ASYMPTOTIC PROPERTIES OF CORRECTED SCORE ESTIMATOR IN AUTOREGRESSIVE MODEL WITH MEASUREMENT ERRORS

UDC 519.21

D. S. PUPASHENKO, S. V. SHKLYAR, AND A. G. KUKUSH

Abstract. The autoregressive model with errors in variables with normally distributed control sequence is considered. For the main sequence, two cases are dealt with: (a) main sequence has stationary distribution, and (b) initial distribution is arbitrary, independent of the control sequence and has finite fourth moment. Here the elements of the main sequence are not observed directly, but surrogate data that include a normally distributed additive error are observed. Errors and main sequence are assumed to be mutually independent.

We estimate unknown parameter using the Corrected Score method and in both cases prove strict consistency and asymptotic normality of the estimator. To prove asymptotic normality we apply the theory of strong mixing sequences. Finally, we compare the efficiency of the Least Squares (naive) estimator and the Corrected Score estimator in the forecasting problem and conclude that the naive estimator gives better forecast.

Анотація. Розглядається модель аторегрессії з похибками у змінних і нормально розподіленою керуючою послідовністю. Для головної послідовності моделі розглянуто два випадки: а) головна послідовность має стаціонарний розподіл; б) початковий розподіл є довільним, не залежить від керуючої послідовності і має четвертий момент. Елементи головної послідовності не спостерігаються безпосередньо, натомість спостерігаються сурогатні дані, що включають нормально розподілену адитивну похибку. Похибки і головна послідовність є незалежними в сукупності.

Коефіцієнт авторегресії оцінюється методом виправленої оціночної функції. В обох випадках доведено строгу конзистентність і асимптотичну нормальність оцінки. Доведення асимптотичної нормальності спирається на властивості коефіцієнта сильного перемішування. В задачі прогнозу порівнюється ефективність (наївної) оцінки найменших квадратів і виправленої оцінки і робиться висновок, що наївна оцінка забезпечує кращий прогноз.

Аннотация. Рассматривается модель аторегрессии с ошибками в переменных и нормально распределенной управляющей последовательностью. Для главной последовательности рассмотрены два случая: а) главная последовательность имеет стационарное распределение; б) начальное распределение является произвольным, не зависит от управляющей последовательности и имеет четвертый момент. Элементы главной последовательности не наблюдаются непосредственно, а вместо них наблюдаются суррогатные данные, включающие нормально распределенную аддитивную ошибку. Ошибки и главная последовательность независимы в совокупности.

Коэффициент авторегрессии оценивается методом исправленной оценочной функции. В обоих случаях доказаны строгая состоятельность и асимптотическая нормальность оценки. Доказательство асимптотической нормальности опирается на свойства коэффициента сильного перемешивания. В задаче прогноза сравнивается эффективность (наивной) оценки наименьших квадратов и исправленной оценки и делается вывод, что наивная оценка обеспечивает лучший прогноз.

1. Introduction

Introduce an autoregressive (AR) sequence

$$X_n - \mu = a(X_{n-1} - \mu) + b\varepsilon_n, \quad n \geq 1, \qquad X_0 \sim N\left(\mu, \sigma^2\right), \tag{1}$$

where

2000 *Mathematics Subject Classification.* Primary 62F12; Secondary 62F10.

Key words and phrases. Autoregressive model, measurement errors, stationary process, strong mixing sequences, least squares estimation, corrected score estimation, efficiency comparison.

- coefficients a, b and mean μ are unknown parameters, such that $|a| < 1$ and $b > 0$,
- $\{X_0, \varepsilon_n, n \geq 1\}$ are independent random variables, $\varepsilon_n \sim N(0,1)$, $n \geq 1$.

Properties and applications of such models were studied, e.g., in McQuarrie and Tsai [10].

We are interested in estimators of the parameters a and μ. In case where there is no errors in variables, estimators of these parameters can be constructed by the Least Squares (LS) method with elementary criterion function

$$q_{LS}(X_k, X_{k-1}; a, \mu) = ((X_k - \mu) - a(X_{k-1} - \mu))^2.$$

Here we consider a situation where elements of the main sequence are not observed directly, but surrogate data that include additive errors are observed. Control sequence of the model is normally distributed and main sequence is stationary distributed, or as a different case, initial distribution is arbitrary, independent of the control sequence and has finite fourth moment.

Estimation of the parameters in autoregressive model with measurement error was considered in Dedecker et al. [7]. They proposed an estimation procedure based on modified least square criterion involving a suitably chosen weight function.

Other consistent estimators exist in this model. Letting $q \to \infty$ as the sample size is increasing, Chanda [6] applies Yule–Walker ARMA(p,q) estimator for errors-in-variables AR(p) model. The estimator does not use the error variance. Moreover, the errors are allowed to be slightly autocorrelated. Under some conditions, Chanda's estimator is consistent and asymptotically normal, but it is not $\sqrt{n}$-consistent.

