157. Bias and heritability of the autocorrelation based on longitudinal data used as resilience indicator

H.A. Mulder

Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; han.mulder@wur.nl

Abstract

To genetically improve resilience, the autocorrelation between subsequent deviations from fitted curves of longitudinal data was proposed. Autocorrelation is related to the speed of recovery after a disturbance: a high autocorrelation means slow recovery, while a low autocorrelation means fast recovery. It is known that estimated autocorrelations are biased, while the genetic properties of the autocorrelation are largely unknown. The aims of this research were to investigate the bias and the heritability of the autocorrelation. Deterministic expressions were derived and evaluated with Monte Carlo simulation. The estimated autocorrelation was 0.2-0.3 lower than the true value when the number of records was 10. The heritability was between 0.05 and 0.10 in most situations. It is recommended to have at least 50 records per animal. This shows good opportunities for genetic improvement. This study is a first step towards better understanding of the mathematical and genetic properties of the autocorrelation.

Introduction

In recent studies on genetics of resilience indicators, the lag-one autocorrelation, or simply called autocorrelation, based on subsequent deviations from fitted curves of longitudinal data from an animal, was found to be a promising resilience indicator. Berghof *et al.* (2019) and Poppe *et al.* (2020, 2021a) showed that the autocorrelation is a heritable trait, with a heritability between 0.05 and 0.1. Autocorrelation based on milk yield deviations was related to recovery from environmental perturbations, e.g. from a heat wave or an unspecified herd disturbance (Poppe *et al.*, 2021b). Cows with genetically a low autocorrelation had a faster recovery than cows with a high autocorrelation. This means that cows with a low autocorrelation need less time until their milk yield reaches normal levels after a disturbance than cows with a high autocorrelation. Genetic correlations between the autocorrelation based on milk yield deviations and health and longevity traits were weak in dairy cattle (Poppe *et al.*, 2020). No relationships were found between the autocorrelation based on body weight deviations and lesion scores or mortality upon a disease challenge in chickens (Berghof *et al.*, 2019). These weak or absent relationships may be partly due to properties of the autocorrelation.

The autocorrelation is new to quantitative geneticists, although it has been studied for many years in resilience research (Scheffer *et al.* 2015) or in research using time series (Arnau and Bono, 2001). It is already known for many years that estimates of autocorrelation are biased, especially with a small number of records (Kendall, 1954; Marriott and Pope, 1954; Arnau and Bono, 2001). The mathematical properties of the autocorrelation are largely unknown for geneticists. Furthermore, there is no equation to understand the relationship between the heritability of the autocorrelation and the number of records. Such an equation would help in optimizing breeding schemes to genetically improve resilience.

The aim of this study was to investigate the bias and the heritability of the estimated autocorrelation. Monte Carlo simulation was used to evaluate the deterministic equations.

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_157, © H.A. Mulder 2022

Materials & methods

Genetic model and Monte Carlo simulation. Phenotypes were generated following a Markov process with autoregressive environmental effects. For simplicity, the phenotype was equal to the environmental effects and did not include additive genetic effects on the level of the phenotype, i.e. the normal additive genetic variance utilized in breeding programs:

$$P_{i,t} = \mu_P + E_{i,t} = \mu_P + r_{a_i} E_{i,t-1} + e_{i,t} \tag{1}$$

where $P_{i,t}$ is the phenotype at time t of individual i, μ_P is the mean of the trait and assumed to be zero, r_{a_i} is the lag-one autocorrelation, or simply called autocorrelation, of individual i, $E_{i,t} = r_{a_i} E_{i,t-1} + e_{i,t}$ is the environmental effect at time t and $e_{i,t}$ is the random environmental effect specific to time t. Since $\mu_P = 0$, $E_{i,t} = P_{i,t}$ here. The initial environmental variance σ_E^2, e.g. for the first environmental effect to be sampled $E_{i,l}$, was 1.0. The variance of $e_{i,t}$ was $1 - r_{a_i}^2$, so that the phenotypic variance is one. The r_{a_i} had a mean μ_{r_a} a random additive genetic effect A_{r_i} ($\mathrm{N}(0, \sigma_{A_r}^2)$) and a random environmental effect E_{r_i} ($\mathrm{N}(0, \sigma_{E_r}^2)$):

