

Table S4 Descriptive statistics for the difference between the heritability estimated using datasets with $(\hat{r}_{g,\hat{g},o}^2)$ and without $(\hat{r}_{g,\hat{g}}^2)$ outliers, taken as the benchmark, for the five methods (M1 to M5) in each of the 10 scenarios.

Scenario		#M1(10)	M2(11)	M3(12)	M4(13)	M5(15)			
		$\hat{H}_{m_1}^2$	$\hat{H}_{m_2}^2$	$\hat{H}_{m_3}^2$	$\hat{H}_{m_4}^2$	$\hat{H}_{m_5}^2$			
Nr	#Gen	Marker effect variance	Outliers	Statistics					
1	177	0.2019	5 σ	Mean	-0.569 ^c	-0.463 ^b	-0.464 ^b	0.004 ^a	0.003 ^a
				Std	0.002	0.002	0.002	0.000	0.000
2	177	0.2019	8 σ	Mean	-0.583 ^d	-0.443 ^b	-0.479 ^c	-0.002 ^a	-0.001 ^a
				Std	0.002	0.002	0.002	0.001	0.001
3	177	0.2019	10 σ	Mean	-0.595 ^d	-0.439 ^b	-0.494 ^c	-0.008 ^a	-0.006 ^a
				Std	0.002	0.002	0.002	0.001	0.001
4	177	0.2019/10	5 σ	Mean	-0.006 ^a	-0.009 ^{ab}	-0.011 ^b	-0.006 ^a	-0.016 ^c
				Std	0.001	0.001	0.001	0.000	0.001
5	177	0.2019/10	8 σ	Mean	-0.012 ^a	-0.019 ^b	-0.022 ^b	-0.010 ^a	-0.029 ^c
				Std	0.001	0.002	0.002	0.001	0.002
6	177	0.2019/10	10 σ	Mean	-0.016 ^a	-0.027 ^b	-0.031 ^b	-0.014 ^a	-0.038 ^c
				Std	0.001	0.002	0.002	0.001	0.002
7	698	0.005892	5 σ	Mean	0.087 ^c	0.091 ^b	0.164 ^a	0.024 ^e	0.013 ^e
				Std	0.001	0.001	0.003	0.001	0.000
8	698	0.005892	10 σ	Mean	0.075 ^a	0.076 ^a	0.074 ^a	0.011 ^b	0.006 ^c
				Std	0.001	0.001	0.001	0.001	0.000
9	698	0.005892/10	5 σ	Mean	-0.004 ^a	-0.002 ^a	-0.004 ^a	-0.002 ^a	-0.004 ^a
				Std	0.000	0.000	0.000	0.000	0.000
10	698	0.005892/10	10 σ	Mean	-0.011 ^b	-0.010 ^b	-0.011 ^b	-0.006 ^a	-0.014 ^b
				Std	0.001	0.001	0.001	0.000	0.000

For each of the 1000 datasets heritability was estimated using five methods for each scenario (1 to 10). Means for pairs of methods within each scenario with the same superscript letter are not significantly different at the 5% level of significance based on the *t*-test. #The number of the equation used in the text is in parenthesis. Methods 1 to 4 use cross-validation but Method 5 does not. σ is the standard deviation of the residual error. Note that positive differences denote overestimation whereas negative differences denote underestimation of the estimated heritability using datasets without outliers.