In present paper we apply Corrected Score (CS) method (see Carroll et al. [4, Ch. 4]). We observe $W_k = X_k + V_k$, $k \geq 0$, where $V_k \sim N(0, \sigma_V^2)$ and $\{X_0, V_k, \varepsilon_k, k \geq 0\}$ are mutually independent. Consider the elementary score function of LS estimator

$$\psi_{0,LS}(X_k, X_{k-1}; a, \mu) = \frac{1}{2}\frac{\partial}{\partial a} q_{LS}(X_k, X_{k-1}; a, \mu).$$

We construct a new score $q_{CS}(W_k, W_{k-1}; a, \mu)$ as a solution to the deconvolution equation

$$\mathsf{E}_{a_0,\mu_0}(\psi_{0,CS}(W_k, W_{k-1}; a, \mu) \mid X_k, X_{k-1}) = \psi_{0,LS}(X_k, X_{k-1}; a, \mu) \quad \text{a.s.},$$

for all $a, \mu \in \mathbb{R}$. Then the CS estimator $(\hat{a}_n, \hat{\mu}_n)^T$ is defined as a solution to equation

$$\sum_{k=1}^{n} \psi_{0,CS}(W_k, W_{k-1}; a, \mu) = 0, \qquad a, \mu \in \mathbb{R}.$$

The true parameter a satisfies $|a| < 1$, and it will be shown below that $|\hat{a}_n| < 1$, for all $n \geq n_0(w)$ a.s.

In this paper we construct the CS estimator explicitly and study its asymptotic properties as $n \to \infty$.

The paper is organized as follows. The CS is given explicitly in Section 2. The strict consistency and asymptotic normality of the estimator are presented in Section 3, and Section 4 concludes. Proofs of the main results are given in Appendix.

We use the following notations. z^T is transposed vector z, E stands for expectation of a random variable, $\stackrel{P1}{\to}$ and $\stackrel{d}{\to}$ denote the convergence a.s. and in distribution respectively, $a_n \stackrel{P1}{\approx} b_n$ means that $a_n - b_n \stackrel{P1}{\to} 0$, as $n \to \infty$.

2. Construction of corrected score estimator

Rewrite model (1) in a more convenient way.

Lemma 2.1. *For the model* (1) *it holds*

$$X_n - \mu = b\sum_{i=1}^{n} a^{n-i}\varepsilon_i + a^n(X_0 - \mu), \qquad n \ge 1. \tag{2}$$

Proof. This statement is straightforward and can be proved by induction. □

From now on we suppose that $\{W_k, k = 0, \dots, n\}$ are observed instead of

$$\{X_k, k = 0, \dots, n\},$$

where the additive error $V_k \sim N(0, \sigma_V^2)$ and $\{V_k, X_k, k \ge 0\}$ are mutually independent.

First, for the unknown AR coefficient a and mean μ we construct the LS estimators (LSEs). To do that we introduce the objective function:

$$Q_{LS}(W_0, \dots, W_n; a, \mu) = \frac{1}{n}\sum_{k=1}^{n}((W_k - \mu) - a(W_{k-1} - \mu))^2,$$

and minimize it with respect to a and μ. Necessary and sufficient conditions for minimizing are:

$$\begin{cases} \frac{\partial Q_{LS}}{\partial a} = \frac{2}{n}\sum_{k=1}^{n}(a(W_{k-1} - \mu) - (W_k - \mu))(W_{k-1} - \mu) = 0, \\ \frac{\partial Q_{LS}}{\partial \mu} = \frac{2}{n}\sum_{k=1}^{n}(a(W_{k-1} - \mu) - (W_k - \mu))(1 - a) = 0. \end{cases}$$

Solving this system of equations, we get the LSE of the mean μ

$$\hat{\mu}_n = \frac{\sum_{k=1}^{n} W_k W_{k-1} \sum_{k=1}^{n} W_{k-1} - \sum_{k=1}^{n} W_{k-1}^2 \sum_{k=1}^{n} W_k}{n(\sum_{k=1}^{n} W_k W_{k-1} - \sum_{k=1}^{n} W_{k-1}^2) + (\sum_{k=1}^{n} W_{k-1})^2 - \sum_{k=1}^{n} W_{k-1}\sum_{k=1}^{n} W_k},$$

provided the denominator is nonzero, and the LSE of the parameter a is

$$\hat{a}_n^{\mathrm{LS}} = \frac{\sum_{k=1}^{n}(W_k - \hat{\mu}_n)(W_{k-1} - \hat{\mu}_n)}{\sum_{k=1}^{n}(W_{k-1} - \hat{\mu}_n)^2}. \tag{3}$$

Because the LSE $\hat{\mu}$ is too complicated to be investigated, we use the sample mean that provides a strict consistent estimator of the mean μ,

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=0}^{n-1} W_k \xrightarrow{P1} \mu, \quad \text{as } n \to \infty.$$

We prove that the $\hat{\mu}_n$ is asymptotically normal using the Central Limit Theorem (CLT) (see Billingsley [1, Th 27.4]) and results of Bosq and Blanke [3, p. 47–48] in order to ensure that we deal with a geometrically strong mixing sequence.

