Time Series Analysis of Copper Prices

Daniel Hopp

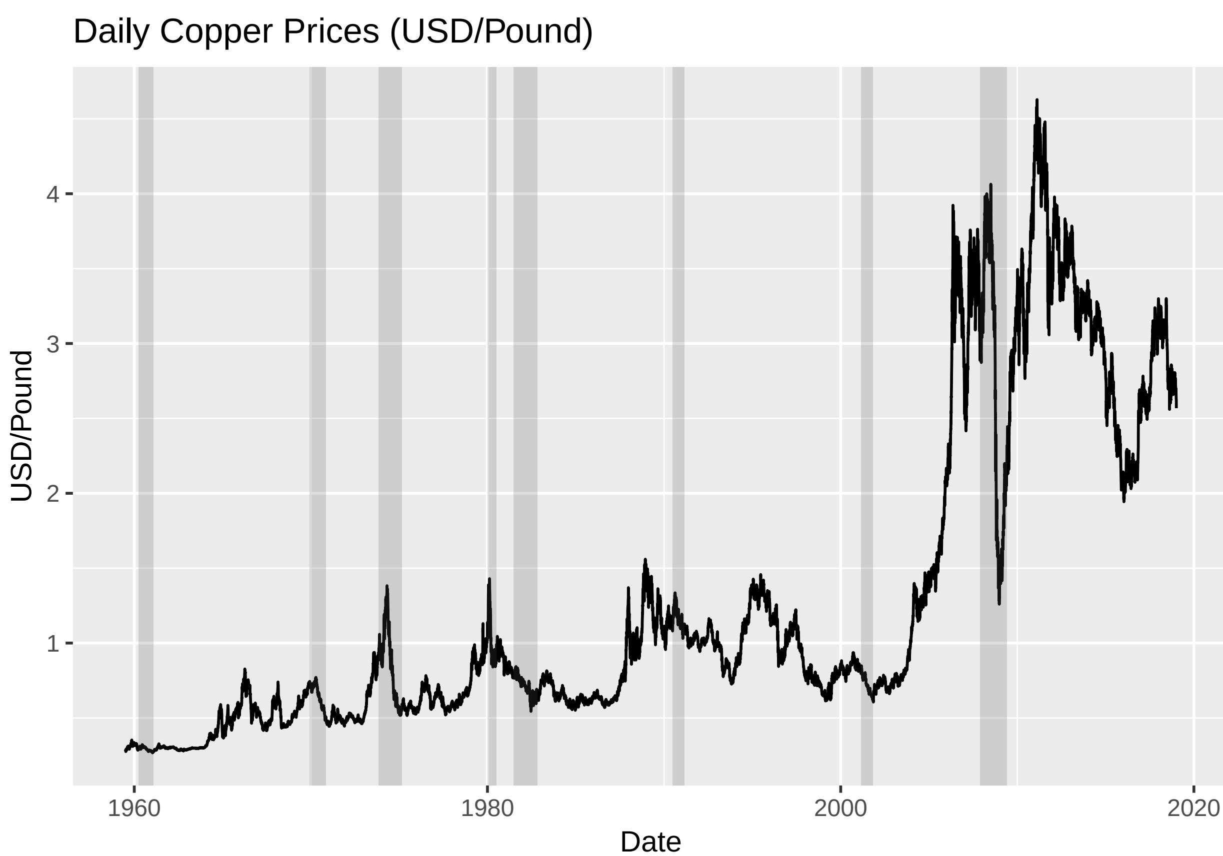
4 January 2019

Exposition of the Problem

The goal of this analysis is to get a better sense of how copper prices behave, with the ultimate goal of being able to identify, with a reasonable degree of certainty, if a day's price is the lowest that can be expected over a 4 week period. Of course the price of copper is dependent on many different factors, including prices/indicators of other commodities, inventory levels, regulatory announcements, geopolitical conditions in source countries, and indicators of downstream industries. However for simplicity and ease of use/interpretation, this analysis will stick with only copper's past prices as indicators of its future direction.

