US Unemployment Forecast

By Meiru Zhong, Linan Zhou, Tianqi Xiong, Shihui Cao

Introduction

Field of Consultancy

- Forecast the US unemployment rate
    - Introduce S&P 500 Index as external regressor

Selected Methods

- Data-driven methods: naive forecasts & smoothing methods
- Model-based methods: ARIMA

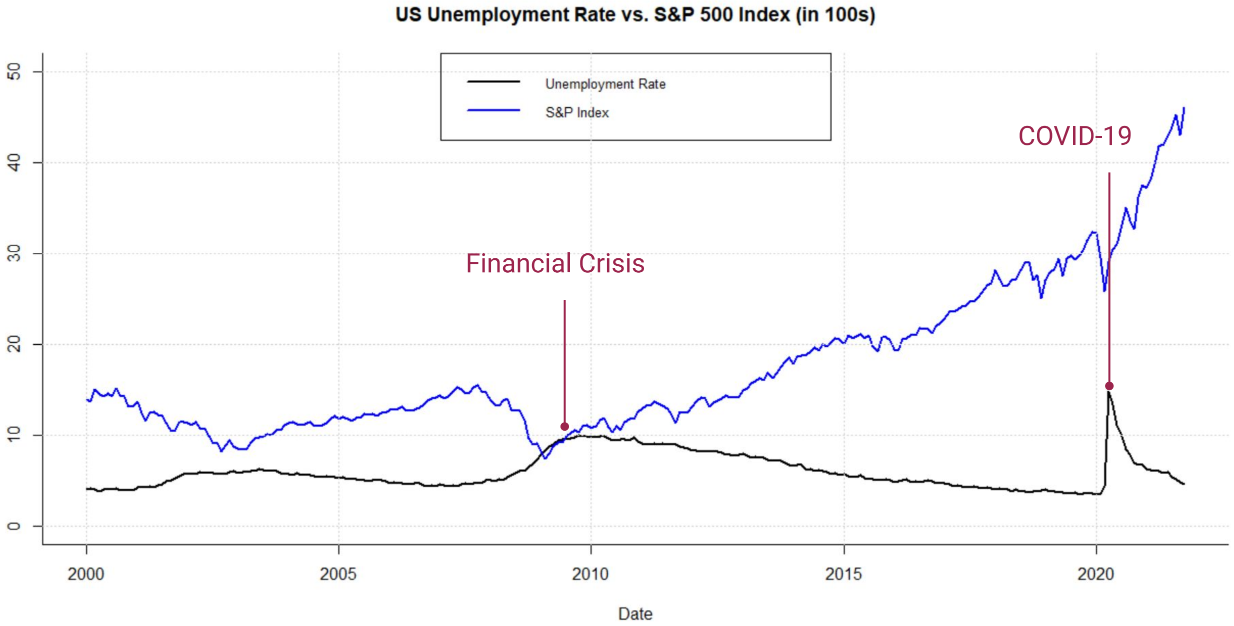
Intended Audience

- Policy makers, economic researchers

Data Analysis

Data Sources

- US unemployment data: Federal Reserve Economic Data
- S&P Index: Yahoo Finance
- Data range: Monthly data from January 2000 to October 2021



Exploratory Data Analysis

- US unemployment rate
  - Significant trend, insignificant seasonality
  - A cyclic pattern corresponds to recessions & expansions
- S&P 500 Index
  - Upward trend with fluctuations, insignificant seasonality
- Negative correlation between unemployment rate and S&P Index (intuitively)

