# ECO 519. Handout on Huber (1967), Lemma 3 (Asymptotic Normality)

## 1 Setup

In econometrics, we are familiar with extremum estimators which satisfy a condition of the type:

$$\sum_{i=1}^{n} \psi(x_i, \widehat{\theta}) = 0. \tag{1}$$

Typically, (1) represents a vector of first-order conditions. If $\psi(\cdot)$ is smooth and differentiable, we can establish $\sqrt{n}$-consistency and asymptotic normality of $\widehat{\theta}$ by performing a Taylor approximation of (1) around $\theta_0$ (the true parameter vector which is assumed to satisfy $E[\psi(x_i, \theta_0)] = 0$. Huber's paper deals with a more general problem: One in which the estimator in question satisfies (1) only asymptotically in probability, and $\psi(\cdot)$ is not required to be smooth (not even continuous (!)).

We will try to stick to the paper's original notation, we will let $\lambda(\theta) = E[\psi(x_i, \theta)]$, where both $\theta \in \mathbb{R}^m$ and $\psi \in \mathbb{R}^m$. Suppose we have an estimator (statistic) $T_n$ that satisfies $T_n = \theta_0 + o_p(1)$ and[1]

$$\sqrt{n}\lambda(T_n) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(x_i, \theta_0) = o_p(1). \tag{2}$$

This would be great news, because if $\lambda(\cdot)$ is $\mathcal{C}^1$, we can investigate the asymptotic properties of $T_n$ by performing a Taylor approximation on $\lambda(\cdot)$ instead of $\psi(\cdot)$. This would allow us to deal with non-smooth functions $\psi(\cdot)$, and only require us to assume that its expected value satisfies the smoothness conditions needed for Taylor approximations. Let $\Lambda(\theta) = \nabla_\theta \lambda(\theta)$. Then, under conditions that guarantee $\Lambda(T_n) \xrightarrow{p} \Lambda(\theta_0)$ (for example, boundedness of $\lambda(\cdot)$ in addition to continuity given that $T_n$ converges in probability to $\theta_0$), and CLT conditions that guarantee

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(x_i, \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

[1] Alternatively, we could have $\sqrt{n}\lambda(T_n) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(x_i, \theta_0) = o_p(1)$. The sign of $\psi(\cdot)$ is irrelevant here because it plays the role of a Jacobian vector, with expectation zero evaluated at $\theta_0$.

we could establish $\sqrt{N}$-asymptotic normality of $T_n$ via a Taylor approximation of $\lambda(\cdot)$ in (1):

$$\underbrace{\sqrt{n}\lambda(\theta_0)}_{\text{zero}} + \Lambda(\widetilde{\theta})\sqrt{n}(T_n - \theta_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) = o_p(1)$$

where as usual $\widetilde{\theta}$ is between $T_n$ and $\theta_0$. If $\Lambda(\theta_0) \equiv \Lambda$ is invertible, we get then

$$\sqrt{n}(T_n - \theta_0) = -\Lambda(\widetilde{\theta})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \Lambda^{-1}V\Lambda^{-1'}) \tag{3}$$

Therefore, we should look for conditions that guarantee (2). Consider the following claim:

**Claim 1** Suppose $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) = O_p(1)$ (which would be satisfied if the conditions for a CLT are met). Then:

$$\text{If} \quad \frac{\sum_{i=1}^{n}\left[\psi(x_i, \theta_0) + \lambda(\tau)\right]}{\sqrt{n} + n\|\lambda(\tau)\|} = o_p(1), \quad \text{then} \quad \sqrt{n}\lambda(\tau) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) = o_p(1).$$

**Proof:** First, we prove that if the condition of the claim is satisfied, then $\sqrt{n}\|\lambda(\tau)\| = O_p(1)$. Take any $1 > \epsilon > 0$, then

$$\underbrace{\Pr\left[\left\|\frac{\sum_{i=1}^{n}\left[\psi(x_i, \theta_0) + \lambda(\tau)\right]}{\sqrt{n} + n\|\lambda(\tau)\|}\right\| > \epsilon\right]}_{\longrightarrow 0\ \forall\ \epsilon>0 \text{ by assumption.}} = \Pr\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) + \sqrt{n}\lambda(\tau)\right\| > \epsilon + \epsilon\sqrt{n}\|\lambda(\tau)\|\right]$$

