Some further remarks on QE based Testing using

→ Null asymptotic equivalence between the $\mathcal{L}_n$ and the World statistic under the null $\mathcal{H}_0 : \theta_0 \in \Theta^*$, when $\theta_0 \in \Theta^\circ$. (* Under the $Var(z_{\theta_0}) = \frac{1}{4} \breve{J}_{\theta_0}$ assumption)

The derived limit theory directly implies that under the null and if θ_0 is an interior point

$$\mathcal{L}_n = z_{\theta_0}' \frac{1}{4} \breve{J}_{\theta_0}^{-1} z_{\theta_0} + o_p(1)$$

and the Wald statistic,

$$W_n = n^2 (\hat{\theta}_n - \theta_0)' V_n^{-1} (\hat{\theta}_n - \theta_0) \text{ and}$$

since $V_n^{-1} = \left(\frac{1}{4} \breve{J}_{\theta_0}^{-1} V_{\theta_0} \breve{J}_{\theta_0}^{-1} \right)^{-1} + o_p(1)$, while $\sqrt{n}(\hat\theta_n-\theta_0) = -\frac{1}{2} J_{\theta_0}^{-1} Z_{\theta_0} + o_p(1)$

it is obtained that

$|d_n - W_n| = o_p(1)$, yielding that

the two statistics, not only have the same limiting distribution under the null, but they are also asymptotically of the "same form".

→ Sequences of Local Alternatives

The consistency property of a testing procedure implies that the test can with probability converging to one distinguish the null hypothesis from any fixed distribution that supports the alternative, when the latter holds. A more strenuous environment for the test, would evaluate the performance of the test for distributions in the alternative, which however converge as $n \to +\infty$, to a distribution that lies in the null. One way to achieve this is to consider the performance of the test under sequences of local (to the null) alternatives.

Suppose for simplicity that the null hypothesis is simple,

$$H_0: \theta_0 = \theta^*, \text{ and for } \delta \in \mathbb{R}^p$$

δ → local parameter

$$H_n: \theta_0 = \theta^* + \delta / n^{1/2}$$

$n^{1/2}$ → rate of convergence to H_0.

Notice that under H_n, θ_0 is now considered as dependent on n (it was up to now considered independent of n), yet it converges to the fixed value appearing in the null. When $\delta = 0_{p \times 1}$, $H_n = H_0$.

The local asymptotic power of a given testing procedure, say τ_n under this local hypothesis structure,

$$\wp_\alpha(\tau_n, \theta^*; \delta, n^{1/2}) := \lim_{n \to \infty} \mathbb{P}\left(\tau_n \text{ Rejects } H_0 \,/\, \text{under } H_n\right].$$

It is expected that for small values of the local parameter δ, whereas H_n is closer to H_0, the test would have more difficulty discerning H_n from H_0 compared to the case for larger values of the local parameter.

Consider for brevity, the case of a Wald type test for H_0:

the statistic is $W_n = n(\theta_n-\theta^*)' V_n^{-1}(\theta_n-\theta^*)$

and we have that:

p.d. matrix

a. $n^{1/2}(\theta_n-(\theta^*+\delta/n^{1/2})) \underset{\mathbb{P}_{\theta^*+\delta/n^{1/2}}}{\rightsquigarrow} Z \sim N(0, V_{\theta^*})$

(where now $\mathbb{P}_{\theta^*+\delta/n^{1/2}}$ denotes a distribution that supports H_n)

b. $V_n \xrightarrow[\mathbb{P}_{\theta^*+\delta/n^{1/2}}]{P} V_{\theta^*}$

Then $n^{1/2}(\theta_n-\theta^*) = n^{1/2}(\theta_n-(\theta^*-\delta/n^{1/2}))+\delta \overset{\text{Slutsky's lem.}}{\rightsquigarrow} Z+\delta \sim N(\delta, V_{\theta^*})$

and thereby $W_n \underset{\mathbb{P}_{\theta^*+\delta/n^{1/2}}}{\rightsquigarrow} (Z'+\delta')V_{\theta^*}^{-1}(Z+\delta)$

But what is the distribution of the quadratic form $(z'+\delta')\, V_{\theta^*}^{-1}(z+\delta)$, given that $V_{\theta^*}^{-1/2}(z+\delta) \sim N(\delta, I_{p\times p})$?