Stationarity Analysis

ADF Test for Unemployment Rate

- The raw data is non-stationary

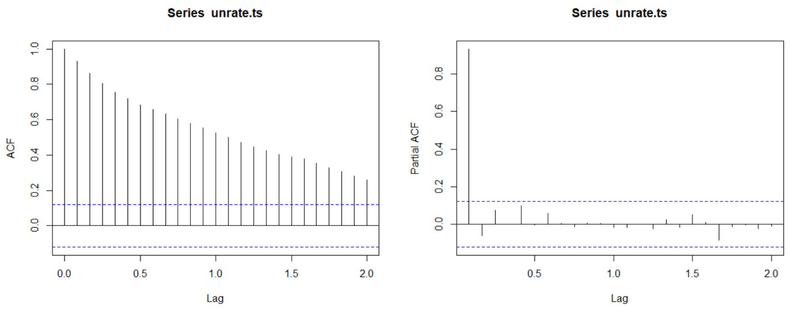
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.006232   0.006945  -0.897    0.370
z.diff.lag  0.032992   0.062248   0.530    0.597
```

Differencing

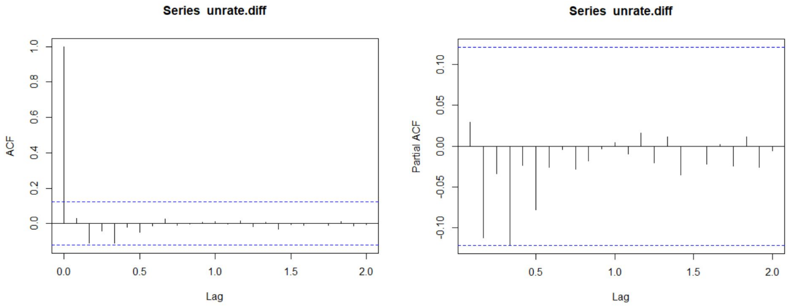
- Take 1st differencing to remove trend
- Unemployment rate after removing trend is stationary

```
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -1.07923    0.08641  -12.49   <2e-16 ***
z.diff.lag  0.11228    0.06203    1.81   0.0714 .
```

ACF and PACF of raw data



ACF and PACF after removing trend



Stationarity Analysis

ADF Test for S&P 500 Index

- The raw data is non-stationary

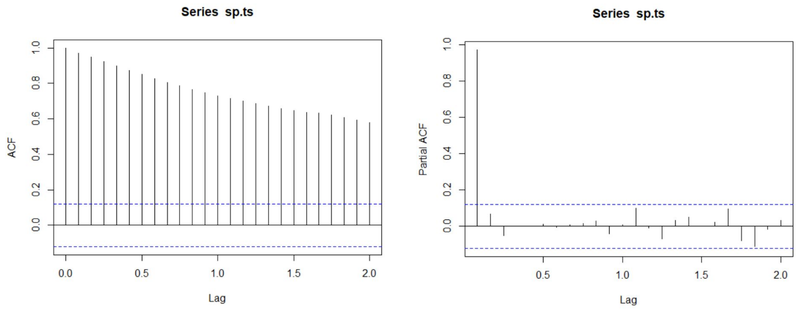
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.207413   0.120779  -1.717   0.0871 .
z.lag.1      0.006238   0.011423   0.546   0.5855
tt           0.001747   0.001216   1.437   0.1520
z.diff.lag  -0.057738   0.064512  -0.895   0.3716
```

Differencing

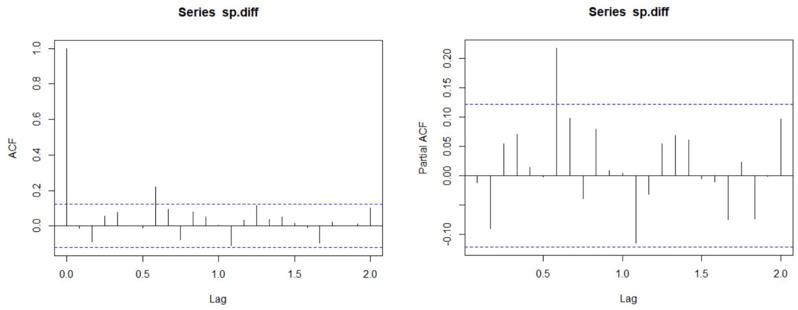
- Take 1st differencing to remove trend
- S&P 500 Index after removing trend is stationary

```
Coefficients:
              Estimate  Std. Error t value Pr(>|t|)
(Intercept) -0.2177205  0.1007829  -2.160   0.0317 *
z.lag.1     -1.2029908  0.0901501 -13.344  < 2e-16 ***
tt           0.0027619  0.0006879   4.015 7.82e-05 ***
z.diff.lag   0.1505271  0.0640258   2.351   0.0195 *
```

