Predictive Analytics Assignment 2

Steven VanOmmeren*

April 10, 2025

1 Introduction

I chose to analyze Monthly Wholesale Trade (Sales and Inventories) for Chemicals and Allied Products in the United States, using data from the U.S. Census Bureau's Business and Industry reports.[1] Chemicals and allied products are inputs to many products in practically all industries, for example petrochemicals used to create synthetic textile fibers for clothing. Predicting wholesale trade of chemicals would be of interest to several parties, including domestic/international chemical manufacturers, clothing producers, and regulators.

To forecast wholesale trade of chemicals, I used the Dow Jones US Chemicals Index as a predictor.[2] The index aggregates the stock prices of major US chemical manufacturers, whose price should reflect both current and future market outlook for wholesale trade. I collected monthly data on the index price, open/high/low prices, and volume traded.

I chose the time period of Jan. 2006 to Dec. 2012 to study wholesale trade of chemicals.[3] Notably, this period includes the Great Recession, which seriously impacted the time series. I deliberately chose my training period (Jan. 2006 to Dec. 2011) to include the Great Recession to see how various models can forecast future behavior following an extreme anomaly

*A complete replication package of this project is available at `https://github.com/svanomm/forecasting-assignment-2`.

[1] `https://www.census.gov/econ/currentdata/?programCode=MWTS&startYear=1992&endYear=2025&categories[]=4246&dataType=SM&geoLevel=US&adjusted=0&notAdjusted=1&errorData=0#line095`

[2] `https://www.investing.com/indices/dj-chemicals-historical-data`

[3] I included more than 5 years of data to ensure that sufficient data is included that was not tainted by the Great Recession.

in the training data. This yields 84 months of data (72 months of training, 12 months of testing).

2 Methods

I downloaded the time series data from FRED and loaded it into R. Graphing the time series, I did not see any indication of missing data or extreme outliers.

The time series for wholesale trade is visually autocorrelated and non-stationary. I confirmed this result using a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) statistic. To remove this effect, I took the first difference of wholesale trade to get the monthly change. The first-differenced time series appears to be fairly stationary and centered around 0, and this is confirmed by another KPSS test.

Besides the Dow Jones index as a control variable, I also created an indicator variable for the Great Recession, which equals 1 from Dec. 2007 to Jun. 2009, and 0 otherwise.[4]

After combining the time series variables and formatting the appropriate date columns, I split the data into training and test subsets, with the first 72 months used as training data.

Before running all of my models of interest, I tested a variety of ARIMAX models to determine which regressors should be included. I ran a total of 11 ARIMA models on the training data which varied the Dow Index variables:

[4]These dates come from `https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath`.

Model	**Description**
0	No exogenous variables
0_1	Great Recession Flag
1	Great Recession Flag, Volume
2	Great Recession Flag, Volume, Price
3	Great Recession Flag, Volume, Price, Open
4	Great Recession Flag, Volume, Price, Open, High
5	Great Recession Flag, Volume, Price, Open, High, Low
6	Great Recession Flag, Volume, Lag(Volume, 1 month)
7	Great Recession Flag, Volume, Lag(Volume, 1-2 months)
8	Great Recession Flag, Volume, Lag(Volume, 1-3 months)
9	Great Recession Flag, Volume, Lag(Volume, 1-4 months)

All models had their PDQ parameters selected automatically using the Akaike Information Criterion.

After selecting the best model from the ARIMA results above, I used the training data to run 5 models of wholesale trade:

