Classical panel data models in R

Christian Kleiber
Quantitative Methods Unit, WWZ, Universität Basel

February 28, 2011

1 Installation of add-on packages

In R, the classical estimators for static and dynamic linear models for panel data are available in the add-on package **plm** (Croissant and Millo 2008). Also, various data sets from econometrics are available in the package **AER** ("Applied Econometrics with R", see Kleiber und Zeileis 2008). These two packages may be installed – provided you are connected to the internet! – directly from within R, using

```
R> install.packages("AER")
R> install.packages("plm")
```

or more compactly

```
R> install.packages(c("AER", "plm"))
```

Alternatively, GUIs for operating systems such as Windows oder Mac OS X have a menu named "Packages", where installation of packages is possible via selection from a list (loading this list may take some time, as there are very many packages these days!).
When installing **AER** and/or **plm** various further packages are installed as well, among them **lmtest**, **sandwich** ..., because **AER** or **plm** depend on these (in the sense that they require certain functions from these packages).
We load **plm** first:

```
R> library("plm")
```

2 Structure of panel data

For illustrating the basic panel data estimators we now use *the* data set of panel data econometrics, the Grunfeld data (from a Chicago Ph.D. thesis of 1958). See Kleiber and Zeileis (2010) for the somewhat peculiar history of these data as well as further details. We load the data using

```
R> data("Grunfeld", package = "AER")
```

The structure of the data can be inspected via

```
R> str(Grunfeld)
```

```
'data.frame':        220 obs. of  5 variables:
 $ invest : num  318 392 411 258 331 ...
 $ value  : num  3078 4662 5387 2792 4313 ...
 $ capital: num  2.8 52.6 156.9 209.2 203.4 ...
 $ firm   : Factor w/ 11 levels "General Motors",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ year   : int  1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 ...
```

```
R> names(Grunfeld)

[1] "invest"  "value"   "capital" "firm"    "year"

R> levels(Grunfeld$firm)

 [1] "General Motors"    "US Steel"          "General Electric"
 [4] "Chrysler"          "Atlantic Refining" "IBM"
 [7] "Union Oil"         "Westinghouse"      "Goodyear"
[10] "Diamond Match"     "American Steel"

R> head(Grunfeld)

  invest  value capital           firm year
1  317.6 3078.5     2.8 General Motors 1935
2  391.8 4661.7    52.6 General Motors 1936
3  410.6 5387.1   156.9 General Motors 1937
4  257.7 2792.2   209.2 General Motors 1938
5  330.8 4313.2   203.4 General Motors 1939
6  461.2 4643.9   207.2 General Motors 1940
```

The Grunfeld data will now be used for replicating some estimates for a certain version of these data. We need to exclude one of the firms first (see Kleiber and Zeileis 2010 for reasons):

```
R> gr <- subset(Grunfeld, firm != "American Steel")
```

In order to use these data as a panel data set, it is necessary to define the identifiers for individuals and the time index. All software packages require some such step, of course the precise form depends on the software at hand. In **plm**, we need the function `plm.data`, essentially a version of `data.frame`, which takes care of the two identifiers:

```
R> pgr <- plm.data(gr, index = c("firm", "year"))
```

Internally, this entails certain changes in the layout of the dataset:

```
R> head(pgr)

            firm year invest  value capital
1 General Motors 1935  317.6 3078.5     2.8
2 General Motors 1936  391.8 4661.7    52.6
3 General Motors 1937  410.6 5387.1   156.9
4 General Motors 1938  257.7 2792.2   209.2
5 General Motors 1939  330.8 4313.2   203.4
6 General Motors 1940  461.2 4643.9   207.2
```

3 Some classical panel data models

3.1 Fixed effects

Classical one-way fixed effects regression:

```
R> gr_fe <- plm(invest ~ value + capital, model = "within", data = pgr)
R> summary(gr_fe)
```

```
Oneway (individual) effect Within Model

Call:
plm(formula = invest ~ value + capital, data = pgr, model = "within")

Balanced Panel: n=10, T=20, N=200

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-184.000  -17.600    0.563   19.200  251.000

Coefficients :
        Estimate Std. Error t-value  Pr(>|t|)
value   0.110124   0.011857  9.2879 < 2.2e-16 ***
capital 0.310065   0.017355 17.8666 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2244400
Residual Sum of Squares: 523480
R-Squared      :  0.76676
      Adj. R-Squared :  0.72075
F-statistic: 309.014 on 2 and 188 DF, p-value: < 2.22e-16
```