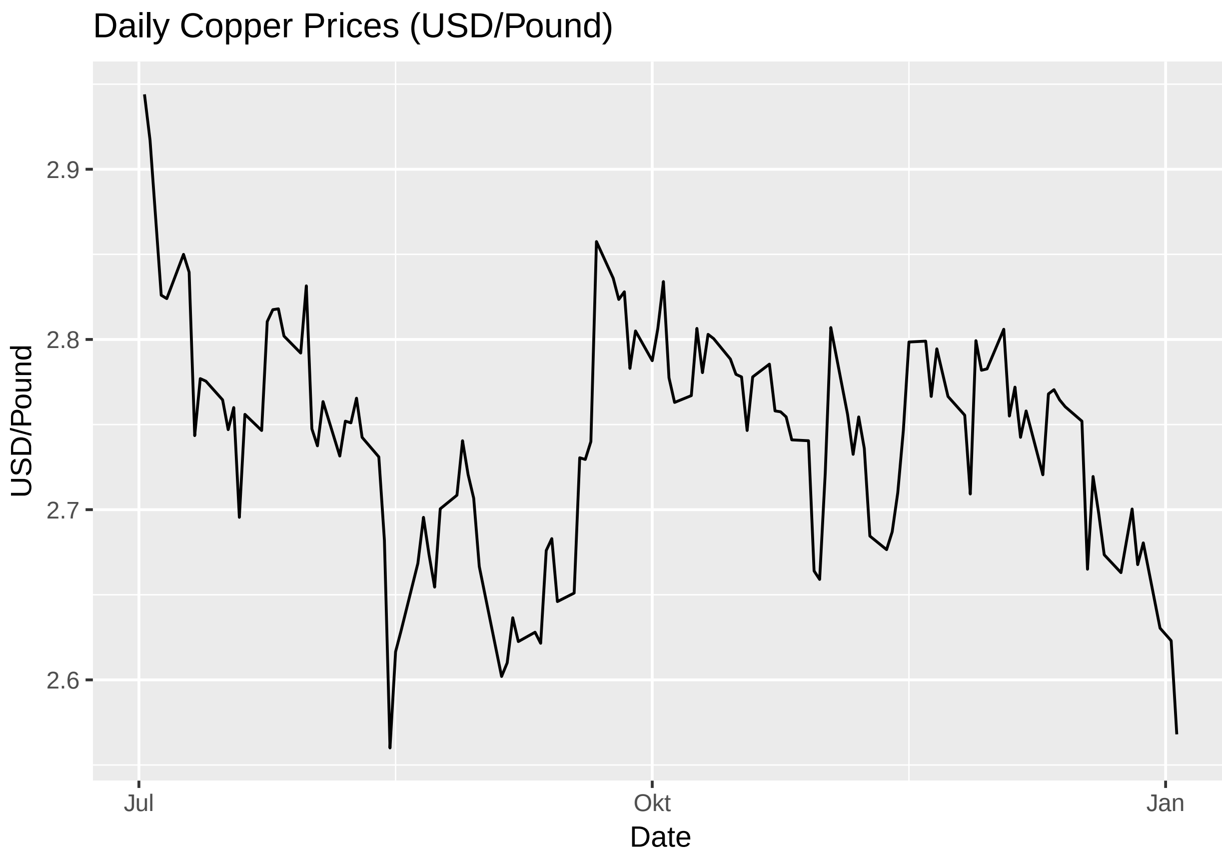
Data

Historical data on copper prices was obtained from: https://www.macrotrends.net/1476/copper-prices-historical-chart-data, which has daily copper prices (USD/pound) starting from 2 July 1959. Though it uses different units than the London Metal Exchange (USD/pound instead of USD/Ton), the shape of the historical data is the same. The source makes no mention of inflation, but we will assume prices are in constant dollars. A simple plot of the raw data will give us an initial idea of how copper prices have developed over time. Recessions are also shaded to give an indication of copper prices' relation to overall economic performance.



