Lecture 9
Local Power

1 Triangular Arrays – the Tools

Triangular arrays are arrays of random variables of the form:

$$\begin{array}{llllll} X_{1,1} & & & & & \\ X_{2,1} & X_{2,2} & & & & \\ X_{3,1} & X_{3,2} & X_{3,3} & & & \\ \ldots & \ldots & \ldots & \ldots & & \\ X_{n,1} & X_{n,2} & X_{n,3} & \ldots & X_{n,n} & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{array}$$

A row-wise i.i.d. triangular array is a triangular array, in which variables in the same row are mutually independent and have the same distribution. Distributions of random variables in different rows are allowed to be different.

LLNs and CLTs are available for triangular arrays of random variables. They typically require slightly stronger moment conditions than the LLNs and CLTs for i.i.d. sequences of random variables. Here I give a WLLN and a uniform WLLN that are simple but assume stronger than necessary conditions. I also give the Lindeberg-Feller CLT and the Lyapounov CLT.

Theorem 8.1 (WLLN for triangular arrays). Let $\{X_{n,i}\}$ be a row-wise i.i.d. triangular array of random variables. If $\sup_n EX_{n,i}^2 < \infty$, then

$$n^{-1}\sum\nolimits_{i=1}^{n}(X_{ni} - EX_{ni}) \to_p 0.$$

Proof. The proof simply applies the Chebyshev inequality:

$$\begin{aligned} \Pr\left(\left|n^{-1}\sum\nolimits_{i=1}^{n}(X_{ni} - EX_{ni})\right| > \varepsilon\right) &\leq \frac{E\left(n^{-1}\sum_{i=1}^{n}(X_{ni} - EX_{ni})\right)^2}{\varepsilon^2} \\ &= \frac{E(X_{ni} - EX_{ni})^2}{n\varepsilon^2} \\ &\leq \frac{\sup_n EX_{n,i}^2}{n\varepsilon^2} \\ &\to 0 \text{ as } n \to \infty \qquad (1) \end{aligned}$$

The above equation holds for any $\varepsilon > 0$. Thus, Theorem 8.1 holds. Notice that $\sup_n EX_{n,i}^2 < \infty$ can be replaced with $\frac{EX_{n,i}^2}{n} \to 0$ and the same proof goes through. ■

Theorem 8.2 (ULLN for triangular arrays): *Suppose* (a) Θ *is compact,* (b) $g(X_{n,i},\theta)$ *is continuous at each* $\theta \in \Theta$ *with probability one,* (c) $g^2(X_{n,i},\theta)$ *is dominated by a function* $G(X_{n,i})$, i.e. $g^2(X_{n,i},\theta) \leq G(X_{n,i})$, *and* (d) $\sup_n EG(X_{n,i}) < \infty$. *Then*

$$\sup_{\theta\in\Theta}\left|n^{-1}\sum\nolimits_{i=1}^{n}(g(X_{ni},\theta)-Eg(X_{ni},\theta))\right| \to_p 0.$$

The proof of Theorem 8.2 is the same as that of the ULLN1 in Lecture 2 and is omitted.

Theorem 8.3 (Lindeberg-Feller Theorem, Ferguson, p. 27): Let $\{X_{n,i}\}$ be a row-wise independent triangular array of random variables with $EX_{n,i} = 0$ and $E\left(X_{n,i}^2\right) = \sigma_{n,i}^2$. Let $Z_n = \sum_{i=1}^n X_{n,i}$, and let $B_n^2 = Var\left(Z_n\right) = \sum_{i=1}^n \sigma_{n,i}^2$. Then $Z_n/B_n \to_d N(0,1)$ if the Lindeberg condition below holds: for every $\varepsilon > 0$,

$$\frac{1}{B_n^2}\sum\nolimits_{i=1}^{n} E\{X_{n,i}^2 1(|X_{n,i}| > \varepsilon B_n)\} \to 0.$$

Note that the Lindeberg-Feller Theorem only requires $X_{n,i}$ to be row-wise independent not identically distributed.

Corollary 8.4 (Lyapounov Theorem): Suppose that $\{X_{n,i}\}$ is row-wise i.i.d. Then the convergence in distribution holds with the Lindeberg condition replaced by the Lyapounov condition: there exists δ such that

$$\sup_n \frac{E\left|X_{n,i}^{2+\delta}\right|}{\sigma_{n,i}^{2+\delta}} < \infty.$$

Corollary 8.4 is implied by Theorem 8.3 by Markov inequality.

