



Estimation of heritability using half-sib model under correlated errors

AMRIT KUMAR PAUL¹, HIMADRI SHEKHAR ROY^{1✉}, RANJIT KUMAR PAUL¹ and M D YEASIN¹

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012 India

Received: 18 August 2022; Accepted: 19 September 2022

ABSTRACT

In general, statistical models for estimation of heritability follow certain assumptions, i.e. random components including the error follow a normal distribution and are identically independently distributed. But in the practical situation, sometimes these assumptions are violated. Thus, from the perspective of plant and animal breeding programs, estimating various genetic variances and inferring their inheritance based on estimations of various genetic parameters is studied. In the present study, estimation of heritability for the half-sib model is considered with correlated error, and sire and error follow a range of different distributions like normal, Cauchy, beta, and t-distribution. Two error structures AR(1) and AR(2) was considered and observations for correlated and uncorrelated cases were generated using a one-way classification model. The developed procedure was applied using the generated observations using simulation. Various heritability ranges, such as high and low (0.5, 0.1), Half-sib AR(1), varied sample sizes (100 and 500), and various correlations of errors between AR(1) and AR, were used to obtain the data (2). $\rho = -1$ to $+1$. It was noticed that correlated errors a significant effect on heritability estimation and are highly affected by the distribution it follows.

Keywords: AR(1), AR (2), Heritability, One-way classification, MSE

In many real-world situations, observations generated from plant and animal breeding are not always randomly distributed, sometimes correlated observations may occur too. The classical assumption of independence between observations is violated in the presence of highly correlated patterns in the data. The information on genetic components of variances of the important characters is the prime interest of plant and animal breeders. In addition to the variability explained by various causes, it is demonstrated that in the presence of correlated observations, something more from the error components can be organized to provide a truly random error component. To get at an independent or truly random error model, it is necessary to investigate a model in which structural variation owing to error may be further accounted for. As a result, the use and development of statistical techniques are advocated, in which the amount of variability due to various causes at the genetic and phenotypic level can be assessed scientifically, factors can be compared with a high degree of precision, and breeding values can be predicted more efficiently. In a time series, serial correlation occurs when a variable and a lagged version of itself are observed to be correlated with one another over time. The number of immediately preceding observations in an autoregressive series that are used to predict the value at the current observation is defined as the order of the autoregressive models. First-

order and second-order autoregressive models i.e. AR(1) and AR(2) error structures have been considered. Fisher (1925) made a significant addition to variance component models by proposing the analysis of the variance technique of estimation. Durbin and Watson (1950) gave a procedure to check the presence of first-order autocorrelation disturbance in the error term. Diblasi and Bowman (2001) proposed a test statistic and a graphical method, which can assess the evidence for the presence of any spatial correlation in the data. Singh *et al.* (2006) estimated variance components when errors are correlated in half-sib data. Costa *et al.* (2009) estimated genetic parameters of test day fat and protein yields using autoregressive multiple lactation animal models. As a result, development of a statistical approach for estimating genetic parameters when errors are correlated is an important genetic research topic in the field of statistical genetics. The correlation present in the observations is ignored in the classical analysis approach of field experimental data. We cannot ignore the correlation effect when there are significantly correlated trends present in the data. Therefore, it becomes vital to create appropriate statistical techniques for these circumstances. Henderson (1963) was the first to formally develop the general theory of mixed model techniques applied to animal breeding (1963). Using an autoregressive test day multiple lactations (AR) animal model, Zuk *et al.* (2011) studied a technique that is not inflated by genetic interactions for estimating heritability from isolated populations. Bernal-Vasquez *et al.* (2014) used ad hoc method to estimate heritability

Present address: ¹ICAR-Indian Agricultural Statistics Research Institute, New Delhi. ✉Corresponding author email: himadriiasri@gmail.com

in genomic prediction study proposed by Piepho and Möhring (2007). Bernal-Vasquez *et al.* (2016) investigated and assessed several techniques for identifying potential outliers in genomic study and applied to hybrid rye genomic prediction programme.

Paul and Wahi (2016) studied the estimation of heritability under correlated errors. Lourenço *et al.* (2020) investigated robust estimation of heritability and predictive accuracy in plant breeding using simulation and empirical data evaluation. Roy *et al.* (2022) studied the estimation of heritability of Karan Fries cattle using the Bayesian procedure.

