

St. Petersburg State University
Institute of Management

DISCUSSION PAPER

Irina Merkuryeva, Elena Paramonova
Julia Bitina, Veronika Gilchenok

**ECONOMIC ANALYSIS BASED ON
MATCHED EMPLOYER-EMPLOYEE
DATA: METHODOLOGY OF DATA
COLLECTION AND RESEARCH**

#43(E)–2006

Saint Petersburg
2006

Irina Merkuryeva, Elena Paramonova, Julia Bitina, Veronika Gilchenok.
Economic Analysis Based on Matched Employer-Employee Data:
Methodology of Data Collection and Research. Discussion Paper #43(E)–
2006. St. Petersburg State University: St. Petersburg, 2006.

This working paper presents methodology for matched employer-employees panel data collection in Russian industrial enterprises. It summarizes the results of the most important Russian and international labor market research based on matched data analysis; introduces sampling procedures and defines possible matching techniques; elaborates the basic questionnaire structure for further data collection.

Irina S. Merkuryeva — Associate Professor, Department of Public Administration, School of Management, St. Petersburg State University
e-mail: merkuryeva@som.pu.ru

Elena N. Paramonova — Ph.D. Student, Department of Public Administration, School of Management, St. Petersburg State University
e-mail: paramonova@som.pu.ru

Julia M. Bitina — Student, Master in International Business Program, School of Management, St. Petersburg State University
e-mail: julia_bitina@rambler.ru

Veronika L. Gilchenok — Ph.D. Student, Department of Public Administration, School of Management, St. Petersburg State University
e-mail: gveronika@mail.ru

CONTENTS

INTRODUCTION.....	4
APPROACHES TO DATA MATCHING	4
REVIEW OF RESULTS FROM RESEARCH BASED ON MATCHED DATA ANALYSIS	11
METHODOLOGY OF DATA COLLECTION.....	20
CONCLUSIONS	28
REFERENCES.....	29

Introduction

The interest of researchers around the world to the collection and application of matched data has emerged relatively recently: the first work and publications in this direction refer to the late 80's. In the beginning of this process some researchers invested significant efforts trying to match independent sources of information on employers and employees that existed at the time (see, *inter alia*, Troske, 1998). In many cases this approach happened to be very useful and productive, yet a whole range of problems did arise and remained unresolved within this work direction (such as, for instance, disproportionate inclusion of different size organizations into the sample).

The next natural step was collection of completely new data or adaptation of existing data sources on employers and employees that by their design allowed for an opportunity of matching. As a rule, the first datasets of this type were created on the basis of national administrative and statistical resources in the countries with well developed public sector (e.g., data collected on the basis of information from national employment offices in the USA, INSEE in France, Nordic countries, and more). Unfortunately, apart from seldom exceptions, these data are not available to the external researchers not employed by the government agencies.

Even after considerable work has been done to collect matched employer-employee data, today there still exist relatively few matched datasets in the world, and access to the majority of them is heavily restricted [Hamermesh, 2007]. Even less of these datasets are panels, and fewer again are dynamically representative. At the moment there are no similar projects implemented in Russia and CIS in spite of the huge interest of researchers to the problems of transition economies and emerging markets. This paper presents methodology developed for collection of matched data on Russian industrial enterprises and their employees.

The paper has the following structure. Section 2 discusses the main matching techniques and provides examples of matched datasets development and their applications in different countries. Section 3 summarizes the most important results of empirical studies obtained with matched data in Russia and internationally. The final section presents the main methodological developments, including suggestions for survey sampling, structure of employer and employee questionnaires, possible approaches to matching, monitoring and quality control.

Approaches to data matching

The term panel data refers to a special category of longitudinal data that contain information on the same sample units observed at different points of time. Data of this type can be used in a very broad range of

applied research, including corporate governance, labor markets, personnel management, social policy, and more. Data matching presume that observations are collected on employing organizations and employed individuals at the same time. Specifics of data collection, in particular availability of matching tools, allows to formulate and solve fundamentally new empirical problems in the above mentioned areas that could not be approached using more traditional data formats.

Research of labor markets and some management problems requires joint analysis of empirical data that characterize both employing organizations and individual employees. Investigation of rules and factors determining relationships between various economic agents in the labor markets is ensured by simultaneous consideration of empirical data on employers and employees, where data matching serves as the crucial condition for the quality of obtained estimates. Intensive work on the collection and analysis of matched data in labor research began around late 80-s. Within the last 10-15 years matched datasets were created in over 15 countries of the world. During this period significant progress was achieved in the development of methodology — both of data collection and estimation.

Matching of data on employers and employees presumes availability of information that allows to link each individual to the employing organization, and vice versa. This data format serves to solve numerous applied problems, including in the first place identification of significant wage determinants relating to individual and employer specific features, estimation of the impact of individual and organization characteristics on job mobility, and wage decomposition.

There exist different approaches to the development of matched employer-employee datasets. Description of existing approaches presented in this section is heavily based on the review made by J. Abowd and F. Kramarz [Abowd, Kramarz, 1999]. There are two main criteria used for classification of matched datasets, both are based on the specifics of data collection. First, matched employer-employee data can be either cross-section or longitudinal. Second, primarily data can be collected either on employees or on employers. In the first case primary units of observation are individuals and households, while employers are matched to individuals later. In the second case the data are primarily collected on organizations that further provide the linkage to their employees. Additional problem relevant to many of matched panel datasets is dynamic representativity, as in many cases applied methods of data collection do not allow to account fully for information on employees and organizations that move in or out from the sample. The main techniques used to match the

data on employers and employees are discussed below in accordance with the underlying principles of data collection and analysis.

1. Representative cross-section on employers with representative data on employees

This section discusses data collection projects that take employing organizations as primary units of observation and supplement collected data with information on representative sample of individual employees selected within these organizations.

The first data of this type were collected in France within Wage Structure Surveys with the purpose of wage structure study, and soon this method became quite popular in many other countries. The Wage Structure Surveys were initiated by the European Statistical Office in 1966 r., two more consecutive surveys were conducted to collect data in 1972 and 1978; the data collection was terminated afterwards to be resumed somewhat later by French National Statistical Institute (INSEE).

This project collected data on the structure and level of wages of employees in manufacturing, construction and service. Data were collected within two stages. In the first stage data were obtained on a sample of employing organizations. The second stage produced a sample of employees in surveyed companies to collect individual data. The underlying population is all organizations with employment of at least ten individuals. Organizations were drawn from a unified database SIRENE that contained information on all establishments registered in France. The sample was stratified by industry, region and organization size. Employees were randomly selected within participating organizations based on the information on month and year of birth.