Next we construct an estimator of the regression coefficient a by the CS method. We introduce a function $\psi_{LS}(X_0, \dots, X_n; a, \mu)$ as

$$\psi_{LS}(X_0, \dots, X_n; a, \mu) = \frac{1}{2}\frac{\partial Q_{LS}}{\partial a} = \frac{a}{n}\sum_{k=1}^{n}(X_{k-1} - \mu)^2 - \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)(X_{k-1} - \mu).$$

We search for a function $\psi_{CS}(W_0, \dots, W_n; a, \mu)$ that satisfies the deconvolution equation

$$\mathsf{E}(\psi_{CS}(W_0, \dots, W_n; a, \mu) \mid X_0, \dots, X_n) = \psi_{LS}(X_0, \dots, X_n; a, \mu) \quad a.s. \tag{4}$$

To do that we obtain polinomial functions $h(W_k; \mu)$ and $g(W_{k-1}, W_k; \mu)$ that solve equations

$$\mathsf{E}(h(W_k; \mu) \mid X_k) = (X_k - \mu)^2 \quad \text{a.s.}, \tag{5}$$

$$\mathsf{E}(g(W_{k-1}, W_k; \mu) \mid X_{k-1}, X_k) = (X_k - \mu)(X_{k-1} - \mu) \quad \text{a.s.} \tag{6}$$

of the following form

$$h(W_k; \mu) = (W_k - \mu)^2 - \sigma_V^2,$$
$$g(W_{k-1}, W_k; \mu) = (W_{k-1} - \mu)(W_k - \mu).$$

Hence we get a polynomial solution to (4)

$$\psi_{CS}(W_0,\dots,W_n;a,\mu) = \frac{a}{n}\sum_{k=1}^{n}\left((W_{k-1}-\mu)^2-\sigma_V^2\right) - \frac{1}{n}\sum_{k=1}^{n}(W_{k-1}-\mu)(W_k-\mu).$$

Plugging-in the sample mean $\hat{\mu}_n$ and equating $\psi_{CS}(W_0,\dots,W_n;a,\hat{\mu}_n)$ to zero we get the CS estimator of a,

$$\hat{a}_n = \frac{\sum_{k=1}^{n}(W_k-\hat{\mu}_n)(W_{k-1}-\hat{\mu}_n)}{\sum_{k=1}^{n}(W_{k-1}-\hat{\mu}_n)^2 - n\sigma_V^2}. \tag{7}$$

Remark 2.1. *The denominator of* (7) *is nonzero starting from certain random number, i.e., for all* $n \ge n_0(\omega)$ *a.s.*

Proof of Remark 2.1 is a part of proof of Theorem 3.2, see Appendix.

3. Main results

Asymptotic properties of CS estimator. We state the consistency and asymptotic normality of the CS estimator (7) as $n \to \infty$.

Theorem 3.1. *In model* (1), *let* $\{X_k, k \ge 1\}$ *be a stationary process. Assume that variables* $\{X_0, \varepsilon_k, V_{k-1}, k \ge 1\}$ *are mutually independent, then the CS estimator* (7) *is strictly consistent.*

For Theorems 3.2 and 3.4, do not assume that X_0 has a stationary distribution of underlying AR sequence. In particular, assume (1) without requirement that $X_0 \sim N(\mu,\sigma^2)$.

Theorem 3.2. *Assume that* $\{X_k, k \ge 1\}$ *in AR* (1) *has an arbitrary initial distribution with finite fourth moment and variables* $\{X_0, \varepsilon_k, V_{k-1}, k \ge 1\}$ *are mutually independent. Then CS estimator* (7) *is strictly consistent.*

Theorem 3.3. *In AR* (1) *let* $\{X_k, k \ge 1\}$ *be a stationary process. Assume that variables* $\{X_0, \varepsilon_k, V_{k-1}, k \ge 1\}$ *are mutually independent. Then the CS estimator* (7) *is asymptotically normal with positive asymptotic variance*

$$\sigma_\infty^2 = 1-a^2+2\left(1-a^2\right)\frac{\sigma_V^2}{\sigma^2}+\left(2a^2+1\right)\frac{\sigma_V^4}{\sigma^4}. \tag{8}$$

Theorem 3.4. *Assume that* $\{X_k, k \ge 1\}$ *in AR* (1) *has arbitrary initial distribution with finite fourth moment and variables* $\{X_0, \varepsilon_k, V_{k-1}, k \ge 1\}$ *are mutually independent. Then the CS the estimator* (7) *is asymptotically normal with positive asymptotic variance*

$$\sigma_\infty^2 = 1-a^2+2\left(1-a^2\right)^2\frac{\sigma_V^2}{b^2}+\left(2a^2+1\right)\left(1-a^2\right)^2\frac{\sigma_V^4}{b^4}. \tag{9}$$

Remark 3.1. *In case of known parameter* μ, *the CS estimator of* a *is defined by* (7) *setting* $\hat{\mu}_n = \mu$. *Then the estimator remains strictly consistent and asymptotically normal with unchanged asymptotic variance* (8).

Proofs of Theorems 3.1 to 3.4 can be found in Appendix.

Comparison of the LS and CS estimators. We compare the efficiency of the LS estimator (3) and CS estimator (7) in the forecasting problem.