$$r_{a_i} = \mu_{r_a} + A_{r_i} + E_{r_i} \tag{2}$$

Monte Carlo simulations were used to evaluate the deterministic expression for the estimated autocorrelation $\widehat{r_a}$. In the simulations to evaluate the expression for the estimated autocorrelation, $\sigma_{A_r}^2 = 0$ and $\sigma_{E_r}^2 = 0$, while μ_{r_a} was varied between -0.9 and 0.9 and the number of records was varied between 10 and 300. In the simulations to evaluate the simple expression for heritability, the maximum heritability was set to 0.1 and $\sigma_{A_r}^2 = 0.003$ and $\sigma_{E_r}^2 = 0.027$, which approximately resemble estimated parameters by Poppe *et al.* (2020). The results were averages of 100 replicates.

Deterministic equation for the estimated autocorrelation. The estimated autocorrelation $\widehat{r_a}$ can be approximated with a Taylor series approximation (Lynch and Walsh, 1998):

$$\widehat{r_a} \cong \frac{\widehat{\sigma_{E_t,E_{t-1}}}}{\widehat{\sigma_E^2}} \left(1 + \frac{var(\widehat{\sigma_E^2})}{\widehat{\sigma_E^2}} - \frac{cov(\widehat{\sigma_E^2}, \widehat{\sigma_{E_t,E_{t-1}}})}{\widehat{\sigma_E^2}\widehat{\sigma_{E_t,E_{t-1}}}}\right) \tag{3}$$

where $var(\widehat{\sigma_E^2})$ is the variance of $\widehat{\sigma_E^2}$, i.e. the variance of a variance estimate, $cov(\widehat{\sigma_E^2}, \widehat{\sigma_{E_t,E_{t-1}}})$ is the covariance between $\widehat{\sigma_E^2}$ and $\widehat{\sigma_{E_t,E_{t-1}}}$ i.e. a covariance between a variance and a covariance estimate. The required expressions for $\widehat{\sigma_E^2}$, $\widehat{\sigma_{E_t,E_{t-1}}}$, $var(\widehat{\sigma_E^2})$ and $cov(\widehat{\sigma_E^2}, \widehat{\sigma_{E_t,E_{t-1}}})$ were derived and will be shown in a full paper.

Simple equation for the heritability. The heritability of the autocorrelation $h_{r_a}^2$ is a function of the number of records. The number of records determines the sampling variance of the autocorrelation, which can be approximated as $\sigma_{r_a}^2 \cong \frac{(1 - r_a^2)^2}{n - 2}$ (Lynch and Walsh, 1998), where n is the number of records. The heritability can therefore be approximated as:

$$h_{r_a}^2 = \frac{\sigma_{A_r}^2}{\sigma_{A_r}^2 + \sigma_{E_r}^2 + \sigma_{r_a}^2} = \frac{\sigma_{A_r}^2}{\sigma_{A_r}^2 + \sigma_{E_r}^2 + \frac{(1 - r_a^2)^2}{n-2}} \tag{4}$$

Equation 4 makes it feasible to quickly evaluate the effect of the number of records on the heritability of the autocorrelation and therefore also the accuracy of selection when used as selection criterion.

