Unobserved Components Models: Applications in Post-COVID Analysis

David J Corliss, PhD is Principal Data Scientist at Grafham Analytics and the founder and Director of Peace-Work, a volunteer cooperative of statisticians and data scientists applying statistical methods to issue-driven advocacy in poverty, education, social justice, and providing analytic support for charitable groups. He writes a column on Data for Good for ASA's Amstat News and serves on the steering committee of the Statistics section of the American Association for the Advancement of Science. He holds a PhD in statistical astrophysics from the University of Toledo.

UNOBSERVED COMPONENTS MODELS: APPLICATIONS IN POST-COVID ANALYSIS

David J Corliss, Peace-Work

Southeast SAS User Group Conference
Charlotte NC
October 22-24, 2023

OUTLINE

Introduction to UCM

Measuring Changes in Baseline Values

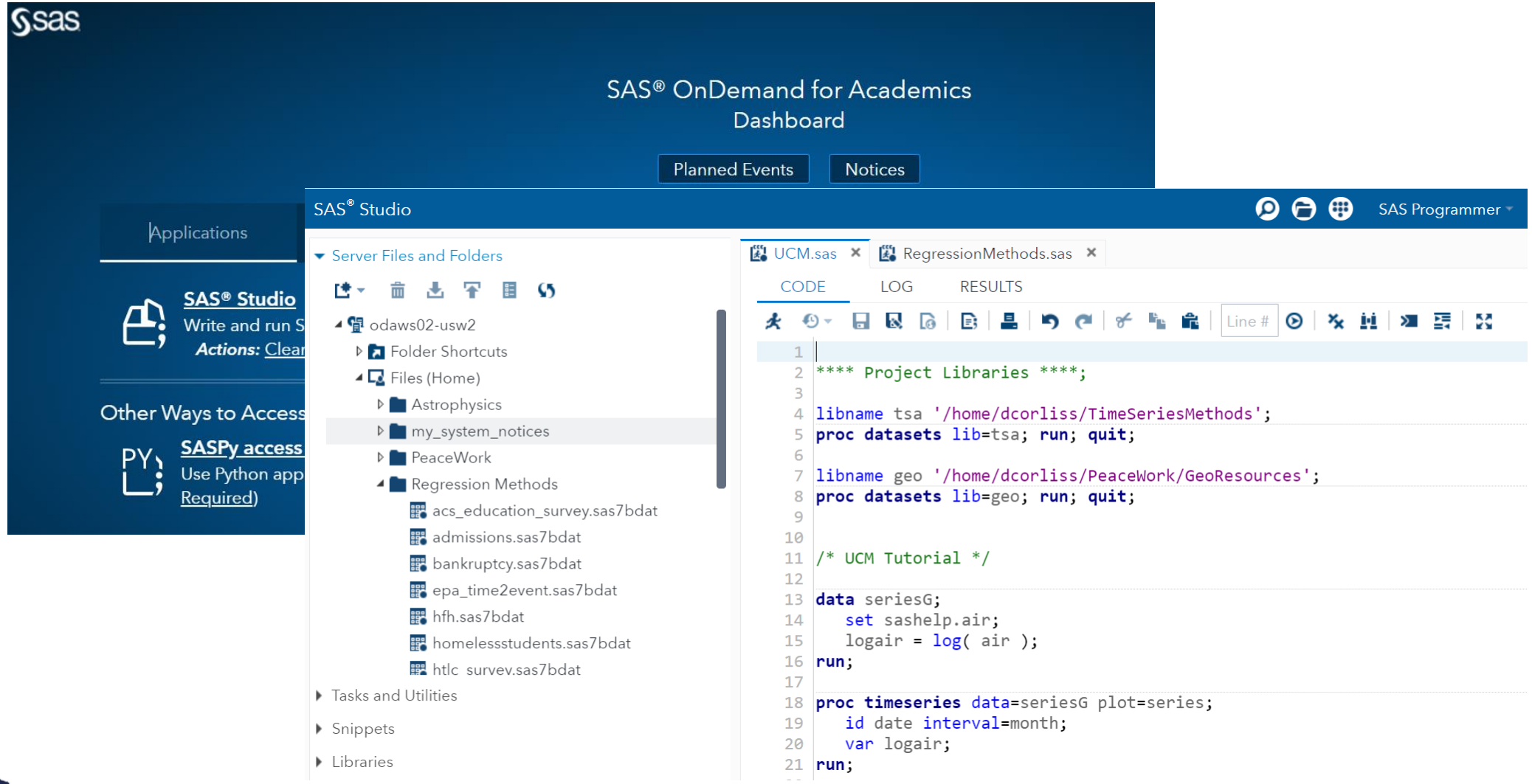
COVID Questions

Limitations of UCM

Conclusions