As two forecasts of the forthcoming observation W_{n+1} we take the values

$$W_{n+1}^{\mathrm{LS}} = \hat{\mu}_n + \hat{a}_n^{\mathrm{LS}}(W_n-\hat{\mu}_n), \qquad W_{n+1}^{\mathrm{CS}} = \hat{\mu}_n + \hat{a}_n^{\mathrm{CS}}(W_n-\hat{\mu}_n).$$

To find an optimal forecast $\mathsf{E}(W_{n+1}|W_n)$ first we calculate the correlation coefficient between W_n and W_{n+1},

$$\rho = \frac{a\sigma^2}{\sigma^2 + \sigma_V^2}.$$

Then we use a theorem from Kartashov [8] which states that for jointly Gaussian random variables $(\xi_1, \xi_2) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, the conditional expectation can be calculated as

$$\mathsf{E}(\xi_1 \mid \xi_2 = y) = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2).$$

Thus, the optimal forecast is

$$\mathsf{E}(W_{n+1} \mid W_n) = \mu + \rho(W_n - \mu) = \mu + \frac{a\sigma^2}{\sigma^2 + \sigma_V^2}(W_n - \mu).$$

But the parameters of the model are unknown, and instead one can use two forecasts constructed above. Because the CS estimator is strictly consistent, i.e. $\hat{a}_n^{\mathrm{CS}} \xrightarrow{P1} a$, as $n \to \infty$, and

$$\hat{a}_n^{\mathrm{LS}} \xrightarrow{P1} a \frac{\sigma^2}{\sigma^2 + \sigma_V^2} \quad \text{as } n \to \infty,$$

we have:

$$W_{n+1}^{\mathrm{CS}} - \mu = (a + o(1))(W_n - \mu) \quad \text{a.s.}$$

and for the LS forecast,

$$W_{n+1}^{\mathrm{LS}} - \mu = \hat{a}_n^{\mathrm{LS}}(W_n - \mu) = \left(a \frac{\sigma^2}{\sigma^2 + \sigma_V^2} + o(1) \right)(W_n - \mu) \quad \text{a.s.},$$

where $o(1)$ is a sequence of random variables that converges to 0 a.s.

Hence like in the example from Cheng and Van Ness [5, p. 70], we conclude that the naive LS estimator yields better forecast.

4. Conclusion

In this paper we considered the autoregressive model with measurement error. We proved the strict consistency and asymptotic normality of the CS estimator. Also we compared the efficiency of the LS (naive) estimator and CS estimator in the forecasting problem and showed that the naive estimator gives better forecast, though the naive estimator is inconsistent as $n \to \infty$.

Appendix

Proof of Theorem 3.1. We suppose that the main sequence of AR (1) has stationary distribution. Initial distribution is $X_0 \sim N(\mu, \sigma^2)$, therefore using stationarity of the process we get that $\sigma^2 = \frac{b^2}{1-a^2}$.

To show the strict consistency rewrite the estimator (7):

$$\begin{aligned}\hat{a}_n = &\frac{\frac{1}{n}\sum_{k=1}^n V_k V_{k-1} + \frac{1}{n}\sum_{k=1}^n (X_k - \hat{\mu}_n)(X_{k-1} - \hat{\mu}_n)}{\frac{1}{n}\sum_{k=1}^n (X_{k-1} - \hat{\mu}_n)^2 + \frac{1}{n}\sum_{k=1}^n V_{k-1}^2 + \frac{2}{n}\sum_{k=1}^n (X_{k-1} - \hat{\mu}_n)V_{k-1} - \sigma_V^2} \\ &+ \frac{\frac{1}{n}\sum_{k=1}^n (X_k - \hat{\mu}_n)V_{k-1} + \frac{1}{n}\sum_{k=1}^n V_k(X_{k-1} - \hat{\mu}_n)}{\frac{1}{n}\sum_{k=1}^n (X_{k-1} - \hat{\mu}_n)^2 + \frac{1}{n}\sum_{k=1}^n V_{k-1}^2 + \frac{2}{n}\sum_{k=1}^n (X_{k-1} - \hat{\mu}_n)V_{k-1} - \sigma_V^2}.\end{aligned} \tag{10}$$

We find the limits as $n \to \infty$ for all terms in (10) separately.

First consider the sequence

$$\frac{1}{n}\sum_{k=1}^n (X_{k-1} - \hat{\mu}_n)^2.$$

Rewriting it as

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)^2=\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2+(\mu-\hat{\mu}_n)\frac{2}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)+(\mu-\hat{\mu}_n)^2$$

and using strict consistency of sample mean $\hat{\mu}_n$, we get that the last two terms are vanishing as $n\to\infty$, hence

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)^2\overset{P1}{\approx}\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2.$$

To get the limit of the sequence

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2,$$

we use the ergodic theorem for stationary processes (see Korolyuk et al. [9]). Conditions of the ergodic theorem can be verified, and we get a limit of the first term in denominator of (10),

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)^2\overset{P1}{\approx}\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2\overset{P1}{\to}\mathsf{E}(X_0-\mu)^2=\sigma^2\quad\text{as } n\to\infty. \tag{11}$$

To get a limit for the second term we use the strong law of large numbers (SLLN):

$$\frac{1}{n}\sum_{k=1}^{n}V_{k-1}^2\overset{P1}{\to}\mathsf{E}\,V_0^2=\sigma_V^2\quad\text{as } n\to\infty. \tag{12}$$

By similar technique we get limits of all terms in (10) as $n\to\infty$:

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)V_{k-1}\overset{P1}{\to}0, \tag{13}$$

$$\frac{1}{n}\sum_{k=1}^{n}V_kV_{k-1}\overset{P1}{\to}0, \tag{14}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_k-\hat{\mu}_n)(X_{k-1}-\hat{\mu}_n)\overset{P1}{\to}a\sigma^2, \tag{15}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_k-\hat{\mu}_n)V_{k-1}\overset{P1}{\to}0, \tag{16}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)V_k\overset{P1}{\to}0. \tag{17}$$

Plugging limits (11)–(17) in expression (10), we get that $\hat{a}_n\overset{P1}{\to}a$ as $n\to\infty$. □

Proofs of Remark 2.1 and Theorem 3.2. We denote stationary distributed random variables satisfying (1) as $\{X_k^{\text{st}},k\ge 1\}$, with initial distribution $X_0^{\text{st}}\sim N(\mu,\sigma^2)$. We assume that $\{X_0,X_0^{\text{st}},\varepsilon_k,V_{k-1},k\ge 1\}$ are mutually independent.

