Supplementary Material for Smooth-threshold multivariate genetic prediction with unbiased model selection

Masao Ueki[*], Gen Tamiya[†]

and for Alzheimer's Disease Neuroimaging Initiative[‡]

1 Smooth-threshold multivariate genetic prediction in linear multiple regression model

First, we present our smooth-threshold multivariate genetic prediction in linear multiple regression model, $y = \mu + \epsilon$, where $\mu = Ey = X\beta$, $\epsilon \sim$

[*] uekimrsd@nifty.com. Biostatistics Center, Kurume University, 67 Asahi-Machi, Kurume, Fukuoka 830-0011, Japan.

[†] gtamiya@genetix-h.com. Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryo-Machi, Aoba-Ku, Sendai 980-8573, Japan.

[‡] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

$N(0, \sigma^2 I_n)$, X is a p-dimensional design matrix and β is the corresponding p regression coefficients. Here and throughout E represents the conditional expectation of y given X.

1.1 Linear multiple regression after marginal association screening

The gene score method (Purcell et al., 2009) and its multivariate generalization (Warren et al., 2013) use upper-ranked SNPs in marginal association. Given a cutoff value $t > 0$, linear multiple regression after marginal association screening uses X_j satisfying $T_j(y, X) > t$ in fitting multiple regression model. Here $T_j(y, X)$ denotes a test statistic for marginal association, taking nonnegative value. The cutoff value t corresponds to quantile of null distribution of $T_j(y, X)$ as in hypothesis test. Without loss of generality, assume that a large value of $T_j(y, X)$ indicates stronger marginal association. Examples of $T_j(y, X)$ include the squared Pearson's correlation and the F-statistics. Multiple regression after marginal association screening can be expressed by

$$\begin{aligned} \hat{\mu}_i &= X_i^T \hat{\beta}, \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_A \\ \hat{\beta}_{A^c} \end{pmatrix} = \begin{pmatrix} (X_A^T X_A)^{-1} X_A^T y \\ 0 \end{pmatrix}, \\ A &= \{j \in M : T_j(y, X) > t\}. \end{aligned} \tag{1}$$

Note that the above procedure is similar to the sure independence screening (Fan and Lv, 2008) which uses predictor variables upper-ranked in marginal association. The procedure (1) is feasible for $p \gg n$ data and is useful in building predictive model.

1.2 Stein's unbiased risk estimation and generalized degrees of freedom

Predictive power largely depends on predictive model choice. We consider unbiased model selection criterion such as the Mallows' C_p and Akaike information criterion (AIC, Akaike, 1973) which are computationally efficient alternatives to cross-validation. Those criteria attempt to correct the bias in residuals of squared, or apparent error, $(y_i - \hat{\mu}_i)^2$, from the squared prediction error $(y_{0,i} - \hat{\mu}_i)^2$, in which $y_{0,i}$ is an independent future observation from the same distribution of y_i given X_i. In other words, residuals become optimistic because the data y are used twice for training (i.e. building predictive model) and testing (i.e. evaluating predictive power of the model). We attempt to select an optimal threshold t in (1) from a model selection perspective. If A is deterministic with $|A| < n$, and $y_i \sim N(\mu_i, \sigma^2)$, the following well-known unbiasedness holds:

$$E\left\{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2 + 2\sigma^2|A|\right\} = \sum_{i=1}^{n}\left[E\{(\mu_i - \hat{\mu}_i)^2\} + \sigma^2\right], \qquad (2)$$

which leads to the Mallows' C_p criterion. However, screening may violate the unbiasedness property (2) since the screening process depends on y and hence A is no longer deterministic (See Section 4 in this Supplementary Material).

For a general modeling process $\hat{\mu} : y \mapsto \hat{\mu}(y)$, a generalization of (2) is given as follows:

$$E\left[\sum_{i=1}^{n}\{(y_i - \hat{\mu}_i)^2 + 2\sigma^2 \mathrm{cov}(\hat{\mu}_i, y_i)\}\right] = \sum_{i=1}^{n}\left[E\{(\mu_i - \hat{\mu}_i)^2\} + \sigma^2\right], \qquad (3)$$

where $\mathrm{cov}(\hat{\mu}_i, y_i)$ is referred to as a covariance penalty (Efron, 2004). If A were deterministic, $\mathrm{cov}(\hat{\mu}_i, y_i)$ reduces to the degrees of freedom $|A|$, and hence (3) coincides with (2). To apply (3), an unbiased estimator of $\mathrm{cov}(\hat{\mu}_i, y_i)$ is needed, but no readily available such estimator is known for (1). Another concern is the selection bias in regression coefficient estimates produced by screening, which is referred to as the winner's curse effect. Discontinuity in y due to screening may also cause instability in prediction (Breiman, 1996).

For differentiable $\hat{\mu}$ in y, the Stein's lemma (Stein, 1981) gives a convenient formula

$$\mathrm{cov}(\hat{\mu}_i, y_i) = E\{\partial_i \hat{\mu}_i(y)\},$$

in which $\partial_i = \partial/\partial y_i$. The above formula leads to the generalized degrees of freedom (GDF) (Ye, 1998),

$$\mathrm{GDF} = \sum_{i=1}^{n} \partial_i \hat{\mu}_i(y), \qquad (4)$$

which allows Stein's unbiased risk estimation (SURE), namely,

$$E\left[\{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2\} + 2\sigma^2 \mathrm{GDF}\right] = \sum_{i=1}^{n}\left[E\{(\mu_i - \hat{\mu}_i)^2\} + \sigma^2\right]. \qquad (5)$$

In the following, we propose a smoothed version of marginal association screening of (1). It is expected that the continuity in y makes the resulting prediction stable (Breiman, 1996). As a consequence, a C_p-type model selection criterion can be obtained for data-dependent choice of an optimal marginal association cutoff t.

1.3 Smooth-threshold multivariate genetic prediction

Here we introduce the smooth-thresholding which replaces the indicator function appeared in screening process by a continuous function (Ueki, 2009; Ueki and Kawasaki, 2011). In view of the normal equations, it can be seen that $\hat{\beta}$ in (1) satisfies, for $j \in M$,

$$(1 - \hat{D}_j)\{X_j^T(X\hat{\beta} - y)\} + \hat{D}_j\hat{\beta}_j = 0, \tag{6}$$

or, in vector form,

$$(I_p - \hat{D})\{X^T(X\hat{\beta} - y)\} + \hat{D}\hat{\beta} = 0,$$

where $\hat{D}_j = 1_{\{T_j(y,X) \leq t\}}$ and $\hat{D} = \mathrm{diag}(\hat{D}_j : j \in M)$. Obviously, for $j \in A^c$, $\hat{D}_j = 1$ and (6) reduces to $\hat{\beta}_j = 0$, i.e. sparse solution; for $j \in A$, $\hat{D}_j = 0$ and the above normal equation reduces to $X_A^T(X_A\hat{\beta}_A - y) = 0$ because $\hat{\beta}_{A^c} = 0$. This is the normal equation for the ordinary least-squares regression with design matrix X_A. The resulting prediction process forms $\hat{\mu}(y) = X_A\hat{\beta}_A =$

$X_A(X_A^T X_A)^{-1}X_A^T y$, which is discontinuous function in y due to thresholding in $\hat{D}_j$.

