Unobserved Components Models: Applications in Post-COVID Analysis

David J Corliss, PhD is 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. Professionally, Dr. Corliss is managing director at Grafham Analytics, a data science consulting company.

PEACE-WORK

Unobserved Components Models: Applications in Post-COVID Analysis

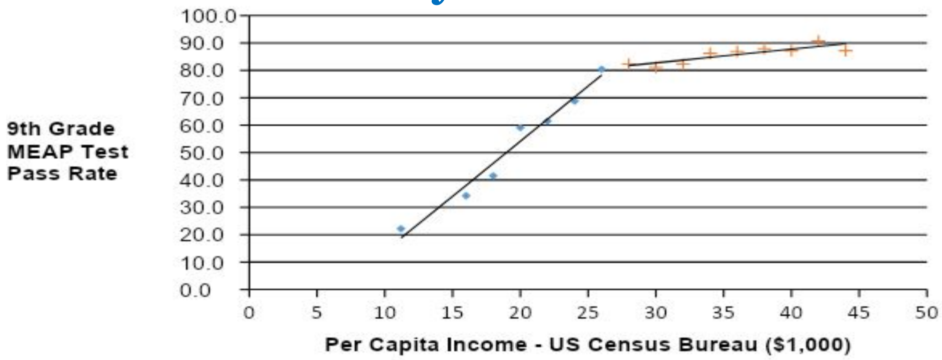
David J Corliss, Grafham Analytics

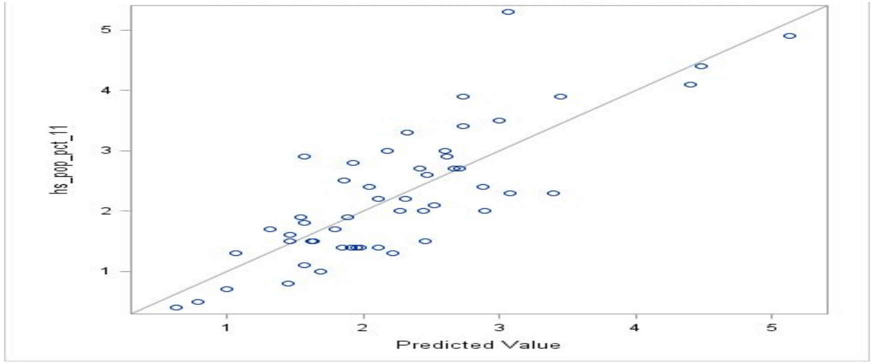
Modern Modeling Methods
University of Connecticut
June 27-28, 2023

PEACE-WORK

Peace-Work: Statistical Volunteers For A Cause

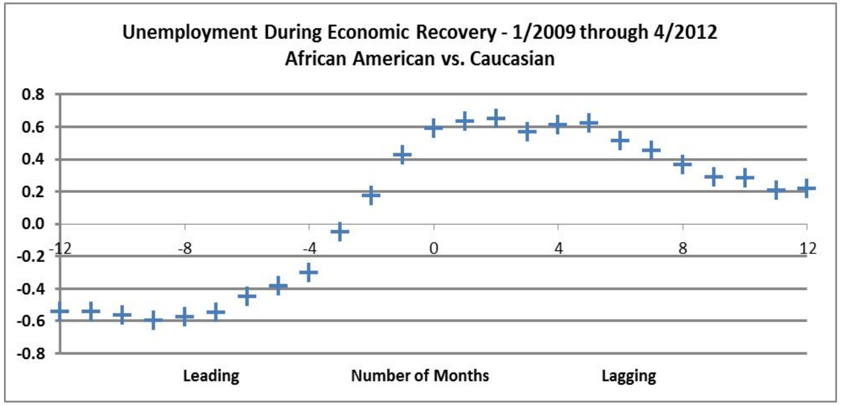
Poverty and Education





Homeless K-12 Students

Impact of Racial Bias



	Riga	Blissfield
Population	1,439	3,340
% under 18	27.2%	34.2%
Population under 18	391	1,142
% in poverty	5.2%	8.4%
Population in poverty	75	281
% under 18 in poverty	3.3%	10.0%
# under 18 in poverty	13	114
% over 65	14.0%	15.5%
Population over 65	201	518
% over 65 in poverty	6.3%	9.4%
# over 65 in poverty	13	49

