Problems

Problem-1: This problem serves as a review of the Big-O notation. Consider two sequences of non-stochastic real numbers $\{a_n\}$ and $\{b_n\}$ with $n \in \mathbb{N}$. Assume that $a_n = O(n^k)$ and $b_n = O(n^l)$ for some $k, l \in \mathbb{R}$.

(a) Prove that $a_n + b_n = O(n^{\max\{k,l\}})$.

(b) Prove that $a_n b_n = O(n^{k+l})$.

(c) Prove that $a_n^l = O(n^{kl})$ whenever a_n^l is well defined.

(Hint: Simply follow the definition of Big-O.)

Solution-1: (a) Since $a_n = O(n^k)$, there exists $B \in \mathbb{R}_+$ such that $|a_n|/n^k \leq B$ for any $n \in \mathbb{N}$. Since $b_n = O(n^l)$, there exists $C \in \mathbb{R}_+$ such that $|b_n|/n^l \leq C$ for any $n \in \mathbb{N}$. Assume without loss of generality that $\max\{k,l\} = k$. Take $A = B + C$, then

$$\frac{|a_n + b_n|}{n^{\max\{k,l\}}} = \frac{|a_n + b_n|}{n^k} \leq \frac{|a_n| + |b_n|}{n^k} \leq \frac{|a_n|}{n^k} + \frac{|b_n|}{n^l} \leq B + C = A.$$

Hence, $a_n + b_n = O(n^{\max\{k,l\}})$.

(b) Since $a_n = O(n^k)$, there exists $B \in \mathbb{R}_+$ such that $|a_n|/n^k \leq B$ for any $n \in \mathbb{N}$. Since $b_n = O(n^l)$, there exists $C \in \mathbb{R}_+$ such that $|b_n|/n^l \leq C$ for any $n \in \mathbb{N}$. Take $A = BC$, then

$$\frac{|a_n b_n|}{n^{k+l}} = \frac{|a_n||b_n|}{n^{k+l}} = \frac{|a_n|}{n^k} \times \frac{|b_n|}{n^l} \leq BC = A.$$

Hence, $a_n b_n = O(n^{k+l})$.

(c) Since $a_n = O(n^k)$, there exists $B \in \mathbb{R}_+$ such that $|a_n|/n^k \leq B$ for any $n \in \mathbb{N}$. Whenever a_n^l is well defined, we have that

$$\frac{|a_n^l|}{n^{kl}} = \frac{|a_n|^l}{n^{kl}} = \frac{|a_n|}{n^k} \leq B.$$

Hence, $a_n^l = O(n^{kl})$.

Problem-2: This problem serves as a review of the Big-Op notation. Consider two sequences of real, stochastic numbers $\{x_n\}$ and $\{y_n\}$ with $n \in \mathbb{N}$. Assume that $x_n = O_p(n^k)$ and $y_n = O_p(n^l)$ for some $k, l \in \mathbb{R}$.

(a) Prove that $x_n + y_n = O_p(n^{\max\{k,l\}})$. (Hint: Use a generic rule of probability that $P(A) = P(A \cap B) + P(A \cap \bar{B})$, where $\bar{B}$ is the complement of B.)

(b) Prove that $x_n y_n = O_p(n^{k+l})$. (Hint: Use a generic rule of probability that $P(A) = P(A \cap B) + P(A \cap \bar{B})$.)

(c) Prove that $x_n^l = O_p(n^{kl})$ whenever x_n^l is well defined. (Hint: Simply follow the definition of Big-Op.)