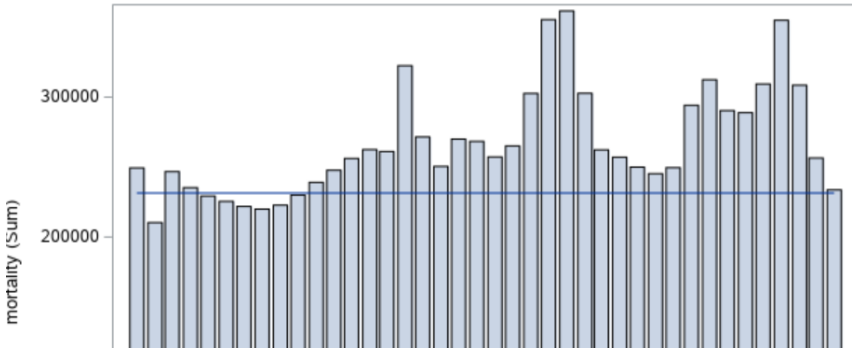
INTRODUCTION TO UNOBSERVED COMPONENTS MODELS

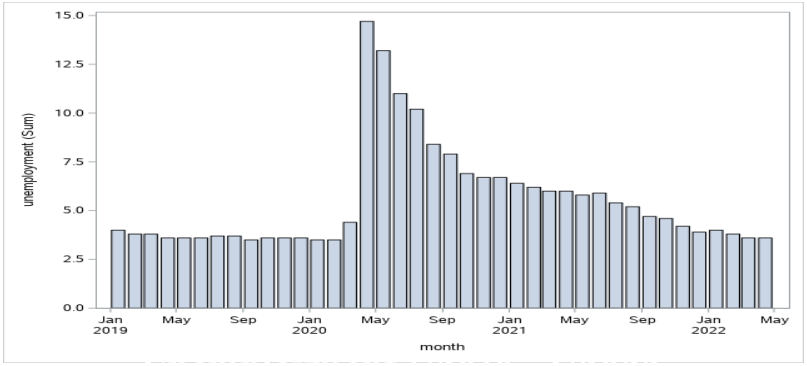
SAS® On Demand for Academics



COVID Data: A Complex Time Series

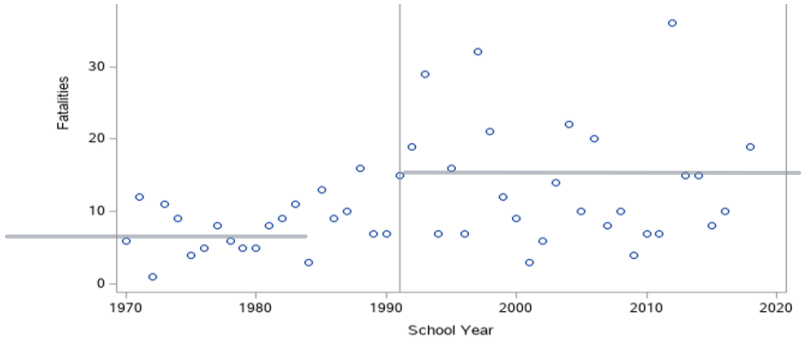
Series of Waves





More than Medical

=> Unobserved Components



Unobserved Components Models

- **Model Type: State Space Time Series Model, A. Harvey 1989**
- **Decomposes a time series into unobserved components that together form the time series, including trends, periodic behavior, and irregular components**
- **Supports measurement of changes in long-term baseline values of the time series => good for modeling high-impact events**
- **SAS: PROC UCM, R Package: rucm**

PROC UCM Source Code and Options

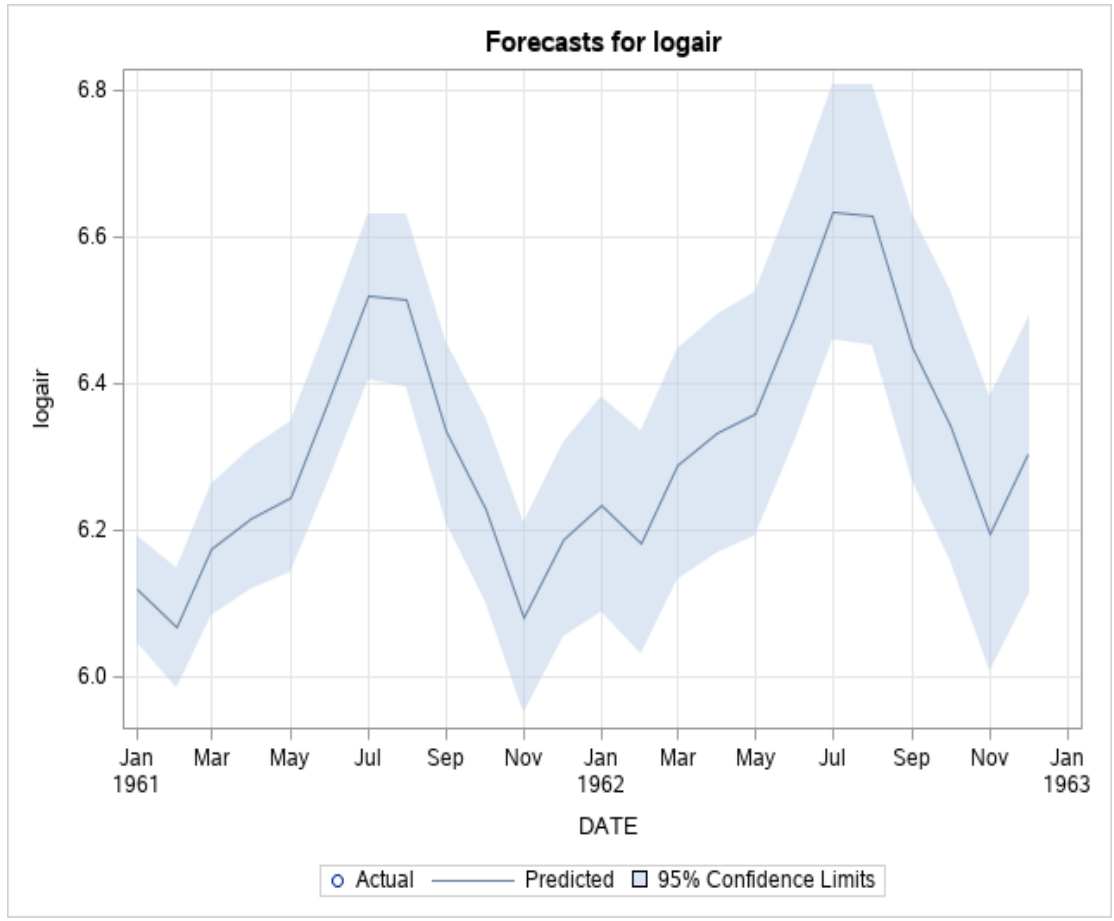
```
proc ucm data=seriesG;
   id date interval=month;
   model logair;
   irregular;
   level;
   slope;
   season length=12 type=trig print=smooth;
   estimate;
   forecast lead=24 print=decomp;
run;
```

