SESUG 2023 - Paper 114

Unobserved Components Models: Applications in Post-COVID Analysis

David J Corliss

Peace-Work, Plymouth, MI

ABSTRACT

Unobserved Components is a type of State Space model used to detect and measure changes in a long-term baseline. The SAS/ETS procedure PROC UCM implements this method to decompose time series into components including long-term trend, periodic variations, and regression features. UCM is often used to analyze variations in baseline values due to change in the state of the system. PROC UCM is applied to the evolution of the COVID-19 pandemic to determine if various factors such as business conditions have returned to pre-pandemic levels. PROC UCM functionality, options, graphical output, and model interpretation.

INTRODUCTION TO Unobserved Components Models

Overview

An Unobserved Components Model (UCM) is a method of time series analysis that decomposes an observed time series into four latent components: Baseline Level, Trend, Periodic, and Irregular. When these unobserved components are added together, they produce the observed time series. While the underlying mathematics are coding are different, UCM is similar to Fourier Analysis in that it identifies the period and amplitude of cyclical behavior in the data while also producing a mean baseline and a trend line like a regression model. When the cycle in question is annual, the periodic component becomes a seasonality. When the periodic, baseline, and linear trend components are subtracted from the original observed time series, the unexplained remainder is termed the Irregular component. While it may be tempting to consider this to be noise, UCM makes no such claim: the irregular component is simply anything that is not periodic, baseline level, or a regression trend.

Unobserved Components Models were first developed by Andrew Harvey in 1989 for use in economic analysis. In SAS, the ETS procedure PROC UCM identifies the parameters of each component, calculates summary statistics, and produces many useful visualizations. In SAS, the components are named Level, Slope, Season, and Irregular. Despite the name of the Season component, it calculates the period and amplitude of any cyclical pattern without regard to the duration of the pattern. PROC UCM supports identifying any or all of the Level, Slope, and Season components in a time series. This SAS procedure also greatly facilitates measurement of changes in baseline level, trend, and periodic behavior, making especially useful for studying the impact of important events affecting a time series.

An Introductory Example

The SAS Help Air dataset provides a time series of monthly airline data from January 1969 – December 1960. This dataset is often used in teaching SAS ETS procedures and offers an excellent example for PROC UCM. In this example, the log of the total volume of flights, designated logair, is used to control the effect of outliers in the data:

```
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;
```

In this example, the titular unobserved components in which the series will be decomposed are Irregular, Level, Slope, and Season. It will be noted that each component has its own statement in the procedure and are individual components are not required to be specified. It is recommended to use all four initially and then remove components not playing a role in the model. id gives the name of the date / time variable in the data and interval is the time between observations. forecast specifies the lead, that is, the number of periods ahead to include in the forecast. As with other time series procedures, it is necessary to use a complete dataset with data for every interval: missing values at any point in time must be imputed. The SAS output provides parameter estimates for each of the components (Figure 1). The forecast is compared the actual values in Figure 2. In this example, the statistical output shows these data do not have any measurable Slope component. With this test result, the model should be run again without a Slope statement.

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	

Fig. 1. Parameter Estimates and statistical tests for an Unobserved Components Model for the log of monthly number of flights in the US in the Sashelp.air dataset

