Big Data Asset Pricing

Exercise 4: Hig-Dimensional Return Predictions

Seyyed Morteza Aghajanzadeh

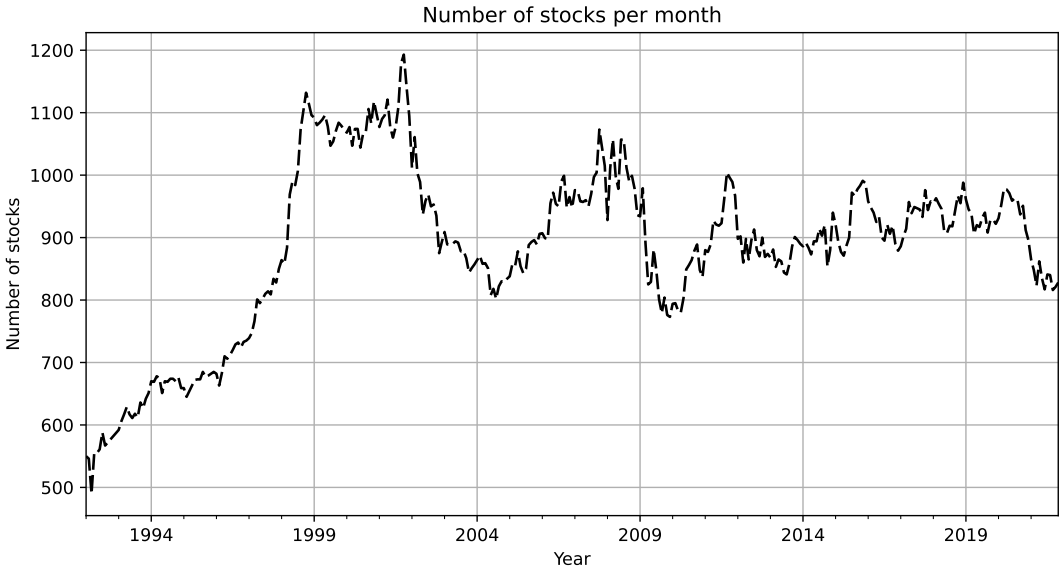
Department of Finance

Stockholm School of Economics

Statement: I certify with my signature that I have solved the exercise according to the Code of Professional Conduct and Ethics. For example, I have not plagiarized others, but, instead, solved the exercise myself (possibly with allowed collaboration with other students), and I have referenced my sources appropriately.