Equality (2) implies that

$$X_n-\mu=(X_n^{\text{st}}-\mu)+a^n\left(X_0-X_0^{\text{st}}\right), \tag{18}$$

hence $X_n-X_n^{\text{st}}\xrightarrow{P1}0$ as $n\to\infty$.

Now we have to find a limit of (10) as $n\to\infty$.

First consider

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)^2 \overset{P1}{\approx} \frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2.$$

We plug expression (18) in the latter sequence, hence

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\mu)^2 &= \frac{1}{n}\sum_{k=1}^{n}\left((X_{k-1}^{\mathrm{st}}-\mu)+a^{k-1}(X_0-X_0^{\mathrm{st}})\right)^2\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(X_{k-1}^{\mathrm{st}}-\mu\right)^2+\frac{2}{n}\left(X_0-X_0^{\mathrm{st}}\right)\sum_{k=1}^{n}a^{k-1}\left(X_{k-1}^{\mathrm{st}}-\mu\right) \qquad (19)\\ &+\frac{1}{n}\left(X_0-X_0^{\mathrm{st}}\right)^2\sum_{k=1}^{n}a^{2(k-1)}.\end{aligned}$$

In the proof of Theorem 3.1, we have shown convergence of the first term in expression (19):

$$\frac{1}{n}\sum_{k=1}^{n}\left(X_{k-1}^{\mathrm{st}}-\mu\right)^2 \overset{P1}{\to} \sigma^2 \quad \text{as } n\to\infty.$$

Since $|a|<1$, we get that

$$\frac{1}{n}\left(X_0-X_0^{\mathrm{st}}\right)^2\sum_{k=1}^{n}a^{2(k-1)} \overset{P1}{\to} 0 \quad \text{as } n\to\infty.$$

For the second term of (19) we proceed as follows. Denote corresponding random sequence as

$$Y_n=\frac{2}{n}\sum_{k=1}^{n}a^{k-1}\left(X_{k-1}^{st}-\mu\right):$$

- First using Chebyshev's inequality

$$\sum_{n=1}^{\infty}\mathsf{P}(|Y_n|>C)\le\sum_{n=1}^{\infty}\frac{\mathsf{E}\,|Y_n|^2}{C^2}<\infty$$

  we show that for each $C>0$, it holds $\sum_{n=1}^{\infty}\mathsf{P}(|Y_n|>C)<\infty$.
- Then Borel–Cantelli lemma implies that $\forall C>0\ \exists n_0\ \forall n\ge n_0\colon |Y_n|\le C$ a.s.

Therefore, to prove that $Y_n \overset{P1}{\to} 0$ as $n\to\infty$, it is enough to show $\sum_{n=1}^{\infty}\mathsf{E}\,|Y_n|^2<\infty$.

After quite cumbersome calculations, we can show that

$$\sum_{n=1}^{\infty}\mathsf{E}\left(\frac{2}{n}\sum_{k=1}^{n}a^{k-1}(X_{k-1}^{\mathrm{st}}-\mu)\right)^2$$

converges.

Hence

$$\frac{2}{n}\left(X_0-X_0^{\mathrm{st}}\right)\sum_{k=1}^{n}a^{k-1}\left(X_{k-1}^{\mathrm{st}}-\mu\right) \overset{P1}{\to} 0 \quad \text{as } n\to\infty.$$

Therefore, plugging all limits found above in (19), we obtain

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)^2 \overset{P1}{\to} \sigma^2 \quad \text{as } n\to\infty. \qquad (20)$$

Similarly we get limits of all terms in (10) as $n \to \infty$:

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)V_{k-1} \overset{P1}{\to} 0, \tag{21}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)(X_k-\hat{\mu}_n) \overset{P1}{\to} a\sigma^2, \tag{22}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_k-\hat{\mu}_n)V_{k-1} \overset{P1}{\to} 0, \tag{23}$$

$$\frac{1}{n}\sum_{k=1}^{n}(X_{k-1}-\hat{\mu}_n)V_k \overset{P1}{\to} 0. \tag{24}$$

We plug (12), (14), (20)–(24) in expression (10) and get that $\hat{a}_n \overset{P1}{\to} a$, as $n \to \infty$, hence the estimator (7) is strictly consistent. A limit of the denominator in (10) is nonzero, therefore, the statement of Remark 2.1 holds true. □

Now, we state lemmas for mixing coefficients and mixing sequences.

First note that for two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, the strong mixing coefficient is defined as follows:

$$\alpha(\mathcal{G},\mathcal{H}) = \sup_{A\in\mathcal{G}, B\in\mathcal{H}} |\,\mathsf{P}(A\cap B) - \mathsf{P}(A)\,\mathsf{P}(B)|.$$

For a random sequence $\{X_k, k\ge 0\}$, denote

$$\alpha^X(m) = \sup_{k\ge 0}\alpha(\sigma(X_0,\ldots,X_k), \sigma(X_{k+m}, X_{k+m+1},\ldots)).$$

The sequence $\{X_k, k\ge 0\}$ is called *a strong mixing process* if $\lim_{m\to\infty}\alpha^X(m) = 0$. It is called *a geometrically strong mixing* (GSM) process if

$$\alpha^X(m) \le br^m, \qquad m\ge 0,$$

for some $0<r<1$ and $b>0$.

