A Statistical Investigation of the Simon-Ehrlich Wager

Brendon Hahm, Nicole Lee, Jason Wong

I. Introduction

Motivation

In 1980 there was a famous wager that pitted Julian L. Simon, a business professor, with Paul Ehrlich, a biologist. Ehrlich was famous for his controversial book *The Population Bomb* in 1968 where he argued that population growth will exceed resources leading to future mass starvation. Simon saw this grim forecast of the future as a repeat of the infamous Malthusian Catastrophe. A Malthusian Catastrophe is when agricultural production can't support population growth and forces a limiting of population back to sustainable levels. Thomas Malthus famously shared his theory in *An Essay on the Principle of Population* and it was clear that innovation from the Industrial Revolution improved production enough to sustain population growth. Just as Malthus seemed to have been wrong, Simon also thought the same reasoning applied to Ehrlich's forecast.

Simon, having faith in improving agricultural production, wagered that should Ehrlich choose any five metals, none of the metals would increase in inflation-adjusted prices. Ehrlich, believing that excess population would increase demand, argued that the prices would increase. He chose nickel, tin, copper, tungsten, and chromium as the five metals to be wagered on and the year 1990 as the year to check the prices. By 1990, all of the metals inflation-adjusted prices decreased leaving Simon the winner of the wager.

What makes this wager particularly interesting are the theoretical backgrounds of the participants. Ehrlich came at the problem from an ecological approach while Simon from an economic approach. However, if we were to build a predictive model, which position would be statistically supported? To be more explicit, if we build a predictive model, would the model forecast an increase or decrease in inflation-adjusted prices?

Dataset

We collected metal and macroeconomic data from two sources. Data on the five metals and market information comes from the US Geological Survey (USGS). We pulled five separate datasets each corresponding to a metal. These datasets have relevant market information such as production levels, scrap levels, imports, prices, and more. We also pulled general macroeconomic and industry level data from the Federal Reserve Economic Data (FRED). The macroeconomic data from this source are mostly used as standard control variables for things such as general economic trends (Gross Domestic Product) and inflation (Consumer Price Index). We also include general industry data. As opposed to the data from USGS, the data from FRED applies to all metals as they are national and industry wide.

The combined dataset was large but had many null values. The handling of these null values were particularly difficult given that which values were null values did not appear random. Given the apparent structurality to the null values, we made conservative measures by removing columns with many nulls and rows that tended to be years with many nulls across the metals. With the remaining nulls, we handled them on a case by case basis.

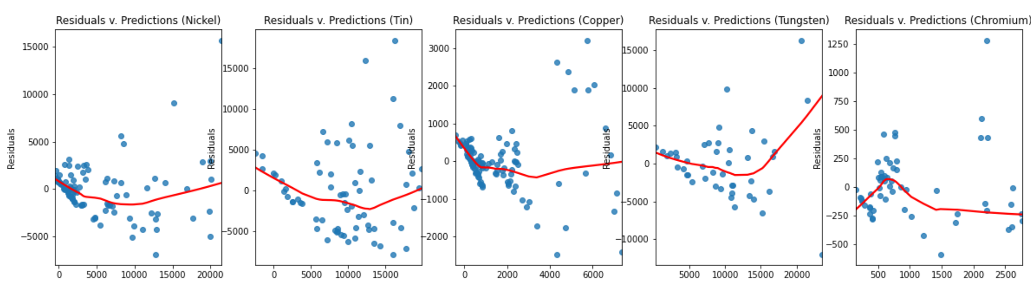
Hypothesis

Could a statistical model have helped inform the answers of the participants of the wager using data only up to 1980? If so, which position would be supported? Can we build a statistical model that can predict metal prices with more modern data? While it is clear in retrospect that Simon was the winner of the wager, it is not clear if a model would have supported Simon's position. Malthusian catastrophes like Ehrlich's are often incorrect because of their inability to account for improving production functions, which are induced by innovation. While we account for capital investments and national Total Factor Productivity (TFP) levels in our data, it is not clear whether our model will be able to accurately account for the degree of innovation that supported Simon's position. We hypothesize that the models will support Ehrlich's position and predict increasing prices.