ACF and PACF of raw data



ACF and PACF after removing trend



Cointegration

Since S&P 500 Index and unemployment rate are non-stationary

↓

Linear regression to estimate long-term cointegration relationship

↓

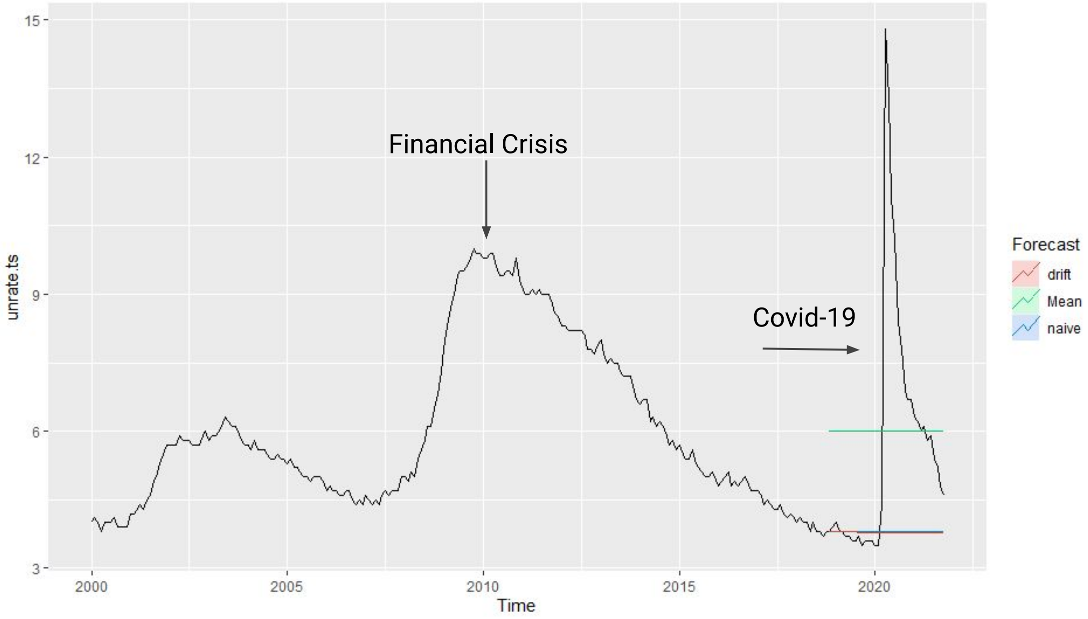
ADF test for the residuals of the relationship