Definition of the Non-Central Chi-Square Distribution

The non-central chi-square distribution is a generalization of the chi-square distribution, arising as the distribution of a weighted sum of squared independent normal variables. Specifically, if $Z_i \sim N(\mu_i, 1)$ are independent, then

$$Q = \sum_{i=1}^{k} Z_i^2$$

follows a non-central chi-square distribution with k degrees of freedom and non-centrality parameter

$$\lambda = \sum_{i=1}^{k} \mu_i^2.$$

Probability Density Function (PDF)

The probability density function (PDF) of the non-central chi-square distribution is:

$$f_Q(x) = \frac{1}{2} e^{-(x+\lambda)/2} \left(\frac{x}{\lambda}\right)^{k/4-1/2} I_{k/2-1}\left(\sqrt{\lambda x}\right),$$

where $I_{k/2-1}(\cdot)$ is the modified Bessel function of the first kind.

Relation to Quadratic Forms

The non-central chi-square distribution is closely linked to quadratic forms in normal random variables. It describes the distribution of a positive semidefinite quadratic form

$$Q = \mathbf{X}^\top \mathbf{A} \mathbf{X},$$

where $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{I})$ and $\mathbf{A}$ is symmetric.

Denote the chi-squared distribution with k degrees of freedom and non-centrality parameter by $\chi^2(k,\lambda)$

Observe that $\chi^2_k = \chi^2(k,0)$.

Notice that $\forall k,\lambda, x>0$, $\mathbb{P}(\chi^2_k \leq x) > \mathbb{P}(\chi^2(k,\lambda) \leq x)$,

$\Leftrightarrow$ $\mathbb{P}(\chi^2_k > x) < \mathbb{P}(\chi^2(k,\lambda) > x)$ $\Rightarrow$

$$\mathbb{P}\left(\chi^2(k,\lambda) > q_{\chi^2_k}(1-\alpha)\right) > 1-(1-\alpha) = \alpha.$$

The previous definition then implies that $W_n \underset{P_{\theta^* + \delta/n^{1/2}}}{\rightsquigarrow} \chi^2(p, \delta'\delta)$.

Remember that the rejection region for the test is based on $q_{\chi^2_p}(1-\alpha)$, and thereby it is obtained that

$$\mathcal{P}_\alpha(W_n, \theta^*; \delta, n^{1/2}) = \mathbb{P}\left(\chi^2(p, \delta'\delta) > q_{\chi^2_p}(1-\alpha)\right) \geq \alpha$$

(where equality is obtained for $\delta = 0_{p\times 1}$)

Hence under the assumption framework that led to the exactness and conservatism results for the Wald test, and if a., b. hold, for any $\delta \in \mathbb{R}^p$, the test is asymptotically unbiased (i.e. the power is greater than equal to α) for any sequence of local alternatives of the form $\theta^* + \delta/n^{1/2}$.

[Something analogous is true for the d_n based test if α holds, $V_{\theta^*} = 4\ddot{J}_{\theta^*}$, and

b^*. $\frac{\partial^2 c_n(\theta_n^*)}{\partial\theta\partial\theta'} \xrightarrow[\mathbb{P}_{\theta^*+\delta/n^{1/2}}]{P} \ddot{J}_{\theta^*}$, for any $\theta_n^* \to \theta^*$ – try it

as an exercise!].

But when conditions like α and b (or b^*) hold?

For example, in the context of the NLLS model where the distribution of (X_n, ε_n) is independent of the local parameter a. would hold

if $\frac{1}{n^{1/2}}\sum_{i=0}^{n} \frac{\partial g(X_{(i)}(\theta+\delta/n^{1/2}))}{\partial\theta}\varepsilon_{(i)} \rightsquigarrow Z_{\theta^*}$,

and it can be proven (using arguments pertaining to the notion of metric entropy), that, additionally to the assumptions that led to the usual ($\delta=0_{p\times 1}$) weak convergence, this holds if $\frac{\partial^2 g(x\,\theta)}{\partial\theta\partial\theta'}$ is continuous (w.r.t. θ) over some B_{θ^*}, and

$\mathbb{E}[\sup_{\theta\in B_{\theta^*}} \| \partial^2 g(X_{(i)}\theta)/_{\partial\theta\partial\theta'}\varepsilon_{(i)}\|] < +\infty$.

b. (or b^*) would follow using exactly the same arguments for the Hessian convergence in the case where $\delta=0_{p\times 1}$.

