



help apc

Title

apc — Estimating age, period and cohort effects

Syntax

```
apc depvar [if] [in], age(varname) period(varname) cohort(varname) [options]
```

<i>options</i>	description
Model	
method (<i>identmethod</i>)	<i>identmethod</i> can be dp or me ; default is dp
lowerbound (#)	lower bound of dependent variable in the population; must be supplied if method is me
upperbound (#)	upper bound of dependent variable in the population; must be supplied if method is me
Estimation	
technique (<i>opttech</i>)	<i>opttech</i> can be nr , dfp , bfgs or nm ; default is nr
ptol (#)	<i>ptol</i> controls the optimization routine; default is ptol(1e-10)
vtol (#)	<i>vtol</i> controls the optimization routine; default is vtol(1e-10)
maxiter (#)	<i>maxiter</i> controls the optimization routine; default is maxiter(2000000)
Reporting	
graph (<i>graphopt</i>)	controls whether to produce and save graphs; <i>graphopt</i> must be either off or on , replace ; default is off

Description

apc estimates the coefficients of the linear age-period-cohort (apc) model with dependent variable *depvar*.

The apc model aims to separate, for some outcome of interest, those influences associated with the process of aging, from those influences associated with the date at which subjects are observed, from those influences associated with a subject's date of birth. In the linear apc model, whose coefficients are estimated by **apc**, the dependent variable is considered to be a linear additive function of dummies representing each age, period and cohort value observed in the sample.

The linear apc model suffers from a well-known fundamental identification problem, namely that there is a perfect linear relationship between these effects: current year equals year of birth plus age (in years).

apc estimates the coefficients using two methods that have been proposed. The first method is that of Deaton & Paxson (1994) which identifies the coefficients by imposing the restriction that the period effect is orthogonal to a trend and sums to zero. The second method is that of Browning, Crawford & Knoef (2010) which is based on a maximum entropy method and can be applied when the dependent variable has a finite support in the population.

Options

Model
method (<i>identmethod</i>); Specifies which method to use for estimation. Allowed options for <i>identmethod</i> are:
dp : the parameters are identified using the restriction proposed by Deaton & Paxson (1994). This is the default.

me: the parameters are chosen based on a maximum entropy method proposed by Browning, Crawford & Knoef (2010). This method is applicable only if the dependent variable has a finite support in the population.

lowerbound(#) The lower bound of the dependent variable in the population. Must be specified if the method used is **me**.

upperbound(#) The upper bound of the dependent variable in the population. Must be specified if the method used is **me**.

Estimation

technique(opttech) If the method of estimation is **me** then, then the parameters are selected in order to maximise the entropy function. For details see Browning, Crawford & Knoef (2010). This maximisation uses STATA's optimize() routine. The optimization technique can be one of:

<i>technique</i>	<i>description</i>
nr	modified Newton-Raphson
dfp	Davidon-Fletcher-Powell
bfgs	Broyden-Fletcher-Goldfarb-Shanno
nm	Nelder-Mead

The default is **nr**.

See optimize init technique() for more details

ptol(#) Controls the convergence criterion used by optimize(). See optimize init conv ptol(). Default is **ptol(1e-10)**.

vtol(#) Controls the convergence criterion used by optimize(). See optimize init conv vtol(). Default is **vtol(1e-10)**.

maxiter(#) Sets the maximum number of iterations allowed by optimize(). See optimize init conv maxiter(). Default is **maxiter(200000)**.

Reporting

graph(graphopt)} controls whether to produce and save graphs. Allowed options for *graphopt* are:

off No graphs are displayed or saved. This is the default.

on, replace If this option is specified, four graphs are produced and saved in the current working directory. These are: a graph of the trajectory of the age effects, the period effects, the cohort effects and all three combined. 95% confidence intervals are included on the graphs. The first three of these are saved as *varname.gph* where *varname* is the name of the age, period or cohort variable in the data. The graph that shows all three trajectories is saved as *apc.gph*. These saved graphs will overwrite graphs of the same name in the current working directory - so for graphs to be produced *replace* has to be specified as part of *graphopt*.