Results

The estimated autocorrelation was substantially lower than the true autocorrelation, indicating bias, as shown in Figures 1A and 1B. The curves show that the bias was as large as 0.2-0.29 when the number of records was 10. The bias decreased slowly when the number of records increased, but even with 200-300 records the bias was 0.005-0.013 when the autocorrelation was 0.3 or 0.5. The bias was larger when the autocorrelation was larger. The deterministic prediction using Equation 3 resembled results from Monte Carlo simulation, while a small deviation was observed when the number of records was 10. If an acceptable bias is smaller than 0.05, the minimum number of records is 40 when the autocorrelation is 0.3 and >50 when the autocorrelation is 0.5. In summary, the estimated autocorrelation can be substantially lower than its true value; therefore it is recommended to have at least 50 records per individual.

The simple equation for heritability (Equation 4) accurately predicted the heritability when the number of records was large, e.g. 80-100, or when the autocorrelation was 0.1 (Figure 2A). However, when the number

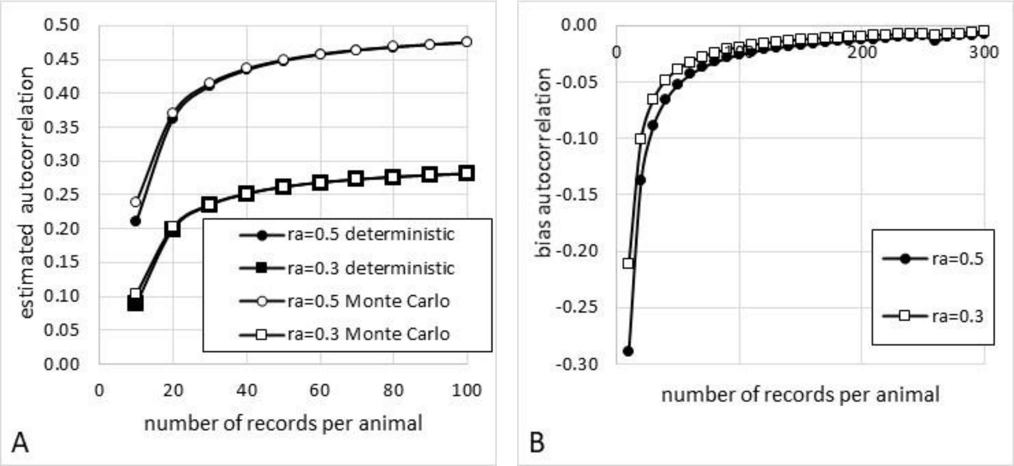

Figure 1. The estimated autocorrelation (panel A) and its bias (estimated − true autocorrelation; panel B) as a function of the number of records per animal when the true autocorrelation (ra) is 0.3 and 0.5 using deterministic Equation 3 (panel A and B) or Monte Carlo simulation (panel A).

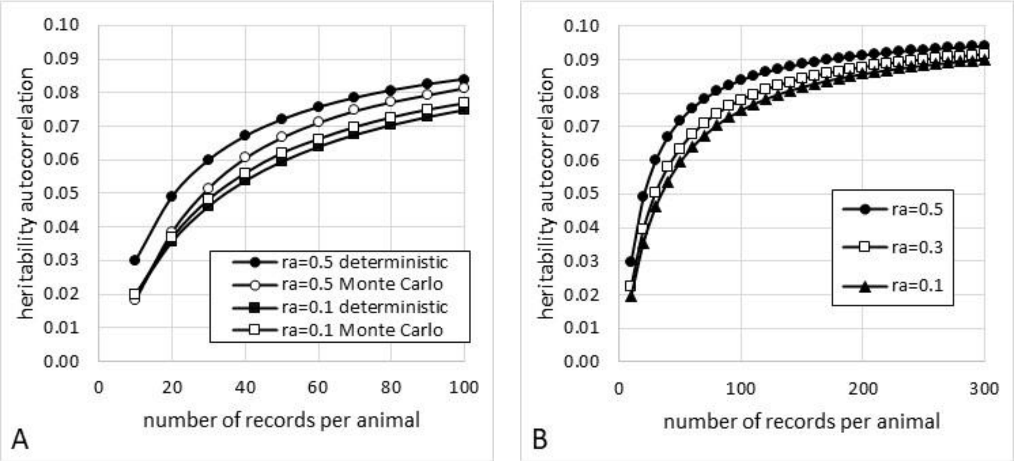

Figure 2. The heritability of the autocorrelation as a function of the number of records when the mean autocorrelation is 0.1, 0.3, 0.5 and the maximum heritability is 0.1 using the deterministic Equation 4 or Monte Carlo simulation.

