

5250 Project: Gold Prices in the U.S. From 1979-2021

Meredith Kadlac, Evan Scheerer, Cameron Sowers

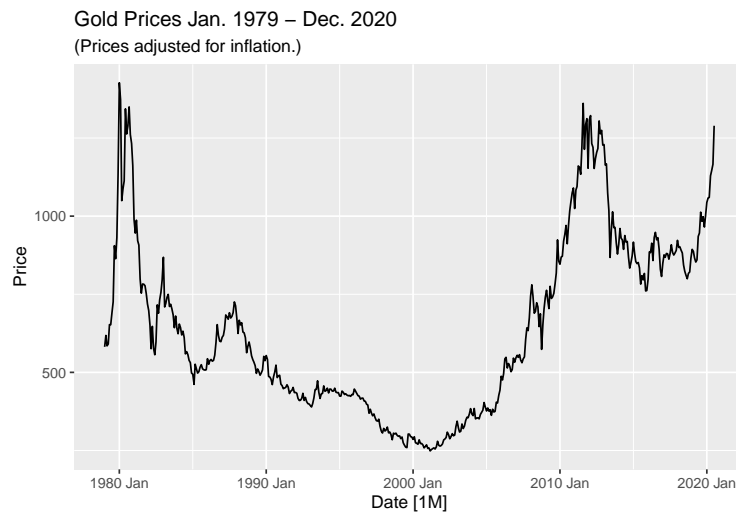
November 24, 2021

Introduction

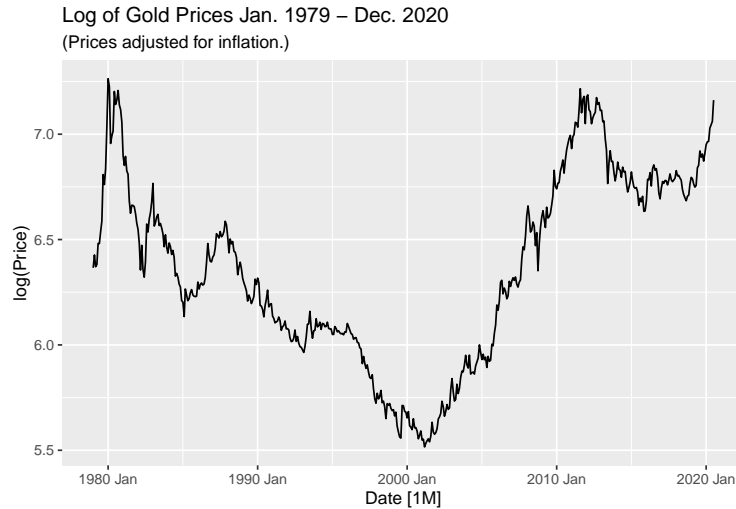
Our data set contains historic gold prices of 18 different countries in their respective currencies. We are only focusing on United States of America, as this will be the most useful and interpretable for us. Because we are dealing with currency, we have adjusted for inflation over time. The data occurs from January 1979 through July 2021 and we have split the data into a training set and a test set to see how accurate our forecasts are for selected models. Since forecasting currency is historically not accurate, we have only cut off the last 12 months of data and reserved it for our test set, while the majority of the data is used to train our forecasting models. The dates for the training set are January 1979 – July 2020 and the dates for the test set are August 2020 – July 2021. Below we will explore the data, fit appropriate ARIMA models, and forecast the last 12 months of the data set as well as an additional 12 months into the future.

Data Exploration

After adjusting our raw data for inflation of the U.S. dollar over time and then splitting our data into a training and test sets, we then continued to work exclusively with the data in the training set in order to explore the data and fit our models. First, we wanted to visualize our time series and look for seasonal patterns and/or trend, and also get an idea if we needed to transform the time series at all.

