

The Roles of Employer and Employee Characteristics for Plant Productivity

Pekka Ilmakunnas
Helsinki School of Economics

Mika Maliranta
Research Institute of the Finnish Economy (ETLA)

Jari Vainiomäki
University of Tampere

*Helsinki School of Economics and Business Administration, Working Papers W-223,
1999*

Correspondence address: Pekka Ilmakunnas, Helsinki School of Economics, P.O. Box 1210,
FIN-00101 Helsinki, Finland, e-mail: ilmakunn@hkkk.fi

Abstract

Using a matched worker-plant data from Finnish manufacturing, the relationships of worker characteristics, wages, and productivity are examined. The process of linking various registers on employees and plants is described in detail. The final data set includes the characteristics of plants and their employees. The plant panel data is used for estimating productivity and wage profiles according to age and seniority. At low seniority productivity increases fast, but starts to decline early. Wage profiles are not related to productivity profiles, but continue to increase with seniority. These results support the hypothesis that human capital is not firm specific, and seniority related wages are used for incentive reasons. Various components of worker turnover have an impact on productivity growth.

1. Introduction

There is an increasing interaction of the fields of industrial and labor economics. Labor markets affect the behavior of firms in their product markets and product market performance is reflected in labor demand and wage formation (see, e.g., Nickell, Vainiomäki, and Wadhvani, 1994). We examine a topic where this connection is particularly clear: the productivity of manufacturing plants. On one hand, the interesting issue from the point of view of industrial economics is how one can explain heterogeneity in plant productivity with the fact that the “quality” of the work force varies across plants. To study this, we need to link information on employees to plant data. On the other hand, from the point of view of labor market research, an interesting issue is whether wage formation is based on productivity or on incentive considerations. To examine this, we have to link plant productivity information to data on employee age and tenure.

So far, a major problem in the research on these questions has been the difficulty of measuring the productivity of individuals, although their earnings can be measured with reasonable precision. However, register-based data sets that match information on individual employees and their employers provide a way of measuring the productivity profile of workers. We assess the effect of various human capital components on productivity by using panel data from the Finnish manufacturing plants that is extended with variables measuring average employee characteristics as well as plant level measures of worker turnover and other plant characteristics.

The creation and use of matched employer-employee data is essential for the analysis of many industrial and labor market issues (see e.g. Haltiwanger et. al., 1999). However, access to this type of information has been rare. An essential feature of the present work is that we have linked employer-employee characteristics data, and we therefore devote some attention to the issues related in creating such data. We describe in this paper the various data sources used in the linking of employees and plants, the process of linking, and the problems encountered. In this work, the solutions made in the production process of the official statistics restrict to some extent the freedom of the researchers who use the data. We discuss how different practices in the various statistics lead to incompleteness of the matching. The research topic

itself may lead to a further loss of data if some key variables are available in sufficient accuracy only for a subset of plants or individuals.

After a brief survey of the relevant economic issues related to the productivity and earnings profiles of age and tenure and the effects of worker turnover, we describe the construction of the data and variables. We then present estimation results on models where plant-level total factor productivity and wages are explained, among other variables, by the age and seniority of the employees. In this way we can compare the wage and productivity profiles obtained from the same data. We also examine how the inflow and outflow of workers have contributed to total factor productivity growth and earnings growth.

2. Age, Tenure, Turnover, and Productivity

It is important to combine information on both plant characteristics and employee characteristics in the analysis of productivity, since the composition of the work force and changes in it have several different influences. The age and experience of the work force are important control variables in explanatory models of productivity, if these characteristics vary greatly across plants. According to the conventional economic wisdom, which is supported by evidence from gerontological studies, age slows down productivity growth. In the beginning of the career of an employee, his productivity increases fast, both because of learning by doing and on-the-job training. In the later career, productivity growth slows down and may turn to a fall. The influence of seniority on productivity is less clear-cut. If skills are general and not firm specific, productivity of an employee should be as high after a job switch as before it. Short tenure should then not be reflected in low productivity, when the age of the employees is controlled. If skills were job specific, new recruits would begin with low productivity irrespective of their earlier experience.

The relationship between productivity and wage profiles has been under much discussion in theoretical and econometric studies (see Hutchens, 1989, for a survey of the issues). Models of firm specific human capital imply that in the early career wage exceeds productivity, but the productivity profile is steeper than the wage profile so that in the later career productivity exceeds the wage. Since the skills are firm specific, the firm pays the cost of training, which has to be compensated by the less steep wage profile. This also discourages the firm from

laying off experienced workers. If skills are not firm specific but general, wage and productivity profiles should be similar. On the other hand, incentive wage models (see, e.g., Lazear, 1995) suggest that to keep working incentives high to the retirement age, wages should in the early career be below productivity and in the later career above productivity. In countries with strong labor unions, wages may also rise with seniority because of the bargaining power of the insiders. The compensation systems may guarantee steady wage increases that are not directly related to productivity.

Another impact from personal characteristics to productivity comes through education. Skills acquired in education before the working career should be reflected in a productivity profile that starts at a higher level than without education. It is likely that skills acquired either through education or experience are complementary to the capital input and/or technology. Technical change can be skill-biased, since educated employees can more easily adapt to the use of new technologies (see, e.g., Bartel and Lichtenberg, 1987). With newer capital stock, a given skill should give a higher productivity. To obtain the productivity profiles of various characteristics of the personnel, it is therefore necessary to control the age or vintage of the plant.

There have been different approaches to measuring age-productivity profiles. One is to use data on wages. If wage were directly related to productivity, the age-earnings profile would also measure the productivity profile. However, this holds only in special cases. The productivity proxies that have been used include performance evaluations (Medoff and Abraham, 1980), expected present value of life-cycle compensation as a measure of the expected present value of life-cycle productivity (Kotlikoff and Gokhale, 1992), and piece rates which are directly related to productive performance (e.g., Shearer, 1996). A few studies have directly measured the productivity profile. Hellerstein and Neumark (1995), Hellerstein, Neumark and Troske (1996), and Hægeland and Klette (1998) have used information on the age distribution of employees (as well as other employee characteristics) to model the quality of labor in a production function estimated at the plant level. By estimating this jointly with a wage equation, they were able to compare age and productivity profiles directly. Our approach is slightly different. We calculate the average employee characteristics of each plant and combine this information with the data on plant characteristics to explain total factor productivity and wage.

It is also important to examine how the flows of employees affect productivity. If there is no turnover in the personnel of the firm, the observable characteristics of the employees can only change because of aging and increases in seniority. Over time, a more important change happens through the inflow and outflow of employees. Trivially, the age and educational level of the employees change when the characteristics of the inflow differ from those of the outflow. Seniority based theories of turnover suggest that because of firm specific human capital, firms would first lay off younger and less experienced workers during recessions, since they are the least productive. This would mean an increase in average age and seniority and a raise in productivity. In an expansion, new hires would include also less productive employees and productivity would decline. If human capital were not firm specific, the layoffs would be determined by productivity only and not by seniority. Worker outflow would again increase average productivity and age, but may not have a clear effect on average tenure. Inflow, in turn, might increase productivity, because firms can benefit from experience gained in other firms. In addition, legal constraints or insider power can influence the order of layoffs, which may prevent any productivity gains.

The worker flows themselves can have more indirect impacts on productivity. For example, in some efficiency wage models there are costs involved with the turnover. In empirical labor demand models, these costs are sometimes modeled through adjustment costs related to worker flows (e.g., Hamermesh and Pfann, 1996). This is not the whole story, however. Turnover is a sign of an ongoing process of matching employers and employees. The fact that some matches are broken shows that either the employer has felt that the quality of the worker does not fit the requirements of the job or the worker has concluded that the job and the wage connected to it do not match his requirements. Through the turnover of workers, firms can find the workers who are the most productive in the available jobs and high productivity workers can find the jobs that compensate for their productivity. Therefore, the matching process leads to higher productivity, as emphasized by Jovanovic (1979), Lazear (1995), McLaughlin (1991), and others. In this sense, high turnover may be positively related to productivity.