of records was smaller than 50 and the autocorrelation was 0.5, the simple equation overpredicted the heritability by 8-65%. Figure 2B shows that the heritability was larger when the mean autocorrelation was higher. Given a number of records, the heritability was lowest when the mean autocorrelation was zero. The increase in heritability was small when the number of records was larger than 200. Figure 2B shows that the heritability was between 0.06 and 0.1 when the number of records was at least 50 and when the maximum heritability was 0.1.

Discussion

In this study, the bias in the autocorrelation and the heritability of the autocorrelation were investigated and deterministic equations were presented. The bias in the autocorrelation was already studied in detail by Marriot and Pope (1954) and Kendall (1954). The bias is due to the dependency between the estimated covariance and the estimated variance, which is shown in Equation 3 by the term $\frac{cov(\widehat{\sigma_E^2},\widehat{\sigma_{E_t,E_{t-1}}})}{\widehat{\sigma_E^2}\widehat{\sigma_{E_t,E_{t-1}}}}$. The estimated autocorrelation may be corrected by adding the predicted bias based on the number of records (Arnau and Bono, 2001). Alternatively, a fixed polynomial regression on the number of records may be included in the model to account for the non-linear effect of number of records on the estimated autocorrelation.

The bias in the mean autocorrelation also affects the heritability of the autocorrelation, because the genetic covariance between the breeding value for the autocorrelation and the estimated autocorrelation, i.e. the numerator of the heritability, was smaller than the true value. Furthermore, the variance of the autocorrelation, the denominator, was underestimated in Equation 4. However, the simple equation gave a very good prediction of the heritability when the number of records was larger than 50. The equation can be used to get more insight in the accuracy of selection for different types of longitudinal data in animal breeding. In general, given the range of heritabilities between 0.05 and 0.10, there are good opportunities for genetic improvement of resilience by selecting on a lower autocorrelation. This study is a step forward in quantitative genetic understanding of the autocorrelation used as resilience indicator in breeding.

References

Arnau, J., Bono, R. (2001) Qual. Quant. 35:365-387. https://doi.org/10.1023/A:1012223430234

Berghof, T.V.L., Bovenhuis, H., Mulder, H.A. (2019) Front. Genet. 10:1216 https://doi.org/10.3389/fgene.2019.01216

Kendall, M.G. (1954) Biometrika 41:403-404. https://doi.org/10.1093/biomet/41.3-4.403

Lynch, M., Walsh, B. (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, US.

Marriott, F.H.C., Pope, J.A. (1954) Biometrika 41:390-402. https://doi.org/10.2307/2332719

Poppe, M., Bonekamp, G., van Pelt, M.L., Mulder, H.A. (2021a) J. Dairy Sci. 104(2):1967-1981. https://doi.org/10.3168/jds.2020-19245

Poppe, M., Mulder, H.A., Veerkamp, R.F. (2021b) J. Dairy Sci. 104(7):8094-8106. https://doi.org/10.3168/jds.2020-19817

Poppe, M., Veerkamp, R.F., Van Pelt, M.L., Mulder, H.A. (2020) J. Dairy Sci. 103(2):1667-1684. https://doi.org/10.3168/jds.2019-17290

Scheffer, M., Carpenter, S.R., Dakos, V., van Nes, E.H. (2015) Annu. Rev. Ecol. Evol. Syst. 46:145-167. https://doi.org/10.1146/annurev-ecolsys-112414-054242