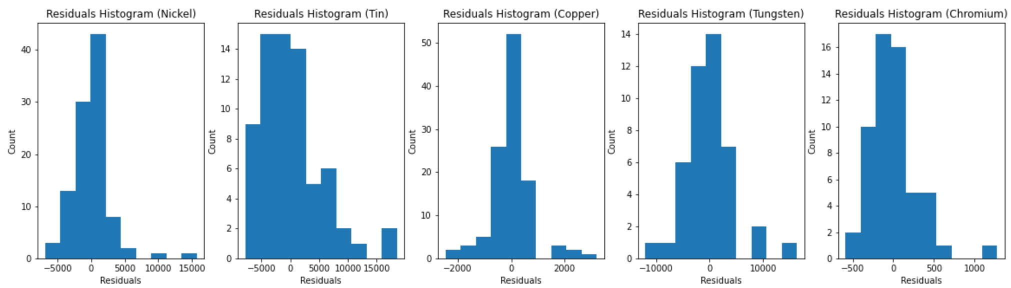
II. Methods

Our primary objective is to forecast metal prices. Therefore, we will be appropriately treating this as a prediction problem. With a held out test set documenting only the most recent data, we will be training a model using our validation and training datasets created during cross validation.

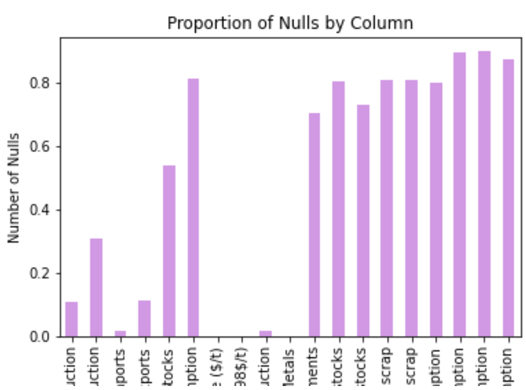
To appropriately choose a model, we first conducted exploratory data analysis. Our results from fitting a simple linear regression model affirmed our original hypothesis that a linear regression model would be the best. The figure below shows prediction vs. residuals plot for each metal, and the random distribution of points is one indication of linearity.

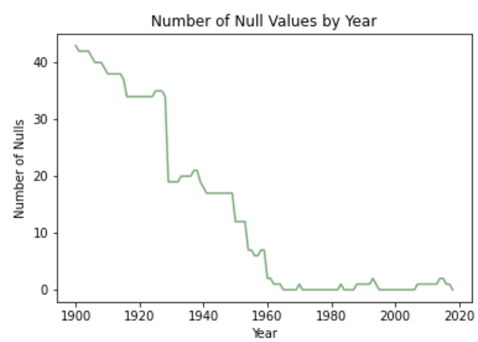


The second indication of linearity is the residual histogram. Based on the figure below, we can see that the variance is normally distributed, which once again affirms that a linear model is an apt representation of our data.



After determining our model, we began to use various approaches to select features. Our first method was to evaluate the nulls in our dataset. We opted to drop columns with a significant portion of the data missing. We determined that we did not need columns like “Reported Consumption” with ~90% nulls for our model, while we could ignore nulls in columns with ~30 percent missing data. Although we set 50% to be our boundary for now, we will continue to perform feature selection below. In addition, we determined that imputing values into null rows based on date would heavily skew our data (as data was *not* missing completely at random), based on the plot below, so we dropped all years with missing data after dropping the columns (mainly rows before 1960). The missing data after 1960 were imputed with the mean of each column.