Research and Fact-Checking

INTRODUCTION TO UNOBSERVED COMPONENTS MODELS

Unobserved Components Models

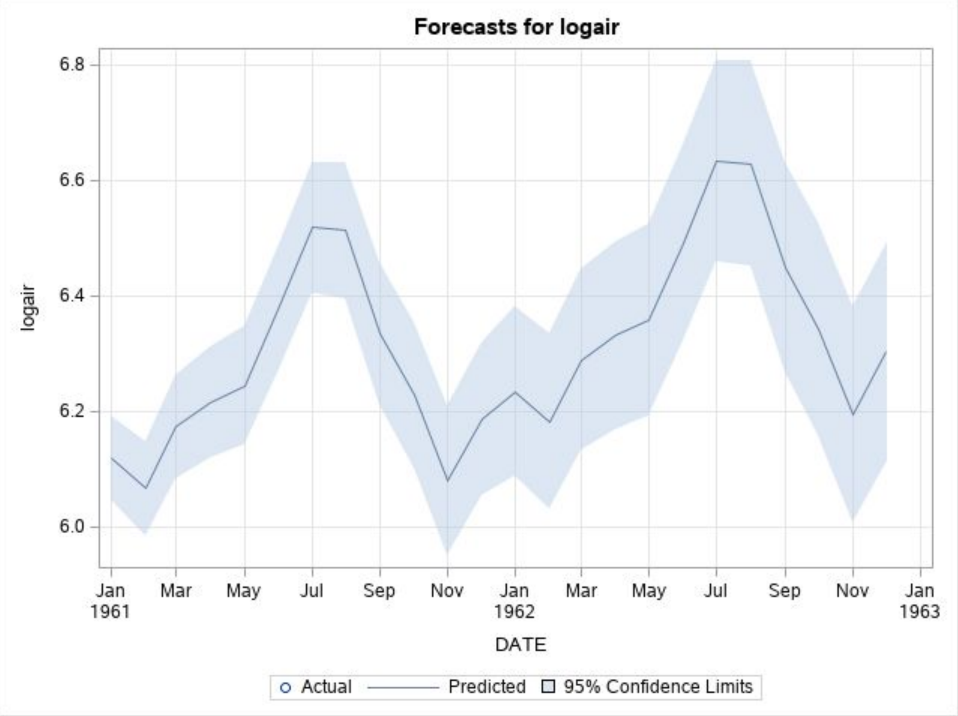
- Model Type: State Space Time Series Model, A. Harvey 1989
- Decomposes a time series into unobserved components that together form the time series, including trends, periodic behavior, and irregular components
- Supports measurement of changes in long-term baseline values of the time series => good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

PEACE-WORK

Unobserved Components Model Results, Output and Plots

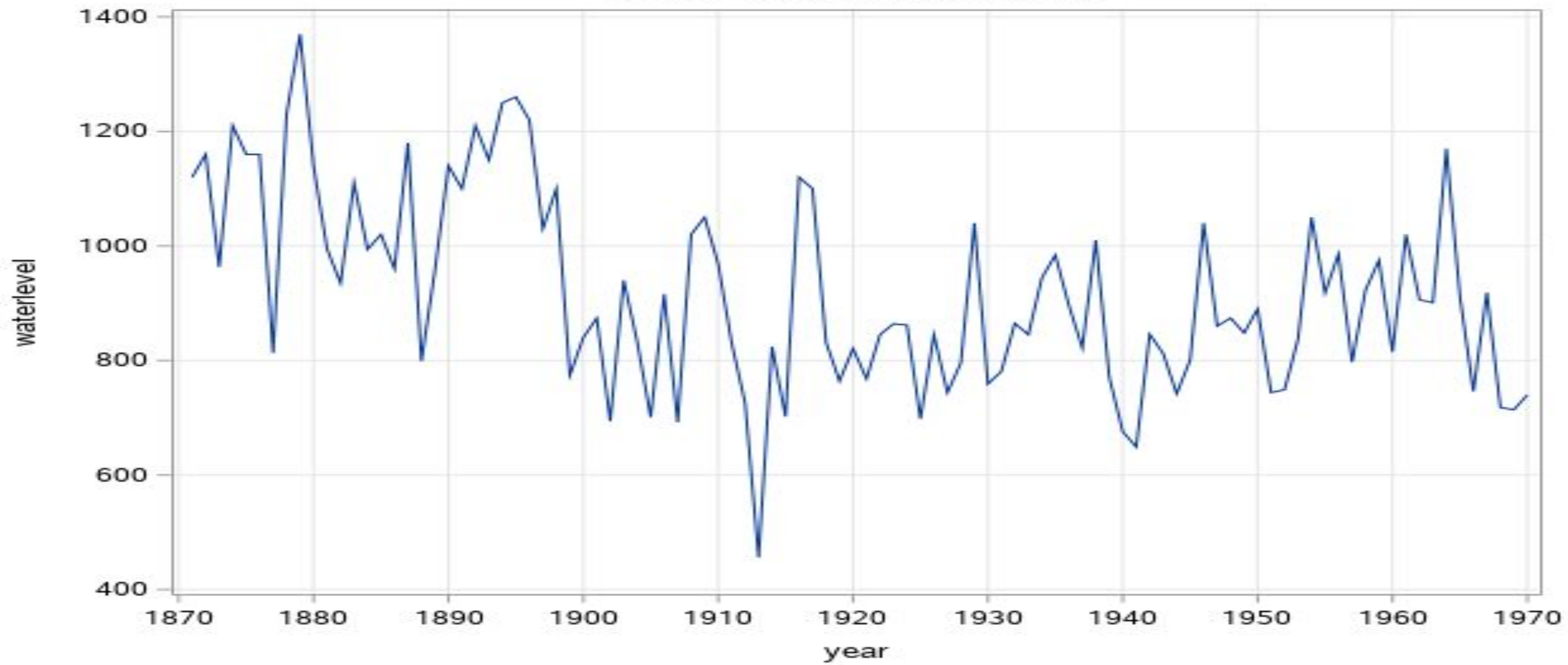
Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	0.00023436	0.0001079	2.17	0.0298
Level	Error Variance	0.00029828	0.0001057	2.82	0.0048
Slope	Error Variance	8.47922E-13	6.2271E-10	0.00	0.9989
Season	Error Variance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