Today similar matched panels exist in the majority of European countries including the UK and Germany, and also in Canada and Japan. Unfortunately, these data are not dynamically representative, as they do not allow to trace individuals who change their employer between different rounds of the survey.

2. Representative cross-section on employers with non-representative data on employees

Similar to the previous case, data of this type are primarily collected on employing organizations, and at the second stage non-representative sample of workers is matched to this dataset. The worker sample usually is not representative because of the sampling rule selected for the second stage of the survey. For example, the number of individual respondents in each surveyed organization can be set based on the quotas assigned in

accordance with criteria such as company tenure, position or other social and demographic characteristics. As in the case with the first matching technique, firm sample is not representative dynamically (i.e. even when the survey contains time dimension, the sample is only representative as a cross-section at each data collection point). This problem arises as the sampling universe does not allow for changes over time.

An example of matched data created within this framework is a Panel Study of Manufacturing Establishments conducted in the UK in 1994. In this survey the sampling unit was considered as actual location of any establishment. The sample was selected based on the British Telecom telephone directory. This approach turned out to be very convenient, as researchers automatically received access to the contact information of future respondents, and also data on establishment size and industrial classification. The final sample included only manufacturing enterprises. The sample was stratified on location, size and industry.

At the next stage of data collection a series of interviews were conducted with management of selected establishments (usually senior manager, personnel manager and financial director). Information collected during the interviews covered different areas of the firm activity: production, markets, ownership structure, innovation, investments, HR policy, and financial performance. In addition, information on two employees of establishment was collected: the latest hire and one randomly selected employee.

Another example of data collected using this principle is Employment Opportunity Pilot Project conducted in the USA in 1982. The sample was based on the list of firms paying unemployment insurance taxes. The majority of surveyed companies were relatively small: 70 % of the firms employed 50 individuals or less, while large firms with the number of employees exceeding 200 accounted for only 12 % of the sample. Interviews were conducted with personnel managers in larger organizations and with owners in smaller ones. Interviews asked questions on organizational characteristics, and also personal information on two employees recently hired for similar positions.

3. Cross-section data employees with matched panel data on employers

It happens sometimes that researchers attempt to match two independent datasets on employers and employees after all data already have been collected for some other purposes. Normally in these cases the primary units of observation are individuals, while employers are matched and surveyed in the second stage. This approach was implemented by K. Troske who matched the USA data on manufacturing enterprises from

the Longitudinal Research Database with 1990 Decennial Census of Population to create Worker-Establishment Characteristics Database.

In this study matching was possible due to availability of information on location and industrial classification of employing organization in the individual records drawn from population census. Corresponding information on manufacturing establishments was taken from the register of active US establishments that also included industrial classification and location. Matching was implemented in four stages. In the first stage, industrial classification and location codes in the two data sources were standardized. Second, all enterprises with coincident location and industrial codes were removed from the sample (the remaining sample presumed one-to-one relation between industrial codes and location). Third, individual respondents were matched to establishments based to the industry-location codes. Only inputs with complete information available were used for matching. Finally, all matches that resulted in employment exceeding the figure officially reported by the company were removed. The resulting dataset contained information on 200 207 workers and 16 197 manufacturing establishments. Its main problem was bias in favor of large enterprises and organizations located in urban areas that were overrepresented in the sample. Due to this fact, the sample produces excessive rates of male, white, educated workers comparing to the original census data. Approach developed by Troske was later applied by K. Bayard et al. [Bayard et al., 1999] to create extended matched cross-section dataset of establishments occupied in manufacturing as well as in other sectors.

In Russia a project matching available individual data with the employers was implemented by D. Brown and J. Earle [Brown, Earle, 2003]. This project traced individuals surveyed in RLMS panel to the employing manufacturing enterprises; subsequent supplementary survey of these employers was conducted to create match dataset that allowed to study the problems of job and workers reallocation in the Russian industry.

4. Administrative representative matched employer-employee panels

Administrative data, such as national statistical and fiscal databases, are often used to create matched panel data. In this section some distinctive examples of matched panels originating from administrative sources are discussed.

All states of the US maintain quarterly databases on employment and earnings in manufacturing covering over 90% of employed individuals. These databases were created in order to manage state programs of unemployment benefits. Additional information contained in these databases include worker and employer identification numbers, earnings

structure, industrial classification, average monthly employment, total wage bill, etc. Based on these files, some states have created and surveyed random samples of the workers. Information from the state unemployment insurance records was also merged to the Current Population Survey data [Lane et al., 2001].

In France similar dataset was developed based on the information collected by the Tax and Social Security Authorities in order to compare individual declarations with the data provided by employers. Statistical office of Denmark maintains a database on firms and workers (IDA) based on annual administrative registration of the population since 1980. This is probably the longest existing matched panel [Hamermesh, 2007]. It contains substantial amount of information on individuals, yet information on employers is much less detailed. In Japan the basis for administrative matched data is annual census of establishments that can be linked to individual earnings records.

German employers annually submit information on their workers for the purposes of social and health insurance. Between 1975 and 1990 about 1 % of these data were used as a sample basis by the Labor Institute. Collected information included gender, nationality, education, earnings, reasons for termination of employment if it took place. Employer related information included identification number, size, etc. This information can be supplemented with administrative records from unemployment database that covers periods of unemployment (duration, amount of benefits received, trainings, etc.). The two sources combined provide for a comprehensive view of the individual career path.

In the post-soviet area an extremely well developed administrative matched dataset is maintained by the Hungarian government dating back to 1986 and including observations on 1.35 million workers and 21 238 employers [Earle and Telegdy, 2007].

However, in general administrative data sources are often limited in terms of recorded information. Access to these data and collection of additional information on individuals or establishments may be restricted legislatively due to the privacy considerations.

5 Representative matched employer-employee panels (non-administrative origin)

Non-administrative matched panels are created as a result of statistical surveys design for implementation of specific projects. One of the most successful datasets of this type developed for research purposes is French Labor Force Survey conducted on annual basis by the French National Statistical Institute. The universe sampled is the households residing in metropolitan France. The sample is annually rotated by 30%. Information

in the dataset includes answers to a number typical labor research questions as well as some less standard things, such as age at completion of education, last occupation of both parents, position in organization and specific aspects if performed work, etc. In addition each respondent provides information on location of employing organization which can be transformed into unique establishment code in SIRENE database. This procedure allows to match the data of the Labor Force Survey with virtually any other source of information on employers. The crucial property of this dataset is that it is dynamically representative, as the sampling procedure is based on housing information that allows to trace individuals regardless of their place of employment.

