STATISTICS IN TRANSITION new series, December 2024
Vol. 25, No. 4, pp. xx–xx, https://doi.org/10.59170/stattrans-2024–xxx
Received – xx.xx.20xx; accepted – xx.xx.20xx

On some statistical properties of a stationary Gaussian process in the presence of measurement errors

Kuntal Bera[1], **M. Z. Anis**[2]

Abstract

Process outputs of many production processes like chemical, food processing and pharmaceutical industry follow a stationary Gaussian process. Some amount of measurement error always present in the measured data due to inaccurate measuring processes. Throughout this paper, we discuss some statistical properties like the mean and variance of a stationary Gaussian process when observed data are affected by measurement errors. As a special case, we discuss a stationary autoregressive process of order one with Gaussian white noise where measurement error follows an independent Gaussian distribution.

Key words: autoregressive process, central moments, measurement errors, white noise.

1. Introduction

Measurement error is a general problem while collecting data and hence the usage of such data may lead to improper inference. The variable of interest, say X, cannot be measured accurately in the presence of measurement error. Koutsoyiannis (1977) presents interesting examples of measurement errors in economics. The presence of measurement error has a profound influence in almost every area dealing with the measurement of the sample. Maleki et al. (2017) pointed out that in spite of refined and sophisticated measuring devices, real-life data are contaminated with measurement errors. Hence, these measurement errors need to be taken into account while monitoring items. There is a large body of literature dealing with the effect of measurement error in many areas including statistical process control (SPC), economics, medical studies, environmental studies, agricultural studies and others. See, for example, Carroll (1998), Linna and Woodall (2001), Noor-ul Amin et al. (2022), Schennach (2016), Blackwell et al. (2017), Abay et al. (2023) and the references therein. Wu (2011) cautioned that if measurement error is ignored, it may lead to unreliable decisions for the process under study. Attention should first be paid to the measurement system to ascertain if the variability significantly increases due to the presence of measurement error in the data.

Autocorrelation is an inherent property of many processes, see, for example, Shumway and Stoffer (2017). The interval between process observations is decreasing due to the

[1]SQC & OR Unit, Indian Statistical Institute, Kolkata-700108, India. E-mail: bera.kuntal12345@gmail.com. ORCID: https://orcid.org/0000-0003-1217-5708.
[2]SQC & OR Unit, Indian Statistical Institute, Kolkata-700108, India. E-mail: zafar@isical.ac.in. ORCID: https://orcid.org/0000-0001-5546-0723.
© K. Bera, M. Z. Anis. Article available under the CC BY-SA 4.0 licence

rampant use of online data acquisition systems, leading to positive autocorrelation as noted by Runger and Willemain (1995). Such trends are more pronounced in the process and chemical industries, the practice of measuring every part produced induces positive autocorrelation even in discrete parts measurement. Statistical tests (e.g. the Durbin-Watson test and the Bartlett test) can be used to detect the presence of first-order and higher-order autocorrelation. Zhang (1998) has estimated the variance and expected value of the sample mean and the sample variance for a stationary Gaussian process.

In many industrial production processes it is found that quite a few of the quality characteristics of output products follow stationary Gaussian process, see Box et al. (2015). To control the production process economically, knowledge of the mean and variance of the respective quality characteristics is necessary. For instance, the knowledge of the mean and variance of a process is required for many statistical process control techniques like estimation of control limits, estimation of process capability, estimation of percentage of conforming quality, etc.

In this work, we discuss some statistical properties of such autocorrelated processes when the observed data is contaminated with measurement error and thereby extend the work of Zhang (1998). The paper is organized as follows. In Section 2, we discuss the statistical properties of a stationary Gaussian process. These results are extended to account for measurement errors in Section 3. The impact of measurement errors in estimating the mean and variance of a stationary $AR(1)$ process is shown graphically as a special case. Some simulation studies are reported and an industrial example is reported in Section 4. An industrial application is mentioned in Section 5 and Section 6 concludes the paper.

2. Statistical Properties of a Stationary Gaussian Process

Let $\{X_t, t \in \mathbb{Z}\}$ be a discrete-time stationary Gaussian process. A time series $\{X_t\}$ is said to be a stationary process if $Var(X_t) < \infty$, expectation $E(X_t)$ is independent of time t and auto-covariance function $Cov(X_{t+h}, X_t)$ depends only on difference between the time points; i.e., h and is independent of t. For a stationary time series X_t, let

$$E(X_t) = \mu_X$$

and

$$Cov(X_{t+h}, X_t) = \gamma_X(h).$$

A time series $\{X_t\}$ is said to be a Gaussian process if the joint distribution of any finite n number of random variables $\{X_1, X_2, X_3, \dots, X_n\}$ from the process follows a multivariate normal distribution; see Kotz et al. (2019) for details about properties of multivariate normal distribution. A process that is stationary and Gaussian simultaneously is said to be a stationary Gaussian process, see Brockwell and Davis (2002) for details. A stationary Gaussian process is also strictly stationary. A time series is said to be *strictly stationary* if the probabilistic behaviour of every finite collection of values $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ is identical with the time shifted values $\{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}\}$.

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of n consecutive observations from a stationary Gaussian process. Then, the sample mean and the sample variance of the process are, respectively, given by $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and $S_X^2 = \frac{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]}{n-1}$. Zhang (1998) found the expected value and variance of $\bar{X}$ and S_X^2. We shall stick to the same notation as used by Zhang (1998); but for completeness, we shall define the functions that will be used subsequently.