Blakemore and Hoffman (1989) have attempted to measure this effect directly. They explain productivity with job tenure and different components of labor flow (quits, layoffs, new hires,

rehires) using aggregate manufacturing and two-digit manufacturing industry data from the US. Their results support the view that firms follow the seniority layoff rules because of productivity concerns. We examine this issue using plant level data on total factor productivity change and explaining it, among other variables, by the inflow and outflow of workers and the changes in the characteristics of the work force.

3. Data Sources and The Linking Process

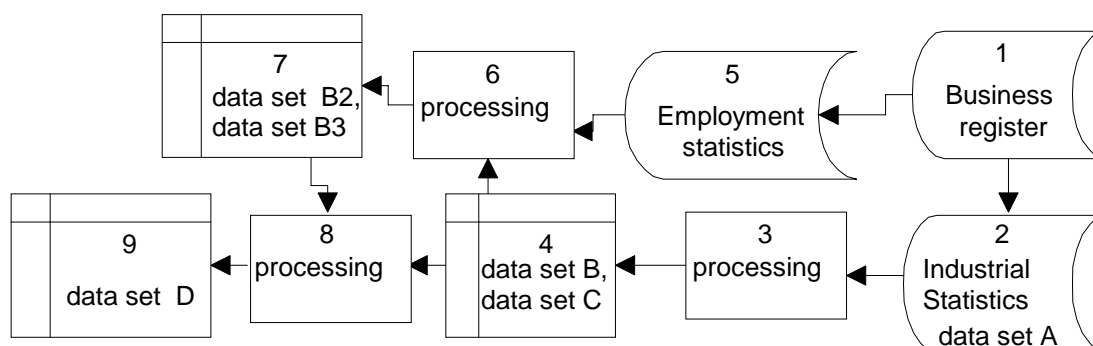
3.1 Registers

The Finnish case shows how administrative data sets can be combined to obtain linked employer-employee data. Finland, like other Scandinavian countries, has some advantages in this kind of work. The size of the country is small, which makes it possible to form registers, which cover practically the whole population of plants and employees. The linking of the registers is manageable, especially when the attention is on one sector of the economy. These complete registers can further be linked with some sample-based statistics on firms. There are also some institutional features that have made it possible to maintain large registers. There is a long tradition of collecting registers on individuals in various forms, starting from population statistics in the 18th century. There is no public hostility towards the registers, although in recent years the legislation on collecting information on individuals has been tightened. Further, the Finnish economy is very corporatist, with centralized wage negotiations and the government playing an important role through the development of the welfare state by social and tax policy. This has made it necessary and possible to maintain good records of different sectors of the economy.

The unique identification codes for persons, enterprises and plants used in different registers form the backbone of the Finnish register network whereby different sources of information can be integrated conveniently for statistics purposes. Business Register, Employment Statistics and Industrial Statistics are three basic registers maintained by Statistics Finland that are relevant for the current exercise to build a linked employer-employee data set. The process of linking is illustrated in Figure 1. The numbers refer to different stages of the work. The letters from A to D refer to different data sets constructed during the work. Data set A is

the original Industrial Statistics data, data set B contains plant characteristics of those plants that have industrial activities, data set B2 worker characteristics of those plants which can be found both in data set B and in Employment Statistics, and data set B3 is a subset of B2 after some restrictions on the data. Data set C is a subset of B for which total factor productivity can be calculated, and data set D those plants which have all the plant and worker characteristics that are needed in the regression analysis. The details of these data sets are explained below in section 3.4.

Figure 1: Registers and the linking process



1. The Business Register data base of Statistics Finland covers registered employers and enterprises subject to VAT and their plants in Finland, and it is the basic source of enterprise and plant codes used in other Statistics Finland registers and statistics. Identification codes for enterprises used in Business Register originate from tax authorities. Identification codes for plants, in turn, are given by Business Register when a new plant is established. Business Register also follows changes in the demographic structure of plants and enterprises, like their death and changes in ownership. Furthermore, Business Register includes information on the contact address, classifications like industry, and some basic variables like turnover and employment. However, this information is too limited for present purposes, so it must be supplemented from other sources. The role of Business Register for our data is that it maintains enterprise and plant codes used in our other data sources.

2. The Industrial Statistics compiles comprehensive information on the economic activity of industrial plants by annual surveys. When a plant in Business Register fulfills the selection criteria to be included in the Industrial Statistics Survey (employing at least five person being

the main criteria up to the year 1994), it is picked into the production system of Industrial Statistics. The enterprise and plant identification codes, industry group, etc. originate at this stage from Business Register. However, since then the plant's identification codes, classifications and contact information are maintained and updated, if need be, in the systems of Industrial Statistics. Therefore, it is possible that the connection with the plant's original counterpart in Business Register may weaken or disappear over time, which causes some problems when matching Industrial Statistics with other data sources that use the codes from Business Register (see below for more details). Industrial Statistics is our main source for plant level variables, like output, total employment and capital stock. The plant level data series from Industrial Statistics are available for the period 1975-1994. After 1994 there is a major break in the data collection practices, which dictates the final year. The full Industrial Statistics data base includes annually about 7000-8000 plants (data set A in Figure 1), but in our analysis we concentrate on active production plants (omitting, e.g., headquarters and auxiliary units), so our basic plant data includes approximately 6000 plants annually (data set B).

3. Information on the structure and characteristics of plants' work force is limited in Industrial Statistics, so it is obtained from our second main data source, the Employment Statistics database. This database compiles information on the economic activity of individuals and their background characteristics from a large number of different administrative registers. The enterprise and plant identification codes, industry and other general information needed in Employment Statistics are taken as such from Business Register. The employer-employee links on which our linked data rests are those determined in the Employment Statistics system. The employer-employee match in Employment Statistics is based on the register of Wages and Pensions, which includes information on all employment spells during a year of all individuals in Finland and is a part of the Employment Statistics production system. For each person a unique plant appearing in Business Register is determined as his/her primary employer during the last week of each year. This connection is traced out using the enterprise identification codes in Wages and Pensions register. For multi-unit enterprises the person-plant matches are determined using a questionnaire asking enterprises to attach persons to their appropriate plants. Furthermore, checkups and corrections are performed by comparing the geographical location of plants and the place of residence of persons. Linking an individual with the proper employer plant is a challenging task, and there remain a number of persons in Employment

Statistics whose plant code is missing or may be improper. However, a great deal of effort is being made in Employment Statistics to seek the correct plant linkage for each individual, so we consider this information to be the best available on which a linked employer-employee data can be built.

The register based Employment Statistics covers effectively the whole population of Finland. There are over 2 million employees in the register, and depending on the year some 350000-450000 work in the manufacturing sector in over 16000 plants. One could use the individuals as basic units and link the plant characteristics to them in order to study, for example, the determinants of individual incomes and changes in employment status. Alternatively, the register can be used as an augmenting source of information on the background characteristics of each plant's work force that is not available in Industrial Statistics, in order to study how they affect plant performance, like productivity in this paper. We therefore formed a plant-level panel data set from the information on individuals by calculating plant-level sums or averages of the background characteristics of the employees (data set B2 in Figure 1). This data set is the same as that used in Vainiomäki (1999), but aggregated to plant totals for the present work. The Employment Statistics database was started in 1987, but on the basis of preliminary investigations and discussions with the Employment Statistics department, there were suspicions about the data quality in the first year of the database. When worker characteristics from this source are combined with the plant level data from Industrial Statistics above, the data period is therefore 1988-94. When merging the linked worker characteristics data with plant level data from Industrial Statistics for the productivity analysis some further data requirements were made. At least two linked employees were required and a total factor productivity index had to be available (data set D in Figure 1).

