Partial regression plots

Brendan Halpin

Feb 10 2018

1 Partial regression plots

Let's explore partial regression plots, using data on mental distress, affected by adverse life effects and SES.

We start by regressing impairment on SES alone, and getting the residuals. These residuals are large where the impairment score is high relative to the prediction based on the SES score, small where they're close and large negative where impairment is much lower than expected. If there are other variables with a positive correlation with impairment, we would expect cases with high residuals to be also high on that variable. See Fig 1.

```
use impair
reg impair SES
predict yres1, res

predict yhat1
scatter impair SES || function _b[_cons] + x*_b[SES], range(SES) ///
 || rspike yhat1 impair SES, title("Residuals, Impairment predicted by SES")
```

If the other variable is correlated with SES, then it might be hard to see whether it has any independent explanatory power. So what we do is regress life events (the other variable) on SES and get the residuals. High residuals indicate levels of life events that are high for that level of SES. The residuals are the variation in LIFEVT that is not associated with SES. See Fig 2.

```
reg lifeevt SES
predict yhat2
predict xres1, res
scatter lifeevt SES || function _b[_cons] + x*_b[SES], range(SES) ///
 || rspike yhat2 lifeevt SES, title("Residuals, LIFEEVT predicted by SES")
```

1.1 Plots

Let's compare the residuals for impairment|SES with LIFEEVT|SES (where "|" means "controlling for"). This focuses on where impairment is high relative to SES (or low) and relating it to where LIFEEVT is high relative to SES (or low). If there is association, it implies that LIFEEVT has explanatory power even after SES is taken into account. See Fig 3.

```
scatter yres1 xres1, title("Residuals, Y|SES vs LIFEEVT|SES") ///
 name(gr1, replace)
```

Since we have two explanatory variables, it makes sense to do the same thing predicting impairment using LIFEEVT and looking at whether SES has explanatory power independent of it. That is, to look at residuals for Y|LIFEEVT vs SES|LIFEEVT. See Fig 4 which combines the two plots. In the left plot, we see a negative relationship: high levels of SES (controlling for LIFEEVT) go with low levels of impairment (controlling for LIFEEVT). In the right plot, we see that high levels of LIFEEVT (controlling for SES) go with high levels of impairment (controlling for SES). Each plot is telling us that the second variable is likely to add explanatory power to

