Supplemental Material for *EigenPrism: Inference for High-Dimensional Signal-to-Noise Ratios*

Lucas Janson, Rina Foygel Barber, Emmanuel Candès

A Inference for θ^2 under non-Gaussian design with known variance

The method of Section 2.1 also works asymptotically under more general conditions than the Gaussianity assumptions of (1.4). Let $\boldsymbol{z} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the statement that $\boldsymbol{z}$ has some distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Consider again the linear model (1.1) but with relaxed assumptions,

$$\boldsymbol{x}_i \overset{i.i.d.}{\sim} (\boldsymbol{\mu}, \boldsymbol{I}_p - \boldsymbol{\mu}\boldsymbol{\mu}^\top), \qquad \varepsilon_i \overset{i.i.d.}{\sim} (0, \sigma^2),$$

again with σ^2 known and $\boldsymbol{X}$ independent of $\boldsymbol{\varepsilon}$. Under this model, we get that

$$y_i^2 \overset{i.i.d.}{\sim} (\theta^2 + \sigma^2, v_1)$$

and the asymptotic distribution in (2.1) in turn becomes, by the CLT,

$$\frac{1}{\sqrt{n}}\|\boldsymbol{y}\|_2^2 - \sqrt{n}\left(\theta^2 + \sigma^2\right) \xrightarrow{\mathcal{D}} N(0, v_1), \tag{A.1}$$

as $n \to \infty$, where v_1 does not depend on n but does depend on the unknown $\boldsymbol{\beta}$, and is given by

$$\begin{aligned} v_1 = \mathbb{E}\left(\varepsilon_i^4\right) + 4\sigma^2\left[\theta^2 - \left(\boldsymbol{\mu}^\top\boldsymbol{\beta}\right)^2\right] + 4\mathbb{E}\left(\varepsilon_i^3\right)\boldsymbol{\mu}^\top\boldsymbol{\beta} + \mathbb{E}\left[\left(\boldsymbol{x}_i^\top\boldsymbol{\beta}\right)^4\right] \\ - \sigma^4 - \theta^4 - \left(\boldsymbol{\mu}^\top\boldsymbol{\beta}\right)^4 + 2\theta^2\left(\boldsymbol{\mu}^\top\boldsymbol{\beta}\right)^2 \end{aligned}$$

In order to be less parametric, we can consider bootstrap confidence intervals based on the above calculations. Corresponding to (A.1) we can get an unbiased statistic,

$$\begin{aligned} T_1 &:= \frac{1}{n}\|\boldsymbol{y}\|_2^2 - \sigma^2, \\ \mathbb{E}\left(T_1\right) &= \theta^2, \\ \mathrm{SD}(T_1) &= \sqrt{v_1/n}, \end{aligned} \tag{A.2}$$

whose distribution we may hope to be close to Gaussian. T_1 can be bootstrapped (potentially with standard bias-correction and acceleration) to obtain bootstrap CIs, nonparametrically dealing with the unknown variance v_1. We ran simulations with $n = 800$, $p = 1500$, $\boldsymbol{X}$ having i.i.d. Bernoulli(0.05) entries (the columns of $\boldsymbol{X}$ were then standardized to have mean 0 and variance 1), $\theta^2 = \sigma^2 = 10$, and ε_i i.i.d. t_5 (rescaled to have variance 10). We generated a single $\boldsymbol{\beta}$ uniformly on the θ-radius sphere and ran 1000 simulations (so that $\boldsymbol{\beta}$ did not change across simulations). Bias-corrected, accelerated 95% bootstrap CIs achieved 93.8% coverage (this is within statistical uncertainty of the nominal 95%, as a 95% CI for the CI coverage is $[0.923, 0.953]$).

B Calculation of variance of EigenPrism estimator

In this section we calculate the variance of the statistic $S = \sum_{i=1}^n w_i z_i^2$ when conditioning on $\boldsymbol{d}$. Here we treat $\boldsymbol{w}$ as fixed, but note that since we condition on $\boldsymbol{d}$, this includes values of $\boldsymbol{w}$ that are calculated as a function of $\boldsymbol{d}$, as in the EigenPrism method.

$$
\begin{aligned}
\operatorname{Var}(S|\boldsymbol{d}) &= \operatorname{Var}\left(\sum_{i=1}^{n} w_i z_i^2 \middle| \boldsymbol{d}\right) \\
&= \sum_{i=1}^{n} w_i^2 \operatorname{Var}\left(z_i^2 \middle| \boldsymbol{d}\right) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} w_i w_j \operatorname{Cov}\left(z_i^2, z_j^2 \middle| \boldsymbol{d}\right) .
\end{aligned}
$$

We now calculate each term. Recall that $\lambda_i := d_i^2/p$ for $i = 1, \ldots, n$. Then

$$
\begin{aligned}
\operatorname{Var}\left(z_i^2 \middle| \boldsymbol{d}\right) &= \mathbb{E}\left(z_i^4 \middle| \boldsymbol{d}\right) - \mathbb{E}\left(z_i^2 \middle| \boldsymbol{d}\right)^2 \\
&= \mathbb{E}\left[(d_i\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle + \epsilon_i)^4 \middle| \boldsymbol{d}\right] - \left(\lambda_i\theta^2 + \sigma^2\right)^2
\end{aligned}
$$

Using $\epsilon_i \sim N(0, \sigma^2)$ (and the fact that $\boldsymbol{\epsilon} \perp\!\!\!\perp \boldsymbol{V}$),

$$
= \mathbb{E}\left[(d_i\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle)^4 \middle| \boldsymbol{d}\right] + 6\sigma^2 \mathbb{E}\left[(d_i\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle)^2 \middle| \boldsymbol{d}\right] + 3\sigma^4 - \left(\lambda_i\theta^2 + \sigma^2\right)^2
$$