- irregular, level, slope, season = potential components in the model
- id is the name of the date / time variable, interval is the time between observations
- forecast specifies the lead = number of periods in the forecast

Unobserved Components Model Results, Output and Plots

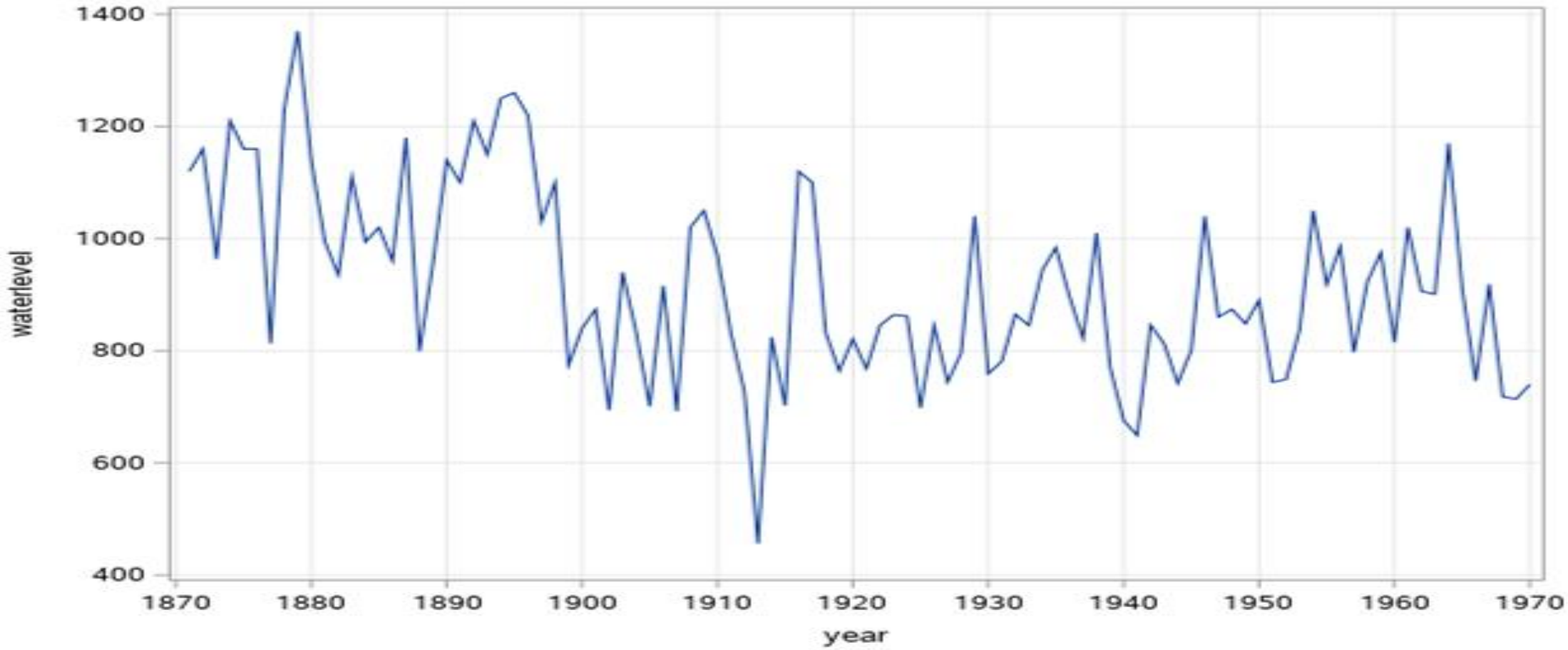
Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	0.00023436	0.0001079	2.17	0.0298
Level	Error Variance	0.00029828	0.0001057	2.82	0.0048
Slope	Error Variance	8.47922E-13	6.2271E-10	0.00	0.9989
Season	Error Variance	0.00000356	1.32347E-6	2.69	0.0072

Fit Statistics Based on Residuals	
Mean Squared Error	0.00147
Root Mean Squared Error	0.03830
Mean Absolute Percentage Error	0.54132
Maximum Percent Error	2.19097
R-Square	0.99061
Adjusted R-Square	0.99039
Random Walk R-Square	0.87288
Amemiya's Adjusted R-Square	0.99002
Number of non-missing residuals used for computing the fit statistics = 131	



