{n,t} = 0$ and $\tilde{\theta}_n \overset{p}{\to} \theta_0$, it follows $E[\Psi_{n,t}(\tilde{\theta}_n)] \to 0$. Therefore $\tilde{S}_n(\tilde{\theta}_n) = 1/n \sum_{t=1}^n \psi_{n,t}^2 Z_t Z_t' + o_p(1)$.

Next, by boundedness A.4 and Lemma A.1 $1/n \sum_{t=1}^n \psi_{n,t}^2 Z_t Z_t' - E[\psi_{n,t}^2 Z_t Z_t'] \overset{p}{\to} 0$. Finally, continuity and $c_n \to r$ imply $E[\psi_{n,t}^2 Z_t Z_t'] = E[\psi(\epsilon_t, c_n)^2 Z_t Z_t'] \to E[\psi(\epsilon_t, r)^2 Z_t Z_t'] = S$. See Footnote 2 for verification that $S$ is positive definite. This completes the proof.

We exploit two central limit theorems: a UCLT for $m_{n,t}^*(\theta)$ helps prove $n^{1/2}$-consistency of $\hat{\theta}_n$, and a CLT for $\psi(\epsilon_t, c_n) Z_t + a \{I(|\epsilon_t| > c_n) + E[I(|\epsilon_t| > c_n)]\}$ with $a \in \mathbb{R}^d$ is used to prove asymptotic normality for $\hat{\theta}_n$.

**Lemma A.6 (UCLT for $m_{n,t}^*(\theta)$)** \{1/n^{1/2} \sum_{t=1}^n \zeta'(m_{n,t}^*(\theta) - E[m_{n,t}^*(\theta)]) : $\theta \in \Theta$\} $\overset{d}{\to}$ \{M($\theta$, $\zeta$) : $\theta \in \Theta$\}, for any conformable $\zeta' \zeta = 1$ a Gaussian process with almost surely uniformly continuous sample paths. Further $E[\sup_{\theta \in \Theta} n^{-1/2} \sum_{t=1}^n \zeta'(m_{n,t}^*(\theta) - E[m_{n,t}^*(\theta)])] = O(1)$ where $| \cdot |$ is the $l_1$-norm.

**Proof.**

**Step 1.** Define

$$M_{n,t}^*(\theta, \zeta) \equiv \zeta'(m_{n,t}^*(\theta) - E[m_{n,t}^*(\theta)]) \quad \text{and} \quad \varphi_{n,t}(\theta) \equiv \varphi(\epsilon_t, c_n(\theta))$$

$$\delta_{n,t}(\theta) \equiv \frac{\partial}{\partial \theta} \varphi_{n,t}(\theta) \quad \text{and} \quad I_{n,t}(\theta) \equiv I(|\epsilon_t(\theta)| \leq c_n(\theta)),$$

Hide dependence on $\zeta$: $M_{n,t}^*(\theta) = M_{n,t}^*(\theta, \zeta)$ and $M(\theta) = M(\theta, \zeta)$. By construction and B.6 $M_{n,t}^*(\theta)$ is $L_2$-bounded uniformly on $1 \leq t \leq n$, $n \geq 1$, and $\Theta$, and is geometrically $\beta$-mixing and stationary over $1 \leq t \leq n$. We show in Step 2 that $\{M_{n,t}^*(\theta) : \theta \in \Theta\}$ satisfies the metric entropy with $L_2$-bracketing bound $\int_0^1 \ln(N_{[\cdot]}(\epsilon, \Theta, || \cdot ||_2))d\epsilon < \infty$ with $L_2$-bracketing numbers $N_{[\cdot]}(\epsilon, \Theta, || \cdot ||_2)$. The first claim therefore follows from Doukhan et al. (1995, Theorem 1, eq. (2.17). Application 4). The second claim is a consequence of Theorem 2 in Doukhan et al. (1995) since their required bound (2.10) holds under their (2.17), which $\int_0^1 \ln(N_{[\cdot]}(\epsilon, \Theta, || \cdot ||_2))d\epsilon < \infty$ ensures.

**Step 2.** The bracketing bound holds if $E[(M_{n,t}^*(\theta) - M_{n,t}^*(\theta))(\theta))^2]^{1/2} \leq K||\theta - \tilde{\theta}||$ for any $\theta, \tilde{\theta} \in \Theta$ (see, e.g. Pollard, 1984). By Minkowski’s inequality, $m_{n,t}^*(\theta) = \varphi_{n,t}(\theta) I_{n,t}(\theta) Z_t$, and $||Z_t|| \leq K$ a.s. by B.5:

$$\left( E\left[ (M_{n,t}^*(\theta) - M_{n,t}^*(\theta))^2 \right] \right)^{1/2} \leq \left[ E[|m_{n,t}^*(\theta)| - E|m_{n,t}^*(\tilde{\theta})|^2] \right]^{1/2} + K \left( E\left[ \left\{ \varphi_{n,t}(\theta) I_{n,t}(\theta) - \varphi_{n,t}(\tilde{\theta}) I_{n,t}(\tilde{\theta}) \right\}^2 \right] \right)^{1/2}.$$

Note $\sup_{\theta \in \Theta} ||E[m_{n,t}^*(\theta)]|| = O(1)$ and $\sup_{\theta \in \Theta} ||J_n(\theta)|| = O(1)$ from A.2, A.4, $\sup_{\theta \in \Theta} |c_n(\theta) - r(\theta)| \to 0$ and $\sup_{\theta \in \Theta} r'(\theta) \leq K$. Hence by the definition of a derivative $E[m_{n,t}^*(\theta)] - E[m_{n,t}^*(\tilde{\theta})] = J_n(\theta) \times (\theta - \tilde{\theta}) \times (1 + o(1))$, thus $||E[m_{n,t}^*(\theta)] - E[m_{n,t}^*(\tilde{\theta})]|| \leq K||\theta - \tilde{\theta}||$.

For the second term, note that, similar to the proof of Lemma A.3, $I_{n,t}(\theta) = I(|\epsilon_t(\theta)| \leq c_n(\theta))$ can be approximated by a continuous, differentiable, uniformly bounded function $I_{n,t}^{(N)}(\theta) = I^{(N)}(|\epsilon_t(\theta)| \leq c_n(\theta))$ that has a uniformly bounded derivative $(\partial/\partial \theta) I_{n,t}^{(N)}(\theta)$, where $N \in \mathbb{N}$ guides the approximation: $\lim_{N \to \infty} \sup_{\theta \in \Theta} |I_{n,t}^{(N)}(\theta) - I_{n,t}(\theta)| = 0$ a.s. and $\lim_{N \to \infty} \sup_{\theta \in \Theta} ||(\partial/\partial \theta) I_{n,t}^{(N)}(\theta)|| = 0$ a.s.
for each $n$. Specifically, for some double array of positive non-random functions \( \{\epsilon_n(N) \theta, \tilde{\theta}\} : n, N \in \mathbb{N} \} \), where $\sup_{\theta, \tilde{\theta} \in \Theta} |\epsilon_n(N) \theta, \tilde{\theta}| \to 0$ as $N \to \infty$ for each $n$:

$$
\left( E \left[ \left\{ \varphi_{n,t}(\theta) I_{n,t}(\theta) - \varphi_{n,t}(\tilde{\theta}) I_{n,t}(\tilde{\theta}) \right\}^2 \right] \right)^{1/2} = E \left[ \left\{ \varphi_{n,t}(\theta) I_{n,t}^{(N)}(\theta) - \varphi_{n,t}(\tilde{\theta}) I_{n,t}^{(N)}(\tilde{\theta}) \right\}^2 \right]^{1/2} + \epsilon_n(N) \theta, \tilde{\theta}.
$$

