

Fixed Effects and Sample Sizes: LPMs & PGA Putting

```
. use "PGA2017 Putting v2.dta", clear

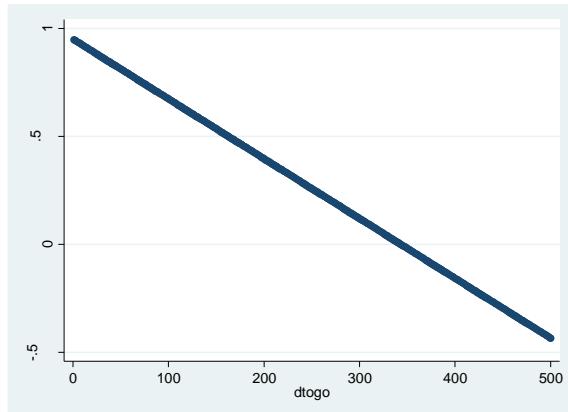
. reg madeit dtogo if dtogo <=500

      Source |       SS           df          MS
-----+-----
    Model |  44894.0368         1  44894.0368
  Residual | 47220.8307401265   .117679914
-----+-----
    Total | 92114.8675401266   .229560609

      Number of obs = 401267
      F( 1, 401265) = .
      Prob > F = 0.0000
      R-squared = 0.4874
      Adj R-squared = 0.4874
      Root MSE = .34305

-----+
      madeit |     Coef.    Std. Err.      t    P>|t|    [95% Conf. Interval]
-----+
      dtogo | -.00277  4.48e-06 -617.65  0.000  -.0027788  -.0027613
      _cons | .9507807 .000736 1291.90  0.000  .9493382  .9522231
-----+
```

```
. predict yhat0
. egen plotflag=tag(dtogo)
. scatter yhat0 dtogo if plotflag==1 & dtogo <=500
```



... hmmm, predicted values below 0? Really?

Looking at a small sample I: About 40 observations

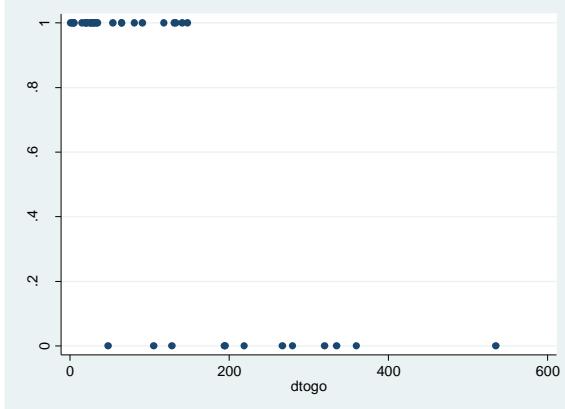
```
. *gen flag1 = (runiform()<.0001)
. tab flag1

      flag1 |       Freq.        Percent        Cum.
-----+
      0 |  421,756      99.99      99.99
      1 |       37       0.01      100.00
-----+
    Total |  421,793     100.00
```

Fixed Effects and Sample Sizes: *LPMs & PGA Putting*

Let's try fixed effects:

```
. qui: reg madeit i.dtogo if flag1==1  
. predict yhat1  
. scatter yhat1 dtogo if flag1==1
```



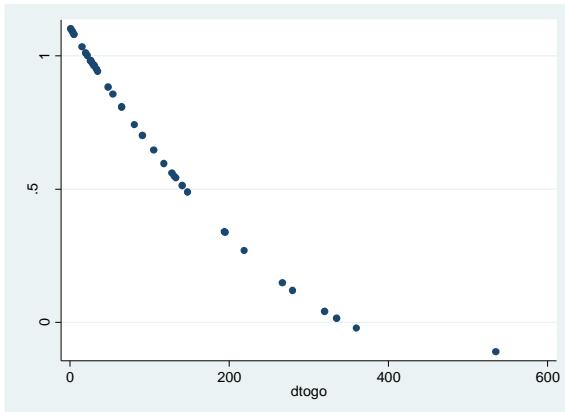
Not sure this tells us much, eh?