Solution-2: (a) Since $x_n = O_p(n^k)$, for any $\alpha/2 \in \mathbb{R}_{++}$, there exists $B_\alpha \in \mathbb{R}_+$ such that $P[|x_n|/n^k > B_\alpha] < \alpha/2$ for any $n \in \mathbb{N}$. Since $y_n = O_p(n^l)$, for any $\alpha/2 \in \mathbb{R}_{++}$, there exists $C_\alpha \in \mathbb{R}_+$ such that $P[|y_n|/n^l > C_\alpha] < \alpha/2$ for any $n \in \mathbb{N}$. Assume without loss of generality that $\max\{k,l\}=k$. Take $A_\alpha = B_\alpha + C_\alpha$, then

$$\begin{aligned}
P\left[\frac{|x_n+y_n|}{n^{\max\{k,l\}}} > A_\alpha\right] &= P\left[\frac{|x_n+y_n|}{n^k} > A_\alpha\right] \\
&= P\left[\frac{|x_n+y_n|}{n^k} > A_\alpha, \frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|x_n+y_n|}{n^k} > A_\alpha, \frac{|x_n|}{n^k} \leq B_\alpha\right] \\
&\leq P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|x_n|+|y_n|}{n^k} > B_\alpha + C_\alpha, \frac{|x_n|}{n^k} \leq B_\alpha\right] \\
&\leq P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|y_n|}{n^k} > C_\alpha\right] \\
&\leq P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|y_n|}{n^l} > C_\alpha\right] \\
&< \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha,
\end{aligned}$$

where the second equality holds from the generic rule of probability that $P(A) = P(A \cap B) + P(A \cap \bar{B})$. Hence, $x_n + y_n = O_p(n^{\max\{k,l\}})$.

(b) Since $x_n = O_p(n^k)$, for any $\alpha/2 \in \mathbb{R}_{++}$, there exists $B_\alpha \in \mathbb{R}_+$ such that $P[|x_n|/n^k > B_\alpha] < \alpha/2$ for any $n \in \mathbb{N}$. Since $y_n = O_p(n^l)$, for any $\alpha/2 \in \mathbb{R}_{++}$, there exists $C_\alpha \in \mathbb{R}_+$ such that $P[|y_n|/n^l > C_\alpha] < \alpha/2$ for any $n \in \mathbb{N}$. Take $A_\alpha = B_\alpha C_\alpha$, then

$$\begin{aligned}
P\left[\frac{|x_n y_n|}{n^{k+l}} > A_\alpha\right] &= P\left[\frac{|x_n y_n|}{n^{k+l}} > A_\alpha, \frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|x_n y_n|}{n^{k+l}} > A_\alpha, \frac{|x_n|}{n^k} \leq B_\alpha\right] \\
&\leq P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|x_n||y_n|}{n^{k+l}} > A_\alpha, \frac{|x_n|}{n^k} \leq B_\alpha\right] \\
&= P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|x_n|}{n^k} \times \frac{|y_n|}{n^l} > B_\alpha C_\alpha, \frac{|x_n|}{n^k} \leq B_\alpha\right] \\
&\leq P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[B_\alpha \times \frac{|y_n|}{n^l} > B_\alpha C_\alpha\right] \\
&= P\left[\frac{|x_n|}{n^k} > B_\alpha\right] + P\left[\frac{|y_n|}{n^l} > C_\alpha\right] \\
&< \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.
\end{aligned}$$

Hence, $x_n y_n = O_p(n^{k+l})$.

(c) Since $x_n = O_p(n^k)$, for any $\alpha \in \mathbb{R}_{++}$, there exists $B_\alpha \in \mathbb{R}_+$ such that $P[|x_n|/n^k > B_\alpha] < \alpha$ for any $n \in \mathbb{N}$. Whenever x_n^l is well defined, we have that

$$P\left[\frac{|x_n^l|}{n^{kl}} > B_\alpha\right] = P\left[\frac{|x_n|^l}{n^{kl}} > B_\alpha\right] = P\left[\frac{|x_n|}{n^k} > B_\alpha\right] < \alpha.$$

Hence, $x_n^l = O_p(n^{kl})$ whenever x_n^l is well defined.

