**Finance Lab 2021**

**Data Analysis and Reporting for Your Quant Research Project**

G. Rao, B. Vartak

Jan 19, 2021

Please include the following analysis in your slide deck.

If you are doing ONLY back tests, you will not need the regression diagnostics. **All projects should conduct and report these test results.**

1. First table (**EVERYONE)**:

**Descriptive Statistics for Sample**

**(a**) Sample Size (ie number of observations), Time Period Covered

For all variables in analysis: Mean, median, min, max, 25th and 75th percentiles, standard deviation, skewness, kurtosis

(b). Sector and industry composition – percentage of sample observations by GICS sectors

(c) Market cap breakdown – percentage of sample by cap categories

2**. Outliers** **(EVERYONE)**

Calculate Gaussian limits i.e. median +/- 3.5\* (IQR) where IQR = 75th percentile – 25th percentile. Report number of observations beyond Gaussian limits.

Winsorize outliers 🡪 truncate at upper and lower bounds.

3. **Autocorrelation structure** (**EVERYONE)**

Autocorrleation of returns using partial autocorrelation tests.

4. **For back tests:**

Report following for total strategy returns as well as excess returns.

Histogram of returns

Mean return, median return along with p-values.

Standard deviation of returns.

Hit ratio: % of times strategy return is above benchmark, along with binomial test of significance.

**For long short strategy**: report long and short separately, as well as Long-Short. For each, mean return and excess return vs benchmark, also hit ratio. Report p-values for all these stats.

5. **For regression analysis.**

NOTE: If you detect non-normality in returns, taking natural logs will help smooth the distribution.

(a). For your predicted variable (i.e. “Y” variable in regression), report following WITH P-VALUES

- normality test of Y variable, and autocorrelation.

- **Outliers – report, identify and treat as in #2 above 🡪 VERY IMPORTANT AS THIS CAN SKEW YOUR RESULTS**

(b) Regression diagnostics for residuals

- normality – Kolmogorov-Smirnoff

- heteroskedasticity – White test

- autocorrelation of residuals – Durbin-Watson (DW) test

- multicollinearity – VIF and condition index

**If needed, stationarity test 🡪 Dickey-Fuller test if you have autoregressive residuals.**

**TREAT IF NECESSARY !!! Otherwise, regression results can be spurious. T-statistics can be inflated.**