1. ETS: an exponential smoothing model with additive trend and additive seasonality (Holt-Winters)

2. ARIMA: does not include any external regressors

3. ARIMAX: includes a Great Recession flag and Dow Index volume

4. ARIMAX Manual: a seasonal ARIMA(1,1,0)(1,1,0)[12] model with Dow Index volume as a regressor

5. Ensemble: a simple average of the above models.

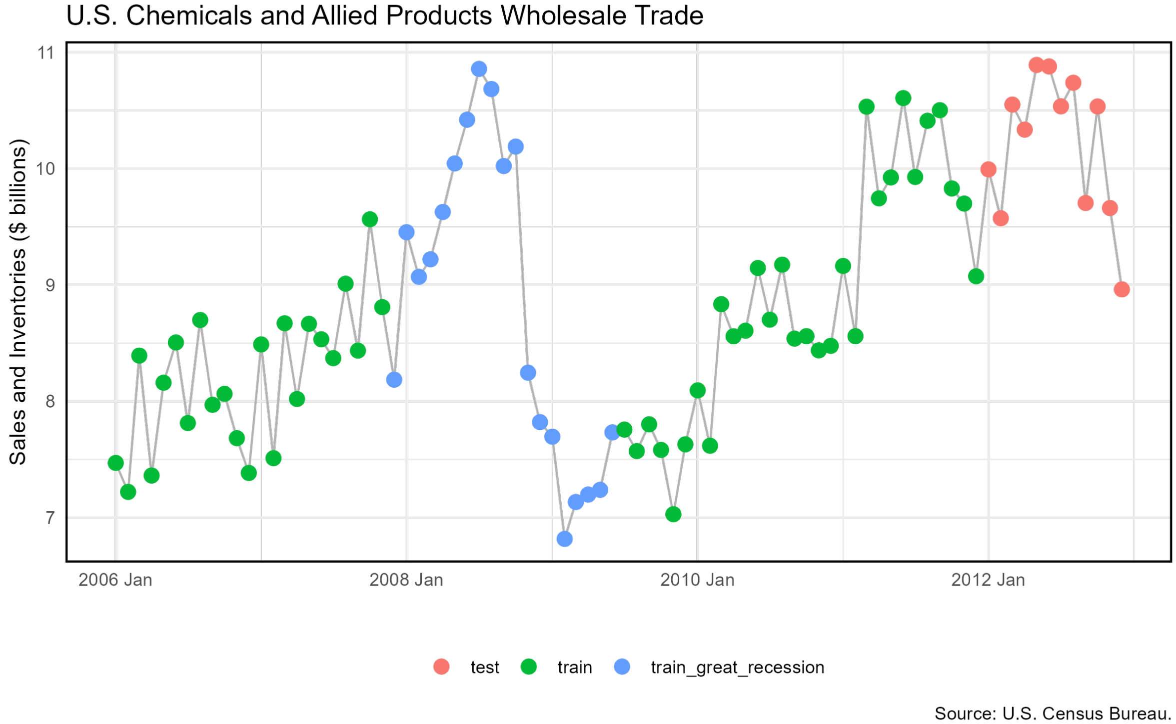
All models had their parameters selected automatically using the Akaike Information Criterion.

After running these models on the training data, I forecasted the next 12 months of trade volume (Jan.-Dec. 2012) and compared the forecasts to the actual values in the test data.

3 Results

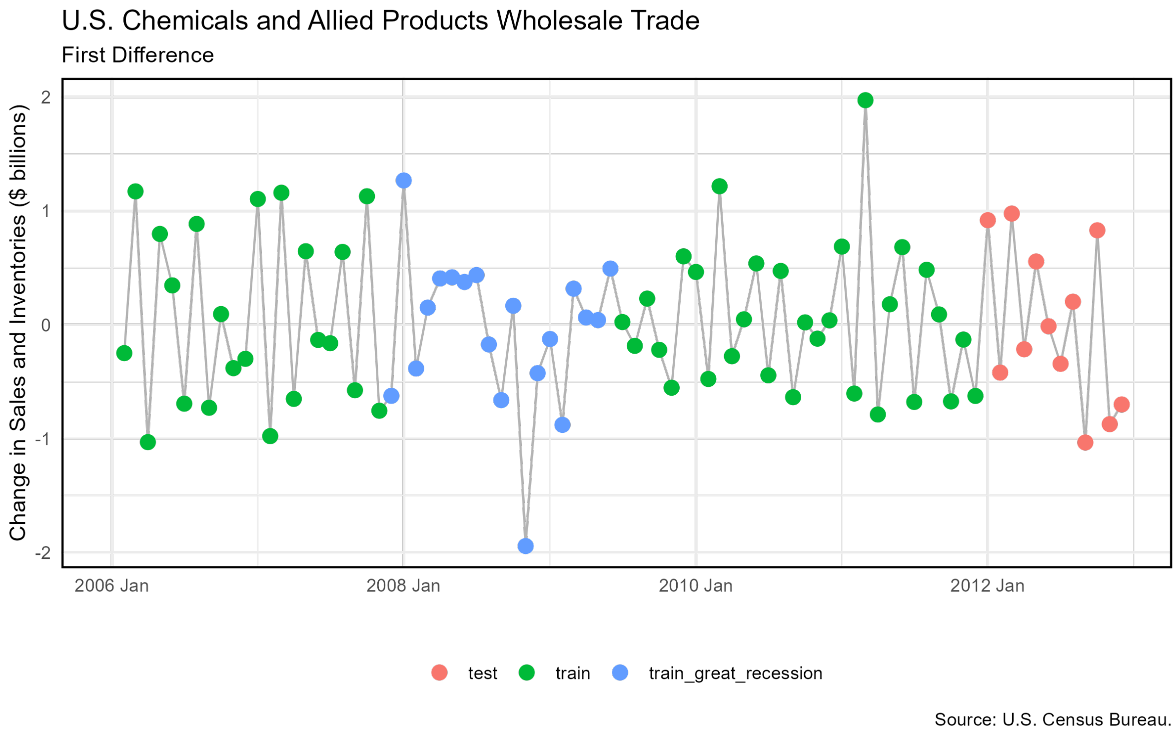
The time series used in my analysis is shown below. I separated wholesale trade into three groups: the testing period, the training period, and the part of the training period affected by the Great Recession. As we can see, the Great Recession at first dramatically increased wholesale trade and inventories, but then plummeted by 2009. 2010 sees a sharp increase, followed by another sharp increase in 2011. The time series is highly non-stationary, exhibiting multiple level changes. Overall, this presents a very difficult forecasting problem for my models of interest.

