Readme file for:

"On the Evolution of Firm Size Distributions"
by Paolo Angelini and Andrea Generale

The raw data are in a dataset called **datasur.dta,** a Stata data file. It contains the variables of the four Mediocredito surveys and the relevant balance-sheet variables used in the paper. The Mediocredito dataset can be requested from the owner for research purposes - see the reference given in the paper. The 1992 and 1995 waves were made available to the Bank of Italy by Mediocredito in anonymous form, so that the identity of the firms cannot be traced. Firms in the 1998 and 2001 waves had a social security code, which allowed us to merge the data with another source, Cerved, to control and fill missing values in the age field. In datasur.dta we eliminated the social security code and replaced it with an anonymous firm identifier code.

THE MEDIOCREDITO ANALYSIS: VARIABLES, PROGRAMS AND DATASETS

ANNO   year
IDENTIF anonymous firm identifier. Integer number.

****
**** Variables used to obtain the survey-based definitions of financially constrained firms
****

***** 1992 SURVEY *******

d17_1,d17_2,d17_3,d17_4,d17_5: codify the answers to the following question: "The firm encountered difficulties in financing the last investment project? (multiple answers are allowed with an intensity ranking from 1 to 3; 9=missing):
d17_1   Insufficient cash flow or risk capital
d17_2   Insufficient collateral
d17_3   Insufficient long-term finance
d17_4   High cost of debt
d17_5   Other

***** 1995 SURVEY *******

NONOTTEN is equal to 1 (to 0) when the firm in the 1995 survey answered "Yes" (answered "No") to the following question: "In 1994 the firm wanted more credit, asked for it and was turned down by the bank".
AVREBACC is equal to 1 (to 0) when the firm in the 1995 survey answered "Yes" (answered "No") to the following question: "In 1994 to obtain more credit the firm would have accepted either to pay a higher loan rate or to pledge more collateral".

***** 1998 and 2001 SURVEYS *******

DESI is equal to 1 (to 2) if the firm answered "Yes" (if it answered "No") to the following question: "In the previous year the firm wanted more credit at the prevailing market conditions".

CHIE   is equal to 1 (to 2) if the firm answered "Yes" (if it answered "No") to the following question: "In the previous year the firm asked for more credit but was turned down by the bank".

PAGA   is equal to 1 (to 2) if the firm answered "Yes" (if it answered "No") to the following question: "In the previous year to obtain more credit the firm would have accepted to pay a slightly higher loan rate".

****
***** Other firm level variables
****

PAV   classification of manufacturing sectors according to Pavitt. 1=traditional sectors; 2= scale-intensive; 3=specialized suppliers; 4=high-technology.

DIPxx   where xx is the year, is the number of employees drawn from the Mediocredito surveys. Since each survey asks for no. employees in the previous 3 years, xx runs from 89 through 00 (for 2000).

ETAxx   where xx is the year (92,95,98,01) is the age of the firm in the survey year.

ETAxxn   where xx is the year (98 or 01), is the age of the firm in the survey year and is drawn from the Cerved database to fill missing values in the age field in the 1998 and 2001 surveys.

INDxx   where xx is either 92,95,98,01, flags each survey and the companion balance-sheet data. For example ind92=1 identifies the 1992 survey (referred to year 1991) and the prior 3 years of balance-sheet data.

ACQUI   equal to 1 if the firm was party to mergers and acquisition.

SCORPO   equal to 1 if the firm was created through a spinoff.

ISGRU   equals 1 if the firm is part of an industrial group.

****
**** Balance sheet variables (in thousands euro)
****

ACO   current assets
AIM   net fixed assets
SALES   sales
UTIL   earnings before financial expense
UTILN   earnings before taxes and extraordinary items
OF   financial expense
IMIM   net intangible assets
IMTE   net tangible assets

****
***** List of programs and datasets (Stata)
****

dataset.do: assembles the main datasets that are used for the analysis of the Italian data in the paper:
- loads data from datasur.dta
- creates proxies for financially constrained firms, sectoral dummies, etc.

            - drops missing, etc.

            - saves a series of intermediate datasets - csurxx.dta (xx=92, 95, 98, 01) etc.

            - saves a series of final datasets, to be used by subsequent programs**.**

open9201.dta and **closed9201.dta**: graphical analysis (figures 1 and 2 and table 1) contain only 4 years of data, corresponding to the 4 Mediocredito surveys. closed9201 contains the 2 balanced panels created using the 1992 and 1995 survey, and the 1998 and 2001 survey, respectively. A balanced panel could not be created using the 1995 and 1998 surveys because the firm identifier in the first two surveys is different from the identifier provided in the last two.

tempstime_1.dta: dataset used for ols and gmm regressions presented in table 2.
exploits the fact that in each survey some questions refer to each of the last 3 years.

fig12tav1.do: creates figures 1 and 2 of the paper and table 1.

ols.do: produces the ols regressions presented in table 2 (columns a) trough c)).

gmm.do: estimates the gmm regressions presented in table 2, columns (d) and (e).

THE WBES ANALYSIS: PROGRAMS

ocse.do: produces panel (i) of figure 3

nocse.do: produces panel (ii) of figure 3

both programs read data in wbes.dta, generated from the excel file downloaded from IFC webpage (as of May 29, 2007: http://www.ifc.org/ifcext/economics.nsf/Content/IC-wbes). The World Business Environment Survey (WBES) ©2000 The World Bank Group. In this page one also finds variable codes and country codes.