To smooth the prediction process $\hat{\mu}$, the expression (6) suggests to replace the indicator function $\hat{D}_j$ by a smooth function. Following to Ueki (2009), $\hat{D}_j = 1_{\{T_j(y,X)\leq t\}}$ is replaced by an adaptive lasso smooth-thresholding function

$$\check{D}_j = \min[1, \{t/T_j(y,X)\}^{\frac{1+\gamma}{2}}], \tag{7}$$

where $\gamma > 0$ is a tuning parameter. This smooth-thresholding function is chosen so as to be identical to the adaptive lasso estimator under the simplest least-squares regression of $y = \beta + \epsilon$ (Ueki, 2009). Figure S1 depicts the smooth-thresholding function and the resulting estimator in this simplest setting. If $T_j(y,X) \leq t$ or $j \in A^c$, then $\check{D}_j = 1$ giving zero regression coefficient; if $T_j(y,X) > t$ or $j \in A$, $\check{D}_j < 1$ giving nonzero regression coefficient. Therefore, the condition for sparse solution with $\check{D}_j$ is the same as that with $\hat{D}_j$. Note that $\check{D}_j$ is monotonically decreasing in $T_j(y,X)$, making the regression coefficients having large $T_j(y,X)$ less penalized than the regression coefficients having small $T_j(y,X)$. From the fact that the winner's curse effect produces larger selection bias for small regression coefficient (Zhong and Prentice, 2008), it is expected that the above feature of penalization decreases the selection bias.

Our proposed estimation equations where $\hat{D}$ is replaced by $\check{D}$ are as follows:

$$(I_p - \check{D})\{X^T(X\check{\beta} - y)\} + \tau\check{D}\check{\beta} + \lambda(I_p - \check{D})\check{\beta} = 0, \tag{8}$$

in which the solution is denoted by $\check{\beta}$. Here $\lambda > 0$ is a small constant to ensure invertibility of matrix as in ridge penalization (See (10)). We empirically found that introducing an additional tuning parameter $\tau > 0$ which controls the extent of penalization for coefficients for A improves prediction performance. Recalling that $A = \{j \in M : T_j(y, X) > t\}$, the resulting regression coefficients $\check{\beta}$ are written explicitly as follows:

$$\begin{pmatrix} \check{\beta}_A \\ \check{\beta}_{A^c} \end{pmatrix} = \begin{pmatrix} \check{G}_A(I_{|A|} - \check{D}_A)X_A^T y \\ 0 \end{pmatrix}, \tag{9}$$

and the prediction of y_i turns out to be $\check{\mu}_i(y) = X_i^T\check{\beta}$. Here $\check{G}_A = \{(I_{|A|} - \check{D}_A)(\Sigma_{AA} + \lambda I_{|A|}) + \tau\check{D}_A\}^{-1}$, $\Sigma = X^TX$, and $\Sigma_{AA} = (\Sigma_{jk})_{j\in A, k\in A}$. Alternatively, from (8), the regression coefficient for the screened set in (9), $\check{\beta}_A$, can be considered as a solution to

$$X_A^T(X_A\check{\beta}_A - y) + W_A\check{\beta}_A = 0, \tag{10}$$

with $W_A = \text{diag}(W_j : j \in A)$ where $W_j = \lambda + \tau\check{D}_j/(1 - \check{D}_j)$, which is equivalent to the following generalized ridge regression problem:

$$\min_{\beta_A}\left\{||y - X_A\beta_A||^2 + \sum_{j\in A}\beta_j^2 W_j\right\}.$$

Ridge weight for each predictor variable, W_j, represents uncertainty of marginal association screening, which is used for penalizing each regression coefficient. If the marginal association is very weak, we have $\check{D}_j \approx 1$ and large W_j, then the corresponding regression coefficient is strongly shrunken towards zero. If the marginal association is strong, we have $\check{D}_j \approx 0$ and $W_j \approx \lambda$, then the corresponding regression coefficient is less penalized.

1.4 GDF for smooth-threshold multivariate genetic prediction

The Stein's lemma is now applicable to $\check{\mu}_i(y)$ to obtain a closed-form formula for GDF as follows.

Proposition 1 *The GDF for $\check{\mu}(y)$ (4) is equal to*

$$\sum_{i=1}^{n} X_i^T \partial_i \check{\beta} = \mathrm{tr}(\check{G}_A \check{L}_A^T X_A) + \mathrm{tr}\{\check{G}_A (I_{|A|} - \check{D}_A)\Sigma_{AA}\}, \tag{11}$$

where $\check{L}_A$ is an $n \times |A|$ matrix whose (i,j)-element is $(\partial_i \check{D}_j)\check{m}_j$. Here,

$$\partial_i \check{D}_j = -\frac{1+\gamma}{2}\frac{\partial_i T_j(y,X)}{T_j(y,X)}\{t/T_j(y,X)\}^{\frac{1+\gamma}{2}},$$

and $\check{m}_j$ is the jth component of $\check{m} = \{X^T X - (\tau - \lambda)I_p\}\check{\beta} - X^T y = -X^T(y - X\check{\beta}) - (\tau - \lambda)\check{\beta}$.

It is shown later that this is a special case of Proposition 2, and the derivation is omitted. If $\check{D}_A \approx O$, then $\check{G}_A \approx (X_A^T X_A)^{-1}$, and the second term in the right-hand side of (16) reduces to

$$\mathrm{tr}\{\check{G}_A(I_{|A|}-\check{D}_A)X_A^TX_A\} \approx \mathrm{tr}\{X_A(X_A^TX_A)^{-1}X_A^T\} = \mathrm{rank}(X_A),$$

i.e., the usual degrees of freedom. On the other hand, the first term $\mathrm{tr}(\check{G}_A\check{L}_A^TX_A)$ represents the effect of screening which does not appear in the multiple regression with deterministic A.

If one includes unpenalized components in X such as intercept or covariates such as sex, age and body mass index, the formula of GDF is still valid by setting $\partial_i\check{D}_j = 0$ for index j corresponding to these unpenalized components.

1.5 F-test screening

In SNP-GWAS, F-test can be used for testing association for quantitative traits. Covariates, Z say, can be adjusted for in the F-test. Inclusion of intercept only corresponds to Z being an n-vector of ones. F-statistic for jth SNP is

$$T_j(y,X) = \frac{(n-d)y^T(P_{(Z,X_j)}-P_Z)y}{y^T(I_n-P_{(Z,X_j)})y} = \frac{(n-d)y^TP_{\tilde{X}_j}y}{y^T(I_n-P_Z-P_{\tilde{X}_j})y},$$

where $d=\mathrm{rank}(Z)$, P_X denotes the projection onto column space of X and $\tilde{X}_j = (I_n-P_Z)X_j$. Threshold t is the $(1-\alpha)$th quantile of $F(1,n-d)$-distribution, $q_{1-\alpha}$ say. Instead of the cutoff value t for $T_j(y,X)$, the significance level α is considered as a tuning parameter to be optimized. There is a one-to-one correspondence between α and t.