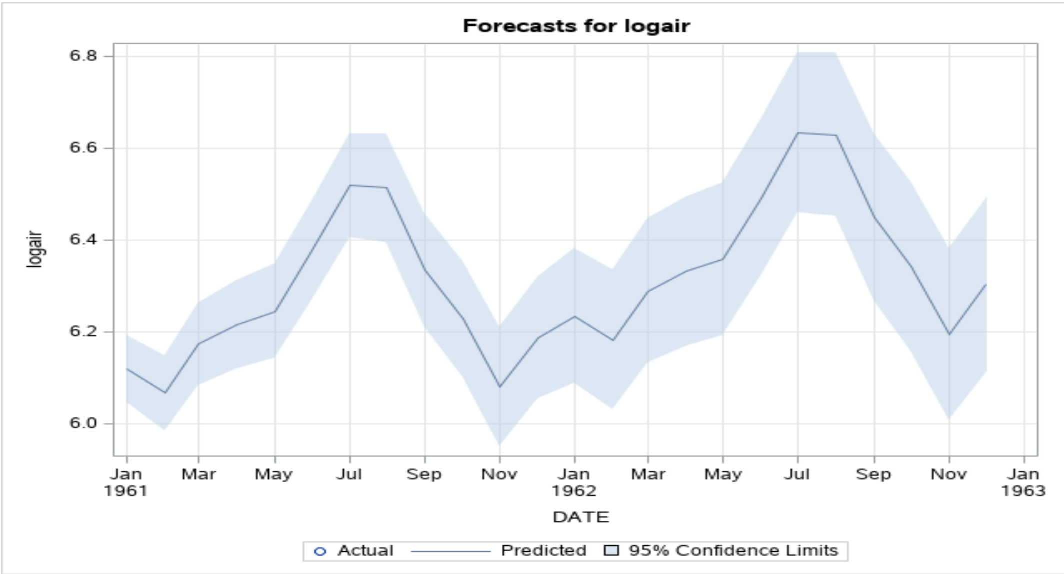

Fig. 2. UCM forecast of the monthly number of flights in the US in the Sashelp.air dataset

Types Measuring Changes in Baseline Vaues

Depth on the Nile

A common example given in introducing UCM is a study of the water level on the Nile River, which was measured annually at Aswan in southern Egypt from 1871 to 1970 (Cobb 1978). In Figure 3, PROC TIMESERIES is used to plot the data over this 100-year period.

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

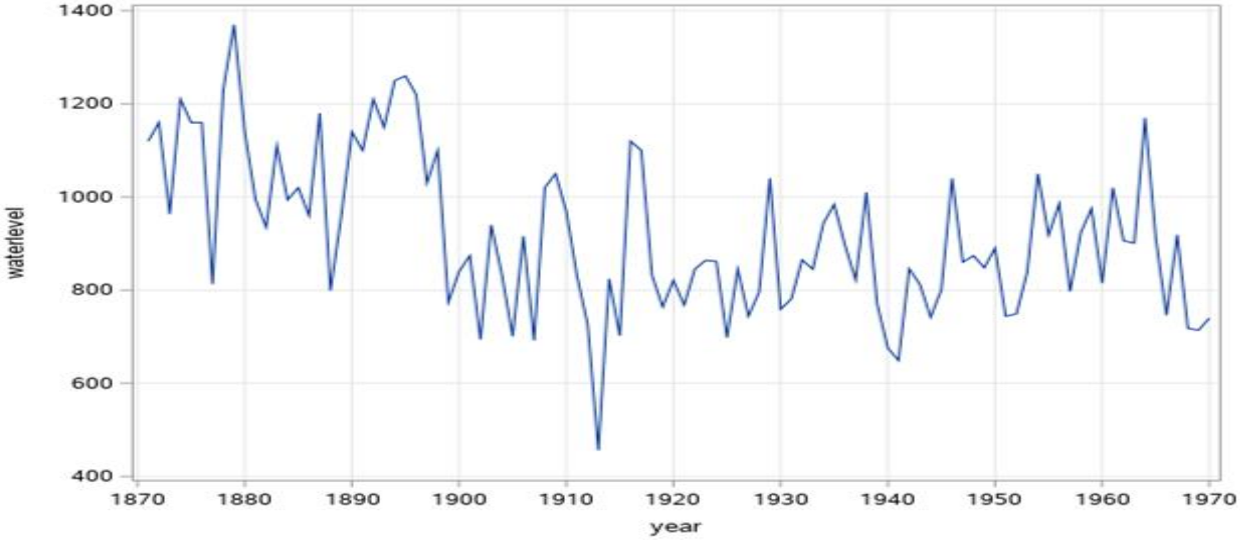

Fig. 3. Water level of the Nile River, measured at Aswan annually on January 1 from 1871-1970