3.2 Plant Codes

The linking of worker characteristics to plants in Industrial Statistics is done using the plant codes in the two sources. As indicated above, they originate from the same source, but after the initial appearance of a new plant in Business Register, the plant codes are maintained in the Industrial Statistics system. Since the objectives and data needs differ between Business Register and Industrial Statistics this may lead to some differences in plant delineation and plant identification codes in these two systems. Industrial Statistics strives for providing a

comprehensive description of the industrial activities in different industries and regions. To this end the information should ideally be surveyed from a unit that engages in one or predominantly one kind of activity at a single location, which is the basic 'definition' of a plant in Industrial Statistics. Although in theory an establishment-based survey is conducted for Industrial Statistics, in practice an "establishment-type of unit" may be used, which means in many cases some kind of mixture of a local unit and a kind-of-activity unit. In some special cases it is even allowed that an integrated whole that is defined as the statistical unit, consists of parts locating geographically detached. When the unit covers such an integrated whole it is often easier for an enterprise to give comprehensive, relevant and reliable information for the purposes of Industrial Statistics. Business Register in turn keeps record of the 'population' of business units with a limited information content on enterprises and establishment, so it is substantially easier for Business Register to stick to a stricter definition of an establishment unit.

There are also some differences in treating demographic events in Business Register and Industrial Statistics. Generally, Industrial Statistics is more reluctant to change the code of a unit that is continuing activities after a demographic event. For example, in cases when two or more units are merged, generally the code of the oldest plant is kept and the other are incorporated under it in Industrial Statistics. However, this treatment is not fully formal and also the size and the industry of the units are considered in making the decision. On the other hand, when the Business Register considers changing a plant code, it takes into account three criteria: industry, address and ownership. In principle, the plant code is changed if at least two of the above criteria change. However, in practice these criteria have been used only as guidelines for decisions made case by case. Plant code may be renewed, if a 'substantial' change has occurred only in industry or only in location. When a plant is transferred to a new owner (simple change in ownership), the plant code does not change. In cases where two or more plants are combined, the practices in Business Register have varied to some extent. In some cases, a new plant code is given to the new combined unit.

Because Industrial Statistics generally follows a more conservative policy than Business Register in changing the plant codes, there are a number of codes in Industrial Statistics that cannot be found in Business Register. Especially the older plants (which are likely to be larger as well as likely to have been involved in demographic changes) are exposed to a greater risk

that the connection with any Business Register code, and therefore with Employment Statistics, is broken down. Conversely, there are a large number of plant codes in Business Register and Employment Statistics that cannot be found in Industrial Statistics, because they do not fall within the criteria defined for the units to be included in the Industrial Statistics Survey (i.e., the small plants employing less than five persons). Furthermore, due to differences in definitions of plant delineation, a plant in Industrial Statistics may have a specific Business Register code, but in practice is a composite of several Business Register plants. In principle, also a converse situation is possible, where a Business Register plant is divided into two separate plants in Industrial Statistics. These differences in plant coding practices clearly cause some matching problems when using plant codes from Business Register and Industrial Statistics. However, the differences should not be given too much emphasis. In most cases, there are no discrepancies between the two systems, and the simple reliance on ‘administrative’ plant codes yields a correct match. More refined procedures to unify plant coding in Business Register and Industrial Statistics, based on historical records of coding changes and/or using data on individuals to form consistent plant identifiers, is a major task and was considered outside the present work.

3.3 Construction of Variables on Plant and Worker Characteristics

The list of variables available in the register-based Employment Statistics is too extensive to go through in full length here. Among other things, for each person the following information is included or can be inferred: personal identification code, identification code for the employer enterprise and plant, industry of the plant, age, marital status, education (Statistics Finland educational classification in great detail), experience (general and firm-specific), income from employment, other income, and labor force status (employed, unemployed or out of labor force). In other words, this data set offers many opportunities for investigating interesting hypotheses about the connection of employee characteristics and plant performance. In this paper, we take interest especially in such characteristics as age, education, experience, and changes in employment or labor force status. For those plants from which we have information on at least two employees, we have calculated the following average employee characteristics (in years): age (AVAGE), experience in the plant (SEN) and schooling (EDUY). The schooling years are based on detailed information on the educational degrees held by persons, which are transformed to years using years to complete the degree.

We have also measures of worker flows for each plant during successive pairs of years from the Employment Statistics database. We have calculated the number of persons who appear in the same plant in both years (stayers). Similarly, we have counted those who have exited from the plant (worker outflow) and those who have entered during the period (worker inflow). Worker inflow rate (WFIN) and worker outflow rate (WFOUT) were calculated by dividing the respective flows by the average employment in two successive years (following Davis, Haltiwanger and Schuh, 1996). Total worker flow rate (WF) is the sum of WFIN and WFOUT, and the net rate of employment change (NET) is the difference of WFIN and WFOUT. The churning rate (CHURN) measures excessive worker turnover, defined as the difference of worker flow rate WF and NET (Burgess, Lane, and Stevens, 1994), so that $CHURN = WFIN + WFOUT - |WFIN - WFOUT|$.

Industrial Statistics includes a wide variety of variables on output and inputs of plants. Output can be measured with gross output and value added. These variables are converted into real terms by using corresponding (2- or 3-digit) industry level implicit price indices obtained from Finnish National Accounts. Labor and capital inputs are of particular interest in productivity analyses. The former can be measured by hours worked or the number of persons (separate figures for production workers and salaried staff are available). Since the number of employees includes for example temporarily laid off and those on maternity leave, it is an imperfect indicator of the labor input in production. Therefore, we use the actual hours worked as the labor input measure. As for capital stock measures, they have not been included in the questionnaires since 1985. Capital input estimates were derived for a vast majority of plants with a perpetual inventory method that makes use of investment figures in Industrial Statistics. Investments are converted into real terms with implicit price deflators obtained from National Accounts. Two estimates were constructed: one for machinery and equipment and another for buildings and constructions (see Maliranta, 1997, for details). It seems that the quality of the machinery and equipment measure is superior to that of buildings and constructions. As the capital services from the former are substantial, machinery and equipment capital is preferred as a proxy of the total capital input.

As we are seeking factors affecting the productive performance of the plants, we need a suitable indicator for it. The total factor productivity is a useful measure as it incorporates

efficiency both in labor and capital usage. If this was estimated using a production function, we would encounter several difficulties, like choice of functional form, endogeneity of inputs, and errors-in-variables problems especially in the capital input. We chose instead to measure total factor productivity directly using the translog multilateral productivity index introduced by Caves, Christensen, and Trethway (1981) and Caves, Christensen, and Diewert (1982). It allows the factor elasticities to vary across plants and industries. When using this procedure we are able to pool conveniently different industries.

We derived the TFP index separately for plants in each 2- or 3-digit industry for the period 1988 through 1994. When two types of inputs are used in the production and there are constant returns to scale, the index can be calculated by the following formula:

$$\ln TFP_{it} = \ln \left(\frac{Y_{it}/L_{it}}{\tilde{Y}/\tilde{L}} \right) - \frac{S_{it} + \bar{S}}{2} \cdot \ln \left(\frac{K_{it}/L_{it}}{\tilde{K}/\tilde{L}} \right),$$

where Y denotes real value added, L labor hours and K capital input. Labor productivity of the benchmark plant is denoted here by \tilde{Y}/\tilde{L} and capital intensity by \tilde{K}/\tilde{L} . The benchmark plant is defined by the geometric means of the output (\tilde{Y}) and inputs (\tilde{L} and \tilde{K}). S_{it} , the cost share of capital input in plant i in year t , is calculated as

$$S_{it} = \frac{KCOST_{it}}{KCOST_{it} + LCOST_{it}},$$

where $KCOST_{it}$ is the (nominal) capital cost that includes depreciation and rental costs and $LCOST_{it}$ is the cost of labor, covering wages, social security and other supplementary payments. \bar{S} denotes the average capital cost share among all plants in a given industry in the whole period, and it is calculated from National Accounts by assuming 5 percent interest rate. Factor shares of capital derived in this manner are somewhat smaller than those measured with income shares. For the cost shares to agree with income shares, the implied interest rate should be some 10 – 15 percent depending on the industry. We proxy K_{it} with machinery and equipment capital stock. Total capital cost share S_{it} is estimated using information on the plant's machinery capital costs only. We have adjusted the figures so that in each industry and

in each year total capital input share in our sample of plants is in line with that calculated from National Accounts (see details in Maliranta, 1997).

