

Modeling Florida and Illinois Exports

Econ 144 - Project 2

Jacob Titcomb

Spring 2023

Contents

I. Introduction	2
II. Results	3
(a) Time Series Plots	3
(b) STL Decomposition	6
(c) Trend, Seasonality, and Cycles	7
(d) Residual Plots	9
(e) Residual ACF and PACF	10
(f) Cumulative Sum Plots	11
(g) Model Diagnostic Statistics	12
(h) 12-step-ahead Forecast	13
(i) Comparing to ARIMA Model	14
(j) Combining forecasts	16
(k) VAR model	17
(l) IRF	20
(m) Granger causality	21
(n) Forecasts Using VAR Model	22
III. Conclusion and Future Work	23
IV. References	24
V. Appendix	25

I. Introduction

Our project attempts to model and forecast the exports of goods for Florida and Illinois. More specifically, according to the St. Louis Federal Reserve, these statistics measure the “[e]xports of manufactured and non-manufactured commodities based on origin of movement.” Measuring exports for an economy gives insight into its health and productivity, especially since foreign trade is taken into direct consideration when calculating gross domestic product (GDP). Modeling and predicting export levels of a state has many benefits, such as informing state and federal policy, and assessing the economic climate of an area for the decisions of both domestic and foreign firms. In particular, forecasting the exports for Illinois and Florida would provide insight into the future economies of both states, as well as the economy of the United States at large.

We selected Florida and Illinois as the two American states of interest since they are among the American states with the highest GDPs in 2022, ranking 4th and 5th, respectively (Bureau of Economic Analysis). As far as exports, they both have a similar level of exports, with Illinois being the 5th largest exporter out of all the states and Florida being the 6th.

The largest proportion of Florida exports from March 2023 were in a variety of industries: aircraft parts, medicaments, smartphones, and integrated circuits (OEC). The most common destinations for Florida’s exports were Canada, Brazil, Mexico, and the United Kingdom (OEC).

Similarly for Illinois, their most common exports from March 2023 were in varied industries: petroleum, medicaments, antisera (blood serum containing antibodies), and soybean (OEC). Illinois’ exports were most often sent to Canada, Mexico, China, and Germany (OEC).

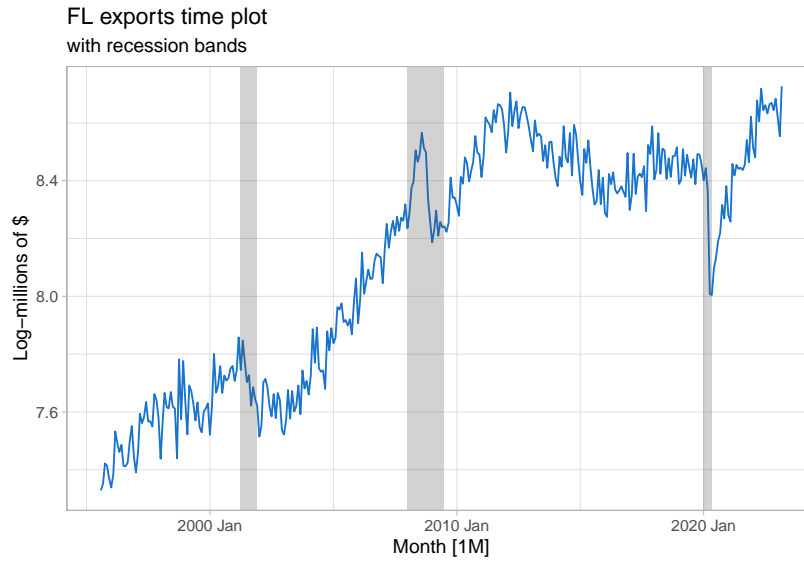
The export statistics for Florida and Illinois cover the time period from August 1995 to March 2023, recorded monthly. The data was sourced from the Federal Reserve Bank of St. Louis, and we performed an initial transformation of taking the logarithm, hopefully reducing variation in amplitude for both series. Thus the units for both time series is log-millions of American dollars.

II. Results

(a) Time Series Plots

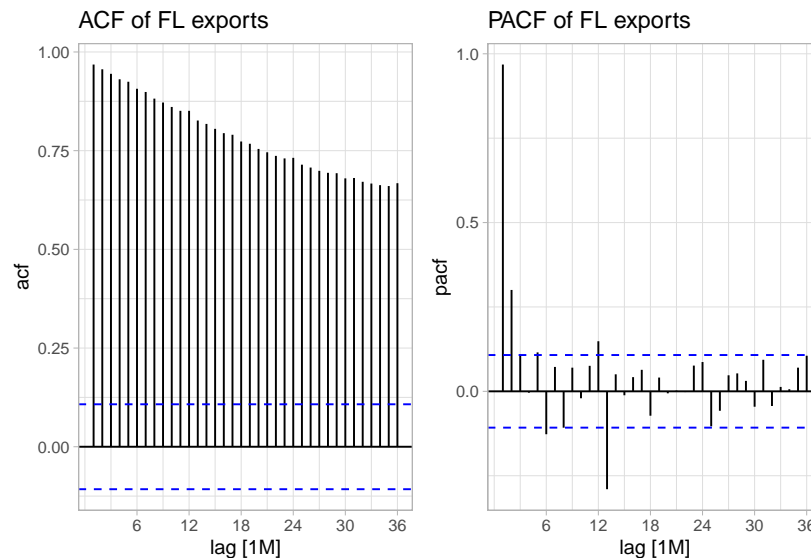
Florida

In the time plot below, we see a general upward trend to the data, with a possible seasonal component, though difficult to verify without further analysis. In addition to trend, the time series shows some stochastic cyclic behavior. With respect to ARMA models, the time series exhibits traits of an AR process, like possible white noise behavior and strong persistence; we will study this possibility later.



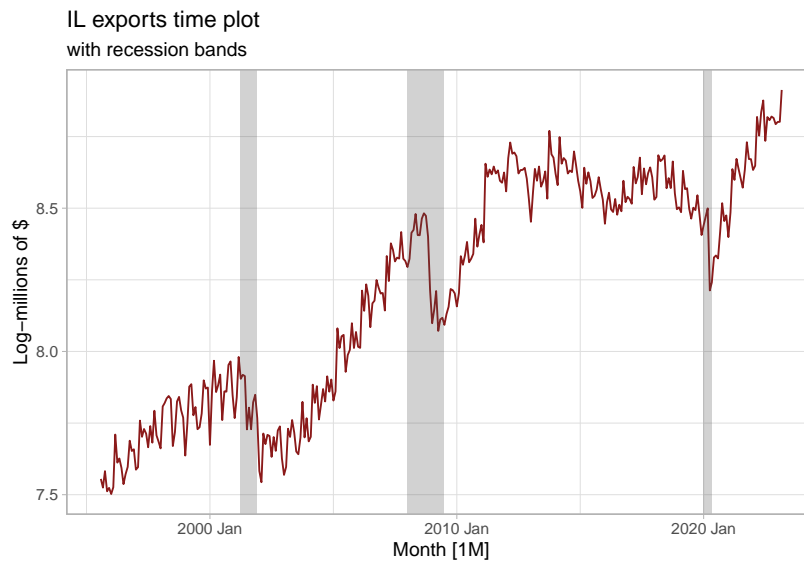
As far as the ACF and PACF for the data, we see gradual decay in the ACF with very slight spikes at every 12 lags, indicating a both seasonal and non-seasonal AR process. We note that the slow decay in the ACF confirms the strong persistence seen in the time plot.

For the PACF, we see significant spikes at lags 1, 2, and 13, challenging the seasonal AR theory. While one might initially guess an AR(13) process, we suspect that differencing (possibly of the first order) will likely be incorporated if we opt for an ARIMA model; thus any decision at this point would be hasty.