Let

$$\rho_i = \rho_X(i) = \frac{\gamma_X(i)}{\gamma_X(0)}$$

for $i = 1, 2, 3, \ldots, n$ be the autocorrelation of X_t at lag i. Define,

$$f(n, \rho_i) = 1 - \frac{2}{n(n-1)} \sum_{i=1}^{n-1} (n-i)\,\rho_i, \tag{1}$$

$$F(n, \rho_i) = n + 2\sum_{i=1}^{n-1} (n-i)\,\rho_i{}^2 + \frac{1}{n^2}\left[n + 2\sum_{i=1}^{n-1} (n-i)\rho_i\right]^2 - \frac{2}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{n-i} (n-i-j)\,\rho_i\rho_j \tag{2}$$

and

$$g(n, \rho_i) = 1 + \frac{2}{n}\sum_{i=1}^{n-1} (n-i)\rho_i. \tag{3}$$

Observe that when the process $\{X_t\}$ is identically and independently normally distributed, then $\rho_i = 0$ for $i \geq 1$. In this case $f(n, \rho_i) = 1$, $g(n, \rho_i) = 1$ and $F(n, \rho_i) = (n-1)$.

Since $\{X_t\}$ is a Gaussian process, therefore $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2 g(n,\rho_i)}{n}\right)$. Hence, the $r-$th order central moment of $\bar{X}$ is

$$\mathbb{E}\left(\bar{X} - \mu_X\right)^r = \begin{cases} 0 & \text{when } r \text{ is an odd integer} \\ 1.3.5\ldots(2k-1)\left[\frac{\sigma_X^2 g(n,\rho_i)}{n}\right]^{2k} & \text{when } r = 2k \text{ for } k = 1, 2, 3, \ldots. \end{cases} \tag{4}$$

Here, we discuss the particular case of an $AR(1)$ process. $AR(1)$ process is very popular and often occurs in many chemical and process industries for modeling autocorrelation structures. Note that the $AR(1)$ process is a special case of stationary Gaussian process. An $AR(1)$ process with mean μ is defined as

$$X_t - \mu = \phi\,(X_{t-1} - \mu) + a_t, \qquad |\phi| < 1. \tag{5}$$

When a_t is a Gaussian white noise, then $\{X_t\}$ will be a stationary Gaussian process. Suppose $a_t \sim IIDN(0, \sigma_a^2)$. For an $AR(1)$ process defined by equation (5), $\rho_i = \phi^i$. So, the expected value and variance of the sample mean and the sample variance, in this case, is

$$\mathbb{E}\,(\bar{X}) = \mu_X,$$

$$Var\,(\bar{X}) = \frac{\sigma_X^2}{n} g(n,\phi),$$
$$\mathbb{E}\left(S_X^2\right) = \sigma_X^2 f(n,\phi),$$

and

$$Var\left(S_X^2\right) = \frac{2\sigma_X^4}{(n-1)^2} F(n,\phi)$$

where

$$\sigma_X^2 = \frac{\sigma_a^2}{(1-\phi^2)}, \tag{6}$$

$$f(n,\phi) = 1 - \frac{2}{n(n-1)} \frac{\phi\,[n-1-n\phi+\phi^n]}{(1-\phi)^2}, \tag{7}$$

$$F(n,\phi) = n + 2\sum_{i=1}^{n-1} (n-i)\,\phi^{2i} + \frac{1}{n^2}\left[n + 2\sum_{i=1}^{n-1} (n-i)\phi^i\right]^2 - \frac{2}{n}\sum_{i=0}^{n-1}\sum_{j=0}^{n-i} (n-i-j)\,\phi^{i+j}, \tag{8}$$

and

$$g(n,\phi) = 1 + \frac{2}{n} \frac{\phi\,[n-1-n\phi+\phi^n]}{(1-\phi)^2}. \tag{9}$$

Thus, the sample mean is an unbiased estimator of the population mean while the sample variance is a biased estimator of population variance. The variance of the sample mean can be expressed as

$$Var(\bar{X}) = \frac{\sigma_a^2 g(n,\phi)}{n(1-\phi^2)}$$
$$= \sigma_a^2 R_1(n,\phi)$$

where

$$R_1(n,\phi) = \frac{g(n,\phi)}{n(1-\phi^2)}.$$

The bias and mean square error of the sample variance are given by

$$Bias(S_X^2) = \sigma_X^2\,[f(n,\phi) - 1]$$
$$= \frac{\sigma_a^2[f(n,\phi)-1]}{(1-\phi^2)}$$
$$= \sigma_a^2 R_2(n,\phi)$$

$$
\begin{aligned}
MSE(S_X^2) &= Var\left(S_X^2\right) + \left\{Bias(S_X^2)\right\}^2 \\
&= \frac{2\sigma_a^4}{(1-\phi^2)^2}\left[\frac{F(n,\phi)}{(n-1)^2} + \frac{\{f(n,\phi)-1\}^2}{2}\right] \\
&= \sigma_a^4 R_3(n,\phi)
\end{aligned}
$$

where

$$R_2(n,\phi) = \frac{[f(n,\phi)-1]}{(1-\phi^2)},$$

$$R_3(n,\phi) = \frac{2}{(1-\phi^2)^2}\left[\frac{F(n,\phi)}{(n-1)^2} + \frac{\{f(n,\phi)-1\}^2}{2}\right].$$

Similarly, the $r-$th order central moment of $\bar{X}$ in this special case will be,

$$\mathbb{E}\left(\bar{X}-\mu_X\right)^r = \begin{cases} 0 & \text{when } r \text{ is an odd integer} \\ 1.3.5\ldots(2k-1)\left[\frac{\sigma_X^2 g(n,\phi)}{n}\right]^{2k} & \text{when } r = 2k \text{ for } k = 1,2,3,\ldots. \end{cases} \tag{10}$$

We graphically present $R_1(n,\phi)$, $R_2(n,\phi)$ and $R_3(n,\phi)$ to show the effect of autocorrelation and the sample size on estimating mean and variance when process observations are autocorrelated. Note that when sample observations are independent, then $\phi = 0$ and in this case $R_1(n,\phi)$, $R_2(n,\phi)$, $R_3(n,\phi)$ will become only the function of the sample size n.