1.6 C_p-type criterion for smooth-threshold multivariate genetic prediction

Using the expression (19) in Section 3 in this Supplementary Material, the GDF for F-test screening is computed by applying Proposition 1. Denoting the dependency of $\check{\mu}$ on α explicitly, using (5), the proposed unbiased C_p-type model selection criterion is as follows:

$$C(\alpha) = \sum_{i=1}^{n} \{y_i - \check{\mu}_i(\alpha)\}^2 + 2\sigma^2 \mathrm{GDF}(\alpha).$$

An optimal α is obtained by minimizing the above quantity within a range of α for search. From (5), the expectation of the above criterion equals to $\sum_{i=1}^{n}[E\{(\mu_i - \hat{\mu}_i)^2\} + \sigma^2]$, and hence unbiased property holds. Therefore, the selection bias in RSS due to screening is accounted for, and the model selection with C_p-type criterion is expected to work properly. In principle, the two tuning parameters (γ and τ) other than α may be selected by the above criterion. However, from our experiences through simulations as well as real data applications, we suggest using fixed γ and τ to achieve stable predictive power. In particular, we suggest fixed $\gamma = 1$ and $\tau = n/\sqrt{\log n}$ throughout. As a consequence, we consider single tuning parameter α to be optimized.

In univariate case in Figure S1, the adaptive lasso smooth-thresholding yields the adaptive lasso solution when $\tau = 1$ (Ueki, 2009), and hence, we interpret γ as that in the adaptive lasso. Existing works (Zou, 2006; Ueki, 2009) examined a few candidate parameters (e.g. $\gamma \in \{0.5, 1, 2\}$). Some literature (e.g. Bühlmann and van de Geer (2011)) defines the adaptive lasso by $\gamma = 1$. We also confirmed that, through simulations, $\gamma = 1$ generally gives good performance. On the other hand, for τ parameter, Ueki (2009) used $\tau = 1$ for smooth-threshold estimating equation in $p \ll n$ setting. In the genetic prediction in $p \gg n$ setting, through simulation studies, we observed that $\tau > 1$ was needed for stable prediction and estimation of the C_p-type criterion. (One can use cross-validation instead of the C_p-type criterion for tuning parameter selection if C_p estimation fails.) We consider that ultrahigh-dimensionality required stronger penalization. A role of increasing τ on the resultant estimate can be understood by the right panel of Figure S1. Examining various candidate values for τ through simulation studies and real data applications, we found that $\tau = n/\sqrt{\log n}$ works well. We also found that $\tau = n\omega$ with some $\omega \in (0, 1)$ generally gives good performance where ω is a constant not close to the boundaries of $(0, 1)$. For high-dimensional data other than SNP-GWAS, there is a possibility that other choice of τ is appropriate. Further works are needed on the choice of τ.

The C_p-type criterion contains σ^2 which is often unknown. In such cases, according to Theorem 3 of Ye (1998), the following surrogate of σ^2 can be used,

$$\hat{\sigma}_1^2 = \frac{1}{n - \mathrm{GDF}(\alpha_1)} \sum_{i=1}^{n} \{y_i - \check{\mu}_i(\alpha_1)\}^2,$$

where α_1 is a pre-specified threshold α that gives sufficiently complex model. From simulation studies and real data applications, we found that $\alpha_1 = 3n/(p \log n)$ works well. If all test statistics $T_j(y, X)$ follows their null distribution, the expected number of screened predictors at $\alpha_1 = 3n/(p \log n)$ is $3n/\log n$, which is the same order of the pre-specified number of screened variables in sure independence screening (Fan and Lv, 2008; Fan et al., 2009).

2 Smooth-threshold multivariate genetic prediction in generalized linear models

We give an extension of the above result for linear multiple regression to generalized linear models. It includes logistic regression as a special case. Suppose that n response variables $y = (y_1, \ldots, y_n)^T$ follow independently from the generalized linear model $p(y_i; \theta_i, \phi) = \exp[\{y_i\theta_i - b(\theta_i)\}/a(\phi) + c(y_i, \phi)]$, where the canonical parameter θ_i, the dispersion parameter ϕ and $a(\cdot)$, $b(\cdot)$ and $c(\cdot, \cdot)$ are scalar-valued functions specific to the corresponding model. Throughout, a linear regression model with the canonical link $b(\mu_i) = \theta_i = X_i\beta$ is considered, where $\mu_i = Ey_i$, or in vector forms, $b(\mu) = \theta = X\beta$ and $\mu = Ey$. Then $\mu(\theta) = b'(\theta)$. Let the $-2\times$ loglikelihood function excluding the term that is constant regardless of the models be $q(y; t) =$

$-2\{yt - b(t)\}/a(\phi)$. For a modeling process $\hat{\theta} : y \mapsto \hat{\theta}(y)$, as in the case of the sum of squared residuals, the apparent error $Q(y; \hat{\theta}) = \sum_{i=1}^{n} q\{y_i; \hat{\theta}_i(y)\}$ is biased from $Q(y_0; \hat{\theta}) = \sum_{i=1}^{n} q\{y_{0,i}; \hat{\theta}_i(y)\}$ where $y_{0,i}$ is an independent future observation from the same distribution of y_i given X_i. Specifically, the following optimism theorem (Efron, 2004; Augugliaro et al., 2013) holds:

$$EQ\{\mu; \hat{\theta}(y)\} = E[Q\{y; \hat{\theta}(y)\} + 2\Omega_\phi],$$

where $\Omega_\phi = a(\phi)^{-1}\text{tr}[\text{cov}\{\hat{\theta}(y), y\}]$. Hence, having an unbiased estimator of Ω_ϕ, an unbiased estimator of $EQ\{\mu; \hat{\theta}(y)\}$ can be obtained from the apparent error, and model selection can be carried out. However, the covariance in Ω_ϕ depends on $\mu = Ey$ which is unknown. Following to Augugliaro et al. (2013), one-term Taylor approximation of $\hat{\theta}(y)$ around $\mu = Ey$ is applied:

$$\Omega_\phi \approx a(\phi)^{-1}\text{tr}[\text{cov}\{\hat{\theta}(\mu) + \partial\hat{\theta}(\mu)^T(y - \mu), y\}] = a(\phi)^{-1}\text{tr}\{\partial\hat{\theta}(\mu)^T\text{var}(y)\},$$

where $\partial = (\partial_i)$. In the generalized linear model, $\text{var}(y) = a(\phi)\text{diag}[b''\{\theta_i(\mu)\}]$. Unknown μ in $\partial\hat{\theta}(\mu)$ and in $b''\{\theta_i(\mu)\}$ are replaced, respectively, by y and by the sufficiently complex estimator, $\hat{\theta}_1$ say. The above approximation is similar to that in Augugliaro et al. (2013) who used the maximum likelihood estimator (MLE) on a saturated model. As a consequence, an approximate unbiased estimator of Ω_ϕ for the modeling process $\hat{\theta} : y \mapsto \hat{\theta}(y)$ is as follows:

$$\mathrm{GDF} = \sum_{i=1}^{n} b''(\hat{\theta}_{1,i}) X_i \partial_i \hat{\beta}, \tag{12}$$

which we refer to as the GDF for the generalized linear model. We expect that the following analogous identity to (5) approximately holds:

$$EQ\{\mu; \hat{\theta}(y)\} \approx E[Q\{y; \hat{\theta}(y)\} + 2\mathrm{GDF}]. \tag{13}$$