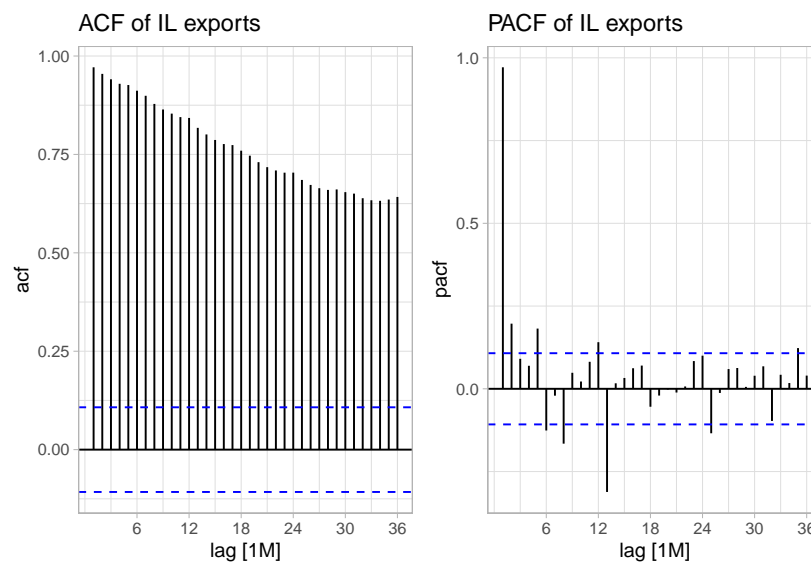


Illinois

The time plot for Illinois exports shows an increasing trend, with some possible seasonality (again, difficult to say definitively). Like Florida, Illinois shows some stochastic cyclic behavior. Strong persistence in the data indicates an underlying AR process, though more analysis must be done.

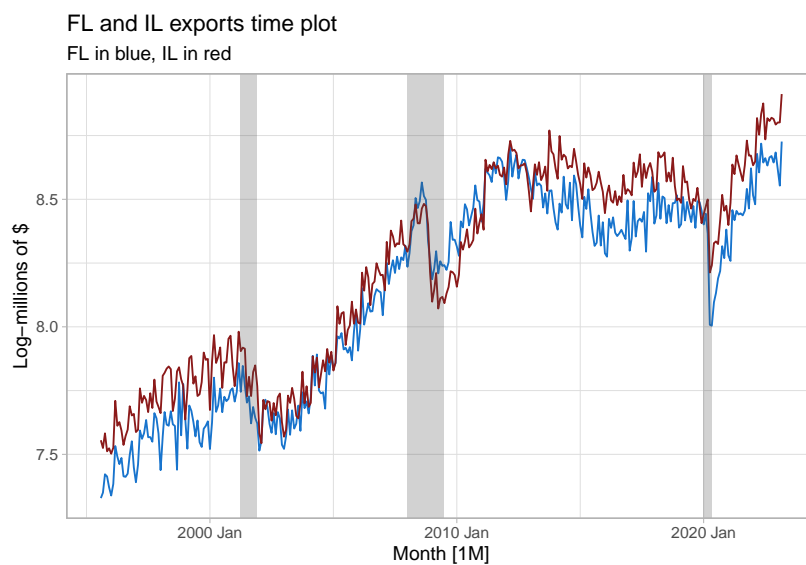


The ACF and PACF for Illinois appear to be very similar to Florida's. There appears to be decay in the ACF and slight scalloping every 12 lags in the ACF, indicating both an AR and seasonal AR process. The PACF has significant lags at 1 and 13; other spikes appear significant but are still close to zero, so they are not as conclusive and 1 and 13. The spikes at 1 and 13 indicate a possible AR(13) process at this point, but it is likely that including a differencing component would be beneficial.



Florida and Illinois Together

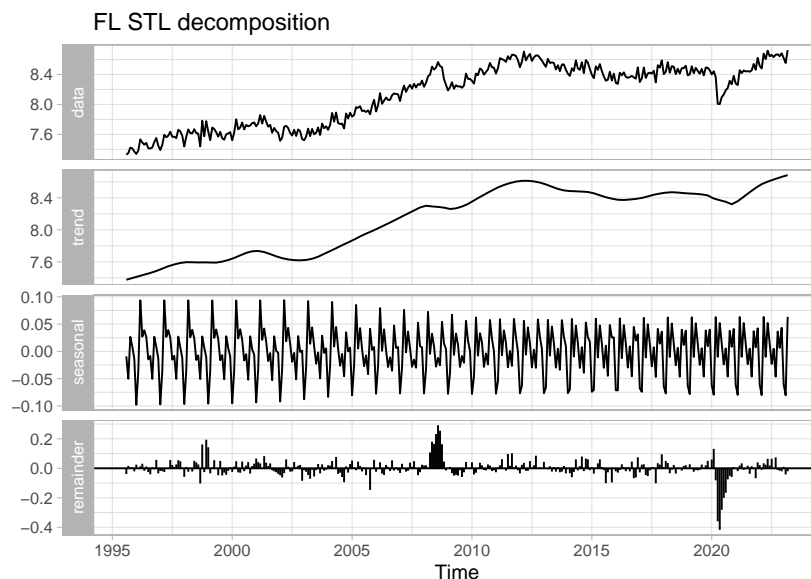
For completeness, we will graphed both time series on the same plot below. We see that they tend to follow each other fairly well.



(b) STL Decomposition

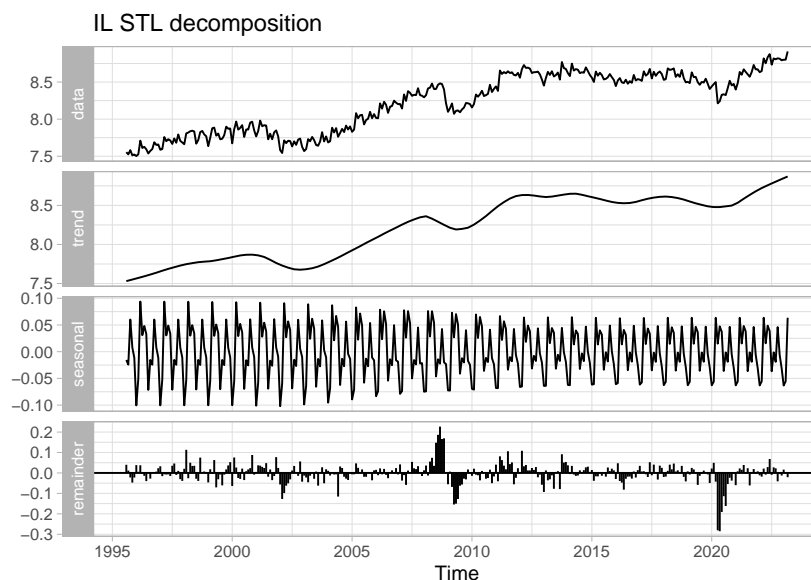
Florida

We observe the positive trend component to the data, agreeing with our description of the trend from earlier. The seasonal component shows strong seasonality which changes slightly over time, decreasing in amplitude. There does seem to be a slight pattern to the remainder component, indicating possible cycles. There are clear deviations in the residuals around 2008 and 2021, understandable given the economic climates at the time.



Illinois

The behavior for Illinois appears to be very similar to Florida, with an upward trend, a seasonal component with a decreasing magnitude, and some slight patterns in the remainder, indicating cycles. The periods of 2008 and 2021 also show large remainder components as well.



(c) Trend, Seasonality, and Cycles

Florida

We propose the following model:

$$Y_t = T_t^{Loess} + \sum_{i=1}^{12} \delta_i M_{i,t} + R_t \quad \text{Trend and Seasonality}$$

$$R_t = \phi R_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \quad \text{Cycles}$$

The first term (T_t^{Loess}) is the Loess fitted model for trend, found using an STL decomposition. The second component of the model captures seasonality using 12 seasonal dummy variables, one for each month. The parameters for the seasonality were determined also with in the STL decomposition. We note that we designed the seasonality to be fixed for the entirety of the data, i.e. the window for the seasonality was set to "periodic" so the values for each month do not fluctuate over time. Lastly, we selected an ARMA(1, 1) model for the cycles, represented with the equation with R_t . We chose the order of AR(1) and MA(1) using the `auto.arima()` function, which we also used to determine the estimates for ϕ and θ .

