Analysis of Time Series

Chapter 9: Case study: An association between unemployment and mortality?

Edward L. Ionides

Outline

1. Historical background
2. Supplementary analysis
   - AIC

Historical investigations linking business cycles to mortality

- Ogburn and Thomas (1922) were among the first to report **pro-cyclical** mortality: death rates are statistically above trend when economic activity is above trend.
- Procyclical mortality, if it exists, shows that a key measure of population health is worse in economic booms than in recessions.
- Both the economy and life expectancy have grown over the last century. However, these phenomena have not always occurred simultaneously. For example, 1950–1980 saw rapid growth in life expectancy in India and China, combined with relatively slow economic growth. Improvement in life expectancy has slowed during their recent economic surges.
- The link between economic growth and health improvement is controversial, since it has political implications. Economists and epidemiologists have argued both sides of this debate, using time series methods.

Implications of pro-cyclical mortality

- If our goal is population health and happiness, how much should our policies focus on gross domestic product (GDP) growth?
- Evidence supporting the view that economic growth is the critical engine for other improvements in living conditions would make a moral argument in favor of economic growth.
- Evidence that there are other major factors involved in improving living conditions suggest that economic growth should be only one political consideration, among others.

A time series of life expectancy in the USA

```
e_data <- read.table(file="life_expectancy_usa.csv",header=TRUE)
head(e_data,n=4)

 Year   e0F   e0M    e0
 1933 62.80 59.19 60.90
 1934 62.30 58.29 60.19
 1935 63.05 58.98 60.91
 1936 62.60 58.35 60.35
```

- Data are from the Human Mortality Database.
- e0 is **life expectancy at birth (LEB)** for civilians,
- `e0F` and `e0M` are LEB for females and males, but we focus on `e0`.
- LEB is an actuarial calculation based on a fictitious individual having mortality rates at each age matching census age-specific mortality rates for the current year.
- LEB is a standard way to combine all the age-specific mortality rates into a single number.

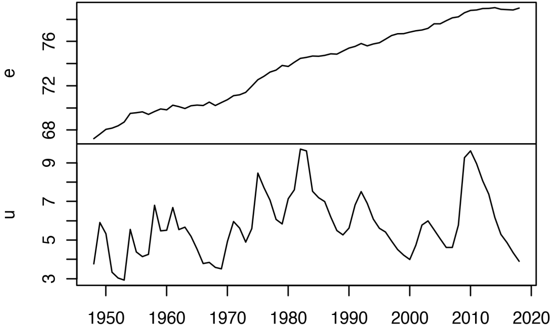
```
u_data <- read.table(file="unadjusted_unemployment.csv",
  sep=",",header=TRUE)
head(u_data,4)

 Year Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
 1948 4.0 4.7 4.5 4.0 3.4 3.9 3.9 3.6 3.4 2.9 3.3 3.6
 1949 5.0 5.8 5.6 5.4 5.7 6.4 7.0 6.3 5.9 6.1 5.7 6.0
 1950 7.6 7.9 7.1 6.0 5.3 5.6 5.3 4.1 4.0 3.3 3.8 3.9
 1951 4.4 4.2 3.8 3.2 2.9 3.4 3.3 2.9 3.0 2.8 3.2 2.9
```

- We consider unadjusted unemployment from Bureau of Labor Statistics.
- Unemployment is just one component of the state of the economy. One could consider other measurements.
- Write e_n for life expectancy in year $t_n = 1947 + n$.
- Write u_n for mean unemployment in year t_n.

A time plot of the raw data

```
t <- intersect(e_data$Year,u_data$Year)
e <- e_data$e0[e_data$Year %in% t]
u <- apply(u_data[u_data$Year %in% t, 2:13],1,mean)
```



- We are interested in changes over business cycle timescales, once trends have been removed.

The Hodrick-Prescott filter

- To extract the cyclical component, we use an econometric method: the **Hodrick-Prescott (HP) filter** (Hodrick and Prescott, 1997).
- Specifically, for a time series $y^*_{1:N}$, the HP filter is the time series $s^*_{1:N}$ constructed as

$$s^*_{1:N} = \underset{s_{1:N}}{\arg\min} \left\{ \sum_{n=1}^{N} \big(y^*_n - s_n\big)^2 + \lambda \sum_{n=2}^{N-1} \big(s_{n+1} - 2s_n + s_{n-1}\big)^2 \right\}. \quad (1)$$