Figure 1: Time Series of Interest



However, after taking the first difference of wholesale trade, we can see that the data is mostly stationary.

Figure 2: First Difference Plot



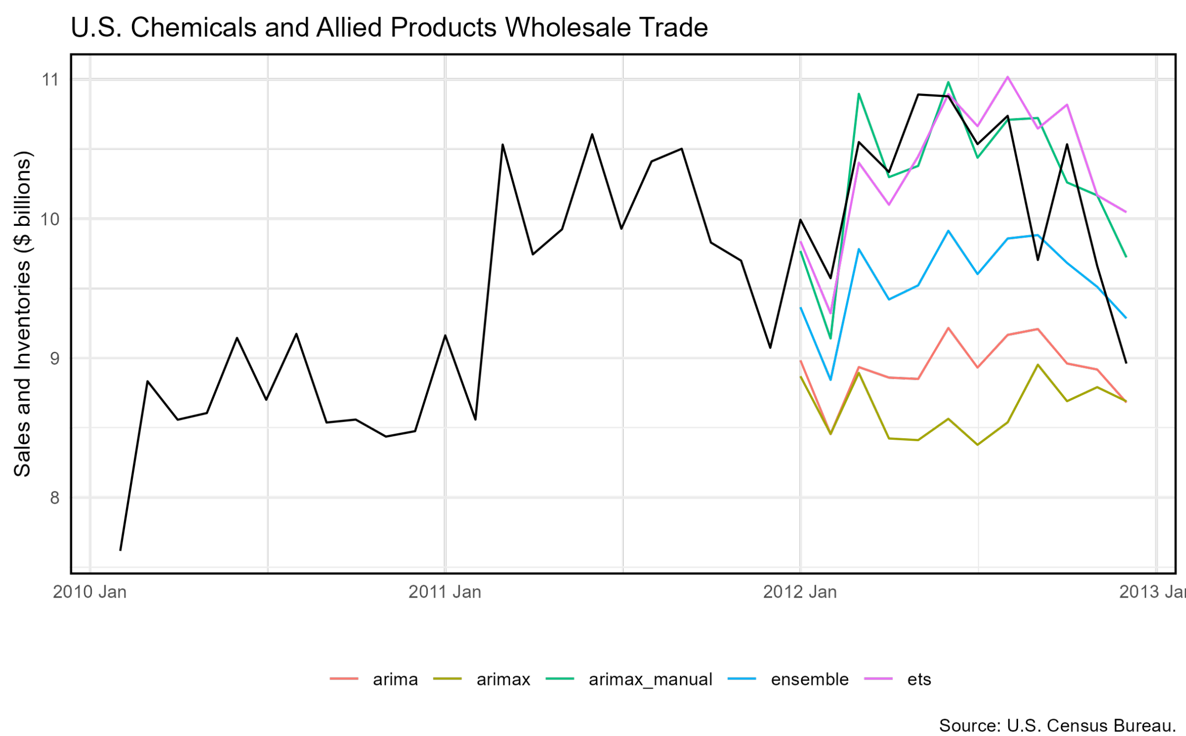
KPSS statistics confirm the visual results above. The KPSS test statistic p-value is 0.01 for wholesale trade, indicating a non-stationary time series with $\alpha = 0.05$. However, the KPSS test statistic for the first-difference time series is 0.1, so we cannot reject the null hypothesis of stationary. Taking one difference is sufficient to make the time series reasonably stationary.

After running my ARIMA model selection tests, I found that the AIC was lowest for model 9 (which included 4 lags of the Volume regressor). However, the AICc (which corrects for small sample sizes) was approximately equal to much simpler models which did not include lagged regressors. I selected model 1 (Great Recession Flag, Volume) because its AICc is tied at the best value, and it is simple with only 2 regressors.

When forecasting the automatically selected ARIMA model, I found that the out-of-sample fit was extremely poor. Reviewing the model parameters, the R package's automated selection criterion chose an ARIMA(4,0,0)(0,0,1)[12] model, meaning 4 lagged terms, and 1 seasonal moving average term. It did not include a differencing term, which I found to be at odds with my visual analysis of the data. To address this concern, I manually specified an ARIMAX(1,1,0)(1,1,0)[12] model, i.e. forcing the model to include a first-difference term. As we can see below, this model far exceeded the automated selection in out-of-sample performance.