CHANGES IN BASELINE LEVELS WITH UCM

A UCM Classic Example: Depth of the Nile River




```
proc timeseries data=nile plot=series;
   id year interval=year;
   var waterlevel;
run;
```

A UCM Classic Example: Depth of the Nile River

```
proc ucm data=nile;
   id year interval=year;
   model waterlevel;
   irregular;
   level plot=smooth checkbreak;
   estimate;
   forecast plot=decomp;
run;
```

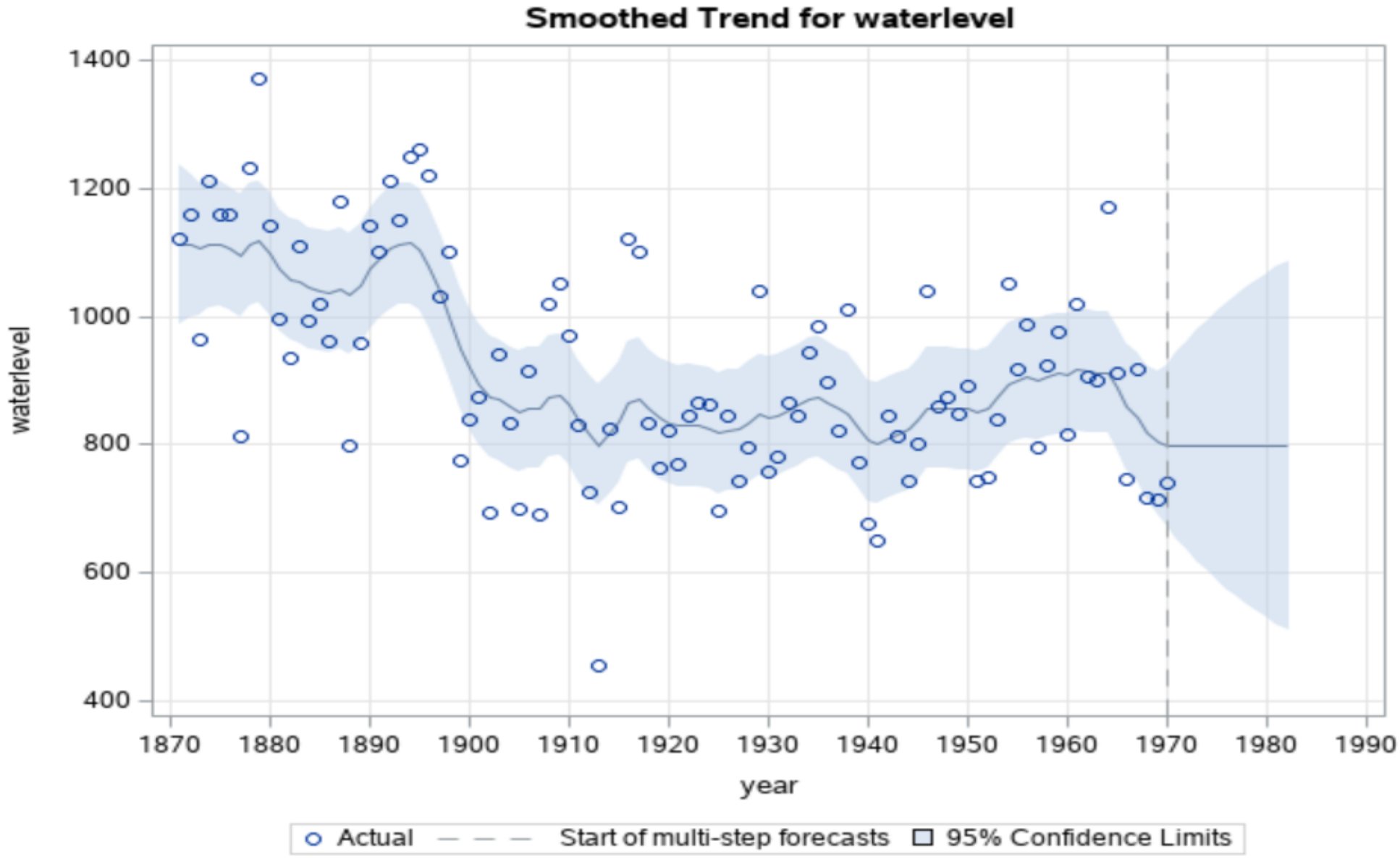
Final Estimates of the Free Parameters

Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	15099	3145.5	4.80	<.0001
Level	Error Variance	1469.17636	1280.4	1.15	0.2512

Fit Statistics Based on Residuals

Mean Squared Error	20689
Root Mean Squared Error	143.83609
Mean Absolute Percentage Error	13.09656
Maximum Percent Error	32.91501
R-Square	0.26706
Adjusted R-Square	0.25950
Random Walk R-Square	0.26066
Amemiya's Adjusted R-Square	0.23684
Number of non-missing residuals used for computing the fit statistics = 99	

A UCM Classic Example: How has the Depth Changed?