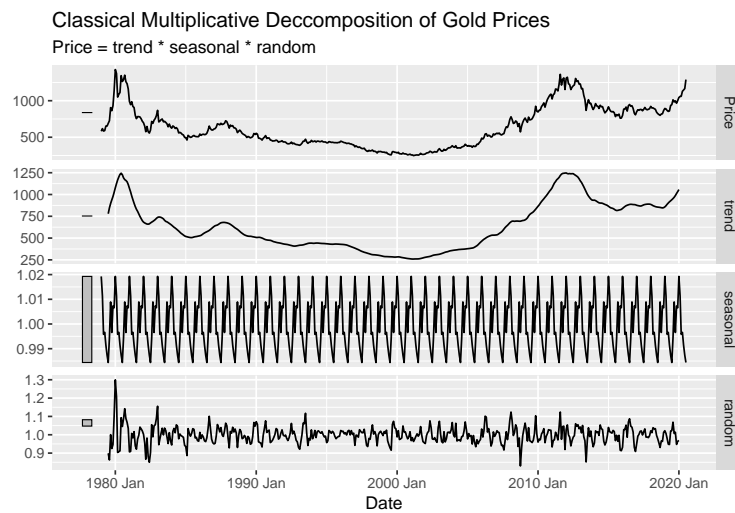


Looking at our data, it made sense to perform a log transform because of the periods of exponential growth we can observe in the raw inflation-adjusted data. Below is the data after taking a log transform, and we continued fitting our models using this transformed data.



Classical Decomposition

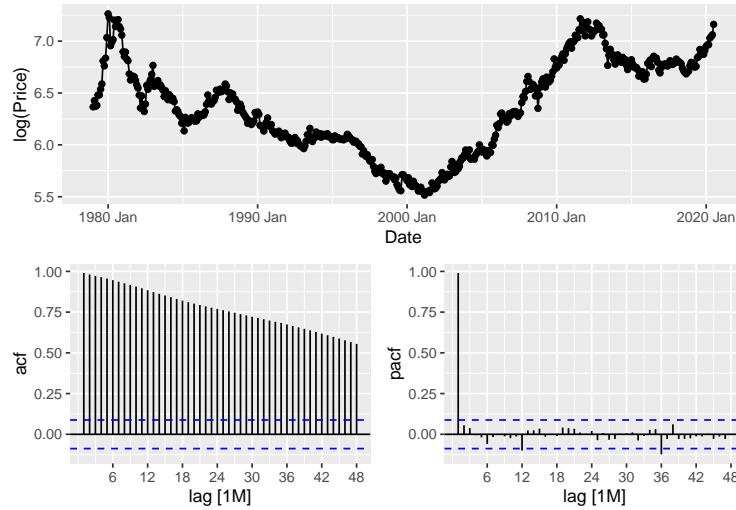
Since we decided to use a log transform, we further visually broke down our data using the multiplicative form of a classical decomposition method.



We see above there is an extremely large trend component to our data, while seasonality makes up a very small amount of change in the monthly gold price. The random component is not too large and seems to be white noise.

ACF and PACF

We then looked at the autocorrelation function and the partial autocorrelation function of our log transformed data.



Based on the above plots, it was clear that the series is not stationary and so differencing was necessary to remove trend from the data. This is consistent with what we inferred from the decomposition before.