Interpretation

We see that on a macro level, copper prices don't seem to follow any readily discernible pattern, beyond the fact that they go down (maybe with some delay) during recessions, so some quantitative time series analyses will need to be done. But first, to get a better look at how the prices behave on a smaller time scale, we will look at the plot of the last 6 months.

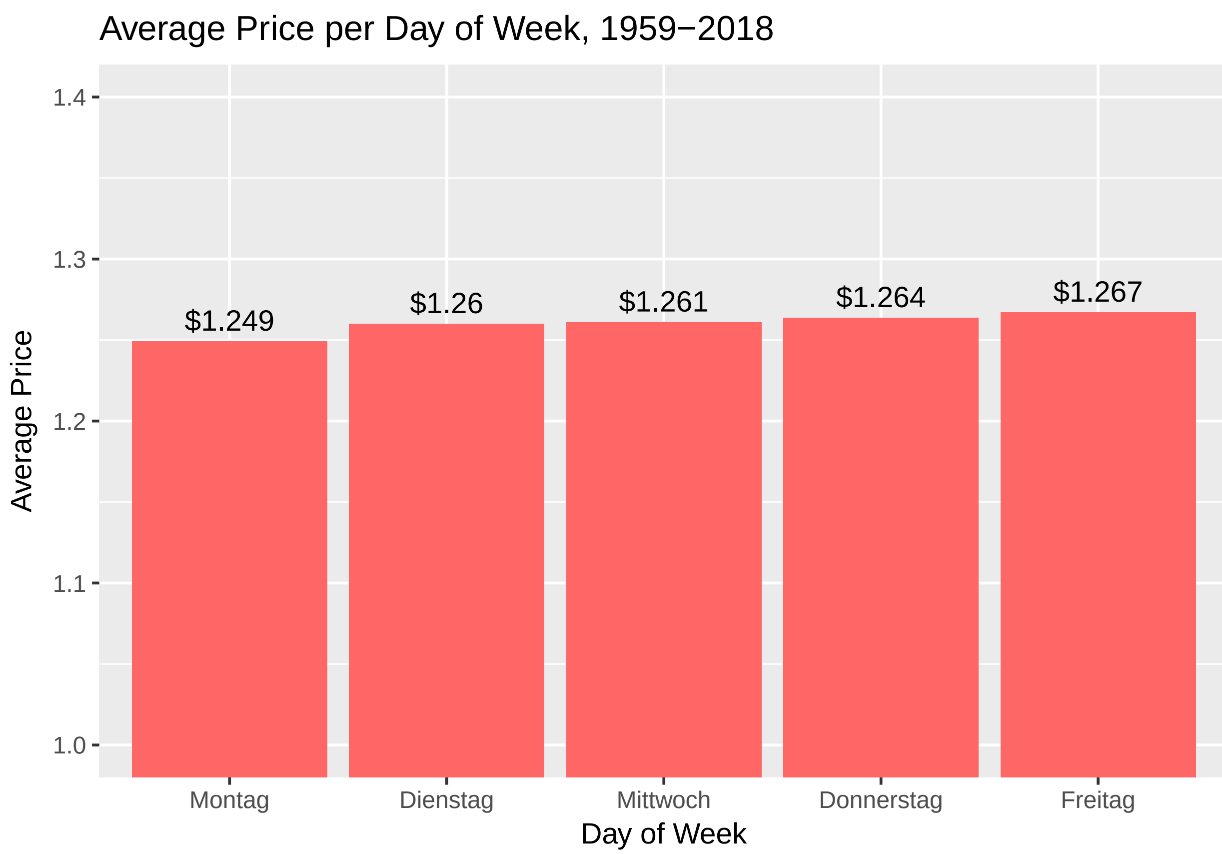


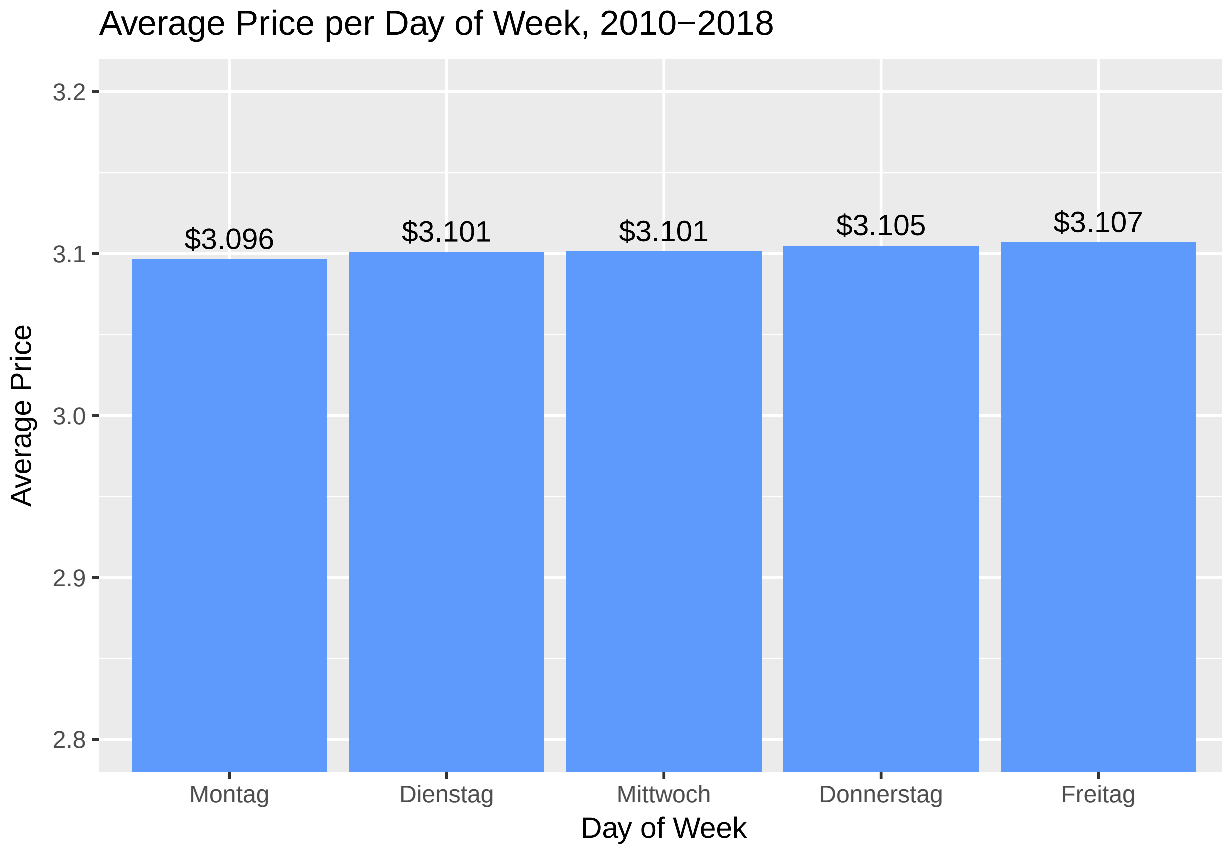
Interpretation

On a micro level, the series seems to have a certain degree of mean correction, however a random walk still looks possible. To get a better idea of the time series' nature, more analyses will need to be performed.

Copper Prices by Day of Week

Before estimating a time series model, we will look at the development of copper prices by day of week. The hypothesis is that prices may display a weekly seasonality and thus there is an ideal day of the week to buy. To examine this we'll first look very generally at mean prices by day of week.


Average Price per Day of Week, 1959−2018
Average Price
1.4
1.3
1.2
1.1
1.0
$1.249
$1.26
$1.261
$1.264
$1.267
Montag
Dienstag
Mittwoch
Donnerstag
Freitag
Day of Week