2024-01

1



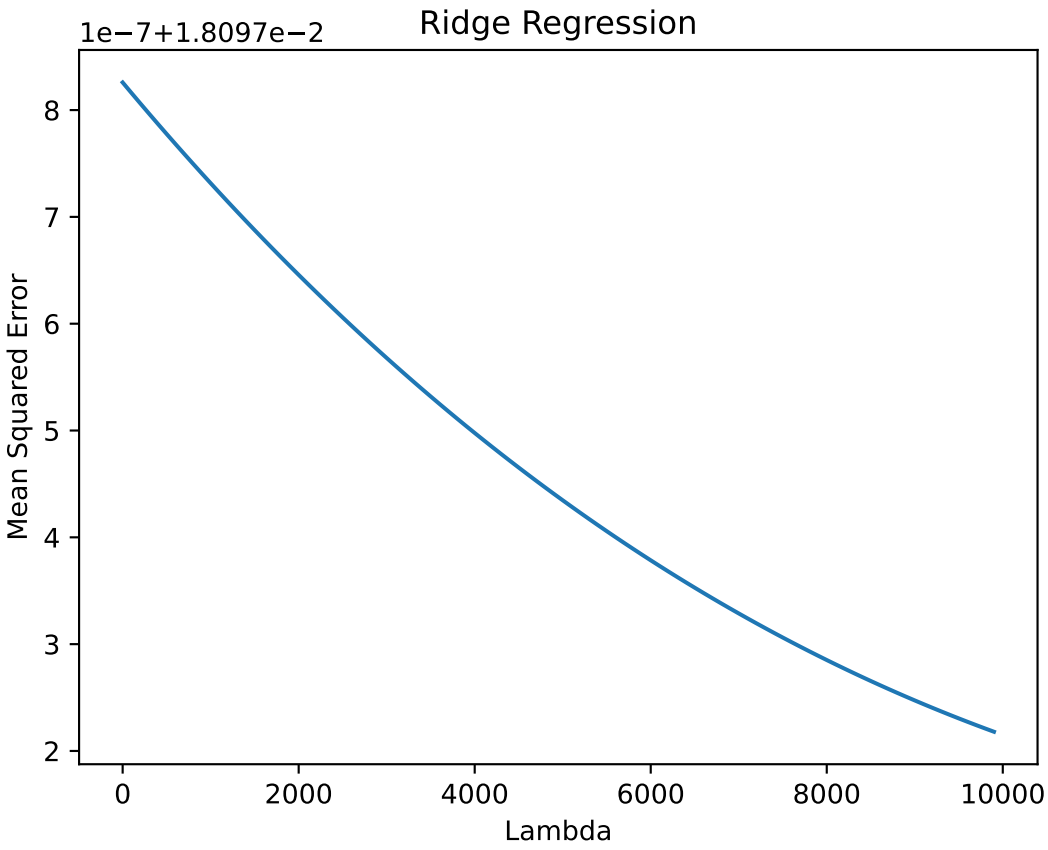
2

2.a

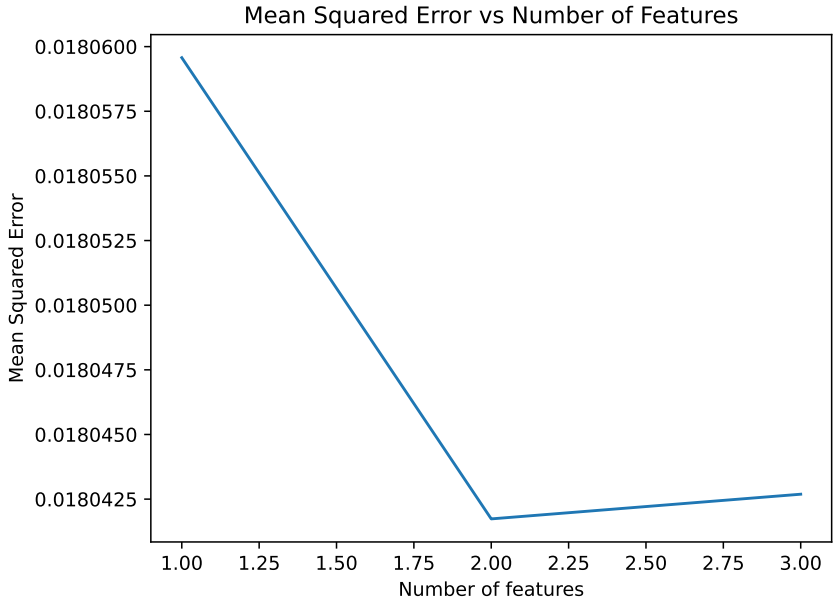
Table 1: The table shows the results of the Fama-Macbeth regression.

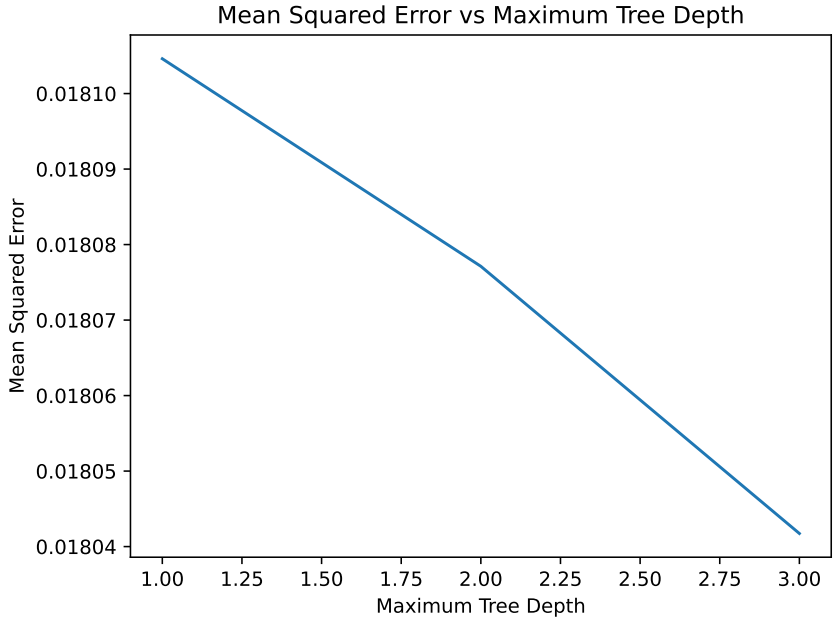
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
be_me	0.0316	0.0263	1.2032	0.2289	-0.0199	0.0831
ret_12_1	-0.0239	0.0211	-1.1340	0.2568	-0.0652	0.0174
market_equity	-0.0316	0.0276	-1.1424	0.2533	-0.0858	0.0226

