On using sample average and sample variance for time series

JEN-WEN LIN, PhD, CFA

September 11, 2017

1 Using sample average as the estimator to population mean

Let z_t denote a weakly stationary process. A natural estimator of the mean $\mu = E(z_t)$ is the sample mean (time average of observations)

$$\bar{z} = \frac{1}{n} \sum_{t=1}^{n} z_t.$$

We are interested in knowing whether the sample mean is a valid or good estimator.

1. Unbiasedness:
$$E(\bar{z}) = \frac{1}{n} \sum_{t=1}^{n} E(z_t) = \frac{1}{n} \cdot n\mu = \mu.$$

2. Consistency:
$$var(\bar{z}) = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{s=1}^{n} cov(z_t, z_s) = \frac{\gamma_0}{n} \sum_{k=-(n-1)}^{n-1} (1 - \frac{|k|}{n})\rho_k, \qquad (1)$$
where we let $k = (t - s)$. Thus, if
$$lim_{n\to\infty} \left[\sum_{k=-(n-1)}^{n-1} (1 - \frac{|k|}{n})\rho_k \right]$$
is finite, then $var(\bar{z}) \to 0$ as $n \to \infty$, and $\bar{z}$ is a consistent estimator of μ. That is,
$$lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} z_t = \mu \qquad (2)$$
in mean squares.[1]

[1] $\{z_t\}$ is said to be ergodic for the mean if Equation (2) holds. A sufficient condition for this result to hold is that $\rho_k \to 0$ as $k \to \infty$. Intuitively, these results simply said that if z_t and z_{t+k} sufficiently far apart are almost uncorrelated, then some useful new information can be continually added so that the time average will approach the ensemble average.

2 Sample variance is a biased estimator for population variance

The commonly used sample variance estimator is

$$\hat{s}^2 = \frac{\sum_{t=1}^{n}(z_t - \bar{z})^2}{n-1}. \quad (3)$$

Expanding the right hand side of Equation (3), we have

$$\hat{s}^2 = \frac{1}{n-1}\sum_{t=1}^{n}(z_t-\mu)^2 - \frac{2}{n-1}(\bar{z}-\mu)\sum_{t=1}^{n}(z_t-\mu) + \frac{n}{n-1}(\bar{z}-\mu)^2.$$

Using the fact that $n(\bar{z}-\mu) = \sum_{t=1}^{n}(z_t-\mu)$, we can simplify the above equation as

$$\hat{s}^2 = \frac{1}{n-1}\sum_{t=1}^{n}(z_t-\mu)^2 - \frac{n}{n-1}(\bar{z}-\mu)^2. \quad (4)$$

Taking expectation on both sides of the above equation, we have

$$E(\hat{s}^2) = \frac{n}{n-1}\gamma_0 - \frac{n}{n-1}var(\bar{z}), \quad (5)$$

where $\gamma_0 = var(z_t)$. If $\{z_t\}$ are independent or uncorrelated over time, we have

$$E(\bar{z}-\mu)^2 = var(\bar{z}) = \frac{1}{n}\gamma_0. \quad (6)$$

Substituting the above result to Equation (5), we have

$$E(\hat{s}^2) = \frac{n-1}{n-1}\gamma_0 = \gamma_0.$$

The above derivation shows that, under the independence assumption, the commonly used sample variance is an unbiased estimator for $var(z_t)$. However, this is not the case for dependent time series. This bias gets bigger when we estimate standard deviation using the square root of the sample variance. The causes of the bias are discussed in details in the next two sections.

2.1 Bias due to autocorrelation

For dependent time series, the autocorrelation functions are not zero so $var(\bar{z}) \neq \gamma_0/n$ and according to Equation (5), $\hat{s}^2$ is a biased estimator.[2] The magnitude of this bias

[2] Assuming that $\rho_k > 0,\ k = 1, 2, \ldots, n-1$, we have

$$var(\bar{z}) = \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1}(1-\frac{|k|}{n})\rho_k \ > \frac{1}{n}\gamma_0.$$

Hence, we would underestimate the actual variance using the sample variance.

may be gauged using

$$\frac{E(\hat{s}^2)}{\gamma_0} = 1 - \frac{2}{n-1}\left[\sum_{k=1}^{n-1}\rho_k - \frac{1}{n}\sum_{k=1}^{n-1}k\rho_k\right]. \qquad (7)$$

For simplicity and for the demonstration purposes only, let's assume z_t follows an autoregressive process of order one

$$z_t = \phi z_{t-1} + \xi_t, \qquad \xi_t \sim NID(0, \sigma_a^2)$$

For an AR(1) process, the autocorrelation at lag k is simply ϕ^k, and Equation (7) becomes

$$\frac{E(\hat{s}^2)}{\gamma_0} = 1 - \frac{2}{n-1}\left[\sum_{k=1}^{n-1}\phi^k - \frac{1}{n}\sum_{k=1}^{n-1}k\phi^k\right].$$