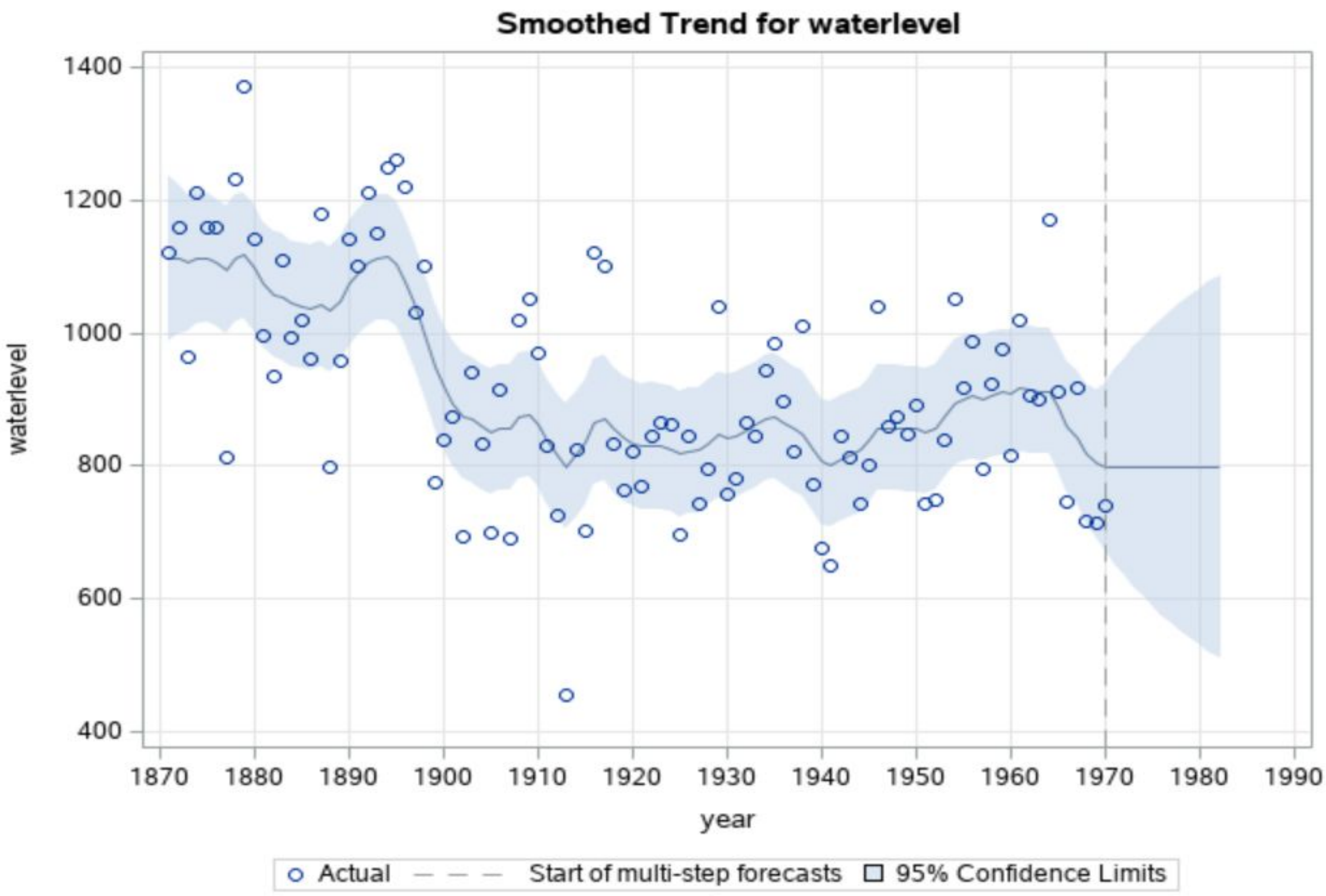
CHANGES IN BASELINE LEVELS WITH UCM

A UCM Classic Example: Depth of the Nile River



PEACE-WORK

A UCM Classic Example: How has the Depth Changed?