We are unable to extract the explicit parameters used for the Loess trend component, but we can extract the seasonality and cycle components. They are summarized in the following two tables:

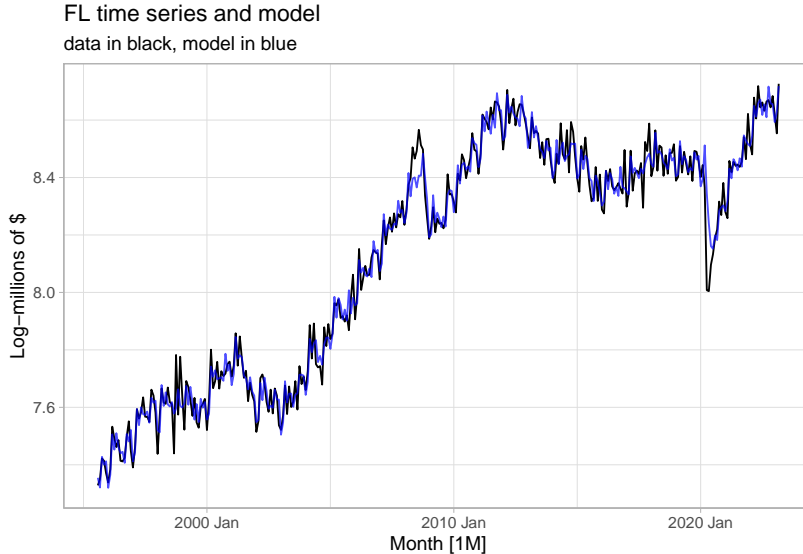
Table 1: FL Seasonal Component

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-0.075	-0.051	0.074	0.003	0.043	0.017	-0.021	-0.007	-0.039	0.041	0.008	0.006

Table 2: FL Cycle Component

	AR(1)	MA(1)
Coefficients	0.655	-0.219

To illustrate the fit of the model, the following graph shows the original time series (black curve) and the model (blue curve).



Illinois

We propose a similar model for Illinois:

$$Y_t = T_t^{Loess} + \sum_{i=1}^{12} \delta_i M_{i,t} + R_t \quad \text{Trend and Seasonality}$$

$$R_t = \phi R_{t-1} + \varepsilon_t \quad \text{Cycles}$$

Again, the trend and seasonality were determined using an STL decomposition, with seasonality being fixed for the entirety of the time series. The biggest difference is in the cycles: we decided to use an AR(1) model rather than ARMA model since `auto.arima()` determined a first-order AR process to be the most ideal. The explicit parameters for the seasonality and cycles are summarized in the following two tables:

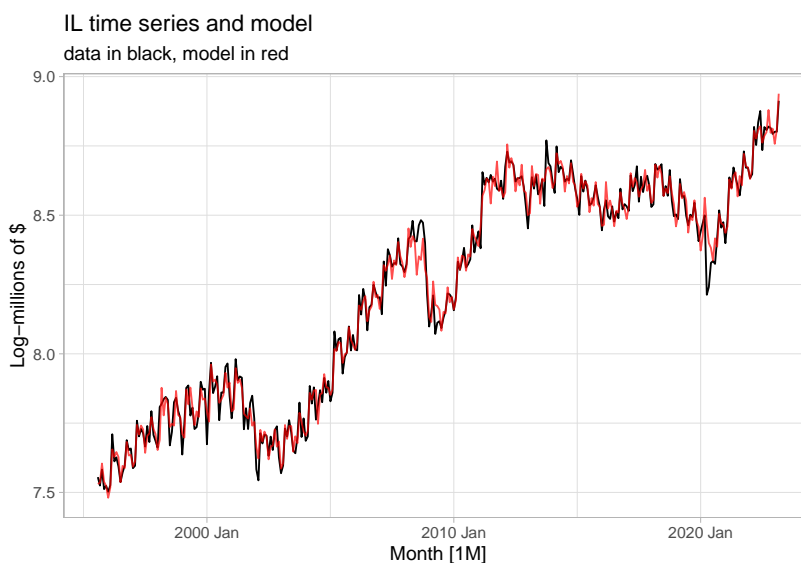
Table 3: IL Seasonal Component

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-0.077	-0.058	0.074	0.022	0.053	0.037	-0.05	-0.009	-0.019	0.055	-0.006	-0.02

Table 4: IL Cycle Component

AR(1)	
Coefficients	0.4187

To illustrate the fit of the model, the following graph shows the original time series (black curve) and the model (red curve).



(d) Residual Plots

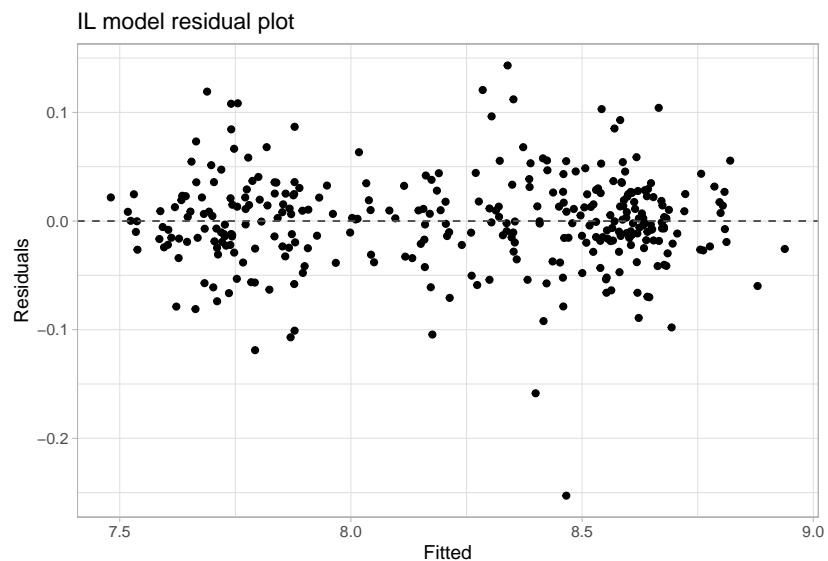
Florida

In the below residual plot, we see fairly constant scattering, with two or so outliers very far from 0. The variance in the residuals seems to decrease a bit on the ends, which could indicate an issue with the model's design. There are no obvious patterns, so our model is relatively successful at capturing the general behavior of the time series.



Illinois

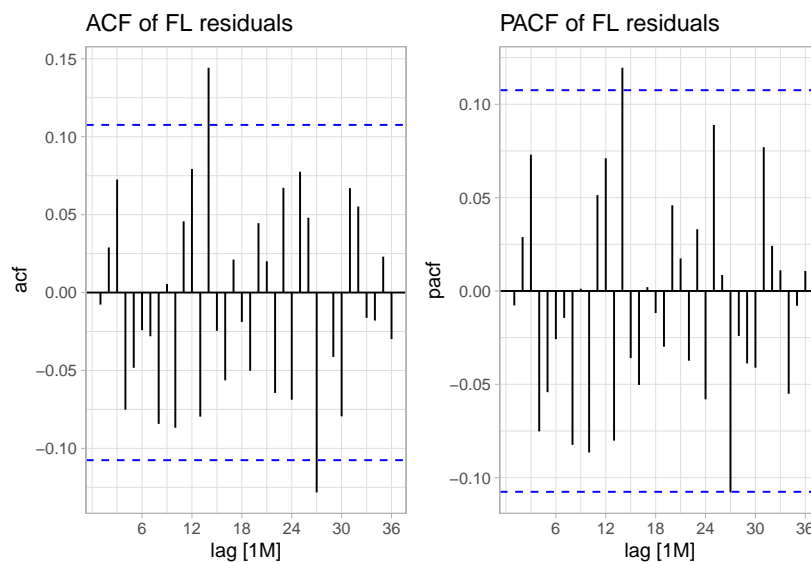
The residual plot below looks very similar to that of Florida's, except Florida's has a slightly larger range on the y -axis. Again, the residuals show fairly constant scattering, with only one outlier being very far from 0. The variance in the residuals decreases towards the ends, indicating a possible error in the model; however, the lack of an obvious pattern indicates our model captured the general behavior of the time series.



(e) Residual ACF and PACF

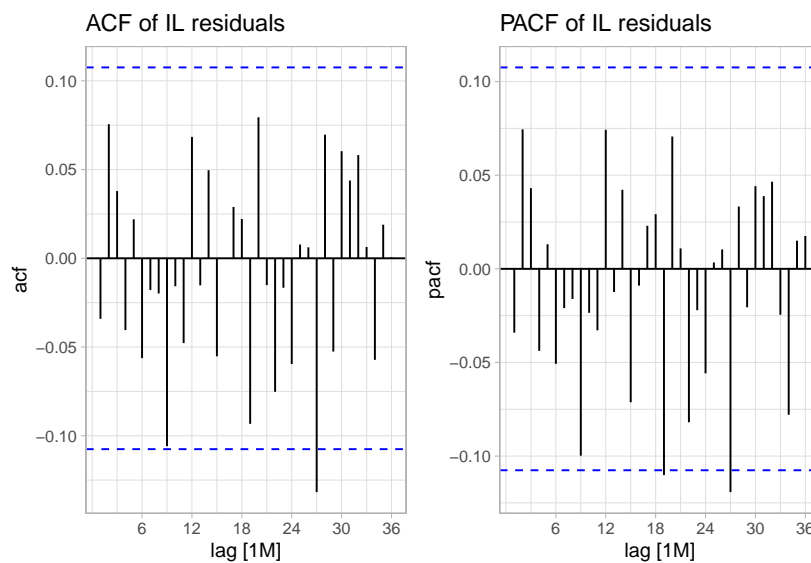
Florida

In the below ACF and PACF, we see close very few statistically significant spikes. We will verify with a statistical test in part (g) if the residuals follow a white noise process.