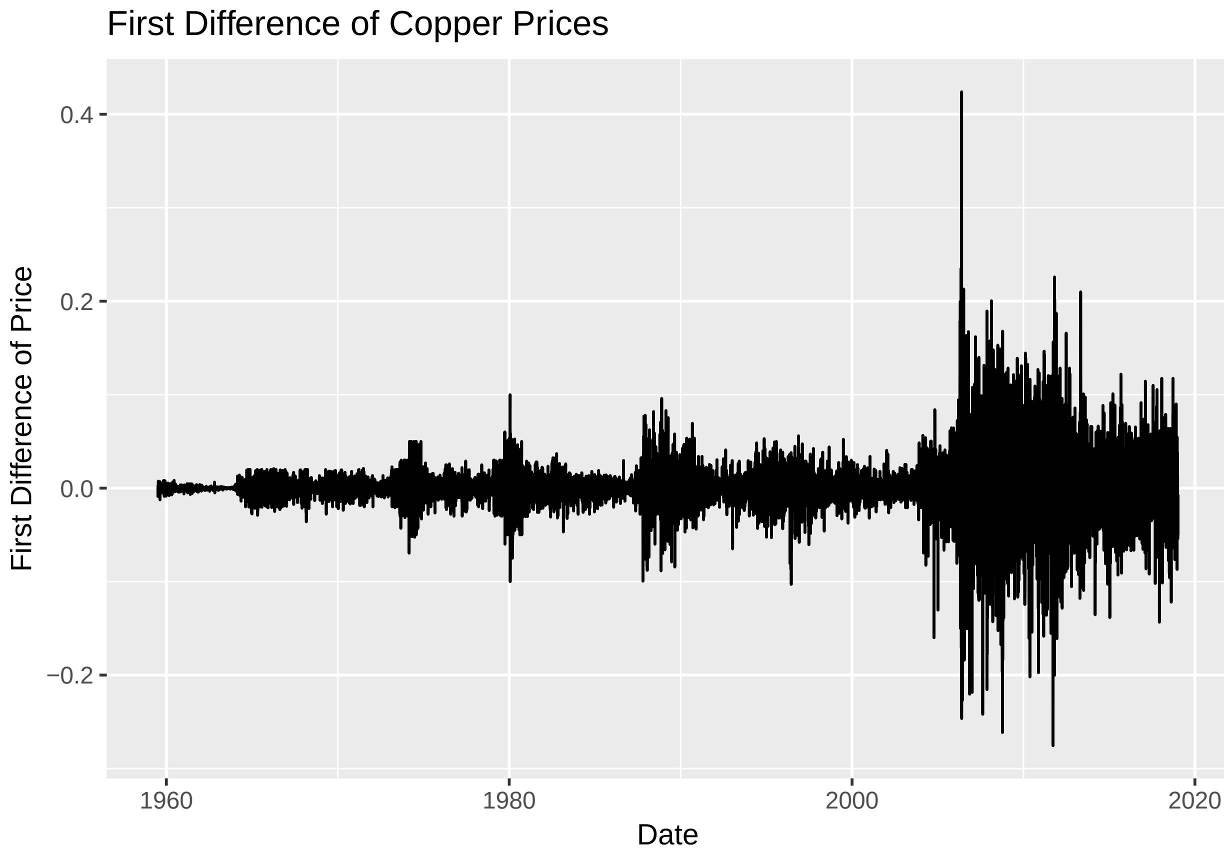


Interpretation

Over the whole data set we see a slight trend of increasing prices during the course of the week, with cheapest prices on Monday (by about only 1 or 2 cents) and most expensive on Friday To make sure this wasn't a time dependent anomaly, the same graph was produced for only 2010-2018. We see the same trend but weaker. Despite the existance of this trend, the evidence is not strong enough to be able to make a blanket recommendation for simply buying earlier in the week. So we move on to the time series analysis.