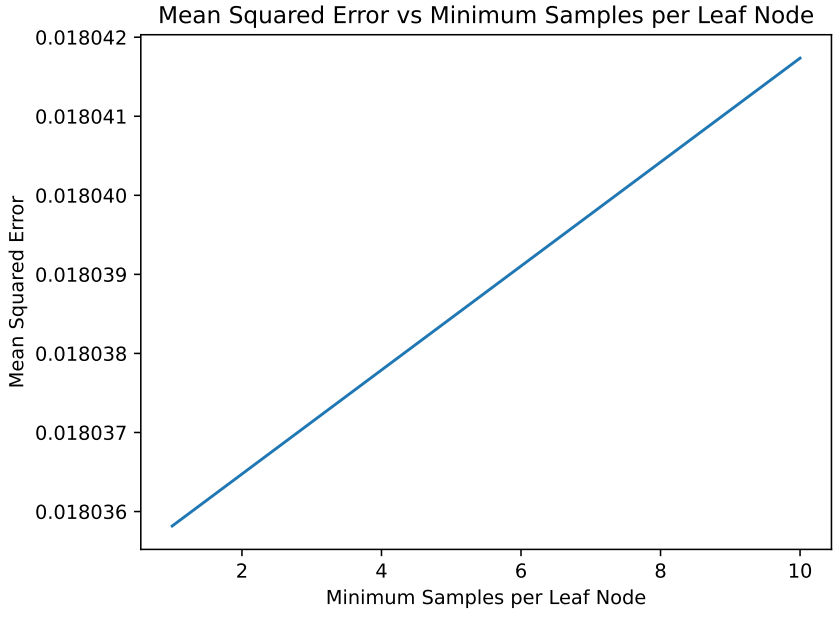
2.b


Ridge Regression
1e−7+1.8097e−2
Mean Squared Error
8
7
6
5
4
3
2
0
2000
4000
6000
8000
10000
Lambda

2.c


Mean Squared Error vs Number of Features
Mean Squared Error
0.0180600
0.0180575
0.0180550
0.0180525
0.0180500
0.0180475
0.0180450
0.0180425
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Number of features


Mean Squared Error vs Maximum Tree Depth
Mean Squared Error
0.01810
0.01809
0.01808
0.01807
0.01806
0.01805
0.01804
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Maximum Tree Depth