The other dependent variable in the present study is the average wage in the plant (WAGE) calculated by dividing total wages paid by hours of employees. Other plant-level variables from Industrial Statistics include geographical location, the ratio of rents paid to the value of machinery, an indicator of foreign ownership, recent investments, an indicator for plants that are going to disappear ('the shadow of the death' à la Griliches and Regev, 1995), average hours per worker, and capacity utilization. For the analysis, we also classified the plants to groups according to their age. We formed six generation groups (cohorts) separately for each industry on the basis of the order of appearance of plants to Industrial Statistics. The newest two groups are decile classes and the rest of the groups are quintile groups. The generation is indicated by dummy variables GENA (newest) to GENF (oldest).

3.4 Descriptive Statistics on The Matching Process

The process of matching workers in Employment Statistics to plants in Industrial Statistics proceeded as follows. First, those persons in the full Employment Statistics data base were chosen who are over 15 years old, whose employer plant's industry is manufacturing, and for whom the plant identification code exists. This can be treated as the full Employment Statistics based 'population' of all manufacturing workers. The number of these employees has a downward trend, which has been strengthened by the recession. Starting with 446000 in 1988, the number of employees reached its minimum of 346000 in 1993 (Table 1, line 1). For several reasons, a matching plant in Industrial Statistics cannot be found for all these employees. First, Industrial Statistics includes only plants employing at least five workers, whereas Employment Statistics also includes workers in smaller plants. Second, the group of plants for which workers are linked is restricted to those plants that have production activities in Industrial Statistics (omitting plants that are headquarters, auxiliary units, etc.). Finally, due to some differences in plant coding in the two statistics, as discussed above, we include only those plants (and their employees) which had exactly the same plant code in both systems. These restrictions decrease the number of individuals in the data set by 92000 in 1988 (line 2) and somewhat less towards the end of the period. Depending on the year, this is a drop of 18 to 23 percent in the number of employees.

Table 1.**Matching Workers to Plants: Employment Statistics (ES) and Industrial Statistics (IS).**

	1988	1989	1990	1991	1992	1993	1994
1. All persons in ES	446125	445986	442819	387695	352048	346680	369583
2. ES persons in IS plants	353922	354629	352714	318476	287393	273224	285388
* share of line 1.	79.3	79.5	79.7	82.1	81.6	78.8	77.2
3. Linked persons	283831	303100	311883	279383	257425	247808	255928
* share of line 4.	65.0	70.8	75.6	72.9	74.7	76.3	78.5
4. Persons in IS	436484	427950	412737	383428	344388	324765	326217
5. Plants in IS	6316	6237	6101	6480	5941	5595	5379
6. Linked plants	5530	5651	5565	5831	5243	4943	4821
* share of line 5.	87.6	90.6	91.2	90.0	88.3	88.3	89.6

Notes for the rows:

1. All persons from Employment Statistics database: aged at least 15, industry of employment is manufacturing, plant code not empty.
2. Persons whose plant codes in ES and IS are the same, ES codes compared to the list of plant codes of active production plants in IS (data set B).
3. Persons fulfilling restrictions to be included in calculations for plant level variables: a) employed, b) wage and salary earner, c) length of employment > 1 month, d) monthly wage available and between min-max bounds.
4. All persons in active production plants in IS (data set B).
5. Active production plants in IS (data set).
6. Plants that had at least one linked worker fulfilling the restrictions required on line 3.

Before the calculation of the employee based plant level variables, individuals with very short spells of employment (under 1 month) and wage income that was likely to be erroneous (average monthly wage outside certain minimum and maximum bounds) were dropped. These restrictions amount to a further loss of persons, which was about 70000 in 1988 but clearly less in the other years (line 3). The remaining employees are used in forming the plant level variables on work force characteristics of the plants in Industrial Statistics (data set B2 in Figure 1). These linked employees account for some 65 to 78 percent of the employment figures in Industrial Statistics (line 4). In the process of the linking, we also lose some plants in Industrial Statistics because no employees can be matched to them. The number of plants lost varies from 540 to 780 plants, or from 8 to 12 percent of the number of active production plants in Industrial Statistics. The remaining plants cover around 90 percent of active plants (lines 5 and 6).

Because of the loss of employees and plants in the matching, there may be problems with the representativeness and quality of the linked data. First, the loss of employees is proportionately larger than the loss of plants, since the share of plants with linked employees is about 90 percent, but the share of linked employees is 65 to 78 percent of employment in these plants (as measured in Industrial Statistics). Either the plants that are lost in the linking process are larger than average, or the loss of employees is greater in larger plants. We explored this by examining how the number of employees of the plants differed as measured in Industrial Statistics and in Employment Statistics. The correlation between these two measures in 1994 was slightly over 0.8, so it seems that on average the matching is reasonably accurate. However, some plants are much larger according to Industrial Statistics than according to the matched employees, indicating that a substantial proportion of these plants' workers is lost in the linking process. These problems are, however, more prevalent in medium sized and small plants than in large plants. It therefore seems that large plants are under-represented in the linked data, but for the linked large plants the matching is fairly accurate. The difference in definition of plant delineation or changes in firm structures and ownership, leading to different plant codes in the two systems, may be more likely to occur in large plants. This could explain their greater loss.

In cases where the number of linked workers is smaller than in Industrial Statistics it is plausible to think that plant level variables of employee characteristics are based on a ‘sample’ of all employees in the plant. A further complication is that we sometimes have more matched workers in a plant compared to its employment in Industrial Statistics. One possible explanation for this is the different concept of employment in Employment Statistics (end of year situation) and in Industrial Statistics (average annual employment). Second, employees have been linked to one plant only, although they may have a second job at another plant. Finally, it is possible that the attachment of persons to plants in the Employment Statistics data base is simply incorrect despite all efforts to form correct matching, or the difference in plant coding systems between the two sources causes some incorrect matching. Despite these problems, this linked data provides a rich source of information on the structure of the labor force of the plants, which is unparalleled to information from any other sources.

Some descriptive statistics on plant characteristics at various data steps are presented in Table 2 for the years 1990 and 1994. The sample of plants shrinks step by step as we are making more requirements for the content of variables. The starting point is **data set A** that includes all plants in Industrial Statistics. **Data set B** excludes other plants than those having industrial activities (headquarters, auxiliary units etc.). This is the subset of plants in Industrial Statistics for which linked worker characteristics were searched. The share of active plants from all plants is about 75 to 85 percent, but in terms of employment they are larger than average, so their employment share is about 95 percent each year. **Data set C** retains only plants for which it was possible to construct the total factor productivity indicator. At this stage, the number of plants drops substantially (to about 60 to 70 percent) and average characteristics change for many reasons. First, an appropriate estimate of capital input (machinery stock), which is needed for the total factor productivity (TFP) measure, is lacking for a number of plants. These plants are typically smaller than average, so average plant size increases and the employment share remains at about 85 to 90 percent. Furthermore, we constructed the TFP indicator only for those plants whose $\ln(\text{real value added per hour})$ and $\ln(\text{real value added per machinery stock})$ did not differ too much from the corresponding industry average in that year.

Table 2.
Descriptive Statistics for plants at various data steps.

