Browse IFS

Research topics
Choose a subject

Publication types
Choose a type

Advanced search

Quick links

Fiscal Facts
Facts and figures about UK taxes, benefits and public spending.
Budgets & Pre-Budget Reports
IFS researchers analyse the Government's fiscal forecasts and tax and spending decisions.
Where do you fit in?
Find out where you are in the income distribution.
Mirrlees Review
Reforming the tax system for the 21st century.
Fiscal Studies
A peer-reviewed quarterly journal publishing articles by academics and practitioners.
Educational Resources
Resources for schools and students.

I'd like to...

Sign up to IFS news
Become a member of the IFS
Attend an event
Make a press enquiry
Get in touch

Datasets

October 2008

Alternative approaches to evaluation in empirical microeconomics: resources

Type: Datasets
Authors: Richard Blundell and Monica Costa Dias
Papers using this dataset: Alternative approaches to evaluation in empirical microeconomics (Mimeos) ; Alternative approaches to evaluation in empirical microeconomics (cemmap Working Papers)

E-mail to a friend

The results displayed in the paper can be reproduced with the simulated datasets and dofiles in this webpage. The following details some information to frame and facilitate the use of these resources. Please address any queries to monica_d at ifs.org.uk.

Simulations are based on the individual life-cycle model of education investment and earnings described in the paper and are made available under alternative assumptions:

- Policy environment: (i) whether or not a subsidy to advanced education is available and (ii) whether or not the agent is informed about the existence and rules of such subsidy;
- Selection mechanism: whether or not unobservables in the selection process and outcomes equation are related - that is, whether there exists selection on unobservable factors other than ability.

To account for all alternatives and allow for all estimation procedures, we include four STATA datasets. They all include 200 Monte-Carlo replications of samples of 2,000 observations each. In total, each dataset contains 400,000 simulated individuals, corresponding to the same number of observations. There are two main datasets, MCdta-corr.dta and MCdta-nocorr.dta, and two auxiliary datasets, MCdta-corr-noS.dta and MCdta-nocorr-noS.dta. The first versions listed are for Stata 10; the second are for Stata 8 or 9.

- MCdta-corr.dta [10 MB]; MCdta-corr-v9.dta [10 MB]
- MCdta-nocorr.dta [10.5 MB, Stata 10]; MCdta-nocorr-v9.dta [10.5 MB]
- MCdta-corr-noS.dta [8.7 MB]; MCdta-corr-noS-v9.dta [8.7 MB]
- MCdta-nocorr-noS.dta [8.7 MB]; MCdta-nocorr-noS-v9.dta [8.7 MB]

MCdta-corr.dta and MCdta-nocorr.dta contain data for all three policy scenarios (depending on whether a subsidy to advanced education exists and is expected). The former represents the case of selection on unobservables other than ability and the latter represents the case of selection on observables and ability only. These two datasets are the basis for all estimation procedures. The following is a list of variables in each dataset:

Variable	*Description*
MCrep	Monte-Carlo replication index
i	Individual id in Monte-Carlo sample (1 to 2000)
theta	Individual ability (ranging between 0 and 1)
z	Observable in selection rule - family background (ranging between -2 and 2)
x	Observable in earnings equation - region (dummy)
y0	Potential earnings if dropping off education before advanced level
y1	Potential earnings if investing in advanced education
e_noS	Effort in preparation for test in the absence of subsidy
e_eS	Effort in preparation for test in the presence of expected subsidy
e_uS	Effort in preparation for test in the presence of unexpected subsidy
s_noS	Test score in the absence of subsidy
s_eS	Test score in the presence of expected subsidy
s_uS	Test score in the presence of unexpected subsidy
d_noS	Education attainment in the absence of subsidy (dummy)
d_eS	Education attainment in the presence of expected subsidy (dummy)
d_uS	Education attainment in the presence of unexpected subsidy (dummy)

we have both because this is generated

1 if one get educ

The two potential earnings, *y0* and *y1*, are included in the dataset. They depend on education attainment only, not on the policy scenario, and can be used together with the education variable, *d_**, to construct the observed earnings in each case.

MCdta-corr-noS.dta and MCdta-nocorr-noS.dta are used together with DID to explore the use of repeated cross sections in the estimation of returns to education. They represent a time period before the occurrence of a policy intervention amounting to the introduction of a subsidy to advanced education. Therefore, only the policy scenario with no education subsidy is considered in these datasets. MCdta-corr-noS.dta represents the case of selection on unobservables other than ability and MCdta-nocorr-noS.dta represents the case of selection on observables and ability only. The following is a list of the variables in each dataset:

Variable	*Description*
MCrep	Monte-Carlo replication index
i	Individual id in Monte-Carlo sample (1 to 2000)
theta	Individual ability (ranging between 0 and 1)
z	Observable in selection rule - family background (ranging between -2 and 2)
x	Observable in earnings equation - region (dummy)
e_noS	Effort in preparation for test in the absence of subsidy
s_noS	Test score in the absence of subsidy
d_noS	Education attainment in the absence of subsidy (dummy)

Related research areas
Econometric methods

y0	Potential earnings if dropping off education before advanced level
y1	Potential earnings if investing in advanced education

A set of STATA .do files implement each estimation procedure:

- download the STATA .do files [12 KB]

Two .do files were created for each method, labelled

- *name-of-the-method*.do and
- *name-of-the-method*-programs.do.

The former contains the main routine, which defines the dataset and variables being used, calls the estimation routines and displays the results. The latter contains two main estimation routines (together with other auxiliary routines in some cases). The first routine implements the respective estimator in a given dataset for a certain set of variables provided by the user. The second routine repeatedly applies the estimation procedure to a series of datasets to produce the Monte-Carlo results.

Back to Datasets

Search

Title (or part of title) Author surname (or part of surname) Search

Publications & Research
- Search
- Links

Events
- Coming soon
- Archive
- Terms & Conditions

In the news
- News about IFS
- Sign up to IFS news
- Newsletters

About IFS
- What we do
- Who we are
- Who we engage with
- What people say
- Contact us
- Jobs
- Membership
- Funding

Financial & Legal
- Disclaimer
- Copyright
- Accessibility
- Code of conduct

Centres
- CPP
- CEE
- EDePo
- TLRC
- ELSA
- CEMMAP
- CERA

The IFS hosts an ESRC research centre.