# Ridge regression for risk prediction

## with applications to genetic data

Erika Cule and Maria De Iorio

Imperial College London
Department of Epidemiology and Biostatistics
School of Public Health

May 2012

Imperial College London

# Outline

# Outline

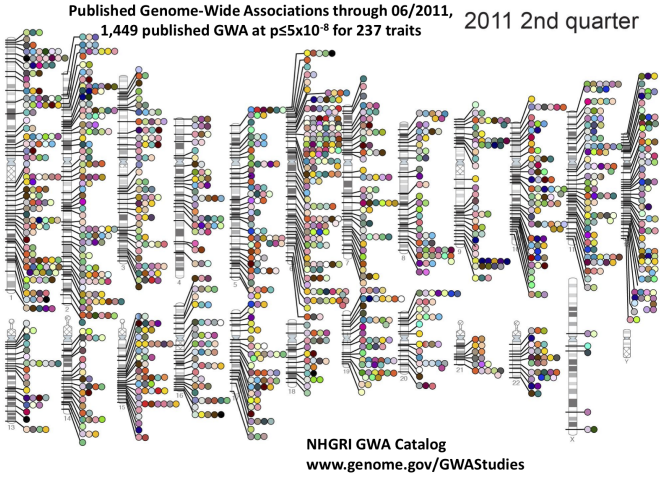
# Risk Prediction using Genetic Data


In the decade following the publication of the first draft of the Human Genome Sequence...

# Risk Prediction using Genetic Data


In the decade following the publication of the first draft of the Human Genome Sequence...

...genome-wide association studies have identified thousands of genetic variants associated with hundreds of diseases and traits.



# Risk Prediction using Genetic Data

However, clinicians are getting impatient about the utility of these identified variants for risk prediction in complex diseases:

NEWS & VIEWS

DIABETES

## GATTACA—are we there yet?

*Jeremy B. M. Jowett*

**In the past few years, new genetic risk factors for type 2 diabetes mellitus have been identified in large-scale studies of genetic associations. But can their inclusion in clinically based risk models improve disease prediction?**

NATURE REVIEWS | ENDOCRINOLOGY VOLUME 5 | APRIL 2009 | 187

## Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

N ENGL J MED 360;17 APRIL 23, 2009

# Risk Prediction using Genetic Data

However, clinicians are getting impatient about the utility of these identified variants for risk prediction in complex diseases:

NEWS & VIEWS

DIABETES

## GATTACA—are we there yet?

*Jeremy B. M. Jowett*

**In the past few years, new genetic risk factors for type 2 diabetes mellitus have been identified in large-scale studies of genetic associations. But can their inclusion in clinically based risk models improve disease prediction?**

NATURE REVIEWS | ENDOCRINOLOGY VOLUME 5 | APRIL 2009 | 187

## Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

Contemporary Reviews in Cardiovascular Medicine

N ENGL J MED 360;17 APRIL 23, 2009

## Genetic Cardiovascular Risk Prediction
### Will We Get There?

George Thanassoulis, MD; Ramachandran S. Vasan, MD

*Circulation* November 30, 2010

# Risk prediction using genetic data

- Recently, questions have been raised about the potential utility of genetic risk prediction for complex diseases (Clayton, 2009).

# Risk prediction using genetic data

- Recently, questions have been raised about the potential utility of genetic risk prediction for complex diseases (Clayton, 2009).
- The aim here is to make ridge regression possible for genetic data in a semi-automatic way

# Risk prediction using genetic data

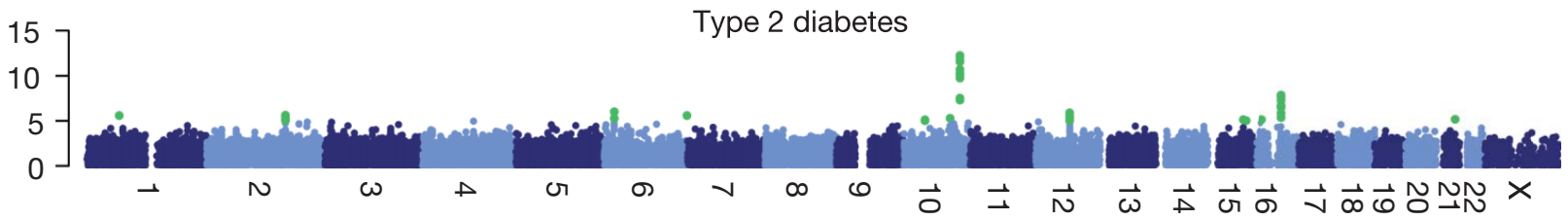
- Recently, questions have been raised about the potential utility of genetic risk prediction for complex diseases (Clayton, 2009).
- The aim here is to make ridge regression possible for genetic data in a semi-automatic way
- The framework that we propose allows for the simultaneous inclusion of all predictors genome-wide in a regression model.

# Risk prediction using genetic data

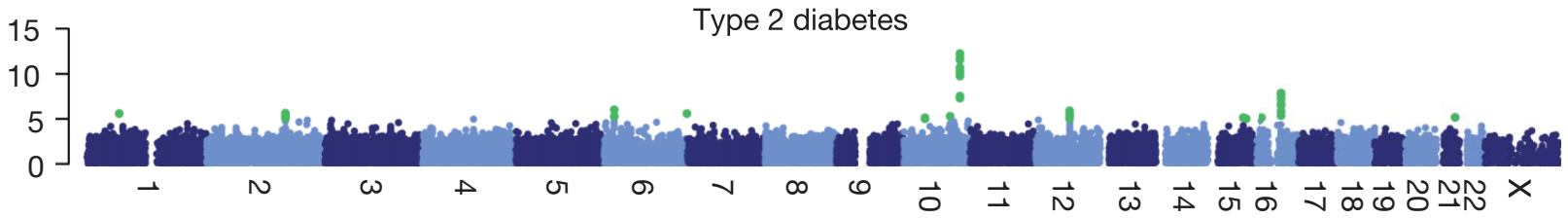
- Recently, questions have been raised about the potential utility of genetic risk prediction for complex diseases (Clayton, 2009).
- The aim here is to make ridge regression possible for genetic data in a semi-automatic way
- The framework that we propose allows for the simultaneous inclusion of all predictors genome-wide in a regression model.
- Our approach is appropriate where there are many predictors of small effect size,which is thought to be the case in genetic data.