↓

p-value < 0.05, indicating H0 can be rejected and two variables are cointegrated

```
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-1.5053 -0.1757 -0.0540  0.0955 10.4614

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.009089   0.043388   0.209  0.83425
z.lag.1     -0.075768   0.023334  -3.247  0.00132 **
z.diff.lag   0.043602   0.062161   0.701  0.48367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6996 on 257 degrees of freedom
Multiple R-squared:  0.03946,	Adjusted R-squared:  0.03198
F-statistic: 5.279 on 2 and 257 DF,  p-value: 0.005667


Value of test-statistic is: -3.2472 5.2929

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

Model Selection: simple forecasting methods



We use the following three forecasting methods as benchmark:

- **Average method**: the forecasts are equal to the mean of the historical data.
- **Naive method**: the forecasts will be the value of the last observation.
- **Drift method**: based on the naive method, the forecasts can drift at the amount of the average change.

Comparison between benchmark models

Performance metrics	ME		RMSE		MAE	
	Training Set	Test Set	Training Set	Test Set	Training Set	Test Set
Average Method	-5.4598e-17	0.3076	1.7745	2.7840	1.4545	0.636
Naive Method	-0.0009	1.9083	0.1636	3.3612	0.1218	2.0306
Drift Method	-1.8945e-16	1.9248	0.1636	3.3731	0.1219	2.0410

Drift model is good for fitting,

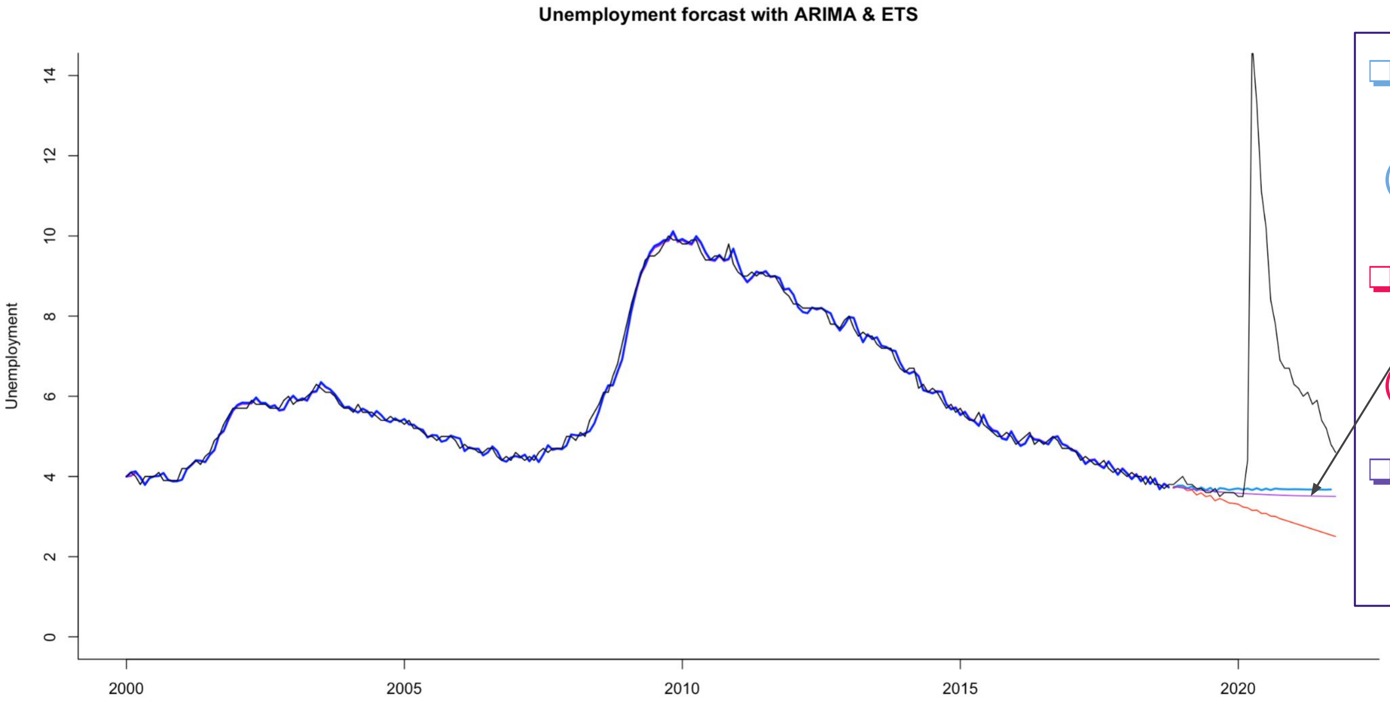
but average model is good for predicting

Granger Causality

- Granger Test: **S&P 500 is granger cause to unemployment.**
- Pick: **ARIMA model** (Two variables are not endogenous)

```
Granger causality test

Model 1: unrate.ts ~ Lags(unrate.ts, 1:1) + Lags(sp.ts, 1:1)
Model 2: unrate.ts ~ Lags(unrate.ts, 1:1)
  Res.Df Df      F  Pr(>F)
1    258
2    259 -1 4.0862 0.04427 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Selection: ARIMA & ETS