Illinois

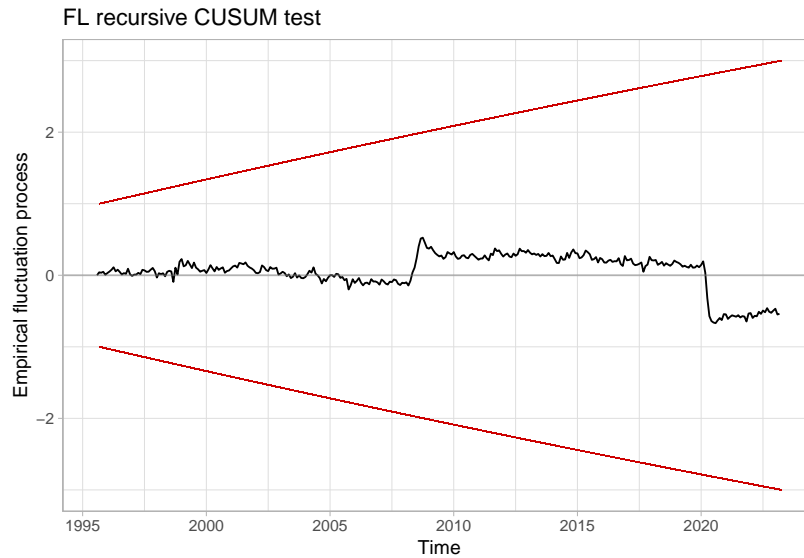
The ACF and PACF show very few statistically significant spikes. We suspect that a formal statistical test – performed in part (g) – will confirm that the spikes that appear to be significant are actually not significant.



(f) Cumulative Sum Plots

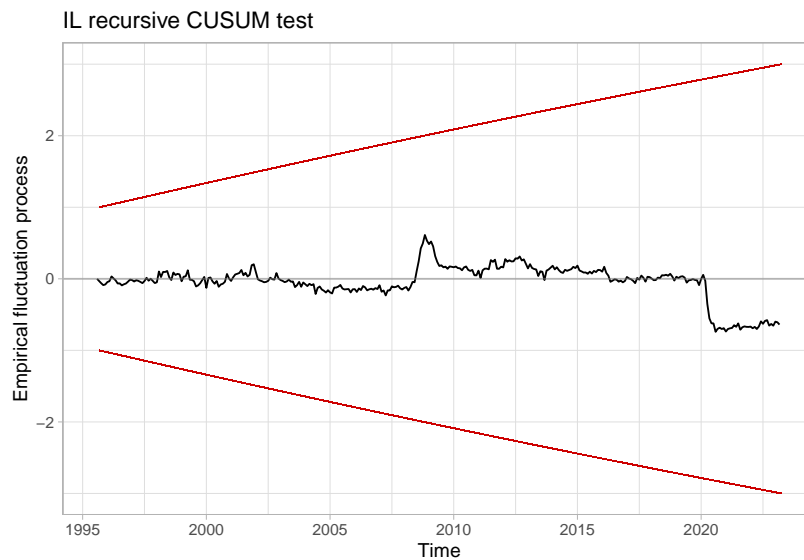
Florida

In the CUSUM plot for Florida, we see very little divergence from 0. In fact, the cumulative sum stays well within the red bands, indicating that there are no structural breaks in the model.



Illinois

In the CUSUM plot for Illinois, we again see very little divergence from 0. The cumulative sum stays well within the red bands, indicating that there are no structural breaks in the model.



(g) Model Diagnostic Statistics

Before discussing diagnostics for the models, we note that we do not include the F statistic (and thus no test of overall significance) and adjusted R^2 for our models. That is because the component for trend was calculated implicitly in an STL decomposition rather than explicitly in an `lm` object. Thus we are unable to extract the relevant details about the trend (e.g. degrees of freedom) and so both measures – F statistic and \bar{R}^2 – cannot be calculated at this point.

Florida

```
## MAPE: 0.4641861
## RMSE: 0.05173537
## ME: 0.0001243753

##
## Box-Ljung test
##
## data: FL_resid_tsib$resid
## X-squared = 13.171, df = 12, p-value = 0.3568
```

The model for Florida has a MAPE of 0.464, which is below 50%; while we would prefer a MAPE under 25%, below 50% is not terrible for a model. The model's RMSE is 0.0517 (in log-millions of dollars) – fairly low compared to the scale of the time series (a range from about 7 to 9). It is worth noting that the mean error is 0.0001243, so it is close to but not exactly 0, meaning that our model slightly underestimates the data.

To further assess model appropriateness, a Ljung-Box test using 12 lags was performed. The test had a p -value of 0.3568; so at the 0.05 significance level we fail to reject the null hypothesis, meaning the residuals follow a white noise process. Therefore our model successfully captured the serial correlation within the time series.

Illinois

```
## MAPE: 0.376174
## RMSE: 0.04300095
## ME: -0.0004791815

##
## Box-Ljung test
##
## data: IL_resid_tsib$resid
## X-squared = 11.172, df = 12, p-value = 0.5143
```

The model for Illinois has a MAPE of 0.376, which is farther below 50% than Florida's model; again, a MAPE below 50% is not terrible for a model. The model's RMSE is 0.0430 (in log-millions of dollars) – fairly low compared to the scale of the time series (a range from about 7 to 9). It is worth noting that the mean error is -0.000479. It is close enough to 0 that it may be rounding error, but if it is not, it indicates that our model slightly overestimates the data.

To assess serial correlation, a Ljung-Box test using 12 lags was performed. The test had a p -value of 0.5143; so at the 0.05 significance level we fail to reject the null hypothesis, meaning the residuals follow a white noise process. Thus our model successfully captured the serial correlation within the time series.

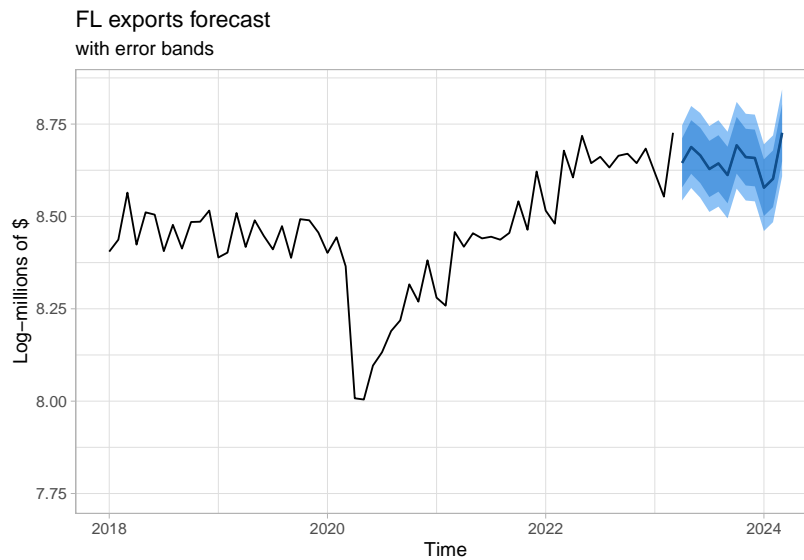
(h) 12-step-ahead Forecast

Since we used an ARMA model to model the remainder component of the STL decomposition, the residuals of the combined model will be the residuals from the ARMA model. Thus the widths of the prediction interval for the ARMA model will be the widths of the prediction interval for the combined model.

The forecasts of the STL decompositions used the naïve forecasting method, and the ARMA components were forecasted normally.