It is apparent from the plot of the data in Figure 3 that some event in the late 1890's resulting in a change of the long-term baseline. The UMC procedure estimates the amount of such changes in the characteristics of a time series, including the baseline level as in this or, or the slope / regression linear trend or the period and amplitude of a cyclical pattern. An unobserved components model was used to model these data, using only the level and irregular components (Figure 4).

```
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	

Fig. 4. UCM model of the water level of the Nile River from 1871 - 1970

In this example, there is a hint of a change in the baseline level but the statistical parameters so substantial uncertainty for this component. PROC UCM creates a plot of the smoothed trend which can be used to estimate the point in time where the change occurs.

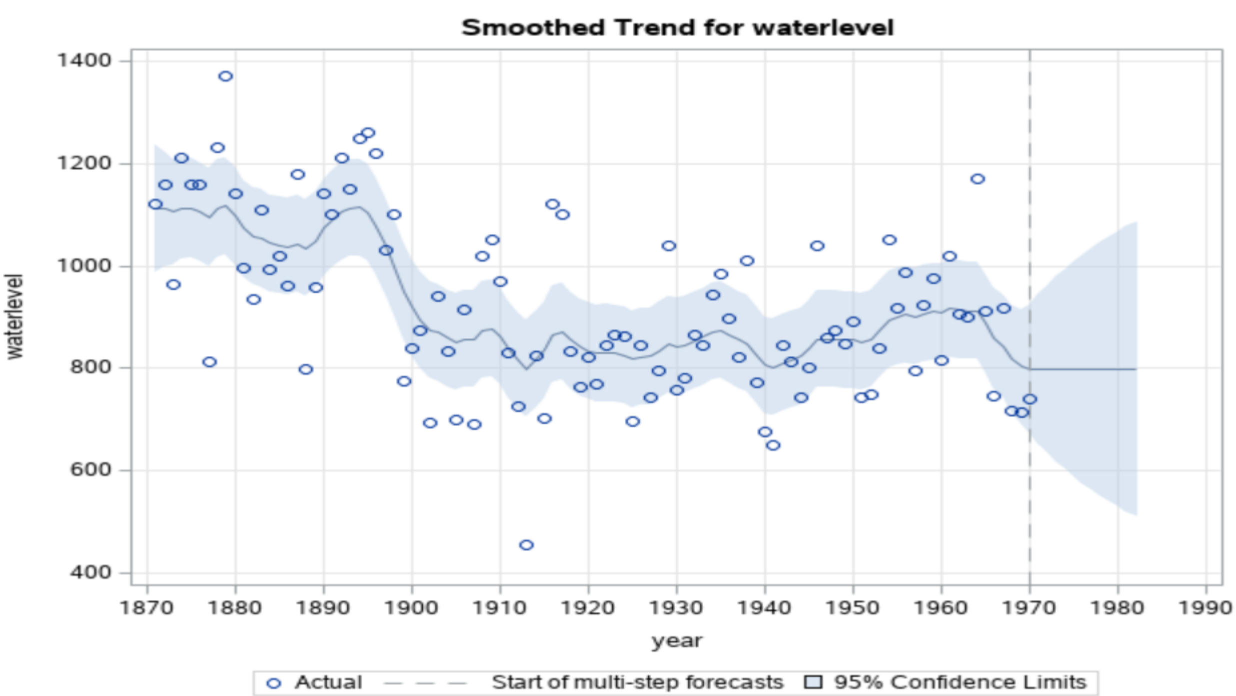

Fig. 5. PROC UCM: Plot of the smoothed trend of the water level of the Nile River

Time series cluster analysis (Corliss 2012) was used to identify1899 as the year that best separates the previous pattern from the new level established after this event.

Estimating a Discrete Change in UCM Component Parameters

Once the point in time best distinguishing the previous from the new pattern, the amount of the change can be estimated. This is done using a binary variable specifying the point in time. The new variable is then added as a predictor in the Model statement of the UCM:

```
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;
```

The resulting UCM clearly distinguishes the distinguishing the change in in the model characteristic – in this case, the baseline level (Figure 6). It is noted from the plot that the 95% confidence level intervals clearly do not overlap, offer strong evidence to reject a null hypothesis that no real change occurred.

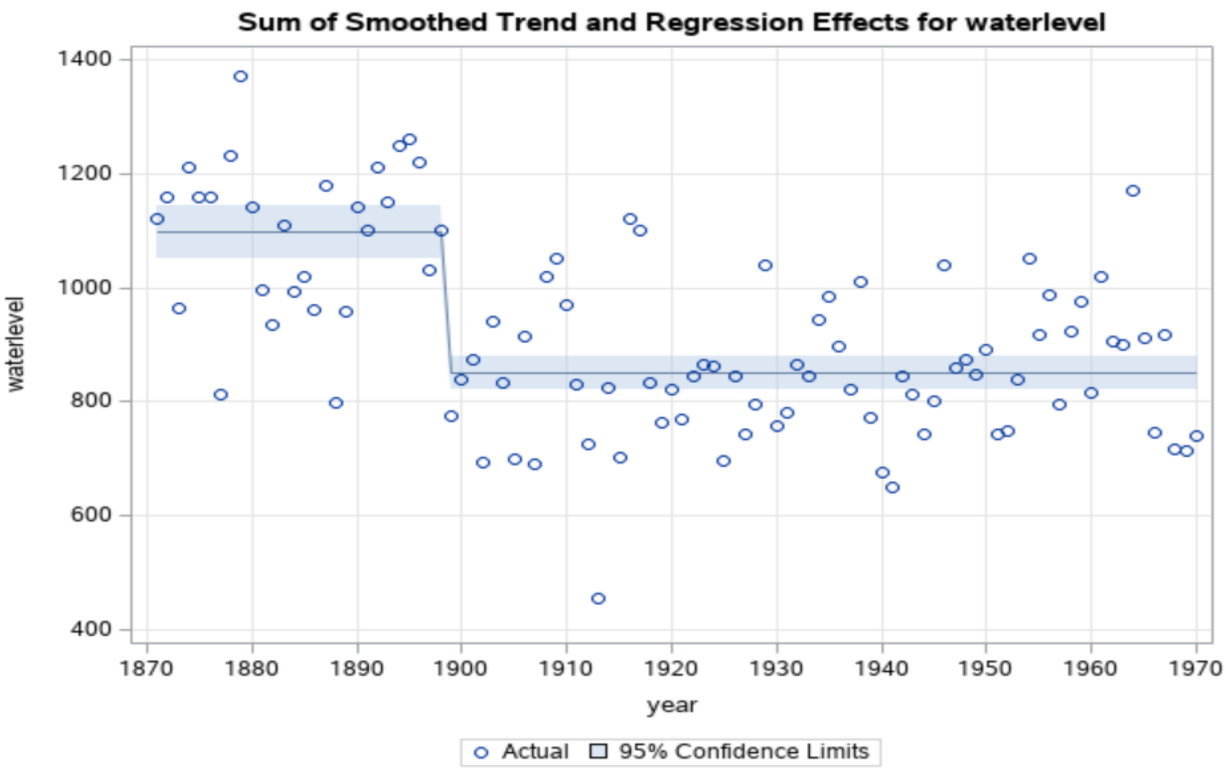

Fig. 6. PROC UCM: Estimating the change in the water level of the Nile River starting in 1899

This ability to identify and estimate point in time changes in the characteristics of a time series makes UCM and effective tool for measuring the impact of other events, such as the COVID-19 pandemic.

The work of Russell Lavery (Lavery 2013) gives a much more detailed description of the basics of PROC UCM and its applications. The focus of this paper is using USM to model impacts of the COVID-91 pandemic.

UCM Modeling of the Impacts of the COVID-19 Pandemic

Economic Impacts: Unemployment

Unobserved Components was originally developed to model changes in econometric time series. One such impact of the COVID pandemic was sharp increase in unemployment due to shutdowns at the start of the pandemic.

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

Figure 7 plots the change in the baseline unemployment level in the US, a key indicator of the economic impact of the pandemic.

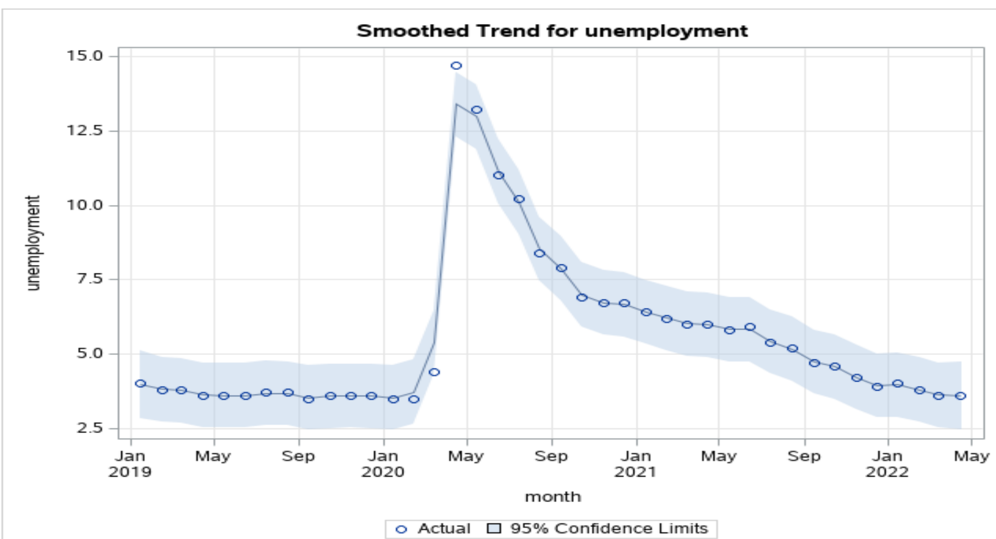

