
Supplementary Material for: Maximum Likelihood for Variance Estimation in High-Dimensional Linear Models

Lee H. Dicker
Rutgers University

Murat A. Erdogdu
Stanford University

S1 DETAILED DESCRIPTION OF FIGURE 1

The plots in Figure 1 from the main text were generated using various estimators of σ_0^2 , each of which was computed for 500 independent datasets (\mathbf{y}, X) . The datasets were generated according to the linear model (1)–(2) (equation references refer to the main text), with $n = 500$, $p = 1000$, $\sigma_0^2 = 1$, and $\eta_0^2 = 4$. We considered settings where β had various sparsity levels, indicated by a parameter $\alpha \in [0, 100]$. In the $\alpha\%$ -sparse setting, β has its first $p(1 - \alpha/100)$ entries all equal to each other and greater than zero; the rest of the entries are set to 0. Detailed results on the performance of the various estimators, corresponding to the plots in Figure 1 are reported in Table S1.

S2 DETAILED DESCRIPTION OF FIGURE 2

The asymptotic variances of various estimators for σ_0^2 and r_0^2 are plotted in Figure 2 of the main text. In the main text, we describe in detail how the asymptotic variances of the MLEs $\hat{\sigma}^2$ and \hat{r}^2 are derived. Here we give a brief derivation of the asymptotic variance of the MM and OLS estimators for σ_0^2 and r_0^2 .

Dicker (2014) proposed the MM estimators $\hat{\sigma}_{\text{MM}}^2$ and \hat{r}_{MM}^2 , and showed that they are asymptotically normal under essentially the same conditions found in Theorem 2 of the main text. Moreover, their asymptotic variances are given by

$$v_{\text{MM}}(\eta_0^2, \rho) = \frac{n}{2\sigma_0^4} \lim_{p/n \rightarrow \rho} \text{Var}(\hat{\sigma}_{\text{MM}}^2) = \rho(1 + \eta_0^2)^2 + 1 + \eta_0^4,$$
$$w_{\text{MM}}(\eta_0^2, \rho) = \frac{1}{2} \lim_{p/n \rightarrow \rho} \text{Var}(\hat{r}_{\text{MM}}^2) = 1 + \rho - \frac{1}{(1 + \eta_0^2)^2},$$

respectively. It is straightforward to check that the asymptotic variance of $\hat{\sigma}_{\text{OLS}}^2$ is

$$v_{\text{OLS}}(\eta_0^2, \rho) = \lim_{p/n \rightarrow \rho} \frac{n}{2\sigma_0^4} \text{Var}(\hat{\sigma}_{\text{OLS}}^2) = \frac{1}{1 - \rho},$$

for $\rho \in (0, 1)$.

Table S1: Numerical results corresponding to the plots in Figure 1 from the main text. Actual value of the residual variance is $\sigma_0^2 = 1$. “Mean” and “Std. Err.” refer to the mean and standard error of the corresponding estimator, computed from 500 independent datasets.

Method	10% Sparsity		40% Sparsity		90% Sparsity		99.8% Sparsity	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
MLE	1.0023	0.2890	1.0051	0.3008	1.0057	0.2978	1.0128	0.2911
MM	1.0136	0.4932	1.0270	0.4915	1.0172	0.5212	1.0242	0.5510
EigenPrism	1.0150	0.4896	1.0268	0.4862	1.0138	0.5168	1.0247	0.5468
AMP	0.9945	0.4165	0.9980	0.4412	1.0144	0.3022	1.0011	0.1143
Scaled-Lasso	4.9634	0.3075	4.9488	0.3100	4.8734	0.2956	1.0500	0.0679
RCV-Lasso	4.8668	0.3513	4.8293	0.3396	3.9171	0.4891	0.9973	0.0660

S3 PREPROCESSING STEPS FOR THE SNP AND GENE EXPRESSION DATA DESCRIBED IN SECTION 6

Preprocessing steps of particular significance included: (i) For each of the 100 genes included in the data analysis, only SNPs within 1 megabase pairs (Mbp) of the gene midpoint were included in the analysis; (ii) we only utilized SNPs with minor allele frequency greater than 5% and no more than 10% missingness; (iii) we imputed missing minor allele counts using the marginal mean from the Han Chinese population; and (iv) highly correlated SNPs were removed from the analysis, by setting a correlation threshold of $2\sqrt{\log(n)/n}$ and discarding SNPs that exceeded this thresholded (one SNP chosen at random from each highly-correlated pair). A correlation screening procedure similar to that described in (iv) was utilized in (Dicker, 2014) and is loosely justified by results on random correlation matrices in (Jiang, 2004).

References

- Dicker, L. H. Variance estimation in high-dimensional linear models. *Biometrika*, 101:269–284, 2014.
- Jiang, T. The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.*, 14: 865–880, 2004.