Topics in Econometrics - Assignment 1

Yixin Sun

April 18, 2021

Question 1b and 1c

We see here that the average $\hat{\beta}_1$ is about the same as the true β_1. The variance of the estimate decreases from $p = 1$ to $p = 5$, and then increases after that. This makes sense as the $p = 1$ regression suffers from omitted variable bias, whereas $p > 5$ is overfitting with variables that are not actually related to the outcome variable or explanatory variable of interest. Even if you included the regressors that are independent of the one you care about, these independent regressors interfere.

Table 1: Average and Variance of $\hat{\beta}_1$

p	mean	variance
1	1.003	0.020
5	1.000	0.009
10	1.000	0.010
50	1.000	0.018
85	1.011	0.065
90	1.010	0.102

Question 1d

Table 2: Frisch-Waugh-Lovell Results

p	x1_2	x_eigen
5	0.041	74.175
10	0.091	55.700
50	0.491	10.487
85	0.840	0.965
90	0.892	0.495

This table shows that the value of the lowest eigenvalue decreases as the number of regressor increase, while $1/N \sum_i \hat{X}_{i1}(\tilde{p})^2$ increases as we add more regressors. $1/N \sum_i \hat{X}_{i1}(\tilde{p})^2$ here is the variance of $\hat{X}_{i1}$.

We're thinking about what this $p \tilde{-} 1$ unrelated regressors are doing to the estimation of β_1, the fitted value of th first regressor we care about. In the DGP, they are all independent regressors, so $\hat{X}_1$ should be zero. But as you add more independent regressors, the estimate gets worse, and hence the variance increases. The lowest eigenvalue addresses the invertibility of the regressors - the lowest eigenvealue is closer to the zero and there is more noise in the inverse. This is the same thing as variance of $\hat{X}_{i1}$ getting bigger, which we see in the second column of table 2.

Question 1e

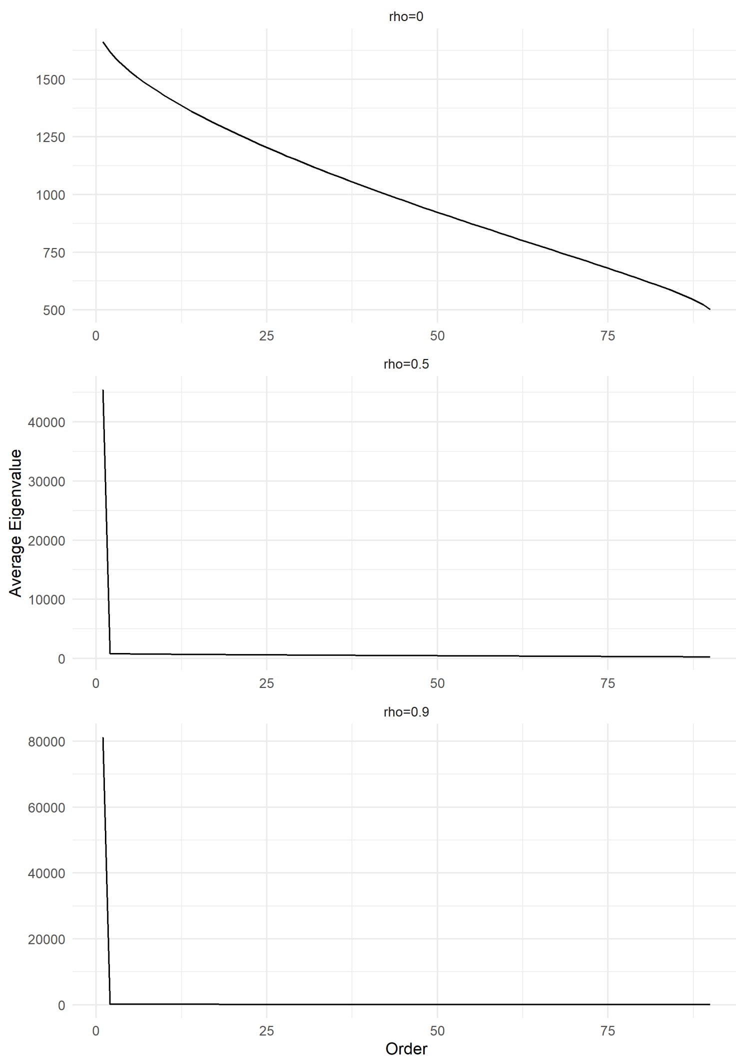
This is even worse than b, c, d. As ρ and N increase, the variance of $\hat{\beta}_1$ increases. The issue here is that the extra covariates are not actually correlated to Y_1. However, the correlation between the extra covariates and X_1 is causing the extra covariates to essentially steal variation that should be attributed to X_1, leading to more noise on the parameter of interest, $\hat{\beta}_1$.