Using $\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 \sim \theta^2 \cdot \mathsf{Beta}\left(\frac{1}{2}, \frac{p-1}{2}\right)$,

$$
\begin{aligned}
&= d_i^4\theta^4 \cdot \frac{1\cdot 3}{p\cdot(p+2)} + 6\sigma^2\lambda_i\theta^2 + 3\sigma^4 - \left(\lambda_i\theta^2 + \sigma^2\right)^2 \\
&= 2\lambda_i^2\theta^4 \frac{p-1}{p+2} + 4\sigma^2\lambda_i\theta^2 + 2\sigma^4
\end{aligned}
$$

Also, for $i \neq j$,

$$
\begin{aligned}
\operatorname{Cov}\left(z_i^2, z_j^2 \middle| \boldsymbol{d}\right) &= \mathbb{E}\left(z_i^2 z_j^2 \middle| \boldsymbol{d}\right) - \mathbb{E}\left(z_i^2 \middle| \boldsymbol{d}\right)\mathbb{E}\left(z_j^2 \middle| \boldsymbol{d}\right) \\
&= \mathbb{E}\left[(d_i\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle + \epsilon_i)^2 (d_j\langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle + \epsilon_j)^2 \middle| \boldsymbol{d}\right] - \left(\lambda_i\theta^2 + \sigma^2\right)\left(\lambda_j\theta^2 + \sigma^2\right)
\end{aligned}
$$

Using $\epsilon_i, \epsilon_j \overset{\text{iid}}{\sim} N(0, \sigma^2)$ (and the fact that $\boldsymbol{\epsilon} \perp\!\!\!\perp \boldsymbol{V}$),

$$
\begin{aligned}
&= \mathbb{E}\left(d_i^2 d_j^2 \langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right) + \sigma^2 \mathbb{E}\left(d_i^2 \langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right) + \sigma^2 \mathbb{E}\left(d_j^2 \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right) \\
&\quad + \sigma^4 - \left(\lambda_i\theta^2 + \sigma^2\right)\left(\lambda_j\theta^2 + \sigma^2\right)
\end{aligned}
$$

Using $\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2, \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \sim \theta^2 \cdot \mathsf{Beta}\left(\frac{1}{2}, \frac{p-1}{2}\right)$,

$$
\begin{aligned}
&= \mathbb{E}\left(d_i^2 d_j^2 \langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right) + \sigma^2\lambda_i\theta^2 + \sigma^2\lambda_j\theta^2 + \sigma^4 \\
&\quad - \left(\lambda_i\theta^2 + \sigma^2\right)\left(\lambda_j\theta^2 + \sigma^2\right) \\
&= d_i^2 d_j^2 \mathbb{E}\left(\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right) - \lambda_i\lambda_j\theta^4 \\
&= d_i^2 d_j^2 \operatorname{Cov}\left(\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2, \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2 \middle| \boldsymbol{d}\right)
\end{aligned}
$$

Using $\left(\langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2, \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2, \theta^2 - \langle \boldsymbol{V}_i, \boldsymbol{\beta}\rangle^2 - \langle \boldsymbol{V}_j, \boldsymbol{\beta}\rangle^2\right) \sim \theta^2 \cdot \mathsf{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}, \frac{p-2}{2}\right)$,

$$
= \frac{-2}{p+2}\lambda_i\lambda_j\theta^4.
$$

Then,

$$
\begin{aligned}
\operatorname{Var}\left(S \middle| \boldsymbol{d}\right) &= \sum_{i=1}^{n} w_i^2 \left(2\lambda_i^2\theta^4 \frac{p-1}{p+2} + 4\sigma^2\lambda_i\theta^2 + 2\sigma^4 \right) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} w_i w_j \left(\frac{-2}{p+2}\lambda_i\lambda_j\theta^4 \right) \\
&= \sum_{i=1}^{n} w_i^2 \left(2\lambda_i^2\theta^4 \frac{p}{p+2} + 4\sigma^2\lambda_i\theta^2 + 2\sigma^4 \right) + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left(\frac{-2}{p+2}\lambda_i\lambda_j\theta^4 \right) \\
&= 2\sigma^4 \sum_{i=1}^{n} w_i^2 + 4\sigma^2\theta^2 \sum_{i=1}^{n} w_i^2\lambda_i + 2\theta^4 \left[\frac{p}{p+2} \sum_{i=1}^{n} w_i^2\lambda_i^2 - \frac{\left(\sum_{i=1}^{n} w_i\lambda_i\right)^2}{p+2} \right].
\end{aligned}
$$

C Proof of Asymptotic Normality of T_2 and T_3

Proof. First consider $T_2(\boldsymbol{y}, \boldsymbol{X})$ as a deterministic function of the random $\boldsymbol{y}$ and $\boldsymbol{X}$. Then for any constant c,

$$T_2(c\boldsymbol{y}, \boldsymbol{X}) = c^2 T_2(\boldsymbol{y}, \boldsymbol{X}). \tag{C.1}$$

Note that $c\boldsymbol{y}$ also follows a linear model, only with θ^2 replaced by $c^2\theta^2$ and σ^2 replaced by $c^2\sigma^2$. Thus by taking $c = 1/\max\{\theta, \sigma\}$ we may treat θ^2 and σ^2 as belonging to $[0, 1]$ in order to prove asymptotic normality of $T_2(c\boldsymbol{y}, \boldsymbol{X})$, which by Equation (C.1) implies asymptotic normality of $T_2(\boldsymbol{y}, \boldsymbol{X})$. The same argument holds for T_3, and so without loss of generality, in the remainder of the proof we assume θ^2 and σ^2 are both bounded. We have assumed $\max\{\theta^2, \sigma^2\} > 0$, as the case $\theta^2 = \sigma^2 = 0$ is immediately identifiable because $\boldsymbol{y} \equiv \boldsymbol{0}$, and trivial.