- Blue line: forecast with S&P 500 regressor (ARIMA(1,2,1)(0,0,2)[12])
- Pink line: forecast without S&P 500 regressor (ARIMA(2,1,1)(0,0,2)[12])
- Purple line:ETS Forecast Model: ETS(M,M,N)

Performance Comparison

Performance Metrics	ARIMA With S&P 500	ARIMA Without S&P 500	ETS
ME	2.050465	2.562338	2.126475
RMSE	3.475854	3.879335	3.513846
MAE	2.109308	2.562338	2.146424
MPE	24.25706	33.85445	26.09967
MAPE	25.91683	33.85445	26.66583

ARIMA model with S&P regressor performs the best

Summary and Interpretations

1. Strong inverse relationship between S&P 500 and unemployment rate

2. S&P 500 is granger cause to unemployment rate but not vice versa

- ➢ Unemployment rate impact on S&P 500?

  - Blanchard (1981) and Orphanides (1992) supported the view that stock price responses to macroeconomic news may depend on the state of the economy

  - Farsio (2013) applied Cointegration Test and Granger Causality Test to prove there is no causal effect from unemployment to stock prices

- ➢ S&P 500 impact on unemployment rate?

  - Granger causality is not necessarily true causality

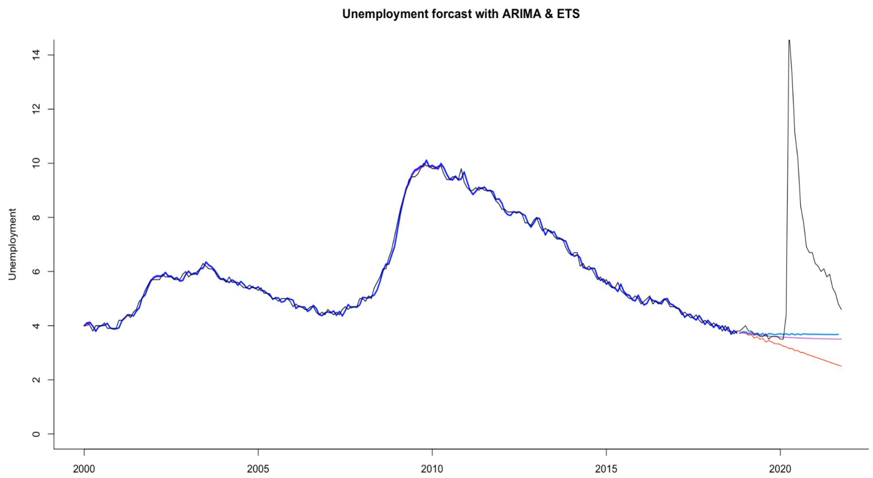
  - Both variables are endogenous whose levels and movements depend on a variety of exogenous factors

Summary and Interpretations

3. Forecasting unemployment rate with S&P 500 regressor improved the model performance

 - Unemployment rate is a lagging indicator

4. Benchmark models slightly better than ARIMA and ETS on metric RMSE: impact of COVID



Performance Metrics	Drift Model	ARIMA With S&P 500	ARIMA Without S&P 500	ETS
RMSE	3.37	3.48	3.88	3.51

Related Research and Further Steps

Forecasts of Professional Forecasters:

- The Federal Reserve Board’s Greenbook
- The Federal Reserve Bank of Philadelphiaís Survey of Professional Forecasters
- Blue Chip panel of economists

Researchers applied various indicators and time series models to forecast the unemployment rate:

- Indicators: Labor Force Flows, Degree of Agreement in Consumer Unemployment Expectations, Job Openings Index, GDP and Inflation, etc.
- Models: SARIMA (seasonal ARIMA), VAR, ETS, SETAR(self-exciting threshold autoregressive), Holt–Winters, NNAR(neural network autoregression), etc.

Using time series models can significantly improve performance of forecasting

Thank You

Questions are welcome