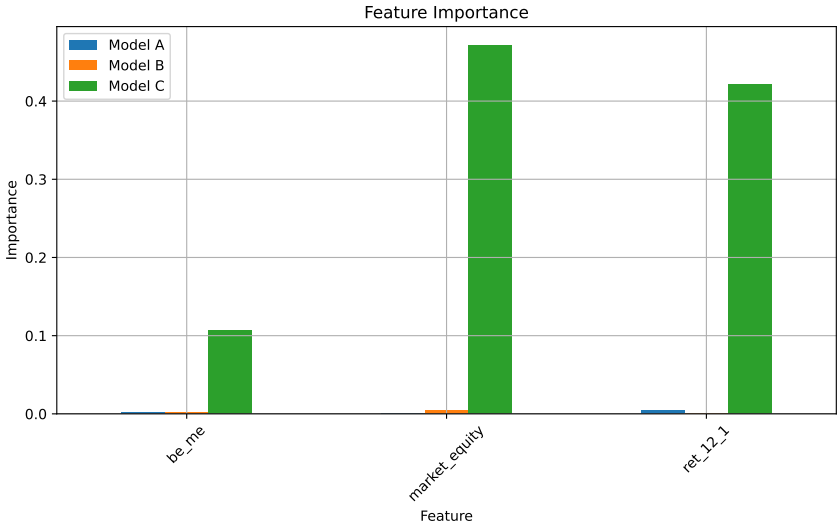

Mean Squared Error vs Minimum Samples per Leaf Node
Mean Squared Error
0.018042
0.018041
0.018040
0.018039
0.018038
0.018037
0.018036
2
4
6
8
10
Minimum Samples per Leaf Node

3

Table 2

Model	In-sample R2
Model A	0.0012
Model B	0.0012
Model C	0.0050

4



5

5.a

Table 3

Model	Out-sample R2
Model A	0.4869
Model B	-0.0007
Model C	0.0002

5.b

Table 4

Panel A: Pooled OLS

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	0.008	1.621	0.008	1.472	0.149	0.139
2	0.008	1.551	0.008	1.447	0.142	0.134
3	0.006	1.160	0.006	1.146	0.106	0.105
4	0.009	1.798	0.009	1.767	0.165	0.162
5	0.003	0.825	0.003	0.938	0.076	0.082
LS	-0.005	-1.515	-0.004	-1.200	-0.139	-0.118

Panel B: Ridge Regression

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	-0.481	-0.787	-0.526	-0.816	-0.072	-0.079
2	0.061	0.802	0.075	0.885	0.074	0.092
3	-0.010	-0.168	-0.012	-0.226	-0.015	-0.019
4	0.003	0.071	0.003	0.080	0.007	0.006
5	0.060	1.770	0.055	1.848	0.162	0.149
LS	0.540	0.884	0.580	0.900	0.081	0.087

Panel C: Random Forest

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	-0.068	-0.380	-0.074	-0.413	-0.035	-0.038
2	-0.019	-1.221	-0.020	-1.214	-0.112	-0.120
3	-0.059	-0.745	-0.032	-0.510	-0.068	-0.038
4	0.008	1.676	0.008	1.605	0.154	0.147
5	0.013	2.508	0.012	2.334	0.230	0.222
LS	0.080	0.452	0.086	0.485	0.041	0.045