From Figure 1 it is clear that the variance of the sample mean increases as the autocorrelation level increases. For small sample size ($n < 50$), this variance is significantly large when autocorrelation is high ($\phi \geq 0.50$). The variance of the sample mean decreases with the increase of the sample size as expected. From the graph of $R_2(n,\phi)$ in Figure 2 we notice that the sample variance is a biased estimate of population variance in the presence of autocorrelation. Note that sample variance is an unbiased estimate of population variance when sample observations are independent. Sample variance is underestimated in the presence of autocorrelation. The bias of the sample variance is significantly large when the sample size is small and autocorrelation is high. From the graph of $R_3(n,\phi)$ in Figure 3 we notice that the MSE of the sample variance increases as autocorrelation increases. This MSE is significantly large when the sample size is small. This happens because uncertainty within the sample increases due to an increase of autocorrelation and as a result, in such cases, a sample of small size cannot properly estimate the mean and variance of the population. Therefore, to avoid estimation error due to autocorrelation, a relatively large sample size is required for estimating mean and variance.

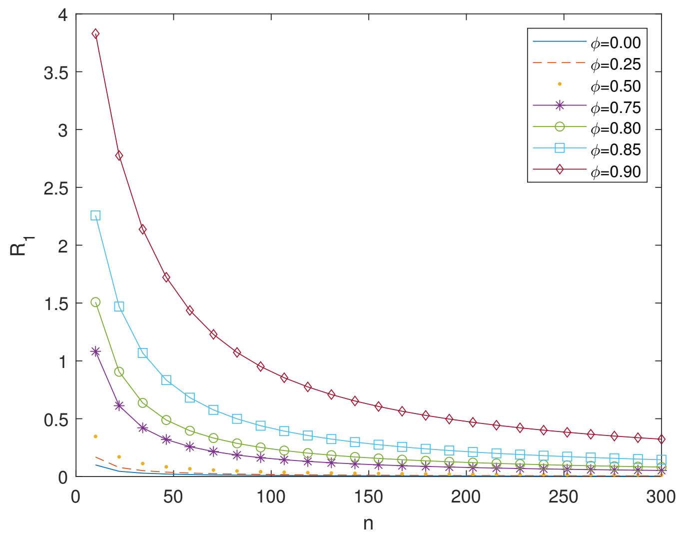


Figure 1: Graph of $R_1(n,\phi)$ corresponding to different values of ϕ.

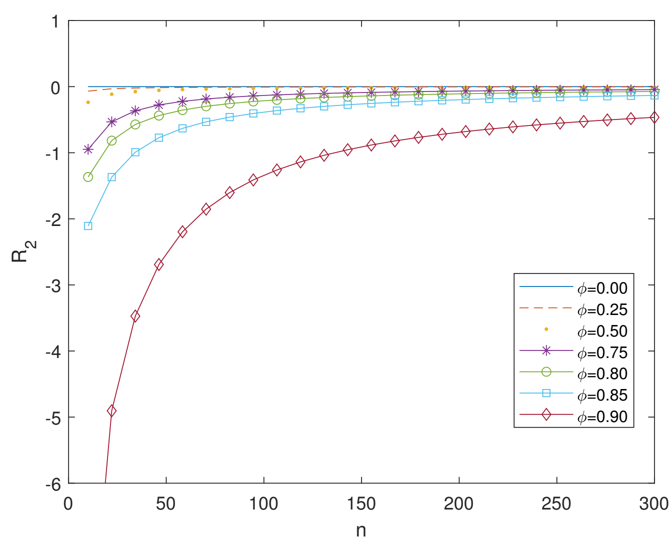


Figure 2: Graph of $R_2(n,\phi)$ corresponding to different values of ϕ.

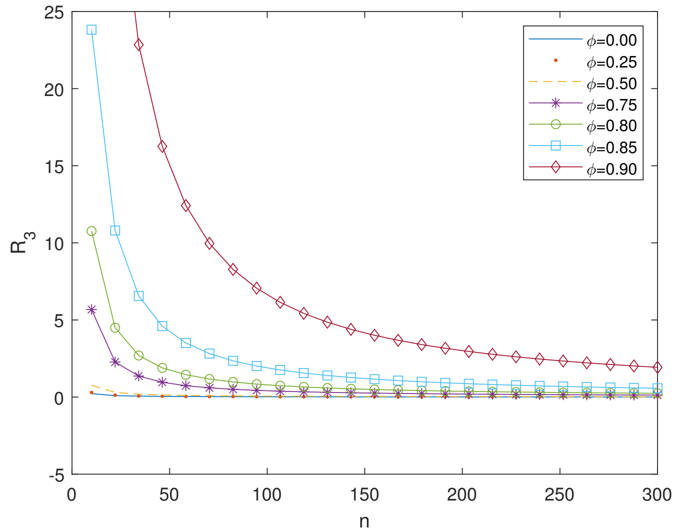


Figure 3: Graph of $R_3(n,\phi)$ corresponding to different values of ϕ.

3. Statistical Properties of a Stationary Gaussian Process in the Presence of Measurement Errors

In practical situations, true values of process outputs are often unobservable as they are contaminated by measurement errors. Instead of the true process $\{X_t\}$, we observe the process $\{Y_t\}$, where Y_t is defined by

$$Y_t = X_t + E_t. \tag{11}$$

Here, E_t is a random measurement error variable. Assume that $E_t \sim N(0,\sigma_E^2)$; X_t and E_t are stochastically independent. Let $\{Y_1,Y_2,\ldots,Y_n\}$ be a random sample of size n from the observable process $\{Y_t\}$. Thus, the sample mean and the sample variance, using the observable data, are given by $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$ and $S_Y^2 = \frac{\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]}{n-1}$ respectively.