Recall that because V is Haar-distributed,

$$(\boldsymbol{V}_1^\top \boldsymbol{\beta}, \ldots, \boldsymbol{V}_n^\top \boldsymbol{\beta}) \stackrel{d}{=} \theta/\|\boldsymbol{u}\| \cdot (u_1, \ldots, u_n),$$

where $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$. From this, we can rewrite T_2 as:

$$\begin{aligned}
T_2 - \mathbb{E}(T_2) &= \sum_{i=1}^n w_i \left(\sqrt{\lambda_i}\theta \frac{u_i}{\|\boldsymbol{u}\|/\sqrt{p}} + \varepsilon_i \right)^2 - \sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2 \right) \\
&= \frac{1}{\|\boldsymbol{u}\|^2/p} \cdot \sum_{i=1}^n w_i \left(\sqrt{\lambda_i}\theta u_i + \varepsilon_i + \varepsilon_i \left(\|\boldsymbol{u}\|/\sqrt{p} - 1 \right) \right)^2 - \left(1 + \frac{1}{\|\boldsymbol{u}\|^2/p} - \frac{1}{\|\boldsymbol{u}\|^2/p} \right) \sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2 \right) \\
&= \frac{1}{\|\boldsymbol{u}\|^2/p} \cdot \left(\sum_{i=1}^n w_i \left(\sqrt{\lambda_i}\theta u_i + \varepsilon_i \right)^2 - \sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2 \right) \right) && \text{(C.2)} \\
&\quad + \frac{2}{\|\boldsymbol{u}\|^2/p} \left(\|\boldsymbol{u}\|/\sqrt{p} - 1 \right) \sum_{i=1}^n w_i \left(\sqrt{\lambda_i}\theta u_i \varepsilon_i + \varepsilon_i^2 - \sigma^2 \right) && \text{(C.3)} \\
&\quad + \frac{2\sigma^2}{\|\boldsymbol{u}\|^2/p} \left(\|\boldsymbol{u}\|/\sqrt{p} - 1 \right) \sum_{i=1}^n w_i && \text{(C.4)} \\
&\quad + \frac{1}{\|\boldsymbol{u}\|^2/p} \left(\|\boldsymbol{u}\|/\sqrt{p} - 1 \right)^2 \sum_{i=1}^n w_i \left(\varepsilon_i^2 - \sigma^2 \right) && \text{(C.5)} \\
&\quad + \frac{\sigma^2}{\|\boldsymbol{u}\|^2/p} \left(\|\boldsymbol{u}\|/\sqrt{p} - 1 \right)^2 \sum_{i=1}^n w_i && \text{(C.6)} \\
&\quad + \left(1 - \frac{1}{\|\boldsymbol{u}\|^2/p} \right) \sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2 \right). && \text{(C.7)}
\end{aligned}$$

Our goal is to show that the right-hand side of (C.2) converges to Gaussian, while (C.3)–(C.7) each converge to zero in probability. In particular, using certain probabilistic properties of the λ_i's and w_i's (which are independent of the other random variables), we will show convergence *conditional* on the λ_i and w_i. We first prove the result for T_2 and then explain the (minor) changes needed to prove the same for T_3 (for which Equation (C.2)–(C.7) also holds).

Before either, however, we need a few tools, including the following Lemma:

Lemma 1. *For both T_2 and T_3, there exist constants a and b such that,*

$$\mathbb{P}\left\{ \forall i,\ n|w_i| \leq \frac{a + b\lambda_i}{\min\{\lambda_i^2, 1\}} \right\} \longrightarrow 1.$$

Proof. We defer the proof to the end of this section.

Note that by convergence of the moments of the λ_i to those of the Marčenko–Pastur (MP) distribution, Lemma 1 implies that

$$\sum_{i=1}^n |w_i|^3 \lambda_i^r \in O_p(n^{-2}) \tag{C.8}$$

for any $r \in \mathbb{R}$. Note also that by the Cauchy-Schwarz inequality,

$$\sum_{i=1}^n w_i^2 \lambda_i^r \geq \frac{\left(\sum_{i=1}^n w_i \lambda_i \right)^2}{\sum_{i=1}^n \lambda_i^{2(1-r)}} = \frac{1}{\sum_{i=1}^n \lambda_i^{2(1-r)}} \in \Omega_p(n^{-1}) \tag{C.9}$$

for any $r \in \mathbb{R}$. Finally, note that $\|\boldsymbol{u}\|/\sqrt{p} \xrightarrow{p} 1$.

Starting from the bottom, (C.7) converges in probability to zero because $\left(1 - \frac{1}{\|\boldsymbol{u}\|^2/p}\right) \xrightarrow{p} 0$ and $\sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2\right) = \theta^2$, a constant. (C.6) and (C.4) equal zero because $\sum_{i=1}^n w_i = 0$.

