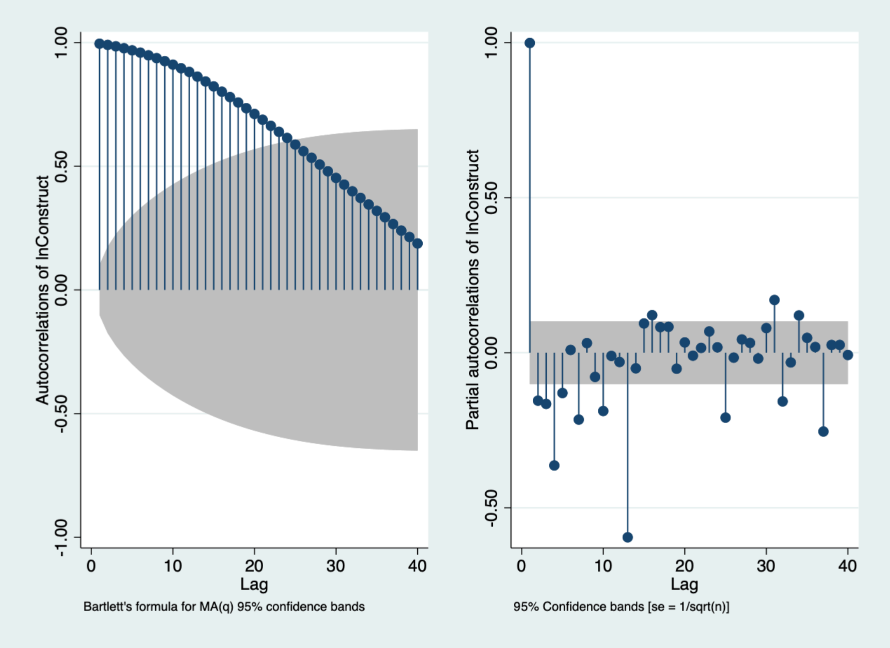
Part 1

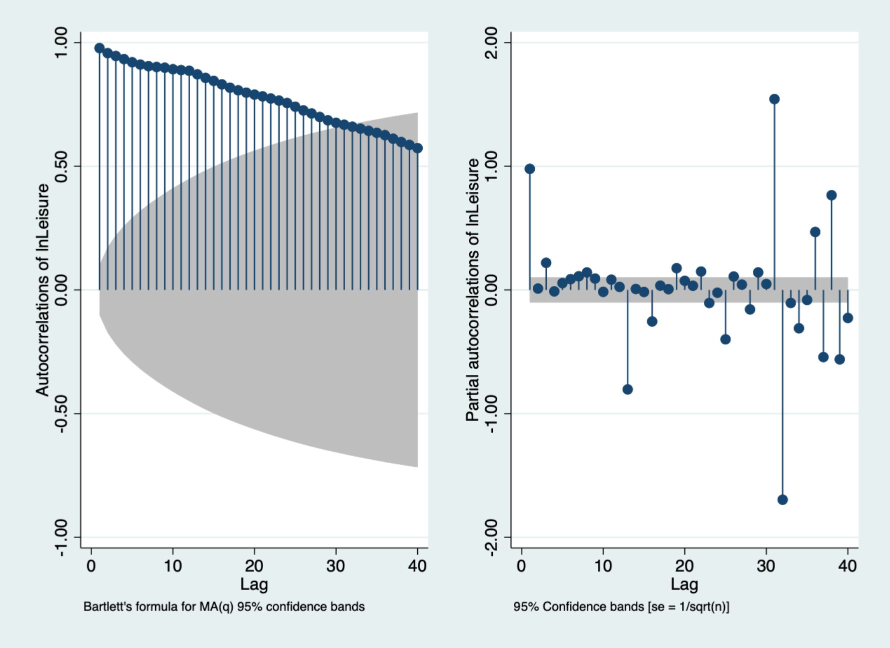
Part 2

Differencing, Log Transforms, and Month Dummies

Differencing

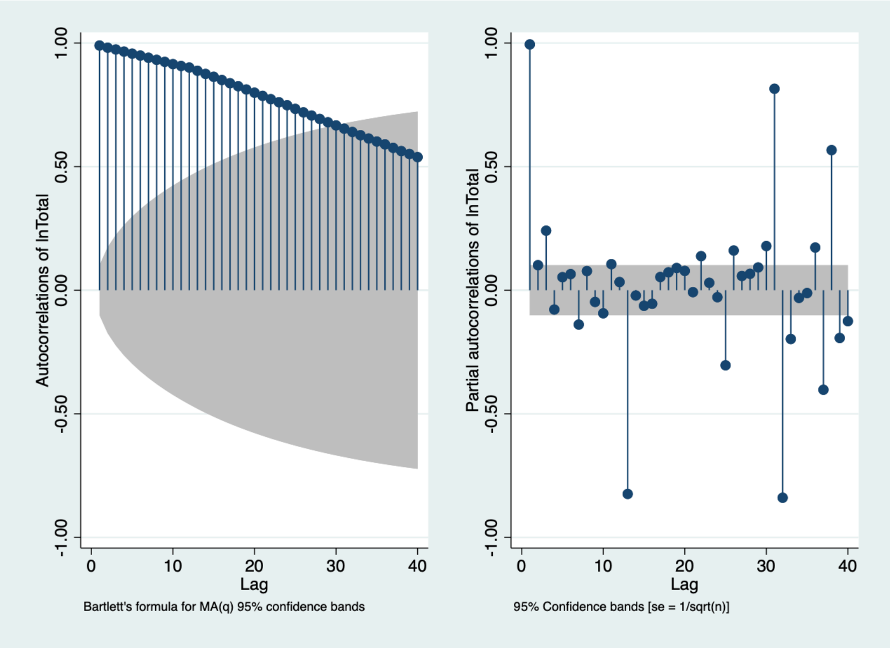
> Construct
>
> 
>
>
> Both first lags are high which means we should difference.

Leisure

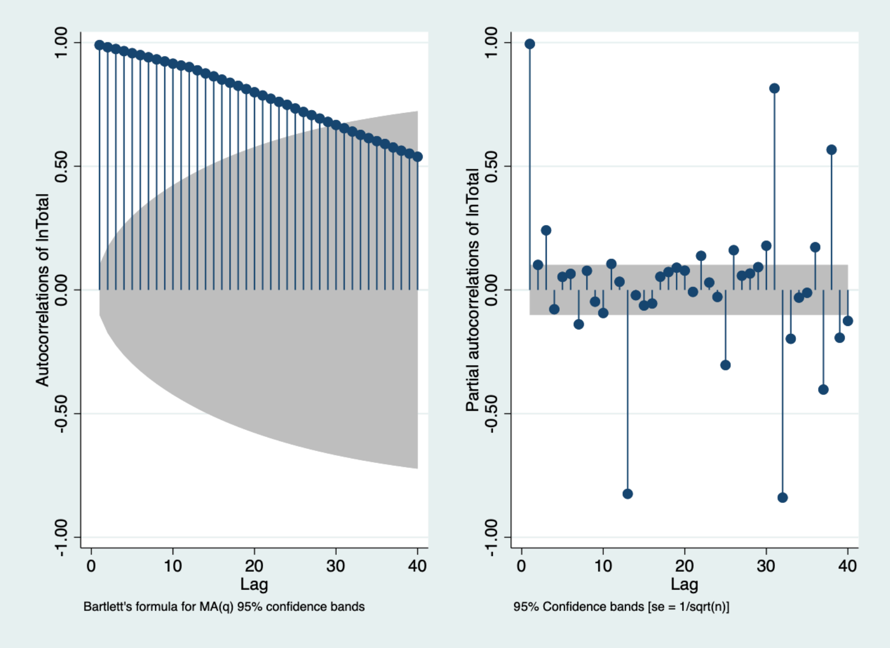


Both first lags are high which means we should difference.

Manufacture



Both first lags are high which means we should difference.

> Total
>
> 
>
>
> Both first lags are high which means we should difference.

Log Transforms

Log transforms make the data not have any values less than zero and forces the data into a normal distribution. It also transforms the data so it has proportional changes rather than absolute changes so that any changes over time can be reported as a percent change.

Month Dummies

There's not any reason to not include month dummies. If your data is monthly or any other form of seasonal, it will help your models because they're now identified to a particular season. If your data isn't seasonal, they won't have any effect.

Content Knowledge and Model Searches

Content Knowledge

Content knowledge can speed up the model selection process because you may already have an idea of what variables or lags have an effect on the dependent variable. For example, hourly wages and hours scheduled per week are probably a very good indication of monthly wages.

GSREG

Global search regression takes all the variables you feed it and runs a regression for any combination of the variables. This is a powerful tool to fine-tune your models, but without filtering the variables through content knowledge, it could take a very long time to run. Rather than just taking the highest scoring model, you should then examine common features of the highest scoring models on the basis of AIC, BIC, and out of sample root mean square error, and choose the most parsimonious one.

What's wrong with *stepwise* model selection?

It's prone to over fitting because it has bad predictive properties. Instead you should use out of sample fitting because it protects against over fitting. Over fitting is caused by dropping the most insignifcant each step which may include variables that should be included in the model but are not relevant on their own.