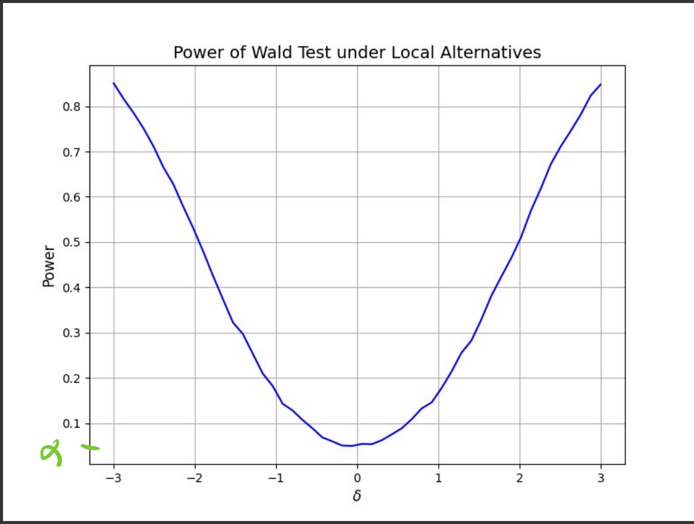
The previous hold trivially for the linear regression case (the Hessian is independent of θ)

Monte Carlo Exp.: Power of Wald test under local alternatives in a simple regression model

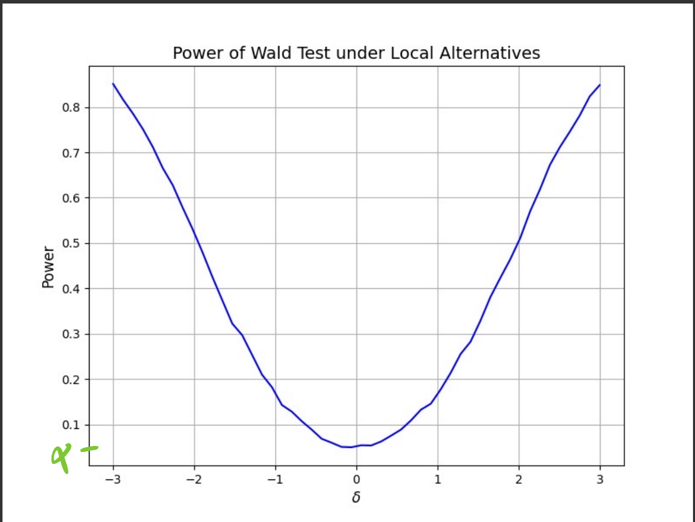
```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import chi2

# Parameters
n = 100  # Sample size
beta_0 = 0  # Intercept
sigma = 1  # Standard deviation of noise
alpha = 0.05  # Significance level
deltas = np.linspace(0, 3, 50)  # Sequence of local alternatives
num_simulations = 8000  # Number of simulations for each delta

# Power calculation
power = []
X = np.random.randn(n, 1)  # Predictor (fixed)

for delta in deltas:
    rejection_count = 0

    for _ in range(num_simulations):
        # Generate data under local alternative
        beta_1 = delta / np.sqrt(n)
        Y = beta_0 + beta_1 * X.flatten() + sigma * np.random.randn(n)

        # Estimate coefficients
        X_aug = np.hstack((np.ones((n, 1)), X))
        beta_hat = np.linalg.inv(X_aug.T @ X_aug) @ X_aug.T @ Y
        beta_1_hat = beta_hat[1]

        # Wald test statistic
        var_beta_1_hat = sigma**2 / (X.T @ X).item()
        wald_stat = beta_1_hat**2 / var_beta_1_hat

        # Compare with chi-square critical value
        if wald_stat > chi2.ppf(1 - alpha, df=1):
            rejection_count += 1

    # Compute power
    power.append(rejection_count / num_simulations)

# Plot power curve
plt.figure(figsize=(8, 6))
plt.plot(deltas, power, 'b-', linewidth=1.5)
plt.xlabel(r'$\delta$', fontsize=12)
plt.ylabel('Power', fontsize=12)
plt.title('Power of Wald Test under Local Alternatives', fontsize=14)
plt.grid(True)
plt.show()
```

Monte Carlo Experiments

$n = 100$



$n = 500$



Notice that the above are MC approximations of the local power of the specific Wald test for fixed n!

(the previous considerations were about their pointwise limits as $n \to +\infty$)