Our present investigation was used to develop a framework for the estimation of heritability under the half-sib model using autocorrelated structures. The developed approach was compared using estimation procedures such as ANOVA (Analysis of Variance), ML (Maximum Likelihood), REML (Restricted Maximum Likelihood), and MIVQUE (Minimum Variance Quadratic Unbiased Estimation).

MATERIALS AND METHODS

One of the basic criteria for studying the statistical properties of genetic parameters is to simulate statistical-biological models with known population parameters. We limit ourselves in this study to estimating heritability through sib analysis. Data for correlated and uncorrelated cases are generated using one-way classification model. In the light of the weaknesses (negativity, lack of distributional properties) of ANOVA estimators developed earlier, other approaches namely MLREML, MIVQUE, etc. emerged.

Models for heritability: Two important models, namely one-way and two-way nested models (Half-Sib model) for the estimation of heritability given by Ronningen (1974) are used in this study. A brief discussion of these Monte Carlo methods is as follows:

One Way Classification

The one-way classification or half-sib analysis model can be written as follows:

$$y_{ij} = \mu + s_i + e_j; i=1,2,\dots,s; j=1,2,\dots,p \quad (1)$$

where, y_{ij} is the observed value on the progeny of the j^{th} dam mated to the i^{th} sire; μ is the general mean; s_i is the effect due to i^{th} sire; e_j is the random effect associated with j^{th} progeny of the i^{th} sire.

The simulation model (Ronningen 1974) to generate half-sib model was carried out as follows:

$$y_{ij} = \mu + \sigma_s a'_i + \sigma_e a'_{ij} \quad (2)$$

Where a'_i and a'_{ij} are random standard normal values. The value of heritability is given by

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

σ_s^2 = sire variance components, σ_e^2 = error variance components

Correlated Case

Suppose that sires are independent but within sire,

progenies are correlated. Further, assume that the correlated errors follow AR(1) i.e.,

$$e_j = \rho e_i(j-1) + \eta_j, \eta_j = \text{random error components} \\ \text{where } |\rho| < 1, \text{Var}(\eta_j) = \frac{\sigma_e^2}{1-\rho^2} \text{ and } \eta_j \sim \text{IIDN}(0,1) \text{ for } j > 1$$

Generate e_j using the equation (3). Then, we can generate the correlated observations y_j 's by using the following modified simulation model:

$$y_j = \mu + \sigma_s a'_i + \sigma_e e_j$$

Notations have the same meaning as defined above, and e_j 's are the values generated from equation (3). In the case of AR(2)

$$e_{ij} = \rho_1 e_i(j-1) + \rho_2 e_i(j-2) + \eta_{ij}, \eta_{ij} = \text{random error components}$$

Similarly, we can generate the correlated data for different error structures other than AR(1) e.g., AR(2), the function of a distance, etc.

RESULTS AND DISCUSSION

Data was generated using different heritability ranges i.e. high and low (0.5, 0.1), Half-sib AR(1):different sample size 100 and 500 and different correlation of errors AR(1) and AR(2). $\rho = -1$ to $+1$. The simulated observations were studied and variance components (Searle 1992) were estimated using SAS *varcomp*proc. ANOVA, ML, REML, and MIVQUE methods were used. Estimates of heritability were obtained using the above formulae are given in different tables.

Half-sib estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for the different parametric values of heritability: The data was generated from a population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE (Means Square Error) were obtained and shown in Table 1. It was seen that value of estimates of heritability changed from negative to positive when ρ changes from -1 to $+1$. MSE value decreased up to $\rho=0$, then again increased ρ is positive. For $\rho = -1$ to $\rho = -0.5$, the estimate of heritability is 0 in the case of ML, REML, and MIVQUE methods and MSE values are not changing. Estimated value of heritability increased from $\rho = -0.4$ to $\rho = 1$. MSE values showed the same trend. With increasing sample sizes, it was noticed that the MSE values decreased. Half-sib estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes in case heritability of 0.10 are presented in Supplementary Table 1.

Half-sib estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes for the different parametric value of heritability: The data was generated from a population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE (Mean Square Error) were obtained and shown in Supplementary Table 2 and Table 3. It was noticed that in the case of AR(2), if fixing AR(1) value changes AR(2) values, the MSE value decrease with increasing the correlation value in general.