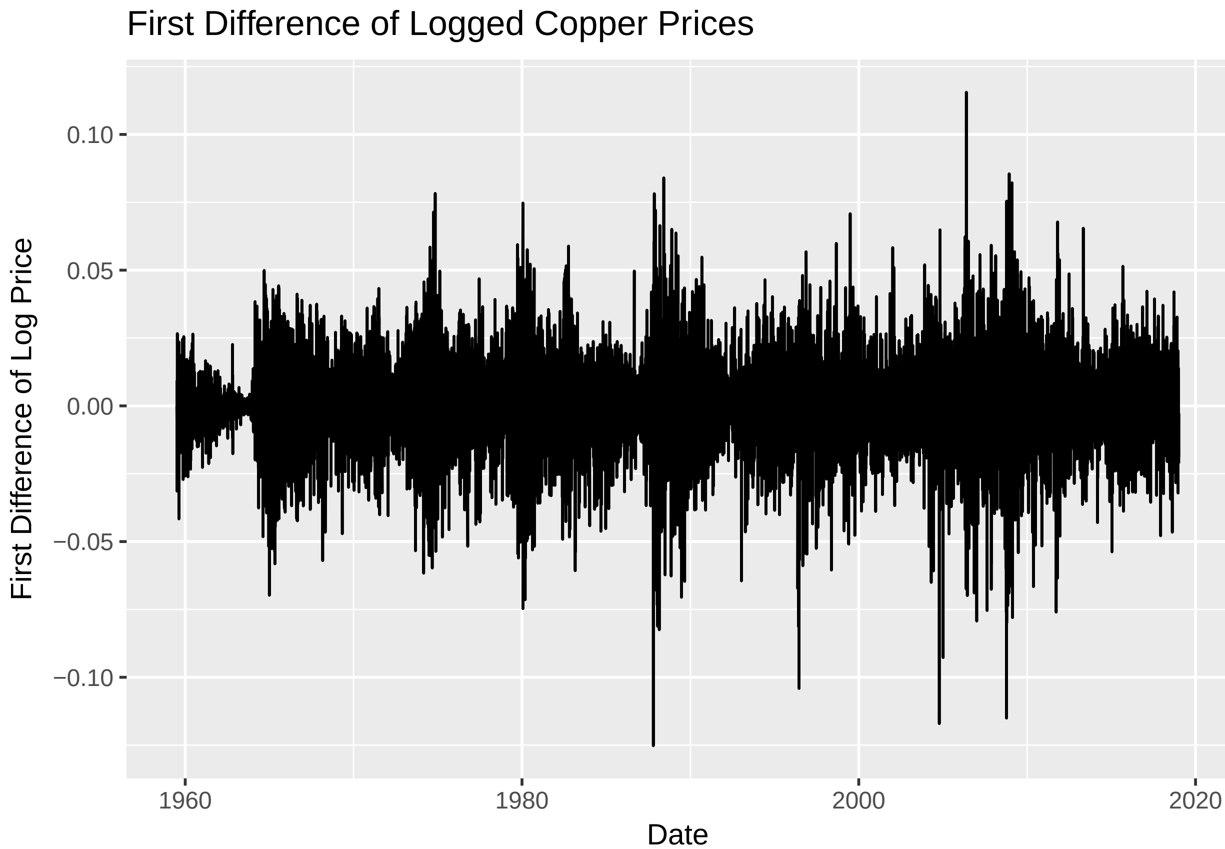
Time Series Model

We will now transform the data to obtain a stationary series. We already saw from our initial plot that the raw data is not a stationary time series, as there are clear trends. Our initial impression can be verified using an Augmented Dickey-Fuller Test for a unit root, obtaining a p-value of 0.22. This is not sufficient to reject the null hypothesis of a non-stationary series, so first we will try first differencing to see if we then obtain a stationary process. The result of that transformation has a Dickey-Fuller p-value of 0.01, indicating stationarity and is plotted below.



Interpretation

Any trends are gone, however we see a high degree of heteroskedasticity even without any tests, with variance increasing with time. This tells us we need to log transform the series, then first difference, which is plotted below.