By the A.4 transform bounds and $\sup_{\theta \in \Theta} c_n(\theta) \leq K$ we can always set $N$ large enough to ensure $\max_{1 \leq t \leq n} \sup_{\theta \in \Theta} |\varphi_{n,t}(\theta) I_{n,t}^{(N)}(\theta)| \leq K$ a.s. and $\max_{1 \leq t \leq n} \sup_{\theta \in \Theta} |(\partial \varphi_{n,t}(\theta) / \partial \theta) I_{n,t}^{(N)}(\theta)| \leq K$ a.s. while $\lim_{N \to \infty} \max_{1 \leq t \leq n} \sup_{\theta \in \Theta} |(\partial \varphi_{n,t}(\theta) / \partial \theta) I_{n,t}^{(N)}(\theta)| = 0$ a.s. Therefore $(E[|\varphi_{n,t}(\theta) I_{n,t}^{(N)}(\theta) - \varphi_{n,t}(\tilde{\theta}) I_{n,t}^{(N)}(\tilde{\theta})|^2])^{1/2} \leq K||\theta - \tilde{\theta}||$ by the mean-value-theorem. Finally, for any choice $\{\theta, \tilde{\theta}\}$ we can take $N \to \infty$ fast enough as $n \to \infty$ that $\epsilon_n(N) \theta, \tilde{\theta} \leq K||\theta - \tilde{\theta}||$. This completes the proof. 

**Lemma A.7 (mixing CLT for $\{\psi(\epsilon_t, c_n) Z_t, I(\epsilon_t > c_n)\}$)** Let $a \in \mathbb{R}^q$. Define $I_{n,t} = I(\epsilon_t > c_n)$ and $M_{n,t} = \psi(\epsilon_t, c_n) Z_t + a \{I_{n,t} - E[I_{n,t}]\}$. Define $\mathcal{Y} = \lim_{n \to \infty} E(1/n^{1/2} \sum_{t=1}^n \{I_{t} - E[I_t]\})^2$, $\mathcal{S} = E[\psi(\epsilon_t, r Z_t Z_t')$ and $\mathcal{R} = a \mathcal{Y} + \sum_{t=1}^\infty E[\varphi(\epsilon_t, r)\{1 - I_t\}(a Z_t + Z_t a')]$. Then $1/n^{1/2} \sum_{t=1}^n M_{n,t} \to N(0, \mathcal{S} + \mathcal{R})$.

The following corollary delivers a CLT for $m_{n,t}^*$, which allows for several key GEL arguments from Newey and Smith (2004) to carry over.

**Corollary A.8 (CLT for $m_{n,t}^* = \psi(\epsilon_t, c_n) Z_t$)** Under the conditions of Lemma A.7 $1/n^{1/2} \sum_{t=1}^n m_{n,t}^* \to N(0, \mathcal{S})$ where $\mathcal{S} = E[\psi(\epsilon_t, r^2) Z_t Z_t']$.

**Proof of Lemma A.7.** By construction and boundedness $E[M_{n,t} = 0$ and $\limsup_{n \to \infty} E(||M_{n,t} - M_t||)^s < \infty$ for any $s > 0$. Further, $M_t = \lim_{n \to \infty} M_{n,t}$ by $c_n \to r$ and dominated convergence. Then, in view of distribution smoothness, stationarity and geometric $\beta$-mixing under B.2-B.4, arguments in Cizek (2008, p. 1529) apply: $1/n^{1/2} \sum_{t=1}^n \{M_{n,t} - M_t\}$ $\mathcal{D} \to 0$. By the Cramér-Wold theorem it remains to prove $1/n^{1/2} \sum_{t=1}^n \zeta^t M_t \to N(0, \zeta'(S + \mathcal{R}))$ for conformable $\zeta$, $\zeta' \zeta = 1$. Since $\zeta'M_t$ is stationary and geometrically $\beta$-mixing, we assume without loss of generality that $q = 1$ hence $a, Z_t, M_t \in \mathbb{R}$. Note $M_t$ is stationary, geometrically $\beta$-mixing and bounded, hence $1/n^{1/2} \sum_{t=1}^n M_t \to N(0, \mathcal{S})$ where $\mathcal{S} = \lim_{n \to \infty} E(1/n^{1/2} \sum_{t=1}^n M_t Z_t)^2 < \infty$ by Theorem 1.6 in Ibragimov (1962).

The proof is complete if we show $\mathcal{S} = S + \mathcal{R}$. By the martingale difference property of $\psi(\epsilon_t, r) Z_t$:

$$
\lim_{n \to \infty} E \left[ \frac{1}{n^{1/2}} \sum_{t=1}^n M_t \right]^2 = \mathcal{S} + a^2 \mathcal{Y} + 2a \lim_{n \to \infty} \frac{1}{n} \sum_{s,t=1}^n E[\psi(\epsilon_s, r) Z_s \{I_t - E[I_t]\}],
$$

hence we must show $\mathcal{R} = a^2 \mathcal{Y} + 2a \lim_{n \to \infty} 1/n \sum_{s,t=1}^n E[\psi(\epsilon_s, r) Z_s \{I_t - E[I_t]\}]$. Use the martingale difference property and $(1 - I_t) I_t = 0$ to deduce:

$$
E[\psi(\epsilon_t, r) Z_t \{I_t - E[I_t]\}] = E[\varphi(\epsilon_t, r)\{1 - I_t\} Z_t \{I_t - E[I_t]\}],
$$

and

$$
E[\psi(\epsilon_s, r) Z_s \{I_t - E[I_t]\}] = E[\psi(\epsilon_s, r) Z_s \{I_t - E[I_t]\} \{I_t - E[I_t]\}] = 0 \text{ for } s > t
$$

$$
= E[\varphi(\epsilon_s, r)\{1 - I_s\} Z_s I_t] \text{ for } s < t.
$$
Hence, in view of stationarity,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{s,t=1}^{n} E [\psi(\epsilon_s, r) Z_s \times \{I_t - E[I_t]\}] = \lim_{n \to \infty} \frac{1}{n} \sum_{s<t}^{n} E [\varphi(\epsilon_s, r) \{1 - I_s\} Z_s I_t] = \\
\sum_{i=1}^{\infty} E [\varphi(\epsilon_1, r) \{1 - I_1\} Z_i I_{i+1}].
\]

This proves \( R \) is identically \( a^2 \Upsilon + 2a \lim_{n \to \infty} 1/n \sum_{s,t=1}^{n} E [\psi(\epsilon_s, r) Z_s \{I_t - E[I_t]\}] \).

Recall \( J_n = (\partial/\partial \theta) E[m_{n,t}^*] \) and \( J = \lim_{n \to \infty} J_n \).

**Lemma A.9 (Jacobian)** Let \( \{\tilde{\theta}_n\}_{n \geq 1} \) be any stochastic sequence that satisfies \( n^{1/2} (\tilde{\theta}_n - \theta_0) \xrightarrow{P} 0 \).