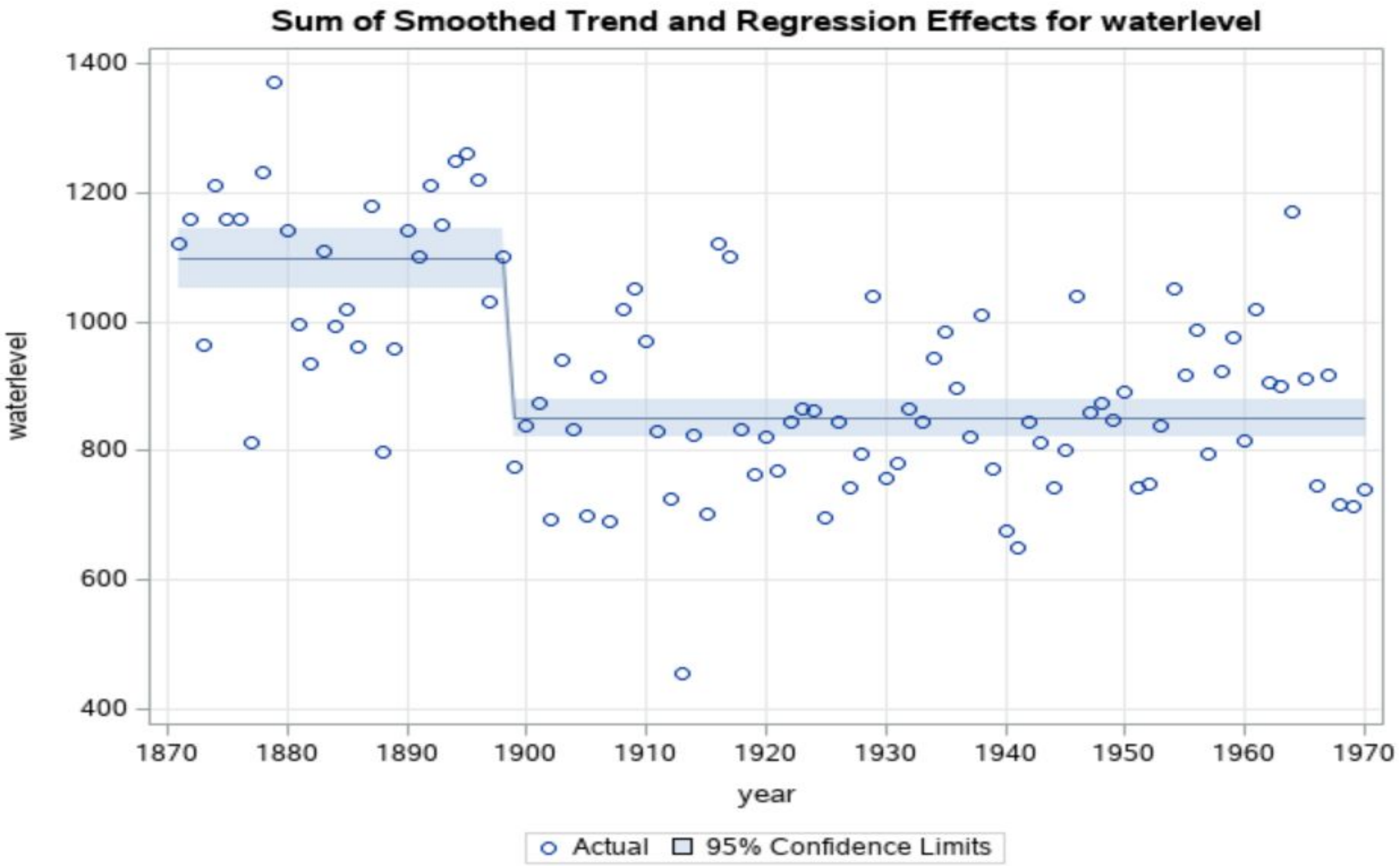


A UCM Classic Example: Depth of the Nile River

```
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;
```