3.1. Statistical Analysis of Sample Mean

It is easy to see that

$$\mathbb{E}(\bar{Y}) = \mu_X; \tag{12}$$

and

$$Var(\bar{Y}) = \frac{\sigma_X^2}{n} g(n,\rho_i) + \frac{\sigma_E^2}{n}. \tag{13}$$

As $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2 g(n,\rho_i)}{n}\right)$; $\bar{E} \sim N\left(0, \frac{\sigma_E^2}{n}\right)$ and $\bar{Y}$ is the sum of two normal variables, it follows that $\bar{Y} \sim N\left(\mu_X, \frac{\left[\sigma_X^2 g(n,\rho_i)+\sigma_E^2\right]}{n}\right)$. Hence, the $r-$th order central moment of $\bar{Y}$ is given by

$$\mathbb{E}\left(\bar{Y}-\mu_X\right)^r = \begin{cases} 0 & \text{when } r \text{ is an odd integer} \\ 1.3.5\ldots(2k-1)\left[\frac{\sigma_X^2 g(n,\rho_i)+\sigma_E^2}{n}\right]^{2k} & \text{when } r = 2k \text{ for } k = 1,2,3,\ldots. \end{cases} \tag{14}$$

In the particular case of an $AR(1)$ process defined by equation (5), we have

$$\begin{aligned} Var\left(\bar{Y}\right) &= \frac{\sigma_X^2}{n} g(n,\phi) + \frac{\sigma_E^2}{n} \\ &= \frac{\sigma_X^2}{n}\left[g(n,\phi) + (1-\phi^2){\tau_a}^2\right] \\ &= \sigma_a^2 R_1^e(n,\phi,\tau_a) \end{aligned}$$

where

$$R_1^e(n,\phi,\tau_a) = \frac{\left[g(n,\phi) + (1-\phi^2){\tau_a}^2\right]}{n(1-\phi^2)}. \tag{15}$$

Here, σ_a^2 is the variance of Gaussian white noise a_t and $R_1^e(n,\phi,\tau_a)$ is defined by the equation (15). Here, τ_a is the ratio between the standard deviation of measurement error E_t and the standard deviation of the Gaussian white noise a_t and is defined by

$$\tau_a = \frac{\sigma_E}{\sigma_a}. \tag{16}$$

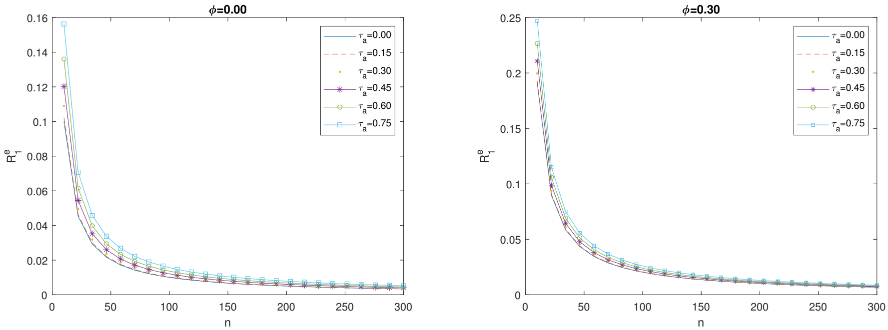
To see the combined effect of measurement error and autocorrelation on the estimation of mean, we graphically analyze the function $R_1^e(n,\phi,\tau_a)$. Note the $R_1^e(n,\phi,\tau_a)$ is a function of autocorrelation parameter ϕ, degree of error contamination τ_a and the sample size n. In case sample observations are free of measurement error then $\tau_a = 0$ and when independent then $\phi = 0$.

It is clear from Figure 4 and Figure 5 that the variance of the sample mean has some increment due to the presence of measurement error. But this increment is not significant. Hence, measurement error does not affect seriously the sample mean.

3.2. Statistical Analysis of Sample Variance

It is easy to see that

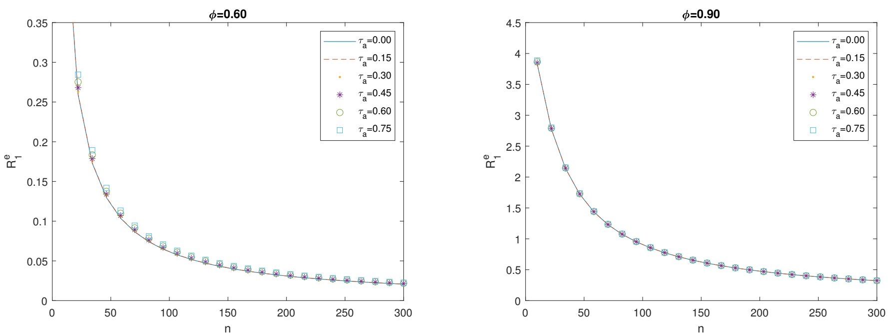
$$S_Y^2 = S_X^2 + \frac{2}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(E_i-\bar{E}) + S_E^2;$$



(a) Graph of $R_1^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.00$.

(b) Graph of $R_1^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.30$.

Figure 4:



(a) Graph of $R_1^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.60$.

(b) Graph of $R_1^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.90$.

Figure 5:

and on taking expectations we get

$$\mathbb{E}(S_Y^2) = \sigma_X^2 f(n,\rho_i) + \sigma_E^2; \tag{17}$$

If $\{X_t\}$ is an AR(1) process defined by equation (5), then

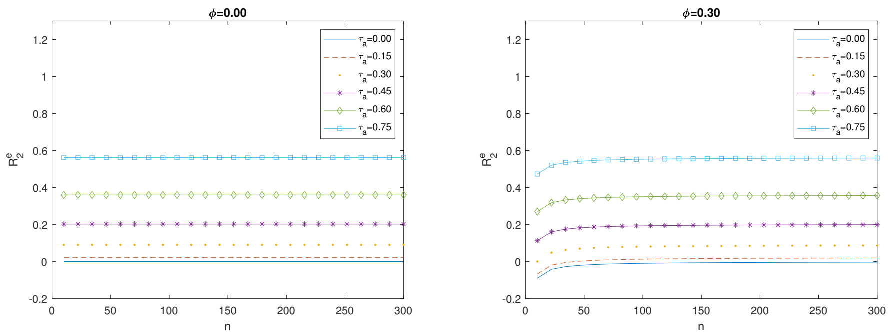
$$\mathbb{E}(S_Y^2) = \sigma_X^2 \left[f(n,\phi) + (1-\phi^2)\tau_a^2 \right].$$