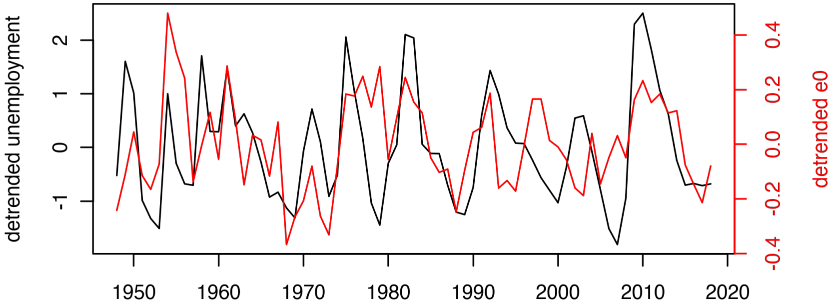
- The HP filter is a **smoothing spline**. Later, we see it can also be viewed as a state space model.
- A standard econometric choice of λ for removing nonlinear trend, for extracting the business cycle component, in annual data is $\lambda = 100$.
- An R implementation of the Hodrick-Prescott filter is `hpfilter` in the R package `mFilter`.

- We use the Hodrick-Prescott filter to define the HP-detrended life expectancy, $e^{HP}_{1:N}$, and unemployment, $u^{HP}_{1:N}$.

```
library(mFilter)
e_hp <- hpfilter(e, freq=100,type="lambda",drift=F)$cycle
u_hp <- hpfilter(u, freq=100,type="lambda",drift=F)$cycle
```

- Plotting two time series on a single graph is not always advisable, but here it is helpful.

```
plot(t,u_hp,type="l",xlab="",ylab="detrended unemployment")
par(new=TRUE)
plot(t,e_hp,col="red",type="l",axes=FALSE,xlab="",ylab="")
axis(side=4, col="red",col.ticks="red",col.axis="red")
mtext("detrended e0",side=4,col="red",line=3)
```



Detrended unemployment (black; left axis) and detrended life expectancy at birth (red; right axis).

- Looking at this figure may suggest that detrended life expectancy and detrended unemployment cycle together.
- We make a formal statistical test to check our eyes are not deceiving us.

A regression with ARMA errors model

- We can investigate the dependence of $e^{HP}_{1:N}$ on $u^{HP}_{1:N}$ using a regression with ARMA errors model,

$$E^{HP}_n = \alpha + \beta u^{HP}_n + \epsilon_n, \qquad (2)$$

where $\{\epsilon_n\}$ is a Gaussian ARMA process. We use an ARMA(1,0) model, as discussed in the supplementary analysis.

```
arima(e_hp,xreg=u_hp,order=c(1,0,0))


Call:
arima(x = e_hp, order = c(1, 0, 0), xreg = u_hp)

Coefficients:
         ar1  intercept    u_hp
      0.4798    -0.0030  0.0713
s.e.  0.1042     0.0304  0.0181

sigma^2 estimated as 0.01825:  log likelihood = 41.26,  aic = -74.52
```

- The standard errors, computed from the observed Fisher information approximation, suggest a statistically significant association between cyclical variation in unemployment and mortality.
- We can also compute a p-value from a likelihood ratio test,

```
log_lik_ratio <- as.numeric(
   logLik(arima(e_hp,xreg=u_hp,order=c(1,0,0))) -
   logLik(arima(e_hp,order=c(1,0,0)))
)
LRT_pval <- 1-pchisq(2*log_lik_ratio,df=1)
```

- This gives a p-value of 0.00018.
- This is clear statistical evidence for a positive association between unemployment and life expectancy.
- As with all observational studies, interpretation of association need care.

Association and causation

- We have been careful to talk about **association**, since observational data giving firm statistical evidence of an assocation between X and Y cannot readily distinguish between three possibilities:
  1. X causes Y.
  2. Y causes X.
  3. Both X and Y are caused by a third variable Z that is unmeasured or has been omitted from the analysis. In this case, Z is called a **confounding variable**.
- Here, it is not considered plausible that mortality fluctuations drive economic fluctuations (the **reverse causation** possibility).
- We think of unemployment as a **proxy variable** for economic fluctuations. We do not claim that increased unemployment itself is necessarily directly causing reduced mortality. Any other omitted variable that fluctuates with levels of economic activity should show similar associations.

Evidence for causality

- A problematic potential confounding variable is lagged economic activity. One could potentially find a pattern where the reduction in mortality for the current economic down-turn is actually caused by the previous economic boom.
- It is hard to entirely dismiss this possibility. However, the association we have found is clearest with no time lag, and (as we have seen previously) economic fluctuations between periods of boom and bust have historically had quite variable duration. A stable lagged relationship between economic activity and life expectancy has not yet been discovered.
- We have found empirical evidence to support a claim that above-trend economic growth CAUSES above-trend mortality.
- We have NOT found evidence that above-trend unemployment causes above-trend mortality, since all cyclical economic phenomena are confounded in this analysis.