Problem-3: Suppose that a true data generating process (DGP) is a *trend stationary* process:

$$y_t = \alpha_0 + \delta_0 \times t + \epsilon_t, \quad t = 1, \dots, n,$$

where $\epsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$ with $E[\epsilon_t^4] < \infty$. (Remark: $\{y_t\}$ behaves like an i.i.d. process around a linear time trend $\alpha_0 + \delta_0 \times t$). Let $x_t = [1, t]'$ and $\beta_0 = [\alpha_0, \delta_0]'$, then the DGP is rewritten as

$$y_t = x_t'\beta_0 + \epsilon_t.$$

Suppose that we fit an exactly specified model:

$$y_t = x_t'\beta + u_t, \tag{1}$$

where $\beta = [\alpha, \delta]'$. Run ordinary least squares (OLS) to get

$$\hat{\beta}_n = \begin{bmatrix} \hat{\alpha}_n \\ \hat{\delta}_n \end{bmatrix} = \left[\sum_{t=1}^n x_t x_t'\right]^{-1} \left[\sum_{t=1}^n x_t y_t\right].$$

(Remark: Running model (1) is called *detrending*. The resulting residual $\hat{u}_t = y_t - x_t'\hat{\beta}_n$, called *detrended series*, is supposed to be stationary.) This problem elaborates the asymptotic property of $\hat{\beta}_n$, which is strikingly different from the standard asymptotics of stationary cases.

(a) Show that

$$\hat{\beta}_n - \beta_0 = \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t\right]. \tag{2}$$

(b) In usual stationary cases, the next step is scaling $\hat{\beta}_n - \beta_0$ by $n^{1/2}$ in order to obtain the asymptotic normality. This approach does *not* work in the trend stationary case. To see why not, consider the usual scaling

$$n^{1/2}(\hat{\beta}_n - \beta_0) = n^{1/2}\left[\sum_{t=1}^{n} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t\right] = \left[\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right]^{-1} \left[\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t \epsilon_t\right]. \tag{3}$$

Show that

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' = \begin{bmatrix} 1 & \frac{1}{2}(n+1) \\ \frac{1}{2}(n+1) & \frac{1}{6}(n+1)(2n+1) \end{bmatrix}. \tag{4}$$

(Remark: This result indicates that $(1/n)\sum_{t=1}^{n} x_t x_t' \to \infty$ as $n \to \infty$ and therefore it is non-invertible. This is why the usual scaling in Eq. (3) does not work.)

(c) In view of Eq. (4), one might be tempted to try another scaling

$$n^{5/2}(\hat{\beta}_n - \beta_0) = \left[\frac{1}{n^3}\sum_{t=1}^{n} x_t x_t'\right]^{-1} \left[\frac{1}{n^{1/2}}\sum_{t=1}^{n} x_t \epsilon_t\right] \tag{5}$$

in order to prevent the divergence of $\sum_{t=1}^{n} x_t x_t'$. This approach does *not* work either, however. To see why not, **show that**

$$\frac{1}{n^3}\sum_{t=1}^{n} x_t x_t' = \begin{bmatrix} \frac{1}{n^2} & \frac{1}{2n}\left(1+\frac{1}{n}\right) \\ \frac{1}{2n}\left(1+\frac{1}{n}\right) & \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) \end{bmatrix} \to \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}. \tag{6}$$

(Remark: $(1/n^3)\sum_{t=1}^{n} x_t x_t'$ converges to a finite but non-invertible matrix. This is why the scaling in Eq. (5) does not work.)

(d) It turns out that a proper scaling factor is $n^{1/2}$ for $\hat{\alpha}_n$ and $n^{3/2}$ for $\hat{\delta}_n$. Define

$$\Upsilon_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}.$$

It turns out that asymptotic normality applies to

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n^{3/2}(\hat{\delta}_n - \delta_0) \end{bmatrix} = \Upsilon_n(\hat{\beta}_n - \beta_0). \tag{7}$$

Below we confirm it via several steps. First, **show that**

$$\Upsilon_n(\hat{\beta}_n - \beta_0) = \left\{ \Upsilon_n^{-1} \left[\sum_{t=1}^{n} x_t x_t' \right] \Upsilon_n^{-1} \right\}^{-1} \times \left\{ \Upsilon_n^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t \right] \right\}. \tag{8}$$

(Hint: Use a generic matrix property that $(AB)^{-1} = B^{-1}A^{-1}$ whenver A and B are invertible.)