In this case, sample variance is a biased estimator of population variance. The bias of the estimator with respect to the true process variance is given by

$$\begin{aligned} Bias(S_Y^2) &= E\left(S_Y^2\right) - \sigma_X^2 \\ &= \sigma_X^2 \left[\{f(n,\phi) - 1\} + (1-\phi^2)\tau_a^2 \right] \\ &= \sigma_a^2 R_2^e(n,\phi,\tau_a) \end{aligned}$$

where

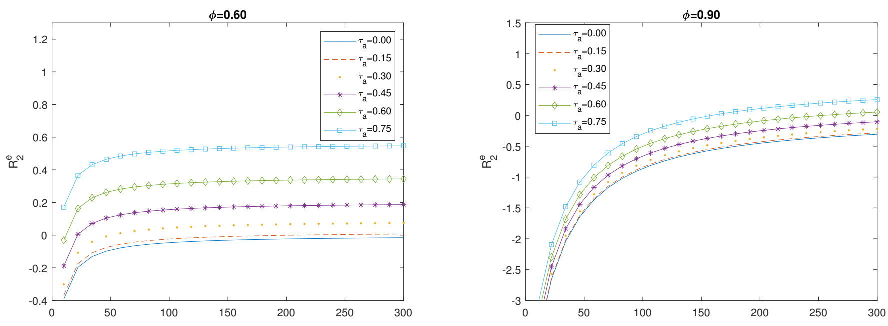
$$R_2^e(n,\phi,\tau_a) = \frac{\left[f(n,\phi) - 1 + (1-\phi^2)\tau_a{}^2\right]}{(1-\phi^2)}. \tag{18}$$



(a) Graph of $R_2^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.00$.

(b) Graph of $R_2^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.30$.

Figure 6:



(a) Graph of $R_2^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.60$.

(b) Graph of $R_2^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.90$.

Figure 7:

We have observed from Figure 2 of Section 2 that the sample variance is underestimated in the presence of autocorrelation. Also, we have noticed that the sample variance is unbiased when the sample is independent. From Figure 6(a) we notice that the sample variance is overestimated in the presence of the measurement error when sample observations are independent. However, a mixed effect is observed in the presence of both autocorrelation and measurement error. Autocorrelation tries to underestimate the sample variance while measurement error attempts to overestimate it. As the effect of underestimation is large when autocorrelation is present and the sample size is small, a relatively small value of $R_2^e(n,\phi,\tau_a)$ is visible for a small sample size.

The expression for the variance of the sample variance is given in the equation (19). Detailed calculation is provided in the Appendix.

$$Var(S_Y^2) = \frac{2\sigma_X^4}{(n-1)^2}\left\{F(n,\rho_i) + (n-1)\frac{\sigma_E^4}{\sigma_X^4} + 2(n-1)f(n,\rho_i)\frac{\sigma_E^2}{\sigma_X^2}\right\}. \tag{19}$$

Now, if $\{X_t\}$ is an $AR(1)$ process defined by equation (5), then we can write the expression (19) as:

$$Var(S_Y^2) = \frac{2\sigma_X^4}{(n-1)^2}\left\{F(n,\phi) + (n-1)(1-\phi^2)^2\tau_a^4 + 2(n-1)(1-\phi^2)\tau_a^2 f(n,\phi)\right\}.$$

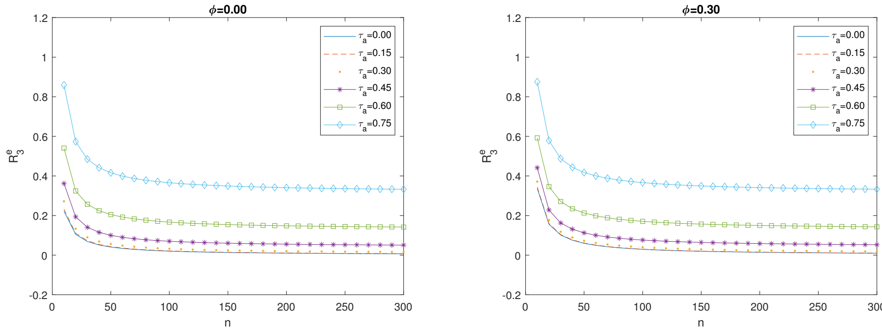
The mean square error of the estimator of the sample variance about the true process variance, in this case, is

$$\begin{aligned} MSE(S_Y^2) = & \frac{2\sigma_X^4}{(n-1)}\left\{\frac{\left[F(n,\phi) + (n-1)(1-\phi^2)^2\tau_a^4 + 2(n-1)(1-\phi^2)\tau_a^2 f(n,\phi)\right]}{(n-1)}\right. \\ & \left. + \frac{(n-1)[\{f(n,\phi)-1\} + (1-\phi^2){\tau_a}^2]^2}{2}\right\} \\ = & \sigma_a^4 R_3^e(n,\phi,\tau_a) \end{aligned} \tag{20}$$

where

$$\begin{aligned} R_3^e(n,\phi,\tau_a) = & \frac{2}{(1-\phi^2)^2}\left\{\frac{\left[F(n,\phi) + (n-1)(1-\phi^2)^2\tau_a^4 + 2(n-1)(1-\phi^2)\tau_a^2 f(n,\phi)\right]}{(n-1)^2}\right. \\ & \left. + \frac{[\{f(n,\phi)-1\} + (1-\phi^2){\tau_a}^2]^2}{2}\right\}. \end{aligned} \tag{21}$$