In (C.5), we seek to show that $\sum_{i=1}^n w_i \left(\varepsilon_i^2 - \sigma^2\right)$ converges in distribution, so that by Slutsky's Theorem, (C.5) converges to zero in probability. The summands $w_i \left(\varepsilon_i^2 - \sigma^2\right)$ are independent and mean-zero with variance $2\sigma^4 w_i^2$. By Lyapunov's central limit theorem (Billingsley, 1995, p. 362), we just need to establish the Lyapunov condition:

$$\frac{\sum_{i=1}^n \mathbb{E}\left(|w_i \left(\varepsilon_i^2 - \sigma^2\right)|^3\right)}{\left(\sum_{i=1}^n \operatorname{Var}\left(w_i \left(\varepsilon_i^2 - \sigma^2\right)\right)\right)^{3/2}} \propto \frac{\sum_{i=1}^n |w_i|^3}{\left(\sum_{i=1}^n w_i^2\right)^{3/2}} \in O_p(n^{-1/2}),$$

where the $\in$ follows from (C.8) and (C.9).

In (C.3), we similarly seek to show that the sum converges in distribution, allowing us to again use Slutsky's Theorem to show (C.3) converges to zero in probability. The argument is nearly the same as that for (C.5), using various different values of r in (C.8) and (C.9) to establish the Lyapunov condition.

Lastly for (C.2), by Slutsky's Theorem (Lehman and Romano, 2005, p. 433), it suffices to show that $\sum_{i=1}^n w_i \left(\sqrt{\lambda_i}\theta u_i + \varepsilon_i\right)^2 - \sum_{i=1}^n w_i \left(\lambda_i \theta^2 + \sigma^2\right)$ converges in distribution to a Gaussian random variable, which can again be established using Lyapunov's central limit theorem in nearly the same way as in the argument for (C.5). Note that the resulting variance expression $\sum_{i=1}^n \operatorname{Var}\left(w_i \left(\sqrt{\lambda_i}\theta u_i + \varepsilon_i\right)^2\right) = \sum_{i=1}^n 2\left(\lambda_i \theta^2 + \sigma^2\right)^2$ is not identical to the variance of S in (2.4), but the quotient of the two expressions converges to 1 as $n, p \longrightarrow \infty$.

Only a few changes to the above proof are needed for establishing asymptotic normality of T_3. First, an analogue to Equation (C.9) can be shown:

$$\sum_{i=1}^n w_i^2 \lambda_i^r \geq \frac{\left(\sum_{i=1}^n w_i\right)^2}{\sum_{i=1}^n \lambda_i^{-2r}} = \frac{1}{\sum_{i=1}^n \lambda_i^{-2r}} \in \Omega_p(n^{-1}). \tag{C.10}$$

Next, in each of (C.4),(C.6), and (C.7), the sum equals a constant while the coefficient in front of the sum converges in probability to zero. The arguments for (C.2), (C.3), and (C.5) take the same form as for T_2 except using (C.8) and (C.10) instead of (C.9) to establish the Lyapunov condition. □

Proof of Lemma 1. We start by slightly rewriting the optimization program $\mathcal{P}_1$:

$$\operatorname*{arg\,min}_{\boldsymbol{w} \in \mathbb{R}^n} t \quad \text{such that} \sum_{i=1}^n w_i^2 \leq t, \ \sum_{i=1}^n w_i^2 \lambda_i^2 \leq t, \ \sum_{i=1}^n w_i = 0, \ \sum_{i=1}^n w_i \lambda_i = 1. \tag{C.11}$$

By the Karush-Kuhn-Tucker conditions for (C.11), the gradient of the Lagrangian with respect to $(t, w_1, \ldots, w_n)$ vanishes, i.e.,

$$1 - \delta_1 - \delta_2 = 0, \tag{C.12}$$

$$w_i \left(2\delta_1 + 2\delta_2 \lambda_i^2\right) + \kappa_1 + \kappa_2 \lambda_i = 0, \tag{C.13}$$

where $\delta_1 \geq 0$ and $\delta_2 \geq 0$ are the dual variables corresponding to the inequalities and κ_1 and κ_2 are the dual variables corresponding to the equalities. Rearranging Equation (C.13),

$$w_i = \frac{-\kappa_1 - \kappa_2 \lambda_i}{2\delta_1 + 2\delta_2 \lambda_i^2}. \tag{C.14}$$

By Equation (C.12) and dual positivity constraints, we have $\delta_1, \delta_2 \in [0, 1]$. Observe that

$$\min_{\delta_1 \in [0,1], \delta_2 = 1 - \delta_1} \delta_1 + \delta_2 \lambda_i^2 = \min\{\lambda_i^2, 1\},$$

establishing a lower-bound on the denominator. Now it suffices to show that $|\kappa_1|, |\kappa_2| \in O_p(1/n)$.

Multiplying Equation (C.13) by w_i and summing over i,

$$2\delta_1 \sum_{i=1}^n w_i^2 + 2\delta_2 \sum_{i=1}^n w_i^2 \lambda_i^2 + \kappa_1 \sum_{i=1}^n w_i + \kappa_2 \sum_{i=1}^n w_i \lambda_i = 0. \tag{C.15}$$

By recalling that Equation (2.11) established that $\max\{\sum_{i=1}^n w_i^2, \sum_{i=1}^n w_i^2\lambda_i^2\} \in O_p(1/n)$ and the constraints $\sum_{i=1}^n w_i = 0$ and $\sum_{i=1}^n w_i\lambda_i = 1$, we have that $|\kappa_2| \in O_p(1/n)$. Next, by just summing Equation (C.13) over i,

$$2\delta_1 \sum_{i=1}^n w_i + 2\delta_2 \sum_{i=1}^n w_i\lambda_i^2 + n\kappa_1 + \kappa_2 \sum_{i=1}^n \lambda_i = 0. \tag{C.16}$$

By Cauchy-Schwarz, $\sum_{i=1}^n w_i\lambda_i^2 \leq \sqrt{\sum_{i=1}^n w_i^2 \sum_{i=1}^n \lambda_i^4} \in O_p(1)$, and using that $|\kappa_2| \in O_p(1/n)$ and $\sum_{i=1}^n \lambda_i \in O_p(n)$, we find that $|\kappa_1| \in O_p(1/n)$ and the Lemma is proved for T_2.