2.1 Smooth-threshold multivariate genetic prediction

We introduce the smooth-threshold multivariate genetic prediction for generalized linear model. This is an extension of the method for multiple linear regression described in previous sections. Let the statistic for screening be $T_j(y, X)$, in which X_j is included if $T_j(y, X) > t$ for a given threshold $t > 0$. Scale parameter $a(\phi)$ is considered as known quantity in the following. Then, the estimator $\check{\beta}$ in the smooth-threshold multivariate genetic prediction is the solution to the equation with respect to β:

$$0 = U(\beta) = (I_p - \check{D})u(\beta) + \tau \check{D}\beta + \lambda(I_p - \check{D})\beta, \tag{14}$$

where $u(\beta) = -X^T\{y - b'(X\beta)\}$ is the score function and $\check{D}$ is the same quantity in (8) with the above $T_j(y, X)$. As in the least-squares case, define the screened set $A = \{j \in M : T_j(y, X) > t\}$. Then, $\check{D}_j = 1$ for $j \in A^c$ and we know that $\check{\beta}_{A^c} = 0$ in advance of computing the solution, giving

$$0 = U_A(\beta_A) = (I_{|A|} - \check{D}_A)u_A(\beta_A) + \tau\check{D}_A\beta_A + \lambda(I_{|A|} - \check{D}_A)\beta_A,$$

and $\check{\beta}_{A^c} = 0$. Here the subscript represents sub-vector with indexes in A. As in (10), the regression coefficient for the screened set, $\check{\beta}_A$, can be obtained by generalized ridge regression problem,

$$u_A(\check{\beta}_A) + W_A\check{\beta}_A = 0, \tag{15}$$

with $W_A = \mathrm{diag}(W_j : j \in A)$ where $W_j = \lambda + \tau\check{D}_j/(1 - \check{D}_j)$. The above equation can be solved by the Newton–Raphson algorithm and no convex optimization is needed.

To compute the GDF (12) for the smooth-threshold multivariate genetic prediction in generalized linear model, a tractable expression of $\partial_i\check{\beta}$ is needed. The following result gives a closed-form formula for the GDF.

Proposition 2 *The GDF for $\check{\theta}(y)$ (12) is equal to*

$$\sum_{i=1}^{n} b''(\hat{\theta}_{1,i})X_i\partial_i\hat{\beta} = \mathrm{tr}\{\check{G}_A\check{L}_A^T b''(\hat{\theta}_1)X_A\} + \mathrm{tr}\{\check{G}_A(I_{|A|} - \check{D}_A)X_A^T b''(\hat{\theta}_1)X_A\}, \tag{16}$$

where $\check{L}_A$ is an $n \times |A|$ matrix whose (i, j)-element is $(\partial_i\check{D}_j)\check{m}_j$, and $\check{m}_j$ is *the* j*th component of* $\check{m} = u(\check{\beta}) - (\tau - \lambda)\check{\beta}$. *Here,*

$$\partial_i \check{D}_j = -\frac{1+\gamma}{2}\frac{\partial_i T_j(y,X)}{T_j(y,X)}\{t/T_j(y,X)\}^{\frac{1+\gamma}{2}},$$

and $\check{G}_A = [(I_{|A|} - \check{D}_A)\{X_A^T b''(X\check{\beta})X_A + I_{|A|}\} + \tau\check{D}_A]^{-1}$.

The derivation is given in Appendix. For Gaussian linear model with canonical link, $b(\theta) = \theta^2/2$. Then, $b'(\theta) = \theta$, $b''(\theta) = 1$, $u_j(\beta) = -X_j^T(y - X\beta)$, and $u_j(\check{\beta}) = X_j^T X\check{\beta} - X_j^T y$. Substituting these quantities, it can be seen that the GDF given in Proposition 1 is a special case of the above GDF in Proposition 2 for generalized linear model in Gaussian linear model. Unpenalized components in X are treated in the same way in the multiple linear regression case.

2.2 Score test screening

In SNP-GWAS for binary traits, a score test for marginal association screening is considered. Adjustment for covariates Z can be easily incorporated, where Z is assumed to be an $n \times d$ matrix. Inclusion of intercept only corresponds to Z being an n-vector of ones. Given a p-value cutoff $\alpha \in (0,1)$, two-sided score test statistic for screening SNPs is

$$T_j(y,X) = u_j^2/v_j,$$

where $u_j = -X_j^T(y - \hat{\mu})$, $\hat{\mu} = b'(Z\hat{\gamma})$, $v_j = X_j^T \hat{W} X_j - (X_j^T \hat{W} Z)(Z^T \hat{W} Z)^{-1}(Z^T \hat{W} X_j)$, $\hat{\gamma}$ is the MLE under model with Z satisfying $Z^T\{y - b'(Z\hat{\gamma})\} = 0$ and $\hat{W} = \mathrm{diag}\{b''(Z_i\hat{\gamma})\}$. Threshold t is the $(1-\alpha)$th quantile of χ^2-distribution of one degree of freedom, $q_{1-\alpha}$ say.

2.3 C_p-type criterion for smooth-threshold multivariate genetic prediction in generalized linear model

Using expression (20) given in Appendix, the GDF for score-test screening (12) is derived by applying Proposition 2. As in the linear regression case, we use fixed $\gamma = 1$ and $\tau = n/\sqrt{\log n}$. Denoting the dependency of $\check{\theta}$ on α explicitly, the unbiased C_p-type model selection criterion based on $-2\times$ loglikelihood is given as follows:

$$C(\alpha) = \sum_{i=1}^{n} q\{y_i; \check{\theta}_i(\alpha)\} + 2\mathrm{GDF}(\alpha).$$

An optimal α is chosen by minimizing the above quantity. Here, we compute $\hat{\theta}_1$ at $\alpha_1 = 3n/(p \log n)$ as a sufficiently complex model as in linear regression case. From (13), the expectation of the above criterion approximates to $\sum_{i=1}^{n} E\{q(Ey_i; \check{\theta}_i)\}$ if the Taylor approximation used in the derivation is sufficiently accurate. The accuracy of the Taylor approximation may depend on the underlying data-generating process. In the simulation studies given in next section, which mimics genetic predictions from GWAS data, we found that the above approximation is accurate (See Figures 3 and 4 in main text).

3 Technical proofs

Here we give technical proofs for earlier sections.

3.1 Proof of Proposition 2

By operating ∂_i on both sides of (14), it follows that

$$\begin{aligned}
0 &= \partial_i U(\check{\beta}) \\
&= -\partial_i \check{D}\{u(\check{\beta}) - (\tau - \lambda)\check{\beta}\} + (I_p - \check{D})\{\partial_i u(\check{\beta})\} + \tau\check{D}\partial_i\check{\beta} + \lambda(I_p - \check{D})\partial_i\check{\beta} \\
&= [(I_p - \check{D})\{X^T b''(X\check{\beta})X + \lambda I_p\} + \tau\check{D}]\partial_i\check{\beta} - \partial_i\check{D}\{u(\check{\beta}) - (\tau - \lambda)\check{\beta}\} - (I_p - \check{D})X_i^T.
\end{aligned}$$

This is equivalent to

$$[(I_p - \check{D})\{X^T b''(X\check{\beta})X + \lambda I_p\} + \tau\check{D}]\partial_i\check{\beta} = \partial_i\check{D}\check{m} + (I_p - \check{D})X_i^T. \qquad (17)$$