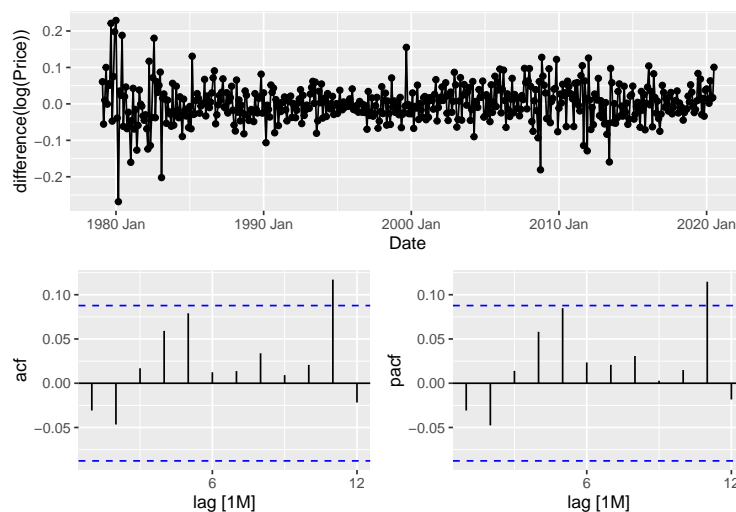
Difference Testing

To be certain of how many times we should difference the data, we conducted a few unit root tests. Our group found that there was significant evidence to difference the data once, but not twice.

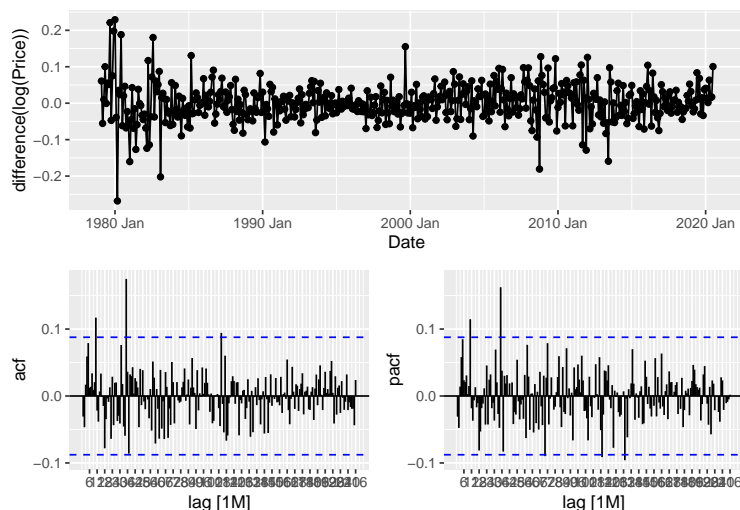
Differences	kpss_pvalue
$\log(\text{Price})$	0.01
$\text{difference}(\log(\text{Price}))$	0.10

ACF and PACF of Differenced Data

Then we looked at the autocorrelation function and the partial autocorrelation function of the appropriately differenced data to assess the plots and suggest a possible model to fit in the next section. Recall that the data is monthly.



Looking at one year of the data, we don't see much indication of a p or q value greater than zero.



The plots show a few significant seasonal values on both the ACF and PACF near the lags of multiples of twelve, with somewhat tailing off in both plots so our suggestion of a potential model is an $ARIMA(0, 1, 0) \times (1, 0, 1)_{12}$.

Model Fitting the Inflation-Adjusted log(Price) Data

After exploring the data and deciding on the appropriate adjustments that the data needed, we then moved on to the model fitting portion of the project. In total, we fit three models to the data: one model of our own choosing and two models produced by different versions of the Hyndman-Khandakar algorithm. We fit these models and then compared them on the basis of AICc to determine which was likely to forecast the time series best.

Model 1 - Guess based on ACF/PACF

As stated previously, we chose to make our first model an $ARIMA(0, 1, 0) \times (1, 0, 1)_{12}$ based on examination of the ACF and PACF plots shown before.

Model 2 - With Algorithm

The second model was decided based on the output of the normal version of the Hyndman-Khandakar algorithm. This algorithm suggested an $ARIMA(0, 1, 1) \times (2, 0, 0)_{12}$ model for the data.

Model 3 - Longer Algorithm

The third and final model was determined based on a longer and more in depth version of the Hyndman-Khandakar algorithm, which searches over a much larger set of models to select the best option. The output here was an $ARIMA(0, 1, 0) \times (0, 0, 0)_{12}$, which is a random walk.