(e) Show that

$$\Upsilon_n^{-1} \left[\sum_{t=1}^{n} x_t x_t' \right] \Upsilon_n^{-1} = \begin{bmatrix} 1 & \frac{1}{2}\left(1 + \frac{1}{n}\right) \\ \frac{1}{2}\left(1 + \frac{1}{n}\right) & \frac{1}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right) \end{bmatrix} \to \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \equiv Q. \tag{9}$$

(Remark: $\sum_{t=1}^{n} x_t x_t'$ is now properly scaled.)

(f) Show that

$$\Upsilon_n^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t \right] = \begin{bmatrix} n^{-1/2} \sum_{t=1}^{n} \epsilon_t \\ n^{-1/2} \sum_{t=1}^{n} (t/n) \epsilon_t \end{bmatrix}. \tag{10}$$

(g) By applying central limit theorem to Eq. (10), it can be shown that[1]

$$\Upsilon_n^{-1}\left[\sum_{t=1}^{n} x_t \epsilon_t\right] \xrightarrow{d} N(0, \sigma^2 Q). \tag{11}$$

Using this result, **show that**

$$\Upsilon_n(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \sigma^2 Q^{-1}). \tag{12}$$

(h) Show that

$$n^{1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{d} N(0, 4\sigma^2) \tag{13}$$

and

$$n^{3/2}(\hat{\delta}_n - \delta_0) \xrightarrow{d} N(0, 12\sigma^2). \tag{14}$$

(Remark: Eq. (13) indicates that $\hat{\alpha}_n - \alpha_0 = O_p(n^{-1/2})$; $\hat{\alpha}_n$ converges to the true value α_0 at the usual rate of $n^{1/2}$.) (**Important Remark**: Eq. (14) indicates that $\hat{\delta}_n - \delta_0 = O_p(n^{-3/2})$; $\hat{\delta}_n$ converges to the true value δ_0 at the rate of $n^{3/2}$, which is faster than the usual rate of $n^{1/2}$. $\hat{\delta}_n$ is therefore said to be *superconsistent*. It is expected that $\hat{\delta}_n$ should be an accurate approximation of δ_0 even for a relatively small sample size n.)

Solution-3: (a) We have that

$$\begin{aligned}\hat{\beta}_n &= \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1}\left[\sum_{t=1}^{n} x_t y_t\right] = \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1}\left[\sum_{t=1}^{n} x_t (x_t'\beta_0 + \epsilon_t)\right] \\ &= \beta_0 + \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1}\left[\sum_{t=1}^{n} x_t \epsilon_t\right].\end{aligned}$$

[1] See Hamilton (1994, pp. 458-460). *Time Series Analysis*. Princeton University Press.

Hence Eq. (2) follows.

(b) We have that

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' = \frac{1}{n}\begin{bmatrix} \sum_{t=1}^{n} 1 & \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \end{bmatrix} = \frac{1}{n}\begin{bmatrix} n & \frac{1}{2}n(n+1) \\ \frac{1}{2}n(n+1) & \frac{1}{6}n(n+1)(2n+1) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \frac{1}{2}(n+1) \\ \frac{1}{2}(n+1) & \frac{1}{6}(n+1)(2n+1) \end{bmatrix}.$$

(c) Eq. (6) is a direct implication of Eq. (4).

(d) Eq. (2) implies that

$$\Upsilon_n(\hat{\beta}_n - \beta_0) = \Upsilon_n \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t\right] = \Upsilon_n \left[\sum_{t=1}^{n} x_t x_t'\right]^{-1} (\Upsilon_n \Upsilon_n^{-1}) \left[\sum_{t=1}^{n} x_t \epsilon_t\right]$$
$$= \left\{\Upsilon_n^{-1}\left[\sum_{t=1}^{n} x_t x_t'\right]\Upsilon_n^{-1}\right\}^{-1} \times \left\{\Upsilon_n^{-1}\left[\sum_{t=1}^{n} x_t \epsilon_t\right]\right\}.$$