Noting that $\partial_i\check{D}_j = 0$ for $j \in A^c$ or $T_j(y, X) \leq t$, (17) is re-expressed as

$$\begin{aligned}
&\left\{\begin{pmatrix} (I_{|A|} - \check{D}_A)X_A^T b''(X\check{\beta})X \\ O \end{pmatrix} + \begin{pmatrix} \lambda(I_{|A|} - \check{D}_A) + \tau\check{D}_A & O \\ O & I_{|A^c|} \end{pmatrix}\right\} \begin{pmatrix} \partial_i\check{\beta}_A \\ \partial_i\check{\beta}_{A^c} \end{pmatrix} \\
= &\begin{pmatrix} \partial_i\check{D}_A & O \\ O & O \end{pmatrix} \check{m} + \begin{pmatrix} (I_{|A|} - \check{D}_A)X_{A,i}^T \\ 0 \end{pmatrix},
\end{aligned}$$

which gives that

$$\begin{pmatrix} \partial_i \check{\beta}_A \\ \partial_i \check{\beta}_{A^c} \end{pmatrix} = \begin{pmatrix} \check{G}_A \left\{ \partial_i \check{D}_A \check{m}_A + (I_{|A|} - \check{D}_A) X_{A,i}^T \right\} \\ 0 \end{pmatrix},$$

where $\check{G}_A = [(I_{|A|} - \check{D}_A)\{X_A^T b''(X\check{\beta})X_A + \lambda I_{|A|}\} + \tau \check{D}_A]^{-1}$. Using the above quantities, it holds that

$$\begin{aligned} \text{GDF} &= \sum_{i=1}^{n} b''(\hat{\theta}_{1,i}) X_i \partial_i \hat{\beta} = \sum_{i=1}^{n} b''(\hat{\theta}_{1,i}) X_{i,A}^T \partial_i \check{\beta}_A \\ &= \sum_{i=1}^{n} b''(\hat{\theta}_{1,i}) X_{i,A}^T \check{G}_A (\partial_i \check{D}_A \check{m}_A) + \sum_{i=1}^{n} b''(\hat{\theta}_{1,i}) X_{i,A}^T \check{G}_A (I_{|A|} - \check{D}_A) X_{i,A} \\ &= \operatorname{tr}\{\check{G}_A \check{L}_A^T b''(\hat{\theta}_1) X_A\} + \operatorname{tr}\{\check{G}_A (I_{|A|} - \check{D}_A) X_A^T b''(\hat{\theta}_1) X_A\}, \qquad (18) \end{aligned}$$

where $\check{L}_A$ is an $n \times |A|$ matrix whose (i,j)-element is $(\partial_i \check{D}_j)\check{m}_j$.

3.2 Explicit formulas of $\partial_i T_j(y, X)$

For F-test screening, $\partial_i T_j(y, X)$ in Proposition 1 can be analytically calculated as follows. It is convenient to re-express $T_j(y, X)$ by

$$T_j(y, X) = (n - d) \frac{y^T P_j y}{y^T Q_j y},$$

in which $Q_j = I_n - P_Z - P_j$ and $P_j = P_{\tilde{X}_j}$. Then,

$$\begin{aligned}
\frac{\partial T_j(y,X)}{\partial y} &= 2(n-d)\left\{\frac{P_j y}{y^T Q_j y} - \frac{(y^T P_j y)Q_j y}{(y^T Q_j y)^2}\right\} \\
&= 2(n-d)\left[\frac{\frac{(\tilde{X}_j^T y)}{||\tilde{X}_j||^2}\tilde{X}_j}{||y - P_Z y||^2 - \frac{(\tilde{X}_j^T y)^2}{|\tilde{X}_j||^2}} - \frac{\frac{(\tilde{X}_j^T y)^2}{|\tilde{X}_j||^2}\{y - P_Z y - \frac{(\tilde{X}_j^T y)}{||\tilde{X}_j||^2}\tilde{X}_j\}}{\{||y - P_Z y||^2 - \frac{(\tilde{X}_j^T y)^2}{|\tilde{X}_j||^2}\}^2}\right] \qquad (19)
\end{aligned}$$

Here, $P_Z = Z(Z^T Z)^- Z^T$. For $Z = 1$, i.e. covariate is intercept only, $P_Z y = \bar{y}1$ where $\bar{y} = \sum_{i=1}^n y_i/n$.

For score-test screening in generalized linear model, the first derivative of $T_j(y,X)$ with respect to y_i is computed as follows.

$$\frac{\partial_i T_j(y,X)}{\partial y_i} = 2\frac{u_j \partial_i u_j}{v_j} - \left(\frac{u_j}{v_j}\right)^2 \partial_i v_j, \qquad (20)$$

where $\partial_i u_j$ is the (i,j)-element of the $n \times p$ matrix

$$\partial u = X - Z\{Z^T b''(Z\hat{\gamma})Z\}^{-1} Z^T b''(Z\hat{\gamma})X,$$

$\partial_i v_j$ is the (i,j)-element of the $n \times p$ matrix

$$\partial v = Z\{Z^T b''(Z\hat{\gamma})Z\}^{-1} Z^T b'''(Z\hat{\gamma})(\partial u \circ \partial u),$$

with $\circ$ denoting the Hadamard product.

3.3 Derivation of (20)

First, by operating ∂_i on both sides of $Z^T\{y-\mu(Z\hat{\gamma})\}=0$, we have that $0=\partial_i[Z^T\{y-b'(Z\hat{\gamma})\}]=Z_i^T-Z^Tb''(Z\hat{\gamma})Z\partial_i\hat{\gamma}$. By letting $B=Z^Tb''(Z\hat{\gamma})Z$, we have $\partial_i\hat{\gamma}=B^{-1}Z_i^T$, or in matrix form,

$$\partial^T\hat{\gamma}=B^{-1}Z^T,$$

which is the $d\times n$ matrix. Substituting into $\partial_i u_j=X_{ij}-X_j^Tb''(Z\hat{\gamma})Z\partial_i\hat{\gamma}$, we have $\partial u=X-ZB^{-1}A^T$. Here $A=Xb''(Z\hat{\gamma})Z^T$, $p\times d$ matrix.

Next consider $\partial_i v_j$. Expressing v_j by

$$v_j=\sum_{a=1}^n b''(Z_a\hat{\gamma})X_{ja}^2-A_jB^{-1}A_j^T,$$

in which $A_j=\sum_{a=1}^n b'(Z_a\hat{\gamma})X_{aj}Z_a$, the jth row of A, it holds that

$$\begin{aligned}\partial_i v_j&=\sum_{a=1}^n b'''(Z_a\hat{\gamma})X_{ja}^2\{Z_a(\partial_i\hat{\gamma})\}-2\left[\sum_{a=1}^n b'''(Z_a\hat{\gamma})X_{aj}\{Z_a(\partial_i\hat{\gamma})\}Z_a\right]B^{-1}A_j^T\\&\quad+A_jB^{-1}\left[\sum_{a=1}^n b'''(Z_a\hat{\gamma})Z_a^TZ_a\{Z_a(\partial_i\hat{\gamma})\}\right]B^{-1}A_j^T\\&=\sum_{a=1}^n\{Z_a(\partial_i\hat{\gamma})\}b'''(Z_a\hat{\gamma})(X_{ja}-Z_aB^{-1}A_j^T)^2\\&=\sum_{a=1}^n(Z_iB^{-1}Z_a^T)b'''(Z_a\hat{\gamma})(\partial_a u_j)^2,\end{aligned}$$

which arrives at the desired expression.