Data set	Year	Number of persons	Share of data set A (%)	Number of plants	Share of data set A (%)	Average size	Nominal value added per hour	Wage per employee
A	1990	434391		7182		60	169	68
	1994	344756		6601		52	235	80
B	1990	412737	95.0	6101	84.9	68	178	67
	1994	326217	94.6	5379	81.5	61	248	78
C	1990	370320	85.3	5005	69.7	74	183	67
	1994	296543	86.0	4317	65.4	69	261	79
D	1990	347387	80.0	4536	63.2	77	183	67
	1994	279181	81.0	3882	58.8	72	261	79

Table 3.
Descriptive statistics on worker characteristics

Variable	YEAR	Plants with worker characteristics estimate (data set B3)					Plants in regression analysis (data set D)				
		Number. of plants	MEAN	MED	P95	P5	Number. of plants	MEAN	MED	P95	P5
AGE	1990	5466	37.8	37.8	45.3	30.2	4536	37.9	38.0	45.2	30.4
	1994	4755	39.1	39.3	46.2	31.7	3882	39.2	39.4	46.1	31.9
SEN	1990	5466	8.3	7.7	17.1	1.9	4536	8.5	8.0	17.3	2.0
	1994	4755	9.4	9.0	18.3	1.8	3882	9.7	9.2	18.4	1.9
EDUY	1990	5466	10.6	10.5	11.9	9.6	4536	10.5	10.5	11.8	9.6
	1994	4755	10.8	10.7	12.4	9.7	3882	10.8	10.7	12.3	9.8

If the value of either of the two indicators differs from group average more than 4.4 standard deviations, the plant is considered to be an outlier. Finally, outliers of the TFP index are picked out in a similar manner. Due to all these restrictions average labor productivity increases further. **Data set D** is the subset of plants used in the regression analysis, and requires that both an appropriate estimate of total factor productivity and characteristics of work force are available. The average size of plants increases somewhat but other plant characteristics do not change. It is notable that labor productivity increases somewhat in each data step, except the last, but there are hardly any differences between the data sets in average wage per worker.

Table 3 presents average worker characteristics for two sets of plants. First, we required at least two individuals to calculate the average labor characteristics for the plant (data set B3 in Table 3). In 1994, there are 4755 such plants with the average labor characteristics data, but only 3882 of these plants were used in the regression analysis (data set D in Table 3). The unavailability of an acceptable total factor productivity indicator (and other variables needed in regressions) drops the number of plants considerably. Table 3 shows that average labor characteristics of plants change quite moderately due to this requirement. The average age is practically the same in the sample of plants used in the regression analysis as among all linked plants. In addition, the distributions are similar, in both data sets there are some two hundred plants annually where the average age of personnel is less than 30 years or more than 45-46 years. The difference in seniority between the two samples of plants is somewhat more notable. There is a wide range of variation in the seniority variable across plants, the 5th percentile being about two years and the 95th some 17- 19 years depending on the year. As for schooling years, the difference between the samples is insignificant. The great majority of the plants fall in the range from 9.5 to 12 years of average schooling of employees.

4. Estimation Results

4.1. Productivity, Earnings and Seniority

The estimated productivity and earnings equations have the following general specification

$$\ln(Y_{it}) = \alpha_0 + \alpha_i + \beta_1 \ln(AVAGE_{it}) + \beta_2 [\ln(AVAGE_{it})]^2 + \beta_3 \ln(SEN_{it}) + \beta_4 [\ln(SEN_{it})]^2 + \beta_5 EDUY_{it} + \beta_6 [EDUY_{it}]^2 + X_{it} \delta + \varepsilon_{it}$$

In the productivity equation the dependent variable is the log of the multilateral total factor productivity indicator $\ln(TFP)$, and in the earnings equation the log of average hourly wage $\ln(WAGE)$. The explanatory variables for plant's labor characteristics are the log of the average age of the employees ($AVAGE$), the log of average seniority (SEN), and the average years spent in schooling by the employees ($EDUY$) and squares of these variables. The only difference between productivity and wage specifications is that the square of education years was not significant in the TFP equations and was dropped from them, but it had a clearly significant coefficient in the wage models. These equations were estimated using both OLS, and random and fixed effects estimators to control for the plant specific fixed effects α_i .

In OLS and random effect estimations with levels the plant-specific other control variables X_{it} include geographical location, the ratio of rents paid to the value of machinery, indicator of foreign ownership, recent investments, 'shadow of the death', average hours per worker, and capacity utilization. In order to control the age of the establishment we use the plant generation variables ($GENA$ - $GENF$). A linear trend was included in the models for the whole period and it was allowed to vary across 4-digit industries. In addition, the recessionary period 1991-1994 was indicated with a dummy variable, and dummies were included for the 4-digit industries. In the fixed effect model the time invariant variables, geographical location and industry dummies were dropped. Since our attention in this paper is on the relationship between worker characteristics, productivity and wage, we do not report the coefficients of these plant and time variables.

Table 5.
Models for earnings levels.
Dependent variable: ln(WAGE)

	(5)		(6)		(7)		(8)	
	Pooled		Pooled		Random effects		Fixed effects	
	coefficient	t-value	coefficient	t-value	coefficient	t-value	coefficient	t-value
Intercept	-7.463	-12.75	-7.896	-12.87	-1.993	-3.42		
ln(AVAGE)	6.012	17.60	6.271	18.39	3.596	11.14	2.288	6.38
[ln(AVAGE)] ²	-0.815	-17.13	-0.851	-17.92	-0.490	-10.88	-0.312	-6.24
ln(SEN)	0.029	6.06	0.039	7.88	0.041	10.03	0.041	9.45
[ln(SEN)] ²	0.002	1.06	0.002	0.98	-0.005	-3.30	-0.012	-7.12
EDUY	0.229	11.87	0.224	11.61	0.078	4.00	-0.015	-0.68
EDUY ²	-0.007	-8.26	-0.007	-8.38	-0.002	-2.70	0.001	0.82
GENEA (ref. group)								
GENEB			-0.023	-2.66	-0.001	-0.13		
GENEC			-0.042	-5.57	-0.022	-2.50		
GENED			-0.067	-8.97	-0.044	-4.93		
GENEE			-0.066	-8.63	-0.042	-4.61		
GENEF			-0.052	-6.73	-0.008	-0.85		
GENEA*t (ref. group)								
GENEB*t			0.006	2.66	0.004	2.16	0.002	1.01
GENEC*t			0.006	2.94	0.004	2.80	0.002	1.58
GENED*t			0.006	2.93	0.003	1.93	0.001	0.68
GENEE*t			0.008	4.15	0.005	3.67	0.003	2.16
GENEF*t			0.011	5.40	0.007	4.88	0.005	3.13
Observations	30062		30062		30062		30062	
R ²	0.566		0.569				0.871	
Log likelihood					14248.08			
Note: t=trend								

We were also concerned about the correct shape of the productivity and wage profiles. The specification was tested by investigating with a non-parametric kernel smoother whether the residuals showed any relationship with the age, seniority or schooling variables. This indicated that our specification works reasonably well with the AVAGE variable at least over the range from 25 to 55 years and with the SEN variable over the whole range to 55 years.

The estimation results for the productivity equations are shown in Table 4. Columns (1) and (2) are pooled OLS estimates without plant effects. In column (3) the plant effects are modeled as random effects and in column (4) as fixed effects. Table 5 presents similar estimation results for the wage equations. In the pooled estimations, both $\ln(\text{AVAGE})$ and $\ln(\text{SEN})$ have an inverse U-shaped relationship to $\ln(\text{TFP})$, but the actual shape of the relationship depends on which control variables are included. EDUY, in contrast, has a linear relationship to $\ln(\text{TFP})$. Also the wage regressions give concave age and seniority profiles. We first consider the impact of age on productivity. Figure 2 shows how the age-productivity profile is altered when other labor and plant characteristics are controlled. Profile (1) shows the relationship between total factor productivity and age, when education, seniority and plant vintage are not included in the model (not reported in Table 4). This profile reaches its peak at 33 years. One potential explanation for the success of the plants that have young personnel comes from the fact that generally the newer generations are more educated than the older ones. The years spent in education are controlled in profile (2). Some of the difference in the productivity performance between the plants where the average age of personnel is, say, 35 years and the plants where the average age is 50 years, can be accounted for by differences in education. It is worth noting that in terms of productivity the returns to schooling seem to be substantial, some 8 percent annually.

The new plants generally have younger personnel. They also have more modern technology that might require better-educated employees than the plants that are stuck with older vintages of capital. In profile (3), the age of the plant is controlled with a dummy variable denoting the generation of the plant, and, furthermore, we have allowed for different trends for each generation. The newer plant generations are superior to the older ones, reflecting the vintage effects (see Maliranta, 1998). When this fact is controlled, the relative performance of the plants that have older personnel improves noticeably.

