To Do/Code:

-Identify outliers/crazy market days on a rolling basis, thus preventing some of the lookahead bias (create definition/metric before 2020?) (possible to have some evaluation metric?) (need to get outliers, but based on going up to the current time)

-some HT testing ideas?

Explain to Archit

(another alternative is number of standard deviations from 0/sample mean) (computationally more efficient?)

-What I did to come to my definition for crazy market day

->extracted the overnight returns plotted for each country’s index in a boxplot and experimented with different IQR (best for finding stuff out of the regular distribution) (last 5 years?)

->intuitively, major global events should be identifiable on the histogram, but even some of the less significant crazy market days, at least make the news (searched on Bloomberg for a fair few of these days)

->mathematically, I think the multiplier chosen prioritises precision over recall, as good for PNL. Assuming idea1 is true, didn’t pick 3x, cos too few data points? (would much rather miss some CMDs, then include non-CMDs in the sample)

-What the alternatives were and I why I chose this one over those

->number of standard deviations… I think it is the most promising, and I can weight it by date with a half-life or something, but… it is more influenced by extreme outliers, so I chose not to.

->percentile (allows more asymmetry between the markets, some markets may experience more craziness, looking for ‘out of regular return distribution’ events) (assumes same number of crazy market days for all countries…)

->abs value (different markets have different volatility, so a 4% return may mean more for a volatile market than a less volatile one

-What are some of the cons and limitations of my approach

-inconsistent timing for Japan + null values, thinking trading days between should take care of that and specify a different mic for each country

- overfit definition/look ahead? (maybe I need to split data for creating the definition?) (edge case where idea is true and H0 is rejected, by H1 is still false, or at least over estimated) (two different HTs) (depends on which hypothesis we care about more…, 2nd is geared more towards PNL, 1st is a more significant result)

(looking at data from 2016 to 2019, would it be possible for me to identify CMDs from 2020 to 2024?)

I need to ask or confirm

-How to HT the first few hypotheses and what is considered a reasonable p-value (come up with ideas first)

-weighted t-stat by liquidity?

-how is beta calculated, if it is calculated at the end of each day, then we have to estimate the beta… also what if on these crazy market days, there are significant alterations to the usual beta of stocks?

To Also work on:

-Strategies for dark, have a configurable/adaptable strategy

-controls/mechanics for dark

-communication of ideas for dark

-hedge fund strategies for dark (reversion, momentum, time lags, rubber band, flash)

-expected shares traded, EV for dark (given GTO)