(e) We have that

$$\Upsilon_n^{-1}\left[\sum_{t=1}^{n} x_t x_t'\right]\Upsilon_n^{-1} = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} n & \frac{1}{2}n(n+1) \\ \frac{1}{2}n(n+1) & \frac{1}{6}n(n+1)(2n+1) \end{bmatrix} \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \frac{1}{2}\left(1+\frac{1}{n}\right) \\ \frac{1}{2}\left(1+\frac{1}{n}\right) & \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) \end{bmatrix} \to Q.$$

(f) We have that

$$\Upsilon_n^{-1}\left[\sum_{t=1}^{n} x_t \epsilon_t\right] = \begin{bmatrix} n^{-1/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{n} \epsilon_t \\ \sum_{t=1}^{n} t\epsilon_t \end{bmatrix} = \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} \epsilon_t \\ n^{-1/2}\sum_{t=1}^{n} (t/n)\epsilon_t \end{bmatrix}.$$

(g) Substitute Eq. (9) and Eq. (11) into Eq. (8) to get Eq. (12).

(h) Eqs. (7) and (12) imply that

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n^{3/2}(\hat{\delta}_n - \delta_0) \end{bmatrix} \xrightarrow{p} N(0, \sigma^2 Q^{-1}). \tag{15}$$

We have that

$$Q^{-1} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix}.$$

Focus on the first element of Eq. (15) to get Eq. (13). Focus on the second element of Eq. (15) to get Eq. (14).

Problem-4: In this problem we perform Monte Carlo simulations on detrending in order to better understand the different convergence rates of $\hat{\alpha}_n$ and $\hat{\delta}_n$. Fix sample size $n \in \{50, 100, 200\}$ and true values $\alpha_0 = 2$, $\delta_0 = 2$, and $\sigma_0^2 = 80$. Execute the following steps for each fixed sample size n, using any statistical software (e.g. Eviews, Excel, Matlab, R, Stata).

Step 1. Generate $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_0^2)$ and construct $y_t = \alpha_0 + \delta_0 \times t + \epsilon_t$ for $t \in \{1, \ldots, n\}$.

Step 2. Run OLS for a linear regression model $y_t = \alpha + \delta \times t + u_t$ to get $\hat{\alpha}_n$ and $\hat{\delta}_n$.

Step 3. Repeat Steps 1-2 $J = 10000$ times to get a set of OLS estimates $\{\hat{\alpha}_n^{(1)}, \ldots, \hat{\alpha}_n^{(J)}\}$ and $\{\hat{\delta}_n^{(1)}, \ldots, \hat{\delta}_n^{(J)}\}$.[2]

(a) Draw histograms of $\{\hat{\alpha}_n^{(1)}, \ldots, \hat{\alpha}_n^{(J)}\}$ and $\{\hat{\delta}_n^{(1)}, \ldots, \hat{\delta}_n^{(J)}\}$. Comment on the histograms in terms of the speed of convergence.

[2]If repeating $J = 10000$ times is computationally hard for your computer, then repeat $J = 5000$ or even $J = 1000$ times. There should not be a drastic change in simulation results.

Table 1: Bias, Variance, and MSE of $\hat{\alpha}_n$ and $\hat{\delta}_n$

	B_n			V_n			MSE_n		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\hat{\alpha}_n$									
$\hat{\delta}_n$									

(b) A standard way of summarizing simulation results is to report the *bias*, *variance*, and *mean squared error (MSE)* of each estimator. Consider $\hat{\alpha}_n$ for example. Compute $\bar{\alpha}_n = (1/J)\sum_{j=1}^{J} \hat{\alpha}_n^{(j)}$. Bias is defined as $B_n = (1/J)\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \alpha_0)$. Variance is defined as $V_n = (1/J)\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n)^2$. MSE is defined as $MSE_n = (1/J)\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \alpha_0)^2$. **Prove a well-known generic equality**:

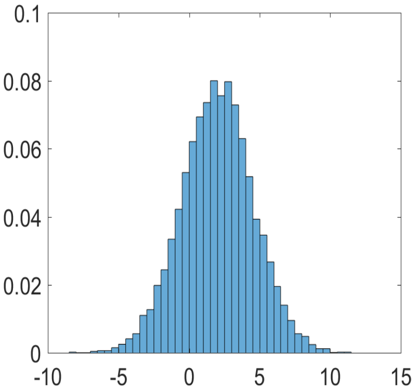
$$MSE_n = B_n^2 + V_n, \tag{16}$$

which is called the *bias-variance formula*. (Remark: A desirable estimator is supposed to have small bias and small variance. Hence, the smaller MSE generally indicates a better performance of the estimator.)