Finally, profile (4) demonstrates the relationship between age and productivity when log of seniority and its square are included in the model. This is model (2) in Table 4. As a large number of seniority years appear to affect productivity negatively and because the seniority years and age are positively correlated, the relative performance of the plants having personnel in advanced years improves after the control of seniority. Profile (4) seems to suggest that age in itself is not necessarily a burden in terms of low productivity, but rather the factors that are often associated with it: technology that is out of date, low turnover of workers and low education. Age-wage profiles have fairly similar concave forms as the productivity profiles, which is consistent with the implications of general human capital. When all the factors are controlled, the wage reaches its peak at the age of 40. This is shown in Figure 2 by profile (5), which is based on model (6) in Table 5.

Next, we examine the impact of seniority on productivity and wage. Figure 3 shows the productivity and wage profiles. The levels of the profiles should not be compared, since wage and total factor productivity are not measured in the same units. Instead, comparison of the slopes of the profiles can give an indication on which theories seem to be supported by the data. Productivity increases initially fast with experience in the plant. However, it reaches its peak already at 2.5 years, and thereafter it declines slowly over time, when all the worker characteristics are controlled (model (1) in Table 4). Long seniority appears to affect productivity negatively. The fact that skills are acquired fast seems to indicate that they are not firm specific. The wage profile is quite different: wages keep on increasing with seniority. The different forms of the wage and experience profiles can be interpreted in alternative ways. On one hand, the results support Lazear's (1995) view that seniority based wage may be used for keeping productivity incentives high. Another interpretation is that there is insider influence on wage determination, which is not related to productivity.

When the plant generation variables are included (model (2) in Table 4), the seniority-productivity profile shifts up. It peaks slightly later, at 3.8 years, and declines more slowly. The main conclusion remains, however, intact: productivity starts declining fairly early in the career, but the wage keeps on increasing. The coefficients of the plant generation variables show that newer vintages have higher productivity. It seems that in model (1) the seniority

Figure 2: Age-productivity and age-wage profiles

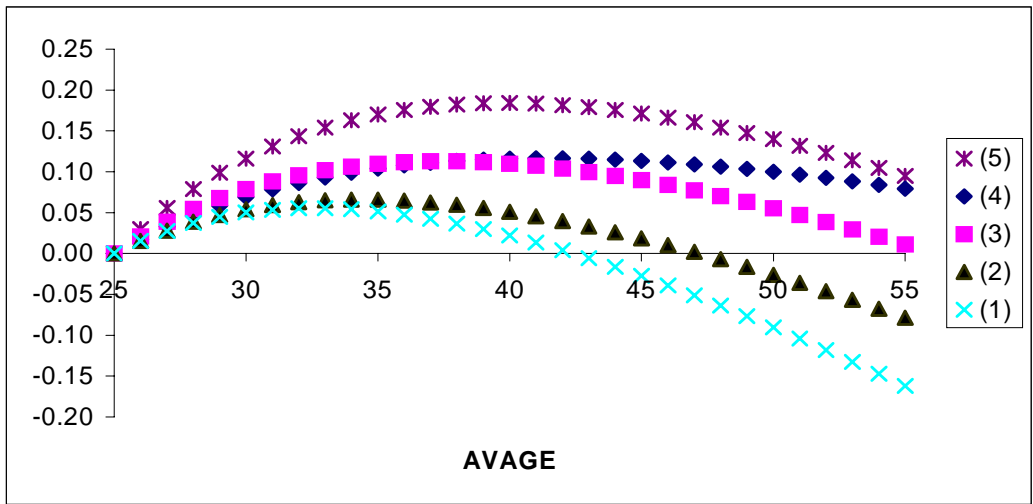
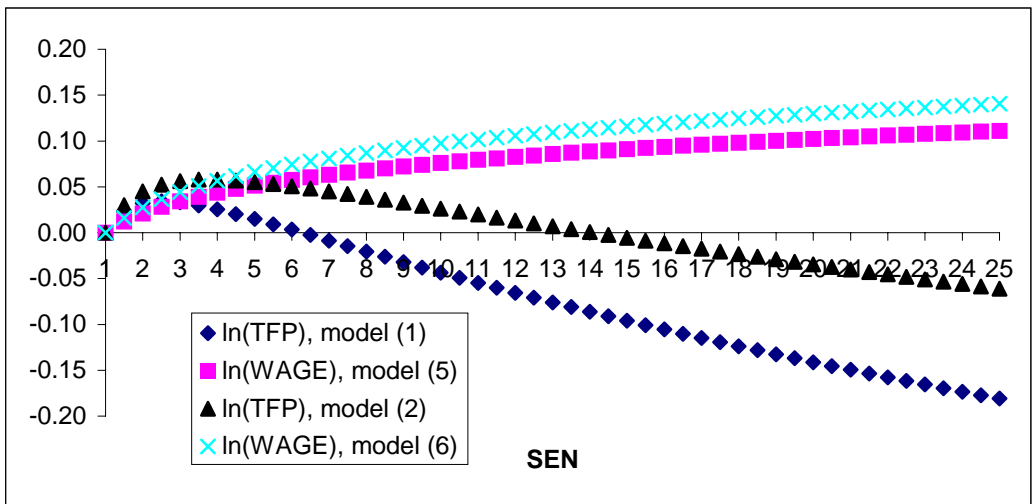


Figure 3: Seniority-productivity and seniority-wage profiles



variables have picked up some of the vintage effect. Older plants have higher average seniority: in 1988 it was 3.3 years in generation A, in contrast to 11.1 years among generation F plants. When the plant generation is controlled, the productivity profile reflects better the true influence of seniority.

Generally, the results indicate that although there are firm-specific skills, they are fairly fast adopted and also fairly quickly eroded. From the point of view of the employees, switching jobs often is not optimal, since with longer seniority within a firm they can benefit from higher pay. From the point of view of the firms, the discrepancy of the wage and productivity profiles makes it profitable to have turnover in the work force. With inflow and outflow of employees, the firms would be able to benefit from the high productivity and low pay of the new workers.

We briefly comment on the other estimates without showing the productivity and wage profiles. The location of the peak in the productivity and wage profiles is sensitive to the estimation method. With random effects, the productivity profiles are fairly similar to the ones obtained with pooling. The inclusion of fixed plant effects shifts the profiles down. The age-productivity profile is fairly flat and reaches its peak already at the age of 32. The seniority-productivity profile is more or less the same as in the pooled estimation. It starts to decline slowly after 2.9 years of seniority. With fixed plant effects, wage is highest at the age of 39 years (the same as with random effects) and at 5.4 years of tenure (60 years with random effects). It is worth noting that the shape of the wage-seniority profile changes considerably when fixed plant effects are controlled. It turns out to be decreasing after a peak, albeit at a slow rate.

4.2. Worker Flows and Productivity Growth

To study the impact of worker flows and changes in worker characteristics on productivity, we estimated models for total factor productivity growth. The growth rate is calculated as two-year differences of $\ln(\text{TFP})$ over the periods 1988-1990, 1990-1992, and 1992-1994. These represent different cyclical situations. A long period of growth in the Finnish economy came to a halt in 1990. The years 1990-1992 were a period when production dropped dramatically. Finally, in the period 1992-1994 output started to increase in the manufacturing

industry. In estimations we pooled these periods, but included period dummies to control the differential productivity growth. The worker characteristics variables for age, seniority and education are included both as lagged levels and their squares, and as changes and their squares. Second, to examine productivity growth effect of worker turnover we included various worker flow variables, which measure the intensity and type of the turnover as averages of the flows in each period. In addition, we also included lagged level of $\ln(\text{TFP})$ in the initial year from which the growth is calculated. The lagged variables are included to control plant differences in the starting level of productivity growth. The differenced variables describe the changes that have happened in the characteristics of the work force. Note that the model is not the difference of the level form model for $\ln(\text{TFP})$, but rather a directly specified dynamic model. We also controlled the plant-specific factors either in levels or in differences. However, in contrast to the OLS models in levels, we dropped geographical location, as it appeared to be insignificant. Also the ‘shadow of the death’ variable is excluded. Any plant specific fixed effects in productivity level that remain constant over time are furthermore controlled by the fact that the dependent variable is differenced. As new variables, we included the interactions of 4-digit industry dummies and period dummies, which allow productivity growth to vary across industries and over time. We also estimated similar models for the growth of $\ln(\text{WAGE})$.