6

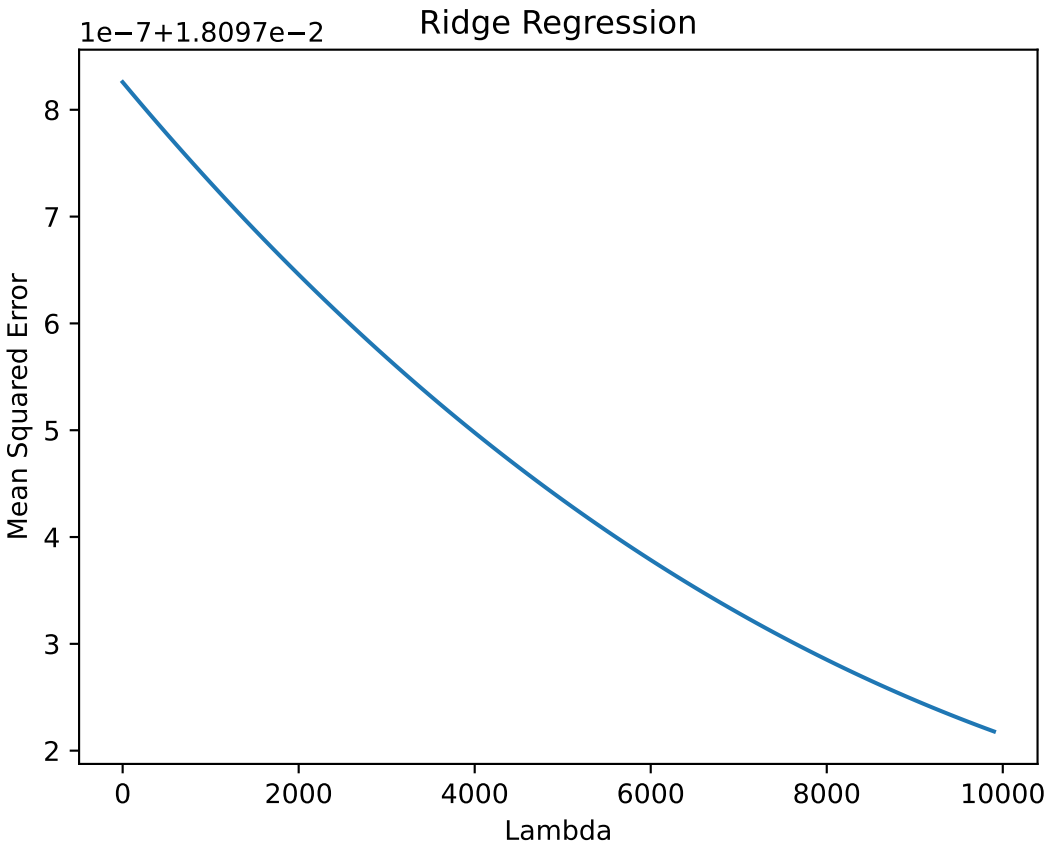
6.2

6.2.a

Table 5: The table shows the results of the Fama-Macbeth regression.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
be_me	0.0041	0.0128	0.3203	0.7487	-0.0210	0.0292
ret_12_1	0.0042	0.0048	0.8691	0.3848	-0.0053	0.0137
market_equity	0.1608	0.1310	1.2274	0.2197	-0.0959	0.4175
ret_1_0	0.0012	0.0048	0.2500	0.8026	-0.0081	0.0105
rvol_252d	0.0001	0.0109	0.0116	0.9907	-0.0213	0.0215
beta_252d	-0.0947	0.0857	-1.1057	0.2688	-0.2627	0.0732
qmj_safety	0.0068	0.0035	1.9140	0.0556	-0.0002	0.0137
rmax1_21d	-0.0931	0.1244	-0.7484	0.4542	-0.3370	0.1507
chcsho_12m	-0.0089	0.0621	-0.1438	0.8856	-0.1307	0.1129
ni_me	-0.0562	0.0472	-1.1901	0.2340	-0.1486	0.0363
eq_dur	43.781	43.188	1.0137	0.3107	-40.865	128.43
ret_60_12	-0.0098	0.0042	-2.3206	0.0203	-0.0180	-0.0015
ope_be	-0.0192	0.0309	-0.6206	0.5349	-0.0798	0.0414
gp_at	0.0024	0.0029	0.8322	0.4053	-0.0033	0.0081
ebit_sale	-0.6719	0.6837	-0.9827	0.3258	-2.0119	0.6682
at_gr1	-1.0733	0.7334	-1.4635	0.1433	-2.5107	0.3641
sale_gr1	0.2397	0.2739	0.8753	0.3814	-0.2971	0.7765
at_be	0.0777	0.0427	1.8208	0.0686	-0.0059	0.1613
cash_at	-0.0002	0.0030	-0.0562	0.9552	-0.0060	0.0057
age	9.9438	7.5646	1.3145	0.1887	-4.8827	24.770
z_score	0.0423	0.0594	0.7122	0.4764	-0.0741	0.1588