Table 1 summarizes the relationship between ϕ and the bias ratio, i.e. $E(\hat{s}^2)/\gamma_0$, when the series length is 240. It is seen clearly in Table 1 that $E(\hat{s}^2)$ and $E(\sqrt{\hat{s}^2})$ underestimates the true variance and standard deviation for postie ϕ.

```r
vol.zbar<-function(phi,n=240) {
        temp1<-sum(phi^(1:(n-1)))
        temp2<-sum((1:(n-1))*(phi^(1:(n-1))))/n
        1-2*(temp1-temp2)/(n-1)
}

phiHat<-c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.99)

var.out<-numeric(length(phiHat))
vol.out<-numeric(length(phiHat))

for(k in 1:length(phiHat)) {
        var.out[k]<-vol.zbar(phiHat[k])
        vol.out[k]<-sqrt(vol.zbar(phiHat[k]))
}
Sigma<-round(cbind(var.out,vol.out),4)
colnames(Sigma)<-c("variance","volatility")
row.names(Sigma)<-paste("phi=", phiHat)
kable(Sigma, format = "latex",
     caption="The relationship of autocorrelation and bias ratio (n=240)")
```

2.2 Bias due to sample standard deviation

Another bias is introduced when we evaluate "volatility" using the square root of the conventional sample variance. Since the square root is a strictly concave function,

Table 1: The relationship of autocorrelation and bias ratio (n=240)

	variance	volatility
phi= 0	1.0000	1.0000
phi= 0.1	0.9991	0.9995
phi= 0.2	0.9979	0.9990
phi= 0.3	0.9964	0.9982
phi= 0.4	0.9945	0.9972
phi= 0.5	0.9917	0.9958
phi= 0.6	0.9876	0.9938
phi= 0.7	0.9807	0.9903
phi= 0.8	0.9672	0.9835
phi= 0.9	0.9278	0.9632
phi= 0.99	0.4858	0.6970

it follows from Jensen's inequality that the square root of the sample variance is an underestimate.

One simple correction of this bias is to multiply a constant to the sample standard deviation. Specifically, some statisticians suggest an unbiased estimator of σ using

$$\hat{\sigma} = c_N \hat{s},$$

where the corrective constant c_N satisfies[3]

$$c_N = \left(\frac{N-1}{2}\right)^{1/2} \frac{\Gamma[(N-1)/2]}{\Gamma(N/2)}.$$

Table 2 shows the relationship between the number of observations and the corrective constants (c_N). Table 2 shows that the effect of Jensen's inequality is negligible when the series length is 240.

```
cN<-function(n){
        gamma(0.5*(n-1))*(0.5*(n-1))^0.5/gamma(0.5*n)
}

out<-numeric(24)
for(l in 1:24) out[l]<-cN(l*10)
Out<-cbind(paste((1:24)*10),round(out,4))
colnames(Out)<-c("number of the series length", "Corrective constant")
kable(Out, digits=4, format="latex",
      caption ="Corrective constants for sample standard deviation" )
```

[3]Holtzman, Wayne H., The unbiased estimate of the population variance and standard deviation. *Americian Journal of Psychology,* 1959, 63, 615-617.

Table 2: Corrective constants for sample standard deviation

number of the series length	Corrective constant
10	1.0281
20	1.0132
30	1.0087
40	1.0064
50	1.0051
60	1.0042
70	1.0036
80	1.0032
90	1.0028
100	1.0025
110	1.0023
120	1.0021
130	1.0019
140	1.0018
150	1.0017
160	1.0016
170	1.0015
180	1.0014
190	1.0013
200	1.0013
210	1.0012
220	1.0011
230	1.0011
240	1.001

- **Remark: Sample autocovariance function**
  Like the sample variance, the sample autocovariance functions are biased estimators. Two types of sample autocovariance functions are used in literature

  $$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z}),$$

  or

  $$\tilde{\gamma}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z}).$$

  Approximating $\sum_{t=1}^{n-k}(z_t - \mu)$ and $\sum_{t=1}^{n-k}(z_{t+k} - \mu)$ by $(n-k)(\bar{z} - \mu)$, we have

  $$E(\hat{\gamma}_k) \approx \gamma_k - \frac{k}{n} - \frac{n-k}{n}var(\bar{z}),$$

  and

  $$E(\tilde{\gamma}_k) = \gamma_k - var(\bar{z}).$$

  In general, $\hat{\gamma}_k$ has a larger bias than $\tilde{\gamma}_k$, especially when k is large with respective to n. However, the estimate $\hat{\gamma}_k$ is always positive semidefinite like γ_k but $\tilde{\gamma}_k$ is not necessary so. Additionally, the variance of $\tilde{\gamma}_k$ is larger than $\hat{\gamma}_k$. So, for some types of processes, $\hat{\gamma}_k$ has smaller man square error than $tilde\gamma_k$.