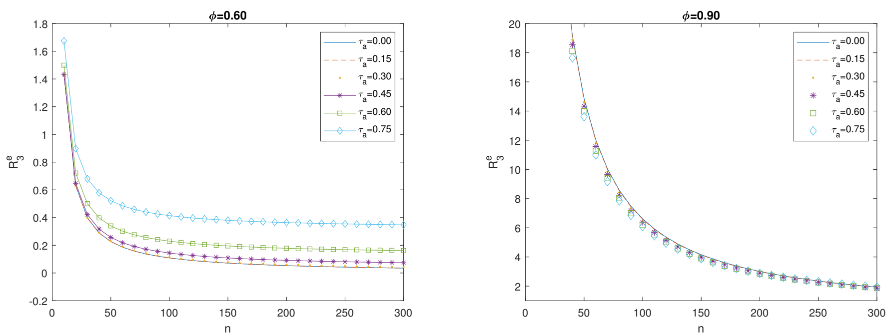
The different behaviour of $MSE(S_Y^2)$ for different values of ϕ, τ_a and n is observable from Figure 8 and Figure 9. We can see from these figures that as measurement error increases, the MSE value also increases. But for highly autocorrelated data ($\phi = 0.90$), as can be seen from Figure 9(b), the value of MSE relatively decreases as measurement error increases.



(a) Graph of $R_3^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.00$.

(b) Graph of $R_3^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.30$.

Figure 8:



(a) Graph of $R_3^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.60$.

(b) Graph of $R_3^e(n,\phi,\tau_a)$ corresponding to different values of τ_a for $\phi = 0.90$.

Figure 9:

4. Simulation

We carried out a simulation exercise to study the behavior of the mean and variance of the sample mean and the sample variance. We simulate 5000 random samples of size $50(25)200$ respectively from each of an IID normal process, a stationary $AR(1)$ process and a stationary $AR(1)$ process with random measurement errors separately. In each of these cases, we compare the sample estimated value of the mean and variance of the estimators with the corresponding theoretical values. Here, `hat` $(\hat{\ })$ indicates the sample estimated value.

4.1. Case of an IID normal Process

Here, we simulate a model of iid random normal variables $X_t \sim N(\mu, \sigma^2)$ where $\mu = 5$ and $\sigma^2 = 4$.

Table 1: Results based on iid normal model X_t.

n	$E(\bar{X})$	$\hat{E}(\bar{X})$	$Var(\bar{X})$	$\hat{Var}(\bar{X})$	$E(S_X^2)$	$\hat{E}(S_X^2)$	$Var(S_X^2)$	$\hat{Var}(S_X^2)$
50	5	5.0068	0.0800	0.0778	4	3.9759	0.6531	0.6700
75	5	5.0058	0.0533	0.0545	4	3.9784	0.4324	0.4337
100	5	5.0056	0.0400	0.0404	4	3.9824	0.3232	0.3235
125	5	5.0045	0.0320	0.0317	4	3.9865	0.2581	0.2580
150	5	5.0038	0.0267	0.0264	4	3.9898	0.2148	0.2145
175	5	4.9967	0.0229	0.0227	4	4.0015	0.1839	0.1833
200	5	4.9968	0.0200	0.0201	4	4.0008	0.1608	0.1598

4.2. Case of a stationary $AR(1)$ process

Here, we simulate a model from a stationary $AR(1)$ model defined by equation (5) where $\mu = 5$, $\phi = 0.5$ and white noise $a_t \sim N(0,1)$.

Table 2: Results based on an $AR(1)$ model X_t.

n	$E(\bar{X})$	$\hat{E}(\bar{X})$	$Var(\bar{X})$	$\hat{Var}(\bar{X})$	$E(S_X^2)$	$\hat{E}(S_X^2)$	$Var(S_X^2)$	$\hat{Var}(S_X^2)$
50	5	4.9981	0.0779	0.0758	1.2811	1.2729	0.1233	0.1072
75	5	4.9980	0.0524	0.0510	1.2983	1.2939	0.0812	0.0746
100	5	4.9989	0.0395	0.0390	1.3069	1.3038	0.0605	0.0575
125	5	4.9992	0.0317	0.0313	1.3122	1.3116	0.0482	0.0461
150	5	4.9995	0.0264	0.0261	1.3157	1.3139	0.0400	0.0386
175	5	4.9997	0.0227	0.0225	1.3182	1.3163	0.0343	0.0330
200	5	4.9997	0.0199	0.0197	1.3201	1.3201	0.0299	0.0294

4.3. Case of an $AR(1)$ process in the presence of measurement errors

Here, we simulate a stationary $AR(1)$ process in the presence of measurement errors. The model is defined by equation (11). In this model X_t is same as defined in the previous cases and $E_t \sim iid\ N(0, \sigma_E^2)$ where $\sigma_E^2 = 4$.

Table 3: Results based on an $AR(1)$ model in the presence of measurement error, Y_t.

n	$E(\bar{Y})$	$\hat{E}(\bar{Y})$	$Var(\bar{Y})$	$\hat{Var}(\bar{Y})$	$E(S_Y^2)$	$\hat{E}(S_Y^2)$	$Var(S_Y^2)$	$\hat{Var}(S_Y^2)$
50	5	4.9990	0.1579	0.1545	5.2811	5.2856	1.1947	1.1494
75	5	4.9999	0.1057	0.1039	5.2983	5.2996	0.7943	0.7836
100	5	5.0001	0.0795	0.0782	5.3069	5.2945	0.5949	0.5851
125	5	4.9986	0.0637	0.0630	5.3122	5.3116	0.4756	0.4718
150	5	5.0020	0.0531	0.0526	5.3157	5.2947	0.3961	0.3938
175	5	5.0027	0.0455	0.0452	5.3182	5.3093	0.3394	0.3367
200	5	5.0017	0.0399	0.0395	5.3201	5.3159	0.2969	0.2944