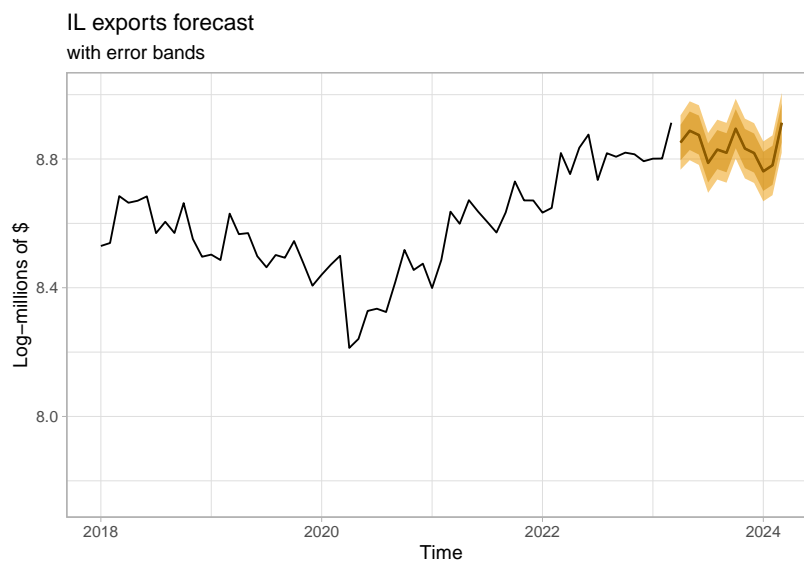
Florida

The below plot has the 12-step-ahead forecast with error bands, starting in April 2023 and ending in March 2024.



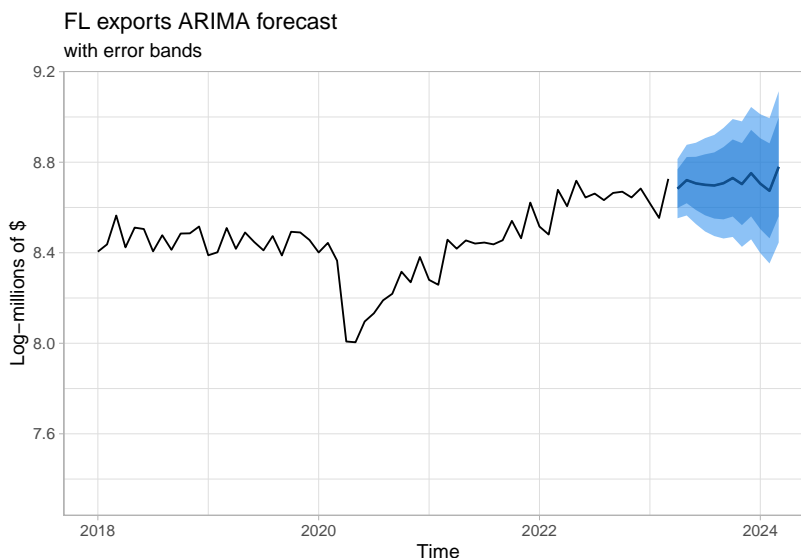
Illinois

The below plot has the 12-step-ahead forecast with error bands, starting in April 2023 and ending in March 2024.



(i) Comparing to ARIMA Model

Florida



Above, we have a 12-steps-ahead forecast for Florida using an ARIMA model. Compared to the forecast from earlier, the ARIMA forecast shows larger error bands and less volatility in the 12-month forecasted window.

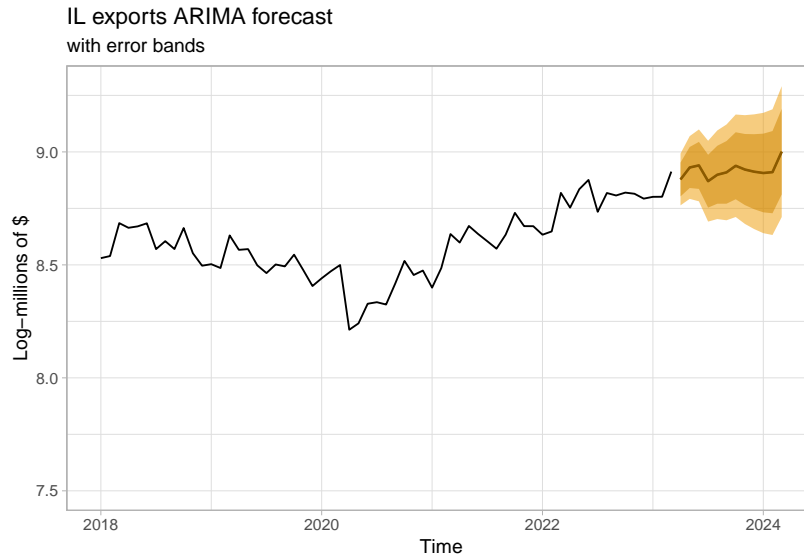
We used the `auto.arima()` function in R to estimate an ARIMA model. The summary of the model is in the code output below. The algorithm determined an ARIMA(2, 1, 1) model with a seasonal AR(2) component ($s = 12$), taking the following form:

$$\begin{aligned} \text{Trend and Cycles:} \quad & (1 + 1.115L + 0.360L^2)(1 - L)y_t = (1 + 0.768L)\varepsilon_t \\ \text{Seasonal:} \quad & (1 - 0.369L^{12} - 0.212L^{24})y_t = \varepsilon_t \end{aligned}$$

From the summary output below, we see the ARIMA model has a higher MAPE: $MAPE_{ARIMA} = 0.6172$ compared to $MAPE_{T+S+C} = 0.4642$. Further inspection of the other performance metrics (e.g. ME and RMSE) reveals that the ARIMA model's metrics have larger magnitudes than the original model in general. Thus our original model (STL decomposition with an ARMA component) performs better than the ARIMA model.

```
## Series: FL_ts
## ARIMA(2,1,1)(2,0,0)[12]
##
## Coefficients:
##          ar1          ar2          ma1          sar1          sar2
##         -1.1152    -0.3601    0.7676    0.3691    0.2125
## s.e.      0.1929     0.0582    0.2059    0.0536    0.0551
##
## sigma^2 = 0.004452:  log likelihood = 426.81
## AIC=-841.63   AICc=-841.37   BIC=-818.82
##
## Training set error measures:
##              ME          RMSE          MAE          MPE          MAPE          MASE
## Training set 0.002677038 0.06611566 0.04997481 0.02970582 0.6171616 0.4246484
##              ACF1
## Training set -0.02319715
```

Illinois



Comparing the above forecast using an ARIMA model to the forecast from earlier, this forecast shows larger error bands and less volatility in the 12-month forecasted window.

We used the `auto.arima()` function in R to estimate an ARIMA model. The summary of the model is in the code output below. The algorithm determined an ARIMA(2, 1, 0) model with a seasonal AR(2) component ($s = 12$), taking the following form:

$$\begin{aligned} \text{Trend and Cycles:} \quad & (1 + 0.314L + 0.108L^2) (1 - L) y_t = \varepsilon_t \\ \text{Seasonal:} \quad & (1 - 0.440L^{12} - 0.246L^{24}) y_t = \varepsilon_t \end{aligned}$$

From the summary output, we see the ARIMA model has a higher MAPE: $MAPE_{ARIMA} = 0.5325$ compared to $MAPE_{T+S+C} = 0.3762$. Further inspection of the other performance metrics (e.g. ME and RMSE) reveals that the ARIMA model's metrics have larger magnitudes than the original model in general. Thus our original model (STL decomposition with an ARMA component) performs better than the ARIMA model.

```
## Series: IL_ts
## ARIMA(2,1,0)(2,0,0)[12]
##
## Coefficients:
##          ar1      ar2      sar1      sar2
##        -0.3135 -0.1076  0.4400  0.2457
## s.e.    0.0546   0.0549  0.0526  0.0539
##
## sigma^2 = 0.003412: log likelihood = 469.18
## AIC=-928.35   AICc=-928.17   BIC=-909.34
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.002257185 0.05797028 0.043713 0.02505639 0.5324907 0.3661273
##              ACF1
## Training set 0.003255715
```

(j) Combining forecasts

Florida

```
## T+S+C model MAPE: 0.4641861
## ARIMA model MAPE: 0.6171616
## Cmb'd model MAPE: 0.4644734
```

We combined the two models (ARIMA model and STL decomposed with ARMA component model) by implementing a rudimentary weighting scheme with a linear regression. When comparing the three models, we observe that the original model (STL decomposed with ARMA component) still has the lowest MAPE. Notably, the combined model performs very similarly to the original model, far better than the ARIMA model.

