**Introduction**

The goal of this data analysis is to determine whether the investment should be encouraged or not.

**Methods**

The available dataset includes weekly S&P 500 indices from March 30, 2020 to June 28, 2021. We transformed the time series data to weekly percent change.

Let $Y\_i$ denote the weekly percent change in the i-th week since March 30, 2020, and let $X\_i$ denote weeks since March 30, 2020. We assumed $Y\_i = \beta\_0 + \beta\_1 X\_i + \epsilon$ (referred to as M1). Under the linear model, $\beta\_1 = 0$ implies a constant expected percent change, and $\beta\_1 \ne 0$ implies a decreasing or increasing percent change on average. We tested for $\beta\_1$ at the significance level $\alpha = 0.05$.

If there would be no statistical evidence to reject $\beta\_1 = 0$, we would determine the recommendation or not based on the linear model without the intercept, $Y\_i = \beta\_0 + \epsilon$ (referred to as M0). Under this model, the parameter $\beta\_0$ is interpreted as the expected percent change which is assumed to be constant with respect to time.

The linearity and constant variance assumptions were assessed by the residual plot against fitted values, and the normality assumption was assessed by the Q-Q plot. In addition, any influential points were detected by the Cook’s distance, and highly influential point(s) were removed from the analysis.

**Results**

One influential point was removed from the data analysis (with a Cook’s distance > 0.5). Under M1, we had a lack of evidence to reject $\beta\_1 = 0$ with a p-value 0.555 and a 95% CI (-0.04, +0.02). Under M0, the point estimate of $\beta\_0$ was +0.70, and there was a statistical significance to conclude $\beta\_0 \ne 0$ with a p-value 0.018 and a 95% CI (0.12, 1.28). In other words, there has been about 0.7% gain per week on average between March 30, 2020 and June 28, 2021, and the p-value and 95% CI support for a positive expected gain during the 15-month period. As shown in figure 1 (top right panel), the percent gain has been fairly consistent during the 15-month period. In figure 1 (the other three panels), all model assumptions seem plausible according to the graphic model assessment.

**Discussion**

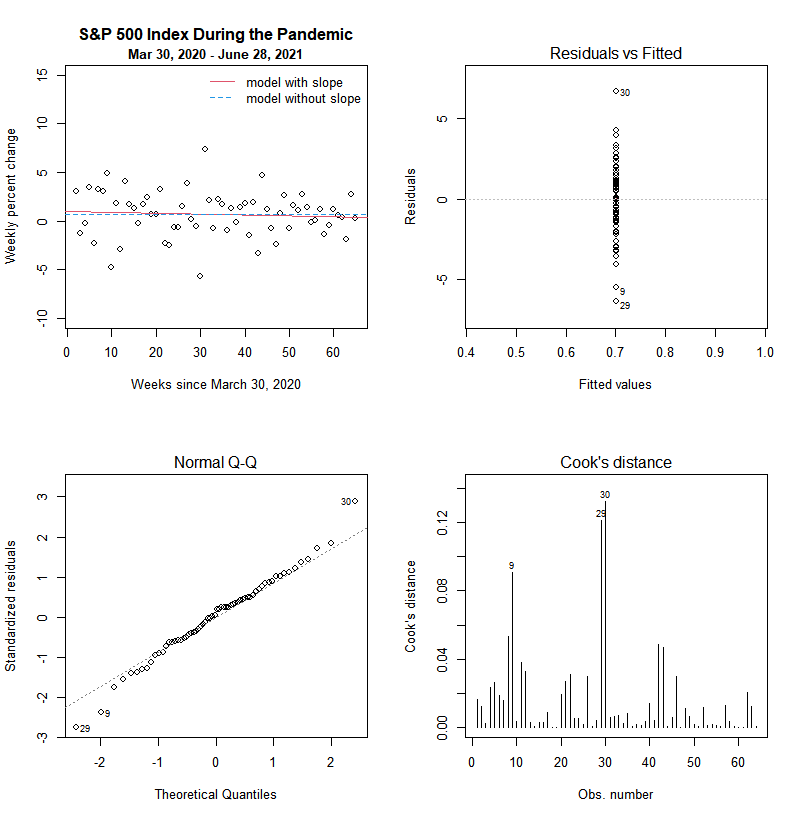
Under the linear model with the slope parameter (M1), no downtrend has been detected yet. Under the intercept model (M0), it seems evident that the positive growth has been maintained for the last 15 months. Though the estimated average weekly growth (0.7%) seems small, it is not a negligible amount if we consider the accumulated growth. There are about 65 weeks in 15 months, and (1.007)65 = 1.57 means a 57% growth over the 65 weeks. (Removing the influential point, a similar estimate could be obtained; 2789 on Apr 6, 2020 and 4291 on Jun 28, 2021).