# Outline

# Univariate tests of association



# Univariate tests of association

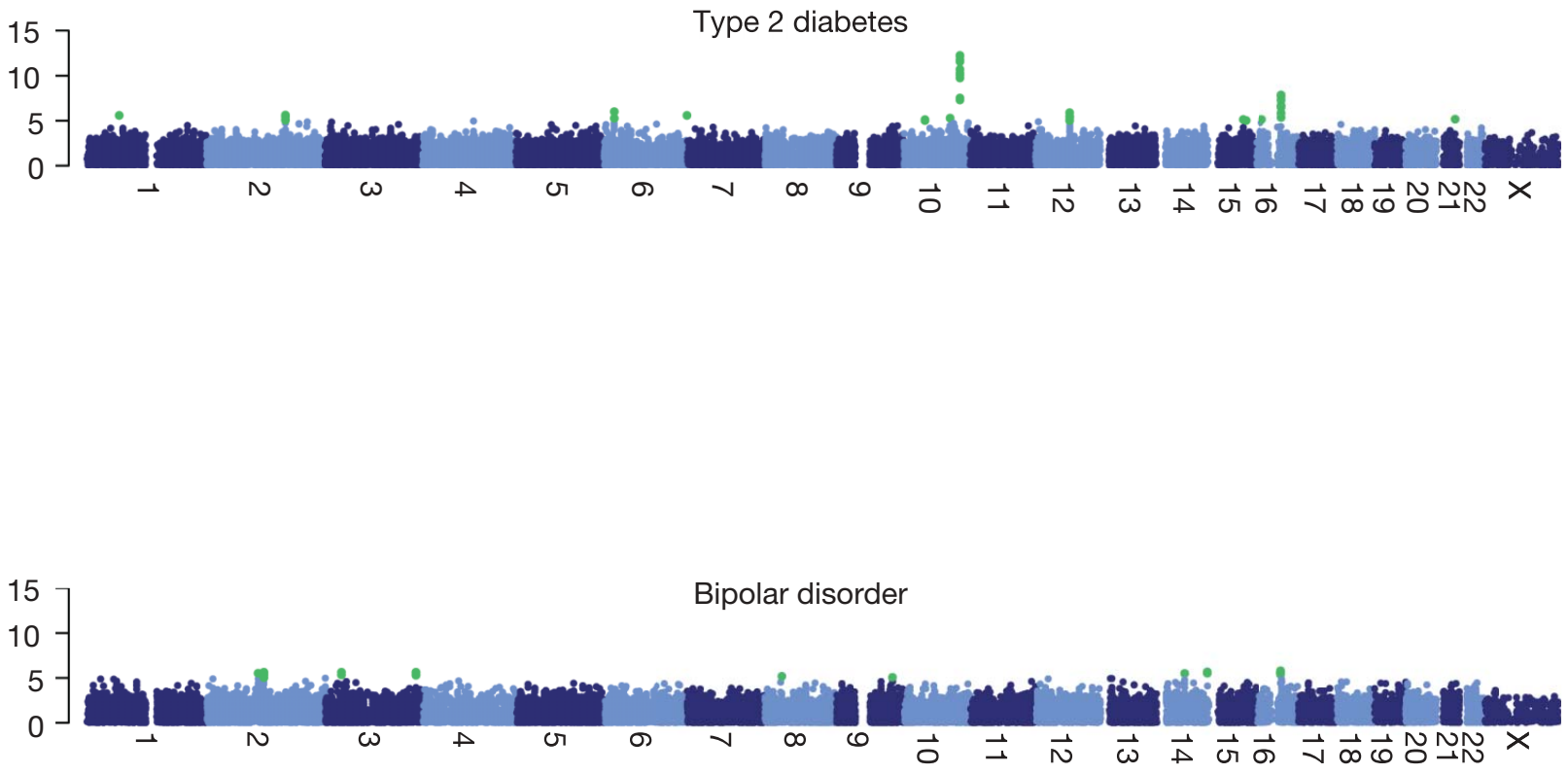



## Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes

James B. Meigs, M.D., M.P.H., Peter Shrader, M.S., Lisa M. Sullivan, Ph.D., Jarred B. McAteer, B.A., Caroline S. Fox, M.D., M.P.H., Josée Dupuis, Ph.D., Alisa K. Manning, M.A., Jose C. Florez, M.D., Ph.D., Peter W.F. Wilson, M.D., Ralph B. D'Agostino, Sr., Ph.D., and L. Adrienne Cupples, Ph.D.

N ENGL J MED 359;21
NOVEMBER 20, 2008

# Univariate tests of association




WTCCC (2007)

# Multivariate methods

- Consider all SNPs jointly

# Multivariate methods

- Consider all SNPs jointly
- Standard multivariate methods cannot be used with modern genetic data sets which have $p \gg n$.
- Typically, additional (non-genetic) covariates are included in the analysis, further increasing the dimensionality of the data.

# Multivariate methods

- Consider all SNPs jointly
- Standard multivariate methods cannot be used with modern genetic data sets which have $p \gg n$.
- Typically, additional (non-genetic) covariates are included in the analysis, further increasing the dimensionality of the data.
- Penalized regression methods constrain the size of the maximum likelihood estimates of regression coefficients.
- Known as shrinkage methods - “shrink” regression coefficients towards zero.

# Multivariate methods

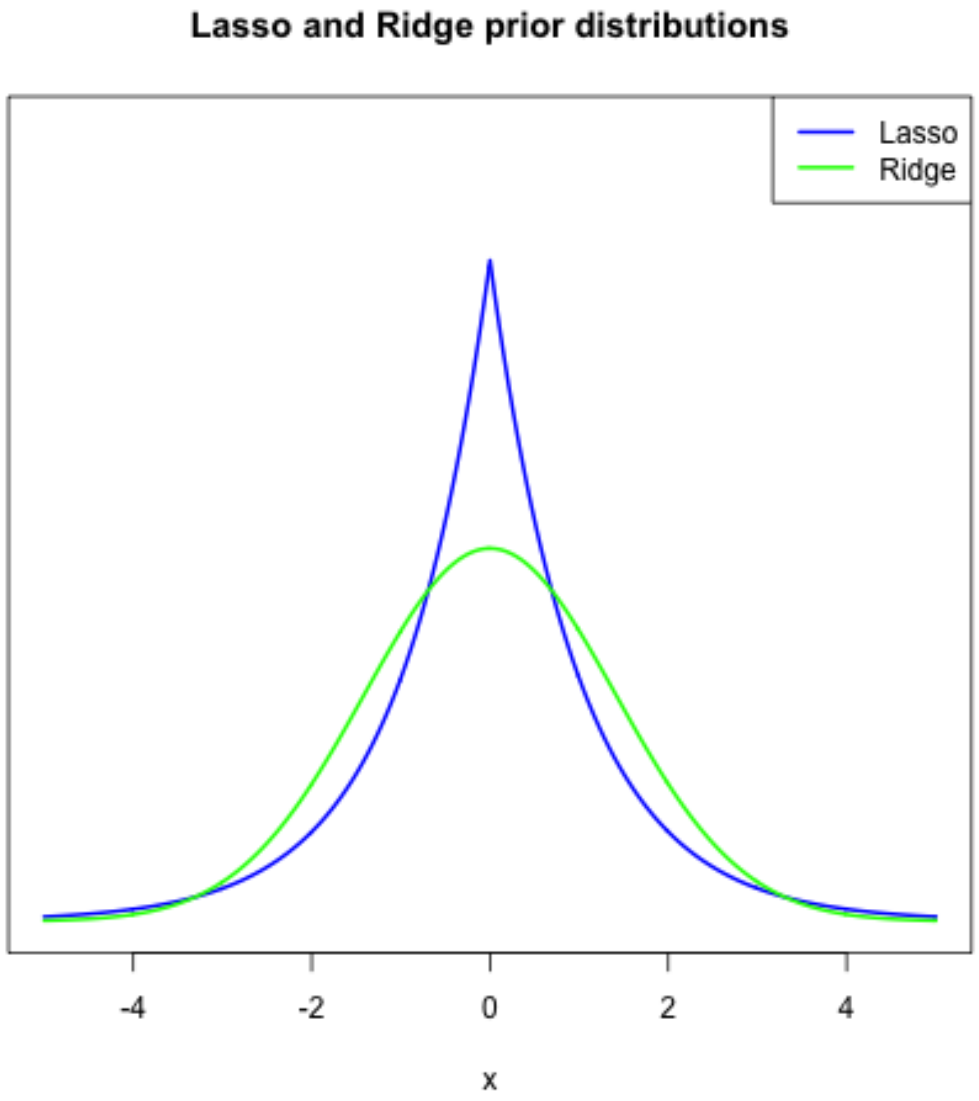
- Consider all SNPs jointly
- Standard multivariate methods cannot be used with modern genetic data sets which have $p \gg n$.
- Typically, additional (non-genetic) covariates are included in the analysis, further increasing the dimensionality of the data.
- Penalized regression methods constrain the size of the maximum likelihood estimates of regression coefficients.
- Known as shrinkage methods - “shrink” regression coefficients towards zero.
- A number of penalized regression approaches have been proposed in the literature: Lasso regression, HyperLasso, Elastic Net...