By the properties of norms $\|\cdot\|$, we know that:[2]

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) + \sqrt{n}\lambda(\tau)\right\| \geq \sqrt{n}\|\lambda(\tau)\| - \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0)\right\| \quad \text{with probability one}$$

Therefore

$$\Pr\left[\sqrt{n}\|\lambda(\tau)\| - \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0)\right\| < \epsilon + \epsilon\sqrt{n}\|\lambda(\tau)\|\right] \longrightarrow 1$$

$$\therefore\ \Pr\left[\sqrt{n}\|\lambda(\tau)\| < \frac{\epsilon}{1-\epsilon} + \frac{1}{1-\epsilon}\underbrace{\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i, \theta_0)\right\|}_{=O_p(1)}\right] \longrightarrow 1$$

[2] This is where I screwed up during lecture. I used the "wrong side" of the triangle inequality and got stuck with an upper bound...

Therefore $\sqrt{n}\|\lambda(\tau)\| = O_p(1)$. Given this, it is straightforward to complete the proof of the claim since

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i,\theta_0) + \sqrt{n}\lambda(\tau) = \underbrace{\frac{\sum_{i=1}^{n}\left[\psi(x_i,\theta_0)+\lambda(\tau)\right]}{\sqrt{n}+n\|\lambda(\tau)\|}}_{=o_p(1)} \times \underbrace{\frac{\sqrt{n}+n\|\lambda(\tau)\|}{\sqrt{n}}}_{=O_p(1)} = o_p(1) \quad \square$$

Now let's go back to our original problem and illustrate exactly how we will exploit the result of Claim 1:

$$\left\|\frac{\sum_{i=1}^{n}\left[\psi(x_i,\theta_0)+\lambda(T_n)\right]}{\sqrt{n}+n\|\lambda(T_n)\|}\right\| \le \left\|\frac{\sum_{i=1}^{n}\left[\psi(x_i,\theta_0)-\psi(x_i,T_n)+\lambda(T_n)\right]}{\sqrt{n}+n\|\lambda(T_n)\|}\right\| + \left\|\frac{\sum_{i=1}^{n}\psi(x_i,T_n)}{\sqrt{n}}\right\|. \tag{4}$$

If $\sqrt{n}^{-1}\sum_{i=1}^{n}\psi(x_i,T_n) = o_p(1)$, then all we have to do in order to verify that the condition of Claim 1 is satisfied is to check that the first term on the right-hand side of (4) is $o_p(1)$. Since $T_n = \theta_0 + o_p(1)$, this amounts to prove that $\exists\ d_0$ such that

$$\sup_{\|\tau-\theta_0\|<d_0}\left\|\frac{\sum_{i=1}^{n}\left[\psi(x_i,\theta_0)-\psi(x_i,\tau)+\lambda(\tau)\right]}{\sqrt{n}+n\|\lambda(\tau)\|}\right\| = o_p(1) \tag{5}$$

This would suffice because $\Pr\big[\|T_n-\theta_0\| < d_0 \to 1$ for any $d_0$. This is what Lemma 3 in Huber (1967) is all about: providing sufficient conditions for (5) to hold. **To summarize, if: (a) $T_n = \theta_0 + o_p(1)$; (b) $\sum_{i=1}^{n}\psi(x_i,T_n) = o_p(\sqrt{n})$; (c) (5) is satisfied, then $\sqrt{n}(T_n-\theta_0)$ is asymptotically normal, via Claim 1, Equation (2) and the conditions needed for Equation (3) to hold.**