PEACE-WORK

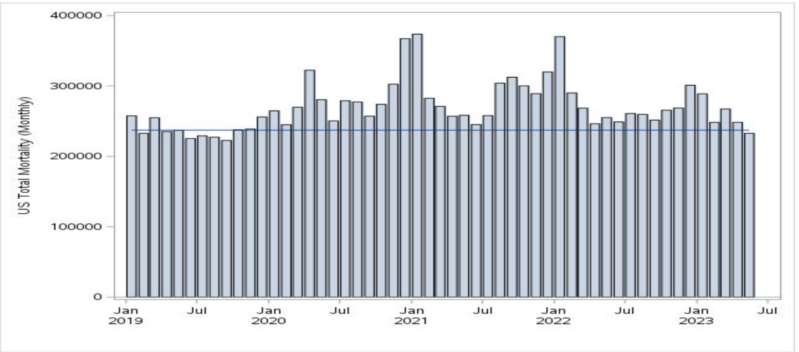
A UCM Classic Example: How has the Depth Changed?



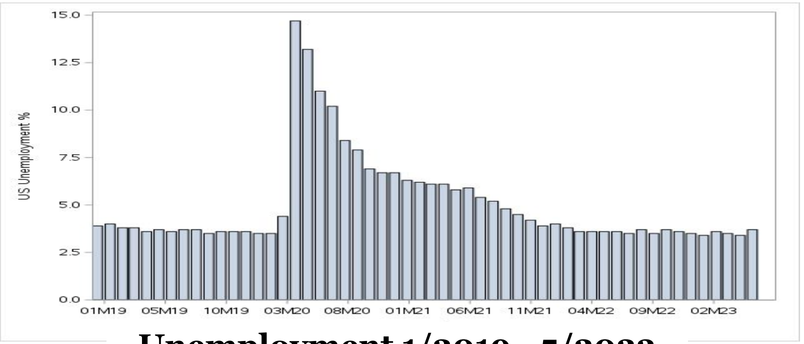
COVID QUESTIONS

COVID Data: A Complex Time Series

Series of Waves

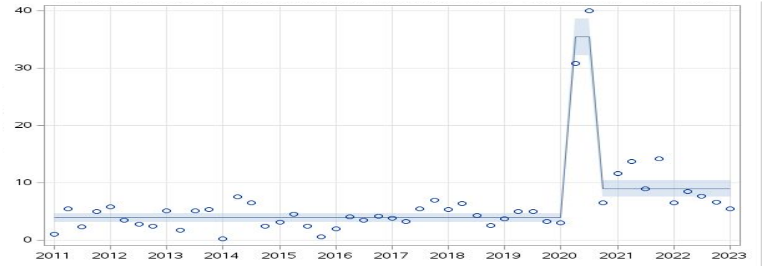


US Mortality 1/2019 - 5/2025



Unemployment 1/2019 - 5/2023

More than Medical

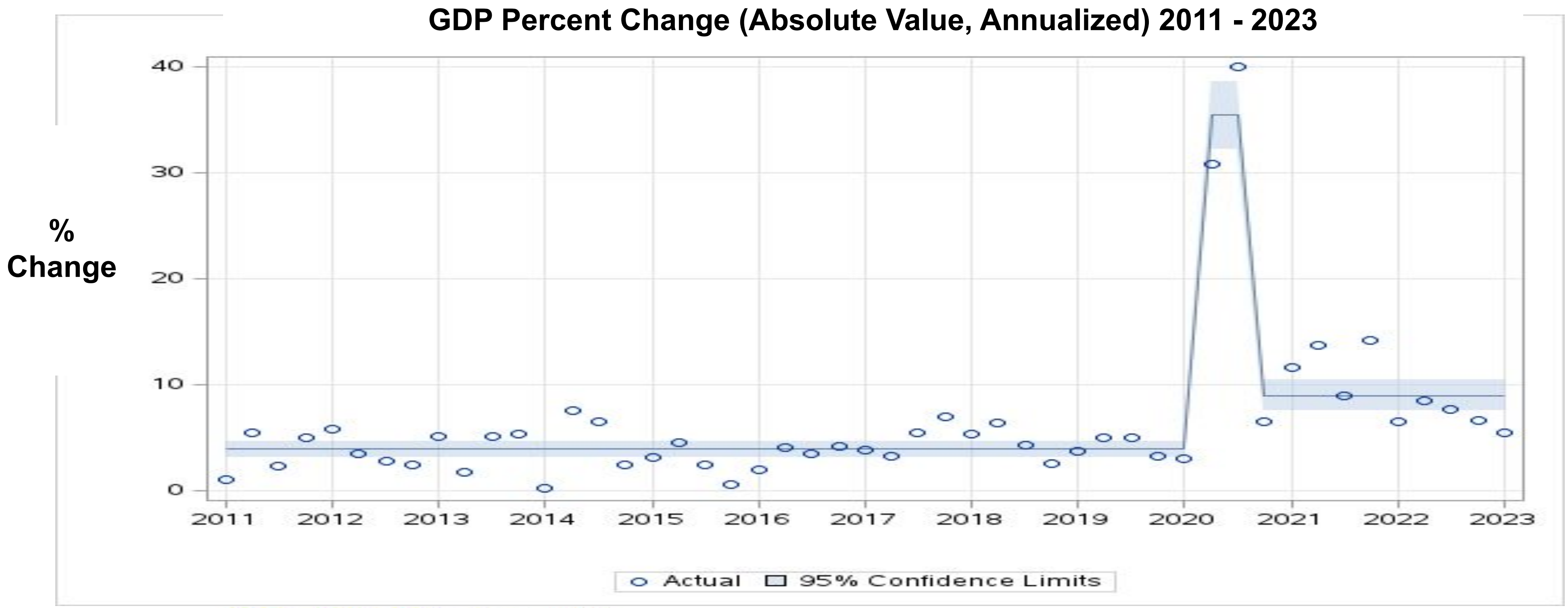


