

The Business Environment and Enterprise Performance Survey (BEEPS) Panel Database, 2002-2009

November 2013

The BEEPS panel dataset is prepared by the World Bank's Enterprise Surveys team as a courtesy to the users. The EBRD added 78 observations for the 2009 survey round in Poland – none of them is a panel observation.

The BEEPS panel dataset covers BEEPS rounds from 2002 onwards. In 2002, it was not possible to interview companies that participated in BEEPS 1999 because the information has not been kept. Kosovo and Mongolia were covered for the first time in the 2008-2009 BEEPS round and are not included in the panel dataset. Turkey is also not included: in 2005, both BEEPS and Enterprise Survey were completed there. Turkey's 2009 panel is based on the Turkey 2005 Enterprise Survey, and is hence not compatible with BEEPS 2002 and 2005 panel. Users who want to include Turkey in their analysis can add it themselves.

The BEEPS survey instrument has changed significantly in 2008 compared to the previous three rounds of BEEPS in order to allow comparison of transition countries (EBRD countries of operation) with other developing countries where the World Bank is conducting Enterprise Surveys. Most surveys conducted after 2006 use stratified sampling and contain weights based on this information. Prior surveys do not contain any information regarding weights as quota sampling was used.

Due to the evolution of the survey instrument as well as methodological changes it is not possible to match all variables in the datasets. It is recommended that users thoroughly familiarize themselves with the questionnaires as well as the ***Reports on sampling and implementation*** from each of the years contained in the dataset before proceeding with the analysis.

- Panel dataset contains a variable “*panel*” that allows easy identification of panel observations. The *idstd* (unique id) variable can be linked to any of the underlying BEEPS data posted on the EBRD website and the Enterprise Surveys portal. Note that 78 observations for Poland 2009 BEEPS do not have *idstd*, but can be linked to the BEEPS data posted on the EBRD website.
- Panel dataset also contains the original country level panel ID variable that is matched across years. This variable differs from country to country but is always one of the first 5 variables in a dataset. (Examples: *serial*).
- Wherever possible, questions (variables) in the older rounds have been matched to variables in the newest round. Questions that have changed significantly over time have not been matched.
- Any variables that could not be matched are retained in their original form or dropped if they can be found in the original raw dataset. You can use *idstd* to link additional variables to the panel dataset.
- All the monetary value variables in the BEEPS panel dataset are in the original currency used in each round. It should be noted that the data from 2002 and 2005 waves have been

converted to USD by the teams implementing the survey. Data from 2007 and later waves have all monetary values expressed in local currency units (LCU).

- There are different scales used in the perceptions questions regarding obstacles to doing business. In the 2002 and 2005 data, they have been re-coded as follows: (1=0) (2=1) (3=2) (4=3) (5=-9). This is not an ideal match as there is no “Very Severe Obstacle” option for the 2002 & 2005 data. It is suggested that you use caution when making comparisons of these variables across time.
- Please note that in some cases where the labels of variables contains “year” or ”fiscal year”, then it could be understand under the context of latest questionnaire and the *year* variable. For example, the label for *d2* could be “Total sales in last fiscal year” and *year* is 2007 for that observation, then the sales value could be understand as “Total sales in 2006”. For the 2008/2009 round (year=2009), last fiscal year refers to 2007 for all observations because fieldwork started in 2008 - before the full-year figures for 2008 were available. Please refer to the questionnaire and the **Reports on sampling and implementation** if available.

The panel ID and year are already set in the data (xtset):

- panel variable: *panelid* (unbalanced)
- time variable: *year*

To understand more about the panel structure, please type in Stata: tab panel year

Panel: Firm interviewed in these years	Year of survey				Total
	2002	2005	2007	2009	
only in 2009	0	0	0	8,096	8,096
only in 2007	0	0	1,945	0	1,945
only in 2005	0	6,130	0	0	6,130
only in 2002	4,754	0	0	0	4,754
only in 2002, 05	1,025	1,025	0	0	2,050
only in 2005, 09	0	1,561	0	1,561	3,122
only in 2007, 09	0	0	7	7	14
only in 2002, 05, 09	374	374	0	374	1,122
PanelID can't be matc	0	8	0	0	8
Total	6,153	9,098	1,952	10,038	27,241

where panel can be defined as:

panel:

- 1 only in 2009
- 2 only in 2007
- 3 only in 2005
- 4 only in 2002
- 5 only in 2002, 05
- 6 only in 2005,08/9
- 7 only in 2007, 09
- 8 only in 2002,05,08/9
- 9 only in 2002, 05, 07, 09
- 10 only in 2008
- 11 only in 2005, 08
- 99 PanelID can't be matched

In this example, the panel variable is not overlapped between categories. The panel label “*2002 and 2005*” indicates firms interviewed in 2002 and 2005 only, while “*2002, 2005 and 2009*” indicates firms interviewed in 2002, 2005, and 2009 only. The 2005 and 2009 firms in these two labels are exclusively different.