2 Asymptotic Local Power

We want to find convenient approximations to the power functions of Wald, LM, and QLR tests. If we consider a parameter value θ_1 that is in the alternative and calculate the probability of rejection of the null hypothesis as $n \to \infty$, then often this probability goes to one as $n \to \infty$. [1] Since the limiting rejection probability is one, but the finite sample rejection probability is something less than one, this type of asymptotic calculation does not generate useful approximations.

Instead, we consider asymptotic approximations in which the true parameter drifts towards the null hypothesis at just the right rate so that even in the limit the rejection probability is in $(0,1)$.

[1] If it does, we say that the test is **consistent** against the parameter θ_1.

We suppose the true parameter value is θ_n, where

$$\theta_n = \theta_0 + \lambda/\sqrt{n} \text{ for } n \geq 1.$$

Here, θ_0 is a parameter value that satisfies the null hypothesis, $H_0 : h(\theta_0) = 0$.

We ask: What is the limit distribution of $\mathcal{W}_n$, when the true parameter is θ_n, as $n \to \infty$? We have:

$$\begin{aligned}\mathcal{W}_n &= \sqrt{n}h(\widehat{\theta}_n)'(\widehat{H}_n\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}_n')^{-1}\sqrt{n}h(\widehat{\theta}_n)\\ &= \sqrt{n}(h(\widehat{\theta}_n) - h(\theta_n) + h(\theta_n))'(\widehat{H}_n\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}_n')^{-1}\sqrt{n}(h(\widehat{\theta}_n) - h(\theta_n) + h(\theta_n)).\end{aligned}$$

Now, under the sequence of local alternatives, the probability limit of $\hat{Q}_n(\theta)$ is the same function $Q(\theta)$ as when the true value is θ_0 for all n. To see why this is true, consider the ML criterion function $\hat{Q}_n(\theta)$. The function $Q(\theta)$ in this case is $-\int \log f(w,\theta)f(w,\theta_0)d\mu(w)$. Suppose $\{W_i : i = 1, ..., n\}$ are iid for given n and the true parameter is θ_n for $n \geq 1$. Since the true parameter θ_n changes with n, the observations actually form a **triangular array** $\{W_{ni} : i = 1, ..., n;\ n \geq 1\}$ rather than a sequence. Then,

$$\hat{Q}_n(\theta) - E_{\theta_n}\hat{Q}_n(\theta) = -n^{-1}\sum_{i=1}^{n}(\log f(W_{ni},\theta) - E_{\theta_n}\log f(W_{ni},\theta)) \to_p 0$$

by a weak law of large numbers for a triangular array of random variables that are row-wise iid (e.g. Theorem 8.2 above). Next, we have

$$\begin{aligned}E_{\theta_n}\hat{Q}_n(\theta) &= -\int \log f(w,\theta)f(w,\theta_n)d\mu(w)\\ &= -\int \log f(w,\theta)f(w,\theta_0)d\mu(w)\\ &\quad -\int \log f(w,\theta)\frac{\partial}{\partial\theta'}f(w,\theta_n^+)d\mu(w)(\theta_n - \theta_0)\\ &= Q(\theta) - \int \log f(w,\theta)\frac{\partial}{\partial\theta'}f(w,\theta_n^+)d\mu(w)\frac{\lambda}{\sqrt{n}}\\ &= Q(\theta) + O(n^{-1/2}),\end{aligned}$$

where the second equality holds by a mean-value expansion and the last equality holds given some regularity conditions on $f(w,\theta)$ that ensure that $\int \log f(w,\theta)\frac{\partial}{\partial\theta'}f(w,\theta_n^+)d\mu(w) = O(1)$. Hence, we find that $\hat{Q}_n(\theta) \to_p Q(\theta)$, where $Q(\theta)$ is the same as when the true value is θ_0 for all n. The same sort of argument shows that this occurs not just in the ML example, but for extremum estimators in general.