US GDP % Change Q1/2019 - Q1/2023

=> Unobserved Components

PEACE-WORK

COVID-Era Changes: GDP Baseline Level

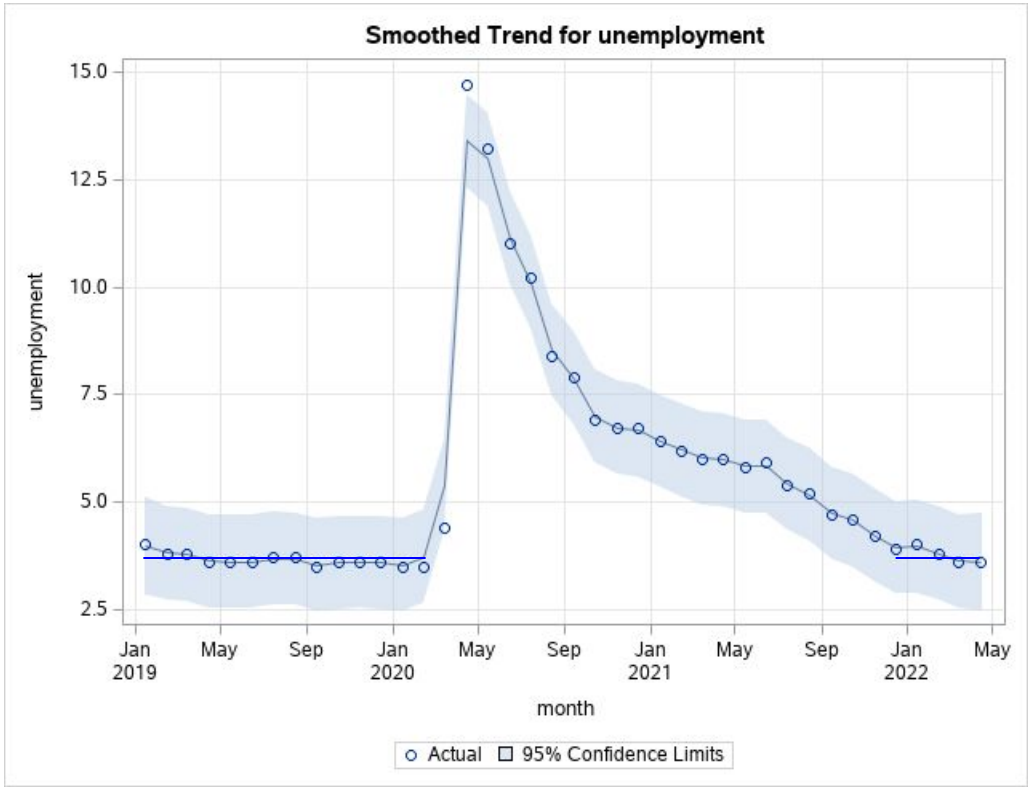



```
proc ucm data=tsa.gdp;
    id qtr interval=qtr;
    model GDP_ABS_Pct_Change = shift2020 shift2021;
    irregular;
    level;
    estimate;
    forecast plot=decomp;
    where qtr ge mdy(1,1,2011) and qtr le mdy(3,31,2023);
```