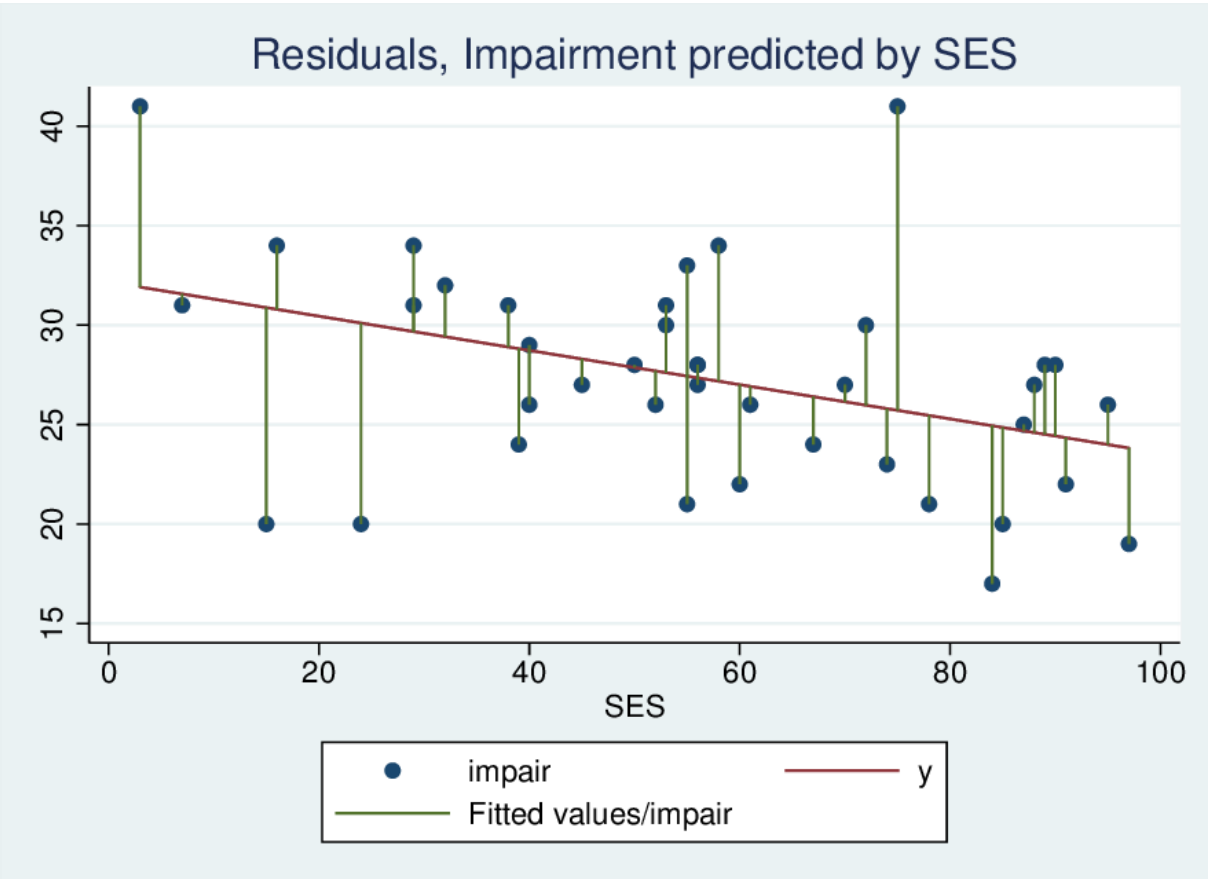

Figure 1: Residuals, Impairment predicted by SES

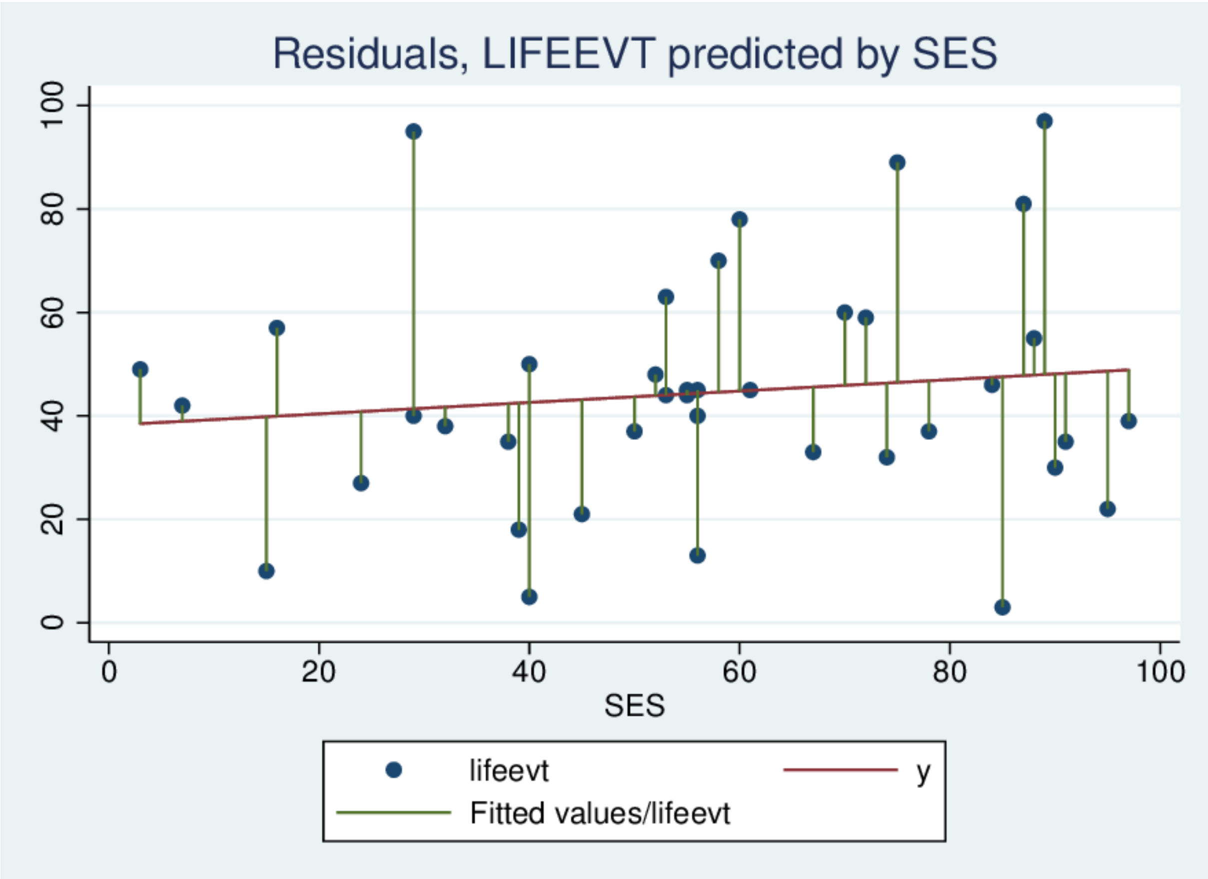

Figure 2: Residuals, LIFEEVT predicted by SES

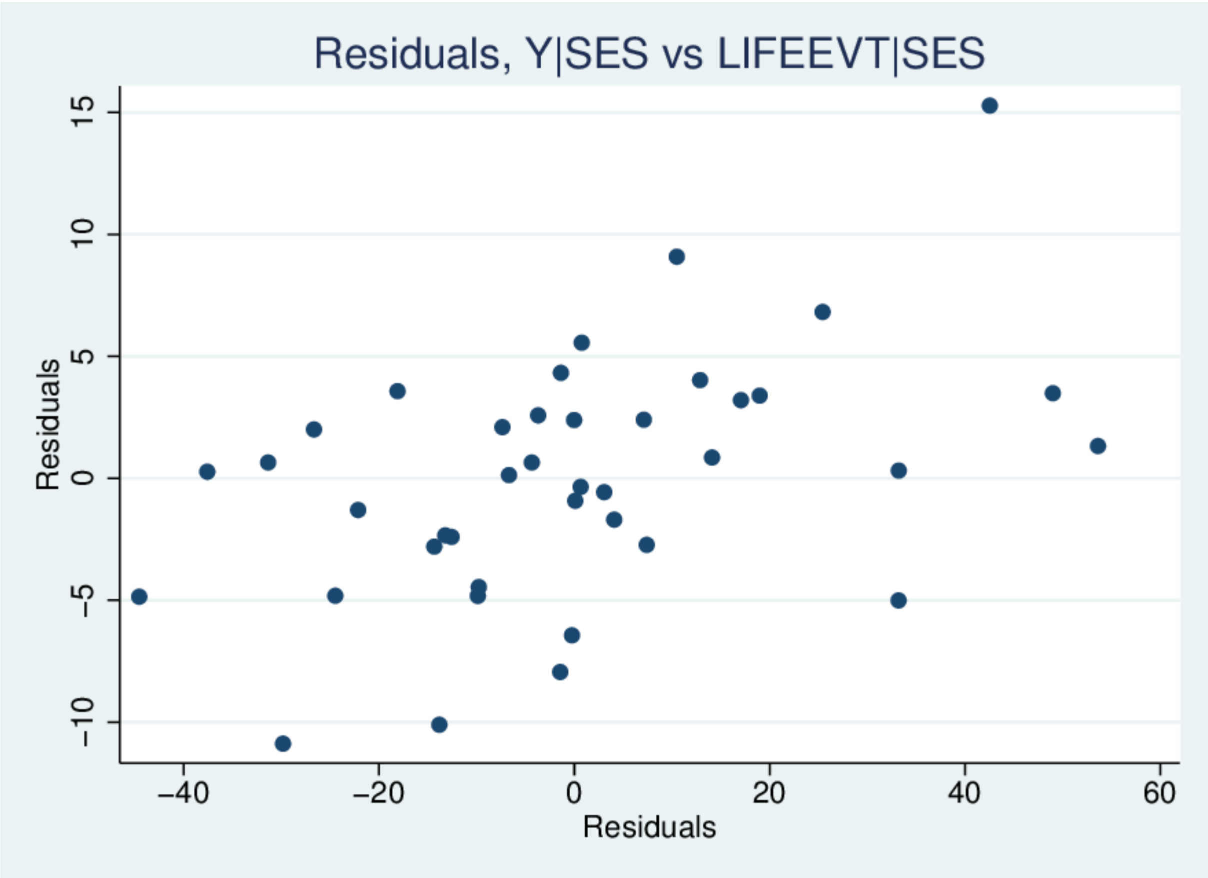

Figure 3: Residuals, Y | SES vs LIFEEVT | SES

the model, even after taking account of the other.