# Multivariate methods

- Consider all SNPs jointly
- Standard multivariate methods cannot be used with modern genetic data sets which have $p \gg n$.
- Typically, additional (non-genetic) covariates are included in the analysis, further increasing the dimensionality of the data.
- Penalized regression methods constrain the size of the maximum likelihood estimates of regression coefficients.
- Known as shrinkage methods - “shrink” regression coefficients towards zero.
- A number of penalized regression approaches have been proposed in the literature: Lasso regression, HyperLasso, Elastic Net...

Ridge Regression

# Prior distributions in Lasso and Ridge Regression



# Outline

# Ridge regression

- Ridge regression (Hoerl & Kennard, 1970) is a penalized regression approach proposed to overcome the problems associated with multicollinearity among predictors in multiple regression.
- Among penalized regression approaches, ridge regression has been shown to offer very good predictive performance (Frank & Friedman, 1993).
- We applied ridge regression to the problem of risk prediction using genetic data obtained from genome-wide association studies.
- Ridge regression shrinks the squared length of the regression coefficient vector - corresponds to a quadratic penalty on the coefficients.

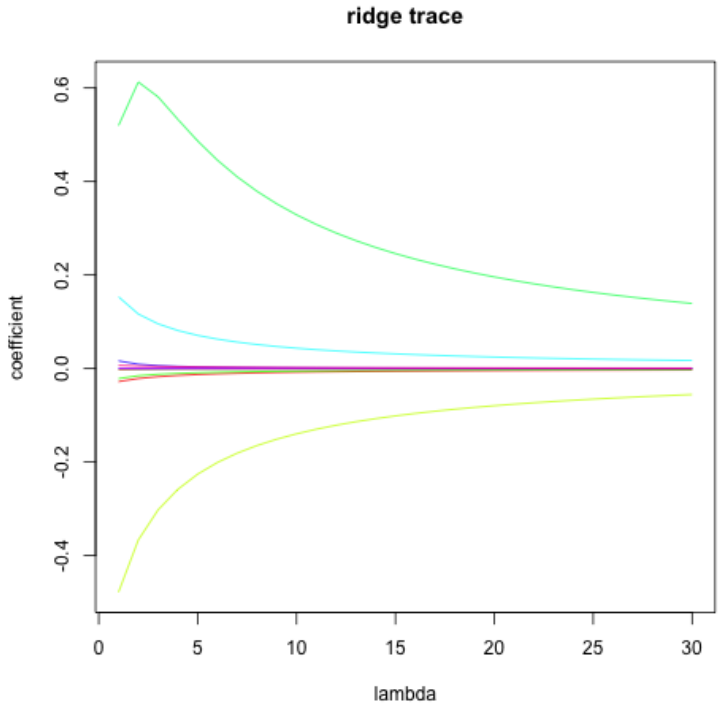
# Outline

# Shrinkage parameter

- Controls the degree of shrinkage of the regression coefficients.
- A larger shrinkage parameter shrinks the coefficients further towards zero.
- Data-driven methods proposed in the literature cannot be applied $p \gg n$, because they depend on the ordinary least squares estimates.

# Shrinkage parameter

- Controls the degree of shrinkage of the regression coefficients.
- A larger shrinkage parameter shrinks the coefficients further towards zero.
- Data-driven methods proposed in the literature cannot be applied $p \gg n$, because they depend on the ordinary least squares estimates.
- Ridge trace (graphical method)



# Our starting point

Linear model:

$$Y = X\beta + \epsilon \qquad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

# Our starting point

Linear model:

$$Y = X\beta + \epsilon \qquad\qquad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Ridge regression:

$$\hat{\beta}_k = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + k \sum_{j=1}^{p} |\beta_j^2| \right\}$$

# Our starting point

Linear model:

$$Y = X\beta + \epsilon \qquad\qquad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Ridge regression:

$$\hat{\beta}_k = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + k \sum_{j=1}^{p} |\beta_j^2| \right\}$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

# Our starting point

Linear model:

$$Y = X\beta + \epsilon \qquad\qquad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Ridge regression:

$$\hat{\beta}_k = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + k \sum_{j=1}^{p} |\beta_j^2| \right\}$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

$\hat{\sigma}^2, \hat{\beta}$ estimated from ordinary least squares (OLS).

# We observed

Linear model:

$$Y = X\beta + \epsilon \qquad\qquad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

# We observed

Linear model:

$$Y = X\beta + \epsilon \qquad = Z\alpha + \epsilon \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

# We observed

Linear model:

$$Y = X\beta + \epsilon \qquad = Z\alpha + \epsilon \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

# We observed

Linear model:

$$Y = X\beta + \epsilon \qquad = Z\alpha + \epsilon \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

$\hat{\alpha}$ are principal components regression coefficients.

# We observed

Linear model:

$$Y = X\beta + \epsilon \qquad = Z\alpha + \epsilon \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Proposed by Hoerl, Kennard & Baldwin (1975):

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

$\hat{\alpha}$ are principal components regression coefficients.

PCR coefficients *are* available when $p >> n$

# We propose

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

Harmonic mean of the “ideal” shrinkage parameters of the PCR coefficients, with coefficients replaced by their ordinary least squares estimates.

# We propose

$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

$$k_r = \frac{r\hat{\sigma}_r^2}{\hat{\alpha}_r'\hat{\alpha}_r}$$

Harmonic mean of the “ideal” shrinkage parameters of the PCR coefficients, with coefficients replaced by their ordinary least squares estimates.