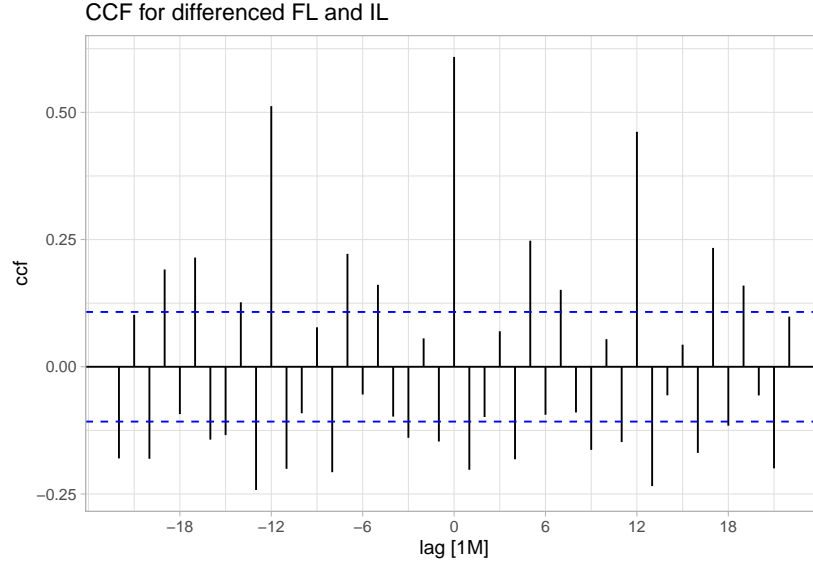
Illinois

```
## T+S+C model MAPE: 0.376174
## ARIMA model MAPE: 0.5324907
## Cmb'd model MAPE: 0.3771858
```

Similar to Florida, the original model – which incorporated an STL decomposition and an ARMA component – had the lowest MAPE. Then the combined model had the second lowest, and the ARIMA model had the highest MAPE. Thus the original model performed the best.

(k) VAR model

First, to make the data stationary, we took the first difference for both time series. We see in the cross correlation function below that the highest cross correlation is for contemporaneous observations. However, other lags appear to be significant too, especially spikes occurring on lags 12 and -12.



Next we will fit a VAR model to the differenced data. From using `VARselect()` to identify the order of a VAR model, the information criteria gave conflicting results. In this case, we will opt for the model with the lowest BIC: the model of order 4. The VAR model parameters can be seen in the table below; the first row is for the model with Florida exports as the endogenous variable, the second row has Illinois exports as the endogenous variable.

With Florida as the endogenous variable, lags 1, 2, and 4 of Florida are all significant, the first two at the 0.001 level and lag 4 at the 0.05 level. Illinois lags 1, 2, and 3 are significant, at levels 0.01, 0.05, and 0.1, respectively.

With Illinois as the endogenous variable, each lag of Illinois is significant at the 0.001 level. For Florida, only lag 3 is significant (at the 0.1 level).

Table 5: VAR lag-order criteria

AIC(n)	HQ(n)	SC(n)	FPE(n)
10	7	4	10

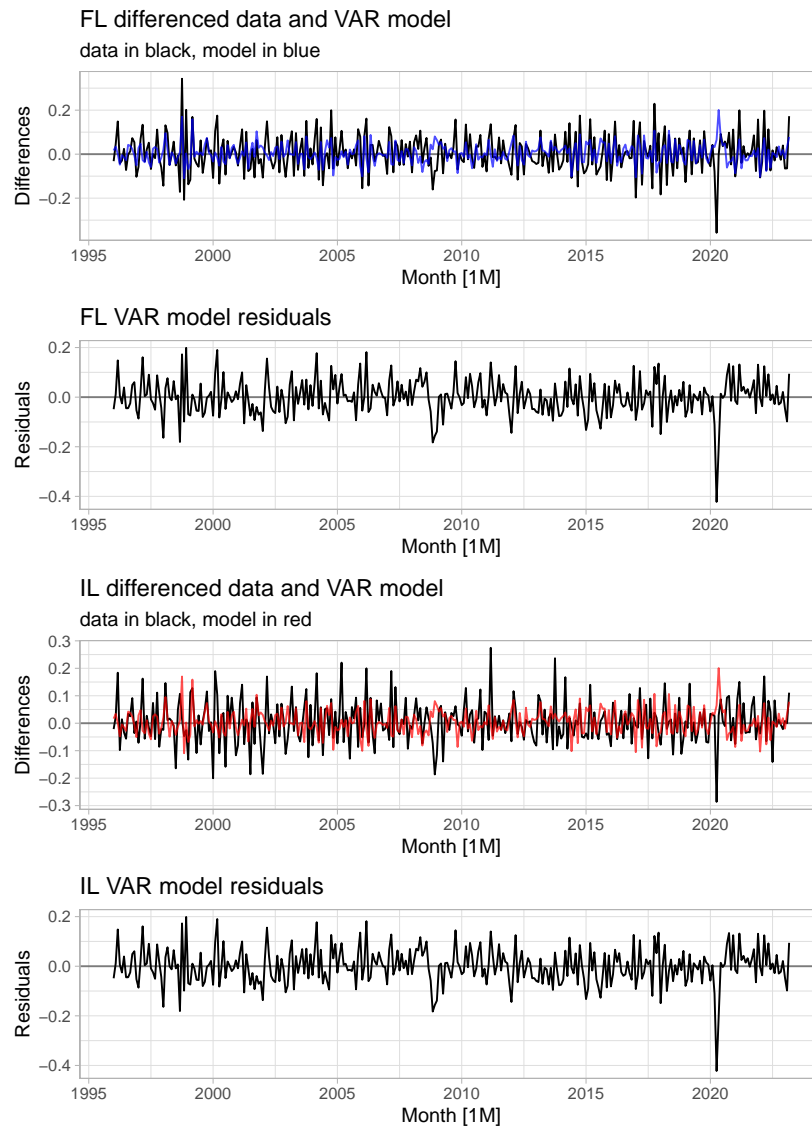
Table 6: VAR model estimates

	FL lag1	IL lag1	FL lag2	IL lag2	FL lag3	IL lag3	FL lag4	IL lag4	Const
FL endog	-0.6355	0.2235	-0.3453	0.1820	-0.0467	-0.1382	-0.1413	-0.1078	0.0078
IL endog	-0.0576	-0.3513	-0.0116	-0.2486	0.1489	-0.3678	0.0302	-0.3695	0.0089

See the appendix for the full summary output of the VAR model.

VAR Model Analysis

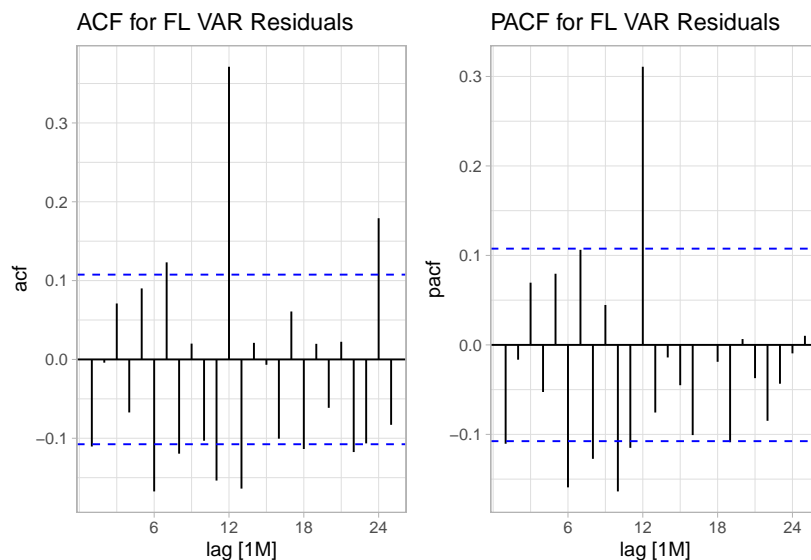
Below we see the VAR fitted model for Florida (the first pair of graphs) and Illinois (the second pair). Note that the data used for the VAR model was differenced, as are the fitted VAR model values. From the graphs, the VAR model does not seem to be entirely accurate, as the model has regions – like after 2020 – where the fitted values and original data diverge significantly. Thus we might not expect the VAR models to perform as well as our earlier models.