In the USA dynamically representative matched longitudinal dataset was constructed for the purposes of the National Longitudinal Survey of Youth 1979 Cohort. Identification of employer was a serious methodological problem in the development of this dataset. To solve this problem, all employer related information submitted by the respondent was collected, accounting for up to five employers within each year (including primary and secondary employment, as well as different types of employers). Company name was used as employer identifier. To verify provided information, the databases Compustat and Dun & Bradstreet were used.

In Russia longitudinal dynamically representative dataset could potentially be constructed on the basis of Population Employment Survey matched to the respective data on employing establishments, however for the purposes of labor studies information on individuals should be supplemented with a number of crucial characteristics, wages to be mentioned in the first place, and longitudinal component of the dataset should be established.

6. Non-representative cross-section and longitudinal employer-employee data

Sometimes when matched datasets are created researchers do not necessarily require them to be representative dynamically. For instant, many business and professional associations conduct regular wage surveys that in fact can be considered as matched datasets and used in research. These datasets usually contain information on the level of employee earnings and their structure (bonuses, premiums, etc.). The data also contain information on organization, such as industry, employment, additional or company specific payments. The scope of information collected in such projects depends on its goal. It can be one or few adjacent industries, region or different professional groups. Industrial wage surveys are very diverse. Usually the dataset includes a whole range of various

occupations, thus allowing to distinguish clearly between different types of professional activity and to obtain detailed information. Projects aimed to study specific activity usually have a broad (mainly national) geographic coverage. This is justified by high level of mobility that leads to the development of national labor market by the type of profession. In such projects a lot of attention is usually paid to the information on education and work experience.

A distinguishing feature of all wage surveys mentioned above is very detailed job description. However, these data may contain little demographic information. Finally, the closest occupations are usually grouped into more aggregated categories. Therefore, in spite of significance of wage surveys for informational purposes of specific market agents, their application in research is quite limited.

An example of dataset that is not representative dynamically is used in the paper by C. Brown and L. Medoff [Brown, Medoff, 1996]. This is consumer survey conducted by research center at the University of Michigan. In this survey respondents were asked supplementary questions on their employer and work. This employer related information was verified using Dun and Bradstreet database. After cleaning large employers were overrepresented in the dataset comparing to the original sample. The data were used to study the impact of organizational age on earnings. Therefore, matched data that are not representative dynamically are still quite actively used in applied labor research.

Review of results from research based on matched data analysis

Analysis of matched firm employee data provides an opportunity to verify models where information on both sides is equally important. Matched data were used to study wage determinants and earnings, employment duration, and relationships between company performance and employee characteristics. Results from the following applications are discussed below:

- *Wage determinants.* In most studies wage determinants can be split into two groups: those related to the characteristics of individuals and of employing firms. Sometimes an additional group — properties of position within organization — is added. One of important organizational factors used in these studies is the size of employing organization. A whole range of works is devoted to the impact firm size has on the wage level, in particular whether and why larger organization tend to pay more to their employees. In addition some attempts are made to decompose wage by the

contribution of various groups of factors in order to identify the one with the highest contribution.

- Determinants of *employment duration* and is another important group of studies. This direction also distinguishes between two groups of factors related to the firm and workers. In particular, dependence between company and individual characteristics and employment spells is studied.

- The studies of *unemployment duration and individual job search* display significant interest to the impact of public policy (such as size and eligibility for unemployment benefits, regulation of employment contracts, etc.) on individual behavior in the labor market.

- The impact of personnel education and training programs on the productivity and wage dynamics.* In this case efficiency of on the job training for the third party — future employer — is studied. The productivity of new entrants is studied depending of previous experience and level of education. In addition, internal mobility of workers depending on various social and demographic characteristics is of particular interest.

- HR management.* In this case the interest is in the impact of organizational life cycle on the company HR policy, as well as HR policy itself on the company performance. The worker flows are studied, and also creation and destruction of job places as an attribute of personnel policy.

- Discrimination in the labor market* can also be studied using matched data. Here significant factors that influence the level of wages, presumably unrelated to the worker productivity or organizational specifics are studied (such as gender, race, age, etc.).

Below we provide a brief review of the main results obtained in empirical research in these areas.

Wage determinants and labor productivity

This group includes studies that investigate the impact of employee characteristics and level of payment on the firm performance. Productivity is usually measured as output or added value per employee, or directly through specification of production function.

J. Hellerstein, D. Neumark and K. Troske [1996] in their work study relative productivity of workers with different characteristics, evaluated on the basis of production function that is compared with relative amount of pay estimated from wage equation. This paper suggests various

specifications of production function that allow to define the impact of individual characteristics of employees (gender, race, marital status, education) on marginal productivity. The main conclusions of this study show that married workers on average earn more than single ones, and that there exists significant difference in their labor productivity. Meanwhile, productivity of older employees (aged 35–54) is not different from the productivity of younger ones, yet they get higher wages. Age premium for employees between 55 and 64 exceeds productivity wage premium. Another conclusion concerned wage differentiation, in particular, lower pay of Afro-Americans is related to the equivalent differences in labor productivity. Difference in pay between males and females (where men get higher wages) is not related to the differences in labor productivity.

T. Haegeland and T. Klette [1999] estimated a model wage determination and labor productivity in similar way using data on Norway. They discovered that education premium, except for the workers with the lowest education level, is directly related to the differences in labor productivity. Therefore, workers with higher level of education get wages that exceed their productivity, while inverse statement is true for workers with short work experience.

Using firm regression estimates, J. Leonard, B. Mulkay and M. Audenrode [1999] investigated relationships between company compensation policy and its productivity measured as added value per one employee. The authors repeatedly estimated the same firm coefficients in different years and discovered that firms with higher level of pay returns on labor for male white collar employees provide an evidence of higher productivity.

D. Blanchflower and S. Machin [1996] investigated the impact of market competition on the level of wages and production in the UK and Australia. Research hypothesis suggested that increasing competition results in labor productivity growth and decline in wage rates. The main result of this research concluded that the impact of market competition on wages and labor productivity is very limited. In the UK labor productivity did not increase with competition, while hypothesis was supported on Australian data. The impact of competition on wages complied with hypothesis (their level dropped at competition increase), but not very strong; significant effect was only noticeable for certain groups of qualified workers.