# We propose

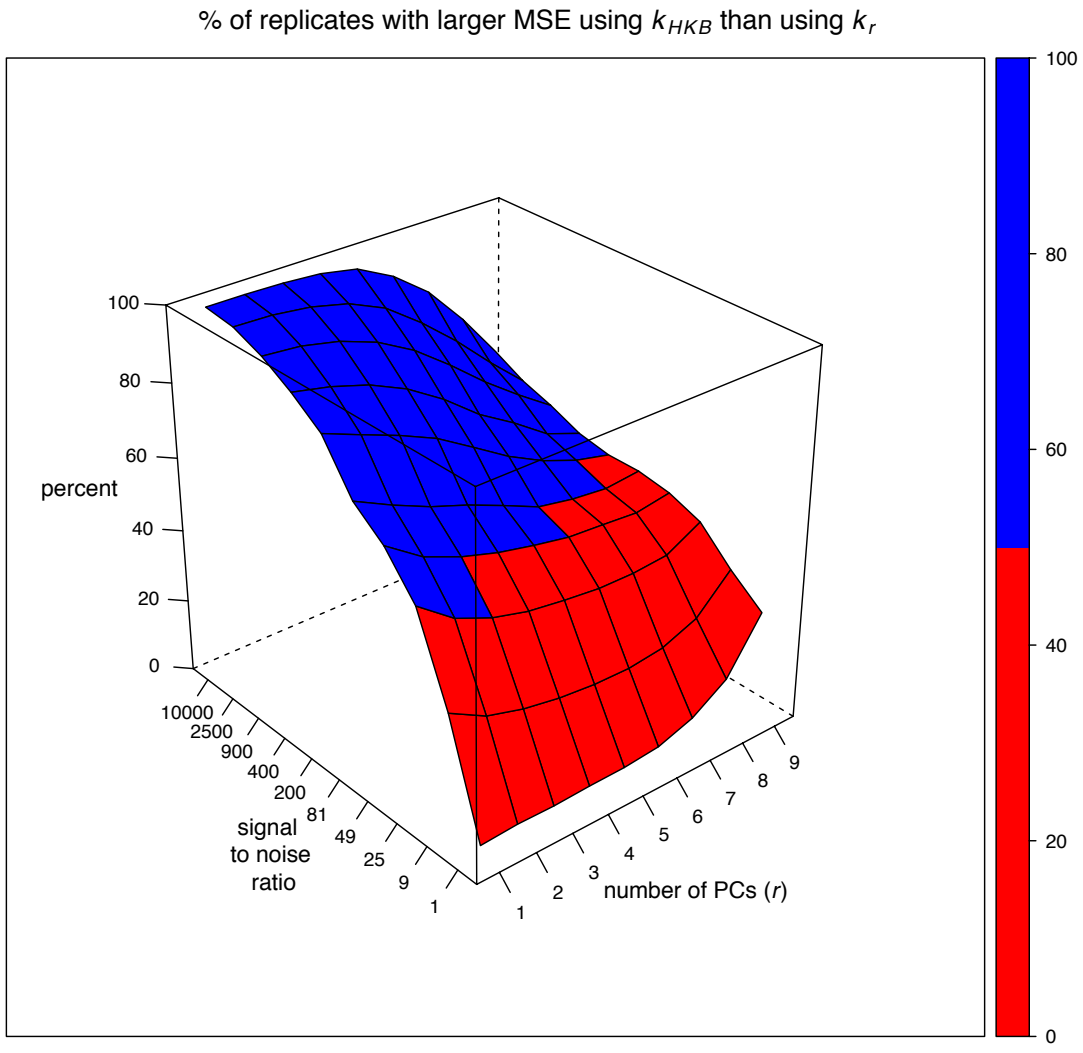
$$k_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

$$k_r = \frac{r\hat{\sigma}_r^2}{\hat{\alpha}_r'\hat{\alpha}_r}$$

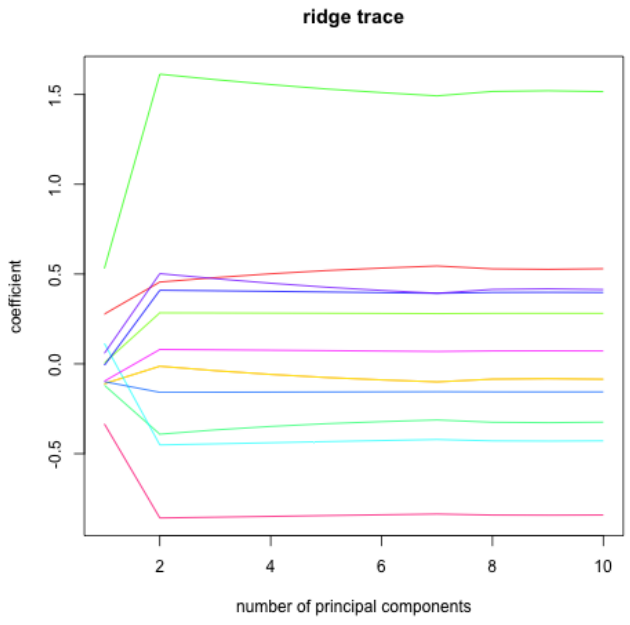
Harmonic mean of the “ideal” shrinkage parameters of the PCR coefficients, with coefficients replaced by their ordinary least squares estimates.

How many components?

# How many components?



# How many components?




Most of the variance in genetic data can be explained by the first few principal components.

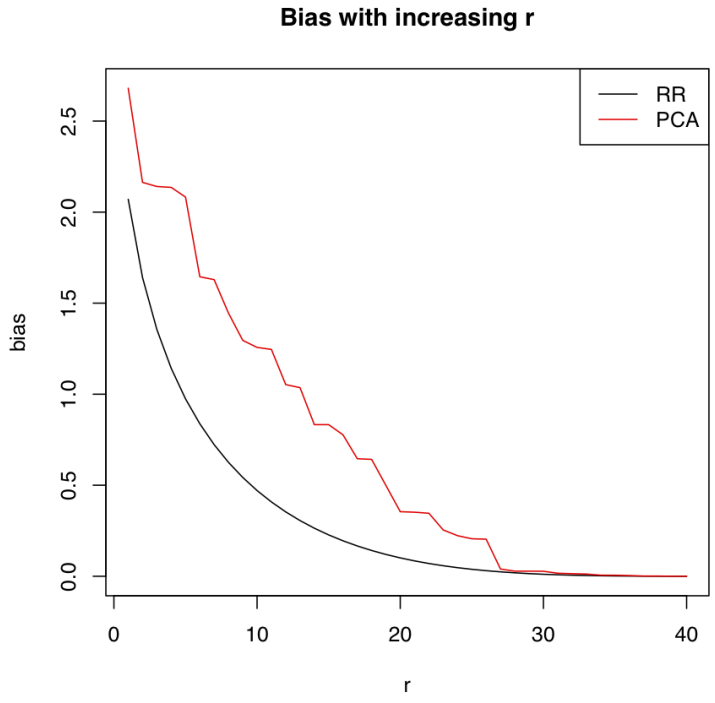
# How many components?

$$\begin{aligned} \text{PSE} &= \left\{1 + \frac{\text{tr}(HH')}{n}\right\}\sigma^2 + \frac{b'b}{n} \\ &= \text{variance} + \frac{\text{bias}^2}{n} \end{aligned}$$

- $H$ is the "hat matrix": $\hat{Y} = HY$
- Degrees of freedom for variance = $\text{tr}(HH')$ (Hastie & Tibshirani (1990) ).