A consequence of the fact that the probability limit of $\hat{Q}_n(\theta)$ is the same function $Q(\theta)$ under the sequence of local alternatives $\{\theta_n : n \geq 1\}$ as under θ_0 is that $\widehat{\theta}_n \to_p \theta_0$ under $\{\theta_n : n \geq 1\}$ as $n \to \infty$. The reason is that, provided we can strengthen pointwise convergence to uniform convergence, Assumption U-WCON holds under $\{\theta_n : n \geq 1\}$ and Assumption ID holds under $\{\theta_n : n \geq 1\}$ provided it holds under θ_0, because it only depends on $Q(\theta)$, which is the same in both cases. Hence, we can use the same consistency proof as in earlier lecture notes to establish that $\widehat{\theta}_n \to_p \theta_0$ under $\{\theta_n : n \geq 1\}$.

In addition, using the same method as we used to prove the asymptotic normality of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ under θ_0, we can establish that

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) \overset{d}{\to} Z \sim N(0, B_0^{-1}\Omega_0 B_0^{-1}) \text{ as } n \to \infty \text{ under } \{\theta_n : n \geq 1\}. \tag{2}$$

Note that the only difference between this result and the asymptotic normality result that holds when the true value is θ_0 is that θ_n is subtracted off, rather than θ_0. To obtain this result, we carry out element by element mean-value expansions of the first-order conditions about the true parameter value θ_n, rather than θ_0 :

$$\begin{aligned} o_p(1) &= n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n) \\ &= n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\theta_n) + \frac{\partial^2}{\partial\theta\partial\theta'}\hat{Q}_n(\theta_n^+)\sqrt{n}(\widehat{\theta}_n - \theta_n), \end{aligned}$$

where θ_n^+ lies between θ_n and θ_0 and, hence, converges to θ_0 as $n \to \infty$. Rearrangement gives

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) = -(B_0 + o_p(1))^{-1} n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\theta_n) \to_d N(0, B_0^{-1}\Omega_0 B_0^{-1}),$$

because $n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\theta_n)$ typically converges in distribution to a $N(0, \Omega_0)$ random variable under $\{\theta_n : n \geq 1\}$. For example, in the ML case,

$$n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\theta_n) = -n^{-1/2}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(W_{ni}, \theta_n) \to_d N(0, \Omega_0) \text{ under } \{\theta_n : n \geq 1\}$$

by a central limit theorem for a triangular array of random variables because $E_{\theta_n}(\partial/\partial\theta)\log f(W_{ni}, \theta_n) = 0$ for all $n \geq 1$.

By (2) and the delta method, we obtain

$$\sqrt{n}(h(\widehat{\theta}_n) - h(\theta_n)) \overset{d}{\to} HZ \sim N(0, HB_0^{-1}\Omega_0 B_0^{-1}H') \text{ as } n \to \infty \text{ under } \{\theta_n : n \geq 1\},$$

where $H = (\partial/\partial\theta')h(\theta_0)$. In addition, we can show that

$$(\widehat{H}_n \widehat{B}_n^{-1} \widehat{\Omega}_n \widehat{B}_n^{-1} \widehat{H}_n')^{-1} \xrightarrow{p} (HB_0^{-1}\Omega_0 B_0^{-1} H')^{-1} \text{ as } n \to \infty \text{ under } \{\theta_n : n \geq 1\},$$

just as occurs under θ_0. In addition, element by element mean value expansions about θ_0 give

$$\begin{aligned} \sqrt{n}h(\theta_n) &= \sqrt{n}(h(\theta_0) + (\partial/\partial\theta')h(\theta_n^+)\lambda/\sqrt{n}) \\ &= H\lambda + o_p(1) \end{aligned}$$

where θ_n^+ lies between θ_n and θ_0 and, hence, satisfies $\theta_n^+ \to_p \theta_0$ as $n \to \infty$. The second equality uses the fact that $h(\theta_0) = 0$ because θ_0 is a null parameter point.