So much for Fixed Effects! Let's try polynomials:

```
. gen dtogo2 = dtogo^2  
. gen dtogo3 = dtogo^3
```

... Let's try a quadratic specification... 2nd order polynomial...

```
. qui: reg madeit dtogo dtogo2 if flag1==1  
. predict yhat1b  
. scatter yhat1b dtogo if flag1==1
```

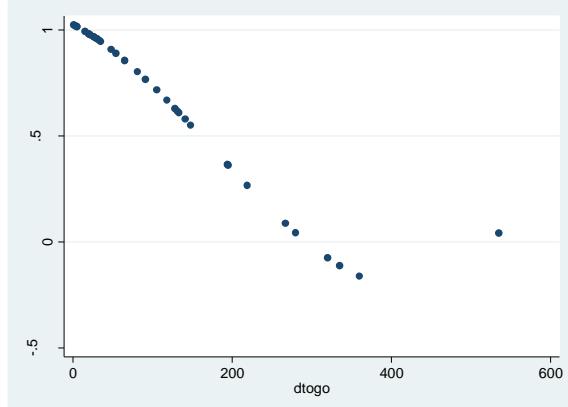


... hmmm, now predicted values above 1 and below 0?

Fixed Effects and Sample Sizes: LPMs & PGA Putting

... Let's try a 3rd order polynomial...

```
. qui: reg madeit dtogo dtogo2 dtogo3 if flag1==1  
. predict yhat1c  
. scatter yhat1c dtogo if flag1==1
```



... predicted values still above 1 and below 0? ... maybe a larger sample will improve things...

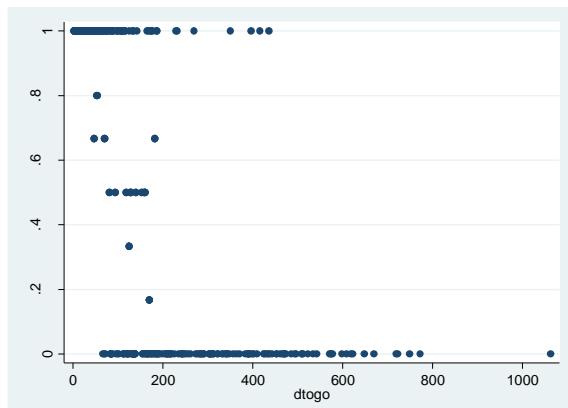
Looking at a small sample II: About 400 observations

```
. *gen flag2 = (runiform()<.001)  
. tab flag2
```

flag2	Freq.	Percent	Cum.
0 421,362	99.90	99.90	
1 431	0.10	100.00	
Total 421,793	100.00		

Let's try fixed effects again:

```
. reg madeit i.dtogo if flag2==1  
. predict yhat2  
. scatter yhat2 dtogo if flag2==1
```

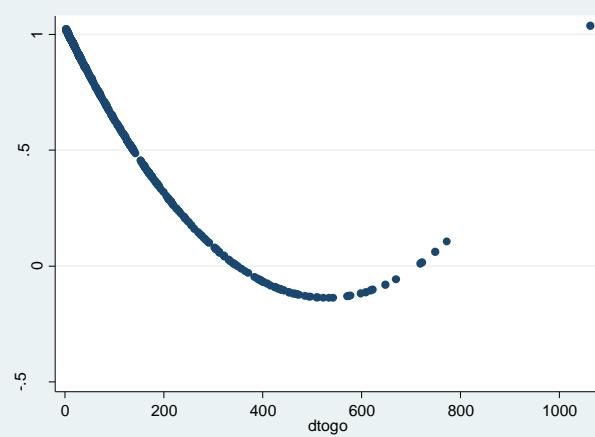


Not sure this tells us much II, eh?