Firm size and employee characteristics

J. Abowd, F. Kramarz and D. Margolis [1999] investigated relationship between the level of pay, employee and firm characteristics, controlling for various elements based on the annual data of Salary

Declarations in France (Declarations Annuelles des Salaires, DAS). Research supported the following hypotheses: (1) hypothesis on the presence of wage differences within and between various organizations that are related to employee specific features and firm size; (2) hypothesis on the relationship between work experience and wage rate; (3) hypotheses on the relationship between pay structure, productivity and profitability of the firm. The following conclusions were made from the study. First, there are significant differences between firms and employees, with individual factors playing more important role. Second, differences between employees within industry usually are more significant than differences between firms within the same industry. Third, at the level of employees the impact of personality has significant impact of differentiation of wage structure, while the impact of company is not so pronounced. Finally, companies that higher more high wage earners are more productive, but not more profitable; they are also more capital and labor intensive in terms of highly qualified labor force. Overall about 90% of wage differences within industry and about 75% within firm are explained by worker specific characteristics.

Productivity and employment duration

F. Kramarz and S. Roux [1998] investigated relationship between the structure of company tenure and firm productivity using 1976-1995 data. The authors were among the first to analyze the impact of hiring and firing decisions on company performance measured by productivity and profitability, as well as by capital structure and labor composition. Using information from the company balance sheet and qualification of employees, the authors estimated parameters of various models capturing the structure of in-company employment tenure and results of its performance. The results of this research have shown that low labor turnover is related to higher performance indicators, while high labor turnover have negative impact on company performance.

New technology

In the 1980s the USA faced significant changes in wage structure, while Western Europe struggled with unemployment growth. Many analysts blamed technology shock that influenced two regions in opposite ways to be responsible for these trends. Many problems were assigned to computerization, especially after some research confirmed the statement that experienced PC users gain higher paid jobs than the workers without computer skills. Researchers on both continents tried to understand the nature of wage premium related to computer literacy. Research that used data from the USA and France confirmed that the users of new technology

used to receive higher wages even prior to the use of technology [Entorf, Kramarz, 1997; Entorf, Gollac, Kramarz, 1999], or that the firms that actively introduce new technologies tend used to hire higher paid workers even before technology implementation [Doms, Dunne, Troske, 1997].

For example, H. Entorf and F. Kramarz [1998] used various sources of French data trying to define whether employees who use new technology get higher wages than the others. In case of positive answer the following question was asked: is higher wage of these workers related directly to their competence in using new technology, or did they also receive higher wages before introduction of technology? Results have shown that the users of new technologies earn higher wage rates, yet when individual characteristics of these employees are controlled for new technology premium partly vanishes, and the remaining part of it is positively related to work experience. This result is explained by the observation that for implementation of new technologies firms prefer to select the best employees who used to get higher wages prior to the introduction of innovations.

Yet in another research H. Entorf, M. Gollac and F. Kramarz [1999] advance hypothesis that the users of computer technologies on average get higher wages, and in comparison with other employees are also better protected from job loss when companies face with temporary difficulties. Research have shown that the users of computer technologies on average receive a premium of 15–20 %, while the highest wage rate growth is associated with the second and the third years of new technology usage. Nevertheless, net wage growth associated with use of computer technologies does not exceed 2%. The users of computer technologies are better protected from job loss in case if the company difficulties do not last for too long.

M. Doms, T. Dunne and K. Troske [1997] used the database of employer-employee characteristics together with the results of research of Manufacturing Technologies conducted in 1988 and containing information on the level of new technology usage on American manufacturing enterprises. New technology comprised manufacturing technological products, such as robots, digital design, lasers, various information networks, automated systems and computers used in manufacturing platforms. In order to estimate the level of technological development in organization, the authors accounted for diversity and quantity of technologies in use. They further checked dependency between technological diversity and composition of the workforce. As information on usage of these technologies by employees was not available, the authors elaborated a set of indicators measuring educational and professional composition of the firm workforce. Research results have shown that

organizations that use more advanced technology as a rule higher more qualified and educated employees. Using the same data, the authors investigated relationship between the level of pay and usage of new technologies. Analysis was conducted for different groups of employees (blue collars, white collars, management and administration) and have shown that the use of new technologies is related to higher wage premiums even when other employee characteristics are controlled for.

Job creation and job destruction

J. Abowd, P. Corbel and F. Kramarz [1999] used data on labor turnover in the French companies to analyze the differences between workers flow, job creation and job destruction, the role of employee skills in hiring and separation decisions, compared importance of long term and short term contracts in the process of labor adjustments and cycles in labor turnover compared to job creation and job destruction. The authors measured labor turnover and job creation/destruction at firm level, controlling for worker skills and availability of contractual agreements at hiring and separation.

K. Albaek and B. Sørensen [1999] investigated in their work labor turnover and dynamics of job creation/destruction in manufacturing sector in Denmark, their relationship, and also relationship between separation and hiring decisions. The sample was drawn from IDA database (Integrated Database for Labour Market Research) that covers all production units and individuals within 1980–91 period. The data contain information on job placement and labor turnover in each plant. Researchers attempted to answer two questions: (1) is it true that employees are less likely to quit growing plants comparing to those with stable hiring rates and (2) do declining plants hire equivalent number of workers to replace separating ones, or do they switch from workers at destructed places to those in continuing ones? The results have shown that, first, both types of employers have roughly equivalent separation rates (for large manufacturers this result is different — when they grow, the number of separations drop sharply). Second, investigation of relationship between the number of workers hired/separated with jobs created/destructed have shown that (1) even large amount of jobs created does not influence significantly the number of separations, (2) declining producers do not switch workers to the remaining jobs, they are more likely to hire new workers to replace old ones. This means that in many cases workers are very highly specialized and can not be easily switched between different jobs (higher heterogeneity between workers). (3) Relationship between separations and employment level is positive and significant only in case of small enterprises, while it can even be inverse for the large ones.

Investments in human capital

Matched data provide broad opportunities for investigation of questions related to human resources management and company investments in the development of human capital. J. Bishop [1994] has studied the impact of work experience and education among newly hired employees in the USA. The theory underlying this research suggested that on the job learning creates externalities that are revealed in the opportunity for one company to use knowledge of employee gained from training with another employer. Research hypotheses were tested on two samples: EOPP-NCRV and data collected by the National Federation of Independent Business. Both datasets comprised survey results of managers and company owners (all companies belong to small and medium business) and allowed to compare various indicators and characteristics of the two most recent company hires. Obtained results have shown that new employees with previous work experience corresponding to the new job, and also previous relevant education and training are more productive in the first period of employment with the new company in comparison with new workers who have irrelevant experience and no specialized education. In addition the costs of subsequent learning of such workers are lower. The study also investigated the level of employer rationality at the new hire. Mainly employers do not have information on specialized education obtained by the candidate, so these employees are not offered higher wages.