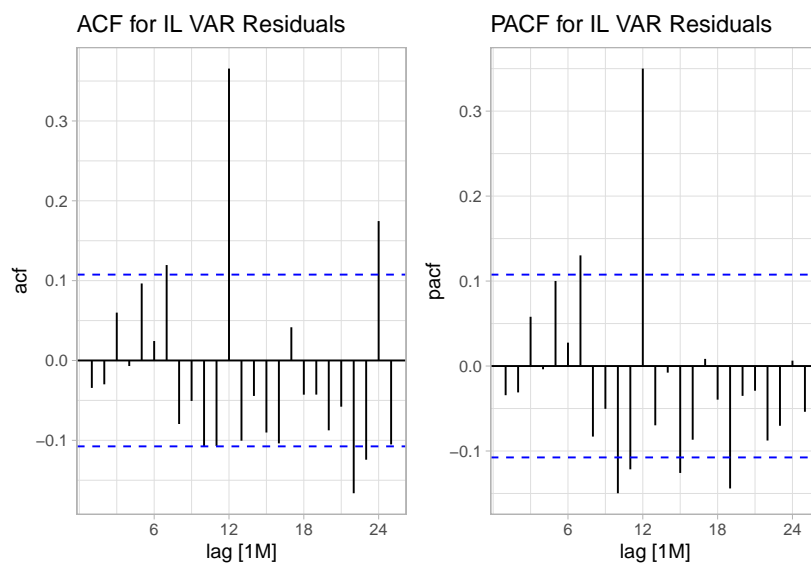
ACF and PACF Analysis

To further assess the VAR model, we will look at the ACF and PACF for both models.

For Florida (the plots below), ACF and PACF show a spike at the 12th lag, indicating seasonality that was not captured. The ACF also shows a significant spike at 24, which might indicate a seasonal AR process in the remainder.

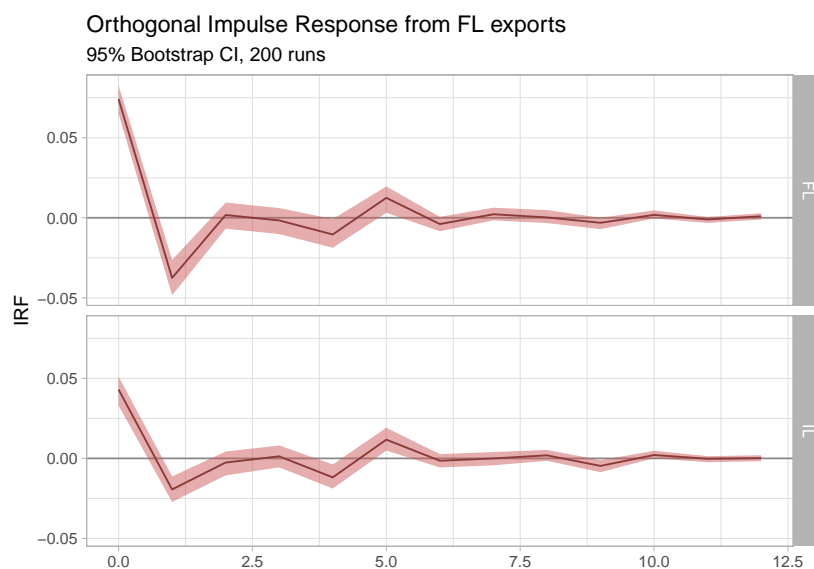


As for Illinois, we see in the below ACF and PACF strong spikes at lag 12, and some other potentially significant spikes at other lags. The spikes at lag 12 indicate seasonal behavior not captured by the model, suggesting a possible seasonal AR process in the remainder.

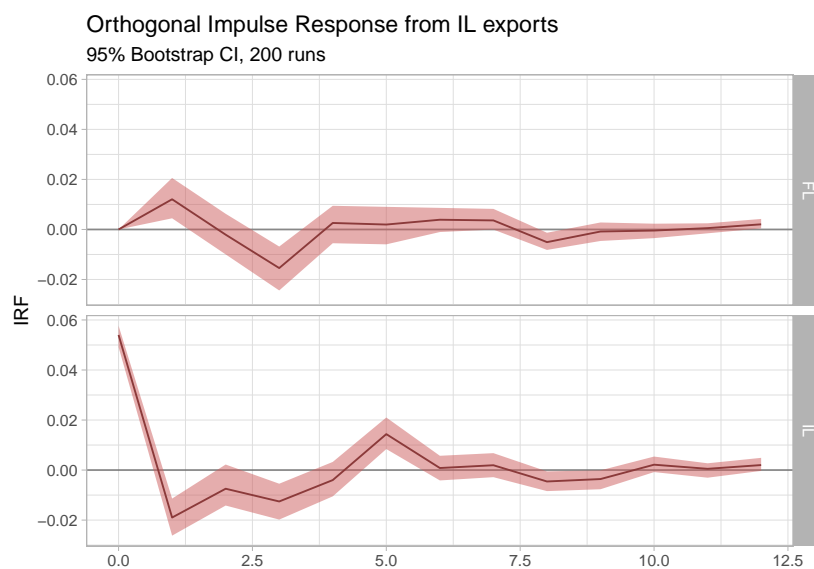


(1) IRF

In the first pair of plots below, we have the impulse response functions for a one unit shock to Florida exports. We see an initial positive effect to both Florida and Illinois exports. After the initial positive start, the impulse response quickly drops past 0 after around one month, then stays near 0 for the future months. It is also worth noting that the magnitudes of both impulse response functions are both very small (around 0.05), indicating that relative to the scale of the data, a shock to Florida exports will have an effect that is not very large.



Then for the next pair of plots, we have the impulse response functions for a one unit shock to Illinois exports. We see a fairly minimal effect to Florida exports, starting and staying near zero, and having fluctuations of a small magnitude. As for the effect on Illinois exports, the effect is similar to that of the shock to Florida's exports: initially positive, dropping below 0 after around one month, then increasing a bit until it stays near 0. Again, the scale of the effect of the shock is very small in comparison to the scale of the actual data.



(m) Granger causality

In the first Granger causality test below, we test if Illinois exports “Granger cause” Florida exports. With a p -value of 0.0001006, we reject the null hypothesis at the 0.001 significance level, meaning there is enough evidence to show that Illinois exports can help predict Florida exports.

The second Granger causality test assesses Florida exports “causing” Illinois exports. With a p -value of 0.0923, we fail to reject the null hypothesis at the 0.05 significance level. Thus there is not enough evidence of Florida exports “Granger causing” Illinois exports.

