Problem Set 7 Solution

May 31st, 2009 by Yang

1. Monte Carlo Simulation of Bootstrap Methods

We estimate $\theta_0 = \exp(\mu)$ by $\widehat{\theta}_n = \exp(\overline{X})$ where μ is the population average and $\overline{X}$ is the sample average. To construct CI for $\widehat{\theta}_n$, we can use the following method.

1. The delta method: Since $\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \sqrt{n}\left(\exp(\overline{X}) - \exp(\mu)\right) \xrightarrow{d} N\left(0, \sigma^2 \exp(\mu)^2\right)$$

Its asymptotic variance can be consistently estimated by $\sigma^2 \widehat{\theta}_n^2$, where σ is not estimated since it is assumed to be known. So the t-statistic is

$$T_n = \frac{\sqrt{n}(\widehat{\theta}_n - \theta_0)}{\sigma \widehat{\theta}_n}$$

The two-sided symmetric CI for θ_0 is obtained as follows.

$$\text{CI}_{0.95} = \left[\widehat{\theta}_n - \frac{1.96\sigma\widehat{\theta}_n}{\sqrt{n}}, \widehat{\theta}_n + \frac{1.96\sigma\widehat{\theta}_n}{\sqrt{n}}\right]$$

The one-sided CI for θ_0 is

$$\text{CI}_{0.95} = \left(0, \widehat{\theta}_n + \frac{1.65\sigma\widehat{\theta}_n}{\sqrt{n}}\right]$$

2. The nonparamteric iid bootstrap based on B bootstrap resamples: For each $b = 1, \cdots, B$, draw n times randomly from $(X_i)_{i=1}^n$. Calculate $\widehat{\theta}^*_{nb} = \exp(\overline{X}^*_{nb})$, then the bootstrap t-statistic is

$$T^*_{nb} = \frac{\sqrt{n}(\widehat{\theta}^*_{nb} - \widehat{\theta}_n)}{\sigma\widehat{\theta}^*_{nb}}$$

As we have B such bootstrap t-statistics, we calculate the empirical quantiles of the numbers to construct CI. For the two-sided symmetric CI for θ_0, we calculate $\widehat{k}$ such that

$$\frac{1}{B}\sum_{b=1}^{B} 1\left(|T^*_{nb}| \le \widehat{k}\right) \approx 0.95$$

and use this to obtain

$$\text{CI}_{0.95} = \left[\widehat{\theta}_n - \frac{\widehat{k}\sigma\widehat{\theta}_n}{\sqrt{n}}, \widehat{\theta}_n + \frac{\widehat{k}\sigma\widehat{\theta}_n}{\sqrt{n}}\right]$$

Following the lecture note, we may order $|T^*_{nb}|$ from smallest to largest, denote them by $|T^*_n|_{(b)}$, and choose $\widehat{k} = |T^*_n|_{(\llcorner 0.95(B+1) \lrcorner)}$. The one-sided CI is

$$\text{CI}_{0.95} = \left(0, \widehat{\theta}_n - \frac{\widehat{k}\sigma\widehat{\theta}_n}{\sqrt{n}}\right]$$

where $\widehat{k}$ is now defined so as to satisfy

$$\frac{1}{B}\sum_{b=1}^{B} 1\left(T_{nb}^{*} \geq \widehat{k}\right) \approx 0.95$$

It seems strange at a first glance, but note that $\widehat{k} \xrightarrow{p} k$ where k satisfies

$$\begin{aligned}
\Pr(T_n \geq k) = 0.95 \Leftrightarrow \Pr\left(\frac{\sqrt{n}(\widehat{\theta}_n - \theta_0)}{\sigma\widehat{\theta}_n} \geq k\right) &= 0.95 \\
\Leftrightarrow \Pr\left(\widehat{\theta}_n - \theta_0 \geq \frac{k\sigma\widehat{\theta}_n}{\sqrt{n}}\right) &= 0.95 \\
\Leftrightarrow \Pr\left(\theta_0 \leq \widehat{\theta}_n - \frac{k\sigma\widehat{\theta}_n}{\sqrt{n}}\right) &= 0.95
\end{aligned}$$

The last expression is consistent with the one-sided CI provided above. Such $\widehat{k}$ can be obtained as follows. Order T_{nb}^{*} from smallest to largest, denote them by $T_{n(b)}^{*}$, and choose $\widehat{k} = T_{n(\ulcorner 0.05(B+1)\urcorner)}^{*}$.

The results are as follows.

	Empirical Cov. Prob.		Av. Lengths of CI	
	2-sided	1-sided	2-sided	1-sided
Delta Method	0.8890	0.8750	4.1238	3.0885
Bootstrap w/ B= 250	0.9220	0.9230	6.4252	4.5707
Bootstrap w/ B=1000	0.9280	0.9290	6.4260	4.5711
Delta Method (t)	0.9110	0.8840	4.6880	3.2648

Basically using the delta method leads us to overreject the null. The average lengths of the confidence intervals are smaller when using the delta method than when using the bootstrap methods, which is why the delta method generates smaller empirical coverage probabilities. Using the critical values from t_n distribution increases the coverage probabilities by a little, but it is still less than those obtained by the bootstrap methods.