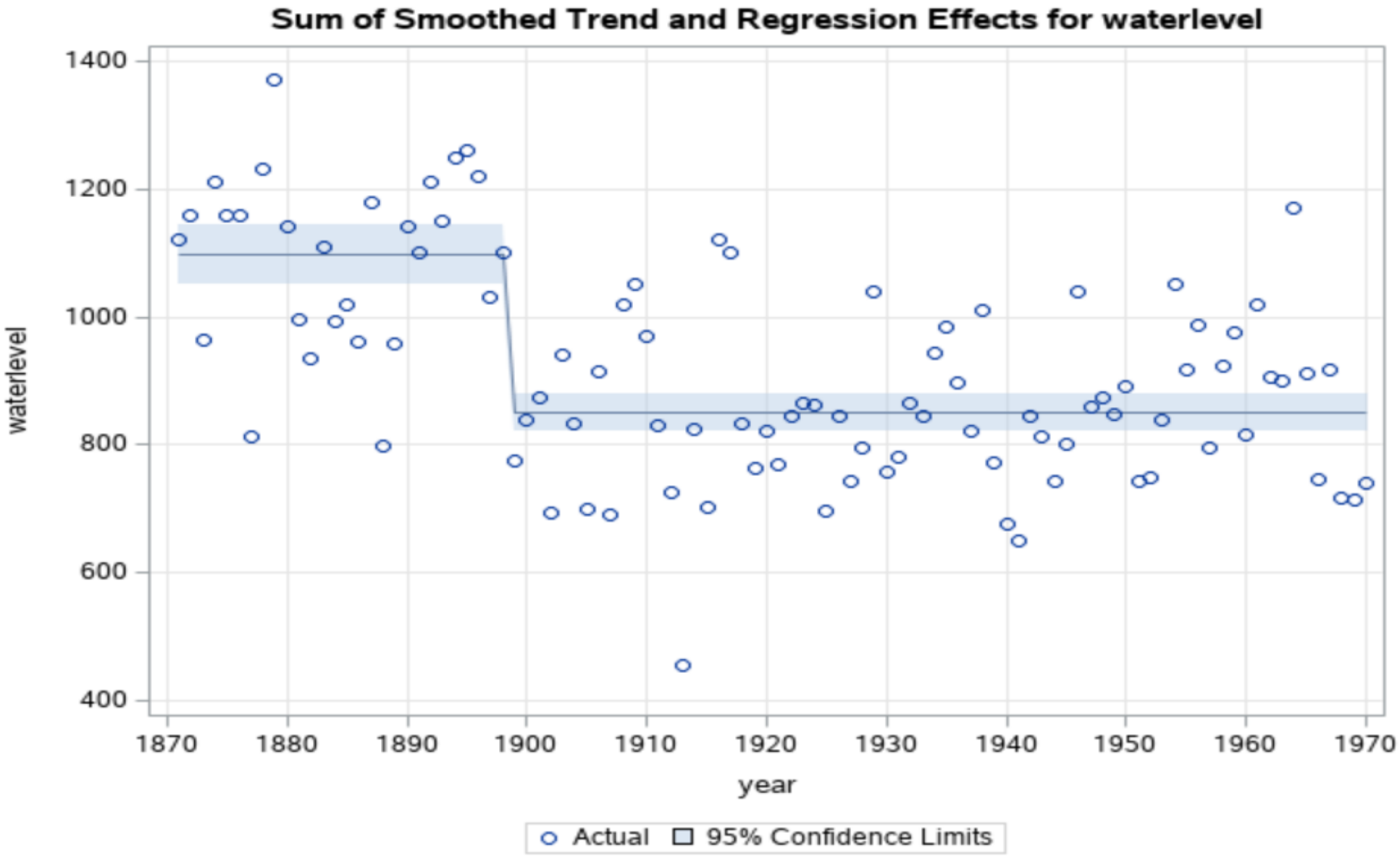


A UCM Classic Example: Depth of the Nile River

```
data nile;
   set nile;
   shift1899 = ( year >= '1jan1899'd );
run;
```

```
proc ucm data=nile;
   id year interval=year;
   model waterlevel = shift1899;
   irregular;
   level;
   estimate;
   forecast plot=decomp;
run;
```