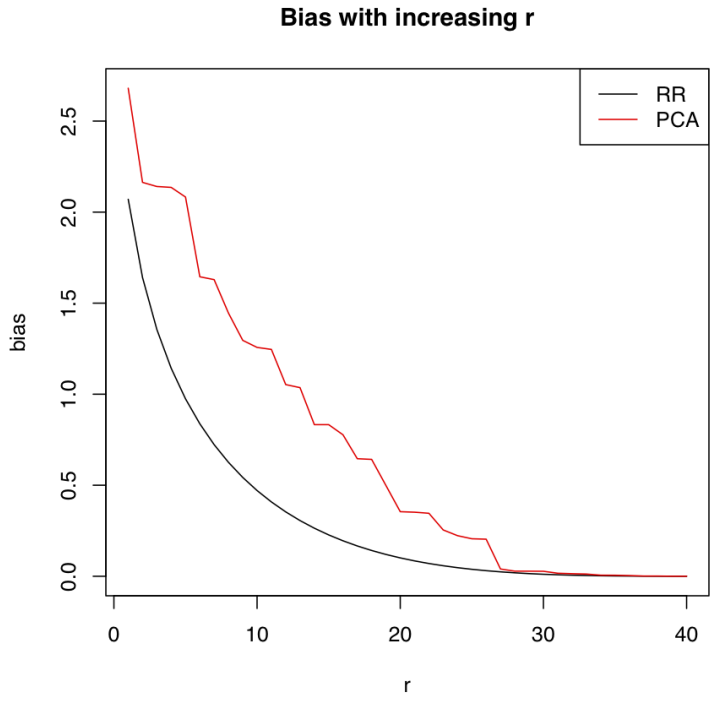
# How many components?

- For given *r*, RR estimates have less bias than PCR estimates.



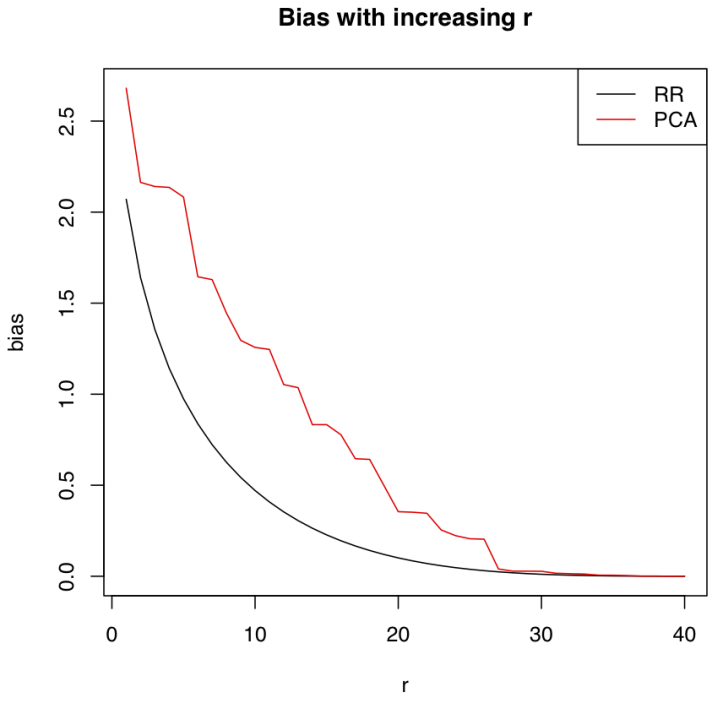
# How many components?

- For given *r*, RR estimates have less bias than PCR estimates.
- PCR using *r* components has *r* degrees of freedom for variance.



# How many components?

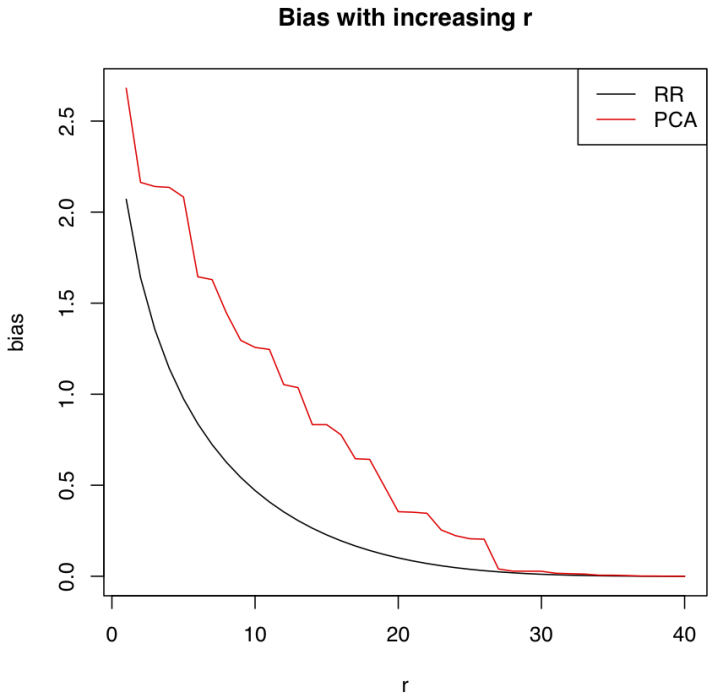
- For given $r$, RR estimates have less bias than PCR estimates.
- PCR using $r$ components has $r$ degrees of freedom for variance.
- We fixed $r$ such that degrees of freedom of the ridge model using $r$ components equals $r$.



# How many components?

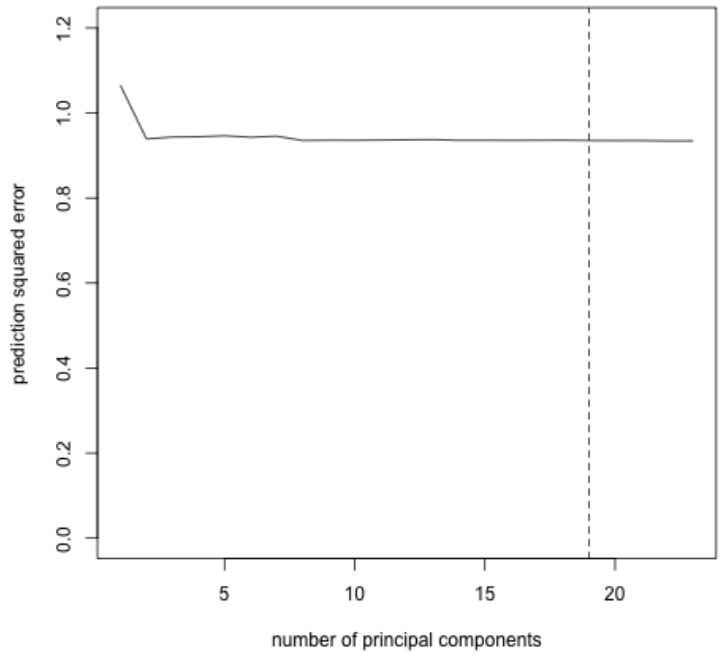
- For given $r$, RR estimates have less bias than PCR estimates.
- PCR using $r$ components has $r$ degrees of freedom for variance.
- We fixed $r$ such that degrees of freedom of the ridge model using $r$ components equals $r$.

$$\text{tr}\left(HH'\right) = r$$



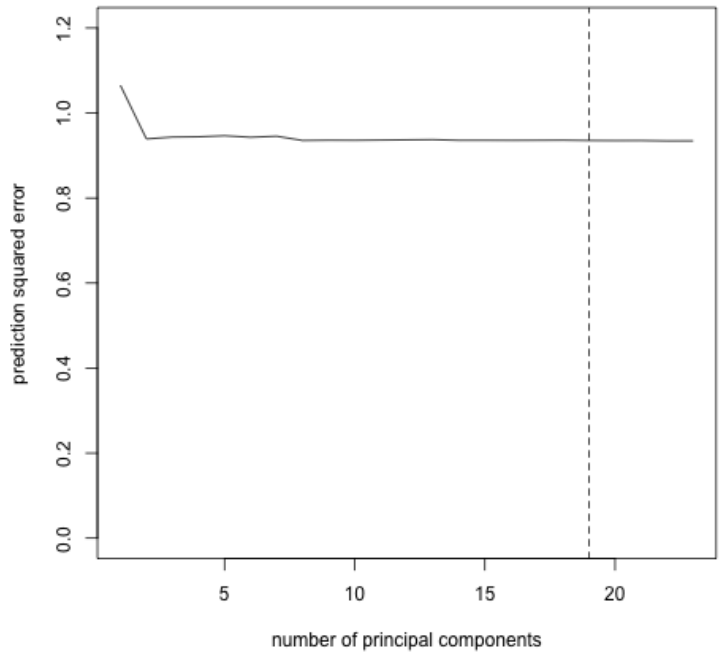
# Simulation Study

Mean prediction squared error:

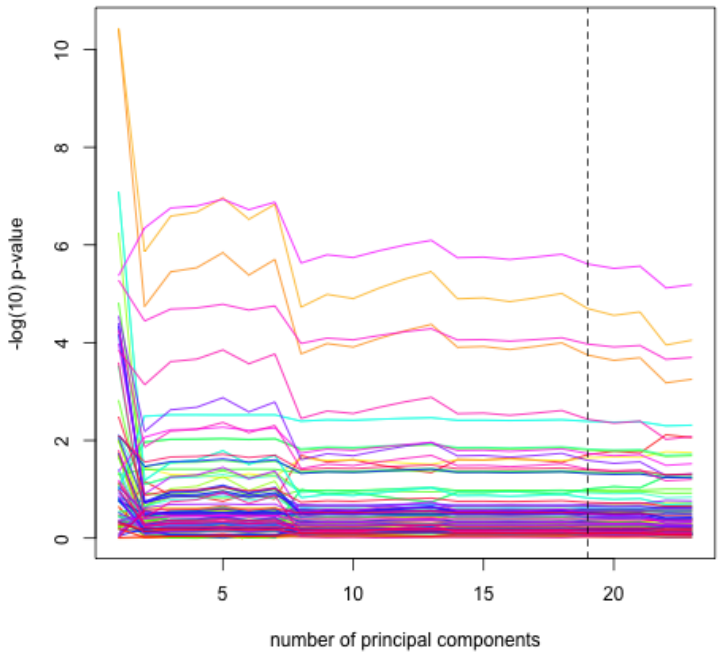


# Simulation Study

Mean prediction squared error:




*p*-value trace:



# Simulation study

- Performance comparison
  - SNP ranking followed by multivariate regression
  - HyperLasso
- Continuous and binary outcomes

| | Univariate | | | | | HLasso | RR |
|---|---|---|---|---|---|---|---|
| % of SNPs ranked by univariate *p*-value | 0.1% | 0.5% | 1 % | 3% | 4% | | |
| Continuous outcomes (mean PSE) | 1.51 | 1.55 | 1.54 | 2.21 | 3.93 | 2.41 | 1.23 |
| Binary outcomes (mean CE) | 0.49 | 0.48 | 0.48 | 0.49 | 0.50 | 0.50 | 0.46 |

# Bipolar Disorder Data

- Two GWAS of Bipolar Disorder: WTCCC and GAIN.
- Case-control studies - model extended to logistic ridge regression.
- SNPs typed on different platforms. Impute2 to obtain common SNPs.
- When determining shrinkage parameter, training data were thinned (1 SNP every 100kb).
- Univariate model - which significance threshold?
- HyperLasso - cross-validation to choose the parameters is computationally intensive.

# Bipolar Disorder Data

- Two GWAS of Bipolar Disorder: WTCCC and GAIN.
- Case-control studies - model extended to logistic ridge regression.
- SNPs typed on different platforms. Impute2 to obtain common SNPs.
- When determining shrinkage parameter, training data were thinned (1 SNP every 100kb).
- Univariate model - which significance threshold?
- HyperLasso - cross-validation to choose the parameters is computationally intensive.

| | Univariate | | | HyperLasso | Ridge Regression |
|---|---|---|---|---|---|
| $p$-value threshold | $10^{-5}$ | $10^{-7}$ | $10^{-10}$ | | |
| Mean Classification Error | 0.489 | 0.491 | 0.490 | 0.492 | 0.465 |

# Outline

# Significance testing in ridge regression

- Ridge regression is not a variable selection method - the shrinkage penalty does not shrink any coefficient estimates to zero.

# Significance testing in ridge regression

- Ridge regression is not a variable selection method - the shrinkage penalty does not shrink any coefficient estimates to zero.
- A test of significance of ridge regression coefficients had been proposed (Halawa & El Bassiouni, 2000) and applied (Malo et al, 2008) but not evaluated.

# Significance testing in ridge regression

- Ridge regression is not a variable selection method - the shrinkage penalty does not shrink any coefficient estimates to zero.
- A test of significance of ridge regression coefficients had been proposed (Halawa & El Bassiouni, 2000) and applied (Malo et al, 2008) but not evaluated.
- We extended the test to be applicable when $p >> n$ and to be applied in logistic ridge regression, and evaluated its performance on simulated and real data sets.

# Significance test

Based on a Wald test:

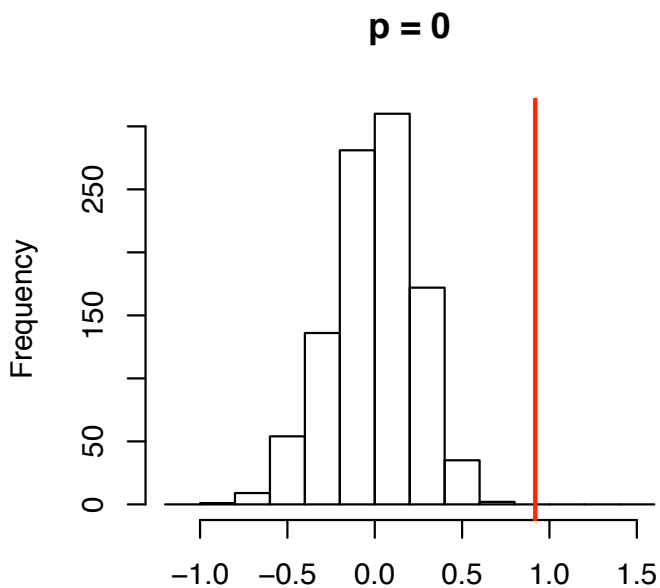
$$T_k = \frac{\hat{\beta}_k}{\text{se}\left(\hat{\beta}_k\right)} \qquad H_0 : T_k \sim \mathcal{N}(0,1)$$

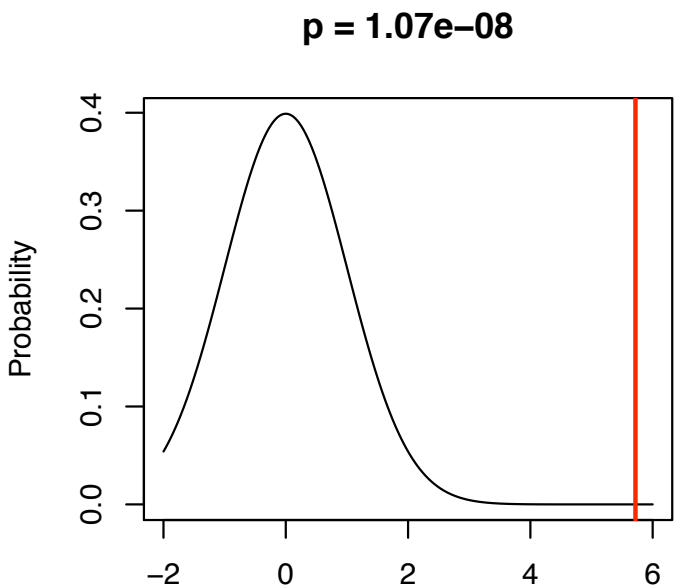
$\text{se}\left(\hat{\beta}_k\right)$ from covariance matrix

$$\text{Var}\left(\hat{\beta}_k\right) = \hat{\sigma}^2 (X'X + kI)^{-1} X'X (X'X + kI)^{-1}$$