# 2 Asymptotic normality (Lemma 3 in Huber)

Consider a parameter space $\Theta \subset \mathbb{R}^m$ which is not necessarily compact (it has to be "locally compact"). We have an underlying probability space $(\mathcal{X},\mathcal{F},P)$, an iid sample $x_1,\dots,x_n$ from the distribution $P$, and a function $\psi : \mathcal{X}\times\Theta \to \mathbb{R}^m$. Consider a sequence of statistics $T_n = T_n(x_1,x_2,\dots,x_n) \in \mathbb{R}^m$ that satisfies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x_i,T_n) \xrightarrow{p} 0$$

we're interested in describing conditions under which *any* such sequence $T_n$ is asymptotically normal. A clear example of such a setting would be an $M$-estimation problem in which some estimator $\widehat{\theta}$ satisfies $\sum_{i=1}^{n}\psi(x_i,\widehat{\theta}) = 0$ (interpret this equation as some "first order conditions"), but obviously our setup is more general than this[3]. We will make the following assumptions:

(N-1) For any $\theta \in \Theta$, the function $\psi(x,\theta)$ has nice measurability properties. In particular, $\psi(x,\theta)$ is "separable"[4].

(N-2) Let $\lambda(\theta) = E\psi(x,\theta)$. Then there exists $\theta_0 \in \Theta$ such that $\lambda(\theta_0) = 0$.

(N-3) Let $\|\cdot\|$ denote a norm in $\mathbb{R}^m$. Then there exist $d_0$, $a$, $b$ and $c$ such that

(i) $\big\|\lambda(\theta)\big\| \geq a \cdot \big\|\theta-\theta_0\big\|$ for all $\big\|\theta-\theta_0\big\| \leq d_0$,

(ii) $E\left[\sup_{\|\tau-\theta\|\leq d}\Big\|\psi(x,\tau)-\psi(x,\theta)\Big\|\right] \leq b\cdot d$ for all $\big\|\theta-\theta_0\big\| \leq d_0 - d$ and any $d \geq 0$,

(iii) $E\left[\left(\sup_{\|\tau-\theta\|\leq d}\Big\|\psi(x,\tau)-\psi(x,\theta)\Big\|\right)^2\right] \leq c\cdot d$ for all $\big\|\theta-\theta_0\big\| \leq d_0 - d$ and any $d \geq 0$

(N-4) $E\left[\big\|\psi(x,\theta_0)\big\|^2\right] < \infty$.

Define the following object:

$$Z_n(\tau,\theta) = \frac{\Big\|\sum_{i=1}^{n}\big[(\psi(x_i,\tau)-\psi(x_i,\theta))-(\lambda(\tau)-\lambda(\theta))\big]\Big\|}{\sqrt{n}+n\big\|\lambda(\tau)\big\|}$$

We have the following result (Lemma 3 in Huber (1967)).

**Lemma 1** *If (N-1), (N-2) and (N-3) are satisfied, then*

$$\mathop{sup}_{\|\tau-\theta_0\|\leq d_0} Z_n(\tau,\theta_0) \stackrel{p}{\longrightarrow} 0$$

Proof: Given Assumption (N-3), things behave nicely in the set $\mathbb{R}^m \supset \|\theta-\theta_0\| \leq d_0$. The basic trick of the proof is to partition this cube into smaller cubes, such that within each one of those cubes, $Z_n(\tau,\theta_0)$ remains bounded in probability. The number of such cubes will

[3]For example, think about a problem in which some first order conditions are satisfied asymptotically, but not necessarily for any fixed sample size. Least Absolute Deviations will be a perfect example as we will see.

[4]Separability ensures that objects like $\inf_{\theta}|\psi(x,\theta)|$ are measurable.

depend on the sample size $n$. Things will go our way if the number of cubes grows slowly enough with respect to $n$. Some version of a "partition trick" is often present when we want to prove uniform consistency or asymptotic normality under weak assumptions, even when we are dealing with infinite-dimensional spaces (we'll talk about this in our discussion about empirical processes).