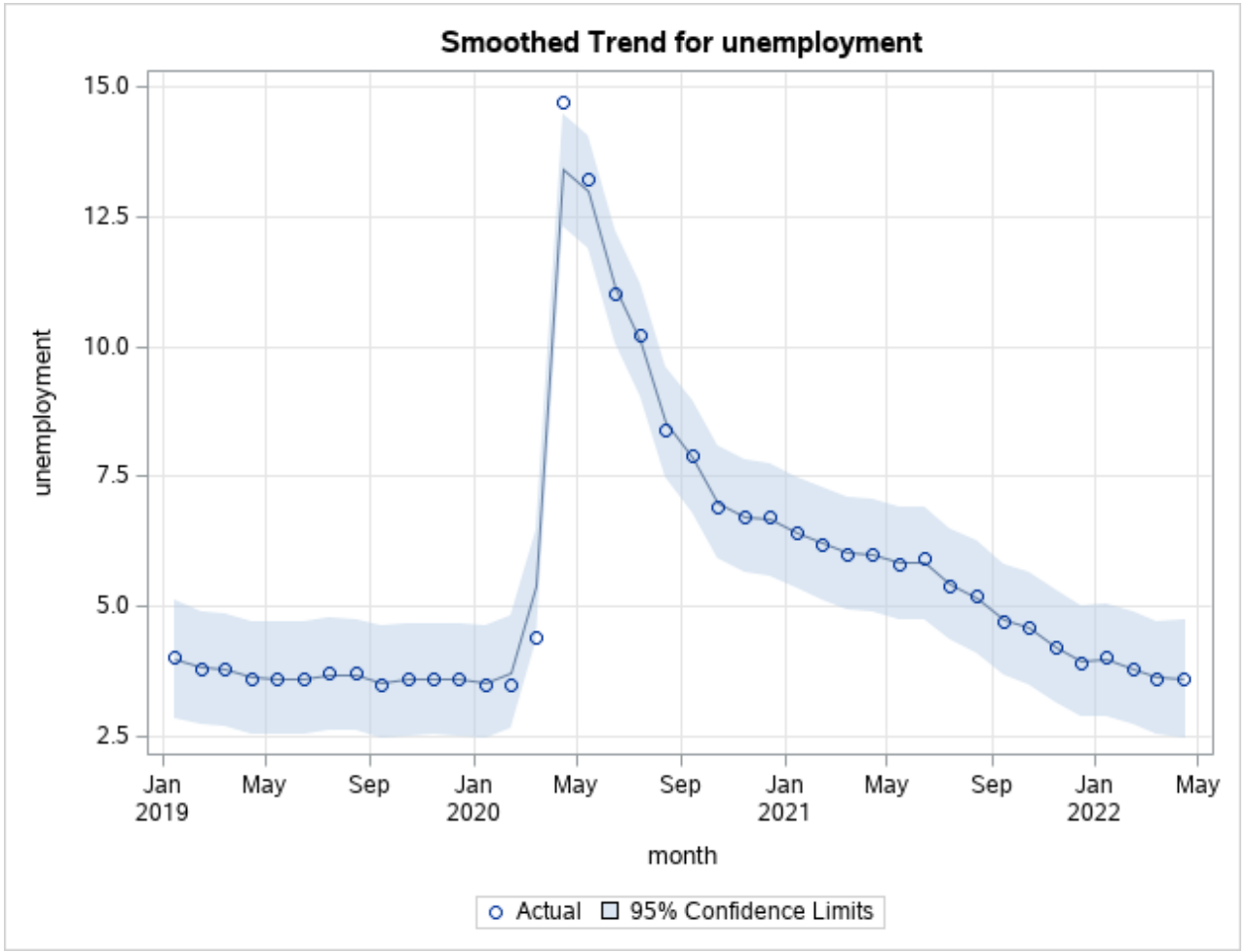
A UCM Classic Example: How has the Depth Changed?



COVID QUESTIONS

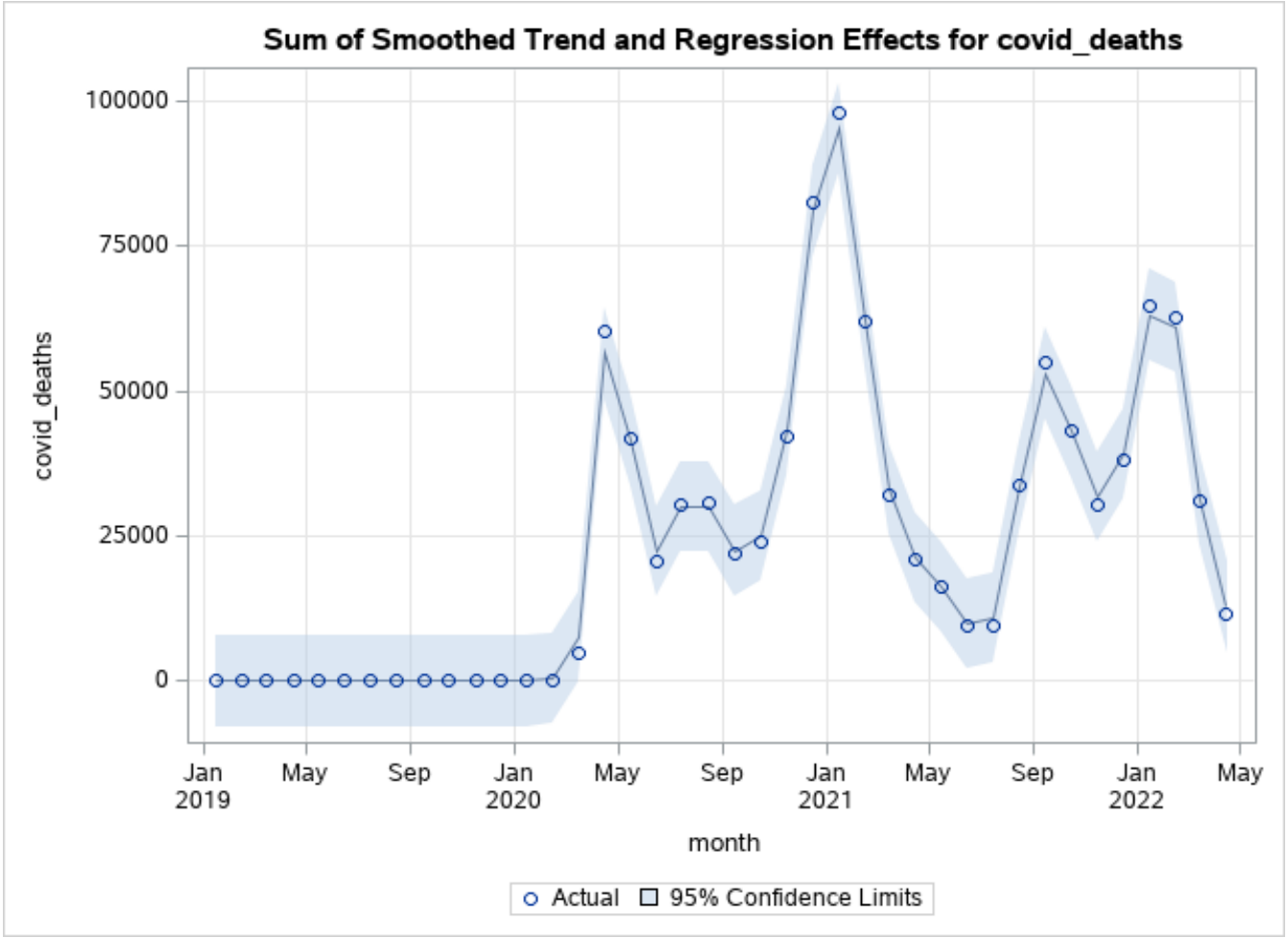
COVID-Era Baseline Changes: Unemployment

```
proc ucm data=work.covid;
   id month interval=month;
   model unemployment = shift2020;
   irregular;
   level;
   estimate;
   forecast plot=decomp;
run;
```