Comparison of the Models

Model	sigma2	log_lik	AIC	AICc
3: ARIMA(0,1,0)(0,0,0)[12]	0.0027300	763.3385	-1524.677	-1524.669
1: ARIMA(0,1,0)(1,0,1)[12]	0.0027385	763.5531	-1521.106	-1521.058
2: ARIMA(0,1,1)(2,0,0)[12]	0.0027400	763.9133	-1519.827	-1519.745

The table above shows our models and various criteria for assessment. On the basis of AICc, the longer version of the Hyndman-Khandakar algorithm gives our best choice, Model 3: $ARIMA(0, 1, 0) \times (0, 0, 0)_{12}$. The second best option is actually our suggested Model 1: $ARIMA(0, 1, 0) \times (1, 0, 1)_{12}$, while Model 2 given by quick version of the Hyndman-Khandakar algorithm is the least suitable: $ARIMA(0, 1, 1) \times (2, 0, 0)_{12}$.

Residual Analysis

Moving forward, we chose to examine the residuals of the top two models: Model 3 and Model 1, to determine whether they were appropriate to use for forecasting.

Ljung-Box Test

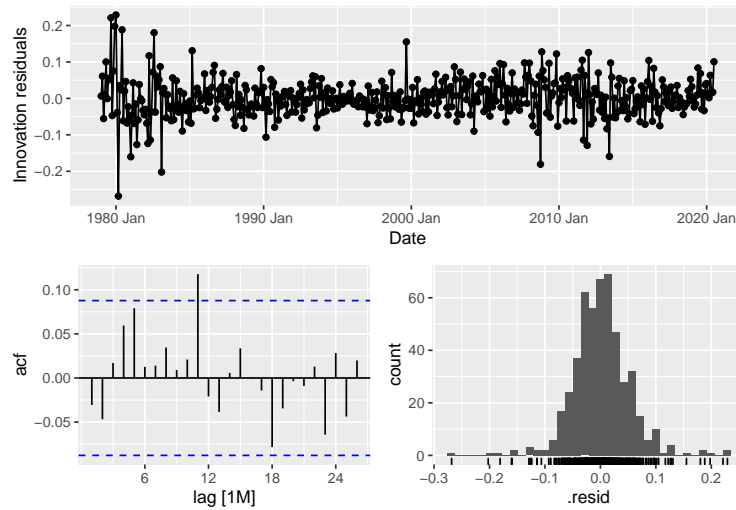
Model	p_value
3: ARIMA(0,1,0)(0,0,0)[12]	0.656
1: ARIMA(0,1,0)(1,0,1)[12]	0.646

We saw that the p-value for both Model 3 and Model 1 is very high at 0.66 and 0.65 respectively, so there is not enough evidence to reject the null hypothesis that the residuals are uncorrelated. This means that the both models pass the Ljung-Box test.

Residuals Plots

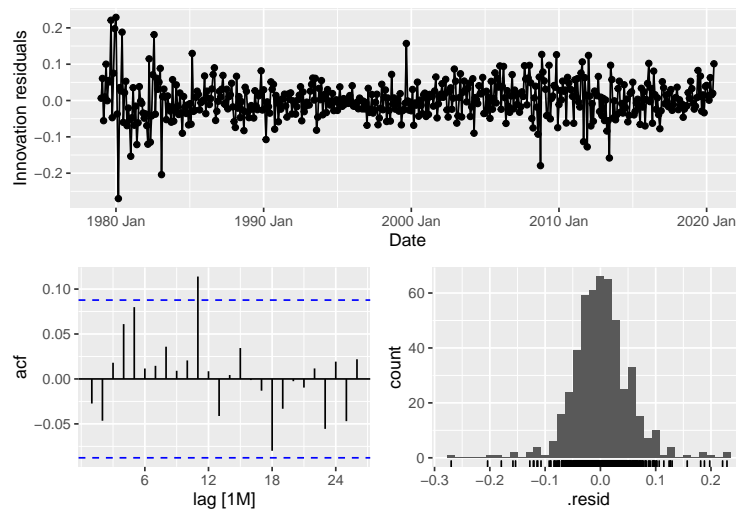
We found that both models 3 and 1 have residuals that indicate forecasting using normally distributed methods would be valid. This is explained in detail below.