From now on, $\|\cdot\|$ will refer to $\|\cdot\|_\infty$, this is without loss of generality, but it saves us the nuisance of carrying on some redundant constants. Take $q \leq 1/2$. Take the following collection of $k_0$ cubes, all of which are centered at $\theta_0$:

$$C_k = \left\{\|\theta - \theta_0\| \leq d_0(1-q)^k\right\}, \quad k = 1, \dots, k_0$$

so the smallest cube is $C_{k_0}$ and the largest one is $\|\theta - \theta_0\| \leq d_0$. Next, we will cover $C_{k-1} \backslash C_k$ with more cubes, for $k = 1, \dots, k_0$. The minimum distance between a point in the border of $C_{k-1}$ and a point in the border of $C_k$ is $d_0(1-q)^{k-1} - d_0(1-q)^k = d_0 q(1-q)^{k-1}$. We will cover $C_{k-1} \backslash C_k$ with more cubes by partitioning each of the sides of $C_{k-1}$ into segments of length $d_0 q(1-q)^{k-1}$. Therefore, the number of new cubes that cover $C_{k-1} \backslash C_k$ is $\leq \left(d_0(1-q)^{k-1}/d_0[q(1-q)^{k-1}]\right)^m = q^{-m}$ for every $k$. After repeating this process for every $k = 1, \dots, k_0$, we end up with the central cube $C_{k_0}$ and a collection of new cubes $C_{(1)}, C_{(2)}, \dots, C_{(R)}$, where $R \leq \sum_{k=1}^{k_0} q^{-m} = k_0 q^{-m}$. The collection of cubes $C_{k_0}, C_{(1)}, \dots, C_{(R)}$ cover the set $C_0 = \left\{\tau : \left\|\tau - \theta_0\right\| \leq d_0\right\}$ entirely.

The object of interest in this lemma is $\sup_{\tau \in C_0} Z_n(\tau, \theta_0)$. Since our collection of cubes covers $C_0$, we easily have

$$\sup_{\tau \in C_0} Z_n(\tau, \theta_0) \leq \sup_{\tau \in C_{k_0}} Z_n(\tau, \theta_0) + \sum_{j=1}^{R} \sup_{\tau \in C_{(j)}} Z_n(\tau, \theta_0)$$

For $j = 1, \dots, R$, take the cube $C_{(j)}$. Then $C_{(j)} \subset C_{k-1} \backslash C_k$ for some $k = 1, \dots k_0$ and denote its center point as $\xi_j$. By (N-3)(i), $\|\lambda(\tau)\| \geq a \cdot \|\tau - \theta_0\| \geq a d_0 \cdot (1-q)^k$ for some $k \leq k_0$. Adding and subtracting $\psi(x_i, \xi_j) - \lambda(\psi_j)$ and using the triangle inequality we have

$$\begin{aligned}
\sup_{\tau \in C_{(j)}} Z_n(\tau, \theta_0) &\leq \sup_{\tau \in C_{(j)}} Z_n(\tau, \xi_j) + \frac{\left\|\sum_{i=1}^n \left[\psi(x_i, \xi_j) - \psi(x_i, \theta_0) - \lambda(\xi_j)\right]\right\|}{\sqrt{n} + n \cdot a d_0 \cdot (1-q)^k} \\
&\leq \sup_{\tau \in C_{(j)}} Z_n(\tau, \xi_j) + \frac{\left\|\sum_{i=1}^n \left[\psi(x_i, \xi_j) - \psi(x_i, \theta_0) - \lambda(\xi_j)\right]\right\|}{n \cdot a d_0 \cdot (1-q)^k}
\end{aligned}$$

Now let us examine $\sup_{\tau \in C_{(j)}} Z_n(\tau, \xi_j)$. The triangle inequality and Jensen's inequality yield the following

$$\sup_{\tau \in C_{(j)}} Z_n(\tau, \xi_j) \leq \frac{\sum_{i=1}^n \sup_{\tau \in C_{(j)}} \Big\| \Big[ \psi(x_i, \tau) - \psi(x_i, \xi_j) - (\lambda(\tau) - \lambda(\xi_j)) \Big] \Big\|}{n \cdot ad_0 \cdot (1-q)^k}$$

$$\leq \frac{\sum_{i=1}^n \Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x_i, \tau) - \psi(x_i, \xi_j) \Big\| + \sup_{\tau \in C_{(j)}} \Big\| \lambda(\tau) - \lambda(\xi_j) \Big\| \Big]}{n \cdot ad_0 \cdot (1-q)^k}$$

$$\text{By Jensen's inequality: } \leq \frac{\sum_{i=1}^n \Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x_i, \tau) - \psi(x_i, \xi_j) \Big\| + E\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \Big]}{n \cdot ad_0 \cdot (1-q)^k}$$