The results above and the continuous mapping theorem yield

$$\begin{aligned} \mathcal{W}_n &= \sqrt{n}(h(\widehat{\theta}_n) - h(\theta_n) + h(\theta_n))'(\widehat{H}_n \widehat{B}_n^{-1} \widehat{\Omega}_n \widehat{B}_n^{-1} \widehat{H}_n')^{-1}\sqrt{n}(h(\widehat{\theta}_n) - h(\theta_n) + h(\theta_n)) \\ &\xrightarrow{d} (HZ + H\lambda)'(HB_0^{-1}\Omega_0 B_0^{-1}H')^{-1}(HZ + H\lambda) \\ &\sim \chi_r^2(\delta) \text{ as } n \to \infty \text{ under } \{\theta_n : n \geq 1\}, \text{ where} \\ \delta &= \lambda' H'(HB_0^{-1}\Omega_0 B_0^{-1}H')^{-1}H\lambda. \end{aligned}$$

Here, $\chi_r^2(\delta)$ denotes a noncentral chi-square distribution with r degrees of freedom and noncentrality parameter δ. Note that

$$\begin{aligned} \delta_n &= nh(\theta_n)'(HB_0^{-1}\Omega_0 B_0^{-1}H')^{-1}h(\theta_n) \\ &\to \delta \text{ as } n \to \infty. \end{aligned}$$

Hence, we can also approximate the distribution of $\mathcal{W}_n$ under θ_n simply by $\chi_r^2(\delta_n)$.

How can we use this result? There are two basic uses. The first is to compare the local power of different tests and to select a good test based on its local power properties. Related to this use is the determination of an optimal test based on local power. Such results, in terms of optimal weighted average power over certain ellipses, are available for Wald, LM, and QLR tests in likelihood contexts (Andrews and Ploberger, 1994).

The second use of local power asymptotic results is to approximate the power of a test in practice for a given sample size n and a given alternative parameter vector θ^* of interest. This is done by approximating the distribution of $\mathcal{W}_n$ by $\chi_r^2(\delta_n)$, where δ_n is defined as above with $\theta_n = \theta^*$. A related calculation is to find out how large δ_n needs to be for the approximate power to be higher than a certain level (say 95%). This gives the researcher useful information in the case of a nonrejection.

Next, we provide two examples to illustrate the usefulness of local power analysis.

Example 1. Let θ be a parameter we are interested in. For example, it can be a regression coefficient. We would like to test

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta \neq 0 \tag{3}$$

Let $\hat{\theta}_n$ be an estimator of θ. Let $\hat{s}_n$ be the standard error of $\hat{\theta}_n$ and $\sqrt{n}\hat{s}_n \to_p \sigma \in (0, 1)$. For example, if θ is a regression coefficient, then σ^2 is proportional to the variance of the regression error term and the variance of the regressor associated with θ.

We can use the t-test:

$$\text{reject if } |t_n| > c_{n,\alpha}, \tag{4}$$

where $t_n = \frac{\hat{\theta}_n}{\hat{\sigma}_n}$ and c_α is the $1 - \alpha/2$ quantile of the standard normal distribution. Under H_0 and reasonable condition, $t_n \to_d N(0, 1)$. Thus, the t-test has correct asymptotic size:

$$\lim_{n\to\infty} \Pr_0 (|t_n| > c_{n,\alpha}) = \Pr\left(|Z_0| > z_{1-\alpha/2}\right) = \alpha, \tag{5}$$

where $\Pr_0$ denotes the probability taken under a distribution of the data with $\theta = 0$, $Z_0 \sim N(0, 1)$ and $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of $N(0, 1)$.

For analyzing the power property of the test, asymptotic power under fixed alternatives is misleading. To see why, consider a fixed alternative $\theta = 0.0001$, then under this fixed alternative (meaning "suppose $EX = 0.0001$"),

$$\frac{t_n}{\sqrt{n}} = \frac{\hat{\theta}_n - 0.0001}{\sqrt{n}\hat{s}_n} + \frac{0.0001}{\sqrt{n}\hat{s}} \to_p \frac{0.0001}{\sigma}. \tag{6}$$

But $c_{n,\alpha} \to z_{1-\alpha/2}$, so

$$\begin{aligned} \lim_{n\to\infty} \Pr_{0.0001} (|t_n| > c_{n,\alpha}) &= \lim_{n\to\infty} \Pr_{0.0001} \left(\left| \frac{t_n}{\sqrt{n}} \right| > \frac{c_{n,\alpha}}{\sqrt{n}} \right) \\ &= 1. \end{aligned} \tag{7}$$