The theory of human capital suggests that education of employees by the companies results in lower starting wage, yet subsequent wage growth rates are higher than in cases with no education. However, many papers on this issue could not reveal significant negative relationship between education and starting wage offer. For example, J. Barron, M. Berger and D. Black [2001] investigated relationship between wages, labor productivity and on the job learning in the USA. The data were collected in two rounds (1980-1982 and 1992) and surveyed one most recent hire from each employer. Research attempted to account for unobserved employee characteristics, measured indirectly from information on the level of difficulty of the tasks performed and company duration of employee search for given position. Research results have shown that relationship between employee education and starting wage offer is significant when factors determining unobservable worker specific characteristics are included. In addition, it was found that education has significant impact on labor productivity. Research have shown that the impact of education on wage growth is much lower comparing to its impact on labor productivity.

Company age and level of wages

C. Brown and L. Medoff [1989] investigated organizational factors that determine the level of compensation. Investigated factors included organization size (number of employees), industry, unionization of workers. This project innovation was in the investigation of the impact of company age on the level of wages and availability of additional benefits. Research hypothesis suggested that older companies tend to provide higher compensation to its employees.

The dataset was based on monthly surveys of US population conducted by Michigan University. A sample of 1 410 individuals employed in private sector was drawn from the households. The results have shown that there exists moderate relationship between organization age and compensation.

U. Brixy, S. Kohaut and K. Schnabel [2004] studied the impact of organization age on the level of employees wages in Eastern and Western Germany. The sample contained information on the companies drawn from two databases (matching was conducted by company identification code available in both databases). The first database contain information of mandatory statistical reporting submitted by all establishments for the purposes of social security: number of employees, age, gender, wages, qualification, etc. The second contains more detailed data on a sample of organizations from the first database (organizational form, profitability, technology, etc.), acquired in course of interviews with managers. The authors conclude that younger companies pay less than older ones, yet significance of the difference disappears within 5 years. It is noticed that the difference in pay between the youngest and the oldest organizations is lower for smaller organizations with employment up to 200 people.

Labor turnover

The impact of various groups of factors on employment duration at one position was studied by J. Lane, L. Salmon and J. Spletzer [2001]. Sample included 2191 British organizations employing more than 10 people. In each organization interview were taken with executive manager, union manager and a random sample of employees. The authors defined two groups of factors: related to the characteristics of employees and related to the characteristics of organization and position (supply and demand factors). Demand factors include among all macroeconomic indicators that influence firm behavior in the labor market. Results revealed dependency between employee mobility and his personal characteristics: mobility increased with education and age. Mobility was also determined by the following factors related to job and organization:

- education of employees within organization reduces working hours, availability of part time work increases employment duration,
- workers employed part time retire earlier,
- workers employed at intellectual jobs work longer.

For macroeconomic characteristics it was determined that growth of labor demand results in decline of employment duration, while unemployment growth gives an opposite effect.

R. Topel and M. Ward [1992] studied mobility and employee behavior in the American labor market during the first 10 years of their careers. Research was conducted on panel data collected for social security authorities quarterly between 1957 and 1972. The dataset contained information on individuals who are matched to their employers (including number of employees, industrial classification, location, etc.). Individual data included gender, race, age, education and working hours. The main conclusions evidence that the first ten years of employment are associated with the most intensive process of job change that gradually slows down in later years. Intensity of job change also declines with wage growth. During the first years of work wage level plays significant role in determining the probability of job change. Visible wage growth is observed for employees working for the same enterprise. Higher match between employer and employee expectations significantly increase the probability of prolonged tenure.

Factors that influence time required to hire employee were also studied by K. Mumford and P. Smith [2002]. The authors used data of British Workplace Employee Relations Survey conducted between 1997 and 1998. In each inspected company an interview was taken with manager, union representative and employees. The authors defined two groups of factors determining search time: supply factors (individual characteristics of workers defining his mobility and opportunity to choose employer) and demand factors (organizational and job characteristics). In addition two demand factors capturing general economic environment were introduced. Results have revealed significance of individual worker characteristics: gender, age, education, children, race. Significant organizational determinants include organization age, number of employees, industry. General economic factors are also significant: declining output and increasing unemployment increase employment duration.

Another research by the same authors was dedicated to cross-country comparison of factors determining employment duration [Mumford, Smith, 2004] based on the UK and Australia data (British Workplace Employee Relations Survey and Australian Work Place Relations Survey, 1995). The

research was made possible by similar methodology of data collection applied in the two countries. Supply and demand factors were distinguished among all factors determining employment duration. Obtained conclusions were quite similar to the previous research. The differences between Australia and the UK was in the most important determinants of employment duration (gender and race) are completely compensated by fixed impact of employer characteristics in the UK and not completely compensated in Australia, probably due to the differences in labor legislation of the two countries.

S. Burgess, J. Lane and D. Stevens [1999] discovered relationship between company personnel policy and its stage of life cycle, and also studied the impact of labor mobility on sustainability and survival of the firms. The data used were collected quarterly by the State of Massachusetts on all firms in accordance with unemployment legislation. The data are only available on the number of employees and payrolls. Apart from that each firm has a code reflecting its registration date. Conclusions suggested that labor mobility is higher for among newly created and dying firms. New companies with lower labor mobility have higher survival chances. As firms develop, requirements to personnel become more elaborated and policies are optimized, resulting in reduced turnover.

Discrimination

The problem of discrimination was touched in research by J. Hellerstein and D. Neumark [2005]. The paper investigates gender and race discrimination in the US labor market based on the hypothesis that employers perceive representatives of discriminated population groups as employees with low productivity. Research matched the data from Worker-Establishment Characteristics Database and Longitudinal Research Database. The results have shown that discrimination was insignificant for African Americans, gender discrimination of women is recognized as significant and unrelated to the differences in productivity.

Methodology of data collection

The overall purpose of the survey is to collect matched worker-employer data on Russian industrial sector to study the determinants of supply and demand in industrial labor market. Suggested survey of enterprises, their managers and workers is supposed to cover a broad range of topics, including corporate governance, ownership, performance, employment, wages, costs, social security at the enterprise, technology and innovation. Discussion of the main survey characteristics follows.