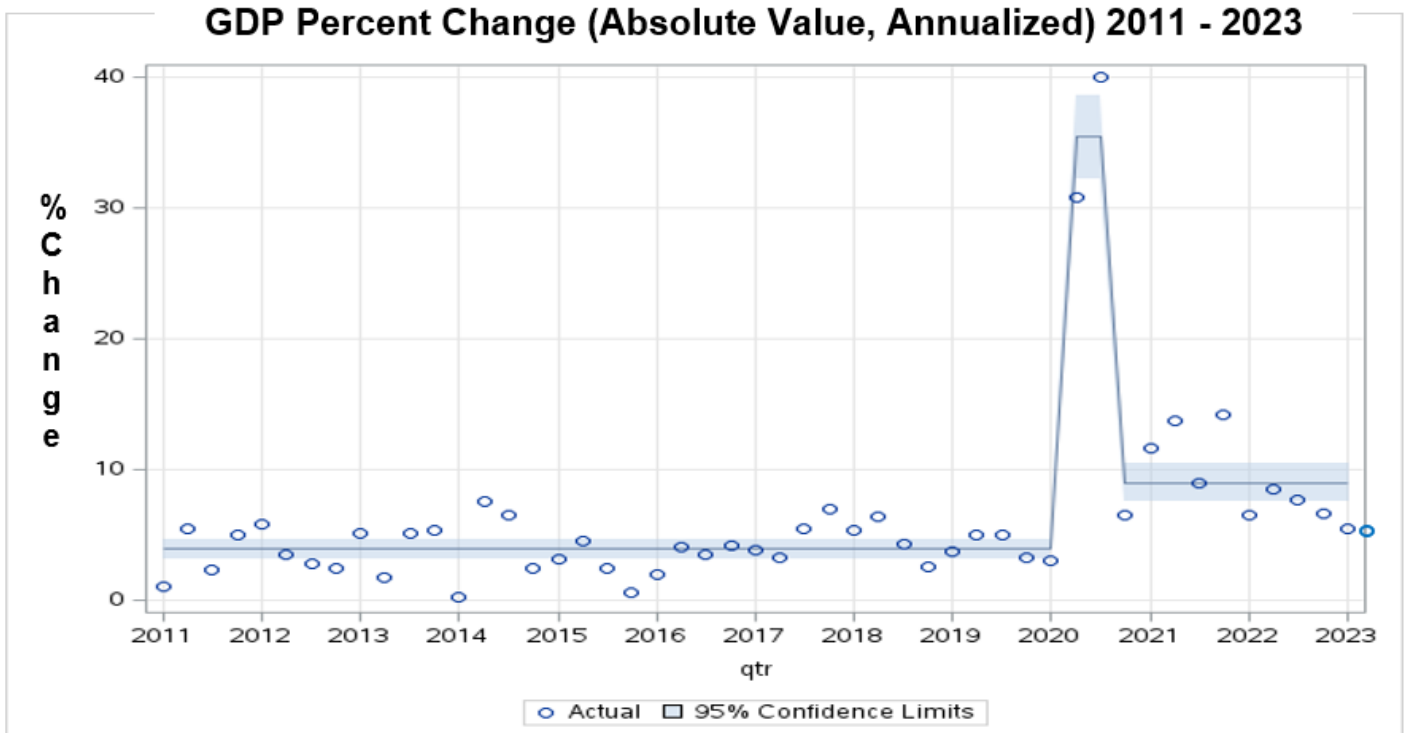
COVID-Era Baseline Changes: COVID Mortality Rate

```
proc ucm data=work.covid;
  id month interval=month;
   model covid_deaths = shift2022;
   irregular;
   level;
   estimate;
   forecast plot=decomp;
run;
```