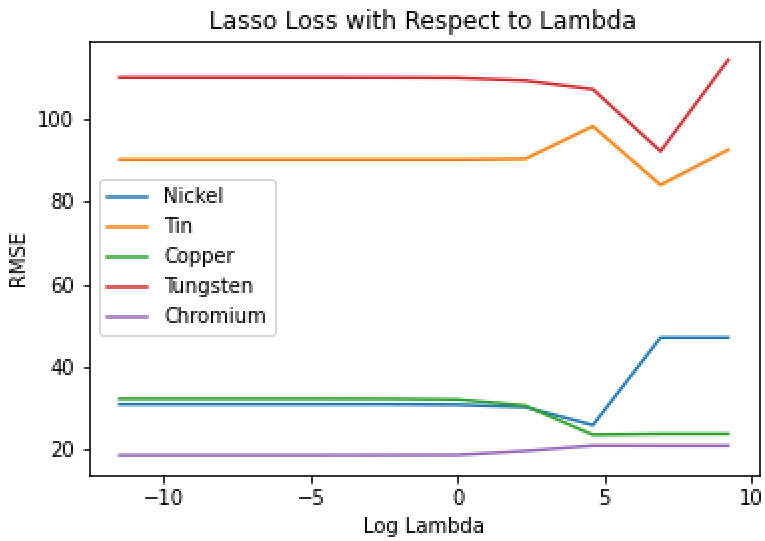




III. Results

Simon-Ehrlich Wager Models

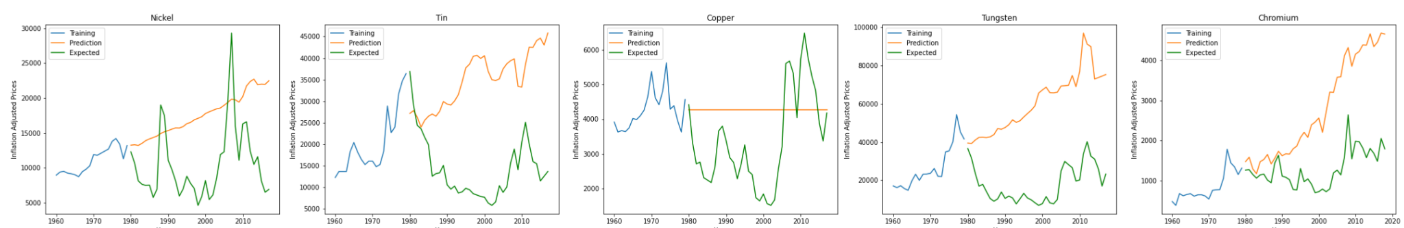
In continuation of our effort to select features, we fit lasso regression models, using cross validation to determine good λ values. After standardizing our data ($\frac{x-\mu}{\sigma}$), we performed lasso regression for each metal with 10 different lambda values from e^{-5} to e^{5}. In order to determine the RMSE to find the best λ value, we used 3-fold cross validation. Each metal yielded a different optimal λ value, as shown in the following plot.



With that, we determined our optimal lambda values to be the following:

Metal	Lambda Value
Nickel	100
Tin	1000
Copper	100
Tungsten	1000
Chromium	1e-05

Our predictions beginning in 1980 are shown in the figure below.

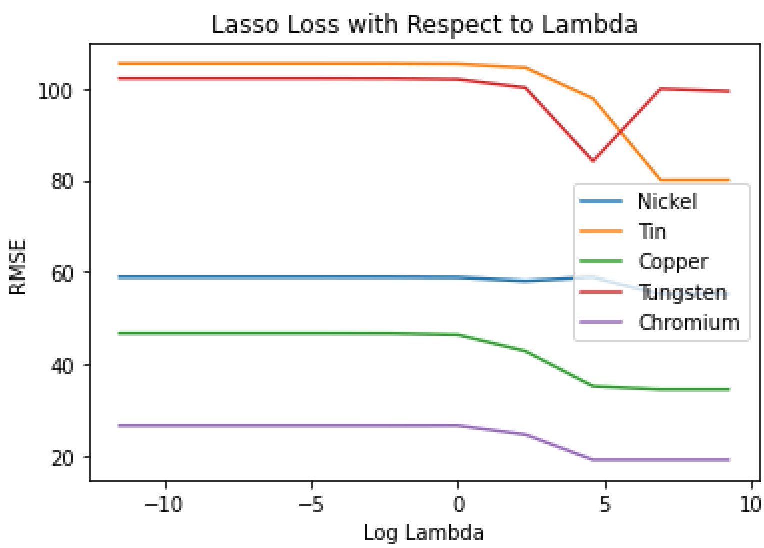


Metal	Train RMSE	Test RMSE
Nickel	787	8781
Tin	5282	23323
Copper	547	1636
Tungsten	6277	44827
Chromium	146	1765