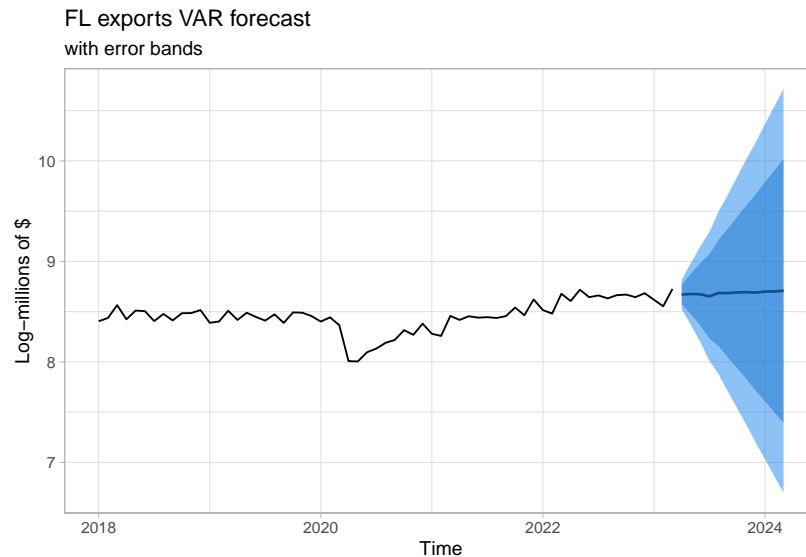
```
## ----- IL causing FL -----
## Granger causality test
##
## Model 1: FL_ts ~ Lags(FL_ts, 1:4) + Lags(IL_ts, 1:4)
## Model 2: FL_ts ~ Lags(FL_ts, 1:4)
##   Res.Df Df       F    Pr(>F)
## 1      319
## 2      323 -4 6.0769 0.0001006 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
##
## ----- FL causing IL -----
## Granger causality test
##
## Model 1: IL_ts ~ Lags(IL_ts, 1:4) + Lags(FL_ts, 1:4)
## Model 2: IL_ts ~ Lags(IL_ts, 1:4)
##   Res.Df Df       F Pr(>F)
## 1      319
## 2      323 -4 2.0139 0.0923 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(n) Forecasts Using VAR Model

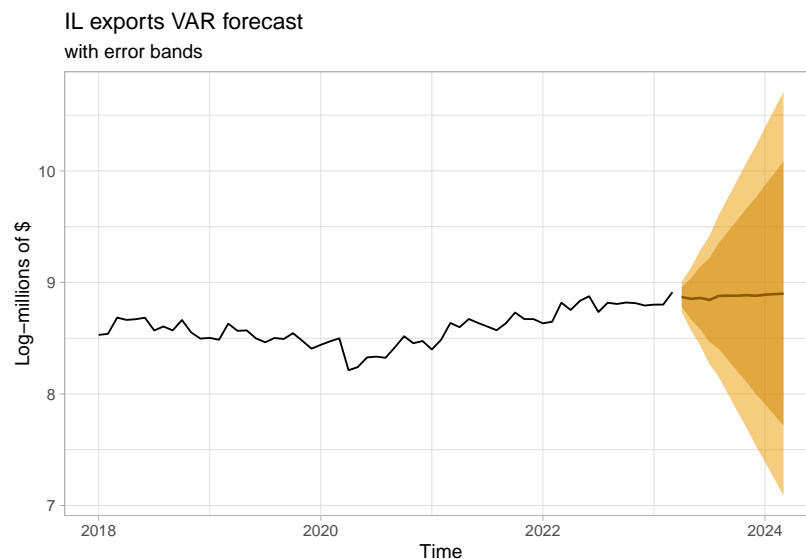
Florida

Below we have the forecast for Florida exports using the VAR model. Note that the forecast is not differenced, rather it is in the original units. Compared to earlier forecasts, this forecast shows the least volatility in the point forecast: a mostly flat point estimate behavior. However, this forecast has the widest error bands, with the width increasing more rapidly than the earlier forecasts. That is due to the accumulating uncertainty associated with forecasting in differenced rather than undifferenced data.



Illinois

Below we have the forecast for Illinois exports using the VAR model. Again, we displayed our forecast in the undifferenced units. Like for Florida, the point estimates show less volatility compared to earlier forecasts, and the error bands are wider with increasing widths (for the same reason as for Florida's VAR forecast).



III. Conclusion and Future Work

Our project consisted of building four models to forecast two different time series: a model incorporating trend, seasonality, and cycles; an ARIMA model; a combined model using the previous two, and a VAR model. Our two time series studied were monthly exports for Illinois and Florida, from August 1995 to March 2023. We initially transformed the data so the units were in log-millions of US dollars.

The first model for both series was an STL decomposition to capture trend and seasonality, and an ARMA component to capture cycles. The forecast for the STL components used the naïve forecast method and the ARMA model was forecasted normally.

The ARIMA model for Florida was an ARIMA(2,1,1) model with a seasonal AR(2) component ($s = 12$). The model for Illinois was an ARIMA(2,1,0) model with a seasonal AR(2) component ($s = 12$).

The combined model used a linear regression to determine the weight allocations for combining the models in a weighted average scheme. For both Florida and Illinois, the original model performed the best out of the three, with the lowest MAPE. The ARIMA model had the highest MAPE and the combined model was in the middle.

For our final model, we applied a first-order difference to the data and fitted a fourth order VAR model. Granger causality tests indicated that Illinois exports “cause” Florida exports but not vice versa.

There are a variety of corrections that could be made to improve our modelling process. From the start, we could have used a Box-Cox transformation of the data using a lambda found with Guerrero’s Method, rather than the standard log-transformation. Doing so would further normalize the variance, making modeling easier. Our STL decomposition assumed unchanging seasonal effects, so a better STL decomposition using a moving seasonality window could increase the accuracy of the model. To make the combined model more dynamic, we could have used a changing (rather than fixed) weighting scheme to optimize our use of the first two models. Lastly, for our VAR model, we could have increased the order beyond 4, using a less strict information criterion than the BIC. Using more lags might have increased the accuracy of the model.

Future work in this area of study would be beneficial, since forecasts for Florida’s and Illinois’ exports could affect business and policy decisions in the long- and short-term. One could also gain further insight into the vitality of American exports at large.

IV. References

- Bureau of Economic Analysis*, “News Release: Gross Domestic Product by State and Personal Income by State, 4th Quarter 2022 and Year 2022.” *US Department of Commerce*, 2023. <https://www.bea.gov/sites/default/files/2023-03/stgdppi4q22-a2022.pdf>.
- Diebold, F.X. (2017), *Forecasting*, Department of Economics, University of Pennsylvania, <http://www.ssc.upenn.edu/~fdiebold/Textbooks.html>.
- Federal Reserve Bank of St. Louis*, “Exports of Goods for Florida (EXPTOTFL).” *United States Federal Reserve System*, 2023. <https://fred.stlouisfed.org/series/EXPTOTFL>.
- Federal Reserve Bank of St. Louis*, “Exports of Goods for Illinois (EXPTOTIL).” *United States Federal Reserve System*, 2023. <https://fred.stlouisfed.org/series/EXPTOTIL>.
- “Florida.” *Forbes*, 2023. <https://www.forbes.com/places/fl/?sh=86f7b0258bcd>.
- Hyndman Rob J., and George Athanasopoulos. *Forecasting: Principles and Practice, 3rd Edition*. Online: OTexts, 2021. <https://otexts.com/fpp3/>.
- “Illinois.” *Forbes*, 2023. <https://www.forbes.com/places/il/?sh=744b452c1777>.
- Observatory of Economic Complexity (OEC)*, “Florida.” *Datawheel*, 2023. https://oec.world/en/profile/subnational_usa_state/fl.
- Observatory of Economic Complexity (OEC)*, “Illinois.” *Datawheel*, 2023. https://oec.world/en/profile/subnational_usa_state/il.

V. Appendix

Full summary output for the VAR model:

```
summary(VAR_model)
```

```
##
## VAR Estimation Results:
## =====
## Endogenous variables: FL_diff, IL_diff
## Deterministic variables: const
## Sample size: 327
## Log Likelihood: 886.051
## Roots of the characteristic polynomial:
## 0.7796 0.7796 0.7756 0.7756 0.6637 0.6637 0.5869 0.5869
## Call:
## VAR(y = combined_df, p = 4)
##
##
## Estimation results for equation FL_diff:
## =====
## FL_diff = FL_diff.l1 + IL_diff.l1 + FL_diff.l2 + IL_diff.l2 + FL_diff.l3 + IL_diff.l3 + FL_diff.l4 +
##
##           Estimate Std. Error t value Pr(>|t|)
## FL_diff.l1 -0.635497   0.070768  -8.980  < 2e-16 ***
## IL_diff.l1  0.223499   0.073142   3.056  0.00244 **
## FL_diff.l2 -0.345304   0.082164  -4.203  3.43e-05 ***
## IL_diff.l2  0.181996   0.080340   2.265  0.02417 *
## FL_diff.l3 -0.046720   0.081827  -0.571  0.56843
## IL_diff.l3 -0.138170   0.081604  -1.693  0.09140 .
## FL_diff.l4 -0.141306   0.070809  -1.996  0.04683 *
## IL_diff.l4 -0.107822   0.074986  -1.438  0.15145
## const      0.007777    0.004153   1.873  0.06205 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.07419 on 318 degrees of freedom
## Multiple R-Squared: 0.2816, Adjusted R-squared: 0.2635
## F-statistic: 15.58 on 8 and 318 DF, p-value: < 2.2e-16
##
##
## Estimation results for equation IL_diff:
## =====
## IL_diff = FL_diff.l1 + IL_diff.l1 + FL_diff.l2 + IL_diff.l2 + FL_diff.l3 + IL_diff.l3 + FL_diff.l4 +
##
##           Estimate Std. Error t value Pr(>|t|)
## FL_diff.l1 -0.057615   0.065879  -0.875  0.382473
## IL_diff.l1 -0.351284   0.068089  -5.159  4.37e-07 ***
## FL_diff.l2 -0.011637   0.076488  -0.152  0.879175
## IL_diff.l2 -0.248590   0.074790  -3.324  0.000992 ***
## FL_diff.l3  0.148866   0.076174   1.954  0.051543 .
## IL_diff.l3 -0.367847   0.075966  -4.842  2.01e-06 ***
## FL_diff.l4  0.030241   0.065917   0.459  0.646713
```

```

## IL_diff.l4 -0.369491    0.069806   -5.293 2.25e-07 ***
## const      0.008923    0.003866    2.308 0.021645 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.06906 on 318 degrees of freedom
## Multiple R-Squared:  0.243,    Adjusted R-squared:  0.224
## F-statistic: 12.76 on 8 and 318 DF,  p-value: 6.312e-16
##
##
## Covariance matrix of residuals:
##          FL_diff  IL_diff
## FL_diff 0.005504 0.003193
## IL_diff 0.003193 0.004770
##
## Correlation matrix of residuals:
##          FL_diff  IL_diff
## FL_diff  1.0000  0.6232
## IL_diff  0.6232  1.0000

```