1. Target population and sampling frame

Development of matched dataset presumes collection of data on employing organizations and on individual workers. In accordance with this dual purpose, two target populations are defined for the survey.

The first is medium and large size industrial firms. Small business (defined by the Russian legislation as commercial organizations with a maximum share of any single owner belonging to the government, NGO or large company limited to 25% and employing at most 100 people for industrial sector) was excluded from the survey due to the deficiencies of available sampling frames addressed below. Industrial firms are defined as those with the main occupation having the codes C, D and E in The Russian classifier of economic activities (OKVED) — see Table 1, regardless of ownership type. The second target population consists of individuals employed by these organizations.

Table 1

Industrial economic activities included in the survey

No	OKVED Code	Economic activity
1	CA	Mining and quarrying of energy producing materials
2	CB	Mining and quarrying, except of energy producing materials
3	DA	Manufacture of food products, beverages and tobacco
4	DB	Manufacture of textiles
5	DC	Manufacture of leather and leather products
6	DD	Manufacture of wood and wood products
7	DE	Manufacture of pulp, paper and paper products; publishing and printing
8	DF	Manufacture of coke, refined petroleum products and nuclear fuel
9	DG	Manufacture of chemicals, chemical products and man-made fibres
10	DH	Manufacture of rubber and plastic products
11	DI	Manufacture of other non-metallic mineral products
12	DJ	Manufacture of basic metals and fabricated metal products
13	DK	Manufacture of machinery and equipment n.e.c
14	DL	Manufacture of electrical and optical equipment
15	DM	Manufacture of transport equipment
16	DN	Manufacturing n.e.c.
17	E	Electricity, gas and water supply

Sampling of enterprises appears to be one of the most complicated methodological tasks in this survey. The difficulties arise in the first place from the lack or limited access to the available listings of registered organizations. Matched data collection projects (France, Great Britain, Germany, Australia, USA, etc.) frequently use the data obtainable from national statistical agencies, tax and employment authorities as a basis for their sampling frames. The main requirements for these sources of data are completeness and availability for use, and in many countries existing resources comply with these requirements. Moreover, some countries (for instance, France) have established consistent system in which statistical data collection at national level includes assignment of individual codes to each organization and employee that allow easy matching anytime a need arises. With this system matching is guaranteed at pre-research stage: researcher can create a sample of organizations and employees after the data were and match them using the identification codes.

In Russia listings are not well developed, consequently information on organizations is often incomplete or even unavailable. For this survey requiring a national sample listing is a crucial problem, as no complete reliable list of Russian industrial organizations exists in open access. Most part of publicly available statistical information includes only insignificant part of organizations, sampled units are not selected randomly and do not cover all regions of the country.

The most complete (though very limited for public access) source for sampling frame is Statistical Registry of Organizations Operating in Russia. This source contains information on around 7 000 000 organizations registered in the Russian Federation and supposedly provides exhaustive coverage. Available information includes registration data, identification codes corresponding to the national statistical standards, the main financial and accounting indicators. The structure of Registry allows to formulate queries, to obtain information on complete or limited set of indicators and to select organizations based on a set of criteria, such as regions, industry or ownership type.

Two main deficiencies prevent us from using Statistical Registry as a sampling frame for this study. First, although the registry covers all operating organizations, procedures required to control quality and to remove defunct units are not well developed. This means that a random sample of firms from the Registry would provide a high share of nonexistent organizations or imprecise locations and contact information, complicating implementation of survey and determination of sample size. Second, information on some important characteristics is not available or not obtainable in full for all registered organizations. In particular, data on

employment that is required for firm weighting is only available from the Industrial Registry that is limited to 30 000 large and medium size industrial enterprises.

Based on these considerations, the resulting sampling frame for firms is medium and large size industrial enterprises listed in the Russian Industrial Registry. Unfortunately, no source of comparable quality was identified for small organizations, and they were dropped from the study. Correspondingly, the main deficiency of this frame is exclusion of small businesses. Another potential frame deficiency is undercoverage that may also occur as it takes substantial time for a newly established organization to be included into the registry; therefore, young organizations are more likely to evade from the survey.

For workers the sampling frame is lists of employees obtainable from personnel departments of the surveyed firms, normally available in electronic formats as it follows from the pilot studies.

2. Sample

The goal of the project is to collect data allowing to make statistical inference on the entire industrial sector in Russia. Accomplishment of this goal requires substantial sample size and application of probability sampling procedures. Drawing a sample of firms from the entire country is especially complicated due to the high regional diversity. Taking into consideration substantial specifics related to the geographic location, market and production structure, legislation and infrastructure, and also the structure of industrial production, regional differences are especially important in the case of Russia. That's why stratified sampling is suggested to account for different specifics of the locations. Stratified sampling is also applied to reduce survey error at given level of costs.

Overall, the sample design can be described as multistage, stratified, clustered area probability sample of primary sampling units; simple random sample of firms drawn within primary sampling units. The target sample size is about 1000 enterprises, 10 000 employees. Oversampling estimated as doubling the target sample size is suggested to account for non-response rates that can be substantial in worker data collection. The survey foresees extension to a longitudinal panel survey of firms. The frequency of the survey is every two years. Levels of observation are firm and worker. The sampling process consists of two levels, both are addressed in more details below: (1) firm sample and (2) worker sample.

Firm sample

The general procedure is based on random selection of organizations all over Russia weighted by industrial employment in given territory. The sample of organizations is constructed as stratified probability sample of industrial firms weighted by the value of regional industrial employment. Stratification procedure consists of two stages. In the first stage the entire population was divided into mutually exclusive strata containing all elements of the population — in this case, 89 federal constituents of the Russian Federation.

After the first stage 30 territories were included into the sample. In addition, the two cities of Federal importance — Moscow and St. Petersburg — were added to the sample with certainty. Altogether, the first level of stratification resulted in the following 32 regions of the Russian Federation to be used as primary sampling units in the survey:

Altaiskii Krai, Krasnodarskii Krai, Krasnoyarskii Krai, Primorskii Krai, Stavropolskii Krai, Amurskaya Oblast, Volgograd Oblast, Nizhegorodskaya Oblast, Tver Oblast, Kaluzhskaya Oblast, Kurganskaya Oblast, St. Petersburg, Leningrad Oblast, Lipetskaya Oblast, Moscow, Moscow Oblast, Orenburgskaya Oblast, Penzenskaya Oblast, Permskiy Krai, Rostovskaya Oblast, Saratovskaya Oblast, Smolenskaya Oblast, Tambovskaya Oblast, Tomskaya Oblast, Tuskaya Oblast, Tyumenskaya Oblast, Chelyabinskaya Oblast, Kabardino-Balkarskaya Rep., Komi Rep., Tatarstan Rep., Udmurtia, Chuvashia.