Fig. 7. PROC UCM: Estimating the change in the US unemployment level starting in March, 2023

Forecasting: COVID Mortality

These models can also be used to make forecasts following a high-impact event, as in this analysis of the times series of COVID mortality rates during the pandemic (Figure 8). In this analysis, the time series is tested for the end of the pandemic using start of 2022 as the key date. In this case, UCM did not find a new baseline; the failure of this test indicates mortality rates had not leveled off and the pandemic had not yet ended as of the test date.

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

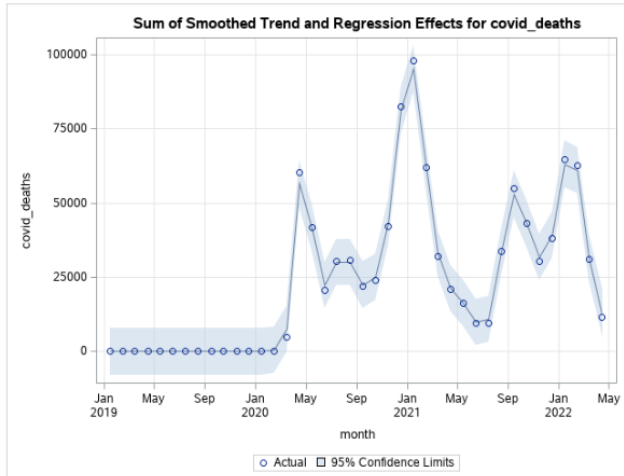

Fig. 8. PROC UCM: Evolving forecast of US COVID mortality rates during the pandemic

Emergence of a New Normal: Change in GDP

The ability of UCM to model changes in the characteristics of a time series can also be used to the emergence of a new baseline level, which can be termed a "New Normal". This can be analyzed using two binary predictive variables: one for the start of the event and one for the end. An application of this during the COVID is modeling the increase in volatility of the economy as a whole caused by the pandemic. One measure of this property is the absolute value of the quarterly percent change in the GDP, which saw wide swings during the pandemic after a long (2010-2019) pattern of gradual recovery from the Great Recession.