4 Deflated RSS due to winner's curse effect

Consider the RSS from multiple regression after marginal association screening which uses all SNPs simultaneously in A. The F-test screening is roughly equivalent to the screening based on the marginal association. Hence the following RSS after marginal screening at a cutoff value $s^2 \geq 0$ follows:

$$\text{RSS}_A = ||y - P_A y||^2 = ||y||^2 - c_A^T G_{AA} c_A, \tag{21}$$

where $P_A = \tilde{X}_A G_{AA} \tilde{X}_A^T$, $G_{AA} = (\tilde{X}_A^T \tilde{X}_A)^{-1}$, $c_A = (c_j)_{j \in A}$, $c_j = \tilde{X}_j^T y$, $\tilde{X}_j = Q_1 X_j$ and $A = \{j \in M : c_j^2 > s^2\}$. Since, in most practical GWASs, the threshold s is taken to be large so as to make the SNP discovery conservative, $|A| < n$ can be assumed. As a result, the above RSS_A is computable and can be used in evaluating a predictive power of the model using X_A. However, the behavior of RSS_A may differ from the behavior when no screening is applied, i.e. sampling variability in selecting A invalidates the theory for RSS_A with deterministic A.

Assume that $y \sim N(\mu, \sigma_y^2 I_n)$. Then, $c = X^T y \sim N(\beta, \Sigma)$, in which $\beta = X^T \mu$ and $\Sigma = \sigma_y^2 X^T X$. For simplicity, assume that $X^T X = I_p$, i.e. an orthogonal case, and hence $G_{AA} = I_{|A|}$ for any $A \subset M$. Under this assumption, $c_1, \ldots, c_p$ are mutually independent: $c = X^T y \sim N(\beta, \sigma_y^2 I_p)$. Meanwhile we assume that σ_y^2 is known. In the case where A is deterministic, the expectation of RSS_A is

$$\begin{aligned} E(\mathrm{RSS}_A) &= E||y||^2 - E(c_A^T G_{AA} c_A) \\ &= E||y||^2 - \beta_A^T G_{AA} \beta_A - \sigma_y^2 |A| \\ &= E||y||^2 - \sum_{j \in A} E(c_j^2) \\ &= E||y||^2 - \sum_{j \in A} (\beta_j^2 + \sigma_y^2). \end{aligned}$$

On the other hand, when A is random, the expectation of RSS_A conditional on $A = \{j \in M : c_j^2 > s^2\}$ is

$$\begin{aligned} E(\mathrm{RSS}_A \mid A) &= E||y||^2 - E(c_A^T G_{AA} c_A \mid A) \\ &= E||y||^2 - E(\sum_{j,k \in M} G_{jk} c_j c_k 1_{\{j,k \in A\}} \mid A) \\ &= E||y||^2 - \sum_{j \in A} E(c_j^2 \mid |c_j| > s) \end{aligned}$$

in which 1_B represents the indicator function of a set B. Note that, at $s = 0$, $E(c_j^2 \mid |c_j| > s) = E(c_j^2)$. By letting $g(s^2) = \log E(c_j^2 \mid |c_j| > s) = \log E(c_j^2 \mid$

$c_j^2 > s^2$), and setting $u = s^2$, the first derivative of $g(u)$ is

$$\begin{aligned}
\frac{dg(u)}{du} &= \frac{d}{du}\log\frac{\int_{x^2>u} x^2\phi_{\sigma_y}(x-\beta_j)dx}{\int_{x^2>u}\phi_{\sigma_y}(x-\beta_j)dx} \\
&= \frac{-u\phi_{\sigma_y}(\sqrt{u}-\beta_j)}{\int_{x^2>u} x^2\phi_{\sigma_y}(x-\beta_j)dx} - \frac{-\phi_{\sigma_y}(\sqrt{u}-\beta_j)}{\int_{x^2>u}\phi_{\sigma_y}(x-\beta_j)dx} \\
&> \frac{-u\phi_{\sigma_y}(\sqrt{u}-\beta_j)}{u\int_{x^2>u}\phi_{\sigma_y}(x-\beta_j)dx} - \frac{-\phi_{\sigma_y}(\sqrt{u}-\beta_j)}{\int_{x^2>u}\phi_{\sigma_y}(x-\beta_j)dx} \\
&= 0,
\end{aligned}$$

and hence, $g(u)$ is monotone increasing in $u = s^2$, which is true for $e^{g(u)} = E(c_j^2 \mid |c_j| > s)$. Consequently, the conditional expectation $E(\mathrm{RSS}_A \mid A)$ is smaller than the expectation of RSS_A with any deterministic A. The larger u is taken, the larger deflation in RSS_A due to screening or the winner's curse effect appears.

5 Additional figures and tables from low heritability polygenic simulations

Here we present additional simulation results under the same polygenic models as that in main text except that lower heritabilities are assumed. To be specific, we repeated quantitative trait simulations 20 times with the following six models: Model P13, $p_0 = 100$, $h^2 = 0.01$; Model P14, $p_0 = 100$, $h^2 = 0.005$; Model P15, $p_0 = 100$, $h^2 = 0.001$; Model P16, $p_0 = 200$, $h^2 = 0.01$; Model P17, $p_0 = 200$, $h^2 = 0.005$; Model P18, $p_0 = 200$,

$h^2 = 0.001$. We also repeated binary trait simulations 20 times with the following six models: Model P19, $p_0 = 100$, $h^2 = 0.01$; Model P20, $p_0 = 100$, $h^2 = 0.005$; Model P21, $p_0 = 100$, $h^2 = 0.001$; Model P22, $p_0 = 200$, $h^2 = 0.01$; Model P23, $p_0 = 200$, $h^2 = 0.005$; Model P24, $p_0 = 200$, $h^2 = 0.001$, which correspond to the models with liability generated from Models P13,…,P18. The results are given in Table S1 and Figures S2 and S3.

References

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. *Pages 267–281 of: Proceedings of the 2nd International Symposium on Information Theory, Petrov, B. N and and Caski, F. (eds.).* Budapest: Akadimiai Kiado.

Augugliaro L, Mineo AM, Wit EC. 2013. Differential geometric least angle regression: a differential geometric approach to sparse generalized linear models. *J Roy Stat Soc B* **75**, 471–498.

Breiman L. 1996. Heuristics of instability and stabilization in model selection. *Ann Stat* **24**, 2350–2383.

Bühlmann P, van de Geer S. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Berlin, Heidelberg: Springer.

Efron B. 2004. The estimation of prediction error: covariance penalties and Ccross-validation. *J Am Stat Assoc* **99**, 619–632.

Fan J, Lv J. 2008. Sure independence screening for ultrahigh dimensional feature space (with discussion). *J Roy Stat Soc B* **70**, 849–911.

Fan J, Samworth R, Wu Y. 2009. Ultrahigh dimensional feature selection: beyond the linear model. *J Mach Learn Res* **10**, 2013–2038.

Purcell SM, Wray NR, Stone JL, Visscher PM, O ’Donovan MC, Sullivan PF, Sklar P, Consortium International Schizophrenia. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–52.

Stein CM. 1981. Estimation of the mean of a multivariate normal distribution. *Ann Stat* **9**, 1135–1151.

Ueki M. 2009. A note on automatic variable selection using smooth-threshold estimating equation. *Biometrika* **96**, 1005–1011.

Ueki M, Kawasaki Y. 2011. Automatic grouping using smooth-threshold estimating equations. *Electron J Stat* **5**, 309–328.

Warren H, Casas JP, Hingorani A, Dudbridge F, Whittaker J. 2013. Genetic prediction of quantitative lipid traits: comparing shrinkage models to gene scores. *Genet Epidemiol* **38**, 72–83.

Ye J. 1998. On measuring and correcting the effects of data mining and model selection. *J Am Stat Assoc* **93**, 120–131.