PEACE-WORK

COVID-Era Baseline Changes: Unemployment

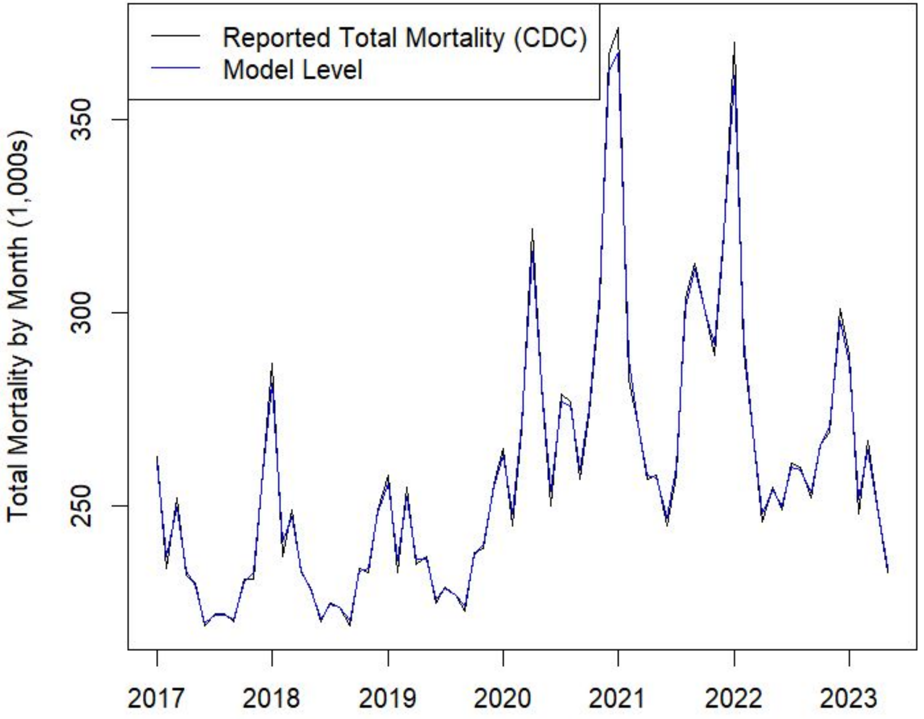
```
import statsmodels.api as sm
covid_ts['Date'] = pd.to_datetime(covid_ts['Date'])

# Unrestricted model, using string specification
unrestricted_model = {
    'level': 'local linear trend', 'cycle': True, 'damped_cycle':
True, 'stochastic_cycle': True}

# The restricted model forces a smooth trend
restricted_model = {
    'level': 'smooth trend', 'cycle': True, 'damped_cycle':
True, 'stochastic_cycle': True}

unemp_restricted_mod = sm.tsa.UnobservedComponents
(covid_ts['Unemployment_Pct'], **restricted_model)
unemp_restricted_res = unemp_restricted_mod.fit
(method='powell', disp=False)

unemployment_mod = sm.tsa.UnobservedComponents
(covid_ts['Unemployment_Pct'], **unrestricted_model)
unemployment_res = unemployment_mod.fit
(method='powell', disp=False)
```