Table 6 shows the estimation results for the productivity growth models (where insignificant squared terms have been dropped). We start by explaining productivity growth only by the past worker characteristics and TFP (and the plant-specific variables, which are not reported). Past productivity has a significant negative coefficient. There is clearly a regression to the mean phenomenon working (e.g., Friedman, 1992). This result should therefore not be interpreted as a convergence of productivity across plants. The estimated contribution of the personal characteristics on productivity growth is more or less consistent with the implications of the level form models. Of the level terms, only the starting levels of age and education have a significant impact, and the contribution of education has a downward trend, as shown by the interactions of education and period dummies in model (6). Among the differenced terms, increases in age have a positive influence, whereas seniority growth decreases TFP growth. However, because of the functional form, small decreases in seniority may not have a positive impact. The estimates of model (9), for example, imply that if

Table 6.
Models for total factor productivity growth.
Dependent variable: $\Delta \ln(\text{TFP})$

	(9)		(10)		(11)		(12)	
	coefficient	t-value	coefficient	t-value	coefficient	t-value	coefficient	t-value
Intercept	-4.029	-1.81	-3.836	-1.73	-3.794	-1.71	-3.856	-1.74
$\ln(\text{TFP})$	-0.417	-51.04	-0.421	-51.67	-0.421	-51.70	-0.422	-51.77
$\ln(\text{AVAGE})$	3.191	2.57	3.054	2.47	3.017	2.44	3.141	2.54
$\ln(\text{AVAGE})^2$	-0.431	-2.49	-0.409	-2.37	-0.402	-2.34	-0.420	-2.43
$d[\ln(\text{AVAGE})]$	0.085	1.37	0.105	1.70	0.108	1.75	0.104	1.68
$\ln(\text{SEN})$	-0.039	-1.59	-0.015	-0.63	-0.009	-0.37	-0.010	-0.38
$\ln(\text{SEN})^2$	0.004	0.57	-0.003	-0.41	-0.003	-0.42	-0.003	-0.45
$d[\ln(\text{SEN})]$	-0.016	-1.17	-0.012	-0.93	-0.006	-0.46	-0.008	-0.57
$[d[\ln(\text{SEN})]]^2$	-0.012	-2.46	-0.009	-1.81	-0.007	-1.45	-0.007	-1.46
EDUY	0.039	6.01	0.040	6.29	0.040	6.20		
EDUY*YEAR 1990							0.050	4.45
EDUY*YEAR 1992							0.046	4.12
EDUY*YEAR 1994							0.026	2.46
$d(\text{EDUY})$	0.001	0.16	0.005	0.55	0.004	0.47		
$d(\text{EDUY})$ *YEAR 1990							0.029	1.87
$d(\text{EDUY})$ *YEAR 1992							-0.013	-0.84
$d(\text{EDUY})$ *YEAR 1994							-0.001	-0.06
WF^2	-0.029	-3.29						
WFIN			0.099	4.29	0.084	3.55		
WFIN*YEAR 1990							0.077	2.19
WFIN*YEAR 1992							0.088	2.13
WFIN*YEAR 1994							0.055	1.17
WFOUT			-0.247	-9.04	-0.283	-9.54		
WFOUT*YEAR 1990							-0.148	-3.02
WFOUT*YEAR 1992							-0.417	-8.11
WFOUT*YEAR 1994							-0.290	-5.39
CHURN					0.091	3.12		
CHURN*YEAR 1990							0.027	0.62
CHURN*YEAR 1992							0.159	2.98
CHURN*YEAR 1994							0.120	2.21
GENEA (ref. group)								
GENEB	0.002	0.13	0.006	0.32	0.004	0.20	0.001	0.04
GENEC	-0.015	-0.98	-0.010	-0.64	-0.012	-0.78	-0.013	-0.85
GENED	-0.025	-1.59	-0.019	-1.21	-0.021	-1.37	-0.022	-1.45
GENEE	-0.010	-0.63	-0.004	-0.27	-0.008	-0.51	-0.009	-0.57
GENEF	-0.010	-0.63	-0.004	-0.24	-0.007	-0.41	-0.009	-0.56
Observations	10403		10403		10403		10403	
R ²	0.278		0.283		0.284		0.285	

Table 7.
Models for earnings growth.
Dependent variable: $\Delta \ln(\text{WAGE})$

	(13)		(14)		(15)		(16)	
	coefficient	t-value	coefficient	t-value	coefficient	t-value	coefficient	t-value
Intercept	-0.671	-0.82	-0.653	-0.79	-0.639	-0.78	-0.668	-0.81
ln(WAGE)	-0.346	-47.29	-0.346	-47.37	-0.346	-47.39	-0.347	-47.40
ln(AVAGE)	0.949	2.06	0.935	2.03	0.926	2.01	0.934	2.03
ln(AVAGE) ²	-0.122	-1.91	-0.120	-1.87	-0.118	-1.85	-0.120	-1.86
d[ln(AVAGE)]	0.078	3.39	0.081	3.55	0.082	3.58	0.084	3.64
ln(SEN)	-0.035	-3.88	-0.032	-3.46	-0.031	-3.32	-0.031	-3.37
ln(SEN) ²	0.010	4.08	0.010	3.78	0.010	3.77	0.010	3.76
d[ln(SEN)]	0.003	0.53	0.004	0.89	0.006	1.10	0.005	0.94
[d[ln(SEN)]] ²	-0.006	-3.46	-0.006	-3.17	-0.006	-2.98	-0.006	-3.06
EDUY	0.031	13.04	0.031	12.95	0.031	12.9		
EDUY*YEAR 1990							0.032	7.66
EDUY*YEAR 1992							0.028	6.92
EDUY*YEAR 1994							0.033	8.51
d(EDUY)	0.016	4.89	0.017	4.97	0.016	4.94		
d(EDUY)*YEAR 1990							0.018	3.08
d(EDUY)*YEAR 1992							0.011	2.02
d(EDUY)*YEAR 1994							0.020	3.52
WF	-0.056	-3.81						
WF ²	0.031	3.35						
WFIN			0.001	0.15	-0.002	-0.21		
WFIN*YEAR 1990							0.008	0.62
WFIN*YEAR 1992							0.009	0.61
WFIN*YEAR 1994							-0.034	-1.92
WFOUT			-0.126	-4.88	-0.142	-5.12		
WFOUT*YEAR 1990							-0.173	-3.43
WFOUT*YEAR 1992							-0.133	-2.89
WFOUT*YEAR 1994							-0.112	-2.31
WFOUT ²			0.134	4.11	0.146	4.36		
WFOUT ² *YEAR 1990							0.181	3.16
WFOUT ² *YEAR 1992							0.133	2.37
WFOUT ² *YEAR 1994							0.117	1.82
CHURN					0.017	1.57		
CHURN*YEAR 1990							0.020	1.12
CHURN*YEAR 1992							0.014	0.72
CHURN*YEAR 1994							0.022	1.06
GENEA (ref. group)								
GENEB	0.004	0.63	0.005	0.77	0.005	0.71	0.004	0.66
GENEC	-0.001	-0.11	0.000	0.03	0.000	-0.03	0.000	-0.04
GENED	-0.011	-1.97	-0.010	-1.82	-0.011	-1.89	-0.011	-1.91
GENEE	-0.003	-0.52	-0.002	-0.37	-0.003	-0.48	-0.003	-0.46
GENEF	0.012	1.95	0.013	2.08	0.012	2.00	0.012	1.95
Observations	10403		10403		10403		10403	
R ²	0.376		0.377		0.377		0.377	

seniority decreases by more than two years, it affects productivity growth positively. Change in education contributes to TFP growth only when it is interacted with the period dummies. Increases in education affected productivity growth positively only in the high demand period 1988-1990.