Zhong H, Prentice RL. 2008. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621–634.

Zou H. 2006. The adaptive lasso and its oracle properties. *J Am Stat Assoc* **101**, 1418–29.

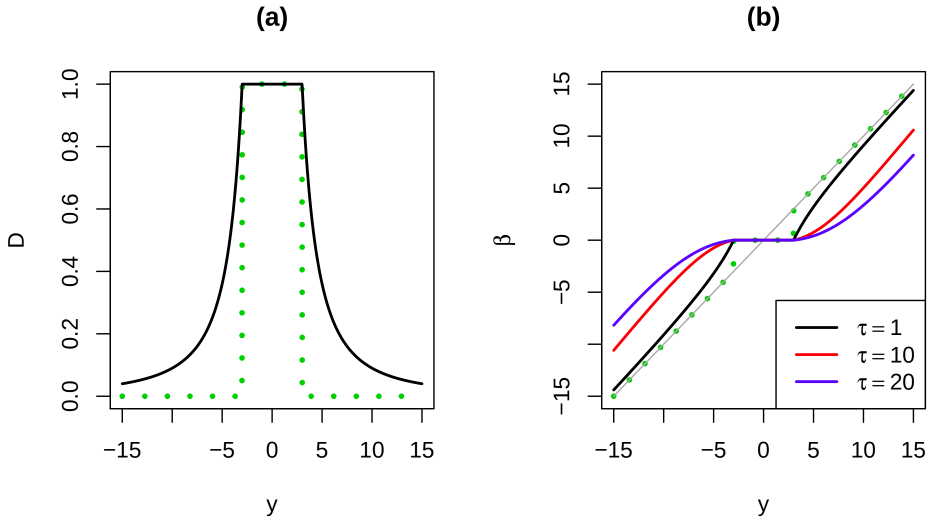

Figure S1: Illustration in a simple least-squares regression, $y = \beta + \epsilon$; (a) indicator function $\hat{D}(y) = 1(|y| \le t)$ with $t = 3$ (green dotted) and its approximation by the adaptive lasso smooth-thresholding function $\check{D}(y) = \min\{1, (t/y)^{1+\gamma}\}$ with $\gamma = 1$ (black solid); (b) plots of $\frac{1-D}{1-(1-\tau)D}y$, which is the solution to the equation $(1 - D)(\beta - y) + \tau D\beta = 0$ with respect to β given $\tau > 0$ and D. Indicator function, $D = \hat{D}(y)$ (Green dotted). (Note: any τ gives an identical solution.) Smooth-thresholding $D = \check{D}(y)$ for $\tau = 1$ (black solid), 10 (red solid) and 20 (blue solid), respectively. When $\tau = 1$, the smooth-threshold estimator reduces to the adaptive lasso estimator.

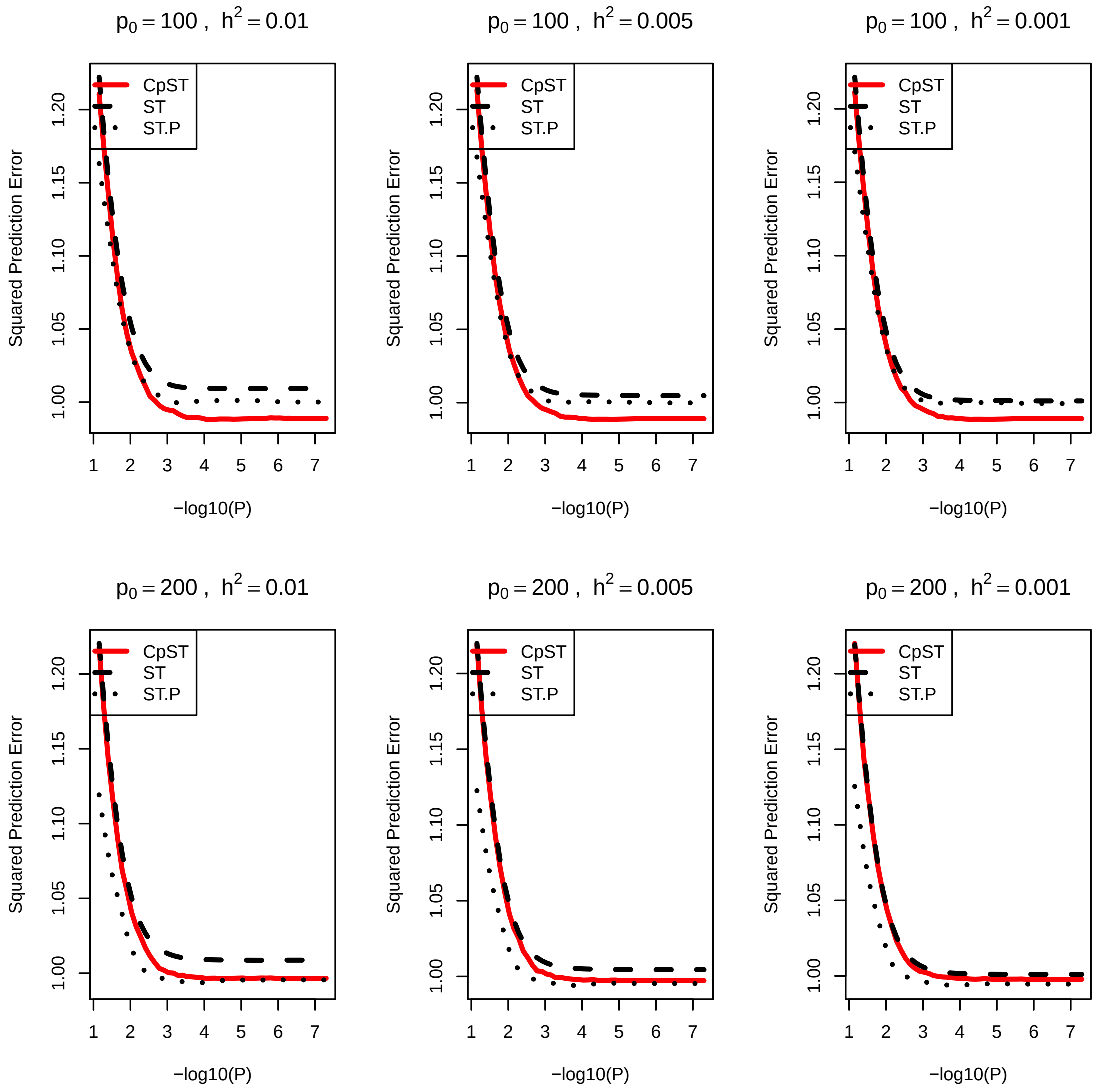

Figure S2: Prediction errors averaged over 20 simulation replicates for quantitative traits in polygenic scenarios (Models P13,...,P18). Black dashed line (ST), average of mean prediction squared error for training data (PSEtr) for predictive models from smooth-threshold multivariate genetic prediction at each p-value threshold in $-\log_{10}$-scale (x-axis). Black dotted line (ST.P), average of prediction squared error for test data (PSEte) for predictive model from smooth-threshold multivariate genetic prediction trained on the training data. Red solid line (CpST), average of the proposed C_p-type criterion (an unbiased estimator of the black dashed line).

