



**Figure S2** Studentized residuals for yield for the large data set ( $n=698$  genotypes) contaminated with an outlier equal to five times the standard deviation of the residual error used to simulate the large datasets.