In finite sample, the power of the test is always less than one and can be quite small. In order to calculate the finite sample power, we make the normality assumption: $\hat{\theta}_n \sim N(\theta, \sigma^2/n)$. Also for simplicity, assume that σ is known: say $\sigma = 3$. Then, $\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \sim N(0, 1)$. Let's try a relatively large alternative: $\theta = 0.1$. The fixed alternative analysis will predict that the power against this alternative is close to one. But the actual power at sample size n is:

$$\begin{aligned}
\Pr_{0.1}\left(|t_n| > c_{n,\alpha}\right) &= \Pr_{0.1}\left(\left|\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} + \frac{0.1\sqrt{n}}{\sigma}\right| > c_{n,\alpha}\right) \\
&= \Pr_{0.1}\left(\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} + \frac{0.1\sqrt{n}}{\sigma} > c_{n,\alpha}\right) + \\
&\quad \Pr_{0.1}\left(\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} + \frac{0.1\sqrt{n}}{\sigma} < -c_{n,\alpha}\right) \\
&= \Phi\left(\frac{0.1\sqrt{n}}{\sigma} - c_{n,\alpha}\right) + \Phi\left(-\frac{0.1\sqrt{n}}{\sigma} - c_{n,\alpha}\right)
\end{aligned}$$

Suppose $n = 100$, $\alpha = 0.05$, then $\Pr_{0.1}(|t_n| > c_{n,\alpha}) = 0.063$. The power increases as we consider further and further alternatives, but only slowly. For example, $\Pr_{0.3}(|t_n| > c_{n,\alpha}) = 0.17$, $\Pr_{0.5}(|t_n| > c_{n,\alpha}) = 0.38$, and $\Pr_{0.6}(|t_n| > c_{n,\alpha}) = 0.52$.

The finite sample analysis requires the normality assumption. The local asymptotics theory allows us to approximate the finite sample power without the normality assumption. As discussed in the previous lecture, we can consider a sequence of alternatives that drifts to the null: $\theta_n \to 0$, and derive asymptotic power under this drifting sequence of parameters. If the sequence drifts to zero at appropriate rate, we get the same local power expression as above. We need the following high-level assumption:

Assumption Asy.Norm: Under any sequence of true parameters $\theta_n \to 0$, we have

$$\frac{\hat{\theta}_n - \theta_n}{\hat{s}_n} \to_d N(0,1) \text{ and } \sqrt{n}\hat{s}_n \to_p \sigma. \tag{8}$$

Under Assumption Asy.Norm, and if $\sqrt{n}\theta_n \to b$,

$$\begin{aligned}
\Pr_{\theta_n}\left(|t_n| > c_{n,\alpha}\right) &= \Pr_{\theta_n}\left(\left|\frac{\hat{\theta}_n - \theta_n}{\hat{s}_n} + \frac{\sqrt{n}\theta_n}{\sqrt{n}\hat{s}_n}\right| > c_{n,\alpha}\right) \\
&= \Pr_{\theta_n}\left(\frac{\hat{\theta}_n - \theta_n}{\hat{s}_n} + \frac{\sqrt{n}\theta_n}{\sqrt{n}\hat{s}_n} > c_{n,\alpha}\right) + \\
&\quad \Pr_{\theta_n}\left(\frac{\hat{\theta}_n - \theta_n}{\hat{s}_n} + \frac{\sqrt{n}\theta_n}{\sqrt{n}\hat{s}_n} < -c_{n,\alpha}\right) \\
&\to \Phi\left(\frac{b}{s} - c_{n,\alpha}\right) + \Phi\left(-\frac{b}{s} - c_{n,\alpha}\right)
\end{aligned}$$

This formula approximate the finite sample power derived under normality (above) pretty well.

Example 2. In this example, we use local power to rank two tests for the general inequality constraints discussed in the previous lecture. Let $h_{F,1}$ and $h_{F,2}$ be two parameters of interest determined by the DGP F. We would like to test the inequality restrictions:

$$\begin{aligned} H_0 \quad : \quad & h_{F,1} \leq 0 \\ & h_{F,2} \leq 0. \end{aligned} \tag{9}$$

Let $\hat{h}_1$ and $\hat{h}_2$ be estimators of $h_{F,1}$ and $h_{F,2}$. Suppose that the estimators are asymptotically normal under any drifting sequence of true parameters $\{(h_{F_n,1}, h_{F_n,2})\}$,

$$\sqrt{n}\left(\left(\begin{array}{c} \hat{h}_1 \\ \hat{h}_2 \end{array}\right) - \left(\begin{array}{c} h_{F_n,1} \\ h_{F_n,2} \end{array}\right)\right) \rightarrow_d N(0, \Sigma) \tag{10}$$

for some Σ.