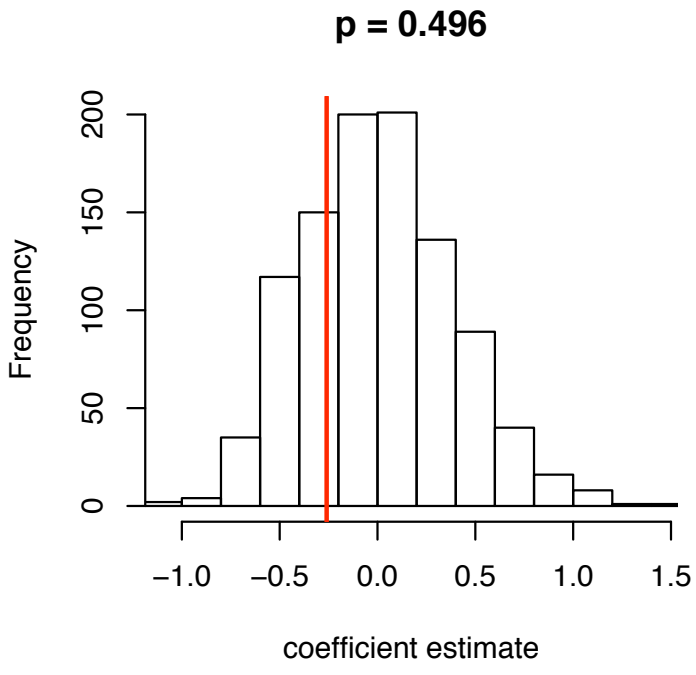
taking into account both correlation in predictors and amount of shrinkage.