```
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;
```

Fig. 8. PROC UCM: Evolving forecast of US COVID mortality rates during the pandemic

As widespread shutdowns due to the pandemic began abound March 18, 2020, this model of the economic volatility uses a binary predictive variable of Q1 2020 as the start of the economic effects of the pandemic and a second variable set to Q1 2022 (Figure 9).

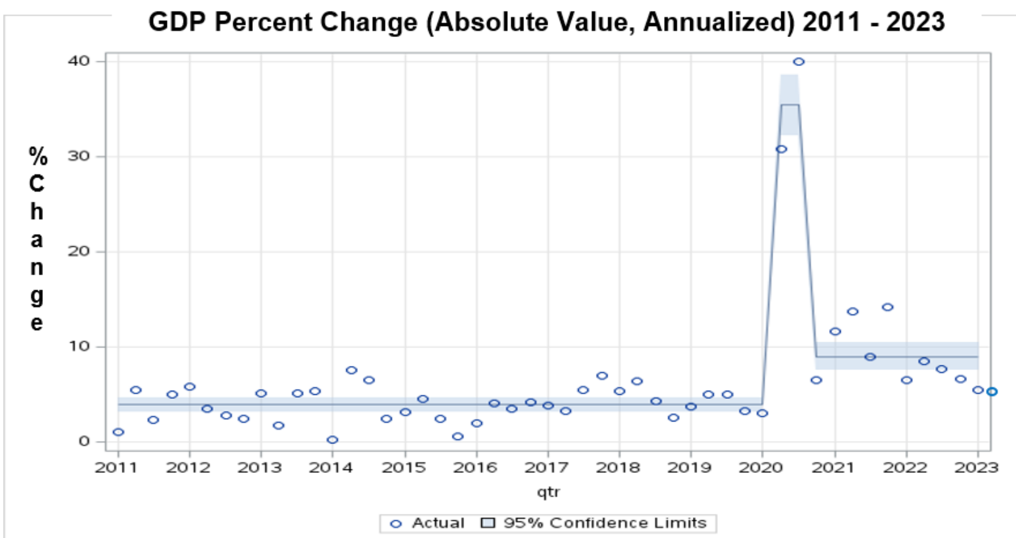

Fig. 9. Economic volatility in the US before and during the COVID-19 pandemic, with evidence of an emerging "New Normal"

During This analysis indicates a fairly short period of extreme economic disruption followed by the emergence of a new baseline level that somewhat higher than that the relatively amount of low turbulence during previous period marked by gradually recovery from the Great Recession.

Limitations of Unobserved Components Models

Like any analytic method, Unobserved Components has important limitations to keep in mind when considering and using this technique. Figure 10 gives an example of data that can be difficult to model with UCM due to rapidly changing non-period behavior.

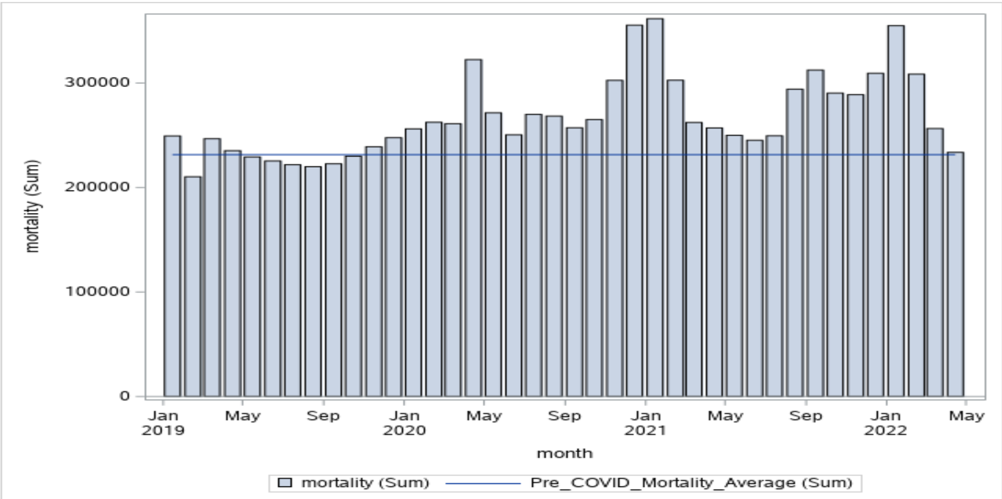

Fig. 10. Time Series of the COVID-19 Mortality Rate.

UCM successfully modeled the beginning and end of the pandemic in terms of mortality and measured the increase in deaths as a whole. However, the technique is not well suited to identifying short-term trends and did not produce good short-range forecasts for mortality. Best practices for using Unobserved Components Models include:

- UCM decomposes a time series into baseline (described by the Level statement is SAS), trend (Slope), periodic (Seasonal) and an Irregular category including everything not described by the others. Where irregular component 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
- UCM 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

Unobserved Components Models decompose time series into baseline level, regression, and cyclical components. In SAS, these features are modeled using PROC UCM, where there are referred to as the Level, Slope, and Season. The remainder of effects not described by these components is called Irregular. These models are effective at identifying and estimating the value of changes in long-term patterns in a time series. Through the use of a binary dummy variable, PROC