2. True or False

(i) QUESTIONABLE.
It is clear that the power of the test gets worse if we use $T_n^* = \widehat{\theta}^*/s(\widehat{\theta}^*)$ instead of $(\widehat{\theta}^* - \widehat{\theta})/s(\widehat{\theta}^*)$. But the null rejection converage probability may also suffer from centering on $\theta_0 = 0$ instead of $\widehat{\theta}$. To see this, note

$$T_n^* = \frac{\widehat{\theta}^*}{s(\widehat{\theta}^*)} = \frac{\widehat{\theta}^* - \widehat{\theta}}{s(\widehat{\theta}^*)} + \frac{\widehat{\theta}}{s(\widehat{\theta}^*)}$$

The first summand converges in distribution to $N(0,1)$, but we have little information on the second summand, even when $\theta_0 = 0$ is true. Apparently, it is not 0 in probability, so there is a possibility that $T_n^* = \widehat{\theta}^*/s(\widehat{\theta}^*)$ does not converge in distribution to $N(0,1)$. In those cases, the null rejection probability using such T_n^* does not converge to α.

(ii) FALSE.
The population moments will be

$$ET_n = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] = Ey_i = \mu$$
$$var(T_n) = var\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}var(y_i) = \frac{\sigma^2}{n}$$

On the other hand, the bootstrap samples $\{y_1^*, \cdots, y_n^*\}$ are randomly drawn from the sample $\{y_1, \cdots, y_n\}$. So the random variable y_i^* follows the multinomial distribution, which means y_i^* takes on $y_1, \cdots, y_n$ with a probability $\frac{1}{n}$ each. Therefore, $E^* y_i^* = \frac{1}{n}\sum_{i=1}^{n} y_i$.

$$E^*T_n^* = E^*\left[\frac{1}{n}\sum_{i=1}^{n} y_i^*\right] = E^*y_i^* = \frac{1}{n}\sum_{i=1}^{n} y_i = T_n$$
$$var^*(T_n) = var^*\left(\frac{1}{n}\sum_{i=1}^{n} y_i^*\right) = \frac{1}{n}var^*(y_i^*) = \frac{1}{n^2}\sum_{i=1}^{n}(y_i - T_n)^2$$

In general, $ET_n \neq E^*T_n^*$ and $var(T_n) \neq var^*(T_n^*)$. But note that $E^*T_n^* \xrightarrow{p} \mu$ and $var^*(\sqrt{n}T_n^*) \xrightarrow{p} \sigma^2$ as $n \to \infty$.

(iii) TRUE.
The optimal linear forecast $\widehat{P}(y_{t+1}|x_t)$ is only linear in y_{t+1} when the bases x_t include a constant term.

(iv) TRUE.
As discussed in the lecture, the HAC covariance matrix estimator may not be positive definite when the truncated kernel is used. While this does not necessarily mean that the estimator becomes negative definite, it is sometimes the case that it becomes a negative definite matrix. For example, suppose that $v_t := x_t u_t$ is an AR(1) process with mean 0 and the first order parameter $|\rho| < 1$. Now suppose we chose $S_T = 1$ probably because T is small. If we use a truncated kernel, the HAC estimator has the following form.

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} v_t^2 + \frac{1}{T}\sum_{t=2}^{T} v_t v_{t-1} + \frac{1}{T}\sum_{t=2}^{T} v_{t-1} v_t$$

Since v_t is an AR(1) process,

$$E\widehat{\Omega} = Ev_t^2 + \frac{2(T-1)}{T}Ev_t v_{t-1} = \left(1 + \frac{2(T-1)\rho}{T}\right)Ev_t^2$$

Let $\rho = -\frac{2T}{3(T-1)}$, then $E\widehat{\Omega} = -\frac{1}{3}Ev_t^2$ which is negative. So $\widehat{\Omega}$ would be negative in many cases.

(v) FALSE.

S_T needs to grow to infinity for consistency of the HAC covariance matrix estimator, but not because it guarantees the variance of the estimator to converge to 0. It guarantess the bias of the estimator to converge to 0 as $T \to \infty$. This is a necessary condition of consistency. On the other hand, $S_T/T \to 0$ should be satsified since it makes the variance of the estimator converge to 0, which is also a necessary condition of consistency.

3. Optimal Forecast

(i) We would like to use the lemma that if Y_{t+1} and X_t are jointly Guassian processes satisfying

$$\begin{pmatrix} Y_{t+1} \\ X_t \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$$

then,

$$Y_{t+1}|X_t \sim N\left(\mu_1 + \Omega_{12}\Omega_{22}^{-1}(X_t - \mu_2), \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)$$

Let us prove the lemma later and proceed to prove the statement first. By the lemma,

$$\begin{aligned} E[Y_{t+1}|X_t] &= \mu_1 + \Omega_{12}\Omega_{22}^{-1}(X_t - \mu_2) \\ &= (\mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2) + \Omega_{12}\Omega_{22}^{-1}X_t \\ &= \begin{pmatrix} \mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2 & \Omega_{12}\Omega_{22}^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ X_t \end{pmatrix} \end{aligned}$$

which means that $E[Y_{t+1}|X_t]$ is a linear function of 1 and X_t. Since $E[Y_{t+1}|X_t]$ minimizes the mean squared error among all forecasts, its mean sqaured error is no greater than that of all linear forecasts. Therefore it is the optimal linear forecast.