Let \( \{\epsilon_n\} \) be a sequence of non-random numbers that satisfy \( \epsilon_n \to 0 \) and \( n^{1/2} \epsilon_n \to \infty \), and denote by \( e_j \) a \((p+1) \times 1\) zero vector with \( j \) as the \( j \)th element. Then \( ||J_n - J|| \xrightarrow{P} 0 \) where
\[
J_n = \left[ J_{n,j} \right]_{j=1}^{p+1} \equiv \left[ \frac{1}{2\epsilon_n} \sum_{t=1}^{n} \left( m_{n,t}^* (\tilde{\theta}_n + e_j \epsilon_n) - m_{n,t}^* (\tilde{\theta}_n - e_j \epsilon_n) \right) \right]_{j=1}^{p+1}.
\]

**Proof.** Since \( (\epsilon_n)_{n \geq 1} \to \infty \) and \( n^{1/2} (\tilde{\theta}_n - \theta_0) \xrightarrow{P} 0 \) it follows by the same argument as approximation Lemma A.3 that for each \( j = 1, \ldots, p+1 \):
\[
J_{n,j} = \frac{1}{2\epsilon_n} \sum_{t=1}^{n} \left( m_{n,t}^* (\tilde{\theta}_n + e_j \epsilon_n) - m_{n,t}^* (\tilde{\theta}_n - e_j \epsilon_n) \right) + o_p(1).
\]

Combine UCLT Lemma A.6 with \( \epsilon_n \to 0 \) and \( (\epsilon_n)_{n \geq 1} \to \infty \) to deduce
\[
\frac{1}{2\epsilon_n} \sum_{t=1}^{n} \left( m_{n,t}^* (\tilde{\theta}_n + e_j \epsilon_n) - m_{n,t}^* (\tilde{\theta}_n - e_j \epsilon_n) \right) = \frac{1}{2\epsilon_n} \left( E \left[ m_{n,t}^* (\tilde{\theta}_n + e_j \epsilon_n) \right] - E \left[ m_{n,t}^* (\tilde{\theta}_n - e_j \epsilon_n) \right] \right) + o_p(1).
\]

Thus, by a first order expansion around \( \theta_0 \), for some \( \tilde{\theta}_n^* \) \( ||\tilde{\theta}_n^* - \tilde{\theta}_n|| \leq ||\tilde{\theta}_n - \theta_0|| = O_p(1/n^{1/2})\):
\[
J_{n,j} = \frac{1}{2\epsilon_n} \left( E \left[ m_{n,t}^* (\theta_0 + e_j \epsilon_n) \right] - E \left[ m_{n,t}^* (\theta_0 - e_j \epsilon_n) \right] \right) + \frac{1}{2\epsilon_n} \left( J_n \left( \tilde{\theta}_n^* + e_j \epsilon_n \right) - J_n \left( \tilde{\theta}_n^* - e_j \epsilon_n \right) \right)' (\tilde{\theta}_n - \theta_0) + o_p(1)
\]
\[
= \frac{1}{2\epsilon_n} \left( E \left[ m_{n,t}^* (\theta_0 + e_j \epsilon_n) \right] - E \left[ m_{n,t}^* (\theta_0 - e_j \epsilon_n) \right] \right) + o_p(1).
\]

The second equality follows from \( \epsilon_n^{-1} ||\tilde{\theta}_n - \theta_0|| = o_p(n^{1/2} ||\tilde{\theta}_n - \theta_0||) = o_p(1) \), and \( \sup_{\theta \in \Theta} ||J_n(\theta)|| = O(1) \). Therefore \( ||J_n - J|| \xrightarrow{P} 0 \) follows from (A.5), \( J_n = (\partial/\partial \theta) E \left[ m_{n,t}^* \right] \), and the definition of a derivative. Since \( ||J_n - J|| \xrightarrow{P} 0 \) by the definition of \( J \), the proof is complete.

Since the estimating equations are non-smooth, we exploit what is essentially a stochastic equicontinuity condition. See, e.g., Pakes and Pollard (1989) and Parente and Smith (2011). Note we use the \( l_1 \)-norm \( || \cdot || \).
Lemma A.10 (stochastic equicontinuity) Write $M_n(\theta) \equiv n^{1/2}(m_n^*(\theta) - E[m_n^*(\theta)])$. For any sequence of positive numbers $\{\gamma_n\}$, $\gamma_n \to 0$:

$$\sup_{|\theta - \theta_0| \leq \gamma_n} \left\{ \frac{|M_n(\theta) - M_n(\theta_0)|}{1 + n^{1/2}} |\theta - \theta_0| \right\} \overset{p}{\to} 0.$$

Proof. Let $\zeta \in \mathbb{R}^q$. Use UCLT Lemma A.6 to deduce boundedness $E[\sup_{\theta \in \Theta} \sup_{|\zeta| = 1} |\zeta'(M_n(\theta)) - M_n(\theta_0)||] = O(1)$. But, boundedness, continuity and $\gamma_n \to 0$ imply $E[\sup_{|\theta - \theta_0| \leq \gamma_n} |\zeta'(M_n(\theta) - M_n(\theta_0)||] \overset{p}{\to} 0$ by Markov’s inequality. This suffices to prove the claim.

Finally, we require limit theory for the Lagrange multipliers, equations and profile weights.

Lemma A.11 (GEL argument) We have $\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' m_{n,t}^*(\theta)| \overset{p}{\to} 0$, and $\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' \hat{m}_{n,t}^*(\theta)| \overset{p}{\to} 0$, and $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ w.p.a.1. $\forall \theta \in \Theta$.

Proof. Recall the criterion function $\rho$ has domain $D$ containing zero. Recall $\hat{m}_{n,t}^*(\theta) = \psi(\varepsilon_t, c_n(\theta))Z_t$, $m_{n,t}^*(\theta) = \psi(\varepsilon_t, c_n(\theta))Z_t$, $|\psi(\varepsilon_t, c)| \leq Kc a.s.$ by A.4, $||Z|| \leq K a.s.$ by B.5, $P(|\varepsilon_t(\theta)| \geq c_n(\theta)) = k_n/n \sim \xi$, and $\sup_{\theta \in \Theta} \{c_n(\theta)\} = O(1)$. Therefore $\max_{1 \leq t \leq n} \sup_{\theta \in \Theta} ||m_{n,t}^*(\theta)|| \leq K \sup_{\theta \in \Theta} \{c_n(\theta)\} = O(1)$. Similarly $\max_{1 \leq t \leq n} \sup_{\theta \in \Theta} ||\hat{m}_{n,t}^*(\theta)|| \leq K \sup_{\theta \in \Theta} \{\varepsilon_t(\theta)\} \leq \sup_{\theta \in \Theta} \{c_n(\theta)\} + \sup_{\theta \in \Theta} |\varepsilon_t(\theta) - c_n(\theta)| = o_P(1)$ given uniform order statistic consistency $\sup_{\theta \in \Theta} |\varepsilon_t(\theta) - c_n(\theta)| \overset{p}{\to} 0$ (Cizek, 2008, Assumption A). The first two claims therefore follow by the construction $\Lambda_n = \{\lambda : ||\lambda|| \leq Kn^{-1/2}\}$:

$$\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' m_{n,t}^*(\theta)| = n^{-1/2} \sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' m_{n,t}^*(\theta)| \overset{p}{\to} 0,$$

and similarly $\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' \hat{m}_{n,t}^*(\theta)| \overset{p}{\to} 0$. The third claim $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ w.p.a.1. $\forall \theta \in \Theta$ follows from the second and $0 \in D$ since $\lambda' \hat{m}_{n,t}^*(\theta) \in D$ w.p.a.1 for all $\theta \in \Theta$ and any $\lambda$: $||\lambda|| \leq n^{-1/2}$.