6.2.b

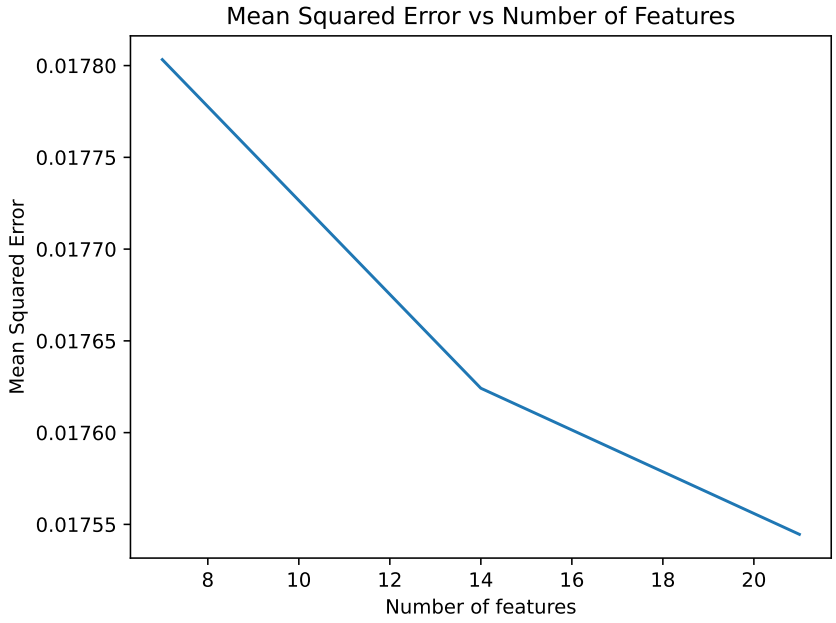


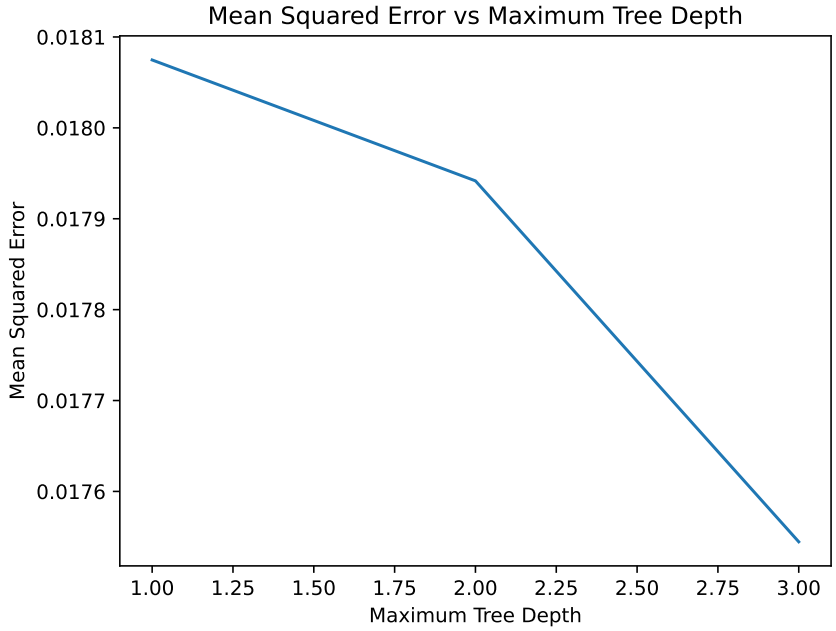
6.2.c

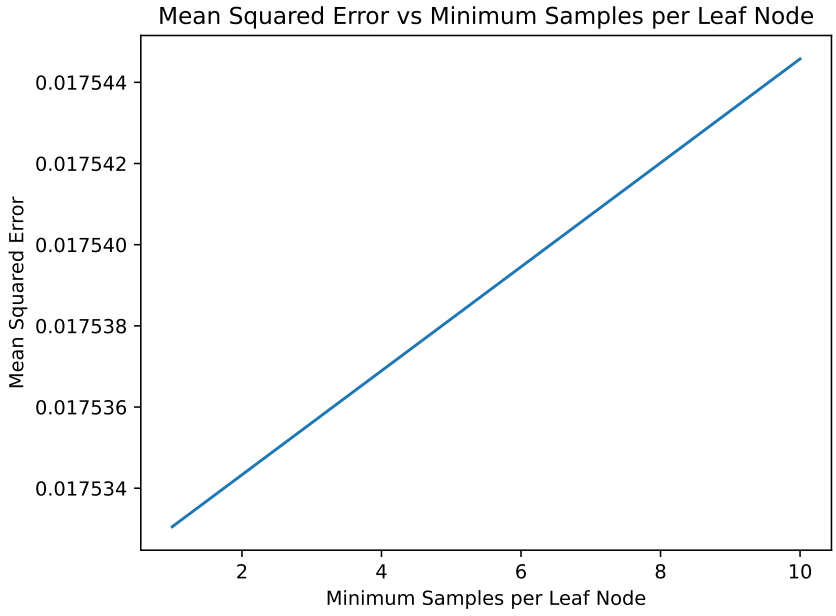
6.3

Table 6

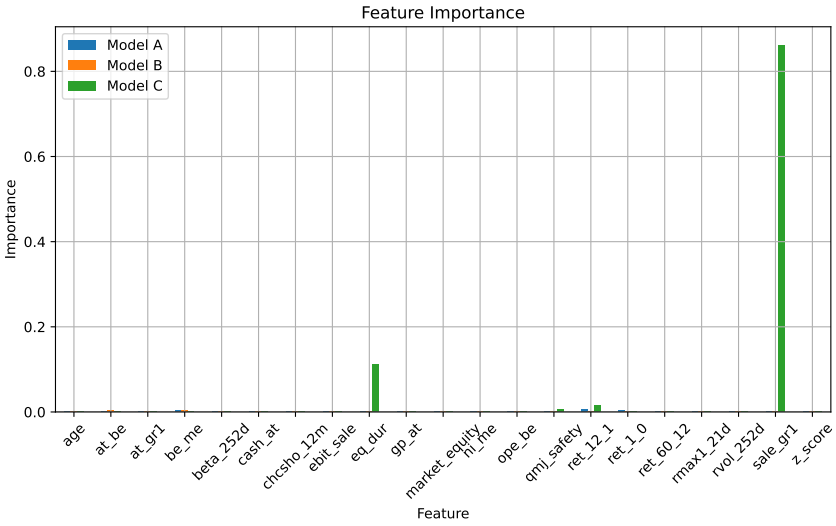
Model	In-sample R2
Model A	0.0023
Model B	0.0023
Model C	0.0324


Mean Squared Error vs Number of Features
Mean Squared Error
0.01780
0.01775
0.01770
0.01765
0.01760
0.01755
8
10
12
14
16
18
20
Number of features


Mean Squared Error vs Maximum Tree Depth
Mean Squared Error
0.0181
0.0180
0.0179
0.0178
0.0177
0.0176
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Maximum Tree Depth


Mean Squared Error vs Minimum Samples per Leaf Node
Mean Squared Error
0.017544
0.017542
0.017540
0.017538
0.017536
0.017534
2
4
6
8
10
Minimum Samples per Leaf Node

7



8

8.1

Table 7

Model	Out-sample R2
Model A	0.4870
Model B	0.0002
Model C	-0.0298

8.2

Table 8

Panel A: Pooled OLS

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	-0.063	-1.113	-0.069	-1.131	-0.102	-0.112
2	-0.005	-0.233	0.008	0.329	-0.021	0.034
3	0.091	1.510	0.084	1.545	0.138	0.129