Now prove the lemma as follows. Define $Z_t := Y_{t+1} - \Omega_{12}\Omega_{22}^{-1}X_t$, then

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} I & -\Omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_{t+1} \\ X_t \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} & 0 \\ 0 & \Omega_{22} \end{bmatrix} \right)$$

Since Z_t and X_t are jointly normal and their covariance is 0, they are independent of each other. Therefore, the conditional distribution of Z_t on X_t is the same as the marginal distribution of X_t.

$$Z_t|X_t \sim N(\mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2, \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})$$

Rewrite this as

$$\left[Y_{t+1} - \Omega_{12}\Omega_{22}^{-1}X_t\right]|X_t \sim N(\mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2, \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})$$

or

$$Y_{t+1}|X_t - \Omega_{12}\Omega_{22}^{-1}X_t \sim N(\mu_1 - \Omega_{12}\Omega_{22}^{-1}\mu_2, \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})$$

Therefore we have

$$Y_{t+1}|X_t \sim N\left(\mu_1 + \Omega_{12}\Omega_{22}^{-1}(X_t - \mu_2), \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\right)$$

(ii) Let $\alpha_1 Y_1 + \cdots + \alpha_n Y_n$ be the optimal linear forecast of Y_{n+h}. The it satisfies

$$E[(Y_{n+h} - \alpha_1 Y_1 - \cdots - \alpha_n Y_n) \begin{pmatrix} Y_1 & \cdots & Y_n \end{pmatrix}] = 0$$

We want to show the optimal linear forecast of X_{n+h} based on $1, X_1, \cdots, X_n$ is $\mu + \alpha_1 Y_1 + \cdots + \alpha_n Y_n$. First, we need to show that it is a linear function of $1, X_1, \cdots, X_n$. Note

$$\begin{aligned} \mu + \alpha_1 Y_1 + \cdots + \alpha_n Y_n &= \mu + \alpha_1 (X_1 - \mu) + \cdots + \alpha_n (X_n - \mu) \\ &= \mu(1 - \alpha_1 - \cdots - \alpha_n) + \alpha_1 X_1 + \cdots + \alpha_n X_n \end{aligned}$$

which proves the claim. Now it satisfies the condition of the optimal linear forecast.

$$\begin{aligned} &E[(X_{n+h} - \mu - \alpha_1 Y_1 - \cdots - \alpha_n Y_n) \begin{pmatrix} 1 & X_1 & \cdots & X_n \end{pmatrix}] \\ &\quad = E[(Y_{n+h} - \alpha_1 Y_1 - \cdots - \alpha_n Y_n) \begin{pmatrix} 1 & Y_1 + \mu & \cdots & Y_n + \mu \end{pmatrix}] \\ &\quad = E[(Y_{n+h} - \alpha_1 Y_1 - \cdots - \alpha_n Y_n) \begin{pmatrix} 0 & Y_1 & \cdots & Y_n \end{pmatrix}] \\ &\qquad + E[(Y_{n+h} - \alpha_1 Y_1 - \cdots - \alpha_n Y_n) \begin{pmatrix} 1 & \mu & \cdots & \mu \end{pmatrix}] \\ &\quad = 0 \end{aligned}$$

where the first summand is 0 by the optimality of $\alpha_1 Y_1 + \cdots + \alpha_n Y_n$ among linear forecasts of Y_{n+h}, and the second summand is 0 by the construction of Y_i so that $EY_i = 0$.

4. HAC estimation

The finite sample rejection probabilities of $H_0 : \delta = 0$ are as follows.

Parameters chosen		T=100		T=5000	
rho1	rho2	HAC estimator	White estimator	HAC estimator	White estimator
-0.50	0.00	0.0625	0.0205	0.0460	0.0125
0.00	0.00	0.0755	0.0495	0.0500	0.0470
0.50	0.00	0.0800	0.1035	0.0480	0.0845
0.90	0.00	0.1000	0.1560	0.0480	0.1465
0.99	0.00	0.0935	0.1555	0.0550	0.1605
1.50	-0.75	0.0820	0.1375	0.0470	0.1220
1.90	-0.95	0.0980	0.1705	0.0520	0.1520
0.80	0.10	0.1005	0.1520	0.0500	0.1450

We can observe that the White covariance matrix estimator is not robust to autocorrelation. When ρ_1 or ρ_2 is not 0, u_t is autocorrelated, resulting in distortion of the null rejection probability of the test using the White estimator. On the other hand, the HAC estimator is robust to autocorrelated errors. When $T = 100$, S_T is small and thus causes a bias, which is why we sometimes observe rejection probabilities far from 0.05. This phenomena disappears when $T = 5000$, proving consistency of the HAC estimator as well as robustness of the test using the HAC estimator.