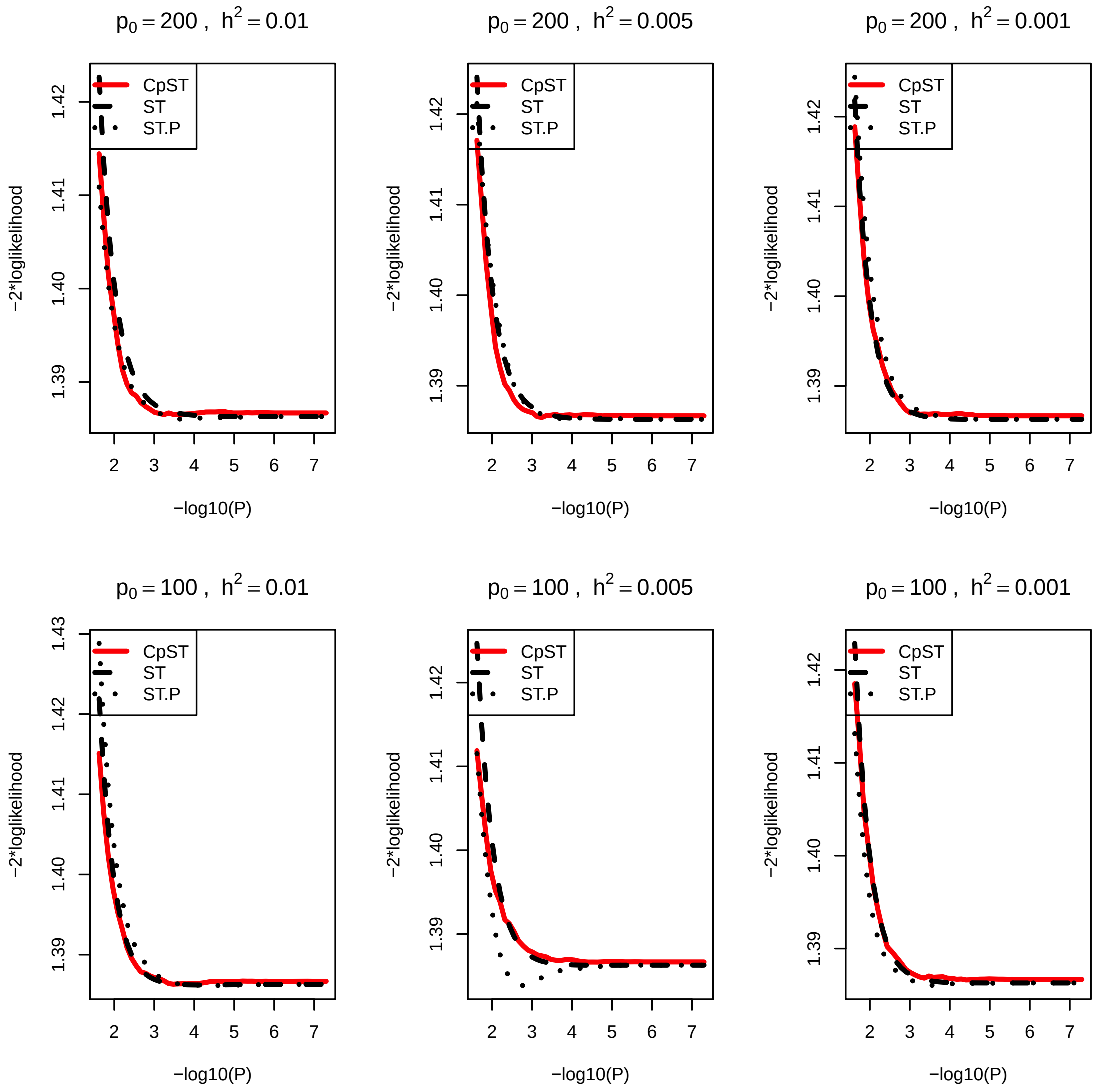

Figure S3: Prediction $-2\times$ loglikelihood averaged over 20 simulation replicates for binary traits in polygenic scenarios (P19,...,P24). Black dashed line (ST), average of mean $-2\times$ loglikelihood for training data (PSEtr) for predictive models from smooth-threshold multivariate genetic prediction at each p-value threshold in $-\log_{10}$-scale (x-axis). Black dotted line (ST.P), average of prediction $-2\times$ loglikelihood for test data (PSEte) for predictive model from smooth-threshold multivariate genetic prediction trained on the training data. Red solid line (CpST), average of the proposed C_p-type criterion (an approximate unbiased estimator of the black dashed line).

Table S1: Simulations under low helitability polygenic scenarios, predictive correlation coefficient (PCC), average AUC with standard deviation in parenthesis, and average number of true/false positive (TP/FP) results for three methods in replicates. STMGP, smooth-threshold multivariate genetic prediction; Enet, elastic net; GS, gene score. The best performing method is emphasized in bold.

p_0	h^2		STMGP	Lasso	Enet	GCTA	GS
				Quantitative traits ($n = 5000$)			
100	0.01	PCC	0.02 (0.07)	0.01 (0.1)	0.01 (0.1)	0.02 (0.11)	**0.03** (0.11)
		TP/FP	0/15	1/87	1/86	-	16/8495
100	0.005	PCC	0 (0.06)	0.01 (0.1)	0.01 (0.1)	0.01 (0.11)	**0.04** (0.13)
		TP/FP	0/6	0/93	0/89	-	15/8547
100	0.001	PCC	-0.01 (0.05)	0 (0.1)	0 (0.1)	**0.01** (0.11)	-0.02 (0.12)
		TP/FP	0/9	0/91	0/81	-	7/5225
200	0.01	PCC	**0.05** (0.1)	0.03 (0.11)	0.03 (0.1)	0.02 (0.11)	0.04 (0.11)
		TP/FP	0/26	1/91	1/80	-	26/8190
200	0.005	PCC	**0.04** (0.09)	0.01 (0.11)	0.01 (0.1)	0.02 (0.11)	0.02 (0.1)
		TP/FP	0/20	0/98	0/80	-	38/12843
200	0.001	PCC	**0.03** (0.09)	0.01 (0.12)	0 (0.1)	0.01 (0.11)	**0.03** (0.1)
		TP/FP	0/28	0/96	0/74	-	18/6052
				Binary traits ($n = 5000$)			
100	0.01	AUC	0.5 (0.03)	0.5 (0.02)	0.5 (0.03)	**0.51** (0.04)	0.5 (0.03)
		TP/FP	1/25	2/1297	5/1663	-	13/6195
100	0.005	AUC	0.5 (0.02)	**0.51** (0.03)	**0.51** (0.03)	0.49 (0.03)	0.5 (0.03)
		TP/FP	0/57	3/2078	6/2952	-	16/9796
100	0.001	AUC	0.5 (0.01)	0.5 (0.03)	**0.51** (0.03)	0.5 (0.04)	0.5 (0.04)
		TP/FP	0/39	3/2166	4/2984	-	13/9234
200	0.01	AUC	0.5 (0.02)	0.5 (0.03)	0.5 (0.03)	0.51 (0.03)	**0.52** (0.03)
		TP/FP	1/88	7/2220	7/2100	-	15/4614
200	0.005	AUC	0.5 (0.03)	0.5 (0.03)	0.5 (0.03)	**0.51** (0.03)	**0.51** (0.03)
		TP/FP	0/78	4/1114	8/2279	-	25/8293
200	0.001	AUC	0.5 (0.02)	**0.51** (0.04)	**0.51** (0.04)	0.5 (0.04)	**0.51** (0.03)
		TP/FP	0/36	7/2642	7/2557	-	21/7656