Table 3: Dependence Structure Between Covariates

mean β_1	var β_1	eig	N	ρ
1.010	0.115	0.421	100	0.0
0.986	0.227	0.207	100	0.5
1.032	1.097	0.043	100	0.9
1.002	0.010	23.627	200	0.0
0.999	0.017	11.887	200	0.5
0.991	0.088	2.386	200	0.9
1.001	0.002	172.734	500	0.0
1.000	0.005	86.569	500	0.5
1.007	0.025	17.328	500	0.9
1.001	0.001	502.059	1000	0.0
1.001	0.002	252.314	1000	0.5
1.009	0.011	50.382	1000	0.9

Question 1f

When $\rho = 0$, there is no relationship between the covariates, and the eigenvalues decrease smoothly, since each additional coviariate adds unexplained variation to the set. However, when we do have correlation between covariates, we see that the eigenvalue starts off very high, and then immediately drops off. This shows the problem of collinearity: even if we have all these extra covariates, the strong correlation between them means that many do not hold any additional information.


rho=0
1500
1250
1000
750
500
0
25
50
75
rho=0.5
Average Eigenvalue
40000
30000
20000
10000
0
0
25
50
75
rho=0.9
80000
60000
40000
20000
0
0
25
50
75
Order

Question 1g

We see a similar relationship between ρ, N and the variance of $\hat{\beta}_1$, but we see what happens with two different rates of convergence of $\frac{p}{N}$. When $p = 0.9N$, $\lim_{N\to\infty} \frac{p}{N} = 0.9$. Thus we see in Table 4 that when N increases, we do not see a huge impact on the lowest eigenvalue. When $p = 20 \times log(N)$, $\lim_{N\to\infty} \frac{p}{N} = 0$. We see that in Table 5, the eigenvalues increase dramatically as N increases, and this translates to a lower variance on $\hat{\beta}_1$.

Table 4: $p = 0.9N$

mean β_1	var β_1	eig	N	ρ
0.993	0.107	0.424	100	0.0
0.994	0.226	0.211	100	0.5
0.983	1.134	0.042	100	0.9
0.998	0.050	0.691	200	0.0
0.990	0.106	0.351	200	0.5
0.977	0.570	0.071	200	0.9
1.000	0.020	1.534	500	0.0
0.994	0.042	0.758	500	0.5
1.002	0.193	0.152	500	0.9
1.001	0.010	2.891	1000	0.0
1.005	0.019	1.450	1000	0.5
0.996	0.106	0.289	1000	0.9

Table 5: $p = 20 \times log(N)$

mean β_1	var β_1	eig	N	ρ
1.014	0.138	0.286	100	0.0
1.039	0.303	0.142	100	0.5
0.967	1.428	0.028	100	0.9
0.999	0.011	16.786	200	0.0
1.000	0.022	8.441	200	0.5
0.995	0.098	1.687	200	0.9
0.999	0.003	131.108	500	0.0
1.000	0.005	65.814	500	0.5
0.998	0.024	13.162	500	0.9
0.999	0.001	404.958	1000	0.0
1.001	0.002	202.733	1000	0.5
0.998	0.011	40.541	1000	0.9