To see the same result for T_3, first note that rewriting $\mathcal{P}_2$ analogously to (C.11) gives the same gradient for the Lagrangian, so that Equations (C.12) and (C.13) still hold with the same implications for the denominator of w_i in Equation (C.14), so all that remains is again showing that $|\kappa_1|, |\kappa_2| \in O_p(1/n)$.

We will need an analogue to Equation (2.11) for T_3 to show that $\max\{\sum_{i=1}^n w_i^2, \sum_{i=1}^n w_i^2\lambda_i^2\} \in O_p(1/n)$. The proof of Equation (2.11) can be found in Appendix D, and follows from the construction of a simple set of weights $\tilde{w}_i$ satisfying the constraints of $\mathcal{P}_1$. By considering instead the set of weights

$$\breve{w}_i := \frac{\lambda_i^{-1}}{\sum_{j=1}^{n/2} \lambda_j^{-1} - \sum_{j=n/2+1}^{n} \lambda_j^{-1}} \cdot \begin{cases} +1, & \text{for } i \leq n/2, \\ -1, & \text{for } i > n/2, \end{cases}$$

satisfying the constraints of $\mathcal{P}_2$, one can follow the same steps to establish $\max\{\sum_{i=1}^n w_i^2, \sum_{i=1}^n w_i^2\lambda_i^2\} \in O_p(1/n)$ for T_3. Using this and the constraints of $\mathcal{P}_2$, Equation (C.15) establishes $|\kappa_1| \in O_p(1/n)$. Using this result and the same methods as for T_2, Equation (C.16) establishes $|\kappa_2| \in O_p(1/n)$, and the Lemma is proved. □

D Variance upper-bound for T_2

In this section we derive the upper bound (2.11) on the variance of the statistic T_2. For simplicity we assume that n is even.

We begin by constructing a vector of weights $\tilde{\boldsymbol{w}}$:

$$\tilde{w}_i := \frac{1}{\sum_{j=1}^{n/2} \lambda_j - \sum_{j=n/2+1}^{n} \lambda_j} \cdot \begin{cases} +1, & \text{for } i \leq n/2, \\ -1, & \text{for } i > n/2. \end{cases}$$

Note that $\tilde{\boldsymbol{w}}$ satisfies the constraints of the optimization problem (2.8), and thus $\mathrm{Var}(T_2)$ is upper-bounded by Equation (2.7) with $\tilde{\boldsymbol{w}}$ plugged in. A second key observation is that we know from random matrix theory that for $n, p \to \infty$ and $n/p \to \gamma \in (0,1)$, the distribution of rescaled eigenvalues, λ_i, converges to the MP distribution with parameter γ.

Recalling the definitions of A_γ, B_γ given in (2.10), this implies that

$$\frac{1}{n}\sum_{i=1}^n \lambda_i \cdot (\mathbb{1}_{i\leq n/2} - \mathbb{1}_{i>n/2}) \to A_\gamma$$

and

$$\frac{1}{n}\sum_{i=1}^n \lambda_i^2 \to B_\gamma \,.$$

Together with the definition of $\tilde{\boldsymbol{w}}$, these imply that as $n, p \to \infty$ with $n/p \to \gamma \in (0,1)$,

$$\begin{aligned} n\sum_{i=1}^n \tilde{w}_i^2 &= \frac{n \cdot n}{\left\{n \cdot \left[\frac{1}{n}\sum_{i=1}^n \lambda_i \cdot (\mathbb{1}_{i\leq n/2} - \mathbb{1}_{i>n/2})\right]\right\}^2} \to \frac{n \cdot n}{(nA_\gamma)^2} = \frac{1}{A_\gamma^2}, \\ n\sum_{i=1}^n \tilde{w}_i^2\lambda_i^2 &= \frac{n \cdot n \cdot \left[\frac{1}{n}\sum_{i=1}^n \lambda_i^2\right]}{\left\{n \cdot \left[\frac{1}{n}\sum_{i=1}^n \lambda_i \cdot (\mathbb{1}_{i\leq n/2} - \mathbb{1}_{i>n/2})\right]\right\}^2} \to \frac{n \cdot n \cdot B_\gamma}{(nA_\gamma)^2} = \frac{B_\gamma}{A_\gamma^2}, \end{aligned}$$

which in turn implies

$$\sqrt{n} \cdot \frac{\sqrt{\mathrm{Var}(T_2)}}{\theta^2 + \sigma^2} \leq \sqrt{2 \cdot \max\left(n\sum_{i=1}^n \tilde{w}_i^2, n\sum_{i=1}^n \tilde{w}_i^2\lambda_i^2\right)} \to \sqrt{2} \cdot \max\left(\frac{1}{A_\gamma}, \frac{\sqrt{B_\gamma}}{A_\gamma}\right). \tag{D.1}$$

E Proof of Theorem 2

Proof. For this proof, we use Le Cam's method (see e.g. Yu (1997, Lemma 1)), which states that

$$\mathbb{P}_{\boldsymbol{Z}\sim P_0}\left[\psi(\boldsymbol{Z})=1\right]+\mathbb{P}_{\boldsymbol{Z}\sim P_1}\left[\psi(\boldsymbol{Z})=0\right]\geq 1-\left\|P_0-P_1\right\|_{\mathsf{TV}}\ ,$$

where $\|\cdot\|_{\mathsf{TV}}$ is the total variation norm:

$$\left\|P_0-P_1\right\|_{\mathsf{TV}}=\sup_{\mathcal{A}\subseteq\mathbb{R}^n}\left|\mathbb{P}_{\boldsymbol{Z}\sim P_0}\left(\boldsymbol{Z}\in\mathcal{A}\right)-\mathbb{P}_{\boldsymbol{Z}\sim P_1}\left(\boldsymbol{Z}\in\mathcal{A}\right)\right|\ ,$$

where the supremum is taken over Lebesgue-measurable sets.

