Chapter 40

THE ANALYSIS OF LABOR MARKETS USING MATCHED EMPLOYER–EMPLOYEE DATA

JOHN M. ABOWD*

Cornell University, NBER and CREST

FRANCIS KRAMARZ

CREST-INSEE and CEPR

Contents

* Abowd acknowledges support from the National Science Foundation (SBER-9618111 to the NBER).

Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

Abstract

Matched employer–employee data contain information collected from households and individuals as well as information collected from businesses or establishments. Both administrative and sample survey sources are considered. Both longitudinal and cross-sectional applications are discussed. We review studies from 17 different countries using 38 different systems for creating the linked data. We provide a detailed discussion of the methods used to create the linked datasets, the statistical and economic models used to analyze these data, and a comprehensive set of results from the different countries. We consider compensation structure, wage and employment mobility, and the relation between firm outcomes and worker characteristics in detail. Matched employer–employee data provide the empirical basis for further refinements of the theory of workplace organization, compensation design, mobility and production; however, the arrival of these data has been relatively recent.

JEL codes: J3; J6; C1; C8

1. Introduction

On the empirical side of these questions, the greatest potential for further progress rests in developing more suitable sources of data on the nature of selection and matching between workers and firms. Virtually no matched worker-firm records are available for empirical research, but obviously are crucial for the precise measurement of job and personal attributes required for empirical calculations. Not only will the availability of such data produce sharper estimates of the wage-job attributes equalizing differences function but also will allow more detailed investigations of the sorting and assignment aspects of the theory, which have not received sufficient attention in past work. (Rosen, 1986, p. 688).

> The recent stress on the role of specific as opposed to general human capital and the development of agency theories of the employee–employer relationship may result in the modification of some of the received doctrines but these theories also serve to enrich the scope of the theory by pointing towards interesting and potentially important connections between wages, job mobility and institutional practices. Future progress in this area will hinge crucially on the development of data which links information on the individual characteristics of workers and their households with data on the firms who employ them (Willis, 1986, p. 598).

In the decade since Sherwin Rosen and Robert Willis wrote these words, economists have made enormous strides in finding and using matched employer–employee data. This chapter reviews about 100 studies from more than 15 different countries. Virtually all of these papers have been written in the last 5 years and many are still only available in working paper form. As this chapter was being prepared more than 40 new papers using matched employer–employee data appeared as a part of a conference organized specifically to investigate this issue.[1]

From the many papers that we discuss below, two broad themes emerge. The first is the relative importance of person and firm variables in the determination of compensation. The second is the relative importance of individual mobility in relation to firm-specific employment adjustments. These questions have now been addressed by dozens of researchers. In contrast to many other areas of empirical labor economics, the results we discuss on these questions have largely been estimated from European, and not American, matched employer–employee data, a situation that was foreshadowed by evident advance of the European statistical systems in providing support for the microeconometric analysis of human resource decision making.[2] It is clear from the degree of professional interest in these research efforts that the availability of the type of data Rosen and Willis called for in the original handbook has already produced many important new results.

2. The different types of matched employer–employee datasets

In order to describe the potential that matched employer–employee datasets offer for labor economists, we begin by describing the datasets that exist and some of the basic applications analyzing compensation, mobility, unemployment insurance and other aspects of the labor market. Table 1 presents a complete summary of each of the datasets we describe as well as basic references for further information and applications.

[1] The International Symposium on Linked Employer–Employee Data was held on May 21–22, 1998 in Washington, DC. The preliminary versions of papers from this conference are discussed in this chapter. See Lane et al. (1997a) for an earlier review.

[2] See Abowd and Kramarz (1996b).

Two important dimensions distinguish the matched employer–employee data that we present. First, some are cross-sectional datasets while others are longitudinal. Second, some sampling designs focus on the employee while others use the firm as the primary unit of analysis.[3] When considering issues of representativeness, we show that certain samples, with a longitudinal component, are representative of the target population in the cross-section without being dynamically representative. In particular, certain sampling techniques do not permit entry and exit of individuals from the labor market and/or entry and exit of firms, phenomena which cannot be ignored with matched employer–employee.

Most labor economists are not familiar with the methods used to construct matched employer–employee data. We have, therefore, taken some care to describe the technical details so that potential users of these data can use this chapter to select data sources that are appropriate for the questions they wish to investigate.

2.1. Representative cross-sections of firms with representative data on workers

We begin with the basic design of datasets in which both the sample of firms and the sample of individuals are cross-sectionally representative of the population under study. We start by describing the French program since it follows closely a structure that has been widely adopted across Europe. The Wage Structure Surveys (Enquête sur la Structure des Salaires, ESS), performed by the French National Statistical Institute (INSEE) in 1986 and 1992, were initiated in 1966 by the European Statistical Office (ESO). However, after the 1966, 1972 and 1978 surveys, the ESS was abandoned by the ESO. INSEE decided to resume this survey because of the importance of the information collected at each round and the uniqueness of the statistical design. The ESS collects data on the structure and amount of individual compensation within a sample of establishments from the manufacturing, construction and service industries.

