# Assignment 5

Ozone

# Preliminary questions

- 1. Under the case $\beta = 0$, as $n \to \infty$, Prob(selecting the true model) = Prob(Not reject|H0 is true)$= 1 - \alpha$, wehre $\alpha$ is the size of the test. Under the case $\beta \neq 0$, as $n \to \infty$, Prob(slecting the true model) = Prob (Reject| H0 is not true) = power = 1
- 2. Post-test estimator for $\alpha$ is consistent. Under the case $\beta = 0$, no matter whether we choose the restricted model or unrestricted model, the estimator $\hat{\alpha}$ is unbiased and consistent estimator for the true $\alpha$. Under the case $\beta \neq 0$, as $n \to \infty$, the Prob(selecting the true model) is 1, then $\hat{\alpha}$ will be consistent with the true $\alpha$. Therefore, the post-test estimator for $\alpha$ is consistent.

# Simulation (Compare Variance)

Table 1:

| | beta | model | n | variance |
|---|---|---|---|---|
| 1 | 0 | res | 50 | 0.021 |
| 2 | 0 | res | 100 | 0.010 |
| 3 | 0 | res | 150 | 0.007 |
| 4 | 0 | res | 200 | 0.005 |
| 5 | 0 | select | 50 | 0.035 |
| 6 | 0 | select | 100 | 0.017 |
| 7 | 0 | select | 150 | 0.010 |
| 8 | 0 | select | 200 | 0.009 |
| 9 | 0 | unres | 50 | 0.071 |
| 10 | 0 | unres | 100 | 0.034 |
| 11 | 0 | unres | 150 | 0.021 |
| 12 | 0 | unres | 200 | 0.017 |

# Compare Variance Comment

- Under the case of $\beta = 0$, holding the sample size $n$ constant, the variance of the restrictive model estimator is the smallest, then the post-test OLS estimator, and the variance of the unrestricted model estimator is the largest. This makes intuitive sense, becasue when $X_2$ is not relevent in the model, including $X_2$ basically adds more noise to the model, which increases the variance of $\alpha$. We see the results shown in class. The post-test OLS estimator is a combination of the restricted and unrestricted estimtors, therefore the size of the post-test OLS estimator variance is between the restricted and unrestricted estimtors.

# Simulation (Bias Computation)

Table 2:

| | beta | model | n | bias | true_bias |
|---|---|---|---|---|---|
| 1 | 0.160 | res | 200 | 0.113 | 0.112 |
| 2 | 0.160 | select | 200 | 0.064 | |
| 3 | 0.160 | unres | 200 | -0.002 | 0 |
| 4 | 0.240 | res | 200 | 0.166 | 0.168 |
| 5 | 0.240 | select | 200 | 0.072 | |
| 6 | 0.240 | unres | 200 | -0.0002 | 0 |
| 7 | 0.500 | res | 200 | 0.347 | 0.350 |
| 8 | 0.500 | select | 200 | 0.011 | |
| 9 | 0.500 | unres | 200 | -0.007 | 0 |

# Bias Computation Comment

- Theoretically, $E[\hat{\alpha}] - \alpha$ is the bias. When the true model is $y = \alpha X + \beta X2 + \epsilon$ (unrestricted model), where $\beta \neq 0$, $X_1$ and $X_2$ is jointly normal distributed, with sd of 1 and cov of 0.7 and we choose the model $y = \gamma X + e$(restricted model), there will be omitted variable bias. $E[\hat{\alpha}^{res}|\beta \neq 0] - \alpha = \beta \frac{cov(X_1, X_2)}{Var(X_1)} = \beta \cdot 0.7$. If we esimate the true model(unrestricted model), then there will be no bias because $E[\hat{\alpha}^{unres}|\beta \neq 0] - alpha = 0$. For the post-test estimator, $E[\hat{\alpha}^{post}|\beta \neq 0] - \alpha \neq 0$ and the bias will be smaller if beta is larger and sample size n is larger.
- Our simulation is line with our theory above. Under the cases $\beta \neq 0$, the bias of the restricted estimator, is around 0.112, 0.168 and 0.35 repectively. The bias of the unrestricted estimator is around 0 as expected. The bias of the post-estimator is bigger than 0, smaller than the restricted estimator, and decreases as $\beta$ and sample size $n$ increases.