Interpretation

To the eye it appears that the series is no longer heteroskedastic, but that can be verified with a test. The Goldfeld-Quandt test for heteroskedasticity obtains a p-value of 0.06, which is $>.05$ and allows us to accept the null hypothesis that variance does not change between segments, and thus our series is homoskedastic. Now we can use R's SARIMA function to estimate a time series model for the data.

```
## Series: ts
## ARIMA(0,0,0) with zero mean
##
## sigma^2 estimated as 0.0002619:  log likelihood=40239.36
## AIC=-80476.72   AICc=-80476.72   BIC=-80469.12
```

Interpretion

The interpretation of the auto.arima output is as follows: the optimal model, based on both AIC and BIC, is an ARIMA(0,0,0) with a mean of 0, meaning the best prediction of log differences in the time series is simply 0. Transforming back (i.e. taking away the first difference and unlogging) to our original variable (copper prices), translates to a random walk with no drift. That is to say that the best estimate of tomorrow's copper price is simply today's copper price. This corroborates Wets findings in his 2012 paper *(Wets and Rios (2012))*, available here: https://www.math.ucdavis.edu/~rjbw/mypage/Mathematics/_of/_Finance_files/WtsR13.pdf. Their logic was that though economically speaking copper prices should be mean reverting due to supply and demand, in reality supply responds so slowly to prices (due to long-term contracts, the excessive time involved to start up or shut down a copper mine) that on a shorter scale mean reversion is negligable. They have similar findings for the explanation of a lack of drift parameter in the random walk. On long term horizons there is most likely a slight positive drift parameter, but on micro scales it is effectively 0.

Implications, Below Rolling Mean Algorithm

Unfortunately this tells us that we cannot glean too much information from the series itself, and that instead copper prices are driven more by external factors like market news, etc. However since the ultimate goal is not to predict copper prices, but rather to select the lowest price in a period of four weeks, there is still further analysis that can be done. What follows is an analysis that will evaluate a day's price compared to the mean price of the past x number of days. The idea is to identify abnormally low priced days, with the two parameters being the number of previous days to include in the mean, and the threshold for considering a day likely to be the lowest, or among the lowest, in a 4 week period (e.g. 15% below mean of last 30 days, 20% below mean of last 15 days, etc.). The models will then be evaluated on two metrics, percentage of time they were correct, i.e. they were the lowest in the four week period, and mean difference from the real lowest. The result of the analysis will be a simple, digestible recommendation in the form of: "If today's copper price is x% below the mean of the past x days, copper should be purchased today to have a high chance of being the lowest in a four week period". Additionally the method will have a digestible justification in the form of: "This method correctly identifies the lowest price in a four week period x% of the time, and when it is wrong it averages x dollars more than the actual lowest price in the period."