Lemma A.12 (constrained GEL) Consider any sequence of random variables $\{\hat{\theta}_n\}$ that satisfies $\hat{\theta}_n \in \Theta$ and $\hat{\theta}_n \overset{p}{\to} \theta_0$, such that $m_n^*(\hat{\theta}_n) = o_P(1/n^{1/2})$. Then $\hat{\lambda}_n \equiv \arg\max_{\lambda \in \Lambda_n(\hat{\theta}_n)} \{\hat{Q}_n(\hat{\theta}_n, \lambda)\}$ exists w.p.a.1, $\hat{\lambda}_n = O_P(1/n^{1/2}) = o_P(1)$, and $\sup_{\lambda \in \Lambda_n(\hat{\theta}_n)} \{\hat{Q}_n(\hat{\theta}_n, \lambda)\} \leq \rho(0) + O_P(1/n)$.

Proof. Smoothness of $\rho(u)$, and $\sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\lambda' \hat{m}_{n,t}^*(\theta)| \overset{p}{\to} 0$ and $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ w.p.a.1. $\forall \theta \in \Theta$ by Lemma A.11, ensure $\hat{\lambda}_n = \arg\max_{\lambda \in \Lambda_n} \{\hat{Q}_n(\hat{\theta}_n, \lambda)\}$ exists w.p.a.1. Further, $\hat{\lambda}_n = O_P(1/n^{1/2})$ by construction of $\hat{\lambda}_n \in \Lambda_n$. Moreover, by covariance consistency Lemma A.5 and the fact that $S$ is positive definite, the smallest eigenvalue of $1/n \sum_{t=1}^n \hat{m}_{n,t}^*(\hat{\theta}_n)\hat{m}_{n,t}^*(\hat{\theta}_n)'$ is bounded away from zero w.p.a.1. Now apply Newey and Smith (2004, proof of Lemma A.2)’s argument to prove each claim.

Lemma A.13 (equation limits) $m_n^*(\hat{\theta}_n) = O_P(1/n^{1/2})$.

Proof. Lemma A.12 trivially holds for $\hat{\theta}_n = \theta_0$. That, in conjunction with ULLN Lemma A.1, CLT Lemma A.8, and uniform GEL argument Lemma A.11 imply $m_n^*(\hat{\theta}_n) = O_P(1/n^{1/2})$ by the same proof Newey and Smith (2004) use for their Lemma A3.
Lemma A.14 (profile functions) Let \{\tilde{\lambda}_n\} be a sequence of random variables, \tilde{\lambda}_n \in \mathbb{R}^q, and define 
\tilde{\pi}_{n,t}(\theta) = \rho^{(1)}(\tilde{\lambda}_{n,t}(\theta))/\sum_{t=1}^{n} \rho^{(1)}(\tilde{\lambda}_{n,t}(\theta)). 
If \tilde{\lambda}_n = O_p(n^{-1/2}) then \sup_{\theta \in \Theta} \max_{1 \leq t \leq n} |\tilde{\pi}_{n,t}(\theta) - 1/n| = O_p(1/n^{3/2}).

Proof. Expand \rho^{(1)}(\tilde{\lambda}_{n,t}(\theta)) around \lambda = 0: 
\rho^{(1)}(\tilde{\lambda}_{n,t}(\theta)) = -1 + \rho^{(2)}(\lambda_{n,t}(\theta)) \times \tilde{\lambda}_n = -1 + \rho^{(2)}(\lambda_{n,t}(\theta)) \times O_p(1/n^{1/2}) for some \|\lambda_{n,t}\| \leq \|\tilde{\lambda}_n\| = O_p(1/n^{1/2}), where O_p(\cdot) is not a function of t or \theta. Lemma A.11, thrice continuous differentiability of \rho, and \rho^{(2)}(0) = -1 ensure \sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_{1 \leq t \leq n} |\rho^{(2)}(\lambda_{n,t}(\theta))| \overset{p}{\to} 1. Therefore 
\[ \sup_{\theta \in \Theta} \max_{1 \leq t \leq n} \left| \rho^{(1)}(\tilde{\lambda}_{n,t}(\theta)) + 1 \right| \leq \sup_{\theta \in \Theta} \max_{1 \leq t \leq n} \left| \rho^{(2)}(\lambda_{n,t}(\theta)) \right| \times O_p\left(1/n^{1/2}\right) = O_p\left(1/n^{1/2}\right). \]

Now use the construction of \tilde{\pi}_{n,t}(\theta) to deduce as claimed: \sup_{\theta \in \Theta} \max_{1 \leq t \leq n} |\tilde{\pi}_{n,t}(\theta) - 1/n| equals
\[ \sup_{\theta \in \Theta} \max_{1 \leq t \leq n} \left\| \frac{\rho^{(1)}(\tilde{\lambda}_{n,t}(\theta)) - 1/n}{\sum_{t=1}^{n} \rho^{(1)}(\tilde{\lambda}_{n,t}(\theta))} \right\| \leq O_p\left(1/n^{3/2}\right). \]

A.2 Proofs of Main Theorems

Proof of Theorem 2.1.

Step 1: We apply a well known argument to prove \( \hat{\theta}_n = \theta_0 + O_p(1/n^{1/2}) \), cf. Pakes and Pollard (1989, Theorem 3.1). Write \( \inf_{\theta \in \Theta, n^{-1/2}|\theta - \theta_0| > \gamma} \) as \( \inf_{n^{-1/2}|\theta - \theta_0| > \gamma} \).

By the definition of a derivative and \( E[m_{n,t}(\theta_0)] = 0 \) in view of the martingale difference property (7), we have \( E[m_{n,t}(\theta)] = \mathcal{M}_t(\theta - \theta_0) \times (1 + o(1)) \) where \( o(1) \) is not a function of \( \theta \). Further, \( \|\mathcal{M}_n\| > 0 \) for all \( n \geq N \) and some large \( N > 1 \) by B.5 and dominated convergence. Hence for every \( n \geq N \) and \( \gamma > 0 \)
\[ \inf_{n^{-1/2}|\theta - \theta_0| > \gamma} \left\{ n^{1/2} \left\| E[m_{n,t}(\theta)] \right\| \right\} \geq K \inf_{n^{-1/2}|\theta - \theta_0| > \gamma} \left\{ \left\| \mathcal{M}_n \times n^{1/2}(\theta - \theta_0) \right\| \right\} \times (1 + o(1)) > 0. \]

Therefore \( \epsilon(\gamma) \equiv \inf_{n \geq N} \inf_{n^{-1/2}|\theta - \theta_0| > \gamma} \left\{ n^{1/2} \left| E[m_{n,t}(\theta)] \right| \right\} > 0 \) for any \( \gamma > 0 \).

By the construction of \( \epsilon(\gamma) > 0 \) it follows
\[ P\left( n^{1/2} \left| \hat{\theta}_n - \theta_0 \right| > \gamma \right) \leq P\left( n^{1/2} \left| E[m_{n,t}(\hat{\theta}_n)] \right| > \epsilon(\gamma) \right). \] (A.6)

Moreover, by Minkowski’s inequality \( n^{1/2} \left| E[m_{n,t}(\theta)] \right| \leq \sup_{\theta \in \Theta} n^{1/2} \left| m_{n,t}(\theta) \right| + n^{1/2} \left| m_{n,t}^*(\theta) \right| \) which is \( O_p(1) \) by applications of UCLT Lemma A.6 and equation limit Lemma A.13.