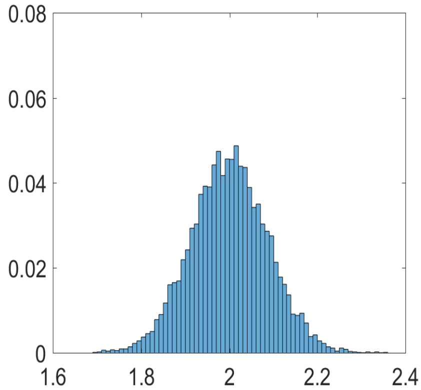
(c) Compute the bias, variance, and MSE for each of $\{\hat{\alpha}_n^{(1)}, \ldots, \hat{\alpha}_n^{(J)}\}$ and $\{\hat{\delta}_n^{(1)}, \ldots, \hat{\delta}_n^{(J)}\}$ (i.e. fill in Table 1). Comment on the results in terms of the speed of convergence.

Solution-4: (a) See Figure 1 for the histograms. For each sample size n, the histogram of $\hat{\delta}_n$ has a much tighter range than the histogram of $\hat{\alpha}_n$. As n increases, the histogram of $\hat{\delta}_n$ collapses to the true value of $\delta_0 = 2$ much faster than the histogram of $\hat{\alpha}_n$ collapses to $\alpha_0 = 2$. These results confirm the key insight from Problem-3 that $\hat{\delta}_n$ has a faster speed of convergence than $\hat{\alpha}_n$.

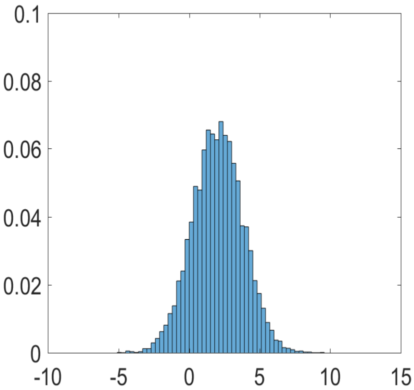
Figure 1: Histograms of $\hat{\alpha}_n$ and $\hat{\delta}_n$