5. An industrial application

As an application, we will discuss the case of estimation of the control limits of the mean chart in statistical process monitoring. Let, the true process $\{X_t\}$ follow a stationary Gaussian process and the observable process $\{Y_t\}$ modeled by equation (11). Then, the upper and lower control limit for the process sample mean based on the observable sample is given by,

$$\mathrm{UCL/LCL} = \mu \pm \mathrm{K}\sqrt{Var(\bar{Y})} \tag{22}$$

where $\mathrm{K} > 0$ is a real positive constant, often taken as $\mathrm{K} = 3$ and $Var(\bar{Y})$ is given by equation (13). In particular, when the process is an AR(1) process

$$\mathrm{UCL/LCL} = \mu \pm \mathrm{K}\sigma_a\sqrt{R_1^e(n, \phi, \tau_a)} \tag{23}$$

where $R_1^e(n, \phi, \tau_a)$ is given by equation (15). As measurement errors increase, the value of $R_1^e(n, \phi, \tau_a)$ also increases, widening the control limits. The probability P of detecting the mean shift from in-control mean μ_0 to out-of-control mean μ_1 is equal to

$$\mathrm{P} = \Phi\left(-K - \delta/\sqrt{R_1^e(n, \phi, \tau_a)}\right) + \Phi\left(-K + \delta/\sqrt{R_1^e(n, \phi, \tau_a)}\right) \tag{24}$$

and the average run length (ARL) of the control chart is equal to

$$\mathrm{ARL} = 1/\mathrm{P} \tag{25}$$

where $\delta = |(\mu_0 - \mu_1)/\sigma_a|$ measures the mean shift. From Figure 4 and Figure 5 it can be noticed that as autocorrelation and measurement error increases, $R_1^e(n, \phi, \tau_a)$ increases. As a result, the power of the control chart decreases and the ARL increases. It can be visualized from Figure 10. Therefore, the performance of the control chart decreases in the presence of both autocorrelation and measurement errors. Some techniques are used to reduce the autocorrelation and measurement error. For example, serial autocorrelation can be reduced by using the s-skip strategy and measurement error can be reduced by taking several measures $(m \geq 1)$ of each observed item or by improving the gauge performance, see, for example, Costa and Castagliola (2011), Shongwe et al. (2021), Shongwe et al. (2019), Shongwe and Malela-Majika (2021), Garza-Venegas et al. (2018).

6. Conclusions

Many industrial process data are autocorrelated and at the same time, the existence of measurement errors due to inadequate measuring devices is common. To estimate some inferential results based on the collected sample, sometimes the mean and variance of the sample are required to be estimated. Hence, statistical properties of the sample mean and the sample variance are required to ascertain how reliable the estimated values are.

Our discussion in Section 2 indicates that the variance of the sample mean and the sample variance increases as autocorrelation increases. Also, the sample variance is biased

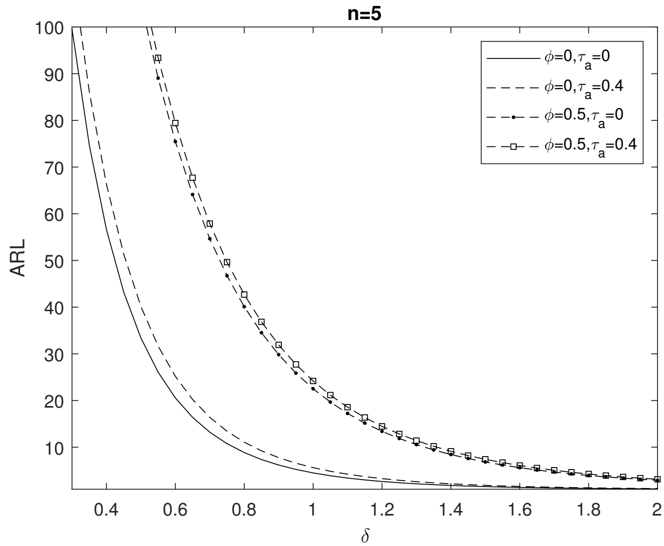


Figure 10: The effect of autocorrelation and measurement error on the ARL of the mean control chart, for $\mathrm{K} = 3$ and $n = 5$.

and underestimated in the presence of autocorrelation. It has been shown that the variance of the sample mean and the sample variance is significantly large when autocorrelation is high and the sample size is small. Therefore, a relatively large sample size is recommended for estimating mean and variance depending on the level of autocorrelation.

Similarly, our analysis in Section 3 shows that the combined effect of autocorrelation and measurement error is found in estimating mean and variance for autocorrelated samples contaminated by measurement errors. Measurement error increases the variance of the sample variance. Sample variance is underestimated in the presence of autocorrelation and overestimated in the presence of measurement error. Therefore, one should be very careful when taking measurements and measuring devices should also be good enough to provide adequate confidence. In Section 4 we compare theoretical values of mean and variance of the estimator of the sample mean and the sample variance with the sample estimated value on the basis of simulated data. These results show that the theoretical values based on our obtained results and the estimated sample values are reasonably close. Here, we have considered a particular case of a stationary Gaussian process, namely an $AR(1)$ process. But there are more stationary Gaussian processes other than the $AR(1)$ process. Also, in this paper, we have assumed that the measurement error follows an independent Gaussian distribution but there are some situations where the measurement error does not follow Gaussian distribution. These situations can be considered in future research.

Disclosure statement

There is no potential conflict of interest involved with this submission.

Acknowledgements

Thanks to anonymous reviewers whose suggestions helped to improve the paper.

References

Abay, K. A., Wossen, T., Abate, G. T., Stevenson, J. R., Michelson, H., & Barrett, C. B., (2023). Inferential and behavioral implications of measurement error in agricultural data. *Annual Review of Resource Economics*, 15(1), pp. 63–83.