Now, by definition of \( O_p(1) \) it follows \( n^{1/2} \left| E[m_{n,t}(\hat{\theta}_n)] \right| = O_p(1) \) if and only if \( \lim_{n \to \infty} P(n^{1/2} \left| E[m_{n,t}(\hat{\theta}_n)] \right| > \epsilon) \in [0, 1) \) for any \( \epsilon > 0 \) and the latter implies \( \lim_{n \to \infty} P(n^{1/2} \left| E[m_{n,t}(\hat{\theta}_n)] \right| > \epsilon(\gamma)) \in [0, 1) \) for any \( \gamma > 0 \). But in view of (A.6) this implies \( \lim_{n \to \infty} P(n^{1/2} \left| \hat{\theta}_n - \theta_0 \right| > \gamma) \in [0, 1) \) for any \( \gamma > 0 \). Finally, \( \lim_{n \to \infty} P(n^{1/2} \left| \hat{\theta}_n - \theta_0 \right| > \gamma) \in [0, 1) \) for any \( \gamma > 0 \) if and only if \( n^{1/2} \left| \hat{\theta}_n - \theta_0 \right| = O_p(1) \).

Step 2: Now consider \( \hat{\lambda}_n \). The conditions of Lemma A.12 are satisfied for \( \hat{\theta}_n = \hat{\theta}_n \) given Step 1

\(^5\)Clearly \( \epsilon(\gamma) \) depends on \( N \), but that has no bearing on the proof since \( N \) is simply chosen to be large.
consistency $\hat{\theta}_n \xrightarrow{p} \theta_0$, and the Lemma A.13 equation limit $m_n^*(\hat{\theta}_n) = O_p(1/n^{1/2})$. Hence $\lambda_n$ exists and $\hat{\lambda}_n = O_p(1/n^{1/2})$ by Lemma A.12.

**Proof of Theorem 2.2.** Drop the trimming parameter argument $\xi$. Recall $J \equiv \lim_{n \to \infty} (\partial / \partial \theta) E[m_{n,t}]$, and by the martingale difference property and dominated convergence $S \equiv \lim_{n \to \infty} E[m_{n,m_{n,t}}] = E[\psi(\epsilon_t, r)^2 Z_t Z_t'] < \infty$ is the long-run asymptotic variance of $1/n^{1/2} \sum_{t=1}^n m_{n,t}$. Hence the standard GEL asymptotic variance components are (cf. Newey and Smith, 2004)

$$V \equiv (J'S^{-1}J)^{-1}, \quad A \equiv \begin{bmatrix} \mathcal{V} & 0 \\ 0 & P \end{bmatrix}, \text{ where } P \equiv S^{-1} - S^{-1} J (J'S^{-1}J)^{-1} J'S^{-1}.$$

Next, define the GEL criterion $\hat{Q}_n(\theta, \lambda) \equiv 1/n \sum_{t=1}^n \rho(\lambda' \hat{m}_{n,t}^*(\theta))$, and a quadratic criterion

$$\hat{L}_n(\theta, \lambda) \equiv -\lambda' J (\theta - \theta_0) - \lambda' \hat{m}_n^* - (1/2) \lambda'S\lambda$$

with solution $\hat{\beta}_n \equiv [\hat{\theta}_n, \hat{\lambda}_n]$ where $\hat{\theta}_n = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^q} \hat{L}_n(\theta, \lambda)$ and $\hat{\lambda}_n = \arg \sup_{\lambda \in \mathbb{R}^q} \hat{L}_n(\theta_n, \lambda)$.

Step 1 proves $n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{p} N(0, A + \mathcal{H}'\mathcal{H}')$. In Step 2 we show

$$\left| \hat{Q}_n(\hat{\theta}_n, \hat{\lambda}_n) - \hat{\lambda}_n(\hat{\theta}_n, \hat{\lambda}_n) \right| = o_n(1/n).$$

We then use (A.7) in Step 3 to prove $n^{1/2}(\tilde{\beta}_n - \hat{\beta}_n) \xrightarrow{p} 0$. Our proof follows arguments in Parente and Smith (2011, p. 102-104), cf. Pakes and Pollard (1989, Theorem 3.1).

**Step 1.** Define

$$M \equiv \begin{bmatrix} 0 & J \\ J' & S \end{bmatrix}, \text{ hence } M^{-1} = \begin{bmatrix} -J'(S^{-1}J)^{-1} & \mathcal{H} \\ \mathcal{H}' & P \end{bmatrix}, \mathcal{H} \equiv (J'S^{-1}J)^{-1} J'S^{-1} \text{ and } \hat{\mathcal{H}} \equiv \begin{bmatrix} \mathcal{H} \\ P \end{bmatrix}.$$  

Since $\hat{L}_n(\theta, \lambda)$ is linear in $\theta$ and quadratic in $\lambda$, and $\Theta$ is compact, the first order conditions for an interior global optimum are $J(\hat{\theta}_n - \theta_0) + \hat{m}_{n,t}^* = 0$ and $J'\hat{\lambda}_n = 0$. Stack this and solve for $n^{1/2}(\hat{\beta}_n - \beta^0)$ to obtain, cf. Newey and Smith (2004, p. 240),

$$n^{1/2}(\hat{\beta}_n - \beta^0) = -M^{-1} \begin{bmatrix} 0 \\ 1/n^{1/2} \sum_{t=1}^n \hat{m}_{n,t}^* \end{bmatrix} = \mathcal{H} \frac{1}{n^{1/2}} \sum_{t=1}^n \hat{m}_{n,t}^*.$$  

(A.8)

Next we require an asymptotic approximation for $1/n^{1/2} \sum_{t=1}^n \hat{m}_{n,t}^*$. Define

$$\delta(\epsilon, c) \equiv \frac{\partial}{\partial \epsilon} \varphi(\epsilon, c), \quad f(r) = \frac{\partial}{\partial r} P(\epsilon_t \leq r), \text{ and } g(r) = f(r) + f(-r),$$

and construct

$$B \equiv E[\delta(\epsilon_t, r) I(\epsilon_t \leq r) Z_t] \times \frac{\xi}{g(r)} \in \mathbb{R}^q.$$  

Also define random variables

$$M_{n,t} \equiv m_{n,t}^* + B \times \{I(\epsilon_t > c_n) - E[I(\epsilon_t > c_n)] \}.$$  

24
In the simple trimming case $\varphi(\epsilon, c) = \epsilon$ hence $\delta(\epsilon, r) = 0$, hence $M_{n,t} \equiv m_{n,t}^*$. Finally, define

$$Y \equiv \lim_{n \to \infty} E \left( \frac{1}{n^{1/2}} \sum_{t=1}^{n} \{ I(|\epsilon_t| > r) - E[I(|\epsilon_t| > r)] \} \right)^2,$$

and

$$R \equiv \mathcal{B}Y\mathcal{B} + \sum_{i=1}^{\infty} E \left[ \varphi(\epsilon_1, r) I(|\epsilon_1| \leq r) \left\{ \mathcal{B}Z_1 + Z_1\mathcal{B}' \right\} \times I(|\epsilon_{i+1}| > r) \right] < \infty.$$ 

By Corollary A.4 we have:

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} m_{n,t}^* = \frac{1}{n^{1/2}} \sum_{t=1}^{n} M_{n,t} \times (1 + o_p(1)). \quad (A.9)$$

Now combine expansions (A.8) and (A.9) and invoke CLT Lemma A.7 to deduce $n^{1/2}(\tilde{\beta}_n - \beta^0) \xrightarrow{d} N(0, \tilde{\mathcal{H}(S + R)}\mathcal{H})$. It is easily verified that $\tilde{\mathcal{H}}S\tilde{\mathcal{H}}'$ is identically $\mathcal{A}$, hence $n^{1/2}(\tilde{\beta}_n - \beta^0) \xrightarrow{d} N(0, \mathcal{A} + \mathcal{H}R\mathcal{H}')$ as claimed.