Table 1. *Half-sib* estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10

r		Methods							
		ANOVA		ML		REML		MIVQUE	
		P=100	P=500	P=100	P= 500	P=100	P=500	P=100	P=500
-1	MSE	0.2293	0.2311	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
	(ht)	-0.3811	-0.3830	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.9	MSE	0.2041	0.2067	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
	(ht)	-0.3539	-0.3566	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.8	MSE	0.1731	0.1764	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
	(ht)	-0.3177	-0.3215	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.7	MSE	0.1416	0.1454	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
	(ht)	-0.2773	-0.2819	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.6	MSE	0.112	0.116	0.010	0.010	0.010	0.010	0.010	0.010
	(ht)	-0.2343	-0.2399	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.5	MSE	0.0845	0.0887	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
	(ht)	-0.189	-0.1954	0	0	0	0	0	0
-0.4	MSE	0.0601	0.0642	0.0095	0.0095	0.0094	0.0094	0.0094	0.0094
	(ht)	-0.1408	-0.1481	0.0004	0.0004	0.0005	0.0005	0.0005	0.0005
-0.3	MSE	0.0392	0.0642	0.0092	0.0095	0.0091	0.0094	0.0091	0.0094
	(ht)	-0.0889	-0.1481	0.0024	0.0004	0.0032	0.0005	0.0032	0.0005
-0.2	MSE	0.0226	0.0258	0.0078	0.0080	0.0076	0.0079	0.0076	0.0079
	(ht)	-0.0320	-0.0408	0.0163	0.0149	0.0186	0.0171	0.0186	0.0171
-0.1	MSE	0.0118	0.0142	0.0065	0.0068	0.0064	0.0068	0.0064	0.0068
	(ht)	0.0314	0.0220	0.0465	0.0430	0.0513	0.0476	0.0513	0.0476
0	MSE	0.0095	0.0109	0.0078	0.0078	0.0083	0.0083	0.0083	0.0083
	(ht)	0.1037	0.0939	0.1006	0.0958	0.1082	0.1031	0.1082	0.1031
0.1	MSE	0.0201	0.0204	0.0180	0.0178	0.0200	0.0197	0.0200	0.0197
	(ht)	0.1879	0.1779	0.1792	0.1715	0.1887	0.1806	0.1887	0.1806
0.2	MSE	0.0515	0.0506	0.0472	0.0464	0.0515	0.0506	0.0515	0.0506
	(ht)	0.2884	0.2784	0.2775	0.2681	0.2884	0.2787	0.2884	0.2787
0.3	MSE	0.1176	0.1157	0.1096	0.1079	0.1176	0.1157	0.1176	0.1157
	(ht)	0.4114	0.4018	0.3991	0.3896	0.4114	0.4018	0.4114	0.4018
0.4	MSE	0.2434	0.2409	0.2301	0.2279	0.2434	0.2409	0.2434	0.2409
	(ht)	0.5659	0.5570	0.5520	0.5432	0.5659	0.5570	0.5659	0.5570
0.5	MSE	0.4753	0.4733	0.4542	0.4524	0.4753	0.4733	0.4753	0.4733
	(ht)	0.7647	0.7571	0.7490	0.7415	0.7647	0.7571	0.7647	0.7571
0.6	MSE	0.8976	0.8980	0.8651	0.8656	0.8976	0.8980	0.8976	0.8980
	(ht)	1.0252	1.0195	1.0077	1.0021	1.0252	1.0195	1.0252	1.0195
0.7	MSE	1.6534	1.6585	1.6050	1.6103	1.6534	1.6585	1.6534	1.6585
	(ht)	1.3663	1.3632	1.3472	1.3442	1.3663	1.3632	1.3663	1.3632
0.8	MSE	2.9268	2.9386	2.8600	2.8719	2.9268	2.9386	2.9268	2.9386
	(ht)	1.7950	1.7948	1.7752	1.7750	1.7950	1.7948	1.7950	1.7948
0.9	MSE	4.7491	4.7670	4.6678	4.6858	4.7498	4.7670	4.7491	4.7670
	(ht)	2.268	2.2703	2.2491	2.2515	2.268	2.2703	2.268	2.2703
1	MSE	6.5406	6.5607	6.4551	6.4753	6.5406	6.5607	6.5406	6.5607
	(ht)	2.6502	2.6538	2.6333	2.6370	2.6502	2.6538	2.65024	2.6538

r, Correlation coefficient; ANOVA, Analysis of Variance; ML, Maximum Likelihood; REML, Restricted Maximum Likelihood; MINQUE, Minimum Variance Quadratic Unbiased Estimation; MSE, Mean squared error; P, Sample size.