```
reg impair lifeevt
predict yres3, res

reg SES lifeevt
predict yhat3
predict xres3, res
scatter yres3 xres3, title("Residuals, Y|LIFEEVT vs SES|LIFEEVT") ///
  name(gr2, replace)
graph combine gr2 gr1, xsize(6) ysize(4)
```

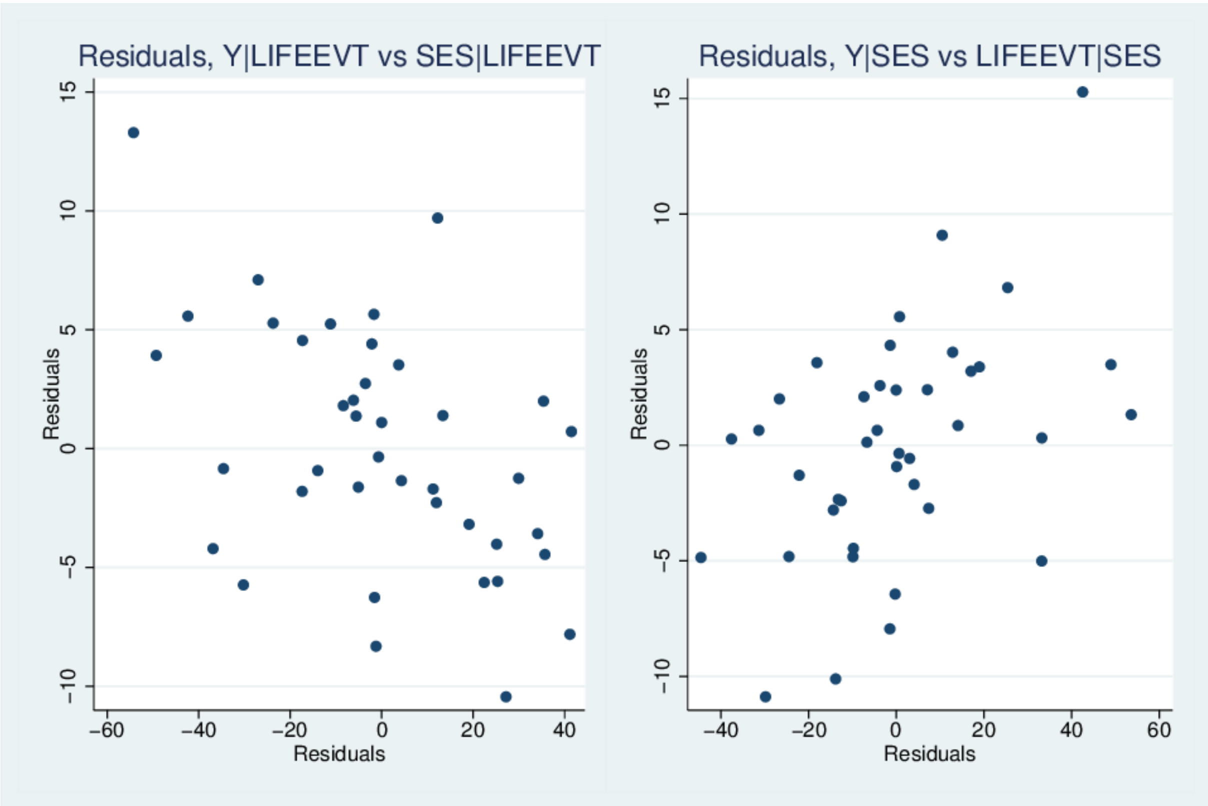

Figure 4: Both Partial Regression Plots

1.2 Stata short cut

There is a Stata shortcut to generate partial regression plots. The code above makes clear exactly how the plots are constructed and what they mean, but in everyday practice it is easier to issue the following commands:

```
reg impair SES lifeevt
avplots
```