- We may also notice from the plot that the relationship seems clearer before the mid 1990s, say in the first 45 years of the time series.

```
tt <- 1:45
arima(e_hp[tt],xreg=u_hp[tt],order=c(1,0,0))


Call:
arima(x = e_hp[tt], order = c(1, 0, 0), xreg = u_hp[tt])

Coefficients:
         ar1  intercept  u_hp[tt]
      0.4863     0.0026    0.0829
s.e.  0.1301     0.0417    0.0232

sigma^2 estimated as 0.02152:  log likelihood = 22.39,  aic = -36.78
```

- There is some suggestion that the association is stronger in the time period 1948–1992, but the difference is not large compared to the standard error on the coefficient.
- It is quite plausible that the relationship between health, mortality,

Conclusions

- There is clear evidence of pro-cyclical mortality at a national level in the USA from 1948 to 2018.
- For example, the Great Recession of 2009-2010 led to high unemployment, but these two years had above-trend values of life expectancy at birth.
- More data, perhaps a state-level or international **panel analysis** combining many time series, might be able to improve the signal to noise ratio and lead to clearer results. This might give us the statistical precision to compare time periods and sub-populations more accurately than can be done with just one national-level dataset.

Model selection by AIC

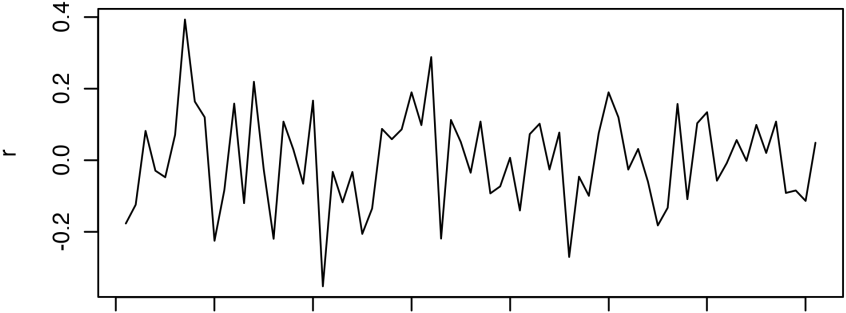
```r
aic_table <- function(data,P,Q,xreg=NULL){
  table <- matrix(NA,(P+1),(Q+1))
  for(p in 0:P) {
    for(q in 0:Q) {
       table[p+1,q+1] <- arima(data,order=c(p,0,q),xreg=xreg)$aic
    }
  }
  dimnames(table) <- list(paste("<b> AR",0:P, "</b>", sep=""),paste
  table
}
e_aic_table <- aic_table(e_hp,4,5,xreg=u_hp)
require(knitr)
kable(e_aic_table,digits=2)
```

	MA0	MA1	MA2	MA3	MA4	MA5
¡b¿ AR0¡/b¿	-58.16	-70.90	-73.08	-74.29	-73.86	-72.33
¡b¿ AR1¡/b¿	-74.52	-72.52	-72.23	-73.19	-79.41	-77.44
¡b¿ AR2¡/b¿	-72.52	-70.52	-72.33	-71.79	-77.46	-78.56

Residual analysis

- We should check the residuals for the fitted model, and look at their sample autocorrelation.

```
r <- resid(arima(e_hp,xreg=u_hp,order=c(1,0,0)))
plot(r)
```



Analysis of temporal differences

- One might model annual changes in life expectancy, rather than difference from a trend. In this case, we consider the variable

$$\Delta e_n^* = e_n^* - e_{n-1}^*. \tag{3}$$

- We compute this as

```
delta_e <- e - e_data$e0[e_data$Year %in% (t-1)]
```

```
plot(t,u,type="l",xlab="Year",ylab="")
par(new=TRUE)
plot(t,delta_e,col="red",type="l",axes=FALSE,xlab="",ylab="")
axis(side=4,col="red")
```

License, acknowledgments, and links

- The materials builds on previous courses.
- Compiled on February 24, 2021 using R version 4.0.4.

Back to course homepage

References

Hodrick RJ, Prescott EC (1997). "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, **29**(1), 1–16.

Ogburn WF, Thomas DS (1922). "The Influence of the Business Cycle on Certain Social Conditions." *Journal of the American Statistical Association*, **18**(139), 324–340.