Take any $\epsilon > 0$. Then

$$\begin{aligned}
&\Pr\Big( \sup_{\tau \in C_{(j)}} Z_n(\tau, \xi_j) \geq \epsilon \Big) \\
&= \Pr\Bigg( \sum_{i=1}^n \Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x_i, \tau) - \psi(x_i, \xi_j) \Big\| + E\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \Big] \geq n \cdot ad_0 \cdot (1-q)^k \Bigg) \\
&= \Pr\Bigg( \sum_{i=1}^n \Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x_i, \tau) - \psi(x_i, \xi_j) \Big\| - E\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \Big] \\
&\qquad \geq n \cdot ad_0 \cdot (1-q)^k \epsilon - 2nE\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \Bigg)
\end{aligned}$$

<u>The last line illustrates why we include $n\|\lambda(\tau)\|$ in the denominator of $Z_n(\tau, \theta)$!</u> By construction of the partition of cubes, for any $\xi_j$ we have:

$$\|\xi_j - \theta_0\| \leq d_0(1-q)^k + \frac{1}{2} d_0 q (1-q)^{k-1} = d_0(1-q)\big[1 - q/2\big] \quad \text{for some } k \leq k_0$$

$$\|\xi_j - \theta_0\| + \underset{\tau \in C_{(j)}}{\|} \xi_j - \tau\| \leq \frac{d_0 q (1-q)^{k-1}}{2} + d_0(1-q)\big[1 - q/2\big] = d_0(1-q)^{k-1} \leq d_0$$

Therefore, (N-3)(ii)-(iii) yield:

$$E\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \leq \frac{b d_0 q (1-q)^{k-1}}{2}$$

$$E\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\|^2 \Big] \leq \frac{c d_0 q (1-q)^{k-1}}{2}$$

And consequently,

$$n \cdot ad_0 \cdot (1-q)^k \epsilon - 2nE\Big[ \sup_{\tau \in C_{(j)}} \Big\| \psi(x, \tau) - \psi(x, \xi_j) \Big\| \Big] \geq n d_0 (1-q)^k \big[ a\epsilon - bq(1-q) \big]$$

Therefore for any $\epsilon > 0$,

$$\Pr\Big(\sup_{\tau\in C_{(j)}} Z_n(\tau,\xi_j) \geq \epsilon\Big) \leq$$

$$\Pr\bigg(\Big|\sum_{i=1}^{n}\Big[\sup_{\tau\in C_{(j)}}\Big\|\psi(x_i,\tau)-\psi(x_i,\xi_j)\Big\| - E\Big[\sup_{\tau\in C_{(j)}}\Big\|\psi(x,\tau)-\psi(x,\xi_j)\Big\|\Big]\Big]\Big| \geq \frac{n(1-q)^{k-1}}{4}\big[a\epsilon(1-q)-2bd_0q\big]\bigg)$$

Using **Chebyshev's inequality** and the fact that $k \leq k_0$ and $q \in (0,1)$ we have

$$\Pr\Big(\sup_{\tau\in C_{(j)}} Z_n(\tau,\xi_j) \geq \epsilon\Big) \leq \frac{qc}{nd_0^2(1-q)^{k_0-1}}\frac{1}{\big[\epsilon a(1-q)-bq\big]^2}$$

Using the assumptions of the Lemma, along with Chebyshev's inequality, it is easier to establish that for any $\epsilon > 0$:

$$\Pr\left[\frac{\Big\|\sum_{i=1}^{n}\Big[\psi(x_i,\xi_j)-\psi(x_i,\theta_0)-\lambda(\xi_j)\Big]\Big\|}{n\cdot ad_0\cdot(1-q)^k} > \epsilon\right] \leq \frac{c}{n\epsilon^2 d_0(1-q)^{k_0}}$$

and therefore

$$\Pr\Big[\sup_{\tau\in C_{(j)}} Z_n(\tau,\theta_0) > \epsilon\Big] \leq \frac{c}{n\epsilon^2 d_0(1-q)^{k_0}} + \frac{qc}{nd_0^2(1-q)^{k_0-1}}\frac{1}{\big[\epsilon a(1-q)-bq\big]^2}$$

The key to make this go to zero is the behavior of $n(1-q)^{k_0}$. We need to make this term go to infinity.