Now, we state a helpful lemma which is a direct consequence of the definition of strong mixing sequences (see Billingsley [2]).

Lemma 4.1. *Let $\{X_n, n\ge 0\}$ be a random sequence and $Z_n = (X_{n-l},\ldots,X_n)^T$, $n\ge l$. Then for α-mixing coefficients associated to sequences $\{X_n, n\ge 0\}$ and $\{Z_n, n\ge l\}$, the following relation holds true:*

$$\alpha^X(m) = \alpha^Z(m+l), \qquad m\ge 0$$

Corollary 4.1. *Let $\{X_n, n\ge 0\}$ be a random sequence and $Z_n = (X_{n-l},\ldots,X_n)^T$, $n\ge l$. For a Borel measurable vector function f, consider a sequence*

$$f(Z) = \{f(Z_n), n\ge l\}.$$

Then

$$\alpha^X(m) \ge \alpha^{f(Z)}(m+l), \qquad m\ge 0.$$

If $\{X_n, n\ge 0\}$ is a strong mixing sequence then $\{f(Z_n), n\ge l\}$ is a strong mixing sequence as well. If $\{X_n, n\ge 0\}$ is a GSM sequence then so is $\{f(Z_n), n\ge l\}$.

Lemma 4.2. *Let $(\Omega_1, \mathcal{F}_1, \mathsf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ be two probability spaces. Let $\mathcal{G}_1$ and $\mathcal{H}_1$ be two sub-σ-algebras of $\mathcal{F}_1$ and let $\mathcal{G}_2$ and $\mathcal{H}_2$ be two independent sub-σ-algebras of $\mathcal{F}_2$. Then*

$$\alpha(\sigma(\mathcal{G}_1\times\mathcal{G}_2), \sigma(\mathcal{H}_1\times\mathcal{H}_2)) = \alpha_1(\mathcal{G}_1,\mathcal{H}_1),$$

Here for the calculating mixing coefficient α_1 we use measure P_1; and for α product measure $\mathsf{P} = \mathsf{P}_1\times\mathsf{P}_2$ is used.

Proof. Denote $\mathcal{G} = \sigma(\mathcal{G}_1 \times \mathcal{G}_2)$, $\mathcal{H} = \sigma(\mathcal{H}_1 \times \mathcal{H}_2)$. Expectation in $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ is denoted as E_2. For $A \in \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$, denote the section $A_{\omega_2} := \{\omega_1 \in \Omega_1 \mid (\omega_1, \omega_2) \in A\} \in \mathcal{F}_1$.

Let $A \in \mathcal{G}$ and $B \in \mathcal{H}$. Then $\mathsf{P}_1(A_{\omega_2})$ and $\mathsf{P}_1(B_{\omega_2})$ are independent random variables. Hence

$$\mathsf{P}(A)\,\mathsf{P}\,(B) = \mathsf{E}_2(\mathsf{P}_1(A_{\omega_2}))\ \mathsf{E}_2(\mathsf{P}_1(A_{\omega_2})) = \mathsf{E}_2(\mathsf{P}_1(B_{\omega_2})\ \mathsf{P}_1(B_{\omega_2})).$$

We have

$$|\,\mathsf{P}_1(A_{\omega_2} \cap B_{\omega_2}) - \mathsf{P}_1(A_{\omega_2})\,\mathsf{P}_1(A_{\omega_2})| \le \alpha_1(\mathcal{G}_1, \mathcal{H}_1) \quad \mathsf{P}_2\text{-a.s.},$$

$$\begin{aligned}|\,\mathsf{P}(A \cap B) - \mathsf{P}(A)\,\mathsf{P}(B)| &= |\,\mathsf{E}_2(\mathsf{P}_1((A \cap B)_{\omega_2})) - \mathsf{E}_2(\mathsf{P}_1(A_{\omega_2})\ \mathsf{P}_1(B_{\omega_2}))| \\ &= |\,\mathsf{E}_2(\mathsf{P}_1(A_{\omega_2} \cap B_{\omega_2}) - \mathsf{P}_1(A_{\omega_2})\ \mathsf{P}_1(B_{\omega_2}))| \le \alpha_1(\mathcal{G}_1, \mathcal{H}_1).\end{aligned}$$

Varying A and B, we get

$$\alpha(\mathcal{G}, \mathcal{H}) \le \alpha_1(\mathcal{G}_1, \mathcal{H}_1). \tag{25}$$

From the other hand

$$\begin{aligned}\alpha(\mathcal{G}, \mathcal{H}) &= \sup_{A \in \mathcal{G},\, B \in \mathcal{H}} |\,\mathsf{P}(A \cap B) - \mathsf{P}(A)\,\mathsf{P}(B)| \\ &\ge \sup_{A_1 \in \mathcal{G}_1,\, B_1 \in \mathcal{H}_1} |\,\mathsf{P}((A_1 \times \Omega_2) \cap (B_1 \times \Omega_2)) - \mathsf{P}(A_1 \times \Omega_2)\,\mathsf{P}(B_1 \times \Omega_2)| \\ &= \sup_{A_1 \in \mathcal{G}_1,\, B_1 \in \mathcal{H}_1} |\,\mathsf{P}_1(A_1 \cap B_1) - \mathsf{P}_1(A_1)\,\mathsf{P}_1(B_1)| = \alpha_1(\mathcal{G}_1, \mathcal{H}_1).\end{aligned} \tag{26}$$