We begin by constructing a related distribution Q_1:

$$\boldsymbol{W}=\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r+\sigma\cdot\boldsymbol{\varepsilon}\ , \tag{E.1}$$

where $\theta=\sigma=\frac{1}{\sqrt{2}}$, and where $r\sim\chi_p/\sqrt{p}$ is independent from $\boldsymbol{V},\boldsymbol{\varepsilon}$. We will bound

$$\left\|P_0-P_1\right\|_{\mathsf{TV}}\leq\left\|P_0-Q_1\right\|_{\mathsf{TV}}+\left\|P_1-Q_1\right\|_{\mathsf{TV}}\ .$$

First, we use the fact that r concentrates tightly near 1 for the following bound:

$$\begin{aligned}\mathbb{E}\left(\left\|\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r-\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\right\|_2\right)&=\mathbb{E}\left[\mathbb{E}\left(\left\|\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r-\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\right\|_2\ \middle|\ r,\boldsymbol{V}\right)\right]\\&=\mathbb{E}\left(\left\|\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\right\|_2\cdot|r-1|\right)\\&\leq\theta\cdot\sqrt{\mathbb{E}\left(\boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^\top\boldsymbol{a}\right)}\cdot\sqrt{\mathbb{E}\left[(r-1)^2\right]}\\&=\theta\cdot\sqrt{\mathbb{E}\left(\boldsymbol{a}^\top\boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^\top\boldsymbol{a}\right)}\cdot\frac{1}{\sqrt{p}}\sqrt{\mathbb{E}\left(\chi_p^2\right)-2\sqrt{p}\cdot\mathbb{E}\left(\chi_p\right)+p}\end{aligned}$$

Using the fact that $\mathbb{E}\left(\chi_p^2\right)=p$ and $\mathbb{E}\left(\chi_p\right)\geq\sqrt{p}-\frac{1}{4\sqrt{p}}$, and that $\mathbb{E}\left((\boldsymbol{V}_i^\top\boldsymbol{a})^2\right)=\frac{1}{p}$ for each $i=1,\ldots,n$,

$$\begin{aligned}\mathbb{E}\left(\left\|\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r-\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\right\|_2\right)&\leq\theta\cdot\sqrt{\sum_i\frac{D_{ii}^2}{p}}\cdot\frac{1}{\sqrt{2p}}\\&=\frac{\theta\sqrt{n}}{\sqrt{2p}}\ ,\end{aligned}$$

since $\frac{1}{p}\sum_i D_{ii}^2=n$. Next, for any measurable set $\mathcal{A}\subseteq\mathbb{R}^n$, we have

$$\begin{aligned}&\left|\mathbb{P}_{\boldsymbol{W}\sim Q_1}\left(\boldsymbol{W}\in\mathcal{A}\right)-\mathbb{P}_{\boldsymbol{Z}\sim P_1}\left(\boldsymbol{Z}\in\mathcal{A}\right)\right|\\&=\left|\mathbb{E}\left[\mathbb{P}\left(\boldsymbol{W}\in\mathcal{A}\mid r,\boldsymbol{V}\right)-\mathbb{P}\left(\boldsymbol{Z}\in\mathcal{A}\mid r,\boldsymbol{V}\right)\right]\right|\\&\leq\mathbb{E}\left[\left|\mathbb{P}\left(\boldsymbol{W}\in\mathcal{A}\mid r,\boldsymbol{V}\right)-\mathbb{P}\left(\boldsymbol{Z}\in\mathcal{A}\mid r,\boldsymbol{V}\right)\right|\right]\\&=\mathbb{E}\left[\left|\mathbb{P}\left(\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r+\sigma\cdot\boldsymbol{\varepsilon}\in\mathcal{A}\ \middle|\ r,\boldsymbol{V}\right)-\mathbb{P}\left(\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}+\sigma\cdot\boldsymbol{\varepsilon}\in\mathcal{A}\ \middle|\ r,\boldsymbol{V}\right)\right|\right]\\&\leq\mathbb{E}\left\{\mathbb{E}\left[\left\|N\left(\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r,\sigma^2\mathbf{I}_n\right)-N\left(\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a},\sigma^2\mathbf{I}_n\right)\right\|_{\mathsf{TV}}\middle|\,r,\boldsymbol{V}\right]\right\}\end{aligned}$$

