

Quant take-home research test #1: Market beta prediction.

Questions in this test is hard and open-ended. Please think creatively and treat them as a real quant research project. The expected workload of this test is about **10 hours**. However, you're more than welcome to spend more time and go for a **more comprehensive and creative solution**.

Please submit your results within 2 weeks after you receive the test. Please finish the questions on page 2, pack your results into a zip file, and send to: Linda.Lui@CubistSystematic.com, including:

1. A document/report that contains all important statistical results (i.e. your conclusions, findings, and comments). We prefer you submit using docx/pdf/ipynb format.
2. All relevant Python code, including:
 - 1) Please include a python driver script (.py file) that allows people to rerun and replicate your results.
 - 2) Please Include necessary Readme/Build file. (i.e. Provide instructions on packages needed to be installed. Please let us know if there are potential version problems.)

Assignment Project Exam Help

We will **score your code and report in the final submission** based on the following criteria:

1. Numerical correctness and robustness.
2. Numerical efficiency.
3. Financial intuition and creativity.
4. Clearness of your report.
5. Coding style. (We hope we can rerun your code with 1 click.)

<https://powcoder.com>

Add WeChat powcoder

The data can be downloaded at:

https://drive.google.com/file/d/1uPPN3GYKG_DnkeQbQRYMyFvgLmrcvBFZ/view?usp=sharing

Please use Python / Pandas function `pd.read_hdf()` to load data. The input data may contain NaNs and you should handle them correctly in the computations. You may assume that for each stock, NaNs appear only at the beginning (not listed yet) or at the end (no longer listed).

Questions:

1. Compute the market (daily average) return and stock's 120-day causal beta¹ to the market return. Your code should take no more than a few seconds to compute the beta for all stocks on all historical dates. Please analyze statistical features of computed variables and comment on your results.

Definitions:

$\text{mkt_ret}(\text{date})$ = daily weighted average stock raw_ret , where $\text{weight}=\text{sp_weight}$, which is the cross-sectional stock's S&P weights.

$\text{beta_120}(\text{date}, \text{stock})$ = the slope of mkt_ret in the running regression:

$$\text{raw_ret}(t < \text{date}, \text{stock}) \sim \text{mkt_ret}(t < \text{date})$$

The running regression uses an exponentially decaying weight of $\text{halflife}=120$ days, i.e.
 $\text{regression_weight}(t) = 0.5^{((\text{date} - t) / 120)}$.

2. Intuitively, a good beta should make $\text{beta}(\text{date}, \text{stock}) * \text{mkt_ret}(\text{date})$ a good estimate of the stock return driven by the market return on that date. Therefore, the univariate regression:

$Y \sim X$, where:

$$Y = \text{raw_ret}(\text{date}, \text{stock}) - \text{mkt_ret}(\text{date})$$

$$X = (\text{beta}(\text{date}, \text{stock}) - 1) * \text{mkt_ret}(\text{date})$$

Let's use the weight provided in the file as the regression weight.

should have:

- 1) A large R^2 .
- 2) A positive coefficient and a high t-stat.

Assuming we only use data provided in the input file, how to construct a better beta? **Please research on models to improve the causal beta estimation.** They should outperform the benchmark beta_120 w.r.t. the statistical measures mentioned above. Please document your method and results. Please comments on the statistical intuition.

Hint:

1. It's good to take a few years (e.g., 2000-2013) as the in-sample period and regard the rest as the out-of-sample period.
2. When is your beta estimation good/problematic? Please try to improve your estimation in bad scenarios.
3. Comment on whether your improvement is statistically significant.

¹ The word "causal" means the variable is computed only using information that has happened at (or before) a given time. In many contexts, "causal" is interchangeable with word "running".