At the second level of stratification administrative territories corresponding to the second level of OKATO classification were identified within selected federal regions, 82 territories were included into the sample as clusters containing the secondary sampling units.

Finally, the elements — large and medium enterprises having OKVED code C, D and E — were randomly selected within identified territories and weighted by industrial employment in the territory to obtain the size of the draw. In each sampled unit an interview with managers (executive director, personnel director, chief accountant) is to be taken. Resulting sample is capable of making inferences to be drawn on the Russian industrial sector in general.

Two possible subgroups can be identified for sampling of firms within strata: proportionate allocation to strata and disproportionate allocation to strata. In the first case the sample size for each stratum is proportional to the share of this stratum in the population. In the second case the sample size for this stratum is also proportional to the share of this stratum and standard deviation of distribution for given characteristic among the elements of the stratum. The main arguments in favor of disproportionate allocation are, first, that large strata should be better represented in the

sample; second, strata with higher standard deviation should be better represented in the sample to increase precision of the estimates.

Worker sample

The second level of sampling is selection of employees within selected firms. Statistical data collected in Russia do not provide sufficient information on the employees of industrial enterprises. Even when these data exist, as it was discussed above, there is no mechanism that allows to match worker to the employing organization. To solve this problem, a random sample of workers may be constructed when interviewing management of the enterprises using internal data on employees. In this case we obtain representative sample matched to the enterprises. The following procedure is suggested to facilitate random selection of workers directly in the course of company visit. Before an interview, a set random numbers is generated within the range 1-30, each representing the day of employee birth. Fither from a list of enterprise employees those born on the selected days are included into the survey. The main deficiency of this approach is that it does not provide a representative sample in the enterprise, yet it gives a sample representative of the population of industrial workers. According to the project requirements, the survey should cover 10 000 employees (on average 10 in each company).

In the course of preliminary investigation the issue of availability of administrative data on employees was discussed with few representatives of HR departments in manufacturing companies and also with some consulting firms that offer accounting and administrative services industrial manufacturers and therefore have more general understanding of established practices than people working in the specific firms. The main result of this study is summarized in Table 2 that shows data commonly available from administration, any additional information might require a personal interview with worker. The most guaranteed administrative source of information on workers is Employee Personal Card, as all employers in Russia are obliged to file these forms by legislation.

Optionally a sample of workers can also be stratified. Variables that can be suggested as a basis of stratification are type of position (blue or white collar worker), gender, age or education level. While making decision on stratification, the following two aspects are usually taken into consideration: heterogeneity of elements and strata and costs. Elements that are included into one stratum should be as homogeneous as possible, while the differences between strata should be maximized.

Table 2

Availability of information on employees in company administrative records

Required information	Availability and sources
1. Possibility to generate lists of employees by birth date	These lists are normally available in electronic form, yet extraction of the day from the date of birth might create additional problems, especially in large organizations.
2. Employee lists by date of birth for previous years	Can be available in larger companies, rather unlikely in smaller ones with low level of IT usage
3. Availability of unique personal codes for employees	Possible identifiers are: (1) payroll IDs – problematic as assignment of these IDs is determined by internal company practice; (2) registration number in social security system; taxpayer number
4. Typical sources of information on employees	<ul style="list-style-type: none"> • Labor books (information on occupied positions through the entire working life with dates of change) • Personal cards of employees (form T-2 attached), compulsory for use in all organizations • Pension forms • Employment agreements (hours, status) • List of staff (form T-3 attached) – pay structure • Work time chart (form T-12 attached) – not used in all organizations, but quite popular in manufacturing • Separation orders (form T-8 attached) – separation date and reason • Electronically: section “Wages and Personnel” of 1C software (standard accounting software used by the majority of organizations)
5. Information usually available electronically	Birth date, address, date of start working for the organization, current position, previous positions, dates of position change, compensation and wages, tenure

3. Questionnaire

Significant part of required information is collected on a regular basis by the Federal Statistical Service (Rosstat). This information is submitted by enterprises within compulsory procedures of data collection and can be partly used to reduce the survey costs, to form the structure of

questionnaires, and to control the quality of obtained results. Usage of standardized forms supports comparability of information obtained from different firms and facilitates the task of the respondents when answering questions, thus reducing the costs of interviewing and potentially increasing response rates.

However, a number of important characteristics are not captured by compulsory statistical and financial reporting. Personal interviews are required to fill this gap and to obtain additional detailed information, mainly of qualitative character. Typical respondents in the company are representatives of top management — executive director, chief accountant and/or personnel director.

The questionnaire used in firm survey contains covers a number of issues that can not be evaluated from standard procedures of statistical and accounting reporting. All questions were included based on the hypotheses of research, covering different aspects of company management. The questions included into firm questionnaire cover ownership structure, corporate governance, production, markets and suppliers, HR policies. These data are not contained in official reporting, however previous research has shown that there exists uniform principal understanding of these questions among the respondents, which guarantees that obtained information will be comparable and standardized across different organizations.

At least two pilot studies are required to test the questionnaires. Each section of the questionnaire is to be answered by a company representative who is the most competent in the given area. Formatting of the questionnaire is another task to be accomplished during the pilot studies.

The mode of administration is face-to-face interviews and collection of administrative records from the companies. Reporting units for different sections of the survey are company management reporting for the company, HR managers reporting for employees and individual employees reporting for themselves. Taking into consideration the need to conduct a national survey with substantial length of interviews, on the job face to face interviews were selected as the most suitable approach to data collection. In each organization an interview is taken with one or two representatives of the management team and a number of employees selected to participate in the survey.

4. Quality control

The quality is controlled at all stages of the field work. Each interviewer completes a registration sheet where she fixes the number of contact attempts, refusal reasons, dates of successful interviews and

characteristics of the respondents. These registration sheets are checked by regional coordinators and project team.

The questions are ranked according to their importance and payment received by the interviewers is differentiated accordingly. The field supervisor of regional surveying representative makes eyeball check on all submitted questionnaires and examines the document quality. Each completed questionnaire is checked by independent controller via phone call or personal visit. Questionnaires are returned for completion and reinterviewing when mistakes are identified, as incomplete questionnaires would imply growth in non-response rates—this means that no questionnaire is dismissed without taking additional try.