The limitation is that we did not account for any other explanatory variables to better understand the variability. Furthermore, the long-term predictability has not been verified in the data analysis, and nobody knows what is going to happen in the next couple of weeks. Over the past year, we have been living in an era of “uncertainty” since the breakout of the pandemic. To mitigate the limitation about the predictability, we could revisit historical data during similar times (e.g., after a sudden economic crisis) and observe how long the growth lasted until the next crisis. Furthermore, any method of detecting a sudden downward trend would be helpful to protect investors.

**Conclusion**

Despite the limitations in this analysis, it is hard to find evidence against discouraging investment at this time if individuals are considering investment. There is no guarantee in any stock market, but an investment can be recommended based on the average growth during the past 15-month data. Providing some safety tools for investors will be helpful to guard against a sudden downward trend.

**Figure 1**. Caption ...



**R Code (Supplement)**

### read data (edit the directory for your machine)

data = read.csv( "sp500sincecovid.csv" )

head(data) ### view the data

dim(data) ### the numbers of rows and columns

plot( data$Index, type="l" ) ### weekly index trend

### sample size of 66 - 1 = 65

x = 1:65 ### weeks since March 30, 2020

index = data$Index ### extract the index

y = ( index[2:66] - index[1:65] ) / index[1:65] \* 100 ### percent change

plot( y ~ x, xlab="Weeks since March 30, 2020", ylab="Weekly percent change", main="S&P 500 Index During the Pandemic" )

fit1 = lm( y ~ x ) ### simple linear model

summary(fit1) ### resulting statistics

plot( fit1, 1 ) ### residual vs. fitted

plot( fit1, 2 ) ### normality assumption

plot( fit1, 4 ) ### Cook's distance

fit2 = lm( y[-1] ~ x[-1] ) ### remove the outlier (obs = 1)

summary(fit2) ### resulting statistics

confint(fit2)

plot( fit2, 1 ) ### residual vs. fitted

plot( fit2, 2 ) ### normality assumption

plot( fit2, 4 ) ### Cook's distance

### conclusion: lack of evidence that the expected percent change has not changed since March 30, 2020

fit3 = lm( y[-1] ~ 1 ) ### remove the slope from the model

summary(fit3) ### resulting statistics

confint(fit3)

plot( fit3, 1 ) ### residual vs. fitted

plot( fit3, 2 ) ### normality assumption

plot( fit3, 4 ) ### Cook's distance

### graphic for report

par( mfrow=c(2,2) )

plot( y[-1] ~ x[-1], ylim=c(-10,15),

xlab="Weeks since March 30, 2020",

ylab="Weekly percent change",

main="S&P 500 Index During the Pandemic" )

mtext( side=3, adj=0.5, line=0.25, text="Mar 30, 2020 - June 28, 2021", font=2, cex=0.8 )

abline( fit2, col=2, lty=1 )

abline( fit3, col=4, lty=2 )

legend( "topright", col=c(2,4), lty=c(1,2), legend=c("model with slope","model without slope"), bty="n" )

plot( fit3, 1 ) ### residual vs. fitted

plot( fit3, 2 ) ### normality assumption

plot( fit3, 4 ) ### Cook's distance