# Simulation (Confidence Interval)

Table 3:

| | beta | model | n | true |
|---|---|---|---|---|
| 4 | 0 | res | 200 | 0.944 |
| 8 | 0 | select | 200 | 0.920 |
| 12 | 0 | unres | 200 | 0.946 |
| 16 | 0.160 | res | 200 | 0.636 |
| 20 | 0.160 | select | 200 | 0.688 |
| 24 | 0.160 | unres | 200 | 0.952 |
| 28 | 0.240 | res | 200 | 0.345 |
| 32 | 0.240 | select | 200 | 0.538 |
| 36 | 0.240 | unres | 200 | 0.951 |
| 40 | 0.500 | res | 200 | 0.002 |
| 44 | 0.500 | select | 200 | 0.861 |
| 48 | 0.500 | unres | 200 | 0.948 |

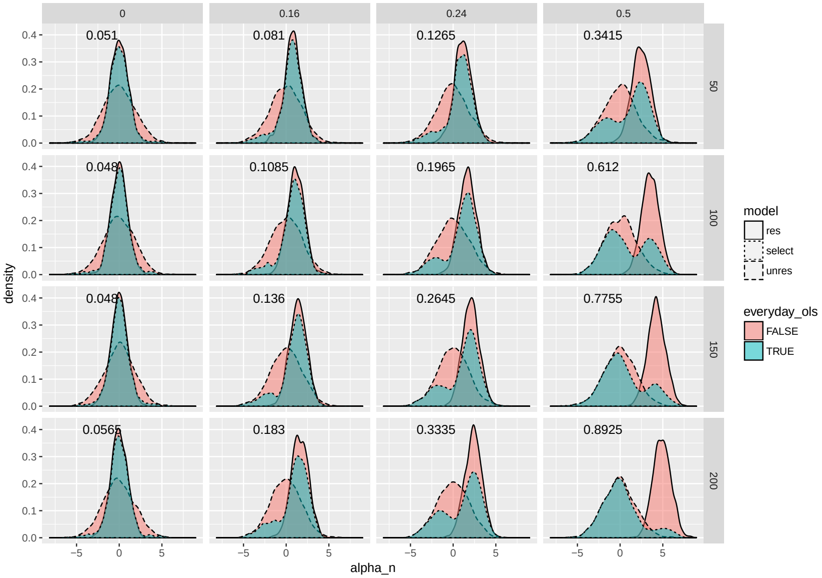
# Confidence Interval Comment

- Under the case $\beta = 0$, about 95% of the time the confidence inverval contains the true $\alpha = 0.2$ no matter we use restricted, unrestriced or everyday OLS. Under the case $\beta \neq 0$, only the the condidence interval of the unrestricted OLS case, contains the true $\alpha$ around 95% of the time. Because both of the two cases, $\hat{\alpha}$ is unbiased and the correspoding T statistics is t-distributed. Therefore the 95% confidence interval contains the true $\alpha$ 95% of the times as constructed.
- When $\beta \neq 0$, the confidence interval of restricted OLS contains the true $\alpha$ all less than 95% of the time. The bigger the beta is, the less times the confidence interval contains the true beta. Notice that when $\beta$ is 0.5, the confidence interval of the restricted model only have 0.2% of the time containing the true $\alpha$. This makes sence, because the confidence interval is conputed as $[\hat{\alpha} - c_{0.25} \cdot se(\hat{\alpha}), \hat{\alpha} + c_{0.25} \cdot se(\hat{\alpha})]$. The bigger the beta is, the more bias the $\hat{\alpha}$ is, the less time the interval computed based on $\hat{\alpha}$ will contain the true $\alpha$.

# Confidence Interval Comment (Continue)

- When $\beta \neq 0$, how many times the confidence interval of post OLS contain $\alpha$ is related to how big the $\beta$ and sample size $n$ is. This could be seen better in the graph in next the slide. Under the case $n = 200$ and $\beta = 0.5$, the distribution of $\hat{\alpha}_{post}$ is closer to the distribution of $\hat{\alpha}_{unres}$ than other $n - \beta$ cases, therefore the times of the confidence interval containing the true *alpha* is the highest, 86% of the time. Under $\beta \neq 0$, the closer the distribution of $\sqrt{n} \cdot (\hat{\alpha}_{post} - 0.2)$ is to the distribution of $\sqrt{n} \cdot (\hat{\alpha}_{unres} - 0.2)$, the time containing the true *alpha* will be more close to 95%.

# Simulation (Plots)



# Simulation Plots Comment

- $\sqrt{n} \cdot (\hat{\alpha} - 0.2)$ is asymptotically $N(0, V_\alpha)$ distributed when $n \to \infty$ and we use the correct model.
- If n is big and the ture $\beta$ is much bigger than zero, the distribution of the post-test OLS is closer to the true distribution.
- The bigger the $\beta$ and $n$ is, the higher the percentage of rejection of the t-test for $\beta$.

# Summary

- Consistency holds no matter whether the ture $\beta$ and the $\beta$ under the null hypothesis are equal or not. Unbiasness does not hold in small sample if the ture $\beta$ and the $\beta$ are different. Efficiency does not hold.
- If "the sample size is big and the ture $\beta$ and the $\beta$ under the null hypothesis are very different" OR "the true $\beta$ is the $\beta$ under the null hypothesis", the usual OLS distribution approximate the distribution of the post-test OLS.