The heads of local interviewing teams are invited for a training session in Moscow that lasts at least few days, additional trainings for supervisor can take place in the process of data collection. Training sessions include dissemination of formal instructions, completion of training questionnaires, explanation, role playing and playing difficult situations in course of an interview. Video trainings supervised by the heads of the teams are organized for those interviewers who were not invited to the centralized training session. Additional training is conducted for data entry personnel.

All questionnaires are double entered into computer to avoid entry mistakes and analyzed to control overall quality of interviewer work. SPSS-DE is used to reduce data entry error. Project managers check and clean the dataset, revealing potential discrepancies and analyzing repetition of individual interviewer answers.

Conclusions

This paper presents the results of study of the international experience in matched worker employer data collection and application of these data in applied social and economic research. It also presents methodological developments for collection of matched data on industrial enterprises and their employees in Russia. Elaborated methodology can be easily implemented to conduct necessary field study, thus representing an opportunity to continue the project. Presented recommendations allow almost immediately to start preparations for the field work in accordance. Transition from development of methodology to actual data collection and usage would allow to increase significantly the sphere of implementation for project results in terms of publications and use in the education process.

References

1. Abowd J. M., Corbel P., Kramarz F. 1999. The entry and exit of workers and the growth of employment: an analysis of French establishments. *Review of Economics and Statistics* **81** (2): 170–187.
2. Abowd J. M., Kramarz F. 1999. The Analysis of Labor Markets Using Matched Employer-Employee Data. In: Ashenfelter O., Card D. (eds.). *Handbook of Labor economics*. Amsterdam: North Holland; 2629–2710.
3. Abowd J. M., Kramarz F., Margolis D. N. 1999. High wage workers and high wage firms. *Econometrica* **67** (2): 251–333.
4. Albaek K., Sørensen B. E. 1998. Worker flows and job flows in Danish manufacturing, 1980-91. *Economic Journal* **108** (451): 1750–1771(22).
5. Barron J. M., Berger M. C., Black D. A. 1999. Do Workers Pay for On-the-Job Training? *The Journal of Human Resources* **34** (2): 235–252.
6. Bayard K., Hellerstein J., Neumark D., Troske K. 1999. *Why are racial and ethnic wage gaps larger for men and for women? Exploring the role of segregation using the new worker-establishment characteristics database*. National Bureau of Economic Research Working Paper W6997.
7. Bishop J. 1994. The impact of previous training on productivity and wages. In: Lynch L. (eds.) *Training and the Private Sector. International Comparisons*. University of Chicago Press: Chicago; 161–200.
8. Blanchflower D., Machin S. 1996. Product market competition, wages and productivity: international evidence from establishment-level data. *Annales d'Économie et de statistique* (41-42): 219–254.
9. Brixy U., Kohaut S., Schnabel C. 2004. *Do newly founded firms pay lower wages? First evidence from Germany*. Institut für Arbeitsmarkt- und Berufsforschung Discussion paper 4.
10. Brown C, Medoff. L. J. 1996. Employer Characteristics and Work Environment. *Annales d'Économie et de Statistique* (41-42): 275–298.
11. Brown J. D., Earle J. S. 2003. The reallocation of workers and jobs in Russian industry. New evidence on measures and determinants. *Economics of Transition* **11** (2): 221–252.
12. Brown C., Medoff L. J. 1989. The employer size-wage effect. *Journal of Political Economy* (97): 1027–1059.
13. Burgess S., Lane J., Stevens D. 2000. The Reallocation of Labour and the Lifecycle of Firms. *Oxford Bulletin of Economics and Statistics* **62** (1): 885–907(23).

14. Cardoso A. R. 1997. *Company wage policies: do employer wage effects account for the rise in market inequality?* European University Institute Working Paper.
15. Doms M., Dunne T., Troske K. R. 1997. Workers, wages and technology. *Quarterly Journal of Economics* **112** (1): 253–290.
16. Erale J., Telegdy A. Ownership and wages: new evidence from linked employer-employee data in Hungary. Upjohn Institute Staff Working Paper 07-134, 2007.
17. Entorf H., Gollac M., Kramarz F. 1999. New technologies, wages and worker selection. *Journal of Labor Economics* **17** (3): 464-491.
18. Entorf H., Kramarz F. 1997. Does unmeasured ability explain the higher wage of new technology workers. *European Economic Review* **41** (8): 1489–1510.
19. Entorf H., Kramarz F. 1998. The impact of new technologies on wages: lessons from matching panels on employees and on their firms. *Economics of Innovation and New Technology* **5** (2-4): 165–197.
20. Haegeland T., Klette T. J. 1999. Do higher wages reflect higher productivity? Education, gender and experience premia in a matched plant-worker data set. In: Haltiwanger J. et al., (eds.). *The creation and analysis of employer-employee matched data*. Amsterdam: North Holland; 231–260.
21. Hamermesh D. Fun with matched firm-employee data: progress and road maps. IZA DP No. 2580, 2007.
22. Hellerstein J. K., Neumark D. 2005. *Using matched employer-employee data to study labor market discrimination*. Institut für Arbeitsmarkt- und Berufsforschung Discussion paper 1555.
23. Hellerstein J. K., Neumark D., Troske K. R. 1996. *Wages, productivity and worker characteristics: evidence from plant-level production functions and wage equations*. National Bureau of Economic Research Working Paper 5626.
24. Kramarz F., Roux S. 1998. *Within-firm seniority structure and firm performance*. Centre for Economic Performance Discussion Paper 420.
25. Lane J. I., Salmon L. A., Spletzer J. R. 2001. *Establishment Wage Differentials*. Urban Institute Research paper 410505.
26. Leonard J. S., Mulkay B., Audenrode M. 1999. Compensation policies and firm productivity, In: Haltiwanger J. et al., (eds.). *The creation and analysis of employer-employee matched data*. Amsterdam: North Holland; 79–114.

27. Mumford K., Smith P. N. 2004. Job Tenure in Britain: Employee Characteristics versus Workplace Effects. *Economica* **71** (281): 275–298.
28. Troske, Kenneth R. 1998. The worker-establishment characteristics database. In: Haltiwanger J., Manser M., Topel R., (eds.). *Labor statistics measurement issues*. University of Chicago Press: Chicago, IL.
29. Topel R. H., Ward M. P. 1992. Job mobility and the careers of young men. *Quarterly Journal of Economics* **107** (2): 439–479.