Blackwell, M., Honaker, J., & King, G., (2017). A unified approach to measurement error and missing data: overview and applications. *Sociological Methods & Research*, 46(3), pp. 303–341.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M., (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brockwell, P. J., Davis, R. A. (Eds.), (2002). *Introduction to time series and forecasting*. New York, NY: Springer New York.

Carroll, R. J., (1998). Measurement error in epidemiologic studies. *Encyclopedia of biostatistics*, 3, pp. 2491–2519.

Costa, A. F., Castagliola, P., (2011). Effect of measurement error and autocorrelation on the X chart. *Journal of Applied Statistics*, 38(4), pp. 661–673.

Garza-Venegas, J. A., Tercero-Gómez, V. G., Lee Ho, L., Castagliola, P., & Celano, G., (2018). Effect of autocorrelation estimators on the performance of the X control chart. *Journal of Statistical Computation and Simulation*, 88(13), pp. 2612–2630.

Kotz, S., Balakrishnan, N., & Johnson, N. L., (2019). *Continuous multivariate distributions, Volume 1: Models and applications* (Vol. 334). John Wiley & Sons.

Koutsoyiannis, A., (1977). Theory of econometrics: an introductory exposition of econometric methods. (No Title).

Linna, K. W., Woodall, W. H., (2001). Effect of measurement error on Shewhart control charts. *Journal of Quality technology*, 33(2), pp. 213–222.

Maleki, M. R., Amiri, A., & Castagliola, P., (2017). Measurement errors in statistical process monitoring: A literature review. *Computers & Industrial Engineering*, 103, pp. 316–329.

Noor-ul-Amin, M., Javaid, A., Hanif, M., & Dogu, E., (2022). Performance of maximum EWMA control chart in the presence of measurement error using auxiliary information. *Communications in Statistics-Simulation and Computation*, 51(9), pp. 5482–5506.

Runger, G. C., Willemain, T. R., (1995). Model-based and model-free control of autocorrelated processes. *Journal of Quality Technology*, 27(4), pp. 283–292.

Schennach, S. M., (2016). Recent advances in the measurement error literature. *Annual Review of Economics*, 8(1), pp. 341–377.

Shongwe, S. C., Malela-Majika, J. C., (2021). A new double sampling scheme to monitor the process mean of autocorrelated observations using an AR (1) model with a skip sampling strategy. *Computers & Industrial Engineering*, 153, p. 107084.

Shongwe, S. C., Malela-Majika, J. C., & Castagliola, P., (2021). A combined mixed-s-skip sampling strategy to reduce the effect of autocorrelation on the X scheme with and without measurement errors. *Journal of Applied Statistics*, 48(7), pp. 1243–1268.

Shongwe, S. C., Malela-Majika, J. C., & Molahloe, T., (2019). One-sided runs-rules schemes to monitor autocorrelated time series data using a first-order autoregressive model with skip sampling strategies. *Quality and Reliability Engineering International*, 35(6), pp. 1973–1997.

Shumway, R. H., Stoffer, D. S., (2017). *Time series analysis and its applications: with R examples*. Springer

Wu, C. W., (2011). Using a novel approach to assess process performance in the presence of measurement errors. *Journal of Statistical Computation and Simulation*, 81(3), pp. 301–314.

Zhang, N. F., (1998). Estimating process capability indexes for autocorrelated data. *Journal of Applied Statistics*, 25(4), pp. 559–574.

Appendix

Derivation of $V\left(S_Y^2\right)$

Here, the observable process is $\{Y_t\}$ where $Y_t = X_t + E_t$, $E_t \sim IID\, N(0, \sigma_E^2)$. Sample variance is,

$$S_Y^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \tag{26}$$
$$= S_X^2 + \frac{2}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})(E_i - \bar{E}) + S_E^2$$

where,

$$S_E^2 = \frac{1}{n-1} \sum_{i=1}^{n} (E_i - \bar{E})^2, \quad \bar{E} = \frac{1}{n} \sum_{i=1}^{n} E_i.$$

Therefore,

$$Var(S_Y^2) = Var(S_X^2) + \frac{4}{(n-1)^2} Var\left(\sum_{i=1}^{n} (X_i - \bar{X})(E_i - \bar{E})\right) + Var(S_E^2). \tag{27}$$

Note that other covariance terms in Equation (27) will become zero as X_t, E_t are independent. From Zhang (1998) we get,

$$Var(S_X^2) = \frac{2\sigma_X^4}{(n-1)^2} F(n, \phi). \tag{28}$$

Also, we have,

$$Var(S_E^2) = \frac{2\sigma_E^4}{(n-1)}. \tag{29}$$

Now,

$$
\begin{aligned}
&Var\left(\sum_{i=1}^{n}(X_i-\bar{X})(E_i-\bar{E})\right)\\
&=\mathbf{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})(E_i-\bar{E})\right)^2\\
&=\mathbf{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\bar{X})(X_j-\bar{X})(E_i-\bar{E})(E_j-\bar{E})\right)\\
&=\sum_{i=1}^{n}\mathbf{E}(X_i-\bar{X})^2\mathbf{E}(E_i-\bar{E})^2+\frac{\sigma_E^2}{n}\sum_{i=1}^{n}\mathbf{E}(X_i-\bar{X})^2\\
&\quad-\frac{\sigma_E^2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}\left((X_i-\bar{X})(X_j-\bar{X})\right)\\
&=\sigma_E^2\sum_{i=1}^{n}\mathbf{E}(X_i-\bar{X})^2\\
&=(n-1)\sigma_E^2\sigma_X^2 f(n,\rho_i).
\end{aligned}
\tag{30}
$$

Substituting Equations (28), (29) and (30) in Equation (27) we finally get,

$$
Var(S_Y^2)=\frac{2\sigma_X^4}{(n-1)^2}\left\{F(n,\phi)+(n-1)\frac{\sigma_E^4}{\sigma_X^4}+2(n-1)f(n,\rho_i)\frac{\sigma_E^2}{\sigma_X^2}\right\}. \tag{31}
$$