Same with robust standard errors (which ones?):

```
R> library("lmtest")
R> coeftest(gr_fe, vcov = vcovHC)

t test of coefficients:

        Estimate Std. Error t value  Pr(>|t|)
value   0.110124   0.014342  7.6783 8.566e-13 ***
capital 0.310065   0.049793  6.2271 3.033e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.2 Random effects

Classical one-way random effects model:

```
R> gr_reswar <- plm(invest ~ value + capital, data = pgr,
+   model = "random", random.method = "swar")
R> summary(gr_reswar)

Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = invest ~ value + capital, data = pgr, model = "random",
    random.method = "swar")

Balanced Panel: n=10, T=20, N=200

Effects:
```

```
                  var std.dev share
idiosyncratic 2784.46   52.77 0.282
individual    7089.80   84.20 0.718
theta:  0.8612

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-178.00  -19.70    4.69   19.50  253.00

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(Intercept) -57.834415  28.898935 -2.0013  0.04674 *
value         0.109781   0.010493 10.4627  < 2e-16 ***
capital       0.308113   0.017180 17.9339  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2381400
Residual Sum of Squares: 548900
R-Squared      :  0.7695
      Adj. R-Squared :  0.75796
F-statistic: 328.837 on 2 and 197 DF, p-value: < 2.22e-16
```

3.3 Pooled OLS

```
R> gr_ols <- plm(invest ~ value + capital, data = pgr, model = "pooling")
R> summary(gr_ols)
```

```
Oneway (individual) effect Pooling Model

Call:
plm(formula = invest ~ value + capital, data = pgr, model = "pooling")

Balanced Panel: n=10, T=20, N=200

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -292.0   -30.0     5.3    34.8   369.0

Coefficients :
               Estimate  Std. Error t-value  Pr(>|t|)
(Intercept) -42.7143694   9.5116760 -4.4907 1.207e-05 ***
value         0.1155622   0.0058357 19.8026 < 2.2e-16 ***
capital       0.2306785   0.0254758  9.0548 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    9359900
Residual Sum of Squares: 1755900
R-Squared      :  0.81241
      Adj. R-Squared :  0.80022
F-statistic: 426.576 on 2 and 197 DF, p-value: < 2.22e-16
```

However . . . these standard errors are not appropriate as they ignore the panel data structure. OLS standard errors are

```
R> sqrt(diag(vcov(gr_ols)))

(Intercept)       value     capital
 9.51167603  0.00583571  0.02547580
```

For comparison, here are clustered standard errors as proposed by Beck and Katz (1995):

```
R> sqrt(diag(vcovBK(gr_ols, cluster = "time")))

(Intercept)       value     capital
6.780964847 0.007212438 0.027886213
```

4 Specification testing

Some diagnostic tests in the pooling model:

```
R> plmtest(gr_ols, type = "bp")

        Lagrange Multiplier Test - (Breusch-Pagan)

data:  invest ~ value + capital
chisq = 798.1615, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects

R> plmtest(gr_ols, type = "honda")

        Lagrange Multiplier Test - (Honda)

data:  invest ~ value + capital
normal = 28.2518, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Hausman test:

```
R> phtest(gr_fe, gr_reswar)

        Hausman Test

data:  invest ~ value + capital
chisq = 2.3304, df = 2, p-value = 0.3119
alternative hypothesis: one model is inconsistent
```

5 Outlook

This is only a brief introduction to some of the fitting functions that are available in the **plm** package for R. There exists much more comprehensive documentation, written by the authors of the package (Croissant and Millo 2008). A more current version, which also documents recent changes as well as additions to **plm**, may be found in an accompanying vignette:

```
R> vignette("plm", package = "plm")
```

Please familiarize yourself with this feature of R: many reasonable packages have accompanying vignettes!

References

Croissant Y, Millo G (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27 (2), 1–43.

Grunfeld Y (1958). The determinants of corporate investment. Ph. D. thesis, University of Chicago.

Kleiber C, Zeileis A (2008). *Applied Econometrics with R*. New York: Springer.

Kleiber C, Zeileis A (2010). The Grunfeld data at 50. *German Economic Review*, 11 (4), 404–417.