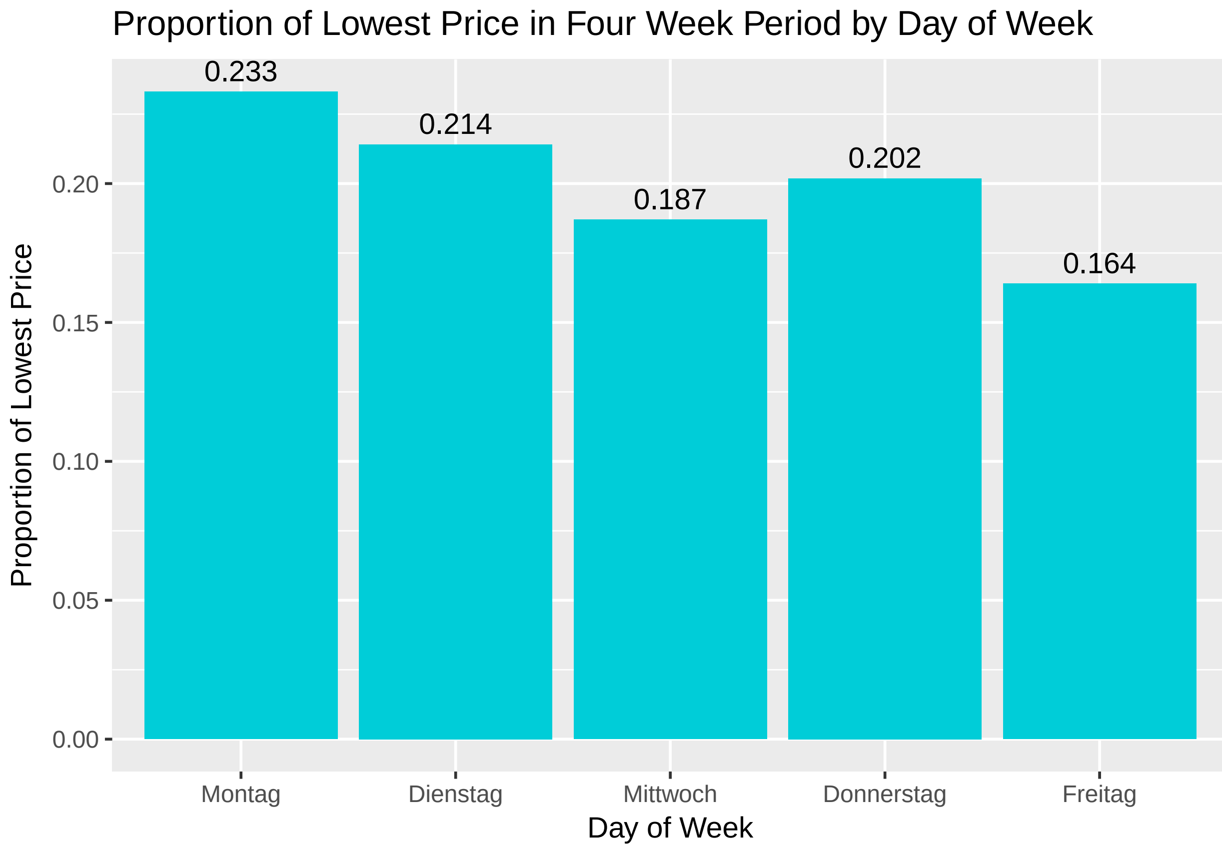
Below Rolling Mean Algorithm Results

A grid search on the two optimization parameters was run on the algorithm described above. A more detailed exposition of the algorithm is below:

- Every four week period since 1960 is split up
- Every day within that period is then evaluated based on the parameters of the algorithm, and if the following equation evaluates to true, that value is taken as the prediction for the lowest in the period and subsequent days are disregarded
  - $|X_t/(\frac{\sum_{i=t-n}^{t-1} X_i}{n}) - 1| \geq p \qquad (1)$
  - where:
  - $X_t = copper\ price\ on\ day\ t$
  - $n = number\ of\ trailing\ days\ to\ consider\ in\ rolling\ mean$
  - $p = threshold\ for\ selection$
- If the formula never evaluates to true in the four week period, the last day's price is simply taken. This can be interpreted as the "risk" factor of waiting for a favorable drop that may never occur

The optimal parameters of the algorithm differ slightly based on evaluation criterion. If best percentage correct is taken, then n = 31 and p = 3%. If lowest mean error is taken, then n = 27 and p = 3%. The first parameters result in a correct decision in 18.7% of the four week periods between 1960 and 2019, and a mean error (i.e. how many cents off it was from the actual lowest price in the period) of 0.059. The second attained results of 18.43% and 0.057. These are good results and are almost 4x better than the results that would be obtained by luck (lowest price correctly picked 5% of the time). Given that the objective is not simply to be "correct" the majority of the time, the n = 27, p = 3% model is most likely preferable, as it minimizes mean error. A final metric important to interpretation/understanding of the algorithm is what percent of the time equation (1) is not fulfilled (i.e. the last price in the period is taken). For the n = 27, p = 3% algorithm that number is 42.55%. That means that this % of the time the conditions will not be filled.

Additional Day of Week Insight A final interesting insight, which corroborates the earlier findings of lower prices on Mondays, is the plot of how often the lowest price of a four week period falls on each day of the week. The plot is below.



We can see a trend that mirrors that of average prices by day of week, with Monday most often having the lowest prices, and Friday least often.

Closing Thoughts and Recommendations

Based on the findings and historical data here, the ideal method of choosing the lowest copper price in a four week period would be to choose a price on a given day if it is 3% below the mean price of the past 27 days. Additionally, the historical data show that this lowest price is most likely to come on a Monday. Combining this analytical approach with market research and heuristics can potentially provide a much stronger purchasing methodology than the two alone.