Clearly from these results, we can see that none of the five LASSO regressions were able to accurately predict the decrease in actual prices after 1980. Just as hypothesized, all models overpredicted the price of metals and made predictions that prices in 1990 would be higher than prices in 1980. There are many possible reasons for these results and we discuss those in the following section.

General Metal Prices Models

A limitation of the previous regression analysis was the limited number of observations accessible. Given that we had to use data before 1980 (the time of the wager), we were given very few data points to train our model. Since we had data all the way up until 2019, we wanted to use more data to train to see if we could accurately predict more recent metal prices. We used all data before 2015 to predict inflation-adjusted prices of 5 metals using LASSO regression. Then we used data after 2015 as our testing set to see whether a more abundant training set will help model performance.

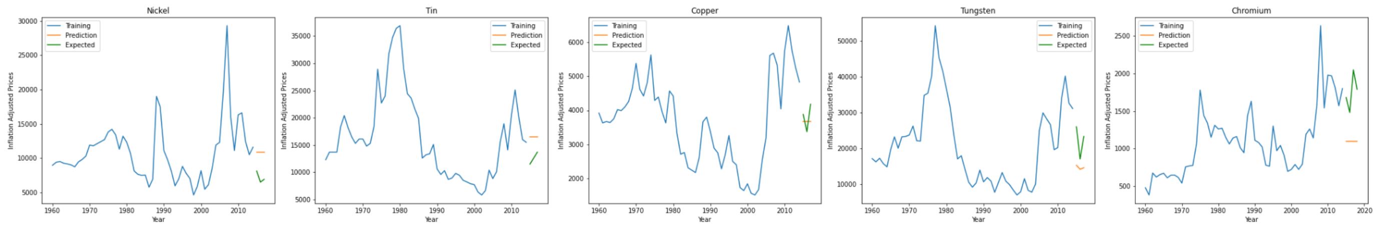


Picking the minimum RMSE, we find that the optimal lambda values to be:

Metal	Lambda Value
Nickel	1000
Tin	1000
Copper	1000

Tungsten	100
Chromium	100

Using the models at the optimal lambda values, we find the predicted values of inflation-adjusted prices in the figures below:



Metal	Train RMSE	Test RMSE
Nickel	4209	3706
Tin	7652	3950
Copper	1259	362
Tungsten	8366	8175
Chromium	446	688

The results of these regression models indicate that there was no clear over or under predicting of true values. While the models predicting Nickel and Tin overpredicted the prices, the models predicting Tungsten and Chromium underpredicted the prices. The models predicting Copper and Chromium happened to perform very well but we discuss later why these results are likely less reliable than these mere RMSE scores indicate.

IV. Discussion

Possible Sources of Biased Predictions

Our analysis did not demonstrate results that correctly predicted inflation-adjusted prices. While this was initially hypothesized, we thought our data might have been able to prove us otherwise. We can see that the predictions of our models overpredict the prices of each of our metals. This is a possible indication that general trends that apply to all five metals may be biasing our results positively where our model should be predicting negative trends. One possible explanation of a general confounder to all metals is the Great Inflation of the 1970s. This event is notable for the soaring inflation rates that came from lax Federal Reserve monetary policy. We were hoping that since our model includes CPI and GDP in our model, such bias would be controlled for.