Sometimes haphazard trends are noticed. We found some good combination of AR(1) and AR(2) values (-0.6,0.4), (-0.5,-0.1), (0, 0), (0.1,0.1) and (0.1, 0.5) combinations gave better estimates of heritability. With increasing sample sizes, the MSE values decreased.

Data were generated using different heritability values

i.e. high, moderate, and low (0.5, 0.25, 0.1) *Half-sib* AR(1): and different sample size 100 and 500 and different correlation of errors A(1) and AR(2) i.e. in the range $\rho=-1$ to $+1$. After generating the data, variance components were estimated using SAS ProcVarcomp. ANOVA, ML, REML, and MIVQUE methods were used. *Half-sib* estimate of

Table 2. *Half-sib* estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for the parametric value of heritability 0.10 when the sire is distributed as Normal but error follows a Beta distribution

Heritability (H_t)	Correlation Coefficient (RC)	Estimated Heritability	M.S.E
0.1	-0.7	-0.779	0.780
		0.000	0.009
		0.000	0.009
		0.000	0.009
0.1	-0.3	-0.323	0.215
		0.004	0.005
		0.006	0.007
		0.006	0.007
0.1	0	0.099	0.049
		0.095	0.018
		0.107	0.020
		0.107	0.020
0.1	0.3	0.578	0.287
		0.554	0.263
		0.579	0.287
		0.579	0.287
0.1	0.7	1.315	1.532
		1.288	1.467
		1.315	1.532
		1.315	1.532

heritability and MSE values in case of correlated errors (AR(1)) and AR(2) different sample sizes for the parametric value of heritability 0.10, 0.25, 0.5 for different sample sizes (100 and 500). The expected mean sum of squares due to error is overestimated when the correlation is negative and they increase as the degree of correlation increases. But the expected mean sum of the square is underestimated if errors are positively correlated and they decrease with an increase in the degree of correlation and approach to 0 as ρ tends to unity. On the other hand, just reverse results were obtained for estimating the mean sums of squares due to sire i.e. expected mean sum of squares is underestimated when ρ is negative and they are overestimated if the correlation is positive. As ρ tends to unity, the expected mean sum of squares due to sire approaches to its maximum value.

Heritability values are overestimated if the correlation is positive. The same trend follows for all levels of heritability. Also, heritability increases from zero to nearly four as autoregressive coefficients increase from minus unity to approximate unity. with and without fixed effects are obtained. The value of estimates of heritability changes negative to positive when ρ changes from -1 to +1. MSE value decreases up to $\rho=0$, then again increasing ρ is positive. Up to $\rho=-1$ to $\rho=-0.5$, the estimate of heritability is 0 in the case of ML, REML, and MIVQUE methods and MSE values are not changing. Estimated value of heritability increased from $\rho=-0.4$ to $\rho=1$. MSE values showed the same trend. With increasing sample sizes, the MSE values decreased.

Table 3. *Half-sib* estimate of heritability and MSE values in case of Correlated errors (AR(1)) and different sample sizes for the parametric value of heritability 0.10 when the sire is distributed as Normal but error follows Cauchy distribution

Heritability (H_t)	Correlation Coefficient (RC)	Estimated Heritability	M.S.E
0.3	-0.7	-0.740	1.196
		0.000	0.089
		0.000	0.089
		0.000	0.089
0.3	-0.3	-0.427	0.589
		0.005	0.090
		0.006	0.090
		0.006	0.090
0.3	0	0.015	0.095
		0.035	0.076
		0.044	0.072
		0.044	0.072
0.3	0.3	-0.427	0.589
		0.005	0.090
		0.006	0.090
		0.006	0.090
0.3	0.7	1.390	1.634
		1.365	1.579
		1.390	1.634
		1.390	1.634

In the case of AR(2) if fixing AR(1) value by changing AR(2) values, the MSE value decrease with increasing the correlation value in general. Sometimes haphazard trends are noticed. We found some good combination of AR(1) and AR(2) values (0, 0), (0.1, 0.1) and (0.1, 0.5) combinations giving better estimates of heritability. With increasing sample sizes, the MSE values decreased.