Model 3



Notice, the residuals are roughly normally distributed with mean zero, and the slight change in variance does not seem to be of concern. While there seems to be a borderline significant value at lag 12 in the ACF, as stated above this model passes the Ljung-Box test so this is not of concern. Therefore, forecasting using normally distributed methods will be valid with this model.

Model 1



Our residuals are similar for the second model. They are roughly normally distributed with mean zero, and do not have steadily increasing variance. While there again seems to be a borderline significant value in the ACF again at lag 12, as stated above this model passed the Ljung-Box test. Therefore, forecasting using normally distributed methods will be valid with this model as well.

Forecasting and Model Assessment

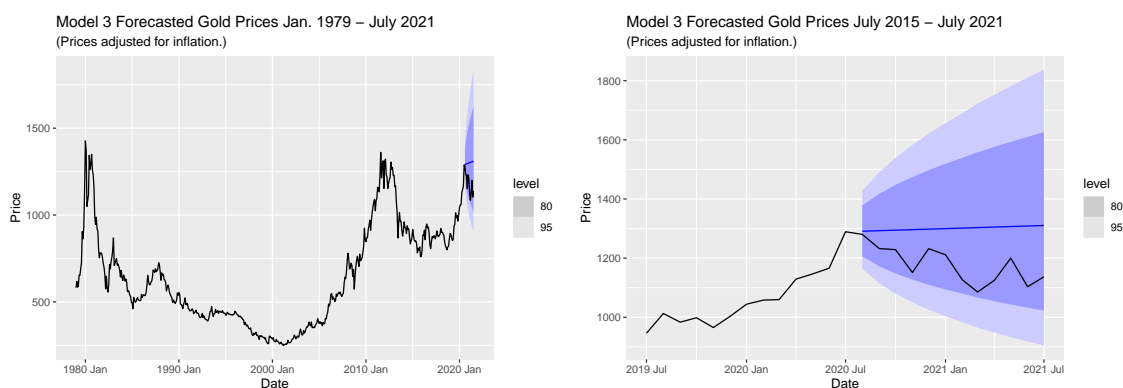
Since both models are reasonable for forecasting, we decided to try both. At this point, we brought in the data from our test set that had previously remained unused. For both model 3 and 1 we used the model

fitted on the training set to then forecast a year's worth of inflated gold prices, which is the same amount of data we removed and placed in the test set. We then compared those forecasted prices to the actual gold prices we stored in the test set. Our evaluation of the models will be based on the calculated root mean square error we find.