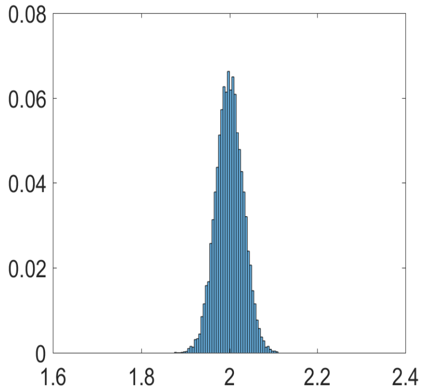
(a) $\hat{\alpha}_n$ $(n = 50)$



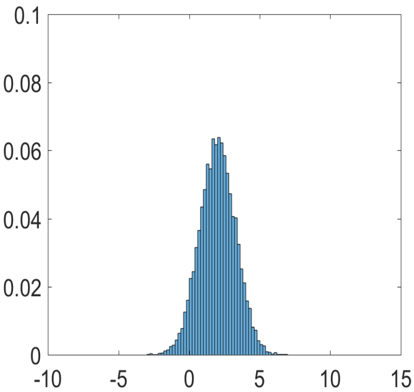
(b) $\hat{\delta}_n$ $(n = 50)$



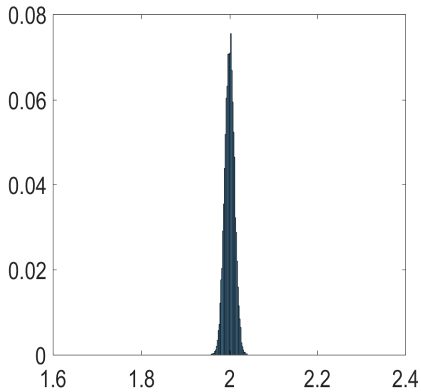
(c) $\hat{\alpha}_n$ $(n = 100)$



(d) $\hat{\delta}_n$ $(n = 100)$



(e) $\hat{\alpha}_n$ $(n = 200)$



(f) $\hat{\delta}_n$ $(n = 200)$

Table 2: Bias, Variance, and MSE of $\hat{\alpha}_n$ and $\hat{\delta}_n$

	B_n			V_n			MSE_n		
	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$
$\hat{\alpha}_n$	0.049	0.019	0.005	6.556	3.249	1.640	6.559	3.250	1.640
$\hat{\delta}_n$	6.3×10^{-4}	9.5×10^{-4}	2.1×10^{-5}	0.008	9.5×10^{-4}	1.2×10^{-4}	0.008	9.5×10^{-4}	1.2×10^{-4}

(b) Note that $B_n = (1/J)\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \alpha_0) = \bar{\alpha}_n - \alpha_0$. We have that

$$
\begin{aligned}
MSE_n &= \frac{1}{J}\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \alpha_0)^2 = \frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n) + (\bar{\alpha}_n - \alpha_0)\right]^2 \\
&= \frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n) + B_n\right]^2 \\
&= \frac{1}{J}\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n)^2 + \frac{2}{J}\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n)B_n + \frac{1}{J}\sum_{j=1}^{J}B_n^2 \\
&= V_n + 2B_n \times \frac{1}{J}\sum_{j=1}^{J}(\hat{\alpha}_n^{(j)} - \bar{\alpha}_n) + B_n^2 \\
&= V_n + 2B_n(\bar{\alpha}_n - \bar{\alpha}_n) + B_n^2 = V_n + B_n^2.
\end{aligned}
$$

(c) See Table 2. MSE of $\hat{\alpha}_n$ is 6.559, 3.250, and 1.640 when sample size n is 50, 100, and 200, respectively. The MSE roughly halves as sample size doubles. MSE of $\hat{\delta}_n$ is 0.008, 9.5×10^{-4}, and 1.2×10^{-4} when sample size n is 50, 100, and 200, respectively. The MSE gets roughly 8 times smaller as sample size doubles. It is thus evident that $\hat{\delta}_n$ has a faster rate of convergence than $\hat{\alpha}_n$.

Problem-5: This problem serves as an empirical illustation of detrending. Visit Federal Reserve Economic Data (`https://fred.stlouisfed.org/`). Download "Consumer Price Index for All Urban Consumers: All Items, Index 1982-1984=100, Monthly, Seasonally Adjusted" from January 1991 through December 1995 ($n = 60$ months). Call this series $\{y_t\}$.

Detrend it by $y_t = \alpha + \delta \times t + u_t$ with $t \in \{1, \dots, 60\}$.

(a) Report OLS estimates $\hat{\alpha}_n$ and $\hat{\delta}_n$.

(b) Draw a time series plot of actual series y_t, fitted series $\hat{y}_t = \hat{\alpha}_n + \hat{\delta}_n \times t$, and detrended series $\hat{u}_t = y_t - \hat{y}_t$ in one figure. (Instruction: Use the left y-axis for y_t and $\hat{y}_t$, and use the right y-axis for $\hat{u}_t$.)

(c) Does $\{\hat{u}_t\}$ seem to be stationary or nonstationary? Explain.

Solution-5: (a) $\hat{\alpha}_n = 134.1$ and $\hat{\delta}_n = 0.336$.

(b) See Figure 2.

(c) It seems to be a marginal case between stationarity and nonstationarity. We at least do not see a clearly explosive path, but $\{\hat{u}_t\}$ owns some strong persistence. We thus need a formal test for judging whether it is stationary or not.

Figure 2: Empirical Results