We may infer that different correlation patterns, such as AR(1), AR(2), and so on, have an impact on heritability estimates and Means Square Error values. Different combinations are identified for better performance in certain instances. Different combinations of distributions, such as normal, beta, Cauchy, and t-distribution with different heritability values, and estimate of heritability and MSE values obtained by four different methods, such as ANOVA, ML, REML, and MIVQUE methods with different parametric values of heritability, are used to generate sire component and error component. From Tables 2 to 3, it can be seen that when the correlation value -ve approaches zero, the MSE values decline.

Half-sib estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for the parametric value of heritability 0.10 when the sire is distributed as t-distribution but error follows t-distribution is given in supplementary Table 4. *Half-sib* estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for the parametric value of heritability 0.90 when sire and error both follow Normal distribution is depicted in supplementary Table

5. Supplementary Table 6 presents *half-sib* estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for the parametric value of heritability 0.90 when the sire is distributed as Cauchy distribution but error follows Cauchy distribution. The MSE values for all choices of distribution grew as we climbed from zero to higher values, i.e. closer to +1.

It can be concluded that there is a significant effect on heritability of correlated error, sire and error follow a variety of different distributions, like normal, Cauchy, beta, and *t*-distributions in the estimation procedure. A one-way classification model is used to generate observations for correlated and uncorrelated cases with respect to two error structures, AR(1) and AR(2). The data were generated using a variety of heritability ranges, including high and low (0.5, 0.1), Half-sib AR(1), different sample sizes (100 and 500), and different correlations of errors between AR(1) and AR (2). $\rho = -1$ to $+1$. Correlated errors are observed to significantly affect heritability estimation and to be greatly influenced by the distribution they follow.

REFERENCES

- Bernal-Vasquez A M, Möhring J, Schmidt M, Schönleben M, Schön C C and Piepho H P. 2014. The importance of phenotypic data analysis for genomic prediction-a case study comparing different spatial models in rye. *BMC Genomics* **15**(1): 1–17.
- Bernal-Vasquez A M, Utz H F and Piepho H P. 2016. Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theoretical and Applied Genetics* **129**(4): 787–804.
- Costa C N, Carvalheira J A, Cobuci A F, Freitas A F and Thompson G. 2009. Estimation of genetic parameters of test day fat and protein yields in Brazilian Holstein cattle using an autoregressive multiple lactation animal model. *South African Journal of Animal Science* **39**: 165–68.
- Dibiasi A and Bowman A W. 2001. On the use of the variogram in checking for independence in spatial data. *Biometrics* **57**: 211–18.
- Durbin J and Watson G S. 1950. Testing for serial correlation in least squares regression. *Biometrika* **37**: 409–28.
- Henderson C R. 1963. Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*. NAS-NRC 982, Washington DC, 141–163.
- Lourenço V M, Ogutu J O and Piepho H P. 2020. Robust estimation of heritability and predictive accuracy in plant breeding: evaluation using simulation and empirical data. *BMC Genomics* **21**(1): 1–18.
- Piepho H P and Möhring J. 2007. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* **177**(3): 1881–88.
- Ronningen K. 1974. Monte-Carlo simulation of statistical biological models which are of interest in animal breeding. *Acta Agriculturae Scandinavica* **24**: 135–42.
- Roy H S, Paul A K, Paul R K, Singh R K, Kumar P and Yeasin M. 2022. Estimation of heritability of Karan Fries cattle using Bayesian procedure. *Indian Journal of Animal Sciences* **92**(5): 645–48.
- Searle S R, Casella G and McCulloch C E. 1992. *Variance Components*. New York, John Wiley.
- Singh N O, Bhatia V K and Paul A K. 2006. Estimation of variance components when errors are correlated by autoregressive of order one. *Journal of Indian Society of Agricultural Statistics* **60**(2): 126–36.
- Estaghrvirou S B O, Ogutu J O and Piepho H P. 2014. Influence of outliers on accuracy and robustness of methods for genomic prediction in plant breeding. *G3* **4**: 2317–28.
- Fisher R A. 1925. *Statistical Methods for Research Workers*, 1st edn. Oliver & Boyd, Edinburgh and London.
- Paul A K and Wahi S D. 2016. *Estimation of heritability under correlated errors*. Project Report. ICAR-IASRI publication.
- Zuk O, Hechter E, Sunyaev S R and Lander E S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* **109**(4): 1193–98.