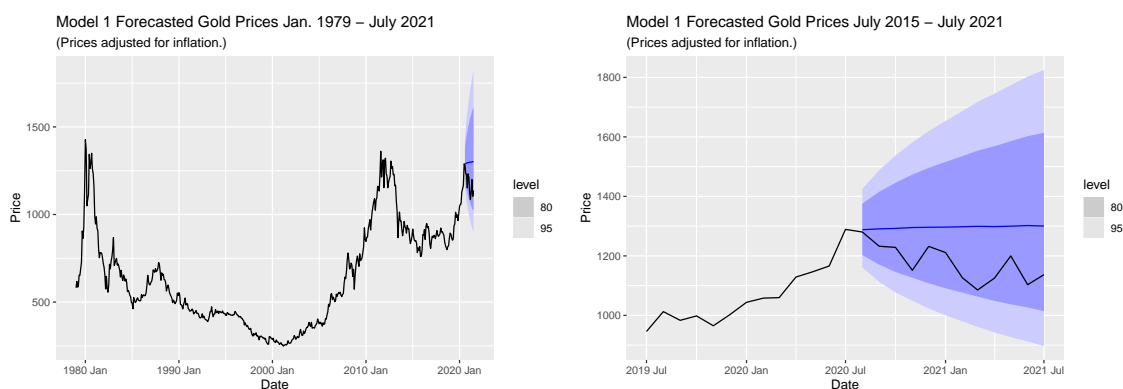
Forecasting our Test Set

The first plot shown visualizes each model's forecasted gold prices for the time period Jan. 1979 - July 2021 and their respective 80% and 95% prediction intervals in blue; also included in black is the true gold prices for that period. The second plot shows only the last two years of data, to give a clearer picture of the forecast's comparison to the actual gold prices.

Model 3



Model 1



Best Model on the Basis of RMSE

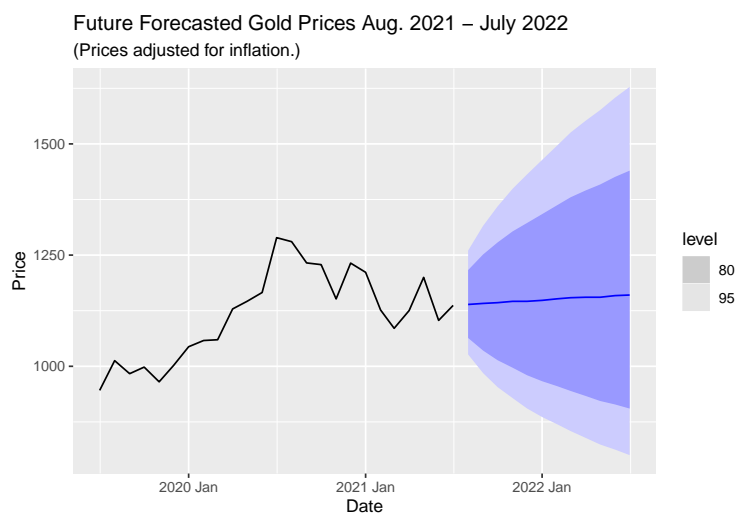
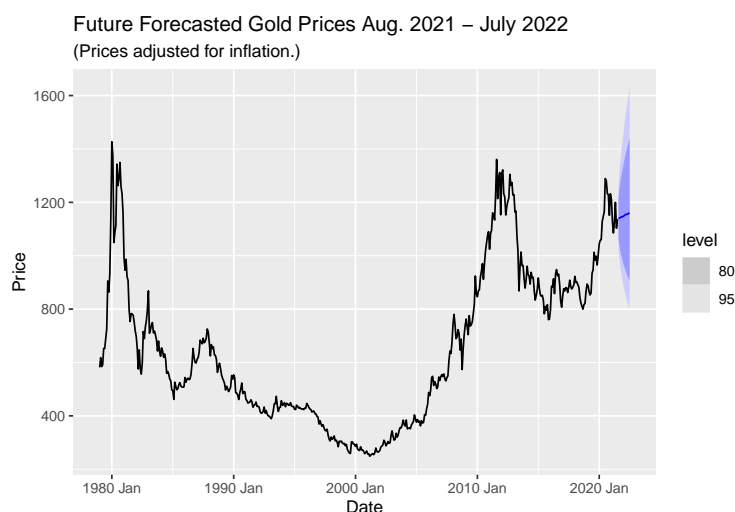
Model	RMSE
1: ARIMA(0,1,0)(1,0,1)[12]	135.715
3: ARIMA(0,1,0)(0,0,0)[12]	140.031

We saw that contrary to what we thought would happen, Model 1 actually has a lower root mean squared error than Model 3, even though on the basis of AICc, Model 3 should've performed the best. So, we determined that using the model we suggested, an $ARIMA(0,0,0) \times (1,0,1)_{12}$, is best (on the basis of RMSE) to model and forecast future inflation adjusted U.S. gold prices. So, Model 1 had lower standard deviation in the prediction errors than Model 3 by \$4.32.

