```
************************************************************************
* This file provides the Stata code used for the analysis of the example
dataset
* used in the paper:
* Interrupted time series regression for the evaluation of public health
* interventions: a tutorial
* IJE 2016
* J. Lopez Bernal, S. Cummins, A. Gasparrini
************************************************************************

clear
set more off
capture log close


****************



*insheet using "sicily.csv", comma
*or: import delimited "sicily.csv"
 import delimited "/Users/alisongemmill/Dropbox/Teaching/SPER
Workshop/Bernal/sicily.csv", encoding(ISO-8859-1)

/* This dataset includes the following variables
year
month
time = elapsed time since the start of the study
aces = count of acute coronary episodes in Sicily per month (the outcome)
smokban = smoking ban (the intervention) coded 0 before the intervention and
1 after
pop = the population of Sicily (in 10000s)
stdpop =  age standardised population
*/


**********************************************
*Step 3: Descriptive analyses
**********************************************
/* Examining the data is an important first step. Looking at the pre-
intervention trend can give an
indication of how stable the trend is over time, whether a linear model is
likely to be approproate
and whether there appears to be a seasonal trend */


*Here we convert the counts into a rate and examine a scatter plot of the
pre-intervention data
gen rate = aces/stdpop*10^5
twoway (scatter rate time) if smokban==0, title("Sicily, 2002-2006")
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year)
```

```
*It is also useful to produce summary statistics for before and after the
intervention
summ, detail

bysort smokban: summ aces
bysort smokban: summ rate



***********************************************
*Step 4: Poisson regression model
***********************************************
/* In step 2 (main paper) we chose a step change model and we also use a
Poisson model as we are using count data
In order to do this we model the count data directly (rather than the rate
which doesn't follow a Poisson distribution)
We then use the population (log transformed) as an offset variable in order
to transform back to rates */


*log transform the standardised population:
gen logstdpop = log(stdpop)


*Poisson with the outcome (aces), intervention (smokban) and time as well as
the population offset offset
glm aces smokban time, family(poisson) link(log) offset(logstdpop) eform


*We generate predicted values based on the model in order to create a plot of
the model:
predict pred, nooffset

*This can then be plotted along with a scatter graph:
gen rate1 = aces/stdpop /*to put rate in same scale as count in model */
twoway (scatter rate1 time) (line pred time, lcolor(red)) , title("Sicily,
2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year) ///
xline(36.5)


*Generate the counterfactual by removing the effect of the intervention
(_b[smokban]) for the post-intervention period
gen pred1 = pred/exp(_b[smokban]) if smokban==1


*Add the counterfactual to the plot
twoway (scatter rate1 time) (line pred time, lcolor(red)) (line pred1 time,
lcolor(red) lpattern(dash)), title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year) ///
xline(36.5)
```

```
*********************************************
*Step 5: methodological issues
*********************************************
* (a) Allowing for overdispersion
/*In the model above we have not allowed for overdispersion - in order to do
this we can add
the scale(x2) parameter to the model which allows the variance to be
proportional rather than
equal to the mean */
glm aces smokban time, family(poisson) link(log) offset(logstdpop) scale(x2)
eform

* (b) Model checking and autocorrelation
*Check the residuals by plotting against time
predict res, r
twoway (scatter res time)(lowess res time),yline(0)

*Further check for autocorrelation by examining the autocorrelation and
partial autocorrelation functions
tsset time
ac res
pac res, yw

* (c) Adjust for seasonality
/* installation of the "circular" package. o find packages select Help > SJ
and User-written Programs,
and click on search */

*we need to create a degrees variable for time divided by the number of time
points in a year (i.e. 12 for months)
gen degrees=(time/12)*360

*we then select the number of sine/cosine pairs to include:
fourier degrees, n(2)

*these can then be included in the model
glm aces smokban cos* sin* time, family(poisson) link(log) offset(logstdpop)
scale(x2) eform

*we can again check for autocorrelation
predict res2, r
twoway (scatter res2 time)(lowess res2 time),yline(0)
tsset time
ac res2
pac res2, yw


*predict and plot of seasonally adjusted model**
predict pred2, nooffset
twoway (scatter rate1 time) (line pred2 time, lcolor(red)), title("Sicily,
2002-2006") ///
```

```
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year) ///
xline(36.5)


/*it is sometimes difficult to clearly see the change graphically in the
seasonally adjusted model
therefore it can be useful to plot a straight line as if all months were the
average to produce a
'deseasonalised' trend. */

egen avg_cos_1 = mean(cos_1)
egen avg_sin_1 = mean(sin_1)
egen avg_cos_2 = mean(cos_2)
egen avg_sin_2 = mean(sin_2)

drop cos* sin*

rename avg_cos_1 cos_1
rename avg_sin_1 sin_1
rename avg_cos_2 cos_2
rename avg_sin_2 sin_2


*this can then be added to the plot as a dashed line
predict pred3, nooffset

twoway (scatter rate1 time) (line pred2 time, lcolor(red)) (line pred3 time,
lcolor(red) lpattern(dash)), title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year) ///
xline(36.5)




**********************************************************
** additional material
**********************************************************
***add a change in slope

*generate interaction term between intervention and time centered at the time
of intervention
gen inter_smokbantime = smokban*(time-36)


*restore fourier variables that were previously changed
drop cos* sin* degrees
gen degrees=(time/12)*360
fourier degrees, n(2)

*add the interaction term to the model
```

```
glm aces smokban inter_smokbantime cos* sin* time, family(poisson) link(log)
offset(logstdpop) scale(x2) eform
*(the coefficient and CI for the interaction term suggests that there is very
little slope change)

*plot seasonally adjusted model with deseasonalised trend**
predict pred4, nooffset

egen avg_cos_1 = mean(cos_1)
egen avg_sin_1 = mean(sin_1)
egen avg_cos_2 = mean(cos_2)
egen avg_sin_2 = mean(sin_2)
drop cos* sin*
rename avg_cos_1 cos_1
rename avg_sin_1 sin_1
rename avg_cos_2 cos_2
rename avg_sin_2 sin_2

predict pred5, nooffset

twoway (scatter rate1 time) (line pred4 time, lcolor(red)) (line pred5 time,
lcolor(red) lpattern(dash)), title("Sicily, 2002-2006") ///
ytitle(Std rate x 10000) yscale(range(0 .)) ylabel(#5, labsize(small)
angle(horizontal)) ///
xtick(0.5(12)60.5) xlabel(6"2002" 18"2003" 30"2004" 42"2005" 54"2006",
noticks labsize(small)) xtitle(year) ///
xline(36.5)
```