Using the fact that $\left\|N(\boldsymbol{\mu},\sigma^2\mathbf{I}_n)-N(\boldsymbol{\mu}',\sigma^2\mathbf{I}_n)\right\|_{\mathsf{TV}}\leq\frac{\|\boldsymbol{\mu}-\boldsymbol{\mu}'\|_2}{\sqrt{2\pi\sigma^2}}$ for any fixed $\boldsymbol{\mu},\boldsymbol{\mu}',\sigma^2$,

$$\begin{aligned}&\leq\mathbb{E}\left[\mathbb{E}\left(\frac{\left\|\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\cdot r-\theta\cdot\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{a}\right\|_2}{\sqrt{2\pi\sigma^2}}\ \middle|\ r,\boldsymbol{V}\right)\right]\\&\leq\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\frac{\theta\sqrt{n}}{\sqrt{2p}}\ ,\end{aligned}$$

where the last step uses our calculations above. Since this is true for any $\mathcal{A}\subset\mathbb{R}^n$, and since $\theta=\sigma=\frac{1}{\sqrt{2}}$ by assumption under the distribution P_1, we have

$$\left\|P_1-Q_1\right\|_{\mathsf{TV}}\leq\sqrt{\frac{n/p}{4\pi}}\ .$$

Next, we bound $\|P_0 - Q_1\|_{\mathsf{TV}}$. By Pinsker's inequality,

$$\|P_0 - Q_1\|_{\mathsf{TV}} \le \sqrt{\frac{1}{2}\mathsf{KL}\left(Q_1\|P_0\right)}\,,$$

where $\mathsf{KL}\left(\cdot\|\cdot\right)$ is the Kullback-Leibler divergence. Note that the distributions P_0 and Q_1 can be reformulated as

$$P_0 : Z_i \overset{\perp\!\!\!\perp}{\sim} N(0,1)$$

and

$$Q_1 : Z_i \overset{\perp\!\!\!\perp}{\sim} N\left(0, \frac{\lambda_i + 1}{2}\right)\,.$$

Writing $p_0(\cdot)$ and $q_1(\cdot)$ to be the densities of the distributions P_0 and Q_1, respectively, we have

$$\begin{aligned}
\mathsf{KL}\left(Q_1\|P_0\right) &= \mathbb{E}_{\boldsymbol{Z}\sim Q_1}\left\{\log\left[\frac{q_1(\boldsymbol{Z})}{p_0(\boldsymbol{Z})}\right]\right\} \\
&= \mathbb{E}_{\boldsymbol{Z}\sim Q_1}\left[\log\left(\frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\left(\frac{\lambda_i+1}{2}\right)}} e^{-\frac{Z_i^2}{\lambda_i+1}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{Z_i^2}{2}}}\right)\right] \\
&= \mathbb{E}_{\boldsymbol{Z}\sim Q_1}\left\{-\frac{1}{2}\sum_i\left[\log\left(\frac{\lambda_i+1}{2}\right) + Z_i^2\cdot\left(\frac{2}{\lambda_i+1} - 1\right)\right]\right\}
\end{aligned}$$

Since $\mathbb{E}_{\boldsymbol{Z}\sim Q_1}\left(Z_i^2\right) = \frac{\lambda_i+1}{2}$,

$$= -\frac{1}{2}\sum_i\left[\log\left(\frac{\lambda_i+1}{2}\right) + \left(1 - \frac{\lambda_i+1}{2}\right)\right]$$

Using the fact that $\log(x) \ge (x-1) - 2(x-1)^2$ for all $x \ge \frac{1}{2}$,

$$\begin{aligned}
&\le -\frac{1}{2}\sum_i\left[\frac{\lambda_i - 1}{2} - 2\left(\frac{\lambda_i-1}{2}\right)^2 + \left(1 - \frac{\lambda_i+1}{2}\right)\right] \\
&= \frac{1}{4}\sum_i(\lambda_i - 1)^2\,.
\end{aligned}$$

Combining everything, we have

$$\|P_0 - Q_1\|_{\mathsf{TV}} \le \sqrt{\frac{1}{2}\mathsf{KL}\left(Q_1\|P_0\right)} \le \sqrt{\frac{1}{2}\cdot\frac{1}{4}\sum_i(\lambda_i-1)^2}\,,$$

and so

$$\|P_0 - P_1\|_{\mathsf{TV}} \le \sqrt{\frac{1}{8}\sum_i(\lambda_i-1)^2} + \sqrt{\frac{n/p}{4\pi}}\,.$$

□

F CVX code for computing the weight vector

The following snippet of code was used with MATLAB Version 8.1 (R2013a) and CVX Version 2.1, Build 1085 on a 64-bit Linux OS. The eigenvalues λ_i are represented by the column vector `lambda`, `t` corresponds to $2\,\mathrm{val}(\mathcal{P}_1)$, and the resulting vector `w` corresponds to $\boldsymbol{w}^*$.

```
cvx_begin
variable t
variable w(n)
minimize t
subject to
    sum(w) == 0;
    sum(w .* lambda) == 1;
    norm([w; (t/2-1)/2])  <= (t/2+1)/2;
    norm([w .* lambda; (t/2-1)/2]) <= (t/2+1)/2;
cvx_end
```

G Bayesian model

The Bayesian model is given explicitly as follows (M, $\boldsymbol{Z}$, σ^2, and $\boldsymbol{\varepsilon}$ are all independent of one another):

$$\begin{aligned} M &\sim \text{Exponential}(\lambda), \\ \boldsymbol{Z} &\sim N(0, I_p), \\ \boldsymbol{\beta} &= \sqrt{M}\boldsymbol{Z}, \\ \frac{1}{\sigma^2} &\sim \text{Gamma}(A, B), \\ \boldsymbol{\varepsilon} &\sim N(0, \sigma^2 I_p), \\ \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \end{aligned} \tag{G.1}$$

where the values for the parameters used were $\lambda = \frac{50{,}000}{p}$ (so $\theta^2 \approx \text{Exponential}(1/2000)$), $A = 14$, $B = \frac{1}{20{,}000}$, and we have used the shape/scale parameterization of the Gamma distribution, as opposed to the shape/rate parameterization. Figure 1 shows the resulting priors for θ^2, σ^2, and $\rho = \frac{\theta^2}{\theta^2+\sigma^2}$.