PEACE-WORK

COVID-Era Changes: Deaths per Month

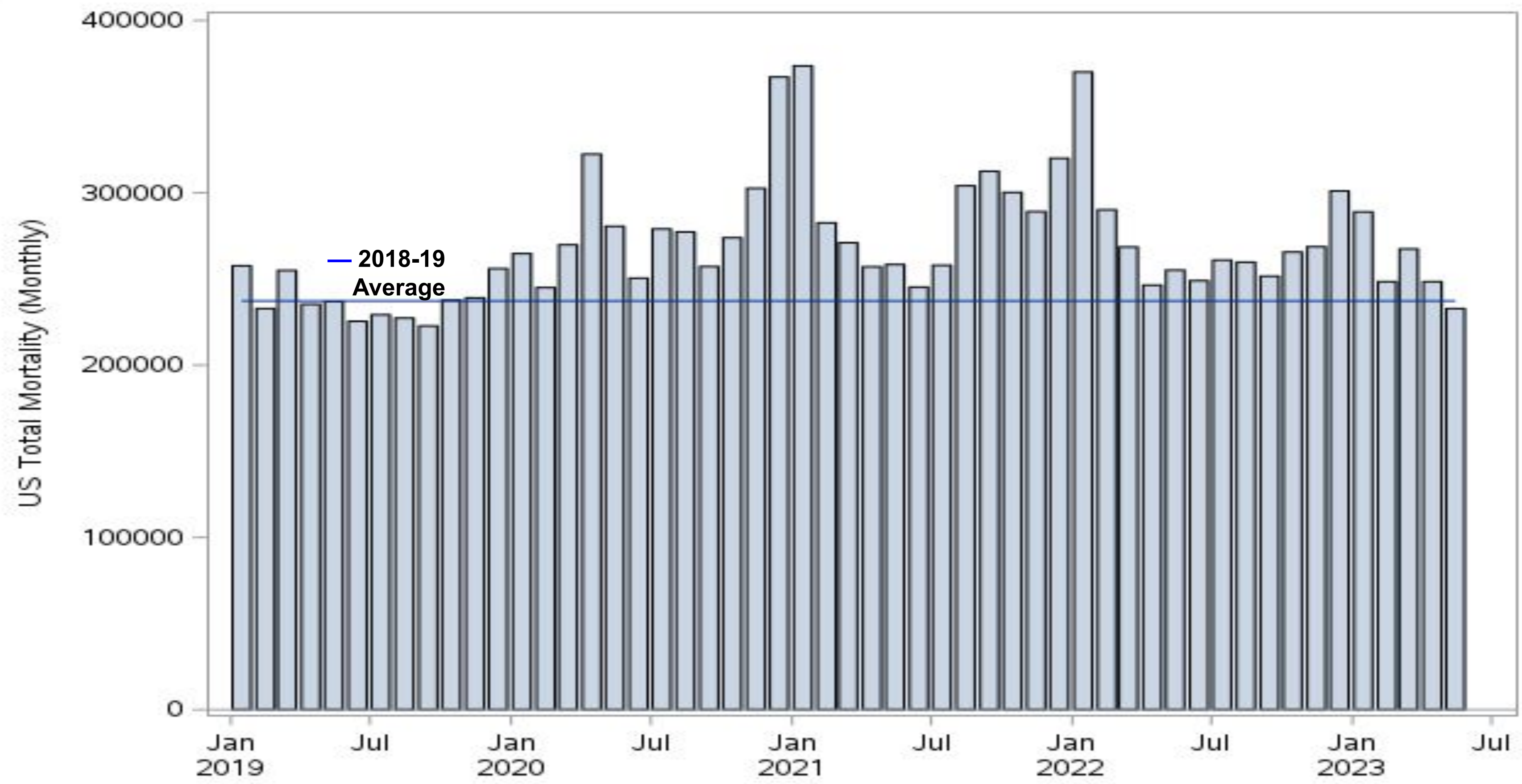
```
mortality <- c(
263,
234,
252,
.
.
.
267,
248,
233)

mortality_ts <- ts(mortality, start = c(2017, 1), frequency = 12)

ucm_mortality <- ucm(formula = mortality_ts~0, data =
mortality_ts, level=TRUE, slope=TRUE, irregular=TRUE)

plot(mortality_ts, ylab = "Total Mortality by Month (1,000s)")
lines(ucm_mortality$s.level, col = "blue")
legend("topleft", legend = c("Reported Total Mortality
(CDC)","Model Level"), col = c("black","blue"), lty = 1)
```



PEACE-WORK

LIMITATIONS OF UNOBSERVED COMPONENTS MODELS

Limitation of UCM: Rapidly Changing Non-Periodic Behavior



US Mortality 1/2019 - 5/2025

PEACE-WORK

UCM Limitations

- This method decomposes a time series into Baseline, Trend, and Periodic components, in addition to 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

- Following a change to some underlying behavior, UCM needs sufficient data in the time series to accurately predict the new parameters - for example, a new baseline level

PEACE-WORK

CONCLUSIONS

Summary

- Unobserved Components models decomposes time series data into level, slope, periodic, and irregular components

- Through the use of a binary dummy variable, Unobserved Components Models 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

- While the medical impacts have changed from pandemic to endemic, the non-medical effects of COVID continue to evolve

PEACE-WORK

References

Box, G. E. P., and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Rev. ed. San Francisco: Holden-Day

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

Chowdhury KR (2015). _rucm: Implementation of Unobserved Components Model (UCM)_. R package version 0.6, <https://CRAN.R-project.org/package=rucm>.

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/

Harvey, A. (1989), Forecasting, structural time series models and the Kalman filter. Cambridge New York: Cambridge University Press

Pant, V. (2020), Time Series Analysis using Unobserved Components Model in Python, Analytics Vidhya, https://medium.com/analytics-vidhya/tsa-ucm-python-5fde69d42e28

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

PEACE-WORK

Questions

?

PEACE-WORK

David J Corliss, PhD

Peace-Work

734.837.9323

davidjcorliss@peace-work.org

@dcorliss_astro

PEACE-WORK