**Step 2.** We use both $l_1$ and spectral norms $| \cdot |$ and $\| \cdot \|$. By a second order Taylor expansion around $\lambda = 0$, and since $\rho(1)(0) = -1$, we have for some $\lambda_n^*, ||\lambda_n^*|| \leq ||\tilde{\lambda}_n||$:

$$\tilde{Q}_n(\tilde{\theta}_n, \tilde{\lambda}_n) = -\tilde{\lambda}_n' \hat{m}_n^*(\tilde{\theta}_n) + \frac{1}{2} \tilde{\lambda}_n' \left( \frac{1}{n} \sum_{t=1}^{n} \rho(2) \left( \lambda_n^* \hat{m}_{n,t}^*(\tilde{\theta}_n) \right) \times \hat{m}_{n,t}^*(\tilde{\theta}_n) \right) \tilde{\lambda}_n.$$ 

GEL argument Lemma A.11, $\rho(2)(0) = -1$ and continuity of $\rho(2)$ imply $\sup_{\theta \in \Theta} \lambda_n^* \max_{1 \leq t \leq n} |\rho(2)(\lambda_n^* \hat{m}_{n,t}^*(\theta))| + 1 \xrightarrow{p} 0$. Further, transform bounds A.4, uniform order statistic consistency (Cizek, 2008, Appendix A), and $\sup_{\theta \in \Theta} c_n(\theta) = O(1)$ imply $\max_{1 \leq t \leq n} \sup_{\theta \in \Theta} ||\hat{m}_{n,t}^*(\theta)|| \leq K \sup_{\theta \in \Theta} c_n(\theta) \times (1 + o_p(1)) = o_p(1)$. Hence $1/n \sum_{t=1}^{n} \rho(2)(\lambda_n^* \hat{m}_{n,t}^*(\tilde{\theta}_n)) \tilde{m}_{n,t}^*(\tilde{\theta}_n) \tilde{\lambda}_n' \tilde{\lambda}_n \xrightarrow{P} S$ by invoking Lemma A.5 with $\tilde{\theta}_n \xrightarrow{P} \theta_0$ from Step 1. Further $\tilde{\lambda}_n = O_p(1/n^{1/2})$ by Step 1. Therefore $\tilde{Q}_n(\tilde{\theta}_n, \tilde{\lambda}_n) = -\tilde{\lambda}_n' \hat{m}_n^*(\tilde{\theta}_n) - (1/2) \tilde{\lambda}_n' S \tilde{\lambda}_n + o_p(1/n)$. Hence, by the definition of $L_n(\tilde{\theta}_n, \tilde{\lambda}_n)$:

$$\left| \tilde{Q}_n(\tilde{\theta}_n, \tilde{\lambda}_n) - \tilde{L}_n(\tilde{\theta}_n, \tilde{\lambda}_n) \right| \leq \left| - \left( \tilde{\lambda}_n' \left( \hat{m}_n^*(\tilde{\theta}_n) - \tilde{m}_n^* \right) - \tilde{\lambda}_n' \mathcal{J} (\tilde{\theta}_n - \theta_0) \right) \right| + o_p(1/n)$$

$$\leq \left| - \tilde{\lambda}_n' \left( \hat{m}_n^*(\tilde{\theta}_n) - \tilde{m}_n^* - E \left[ m_{n,t}^*(\tilde{\theta}_n) \right] \right) \right| + \left| \tilde{\lambda}_n' \mathcal{J} (\tilde{\theta}_n - \theta_0) - E \left[ m_{n,t}^*(\tilde{\theta}_n) \right] \right| \leq \tilde{\lambda}_n' S \tilde{\lambda}_n + o_p(1/n).$$

We need to show $\mathfrak{A}_n$ and $\mathfrak{B}_n$ are $o_p(1/n)$ to prove the desired result (A.7).

Consider $\mathfrak{A}_n$. By Step 1 $|\tilde{\lambda}_n| = O(1/n^{1/2})$. Hence, by approximation Lemma A.3, boundedness $\sup_{\theta \in \Theta} |E[\delta(\epsilon_t(\theta), r(\theta)) I(|\epsilon_t(\theta)| \leq r(\theta)) Z_t]| \leq K \sup_{\theta \in \Theta} r(\theta) \leq K$ by A.4 and $||Z_t|| \leq K$ a.s. by B.5, and Minkowski’s inequality:

$$\mathfrak{A}_n = O_p \left( n^{-1/2} \left| m_{n,t}^*(\tilde{\theta}_n) - m_{n,t}^* - E \left[ m_{n,t}^*(\tilde{\theta}_n) \right] \right| \right)$$

$$+ O_p \left( n^{-1/2} \left| \epsilon_{(k_n)}(\tilde{\theta}_n) - c_n(\tilde{\theta}_n) \right| - \left( \epsilon_{(k_n)} - c_n \right) \right) + o_p(1/n^{1/2}).$$

25
Use stochastic equicontinuity Lemma A.10, and $|\tilde{\lambda}_n| = O(1/n^{1/2})$ and $|\tilde{\theta}_n - \theta_0| = O(1/n^{1/2})$, to deduce
\[
|m_n^*(\tilde{\theta}_n) - m_n^* - E[m_n^*,(\tilde{\theta}_n)]| = |\tilde{\lambda}_n| \cdot o_p\left(n^{-1/2} \left\{ 1 + n^{1/2} |\tilde{\theta}_n - \theta_0| \right\}\right) = o_p(1/n).
\] (A.11)

Further, $\{n^{1/2}(e^{(a)}_{(\theta)}) - c_n(\theta) : \theta \in \Theta\}$ converges weakly to a Gaussian process with almost surely uniformly continuous sample paths, and $n^{1/2}E[\sup_{\theta \in \Theta} |e^{(a)}_{(\theta)}| - c_n(\theta)|] = O(1)$: see Cizek (2008, Appendix A). Therefore $n^{1/2}E[\sup_{|\theta - \theta_0| \leq \gamma_n} |e^{(a)}_{(\theta)}(\theta) - c_n(\theta)| - (e^{(a)}_{(\theta)} - c_n)|] \to 0$ for any sequence of positive numbers $\{\gamma_n\}$ with $\gamma_n \to 0$. Since $|\tilde{\theta}_n - \theta_0| = O_p(1/n^{1/2})$ it then follows
\[
\left|\left(e^{(a)}_{(\theta)}(\tilde{\theta}_n) - c_n(\tilde{\theta}_n)\right) - \left(e^{(a)}_{(\theta)} - c_n\right)\right| = o_p(1/n^{1/2}).
\] (A.12)

Combine (A.11) and (A.12) to deduce $\mathfrak{A}_n = o_p(1/n^{1/2})$.