UCM in SAS can estimate changes in baseline levels and other characteristics of a time series. This makes UCM useful in economic analysis to identify the characteristic properties of a time series impact by important point in time events, including the emergence of a "New Normal" following such an event. Unobserved Components Models support a wide variety of applications in many areas, including econometrics, modeling physical processes, and other time series.

The ability to model changes in baseline levels and identify a New Normal makes UCM an important method for modeling the COVID-19 pandemic and its impacts in biostatistics, econometrics, and social sciences. Analysis of the most recent data indicate the COVID pandemic has ended but the characteristics of the US economy have moved towards new values dissimilar from immediately before the start of the pandemic.

When changes in levels are numerous, large and irregular, UCM tends not to perform well. In these circumstances, Local Regression may represent a better choice for model development.

ACKNOWLEDGEMENTS

Thanks are due to the SAS Institute for making available SAS OnDemand for Academics at no cost to for noncommercial use, including this project. Thanks are due to Nancy Morrison for many helpful conversations and comments on applications of unobserved components models.

REFERENCES

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

Cobb, G. W. (1978). "The Problem of the Nile: Conditional Solution to a Change Point Problem." Biometrika 65:243–251.

Corliss, D., "Dynamically Evolving Systems: Cluster Analysis Using Time", Proc. SAS Global Forum, 2012

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/

De Jong, P., and Penzer, J. (1998). "Diagnosing Shocks in Time Series." Journal of the American Statistical Association 93:796–806.

Harvey, A.C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press

Lavery, R. (2013), “An Animated Guide©: Proc UCM (Unobserved Components Model)”, NESUG 2013, www.lexjansen.com/nesug/nesug04/an/an03.pdf

SAS Institute, (2014), “SAS/ETS® 13.2 User’s Guide, The UCM Procedure”, SAS Institute Inc., Cary, NC

SAS Institute, (2018), “Sashelp Data Sets”, SAS Institute Inc., Cary, NC

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”

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David J Corliss
Peace-Work
davidjcorliss@peace-work.org