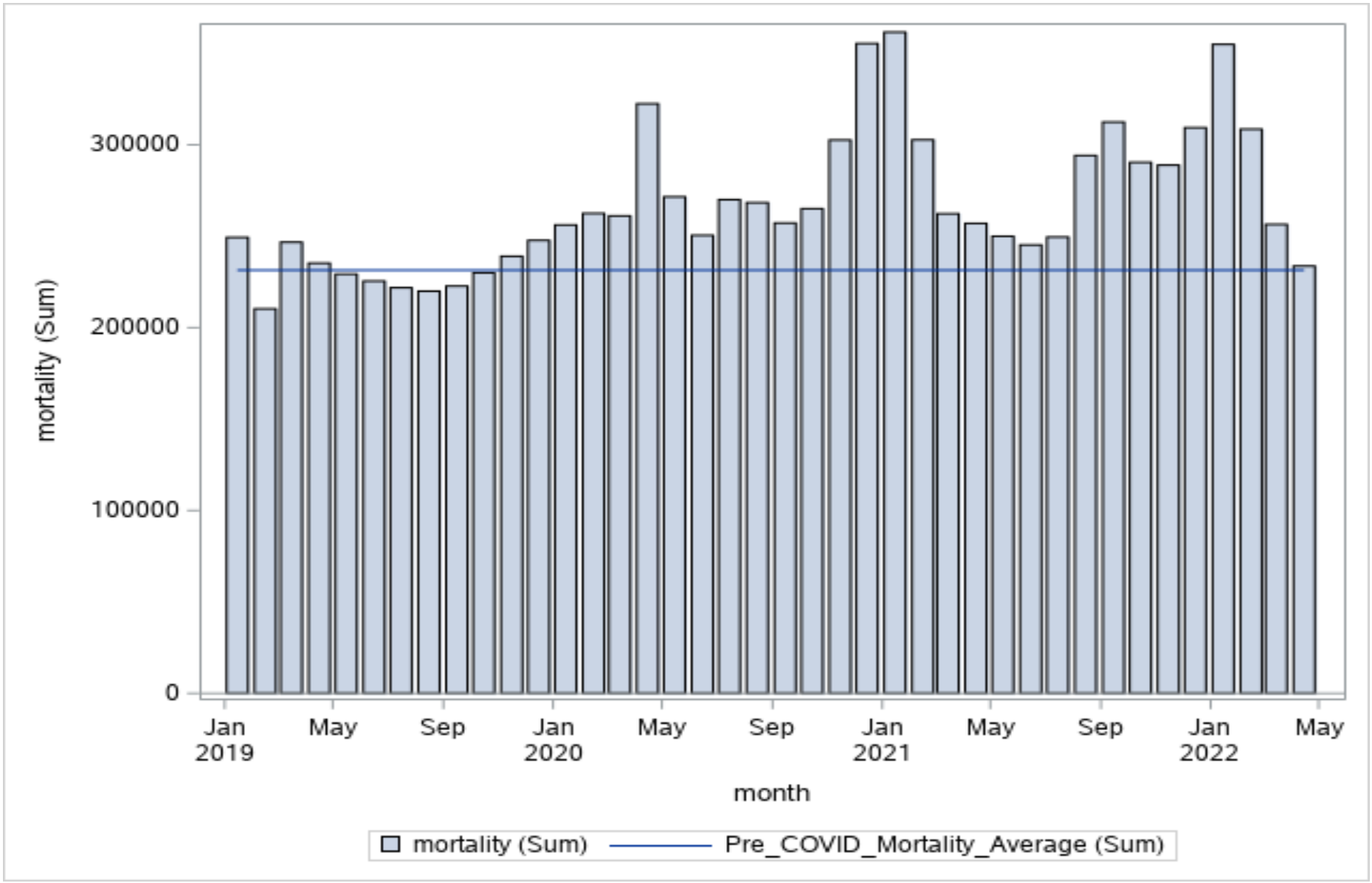
COVID-Era Baseline Changes: GDP

```
proc ucm data=tsa.gdp;
   id qtr interval=qtr;
   model GDP_ABS_Pct_Change = shift2020 shift2021;
   irregular;
   level;
   estimate;
   forecast plot=decomp;
   where qtr ge mdy(1,1,2011) and qtr le mdy(3,31,2023);
run;
```



LIMITATIONS OF UCM

Limitation of UCM: Rapidly Changing Non-Periodic Behavior



Limitations of Unobserved Components Models

- This method decomposes a time series into baseline (it's called "level"), trend ("slope", periodic ('seasonal") in irregular which is everything left. Where irregular dominates, the method isn't very informative – consider local regression

- Noisy or chaotic data often do not model well, as the components are difficult to distinguish

- Needs sufficient data in the time series following a change to the underlying behavior to accurately predict the new parameters e.g. a new baseline

CONCLUSIONS

Summary

- Unobserved Components models decomposes time series data into level, slope, periodic, and irregular components
- Through the use of a binary dummy variable, PROC UCM in SAS can estimate changes in baseline levels
- When changes in levels are numerous, large and irregular, UCM tends not to perform well – Local Regression is a better choice
- The COVID pandemic has ended but the characteristics of the US economy have moved towards new values dissimilar from before the start of the pandemic

References

CDC / NCHS / National Vital Statistics System, US Mortality Data by Year, https://www.cdc.gov/nchs/nvss/deaths.htm

Corliss, D., “Disproportional Impact of COVID-19 on Marginalized Communities”, Proc. SAS Global Forum, 2021

Federal Reserve Bank of St. Louis, GDP and Unemployment data, https://fred.stlouisfed.org/

SAS Institute, (2014), “SAS/ETS® 13.2 User’s Guide, The UCM Procedure

Selukar R (2011). “State Space Modeling Using SAS”. Journal of Statistical Software, 41(12), 1-13. URL http://www.jstatsoft.org/v41/i12/

Petris G, Petrone S (2011). “State Space Models in R”. Journal of Statistical Software, 41(4), 1-25. URL http://www.jstatsoft.org/v41/i04/.re”

QUESTIONS?

CONTACT INFORMATION

David J Corliss, PhD

Peace-Work

E-mail: davidjcorliss@peace-work.org