Choosing Models

Model Type	Model	AIC	BIC	Root Mean Squared Errors
AR only Lags 1-3 Month dummies	`reg d.lnTotal l(1/3)d.lnTotal m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11`	-2382.725	-2323.982	.0129605
AR only Lags 1-3,12,24 Month dummies	`reg d.lnTotal l(1/3,12,24)d.lnTotal m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11`	-2236.218	-2170.633	.01300797
ARDL Lags 1-3 Month dummies	`reg d.lnTotal l(1/3)d.lnTotal l(1/3)d.lnConstruct l(1/3)d.lnLeisure l(1/3)d.lnManufacture m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11`	-2393.008	-2299.019	.01611154
ARDL Lags 1-3,12,24 Month dummies	`reg d.lnTotal l(1/3,12,24)d.lnTotal l(1/3,12,24)d.lnConstruct l(1/3,12,24)d.lnLeisure l(1/3)d.lnManufacture m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11`	-2230.781	-2115.043	.01713897

Which are the best two and why?

Model 1 has the lowest root mean squared error and model 3 has the lowest AIC and BIC. I'm also inclined to believe these are the better ones because they don't include lags 12 and 24 which is a long time for subcomponents of the total employment variable to have an effect on the total employment variable.

Rolling Window

Model 1

```
reg d.lnTotal l(1/3)d.lnTotal m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11
```

Value Type	Value
RWmaxobs12	12
RWminobs12	12
RWrmse12	.0132376

A window width of 12 had the lowest RWrmse. I thought that maybe a smaller window width would be better because the lags did not include lag 12 but I was wrong. Besides 12 months, 6 months had the second lowest.

Model 3

```
reg d.lnTotal l(1/3)d.lnTotal l(1/3)d.lnConstruct l(1/3)d.lnLeisure
l(1/3)d.lnManufacture m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11
```

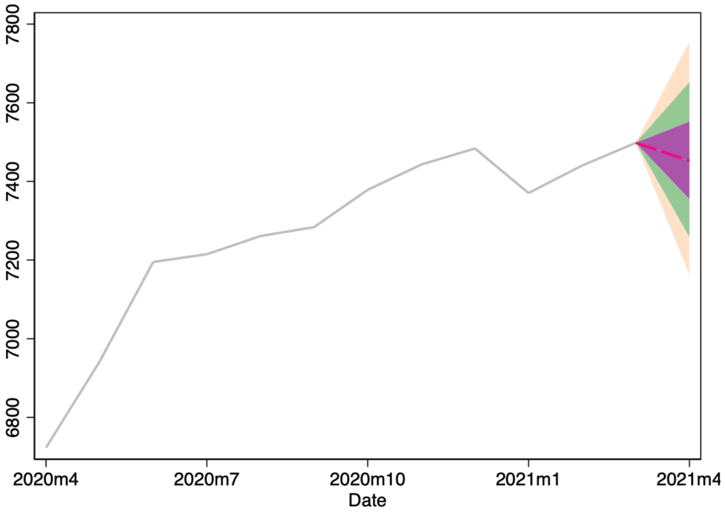
Value Type	Value
RWmaxobs12	12
RWminobs12	12
RWrmse12	.0132376

A window width of 12 had the lowest RWrmse. After my failure in model 1, I tried again hoping for better results. A window width of 12 is still the best.

Ultimately, I'm going to choose model 1 because it is autoregressive and that is what makes ARIMA work and without ARIMA I could not make my pretty fan charts. They have the same RWrmse anyways so I can't imagine the extra variables have too big a difference. And

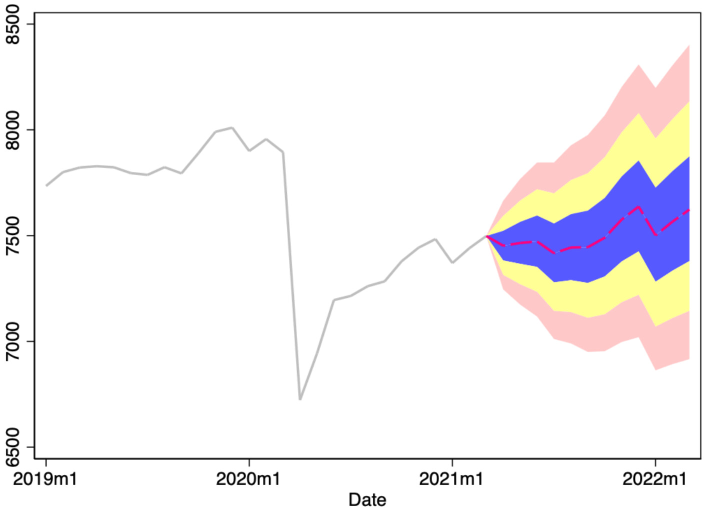
Forecasting

One month ahead



> Sorry for the bad colors. Hailey peer pressured me into it.

One year out



> Ditto my earlier comment on the colors :)

Forecast Evaluation

Empirical

Normal