**Ryazan personnel data: Description of variables**

The data file for analysis has been created based on source data by Ben Kriechel and Karolina Goraus-Tańska. The data consist of three types of information: (i) *Personnel registry records* that contain information on individuals' characteristics such as gender, age, marital status, number of children, and education, as well as information on complete work history including entry and exit date of the employment relationship, position at the firm, training received inside and outside the firm, (ii) *Monthly wages and hours worked*, and (iii) *Monthly consumer price index (CPI)* that is obtained from external data sources.

The final dataset for analysis has an individual-level panel structure with monthly information for every worker between 1990 and 2006 and contains the following variables:

1. *person_id*: The total dataset contains 677,831 observations corresponding to 10,080 distinct values of *person_id*, which reflects the number of workers with non-missing observations at least on one other variable.
2. *year*: The survey year ranges from 1990 to 2006
3. *month*: The survey month ranges from 1 to 12
4. *position_num* refers to the number of positions in which the person is employed in a given month. Upon entry to the firm, the variable *position_num* takes the value of 1. Then, any time the respondent has a job change within the firm, the *position_num* increases by 1. Most workers (62.41%) remain in the same position as when they entered the firm (i.e., *position_number* equals 1 for all years of employment within the firm.
5. *education_year*: The year of graduation
6. *education_level*: The original variable for the education level
7. *education_institution*: The name of the school/university where the highest education level was obtained
8. *marstat_ru*: The original variable for the marital status
9. *chnum_ru*: The original variable for the number of children (string)
10. *residence* contains 5 distinct values. More than 99% equal 1 or 2
11. *married_year*: The year when married
12. *child_year_1*: The year when 1st child was born
13. *child_year_2*: The year when 2nd child was born
14. *child_year_3*: The year when 3rd child was born
15. *child_year_4*: The year when 4th child was born
16. *date_start*: The date of entering the firm
17. *date_exit*: The date of exiting the firm if the respondent is no longer working or missing if still working
18. *birthdate*: The date of birth
19. *chnum*: The number of children
20. *has_children*: A dummy variable for persons that have children
21. *female*: A dummy variable for female
22. *edu*: An ordinal variable constructed based on *education_level* variable takes the values of 1 "basic general", 2 "basic professional", 3 "secondary general", 4 "secondary professional", 5 "higher incomplete", 6 "higher complete". There are also dummy variables for each education level called:

23. *basicgeneral*

24. *basicprofessional*

25. *secondarygeneral*

26. *secondaryprofessional*

27. *higherincomplete*

28. *higher*

29. *schoolyears*: The number of years at school to each education level

30. *marstat*: A categorical variable takes the values of 1 "married", 2 "single" 3 "divorced", 4 "widow". There are also dummy variables for each marital status called:

31. *married*

32. *single*

33. *divorced*

34. *widow*

35. *pos1 – pos9*: The information on work history within the firm, reflecting the type of work performed upon the entry to the firm (*pos1*) and after subsequent changes of positions (*pos2-pos9*). It contains original names in Russian referring to the occupation type (e.g., manager or accountant) and the occupation level for production workers.

36. *position*: The current occupation type (original variable for the occupation type, e.g. manager or accountant)

37. *level*: The current occupation level for production workers. There are 6 levels for auxiliary production workers and 8 levels for primary production workers.

38. *pos_start*: The date when the work in the current position started (for the first position it equals the date of entering the firm and variable *data_start*)

39. *pos_end*: The date when the work in the current position finished (for the last position it equals the variable *date_exit*, namely the date of exiting the firm if the respondent is no longer working or missing if still working)

40. *function*: The name of the function under the current position. This variable provides more detailed information regarding the work currently performed than the "position" variable that has six occupation types and the "level" variable that additionally splits out the production workers. The *function* variable contains 1362 distinct names of functions.

41. *type*: A categorical variable constructed based on "position" variable, reflecting occupation types under the following categories: 1 "Accountant", 2 "Manager", 3 "Technician/Engineer", 4 "Skilled production worker", 5 "Unskilled production worker", 6 "Service staff"

42. *age*: The age of the worker is calculated based on the information on the date of birth and survey year and month.

43. *ch1_age*: The age of the first child is calculated based on the information on the date of birth and survey year and month.

44. *ch2_age* is described the same as [43].

45. *ch3_age* is described the same as [43].

46. *ch4_age* is described the same as [43].

47. *married_t*: A dummy variable that takes the value of 1 if the person is married according to the registry variable "*married*" and if the current year is greater than the year when got married. Thus, this is time varying information on marital status constructed based on the *married_year* variable.

48. *tenure*: Tenure in years in the firm

49. ***ptenure***: Tenure in years in the current position in the firm
50. ***cpi1990***: Monthly CPI for the period of 1991-2006 where prices in 1990 are set to 100.
51. ***hours***: Hours worked in a given month (assuming that this variable reflects all hours worked, including overtime hours)
52. ***hours_leave***: Hours not worked in a given month due to e.g., annual leave, sick leave, or training.
53. ***hours_overtime***: Hours worked overtime
54. ***nightshift***: A dummy variable that takes the value of 1 if the respondent has night shifts in a given month
55. ***premiums_overtime***: This premium variable is observed for the years 1990, 1996, 2003, 2004, 2005, and 2006. The overtime premiums are mostly reported by skilled production workers (65.52%), followed by unskilled production workers (24.5%), service staff (9.38%), managers (0.29%), and technicians (0.31%). In most cases (14,451 out of 17,524, only in the years 2003-2006).
56. ***premiums_share_wage***: The percentage share of wages received by other types of workers than production workers. Positive values are reported in the years 1990-1996 and 2003-2006.
57. ***annual_premium_share_wage***: This type of premium is received for finishing the work plan earlier, thus for greater work efficiency, and also falls in the premiums called (share of wage), but it is given in rubles. This premium variable is mostly reported by the skilled and unskilled production workers (99.9%). Positive values are recorded for 1990-1996 and 2003-2006.
58. ***annual_bonus***: The overall bonus received at the end of the year by the majority of workers and by all types of workers. This bonus variable is reported for 1990 and the period 1996-2006.
59. ***wage_arrears***: Positive values are reported only for the years 1993, 1994, and 1995.
60. ***repayment***: Positive values are reported only for the years 1993, 1994, and 1995. For more than 90% of workers, the sum of their wage arrears equals the sum of repayments.
61. ***wages***: Monthly wages that are assumed to include all the bonuses and do not control for wage arrears (that reflect the wages to be received, while in some cases the receipt of the wage was postponed via wage arrears).
62. ***real_wages***: Wages in 1990 prices (using variable cpi1990).