Inequalities (25) and (26) imply the statement of Lemma. □

Under conditions of Lemma 4.2, a similar relation holds true for ϕ-mixing coefficients:

$$\phi(\sigma(\mathcal{G}_1 \times \mathcal{G}_2), \sigma(\mathcal{H}_1 \times \mathcal{H}_2)) = \phi_1(\mathcal{G}_1, \mathcal{H}_1),$$

where

$$\phi(\mathcal{G}, \mathcal{H}) = \sup_{A \in \mathcal{G},\, B \in \mathcal{H},\, \mathsf{P}(B) \ne 0} \bigl|\mathsf{P}(A \mid B) - \mathsf{P}(A)\bigr|.$$

Proof of Theorem 3.3. Now, the process $\{X_k, k \ge 0\}$ is stationary, $X_0 \sim N(\mu, \sigma^2)$ and $\sigma^2 = \frac{b^2}{1-a^2}$.

From expression (7) for estimator $\hat{\alpha}_n$ we get

$$\sqrt{n}(\hat{a}_n - a) = \frac{\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_{k-1} - \hat{\mu}_n)(W_k - \hat{\mu}_n - a(W_{k-1} - \hat{\mu}_n)) + \sqrt{n} a \sigma_V^2}{\frac{1}{n} \sum_{k=1} n \sum_{k=1}^{n} (W_{k-1} - \hat{\mu})^2 - \sigma_V^2} =: \frac{A_n}{B_n}.$$

From the proof of Theorem 3.1 we get a limit of the denominator:

$$B_n \xrightarrow{P1} \sigma^2. \tag{27}$$

Now, rewrite the numerator. Because $\sum_{k=1}^{n} (W_{k-1} - \hat{\mu}_n) = 0$, we have

$$\begin{aligned}A_n &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_{k-1} - \hat{\mu}_n)(W_k - aW_{k-1}) + \sqrt{n} a \sigma_V^2 \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_{k-1} - \mu)(W_k - \mu - a(W_{k-1} - \mu)) + \sqrt{n} a \sigma_V^2 \\ &\quad - \frac{\hat{\mu}_n - \mu}{\sqrt{n}} \sum_{k=1}^{n} (W_k - \mu - a(W_{k-1} - \mu)).\end{aligned}$$

By the classical CLT,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (W_k - \mu - a(W_{k-1} - \mu)) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (V_k - aV_{k-1} + b\varepsilon_k)$$

converges in distribution. Remember that $\hat{\mu}_n$ is a consistent estimator of μ. Then by Slutsky lemma,

$$\frac{\hat{\mu}_n - \mu}{\sqrt{n}} \sum_{k=1}^{n} (W_k - \mu - a(W_{k-1} - \mu)) \xrightarrow{\mathsf{P}} 0 \quad \text{as } n \to \infty.$$

Denote

$$\begin{aligned} Z_k &= (W_{k-1} - \mu)(W_k - \mu - a(W_{k-1} - \mu)) + a\sigma_V^2 \\ &= (W_{k-1} - \mu)(V_k - aV_{k-1} + b\varepsilon_k) + a\sigma_V^2. \end{aligned}$$

With this notation, $A_n \overset{P}{\approx} \widetilde{A} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$.

The AR process $\{X_k - \mu, k \geq 0\}$ is a GSM sequence, see Bosq, Blanke [3, Ex. 1.5, p. 47]. By Lemma 4.2 $\{(X_k - \mu, V_k)^T, k \geq 0\}$ is a GSM sequence too. Then by Corollary 4.1 $\{Z_k, k \geq 1\}$ is a GSM sequence. Also $\{Z_k, k \geq 1\}$ is a strictly stationary process with $\mathsf{E}\, Z_k = 0$ and $\mathsf{E}\, Z_k^{12} < \infty$. Applying CLT, we get

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k \xrightarrow{d} N\left(0, \sigma_A^2\right)$$

with $\sigma_A^2 = \mathsf{E}\, Z_1^2 + 2\sum_{k=2}^{\infty} \mathsf{E}\, Z_1 Z_k$. After some calculations we have

$$\begin{aligned} \mathsf{E}\, Z_1^2 &= \left(1 - a^2\right) \sigma^4 + 2\sigma^2 \sigma_V^2 + \left(2a^2 + 1\right) \sigma_V^4, \\ \mathsf{E}\, Z_1 Z_2 &= -a^2 \sigma^2 \sigma_V^2, \\ \mathsf{E}\, Z_1 Z_k &= 0, \qquad k \geq 3. \end{aligned}$$

Therefore

$$\sigma_A^2 = \left(1 - a^2\right) \sigma^4 + 2\left(1 - a^2\right) \sigma^2 \sigma_V^2 + \left(2a^2 + 1\right) \sigma_V^4.$$