We study the effects of worker flows by first including the total worker turnover, then decomposing it to inflow and outflow, next adding also churning, and finally using an interaction of the flow components and year dummies. Total worker flow (WF) had the most significant effect in squared form. Increases in worker turnover decrease productivity growth at an increasing rate. Among the components of the turnover, inflow and churning rates have positive coefficients, whereas outflow rate has a negative coefficient. The outflow effect is clearly stronger than the inflow effect. The positive inflow effect is consistent with the general human capital view. Firms can hire new workers who have experience gained in other firms. The negative outflow effect may reflect layoffs that are based on seniority rule. This lowers productivity growth, since the low seniority employees have higher productivity as found in level models above and therefore works against the interest of the employers. Finally, the positive churning effect seems to be in conflict with the impact of the total worker flow. However, the excessive turnover, measured by churning, may be the best indicator for the turnover that happens because of the matching process. Increasing worker flows results in better matches and higher productivity growth. The finding that the churning process affects productivity growth positively may have implications for the interpretations made earlier in the context of level form estimations. We found there that high seniority was connected with a low productivity level. High seniority may be a symptom of a long lasting low churning rate or a high churning rate that has concerned new workers, i.e., a last-in first-out process. Low churning may have led to low productivity growth and eventually to a low productivity level.

We also examined whether the components of the worker flow have different impacts in different phases of the business cycle. The coefficients of the period interaction terms of the flow rates show that the influences of outflow and churning rates are strongest in periods 1990-1992 and 1992-1994, which were years of the deep recession and just after it. This was a time when many firms downsized their work force. The inflow rate, in contrast, had strongest impact in the 1988-1990 period before the recession.

Examination of the wage growth models (Table 7) shows that inflow rate has no impact on wage growth, except for the period from 1992 to 1994 when the effect was negative. Firms that have a high hiring rate may not be able to choose only the young, low wage employees, but may have to hire workers in different age groups. However, after the deepest recession the firms have been able to hire new workers from the pool of unemployed without upward pressures on wages. Outflow rate has a negative impact, but its square a positive impact on wage change. The total impact is close to zero in the relevant range of the outflow variable. Only when the outflow has been very large, is there a positive impact on wage growth. Taking into consideration that outflow had a negative impact on productivity growth, it may be that firms, which need to restructure their work force because of drops in demand, lay off the low wage, high productivity employees who have low seniority. Finally, churning has no impact on wage growth. It seems that productivity gains from better matching are not reflected in wages.

5. Conclusions

We have examined the relationships of worker characteristics (age, education, seniority, and turnover) and plant characteristics to plant performance (productivity and wages), using a matched employer-employee data from Finnish manufacturing. The results support the view that skills are not firm specific. They are accumulated fast, but they also start eroding at low seniority within a plant. On the other hand, wages rise with seniority. This supports incentive wage theories, but may also be a symptom of insider influences on wage formation. The results on the effects of turnover are also consistent with the general human capital view. In addition, they support the hypothesis that turnover leads to better matching and productivity.

An essential part of this research has been the development of the linked employer-employee data. In countries like Finland where register-based information on the whole population of employees and plants is available, it is feasible to create linked data by combining information from different sources and thereby obtain a better picture of the labor market. Even with this kind of data sets, the matching of the data is not an easy task. Different practices in the various statistics, and data needs in the research lead to incompleteness in the linking and to loss of data. For another research topic one would perhaps end up with a different kind of data set. It

seems that it is difficult to form a processed data set, which could be used for multiple purposes. Instead, the best policy might be to form a good infrastructure for making different kinds of linkings from the raw register data. Developments in computer technology have made it possible to process large quantities of micro data to different kinds of data sets at reasonable cost.

References

- Bartel, A. P. and Lichtenberg, F. R., "The Comparative Advantage of Educated Workers in Implementing New Technology." *Review of Economics and Statistics* 69 (1987): 1-11.
- Blakemore, A. E., and Hoffman, D. L., "Seniority Rules and Productivity: An Empirical Test." *Economica* 56 (1989): 359-371.
- Burgess, S., Lane, J., and Stevens, D., "Job Flows, Worker Flows and Churning." CEPR Discussion Paper 1125, 1994.
- Caves, D. W., Christensen, L. R., and Trethway, M. W., "U.S. Trunk Air Carriers, 1972-1977: A Multilateral Comparison of Total Factor Productivity Growth." In *Productivity Measurement in Regulated Industries*, edited by T. G. Cowing and R. E. Stevens. New York: Academic Press, 1981.
- Caves, D. W., Christensen, L. R., and Diewert, W. E., "Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers." *Economic Journal* 92 (1982): 73-86
- Davis, S. J., Haltiwanger, J. C., and Schuh, S., *Job Creation and Destruction*. Cambridge: MIT Press, 1996.
- Friedman, M., "Do Old Fallacies Ever Die?" *Journal of Economic Literature* 30 (1992): 2129-2132.
- Griliches, Z. and Regev, H., "Firm Productivity in Israeli Industry 1979-1988." *Journal of Econometrics* 65 (1995): 175-203.
- Hægeland, T. and Klette, T. J., "Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premia in a Matched Plant-Worker Data Set." Mimeo, 1998.
- Haltiwanger, J., Lane J., Spletzer J., Theeuwes J. and Troske K. (Eds), "The Creation and Analysis of Employer-Employee Matched Data." *Contributions to Economic Analysis Series*, Elsevier Science, 1999 (forthcoming).
- Hamermesh, D. S. and Pfann, G. A., "Turnover and the Dynamics of Labour Demand." *Economica* 63 (1996): 359-367.
- Hellerstein, J. K. and Neumark, D., "Are Earnings Profiles Steeper than Productivity Profiles? Evidence from Israeli Firm-level Data." *Journal of Human Resources* 30 (1995): 89-112.

- Hellerstein, J. K., Neumark, D., and Troske, K. R., "Wages, Productivity, and Worker Characteristics: Evidence from Plant-level Production Functions and Wage Equations." NBER Working Paper 5626, 1996.
- Hutchens, R. M., "Seniority, Wages and Productivity: A Turbulent Decade." *Journal of Economic Perspectives* 3 (1989): 49-64.
- Jovanovic, B., "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (1979): 972-990.
- Kotlikoff, L. J. and Gokhale, J., "Estimating a Firm's Age-Productivity Profile Using the Present Value of Workers' Earnings." *Quarterly Journal of Economics* 107 (1992): 1215-1242.
- Lazear, E., *Personnel Economics*. Cambridge: MIT Press, 1995.
- Maliranta, M., "Plant Productivity in Finnish Manufacturing — Characteristics of High Productivity Plants." Research Institute of the Finnish Economy, Discussion Papers No. 612, 1997.
- Maliranta, M., "Factors of Productivity Performance by Plant Generation — Some Finding from Finnish Manufacturing." Research Institute of the Finnish Economy, Discussion Papers No. 634, 1998.
- McLaughlin, K. J. "A Theory of Quits and Layoffs with Efficient Turnover." *Journal of Political Economy* 99 (1991): 1-29.
- Medoff, J. L. and Abraham, K. G., "Experience, Performance, and Earnings." *Quarterly Journal of Economics* 95 (1980): 703-736.
- Nickell, S., Vainiomäki, J., and Wadhvani, S., "Wages and Product Market Power." *Economica* 61 (1994): 457-73.
- Shearer, B., "Piece-rates, Principal-Agent Models, and Productivity Profiles: Parametric and Semi-parametric Evidence from Payroll Records." *Journal of Human Resources* 31 (1996): 275-303.
- Vainiomäki, J., "Technology and Skill Upgrading: Results from Linked Worker-Plant Data for Finnish Manufacturing." In "The Creation and Analysis of Employer-Employee Matched Data." Ed. by Haltiwanger J. et. al., Elsevier Science, 1999 (forthcoming).