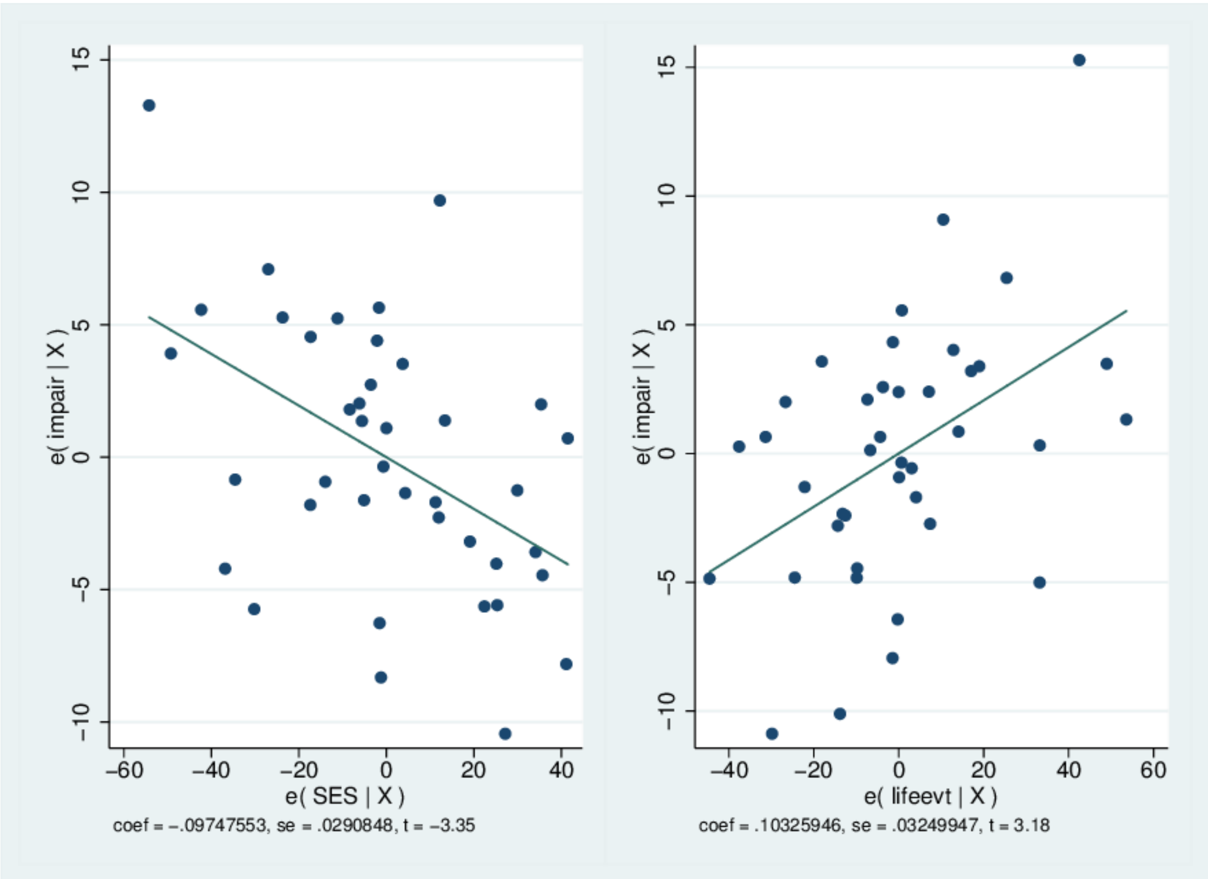

Figure 5: Partial Regression Plots from Stata avplot command