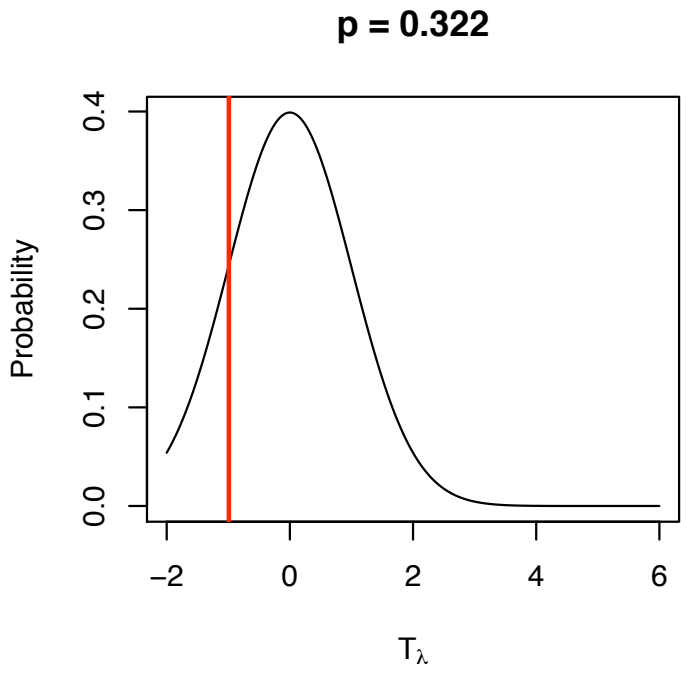
# Simulation study
## Causal SNP



# Simulation study
## Non-causal SNP



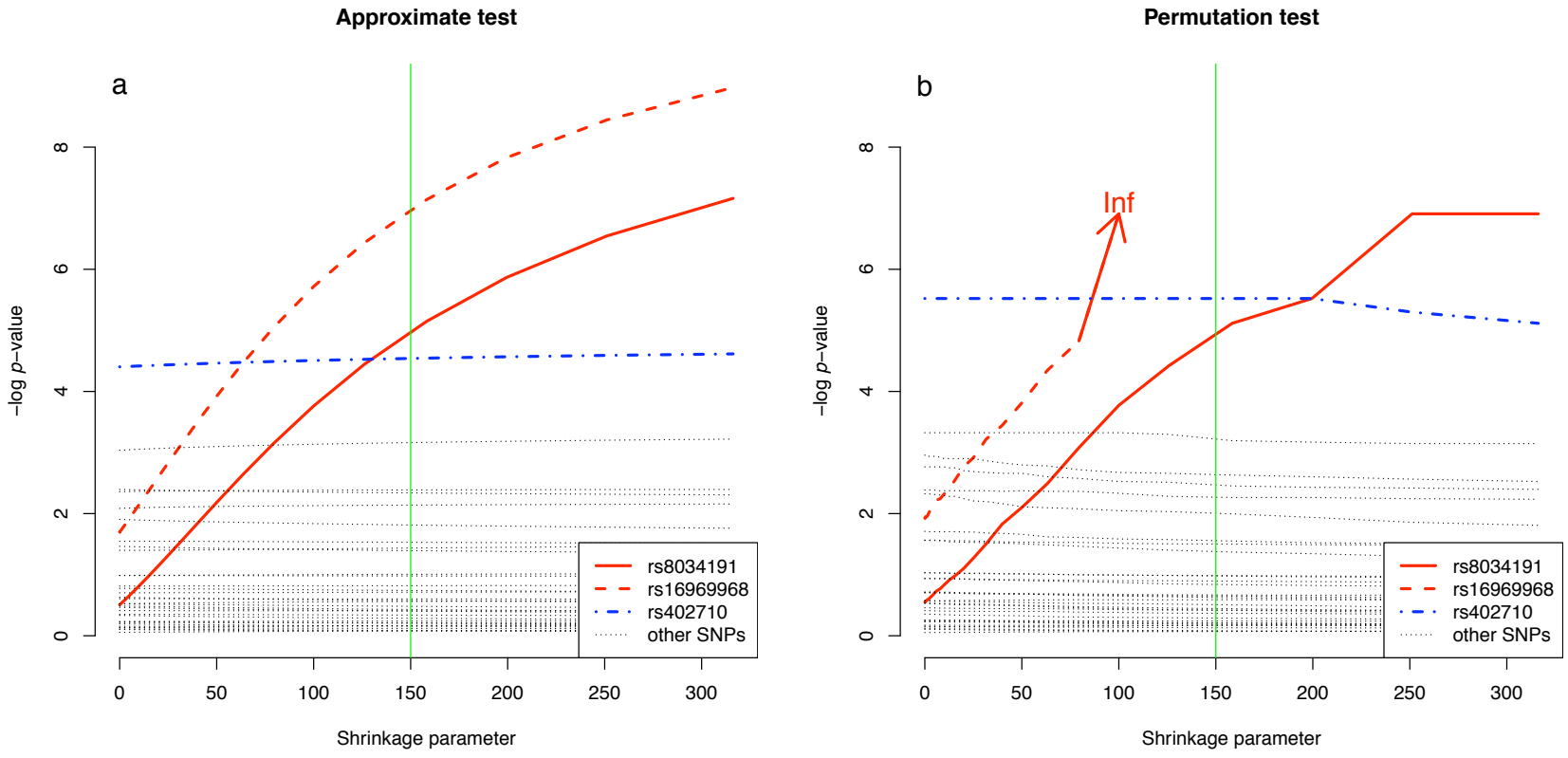



# Simulation study

## True-positive and False-positive rates

| | | | Shrinkage Parameter | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Approximate test | | | | Permutation test | | | |
| | | | 0.1 | 1 | 10 | 100 | 0.1 | 1 | 10 | 100 |
| **Individuals** | **SNPs** | | | | | | | | | |
| 500 | 20 | | | | | | | | | |
| | | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.045 | 0.045 | 0.061 | 0.133 | 0.015 | 0.015 | 0.017 | 0.095 |
| | 100 | | | | | | | | | |
| | | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.056 | 0.054 | 0.071 | 0.141 | 0.015 | 0.018 | 0.024 | 0.074 |
| | 1000 | | | | | | | | | |
| | | TPR | 0.100 | 0.500 | 0.900 | 1.000 | 0.000 | 0.200 | 0.800 | 1.000 |
| | | FPR | 0.038 | 0.045 | 0.049 | 0.080 | 0.007 | 0.006 | 0.010 | 0.029 |
| | ALL | | | | | | | | | |
| | | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.318 | 0.071 | 0.068 | 0.069 | 0.019 | 0.019 | 0.020 | 0.020 |
| 5000 | 20 | | | | | | | | | |
| | | TPR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | FPR | 0.048 | 0.048 | 0.048 | 0.113 | 0.006 | 0.006 | 0.006 | 0.053 |
| | 100 | | | | | | | | | |
| | | TPR | 0.900 | 0.900 | 1.000 | 1.000 | 0.800 | 0.900 | 1.000 | 1.000 |
| | | FPR | 0.055 | 0.052 | 0.062 | 0.100 | 0.003 | 0.001 | 0.007 | 0.055 |
| | 1000 | | | | | | | | | |
| | | TPR | 0.700 | 0.700 | 1.000 | 1.000 | 0.700 | 0.700 | 0.900 | 1.000 |
| | | FPR | 0.046 | 0.046 | 0.045 | 0.060 | 0.006 | 0.007 | 0.008 | 0.014 |
| | ALL | | | | | | | | | |
| | | TPR | 0.400 | 0.500 | 0.900 | 1.000 | 0.300 | 0.900 | 0.900 | 1.000 |
| | | FPR | 0.026 | 0.027 | 0.029 | 0.042 | 0.007 | 0.007 | 0.007 | 0.009 |

# Lung Cancer Data



# Outline

# Summary

- Prediction is a challenging problem!

# Summary

- Prediction is a challenging problem!
- Ridge regression is a popular penalized regression approach that has been shown to perform well for prediction.

# Summary

- Prediction is a challenging problem!
- Ridge regression is a popular penalized regression approach that has been shown to perform well for prediction.
- We propose a semi-automatic method for choosing the shrinkage parameter in ridge regression, which can be applied when $p \gg n$.

# Summary

- Prediction is a challenging problem!
- Ridge regression is a popular penalized regression approach that has been shown to perform well for prediction.
- We propose a semi-automatic method for choosing the shrinkage parameter in ridge regression, which can be applied when $p \gg n$.
- We introduced a method for testing the significance of regression coefficients estimated using ridge regression.

# Summary

- Prediction is a challenging problem!
- Ridge regression is a popular penalized regression approach that has been shown to perform well for prediction.
- We propose a semi-automatic method for choosing the shrinkage parameter in ridge regression, which can be applied when $p \gg n$.
- We introduced a method for testing the significance of regression coefficients estimated using ridge regression.
- We have enabled ridge regression to be a feasible tool for genetic risk prediction on a genome-wide scale.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).
- Where available, multi-core or GPU computation speeds up matrix operations.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).
- Where available, multi-core or GPU computation speeds up matrix operations.
- Flexibility to include non-genetic covariates - penalized or not.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).
- Where available, multi-core or GPU computation speeds up matrix operations.
- Flexibility to include non-genetic covariates - penalized or not.
- Significance test is implemented.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).
- Where available, multi-core or GPU computation speeds up matrix operations.
- Flexibility to include non-genetic covariates - penalized or not.
- Significance test is implemented.
- Graphical outputs - ridge and *p*-value traces.

# R package `ridge`

- Fitting ridge regression models to data comprising hundreds of thousands of predictors presents computational challenges.
- We have written an R package, `ridge`, for fitting such models.
- For large data sets, C code is used (with a user-friendly R interface).
- Where available, multi-core or GPU computation speeds up matrix operations.
- Flexibility to include non-genetic covariates - penalized or not.
- Significance test is implemented.
- Graphical outputs - ridge and *p*-value traces.
- Option for user-specified shrinkage parameter, with our semi-automatic method as the default.

# Acknowledgements


- Maria De Iorio
- Colleagues in the Department of Epidemiology and Biostatistics, Imperial College London
- Colleagues in the Department of Statistical Science, University College London
- ILCO study nested within EPIC
- WTCCC and GAIN studies

Supported by wellcome trust

# References


[1] D. G Clayton.
Prediction and interaction in complex disease genetics: experience in type 1 diabetes.
*PLoS Genetics*, 2009.

[2] Erika Cule and Maria De Iorio.
A semi-automatic method to guide the choice of ridge parameter in ridge regression.
*arXiv*, stat.AP, May 2012.

[3] Erika Cule, Paolo Vineis, and Maria De Iorio.
Significance testing in ridge regression for genetic data.
*BMC Bioinformatics*, 12(1):372, 2011.

[4] Ildiko Frank and Jerome Friedman.
A statistical view of some chemometrics regression tools.
*Technometrics*, 35(2):109–135, May 1993.

[5] A M Halawa and M Y El Bassiouni.
Tests of regression coefficients under ridge regression models.
*Journal of Statistical Computation and Simulation*, 65(1):341–356, 2000.

[6] T Hastie and R Tibshirani.
*Generalized Additive Models*.
Chapman & Hall, 1990.

[7] Arthur E Hoerl and RW Kennard.
Ridge regression: Biased estimation for nonorthogonal problems.
*Technometrics*, 12(1):55–67, 1970.

[8] Clive J Hoggart, JC Whittaker, M De Iorio, and David J Balding.
Simultaneous analysis of all snps in genome-wide and re-sequencing association studies.
*PLoS Genet*, 4(7):e1000130, 2008.

[9] Nathalie Malo, Ondrej Libiger, and Nicholas J Schork.
Accommodating linkage disequilibrium in genetic-association analyses via ridge regression.
*Am J Hum Genet*, 82(2):375–385, Feb 2008.

[10] Robert Tibshirani.
Regression shrinkage and selection via the lasso.
*Journal of the Royal Statistical Society B*, 58:267–288, 1996.

[11] Hui Zou and Trevor Hastie.
Regularization and variable selection via the elastic net.
*Journal of the Royal Statistical Society B*, 67(2):301–320, Jan 2005.