4	0.058	1.653	0.057	1.888	0.152	0.149
5	-0.055	-1.165	-0.059	-1.248	-0.107	-0.114
LS	0.008	0.111	0.011	0.127	0.010	0.013

Panel B: Ridge Regression

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	0.036	1.191	0.034	1.006	0.109	0.105
2	-0.004	-0.410	-0.003	-0.343	-0.038	-0.030
3	-0.451	-1.044	-0.659	-1.299	-0.096	-0.152
4	0.018	1.448	0.017	1.126	0.133	0.120
5	0.017	3.061	0.017	2.804	0.281	0.271
LS	-0.018	-0.619	-0.018	-0.536	-0.057	-0.055

Panel C: Random Forest

Portfolio	$r_i - r_f$	t_{stat}	α_{CAPM}	t_α	Sharpe Ratio	IR
1	-0.061	-0.995	-0.032	-286097475885977.312	-0.091	-3264846760324056.000
2	0.402	0.785	0.142	2248224154820543.750	0.072	7224655780888126.000
3	0.115	1.281	-15.369	-1691176820516556.250	0.117	-10169850206690330.000
4	0.061	7.593	0.151	559173050015646.000	0.696	6160577553020680.000
5	-0.073	-34.301	0.018	275850033277120.750	-3.144	3695901578164656.500
LS	-0.034	-17.384	0.050	1224457471030904.000	-1.594	5112797549406387.000

Appendix

Here you can find the python code that I used to solve the exercise. Link to the GitHub repository.