Now consider $\mathfrak{B}_n$ in (A.10). Observe $E[m_{n,t}^*(\tilde{\theta}_n)] = J(\tilde{\theta}_n - \theta_0) + o_p(||\tilde{\theta}_n - \theta_0||) = J(\tilde{\theta}_n - \theta_0) + o_p(1/n^{1/2})$ by the definition of a derivative given ($\partial/\partial \theta$) $E[m_{n,t}^*] = J$, $E[m_{n,t}^*] = 0$ and $||\tilde{\theta}_n - \theta_0|| = o_p(1/n^{1/2})$. Hence $\mathfrak{B}_n \leq ||\tilde{\lambda}_n|| \times ||J(\tilde{\theta}_n - \theta_0) - E[m_{n,t}^*(\tilde{\theta}_n)]|| = o_p(1/n).$

**Step 3.** By the constructions of $\tilde{\theta}_n$ and $\hat{\theta}_n$ it follows $\hat{Q}_n(\tilde{\theta}_n, \lambda_n) \leq \hat{Q}_n(\tilde{\theta}_n, \hat{\lambda}_n)$ and $\hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) \leq \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n).$ Now use (A.7) twice to deduce:
\[
\hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) - o_p(1/n) \leq \hat{Q}_n(\hat{\theta}_n, \hat{\lambda}_n) \leq \hat{Q}_n(\hat{\theta}_n, \hat{\lambda}_n) \leq \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) + o_p(1/n) \leq \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) + o_p(1/n).
\]
Therefore $\hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) - \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) = o_p(1/n)$, hence $\hat{\lambda}_n J(\tilde{\theta}_n - \hat{\theta}_n) = o_p(1/n)$ in view of the quadratic form of $\hat{L}_n(\theta, \lambda)$. Now use $\hat{\lambda}_n = O_p(1/n^{1/2})$ by Theorem 2.1, and the fact that $J$ has full row rank, to deduce $\tilde{\theta}_n - \hat{\theta}_n = o_p(1/n^{1/2})$.

Lastly, by a similar use of (A.7) it can be verified that $\hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) - \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) = o_p(1/n)$, while by the first order conditions for $[\hat{\theta}_n, \hat{\lambda}_n]$ it follows $\lambda_n J = 0$ and $m_n^* + J(\tilde{\theta}_n - \theta_0) = -S\hat{\lambda}_n$. Hence:
\[
o_p(1/n) = \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n) - \hat{L}_n(\hat{\theta}_n, \hat{\lambda}_n)
= -\hat{\lambda}_n J(\tilde{\theta}_n - \theta_0) - \hat{\lambda}_n \hat{m}_n^* - \frac{1}{2} \hat{\lambda}_n ^s \hat{S} \hat{\lambda}_n + \hat{\lambda}_n ^l \hat{J}(\tilde{\theta}_n - \theta_0) + \hat{\lambda}_n ^l \hat{m}_n^* + \frac{1}{2} \hat{\lambda}_n ^l ^s \hat{S} \hat{\lambda}_n
= \left(\hat{\lambda}_n - \hat{\lambda}_n\right)^l \hat{m}_n^* + \left(\hat{\lambda}_n - \hat{\lambda}_n\right)^l J(\tilde{\theta}_n - \theta_0) + \frac{1}{2} \hat{\lambda}_n ^l ^s \hat{S} \hat{\lambda}_n - \frac{1}{2} \hat{\lambda}_n ^l ^s \hat{S} \hat{\lambda}_n
= - \left(\hat{\lambda}_n - \hat{\lambda}_n\right)^l \hat{S} \hat{\lambda}_n - \frac{1}{2} \hat{\lambda}_n ^l ^s \hat{S} \hat{\lambda}_n = \frac{1}{2} \left(\hat{\lambda}_n - \hat{\lambda}_n\right)^l \hat{S} \left(\lambda_n - \lambda_n\right).
\]

By positive definiteness of $S$ this implies $\hat{\lambda}_n - \hat{\lambda}_n = o_p(1/n^{1/2})$, which completes the proof.

**Proof of Theorem 3.1.** The claim follows from GEL consistency Theorem 2.1, covariance and Jacobian consistency Lemmas A.5 and A.9, and profile weight consistency Lemma A.14.

**Proof of Theorem 3.2.** By Corollary A.4
\[
\frac{1}{n^{1/2}} \sum_{t=1}^n \hat{m}_{n,t}^* = \frac{1}{n^{1/2}} \sum_{t=1}^n m_{n,t}^* + B(\xi) \frac{1}{n^{1/2}} \sum_{t=1}^n \{I(|\xi_t| > c_n) - E[I(|\xi_t| > c_n)]\} + o_p(1),
\]
where $B$ is defined in (8). Thus, $1/n^{1/2} \sum_{t=1}^n \hat{m}_{n,t}^*$ satisfies a mixing central limit theorem by application of Lemma A.7. Furthermore, $1/n \sum_{t=1}^n \hat{m}_{n,t}^* \hat{m}_{n,t}^l$ has a well defined, positive definite probability
limit by Lemma A.5. Finally, \( \max_{1 \leq t \leq n} |\hat{m}_{i,n,t}^*| = o_p(n^{1/2}) \) by boundedness of each \( \hat{m}_{i,n,t}^* \). It is then easy to show Wilks’ theorem applies, hence as claimed the log-likelihood is asymptotically chi-squared distributed by exploiting arguments in Owen (2001, Chapter 11). See also Theorem 2 in Kitamura (1997), Theorem 3.2 in Newey and Smith (2004), and Sections 3 and 4 in Parente and Smith (2011).

References


Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Economet-
rica 50, 1029–1054.


Hill, J. B., 2014. Supplemental material for “robust generalized empirical likelihood for heavy tailed autoregressions with conditionally heteroscedastic errors”.


Robust Empirical Likelihood for AR(1) : Simple Trimming

<table>
<thead>
<tr>
<th>$\xi = .00$</th>
<th>$\xi = .01$</th>
<th>$\xi = .05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_n$</td>
<td>$\xi_n$</td>
<td>$\xi_n$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>$n = 500$</td>
<td>$n = 500$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>AR</th>
<th>AR-GARCH</th>
<th>AR-IG</th>
<th>AR</th>
<th>AR-GARCH</th>
<th>AR-IG</th>
<th>AR</th>
<th>AR-GARCH</th>
<th>AR-IG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>$P_{1.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
</tr>
<tr>
<td>Median</td>
<td>$P_{1.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
</tr>
<tr>
<td>RMSE</td>
<td>$P_{1.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
</tr>
<tr>
<td>95% CI</td>
<td>$P_{1.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
</tr>
<tr>
<td>Cover</td>
<td>$P_{1.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
<td>$P_{2.5}$</td>
</tr>
</tbody>
</table>

Table 1: The trimming fractile is $k_n = [\xi n]$. The case $\xi = .00$ represents no trimming. $P_\kappa$ denotes a Pareto distribution with index $\kappa$. AR is the model with an i.i.d. error; AR-GARCH has a GARCH error; AR-IG is AR-IGARCH. "Cover" is the coverage probability of the 95% confidence region.
Robust Empirical Likelihood for AR(1) with i.i.d. Error: Simple Trimming with Small Weight

Simple Trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$, Instruments $z_t = [y_{t-1}, y_{t-2}]'$, Weight $W_t = 1/\prod_{i=1}^n (1 + y_{t-i}^2)$.