Examples

These examples use the STATA example data set *union.dta*. The dependent variable is *grade completed*. This dependent variable has bounds in the population (between 0 and 18), so the maximum entropy method can be used.

Open data

```
. webuse union, clear
```

Generate a cohort variable as year less age

```
. gen cohort = year - age
```

Estimating coefficients using maximum entropy approach

```
. apc grade, age(age) period(year) cohort(cohort) method(me) lower(0)
  upper(18)
```

Estimating coefficients using maximum entropy approach; displaying and saving graphs

```
. apc grade, age(age) period(year) cohort(cohort) method(me) lower(0)
  upper(18) graph(on, replace)
```

Estimating coefficients using Deaton and Paxson's method; displaying and saving graphs (note, no bounds required)

```
. apc grade, age(age) period(year) cohort(cohort) method(dp) graph(on,
  replace)
```

Estimating coefficients using Deaton and Paxson's method; restricting the sample to those aged less than or equal to 30

```
. apc grade if age<=30, age(age) period(year) cohort(cohort) method(dp)
```

Saved results

apc saves the following in **e()**:

Scalars

e(N) number of observations

Macros

e(cmd) **apc**

Matrices

e(b) coefficient vector

e(V) variance-covariance matrix of the estimators

Functions

e(sample) marks estimation sample

A note on the identification of the apc coefficients

The linear dependence that results in the identification problem comes from the fact that "age + year of birth = current year". However, some characteristics of the data, if present, can result in that equation not holding for everyone in the data. In such a case the coefficients are identified without further restriction. **apc** checks whether the parameters are identified without restriction and if so returns a message informing the user that this is the case. Estimation then proceeds using whichever method the user specified (i.e. **dp** or **me**). Whether one proceeds with a method such as Deaton and Paxson's or the maximum entropy approach or one switches to OLS (or similar) depends on whether the identification restrictions (which may be implicit in the data setup rather than explicitly imposed by the user) are credible.

Two common reasons for the linear apc coefficients being identified without further restrictions are:

1. One of the age, period or cohort variables is denominated using a different time unit to the others. For example, age and period could be represented by dummies representing single years while cohort could be a dummy representing a particular decade. The identifying assumption here is that the cohort effect is the same for all those born in a particular decade. If the user is satisfied that this is a credible assumption, then OLS (or similar) can be used.
2. Due to the time of sampling the equation "age + year of birth = current year" might not hold. For example, if an individual is born in December 1990 (so cohort = 1990), and they are surveyed in January 2012 (so period = 2012), their age in the data is 21. For this individual the equation above doesn't hold. The linear apc coefficients can be identified using OLS. Identification is coming from variation in ages between individuals born in the same year. It will, in many applications, be difficult to justify this as an identification assumption.

Citing apc

The use of **apc** can be cited as follows:

O'Dea, Cormac. 2012. "APC: Stata module to estimate age, period and cohort effects". Available at: www.ifs.org.uk/publications/5998.

Disclaimer

THIS SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

IN NO EVENT WILL THE COPYRIGHT HOLDERS OR THEIR EMPLOYERS, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THIS SOFTWARE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE ABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PART BE ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Author

Stata command **apc** was written by: Cormac O'Dea, Institute for Fiscal Studies. (cormac_o@ifs.org.uk)

References

- Browning, M., Crawford, I. and Knoef, M. 2010. "The Age-Period-Cohort Problem; Set identification and point identification". Working Paper.
<http://www.economics.ox.ac.uk/members/ian.crawford/papers/apc.pdf>
- Deaton, Angus, and Christina Paxson. 1994. "Saving, Growth, and Aging in Taiwan," in D. Wise (ed) Studies in the Economics of Aging, pp. 331-362 National Bureau of Economic Research.