Finally,

$$A_n \xrightarrow{d} N\left(0, \sigma_A^2\right), \tag{28}$$

$$\sqrt{n}(\hat{a} - a) = \frac{A_n}{B_n} \xrightarrow{d} N(0, \sigma_\infty^2)$$

with

$$\sigma_\infty^2 = \frac{\sigma_A^2}{\sigma^4} = 1 - a^2 + 2\left(1 - a^2\right) \frac{\sigma_V^2}{\sigma^2} + \left(2a^2 + 1\right) \frac{\sigma_V^4}{\sigma^4}.$$

Obviously $\sigma_\infty^2 > 0$. Thus, $\hat{\alpha}_n$ is an asymptotically normal estimator of a. □

Proof of Theorem 3.4. Proof of this theorem differs from the proof of Theorem 3.3 only when we deal with numerator $\widetilde{A}_n$. For the case of stationary initial distribution we denote it as $\widetilde{A}_n^{\text{st}}$. Then using relation (18) we rewrite $\widetilde{A}_n$ for an arbitrary distribution as follows:

$$\begin{aligned} \widetilde{A}_n &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} V_k V_{k-1} + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu)(X_{k-1} - \mu) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) V_{k-1} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} V_k (X_{k-1} - \mu) - \frac{a}{\sqrt{n}} \sum_{k=1}^{n} (X_{k-1} - \mu)^2 - \frac{a}{\sqrt{n}} \sum_{k=1}^{n} V_{k-1}^2 \\ &\quad - \frac{2a}{\sqrt{n}} \sum_{k=1}^{n} (X_{k-1} - \mu) V_{k-1} + a\sqrt{n}\sigma_V^2 \\ &= \widetilde{A}_n^{\text{st}} + \frac{1}{\sqrt{n}} (X_0 - X_0^{\text{st}}) \sum_{k=1}^{n} a^{k-1} (V_k - aV_{k-1} + b\varepsilon_k). \end{aligned}$$

Because $\widetilde{A}_n^{\rm st}$ converges in distribution, it remains to prove only that the last term

$$\frac{1}{\sqrt{n}}(X_0 - X_0^{\rm st}) \sum_{k=1}^{n} a^{k-1}(V_k - aV_{k-1} + b\varepsilon_k)$$

converges to 0 in probability. We have

$$\begin{aligned} \mathsf{E}\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} a^{k-1}(V_k - aV_{k-1} + b\varepsilon_k)\right| &\le \frac{1}{\sqrt{n}} \mathsf{E} \sum_{k=1}^{n} |a^{k-1}|(|V_k| + |aV_{k-1}| + |b\varepsilon_k|) \\ &\le \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |a^{k-1}|(\mathsf{E}\,|V_k| + \mathsf{E}\,|aV_{k-1}| + \mathsf{E}\,|b\varepsilon_k|). \end{aligned}$$

Because the sum $(\mathsf{E}\,|V_k| + \mathsf{E}\,|aV_{k-1}| + \mathsf{E}\,|b\varepsilon_k|)$ can be bounded by some constant c and $|a| < 1$, we have:

$$\mathsf{E}\left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} a^{k-1}(V_k - aV_{k-1} + b\varepsilon_k)\right| \le \frac{c}{\sqrt{n}} \frac{1-|a|^n}{1-|a|} \to 0 \quad \text{as } n \to \infty.$$

Hence we obtain that $A_n \overset{\mathsf{P}}{\approx} \widetilde{A}_n^{\rm st}$ and from (27), (28) with Slutsky lemma for all $|a| < 1$ we get:

$$\sqrt{n}(\hat{a}_n - a) = \frac{A_n}{B_n} \xrightarrow{d} \zeta \sim N\left(0, \sigma_\infty^2\right) \quad \text{as } n \to \infty. \qquad \square$$

References

1. P. Billingsley, *Probability and Measure*, John Wiley & Sons Inc., 1995.
2. P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons Inc., 1999.
3. D. Bosq and D. Blanke, *Inference and Prediction in Large Dimensions*, John Wiley & Sons Inc., 2007.
4. R. J. Carroll, D. Ruppert, and L. A. Stefanski, *Measurement Error in Nonlinear Models*, Chapman & Hall., London, 1995.
5. C. L. Cheng and J. Van Ness, *Statistical Regression with Measurement Error*, Chapman & Hall, London, 1999.
6. K. C. Chanda, *Asymptotic properties of estimators for autoregressive models with errors in variables*, The Annals of Statistics **24** (1996), 423–430.
7. J. Dedecker, A. Samson, and M.-L. Taupin, *Estimation in Autoregressive Model with Measurement Error*, Chapman & Hall, London, 2011.
8. M. V. Kartashov, *Probability, Processes, Statistics*, Kyiv University, Kyiv, 2008.
9. V. S. Korolyuk, N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, *A Manual on Probability Theory and Mathematical Statistics*, "Nauka", Moscow, 1985.
10. A. D. R. McQuarrie and C.-L. Tsai, *Regression and Time Series Model Selection*, Singapore, World Scientific, 1998.

Taras Shevchenko National University of Kyiv, Volodymyrska st. 64, 01601, Kyiv, Ukraine
E-mail address: daria.pupashenko@gmail.com

Taras Shevchenko National University of Kyiv, Volodymyrska st. 64, 01601, Kyiv, Ukraine
E-mail address: shklyar@mail.univ.kiev.ua

Taras Shevchenko National University of Kyiv, Volodymyrska st. 64, 01601, Kyiv, Ukraine
E-mail address: alexander_kukush@univ.kiev.ua

Received 01/04/2013