Let $\hat{\Sigma}_n = \left(\begin{array}{cc} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{array}\right)$ be a consistent estimator of $\Sigma = \left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array}\right)$.

Let $\hat{h} = \left(\hat{h}_1, \hat{h}_2\right)$. A simpler version of the Wald test statistic proposed in Lecture 7 is

$$W_n = \left[\frac{\sqrt{n}\hat{h}_1}{\hat{\sigma}_1}\right]_+^2 + \left[\frac{\sqrt{n}\hat{h}_2}{\hat{\sigma}_2}\right]_+^2, \tag{11}$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the two diagonal elements of $\hat{\Sigma}_n$ and $[]_+$ is the positive part operator.

In earlier lectures, two data-dependent critical values are proposed: the Plug-in Asymptotics (PA) and the Generalized Moment Selection (GMS). The PA critical value, c_α^{PA}, is the conditional $1-\alpha$ quantile of

$$[Z_1]_+^2 + [Z_2]_+^2, \text{ where } \left(\begin{array}{c} Z_1 \\ Z_2 \end{array}\right) \sim N\left(0, \left(\begin{array}{cc} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{array}\right)\right), \tag{12}$$

where $\hat{\rho} = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1 \hat{\sigma}_2}$. The GMS critical value, c_α^{GMS}, is the conditional $(1-\alpha)$ quantile of

$$[Z_1]_+^2 \cdot 1(\sqrt{n}\hat{h}_1/\hat{\sigma}_1 > -\log(n)) + [Z_2]_+^2 \cdot 1(\sqrt{n}\hat{h}_2/\hat{\sigma}_2 > -\log(n)). \tag{13}$$

Tests with both critical values are consistent. Consider fixed F such that $h_F \nleq 0$. Without loss of generality, suppose $h_{F,1} > 0$. Under this fixed alternative, both critical values converges in probability to a finite number but

$$W_n \geq \left[\frac{\sqrt{n}\hat{h}_1}{\hat{\sigma}_1}\right]_+^2 = \left[\frac{\sqrt{n}\hat{h}_1 - \sqrt{n}h_{F,1}}{\hat{\sigma}_1} + \frac{\sqrt{n}h_{F,1}}{\hat{\sigma}_1}\right]_+^2 \rightarrow_p \infty, \tag{14}$$

where the divergence holds because $\theta_1 > 0$.

The test with GMS critical value has higher asymptotic local power than the test with PA critical value. It is easy to see $c_\alpha^{GMS} \leq c_\alpha^{PA}$. Thus, the power of the GMS test is no less than the PA test. Under certain drifting sequences of alternatives, the local power of the GMS test is strictly larger than that of the PA test. To see this, consider the drifting sequence of DGPs $\{F_n\}$ such that $\sqrt{n}h_{F_n,1}/\sigma_1 \to b_1 > 0$, $\sqrt{n}h_{F_n,2}/(\sigma_2 \log(n)) \to b_2 < -1$. Under this sequence of true parameters, the first restriction is violated and the second restriction is slack. Asymptotically, we have

$$W_n \to_d (Z_1 + b_1)_+^2 \,. \tag{15}$$

The PA critical value is the $1-\alpha$ quantile of $[Z_1]_+^2 + [Z_2]_+^2$, while the GMS critical value is the $(1-\alpha)$ quantile of $[Z_1]_+^2$ with probability approaching one. Because $[Z_2]_+^2$ is a positive random variable, $c_\alpha^{GMS} < c_\alpha^{PA}$. Moreover, $\text{plim}_{n\to\infty} c_\alpha^{GMS} < \text{plim}_{n\to\infty} c_\alpha^{PA}$. Thus,

$$\lim_{n\to\infty} \Pr_{F_n}\left(W_n > c_\alpha^{GMS}\right) > \lim_{n\to\infty} \Pr_{F_n}\left(W_n > c_\alpha^{PA}\right). \tag{16}$$

The difference between the local power of the GMS test and the PA test are not dramatic when there are only two nonlinear restrictions. When there are many nonlinear restrictions and most of them are typically slack, the difference between GMS and PA tests can be huge.