Even if our model was successful in controlling for the Great Inflation of the 1970s, there is another important consideration to the limitation of our analysis. Malthusian Catastrophes are often predicated on the belief in a static production function. In reality, production functions improve and change based on innovation in the production line of a firm. Ehrlich believed the prices would increase due to excess demand. However, such demand never actualized because innovation and improvements to the production of agriculture materialized in what is known as the Green Revolution (or the Third Agricultural Revolution). This event helped curb anticipated demand that would have led to increased prices. In the economic literature, economists understand Total Factor Productivity to be the measurement of innovation. While we did control for TFP, we could only do so on the national level. This limitation also likely contributed to the bias in our predicted trends.

The final notable limitation of our analysis comes from the lack of data points we could use for training. Given the historical nature of our task, we restricted data points from 1980 for training. This left us with so little data points for training our model that there was likely a strong bias. Given that most of the data we used occurred during the Great Inflation of the 1970s and before the effects of the Green Revolution could be materialized in the price of metals, there are many possible indications that our model had bias. Although this bias from such a small amount of training points seems to be unavoidable and not really something we could have avoided. The later half of the 20th century is often described as the beginning of the Information explosion, the general phenomena of rapid increase in published information and data. Before this explosion, data was very scarce and was not as widely prevalent as it is today. Our restriction of data after 1960 was mostly due to the high prevalence of null values before that year.

The General Models

We also ran the same method as we did with restricted models using more data points for training. We used years up to 2015 as our training set and the years 2016-2019 as our testing set. What we found was that our models did not overall over or under predict prices. We found that the Copper model performed quite well on the testing set. However, we also noticed that the model performed worse on the training set than the testing set, indicating strong underfitting. The best performing model was our model for Chromium, which had the least indication of either underfitting or overfitting with relatively low testing RMSE. Across both the restricted and general models, we can see strong levels of regularization indicated by the lambda values. In the following subsection, we discuss the peculiar underfitting that occurred and possible sources.

Overfitting and Underfitting with LASSO

In a normal OLS regression, the training RMSE will almost always be lower than the testing RMSE. This is because measuring the training RMSE amounts to measuring the model's fit to the data it was trained on. While testing RMSE amounts to measuring the model's fit to different data that has not been trained on. Regularization regressions, like the LASSO regression used in our analysis, does not hold this same relationship with training RMSE being lower than testing RMSE. This is because the estimation of parameters considers the additional cost function of the magnitude of the parameters. We can see in our wager analysis, Copper predicts a completely horizontal line in our visualization. What must be happening is a complete regularization such that all the parameters are being zeroed out by our cost function. This likely occurred because our model predicted a positive trend for prices over time, and as lambda increases, our parameters begin to zero out. Since the actual trend was negative, our LASSO chose the lambda value that resulted in a completely horizontal line since that minimized the testing RMSE (since the testing data points were negatively trended). This unique situation really highlights the limitation of the data we used. While LASSO is a powerful tool that can help combat overfitting, through regularization, our data was simply not well equipped to forecast future prices.

It is notable that our general models, which used more data to train the models, had less instances of underfitting than the restricted models. This makes sense given that more training points lowers the variability of RMSE which may have been causing the underfitting in the restricted models.

Model Estimation and Forecasting

When we increased the training set years to predict more recent year price levels, we saw very minimal contributions to the reliability of the models performances. RMSE only marginally improved and our models were still overfitting and underfitting unpredictably. Due to the severe limitations in our forecasting abilities for prices and poor model performances, a model fitted to all the data points lacked any motivation to do so.

V. Conclusion

There are many considerations for the kind of model building our project attempted. Metals are an integral part of society and play a fundamental role in the production line of many other goods. The demand for metal changes as production of goods change so the prices of this commodity can be very difficult to model correctly. Not to mention general macroeconomic trends and many of the other considerations described in the previous section. While we were not able to construct models that correctly predicted the negative trend of inflation-adjusted prices of the metals in the 1980s, we can conclude that our models supported Ehrlich's position given their predictions of increasing price trends for all five metals.

As for our general models, the performance of these models were only slightly improved by the inclusion of more training data points. In multiple different parts of this report, we discuss various limitations of our analysis. While our Chromium model performed well, the poor performance of the other four models should question the reliability of this model in future prices.