Fixed Effects and Sample Sizes: LPMs & PGA Putting

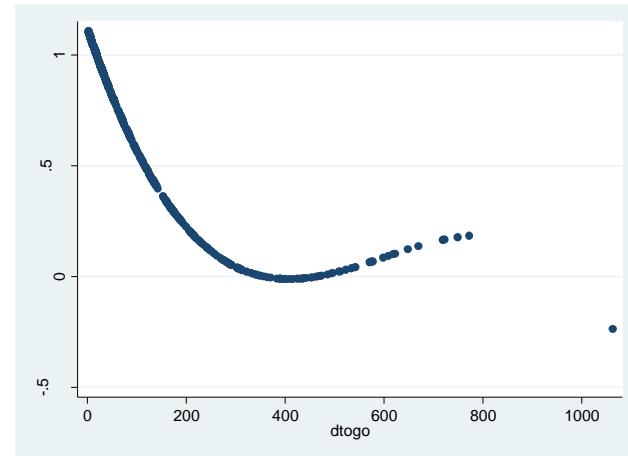
... and 2nd and 3rd order polynomials...

```
. qui: reg madeit dtogo dtogo2 if flag2==1  
. predict yhat2b  
. scatter yhat2b dtogo if flag2==1
```



... Again, predicted values above 1 and below 0... and now U-shaped!

```
. reg madeit dtogo dtogo2 dtogo3 if flag2==1  
. predict yhat2c  
. scatter yhat2c dtogo if flag2==1
```



... well at least there are almost no predicted values < 0!

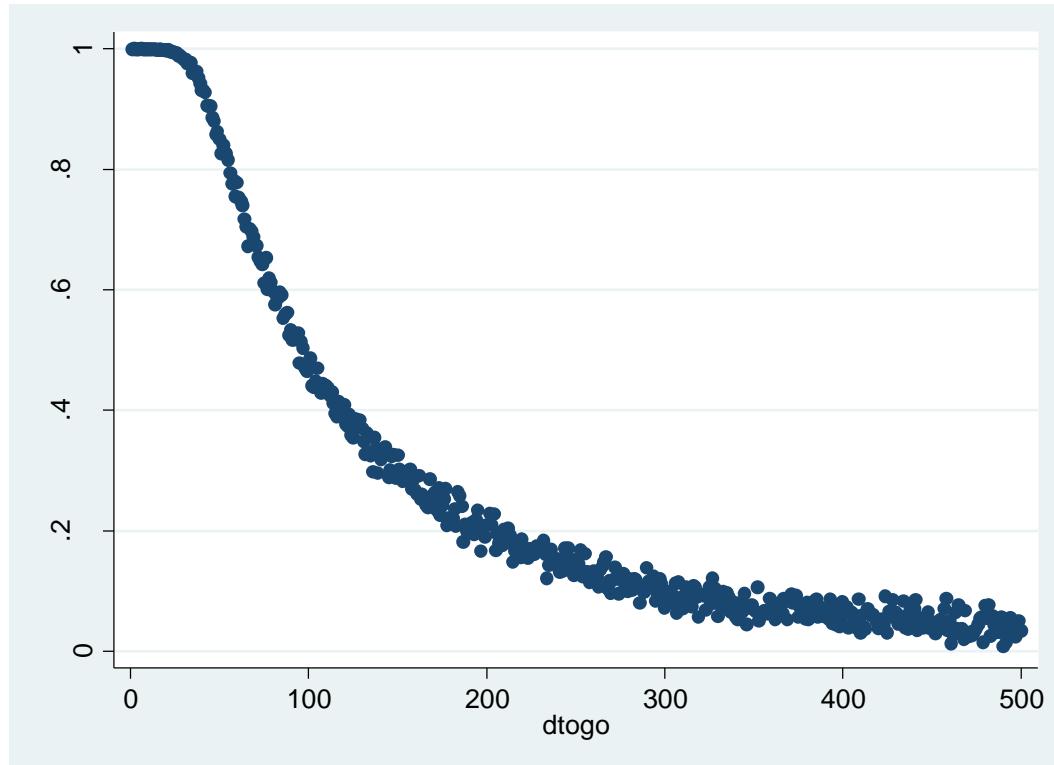
But where did that U-shape come from?

Fixed Effects and Sample Sizes: *LPMs & PGA Putting*

What happens when you have the full dataset... all 400K+ observations!

```
. set matsize 600
. reg madeit i.dtogo if dtogo <=500
. predict yhat3
. egen plotflag=tag(dtogo)

. scatter yhat3 dtogo if plotflag==1 & dtogo <=500
```



Now we're talking! ... no need to fit those polynomials when you have this much data! The (conditional on dtogo) sample means tell the whole story.