We now move to the center cube $C_{k_0}$. Repeating the same Jensen-inequality arguments and simply using the fact that $n\|\lambda(\tau)\| \geq 0$ (this is how we will bound the denominator of $Z_n(\tau,\theta_0)$ for $C_{k_0}$, unlike the way we did it above), we get:

$$\sup_{\tau\in C_{k_0}} Z_n(\tau,\theta_0) \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\sup_{\tau\in C_{k_0}}\big\|\psi(x_i,\theta_0)-\psi(x_i,\tau)\big\| + E\Big[\sup_{\tau\in C_{k_0}}\big\|\psi(x_i,\theta_0)-\psi(x_i,\tau)\big\|\Big]\Big]$$

Take any $\epsilon > 0$. Adding and subtracting $n\cdot E\Big[\sup_{\tau\in C_{k_0}}\big\|\psi(x_i,\theta_0)-\psi(x_i,\tau)\big\|\Big]$, using Chebyshev's inequality and applying (N-3)(ii)-(iii), we get:

$$\Pr\Big[\sup_{\tau\in C_{k_0}} Z_n(\tau,\theta_0) > \epsilon\Big] \leq \frac{n\cdot c(1-q)^{k_0}d_0}{(\sqrt{n}\epsilon - 2nbd_0(1-q)^{k_0})^2}$$

Now, since

$$\sup_{\tau\in C_0} Z_n(\tau,\theta_0) \leq \sup_{\tau\in C_{k_0}} Z_n(\tau,\theta_0) + \sum_{j=1}^{R}\sup_{\tau\in C_{(j)}} Z_n(\tau,\theta_0)$$

then for any $\epsilon > 0$, and using the bounds we have found:

$$
\begin{aligned}
\Pr\Big[\sup_{\tau\in C_0} Z_n(\tau,\theta_0) > \epsilon\Big] &\le \Pr\Big[\sup_{\tau\in C_{k_0}} Z_n(\tau,\theta_0) > \epsilon\Big] + \sum_{j=1}^{R} \Pr\Big[\sup_{\tau\in C_{(j)}} Z_n(\tau,\theta_0) > \epsilon\Big] \\
&\le \frac{n\cdot c(1-q)^{k_0} d_0}{(\sqrt{n}\epsilon - 2nbd_0(1-q)^{k_0})^2} + R\cdot\left[\frac{c}{n\epsilon^2 d_0(1-q)^{k_0}} + \frac{qc}{nd_0^2(1-q)^{k_0-1}}\frac{1}{\big[\epsilon a(1-q)-bq\big]^2}\right] \\
&\le \Pr\Big[\sup_{\tau\in C_{k_0}} Z_n(\tau,\theta_0) > \epsilon\Big] + \sum_{j=1}^{R} \Pr\Big[\sup_{\tau\in C_{(j)}} Z_n(\tau,\theta_0) > \epsilon\Big] \\
&\le \frac{n\cdot c(1-q)^{k_0} d_0}{(\sqrt{n}\epsilon - 2nbd_0(1-q)^{k_0})^2} + q^{-m}k_0\cdot\left[\frac{c}{n\epsilon^2 d_0(1-q)^{k_0}} + \frac{qc}{nd_0^2(1-q)^{k_0-1}}\frac{1}{\big[\epsilon a(1-q)-bq\big]^2}\right]
\end{aligned}
$$

The object on the right hand side depends crucially on how $k_0$ grows with $n$. It will go to zero if $q$ and $k_0$ are chosen so that:

$$(1-q)^{k_0} \le n^{-\gamma} < (1-q)^{k_0-1} \quad \text{for some } \frac{1}{2} < \gamma < 1$$

in this case

$$k_0(n) - 1 < \frac{\gamma\cdot\log n}{|\log(1-q)|} \le k_0(n)$$

which means that in the end, $R = O(\log n)$. The number of cubes in the partition grows slower than $\log n$. $\square$.