Note also that, although not shown, the posteriors achieved under this setup were all unimodal, so that the equal-tailed credible intervals were very close to the minimum-length credible intervals. We used equal-tailed credible intervals to give fair comparison with the EigenPrism CIs, which are also equal-tailed. The interval widths plotted all have nominal coverage of 80%. BCI endpoints were estimated by empirical quantiles of posterior draws from a Gibbs sampler, and thus we were able to much more accurately estimate the 10th and 90th percentiles than, say, the 2.5th and 97.5th percentiles.

H Construction of correlated-column covariance matrices

Dense 10% Correlations used a covariance matrix with ones on the diagonal and 0.1's as all the other entries. The Sparse $100 \cdot P$% Correlations used alternating P and $-P$ as off-diagonal entries in a correlation matrix, then projected that matrix into the positive semidefinite cone and reset the diagonal entries to 1. The resulting matrix has approximately 1/4 of its entries equal to P, 1/2 of its entries equal to 0, and 1/4 of its entries equal to $-2\times10^{-4}\cdot(1-P)$.

I Processing of NFBC1966 dataset

Genotype features from the original data set were removed if they met any of the following conditions:

- Not a SNP (some were, e.g., copy number variations)
- Greater than 5% of values were missing
- All nonmissing values belonged to the same nucleotide
- SNP location could not be aligned to the genome
- A χ^2 test rejected Hardy-Weinberg equilibrium at the 0.01% level
- On chromosome 23 (sex chromosome)

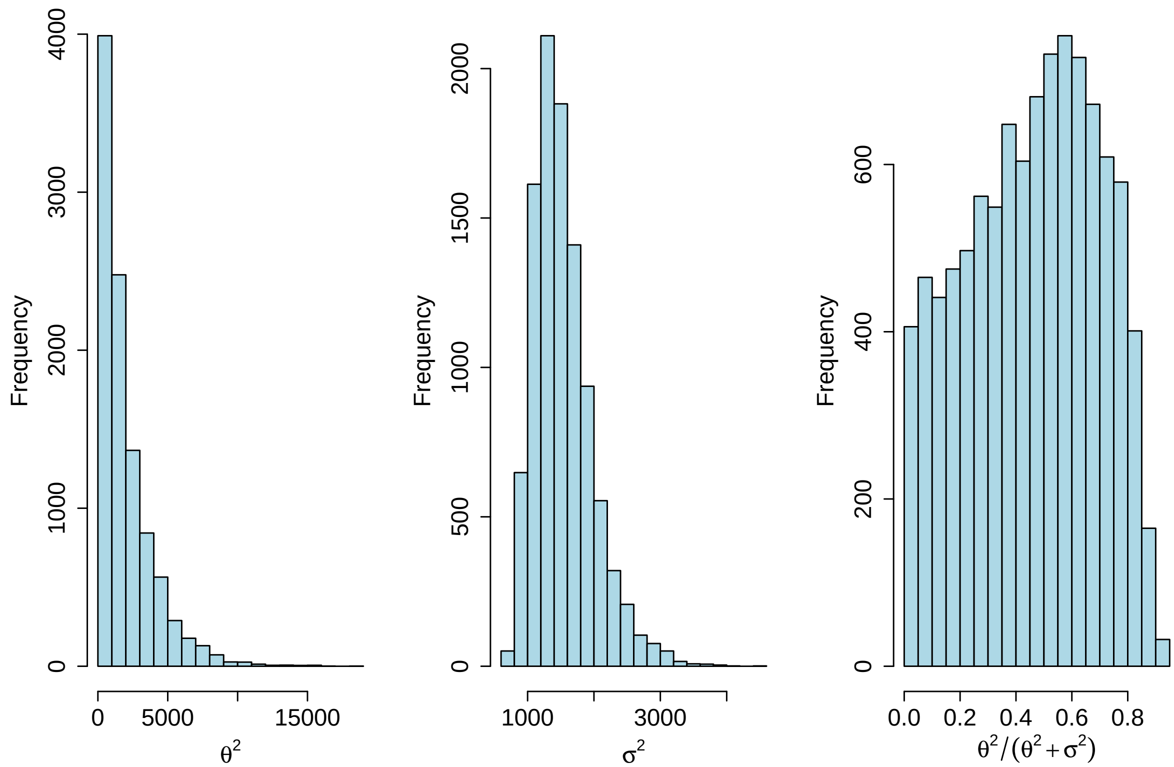

Figure 1: Priors from model (G.1).

The remaining missing values were assumed to take the major allele value (thus were coded as 0's in the pre-centered design matrix).

For each trait, further processing was performed on the subjects. Triglycerides, BMI, insulin, and glucose were all log-transformed. C-reactive protein was also log-transformed after adding 0.002 mg/l (half the detection limit) to 0 values. Subjects were excluded from the triglycerides, HDL and LDL cholesterol, glucose, and insulin analyses if they were on diabetic medication or had not fasted before blood collection (or if either value was missing). Further subjects were excluded from the triglycerides, HDL and LDL cholesterol analyses if they were pregnant or if their respective phenotype measurement was more than three standard deviations from the mean, after correcting for sex, oral contraceptive use, and pregnancy. Subjects whose weight was not directly measured were excluded from BMI analysis. Of course any missing values in each phenotype were also excluded.

References

Billingsley, P. (1995) *Probability and measure.* John Wiley & Sons, third edn.

Lehman, E. L. and Romano, J. P. (2005) *Testing statistical hypotheses.* Springer, third edn.

Yu, B. (1997) Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, 423–435. Springer.