<table>
<thead>
<tr>
<th>Error</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.3322</td>
<td>-0.0732</td>
<td>-0.0117</td>
<td>-0.2324</td>
<td>-0.0469</td>
<td>-0.0165</td>
<td>-0.1836</td>
<td>-0.0403</td>
<td>-0.0183</td>
</tr>
<tr>
<td>Median</td>
<td>0.9178</td>
<td>0.9179</td>
<td>0.9029</td>
<td>0.9543</td>
<td>0.9241</td>
<td>0.8929</td>
<td>0.9223</td>
<td>0.9272</td>
<td>0.8929</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.7285</td>
<td>0.2661</td>
<td>0.0985</td>
<td>0.5932</td>
<td>0.2071</td>
<td>0.0967</td>
<td>0.5134</td>
<td>0.2036</td>
<td>0.1043</td>
</tr>
</tbody>
</table>

| 95% CI | 0.358,928 | 0.487,974 | 0.690,980 | 0.406,945 | 0.517,980 | 0.695,980 | 0.453,954 | 0.540,980 | 0.682,983 |
| Cover  | 0.744     | 0.8973    | 0.9418    | 0.7563    | 0.9294    | 0.9448    | 0.8448    | 0.9376    | 0.9576    |

<table>
<thead>
<tr>
<th>$\xi = .00$</th>
<th>$\xi = .01$</th>
<th>$\xi = .05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>$n = 500$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>$n = 500$</td>
<td>$n = 500$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.3186</td>
<td>-0.0269</td>
<td>-0.0013</td>
<td>-0.1913</td>
<td>-0.0115</td>
<td>-0.0033</td>
<td>-0.1388</td>
<td>-0.0044</td>
<td>-0.0043</td>
</tr>
<tr>
<td>Median</td>
<td>0.8849</td>
<td>0.9125</td>
<td>0.8997</td>
<td>0.8718</td>
<td>0.9075</td>
<td>0.8970</td>
<td>0.8618</td>
<td>0.9075</td>
<td>0.8978</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.8095</td>
<td>0.1614</td>
<td>0.0510</td>
<td>0.624</td>
<td>0.1262</td>
<td>0.0427</td>
<td>0.4798</td>
<td>0.0751</td>
<td>0.0466</td>
</tr>
</tbody>
</table>

| 95% CI | 0.197,930 | 0.612,975 | 0.807,970 | 0.268,958 | 0.685,980 | 0.818,966 | 0.318,971 | 0.744,982 | 0.810,971 |
| Cover  | 0.8533    | 0.9228    | 0.9367    | 0.9145    | 0.9429    | 0.9538    | 0.9524    | 0.9404    | 0.9493    |

<table>
<thead>
<tr>
<th>$\xi = .10$</th>
<th>$\xi = .20$</th>
<th>$\xi = .30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>$n = 100$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>$n = 500$</td>
<td>$n = 500$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.1787</td>
<td>-0.0328</td>
<td>-0.0425</td>
<td>-0.1731</td>
<td>-0.0539</td>
<td>-0.0068</td>
<td>-0.1829</td>
<td>-0.0232</td>
<td>-0.0970</td>
</tr>
<tr>
<td>Median</td>
<td>0.9052</td>
<td>0.9345</td>
<td>0.8721</td>
<td>0.8862</td>
<td>0.9724</td>
<td>1.000</td>
<td>0.8899</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.4906</td>
<td>0.1902</td>
<td>0.1407</td>
<td>0.4681</td>
<td>0.2231</td>
<td>0.2093</td>
<td>0.4870</td>
<td>0.2364</td>
<td>0.1190</td>
</tr>
</tbody>
</table>

| 95% CI | 0.466,957 | 0.529,982 | 0.619,983 | 0.479,958 | 0.428,980 | 0.371,987 | 0.494,958 | 0.243,971 | 0.066,987 |
| Cover  | 0.8665    | 0.9412    | 0.9684    | 0.8946    | 0.9452    | 0.9928    | 0.8965    | 0.9267    | 0.9981    |

<table>
<thead>
<tr>
<th>$\xi = .10$</th>
<th>$\xi = .20$</th>
<th>$\xi = .30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>$n = 100$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>$n = 500$</td>
<td>$n = 500$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
<th>$P_{.75}$</th>
<th>$P_{1.5}$</th>
<th>$P_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.1729</td>
<td>0.0066</td>
<td>-0.0107</td>
<td>-0.1469</td>
<td>-0.0143</td>
<td>0.0575</td>
<td>-0.0880</td>
<td>0.0293</td>
<td>0.0999</td>
</tr>
<tr>
<td>Median</td>
<td>0.8520</td>
<td>0.9031</td>
<td>0.8855</td>
<td>0.8557</td>
<td>0.7736</td>
<td>1.000</td>
<td>0.8563</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.4417</td>
<td>0.0788</td>
<td>0.0698</td>
<td>0.3841</td>
<td>0.1116</td>
<td>0.1392</td>
<td>0.2585</td>
<td>0.1291</td>
<td>0.0999</td>
</tr>
</tbody>
</table>

| 95% CI | 0.355,976 | 0.747,983 | 0.771,979 | 0.396,979 | 0.681,986 | 0.509,989 | 0.452,982 | 0.427,985 | 0.053,990 |
| Cover  | 0.9738    | 0.9465    | 0.9572    | 0.9783    | 0.9860    | 0.9968    | 0.9837    | 0.9679    | 1.000    |

Table 2: The trimming fractile is $k_n = [\xi n]$. The case $\xi = .00$ represents no trimming. $P_\kappa$ denotes a Pareto distribution with index $\kappa$. "Cover" is the coverage probability of the 95% confidence region.
Figure 1: Trimmed EL 95% confidence regions and median $\hat{\theta}_n$ (thick black line), with trimming fractile $k_n(\xi)$ over $\xi \in \{.00, .01, \ldots, .40\}$. The sample size is $n = 100$. Simple trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$, instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1/\prod_{i=1}^2 (1 + y_{t-i}^2)^{1/2}$ are used. The model is $y_t = .9y_{t-1} + \epsilon_t$, where $\epsilon_t$ is i.i.d. $P_{1.5}$ or $P_{2.5}$, or GARCH with an i.i.d. $P_{2.5}$ or $P_{4.5}$ error, or IGARCH with an i.i.d. Normal error.
Figure 2: Trimmed EL bias, median, root mean-squared-error, and and 95% coverage probability, with trimming fractile $k_n(\xi)$ over $\xi \in \{.00, .01, ..., .40\}$. The sample size is $n = 100$. Simple trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$, instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1/\prod_{i=1}^{2}(1 + y_{t-i}^2)^{1/2}$ are used. The model is $y_t = .9y_{t-1} + \epsilon_t$, where $\epsilon_t$ is i.i.d. $P_{1.5}$ or $P_{2.5}$, or GARCH with an i.i.d. $P_{2.5}$ or $P_{4.5}$ error, or IGARCH with an i.i.d. Normal error.
Figure 3: Trimmed EL 95% confidence regions and median $\hat{\theta}_n$ (thick black line), with trimming fractile $k_n(\xi)$ over $\xi \in \{.00,.01,\ldots,.40\}$. The sample size is $n = 100$. Simple trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$, instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1 / \prod_{i=1}^{2}(1 + y_{t-i}^2)$ are used. The model is $y_t = .9y_{t-1} + \epsilon_t$, where $\epsilon_t$ is i.i.d. $P_{.75}$, $P_{1.5}$ or $P_{2.5}$.
Figure 4: Trimmed EL bias, median, root mean-squared-error, and 95% coverage probability, with trimming fractile $k_n(\xi)$ over $\xi \in \{.00, .01, ..., .40\}$. The sample size is $n = 100$. Simple trimming $\psi(\epsilon, c) = \epsilon I(|\epsilon| \leq c)$, instruments $z_t = [y_{t-1}, y_{t-2}]'$, and weight $W_t = 1 / \prod_{i=1}^{2}(1 + y_{t-i}^2)$ are used. The model is $y_t = .9y_{t-1} + \epsilon_t$, where $\epsilon_t$ is i.i.d. $P_{.75}$, $P_{1.5}$ or $P_{2.5}$. 