F-test on Fixed Effects in R (Panel Data)

I am trying to do an F-test on the joint significance of fixed effects (individual-specific dummy variables) on a panel data OLS regression (in R), however I haven't found a way to accomplish this for a large number of fixed effects. Ideally, I would use a function in the `plm` package, however I haven't found anything that specifically does this test.

This is something Stata does automatically when using the `xtreg, fe` command. In Stata, the results looks like this:

```
------------------------------------------------------------------------------
F test that all u_i=0:  F(49, 498) =    12.00   Prob > F = 0.000
```

Again, I am trying to reproduce the Stata result in R for a large number of dummy variables, perhaps specified by `+ factor(us.state)` using `lm()` or `model = "fe"` using `plm()`.

Here is a reproducible example:

```
require(foreign)

voter <- read.dta("http://www.montana.edu/econ/cstoddard/562/panel_hw.dta")

reg1 <- lm(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border
               + factor(state), data=voter)
```

which is equivalent to the following "within" regression using the `plm` package.

```
require(plm)

reg1.fe <- plm(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border,
data=voter, index = c("state","year"), model = "within")
```

So, the test would be the test that all the state dummy variables are jointly different from zero (jointly significant). This is a linear restriction on the unrestricted model (reg1 and reg1.fe above). This F-test is better explained on the following document (see slides 5-7).

http://jackman.stanford.edu/classes/350B/07/ftestforWeb.pdf

Here is one of my feeble attempts at creating an 'R' matrix for the F-test with null hypothesis: Rb = q where b is the matrix of coefficients (beta hat), and q is a vector of zeros.

```
d1 = length(unique(voter$stcode))-1
d2 = length(reg1$coefficients)
R = cbind(matrix(0,d1,d2),diag(d1))

linearHypothesis(reg1,R,rhs=0)
```

This doesn't work! And, I'm hoping there is a streamlined approach to testing for joint significance of all fixed effect dummy variables.

r stata plm

edited Aug 15 '14 at 7:14
landroni
954 7 20

asked May 29 '11 at 23:35
baha-kev
1,214 2 20 29

2 Answers

First, I'd like to suggest that your question could be improved by (1) providing a reproducible example, and (2) describing the precise test to which you refer to when you say 'F test'. A link to the Stata docs maybe? F is the distribution, so there can be a gazillion tests called an 'F test'.

If your substantive interest lies in determining whether the fixed effects model fits the data significantly better than OLS without fixed effects, then you could always use a likelihood ratio test. I'm sure there are many implementations in R, but the one provided by the `lmtest` package is pretty convenient. Here's an example using a dataset distributed with the `plm` package (you seem to have that installed, so it should be easy to try).

```
library(plm)
data(Produc)

library(lmtest)
mod <- lm(pcap ~ hwy + water, Produc)
```

```
mod.fe <- lm(pcap ~ hwy + water + factor(state), Produc)
lrtest(mod, mod.fe)
```

and the output:

```
Likelihood ratio test

Model 1: pcap ~ hwy + water
Model 2: pcap ~ hwy + water + factor(state)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   4 -8038.1
2  51 -6712.4 47 2651.4  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

EDIT: OPTION 2

```
require(foreign)
voter <- read.dta("http://www.montana.edu/econ/cstoddard/562/panel_hw.dta")
reg1 <- lm(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border
               + factor(state), data=voter)

library(plm)
reg1.fe <- plm(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border,
data=voter, index = c("state","year"), model = "within")

reg1.pooling <- plm(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border,
data=voter, index = c("state","year"), model = "pooling")

pFtest(reg1.fe, reg1.pooling)
```

OUTPUT:

```
    F test for individual effects

data:  vaprate ~ gsp + midterm + regdead + WNCentral + South + Border
F = 13.0712, df1 = 45, df2 = 498, p-value < 2.2e-16
alternative hypothesis: significant effects
```

edited May 30 '11 at 3:44

answered May 30 '11 at 1:04

Vincent
2,713 13 24

Hi Vincent, I have provided a reproducible example in my post. Thanks for responding. I am however, trying to replicate precisely the F-test results. From an empirical perspective, I understand there are several statistical tests to compare a FE model with a model without FE. – baha-kev May 30 '11 at 1:32

I think plm's pFtest() function may do what you want (see my edited answer). The outcome is not exactly the same as your Stata output, which is probably due to the fact that the first parameter of the F distribution is different. But when I fit both models individually with lm(), I get degrees of freedom of 543 and 498 (difference of 45), so R seems right here. See if you get the same degrees of freedom in Stata when you fit the pooling and within models individually. The problem with closed source software like Stata is we'll never know precisely how they calculate their F test. – Vincent May 30 '11 at 3:50

I really don't think that this test is useful at all. Instead of estimating what you call a fixed effect (I'll call it no-pooling model), why not a hierarchical model? The hierarchical model (or partial pooling model) will allow for your estimations to shrink to the commom mean for states, but without imposing them to be equal. Morevoer, if you need to assess how much states vary, you just need to use the variance estimated between state and intra-state. If the variance among state is low (near zero), than you aren't gaining so much using a hierarchical model and the intercepts are roughly the same. If the variance is very large (in the limit, when goes to infinity) the hierarchial model adds little and you could run a separated model for each state.

You can estimate a hierarchical model in R with the package lme4. Using your data:

```
require(lme4)
require(arm)
reg1 <- glmer(vaprate ~ gsp + midterm + regdead + WNCentral + South + Border
           + (1|state), data=voter)
display(reg1)
```

The estimated standard deviation of the intercept by states is 4.39 and the standard deviation by individual is 4.19.

answered May 30 '11 at 4:46

Manoel Galdino
860 1 11 24