**Objective:**

We’re trying to predict the log returns of a cross-section of a stock from a set of 6 independent parameters. The basic assumption is that the returns can be explained from past data in the time-series.

When I ran the regressions, I found that some of the variables were not statistically significant. Refer to the below table for coefficients:



**Statistical significance (p-value):**

**Set: 1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| pm\_22D | MA\_5 | MA\_22 | MA\_200 | pm\_68D\_lag | pm\_5D\_lag | y\_val |
| 0.243705 | 0.287777 | 0.350386 | 0.289766 | 0.275134 | 0.190193 | pm\_1D |
| 0.196655 | 0.230113 | 0.263700 | 0.230698 | 0.208616 | 0.001912 | pm\_5D |
| 0.000000 | 0.101086 | 0.125676 | 0.220470 | 0.226428 | 0.075184 | pm\_22D |

**Set: 2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MA\_200 | MA\_22 | MA\_5 | pm\_22D\_lag | pm\_5D\_lag | pm\_68D\_lag | y\_val |
| 0.302345 | 0.382479 | 0.356677 | 0.306682 | 0.245463 | 0.260592 | pm\_1D |
| 0.262476 | 0.193285 | 0.179514 | 0.310503 | 0.242137 | 0.000000 | pm\_5D |
| 0.000000 | 0.035861 | 0.039823 | 0.226234 | 0.220513 | 0.104686 | pm\_22D |