To summarize additional results from my model forecasts:

- The automatically-selected ARIMA models performed extremely poorly because they did not include a first-difference term.
- The ETS model performed remarkably well in capturing the testing data behavior
- Both ETS and the Manual ARIMAX model are clearly the best predictors, and are approximately equal in performance



To quantify the accuracy of my models, I compared both the in- and out-of-sample accuracy metrics for each of them.

Table 1: Accuracy Metrics of Forecasts (In- and Out-of-Sample)

.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
ensemble	Training	-0.01	0.39	0.31	-0.25	3.56	0.24	0.26	0.02
arimax	Training	0.01	0.42	0.33	-0.09	3.76	0.25	0.28	0.00
ets	Training	-0.05	0.43	0.33	-0.79	3.92	0.26	0.29	-0.09
arima	Training	0.02	0.43	0.34	0.05	3.90	0.26	0.29	0.04
arimax_manual	Training	-0.01	0.51	0.38	-0.16	4.36	0.29	0.34	0.03
arimax_manual	Test	-0.09	0.46	0.36	-1.07	3.67			-0.09
ets	Test	-0.17	0.49	0.37	-1.82	3.81			0.50
ensemble	Test	0.64	0.80	0.72	6.06	6.97			0.23
arima	Test	1.27	1.37	1.27	12.17	12.17			0.21
arimax	Test	1.56	1.70	1.56	14.94	14.94			0.36

4 Discussion

Overall, my manually specified ARIMAX model performed slightly better than the ETS model, and the other models are nowhere near as good of predictors. This is best seen in the accuracy metrics table, where the out-of-sample RMSE is 0.46 and 0.49 for Manual ARIMAX and ETS, respectively, and the next best RMSE is 0.8 for the Ensemble model. The automatically selected ARIMAX model has an RMSE of 1.7, or nearly 4 times as large as the Manual ARIMAX model. The Manual ARIMAX performs best on other out-of-sample measures of fit, including MAE and MAPE, though only just beating the ETS model's performance.

As discussed above, the automatically selected ARIMA models performed terribly because they did not include a difference term. The Manual ARIMAX model includes a differencing term, which makes the data stationary and thus easier to forecast.

Examining the ETS Model, it selected $\alpha = 0.764$, $\beta = 0.0001$, and $\gamma = 0.0002$. The large value of α reflects that the model quickly updates to level changes, while the overall trend (β) and seasonality patterns (γ) remain constant throughout the data. The seasonal factors and quick level changes explain why the ETS model performed well, and we can confirm this by seeing that the last year of training data has a similar level to the testing data. The ETS model in this case is nearly approximating a Seasonal Naïve model, which happens to be close to reality.

With the Manual ARIMAX model, the coefficient on Dow Index volume is 0.0925, with a standard error of 0.0624. This gives a t-statistic of approximately 1.5, which is not statistically significant at standard confidence levels ($p \approx 0.14$). However, the point estimate finds a positive relationship between wholesale trade volume and stock trading volumes of US chemical companies.

Because I selected my Manual ARIMA model after the first round of forecasting evaluation, one could argue that I engaged in "p-hacking," choosing a model that would knowingly perform better on the testing data. However, I would argue that the KPSS statistics show that the time series should be differenced, and my specification of the model is not unreasonable. I include first-difference and first-seasonal-difference terms, and order-1 lag and order-1 seasonal lag terms. This does not suggest overfitting the data. In any case, my manually specified model performed approximately as well as the ETS model, which I did not modify after my first round of forecasting evaluation.

To extend the usefulness of my models, I would consider additional variables that could influence the wholesale trade of chemicals. Controlling for chemical demand from producers, and supply factors from chemical manufacturers would likely improve my forecasts. I would modify the models to include more historical data, since the U.S. Census Bureau has data going further back than my period of interest. Lastly, I would experiment with lagged regressors in a first-difference model, something I did not try in this assignment.

5 Conclusion

The Great Recession had a sizeable impact on the wholesale trade of US chemicals. I tested five different time series models to see how well they could forecast trade volumes when the training data included the anomalous behavior of the Recession. I found that the automatically selected ARIMA models failed to include a differencing term, which would have made the time series stationary and easier to forecast. After manually specifying an ARIMA model with first-differences, I found that it substantially outperformed the automated models, and performed slightly better than an ETS model. The ETS model performed well because it weighed the most recent year of training data much more than the Recession period.

Code Appendix

A complete replication package of this project is available at `https://github.com/svanomm/forecasting-assignment-2`.