Forecasting Future Gold Prices

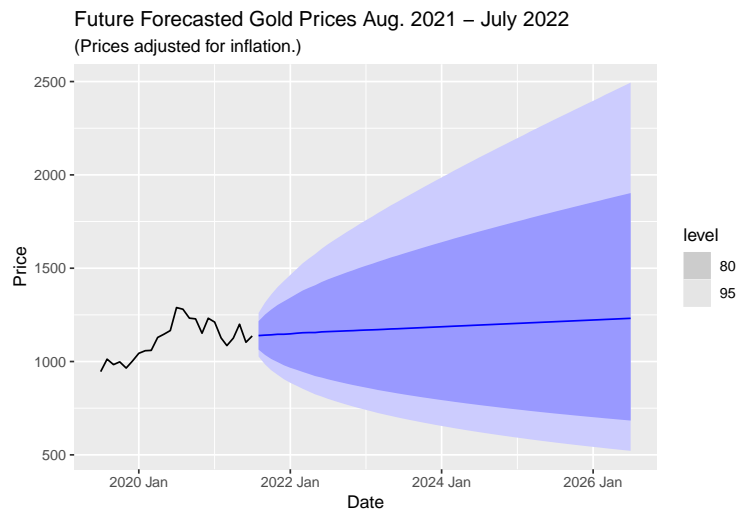
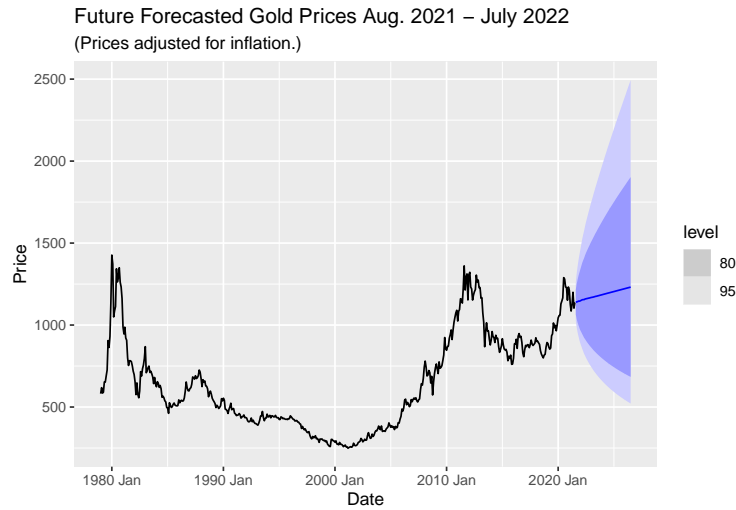
The last part of our project was to predict future inflation adjusted gold prices for the U.S. using our best model. We refit our optimal model (Model 1) on the entire data set instead of just the training set as we did before. Then we forecasted a year into the future (August 2021 - July 2022).

One Year in Future



While the actual forecasted values at each month for the next year do not differ much, the confidence bands for the prediction interval would be useful. Look at the forecasting plot, our 80% PI has approximately a \$500 range at its largest, and the 95% ranges by about at most \$800.

Five Years in Future



Notice when we try to predict a larger amount of 60 months (5 years), our prediction intervals are much larger. The 95% PI has nearly a \$2,000 maximum range and is much less useful, if we want to examine possible gold prices.

Conclusion

From among the three models we considered, the $ARIMA(0, 1, 0) \times (1, 0, 1)_{12}$ was lowest in terms of root mean squared error on the test set. This contradicted our finding that the random walk model $ARIMA(0, 1, 0)$ was best on the basis of AICc. The chosen model thus gave us the best predictive accuracy versus the other two models. We can state with 95% confidence that over the following year, gold prices will be in an \$800 range around the price in July 2021. However, the five year forecast proved to be not so useful due to the wide range of the confidence bands. This information might be of use to economists and investors alike.