The sampling frame has two stages: at the first stage, production units are sampled; at the second stage, individuals employed at these sampled units are sampled. The target population is all establishments with at least ten employees in general industry. In the construction and in the service industries, the first stage sampling unit is the firm. Furthermore, agriculture, transportation, telecommunication and services supplied directly to individuals are excluded from the scope of the ESS except for insurance, banks and all industries where services are also supplied directly to firms. The universe is constructed from the SIRENE system, a unified database recording all existing establishments and firms in France. The sampling rates are stratified according to the industry, the region, and the size of the unit – from unity for the establishments above 500 employees to 1/48 for establishments between 10 and 20 employees. The sampling frame for the employees at sampled units is based on the employee's year and month of birth. The sample is exhaus-

[3] Hildreth and Pudney (1999) provide an interesting methodological discussion of the statistical properties of many of these methods of creating matched datasets.

Table 1
Comparison of matched employer–employee data sources from different countries

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
Algeria	Algiers Regional Manufacturing Establishments	42 manufacturing enterprises in the Algiers area, 1,000 employees of these firms	1992	Daily wage, employee demographics, education, seniority, other work experience; employer information is limited to detailed industry and employer ID	Data for a developing country	Chennouf et al. (1997)
Australia	Australian Workplace Industrial Relations Survey	Probability sample of workplaces with 20 + employees. Probability sample of employees at that workplace	1995, (earlier survey without the employee questionnaire: 1990)	Workplace qualitative and interval information on wages, profits, productivity, competition. Detailed employment and industrial relations data. Individual data on wages, hours, demographics and unions	Focused on industrial relations issues. Close collaboration with the British Workplace Industrial Relations Survey permits comparative analyses	Callus et al. (1991), Morehead et al. (1997), Alexander and Morehead (1999), Blanchflower and Machin (1996)
Austria	Social Security Files Sample (SSFS, Austria)	Probability sample of firms (1/50)	1975–1991	Simple individual demographic variables, detailed earnings and labor force variables; establishment and firm IDs from the SSFS	Exhaustive within establishments; no longitudinal information on individuals in the sample	Winter-Ebmer and Zweimüller (1997)

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
Belgium	Social Security Administrative Records	Universe of private firms	1978–1985	Individual earnings histories, demographic variables, and broad occupation, firm-level accounting data for larger firms	Universe of private employees permits detailed mobility analyses	Leonard and Van Audenrode (1996, 1997), Leonard et al. (1999)
Canada	Workplace and Employee Survey (WES)	Clustered probability sample of establishments	1996 (pilot), 1997	Detailed establishment information from the human resource manager; detailed demographic, labor force, and earnings variables; establishment IDs from Statistics Canada	Designed to collect longitudinal information on establishments (including birth, death, and mergers); no longitudinal information on workers	Picot and Wannell (1996)
Canada	T-4 Supplementary Tax File	1% simple random sample of individuals ever filing a tax return	1975–1993, ongoing	Individual age, sex, taxable earnings, taxpayer ID and employer ID	Very long employment histories	Morissette and Berube (1996), Baker and Solon (1997)
Denmark	Integrated Database for Labor Market Research (IDA)	Universe of the Danish population based on the person ID used in Danish government registers	1970, 1980–1994, ongoing	Detailed demographic and labor force variables; employer reported earnings; employer IDs from the Danish establishment register	Complete census; individuals who are unemployed or not in the labor force are included	Albaek-Sorensen (1999)

Finland	Employment Register matched with manufacturing establishments in Register of Establishments	Census of employed persons and census of manufacturing establishments and plants with 5 + employees	1988–1994, ongoing but 1995 changes in industrial register present comparability problems	Earnings, other income, education, demographics, employment history for workers; Output, value added, inputs, price indices, some capital measures	Because both sides of the match are based on registers, the coverage is very good when supplemental data from other sources are added	Laaksonen et al. (1998)
France	Déclaration Annuelle de Données Sociales (DADS), formerly DAS; Bénéfices Industriels et Commerciaux (BIC); Bénéfices Réels (BRN), Echantillon Démographique Permanent (EDP); Enquête Structure des Emplois (ESE)	1/25th of private and semi-public workforce (born October, even years); supplemental individual data for 1/10th from the EDP; employer information from BIC (larger enterprises), BRN (smaller enterprises) and ESE (establishments)	1976–1995, ongoing	Individual data: earnings, days worked, payroll taxes, occupation, industry, demographic data, education, detailed individual data from EDP; longitudinal firm data from sources keyed to Siren/Siret: production, value added, operating income, assets, employment, imports, exports, prices	Based on a set of databases that permits any data coded by firm or person identifier to be added. Longitudinal for firms and individuals	Abowd et al. (1999a), Kramarz and Roux (1998), Margolis (1999)
France	Déclaration Mensuelle de Mouvements de Main d'Oeuvre (DMMO), matched to the sources cited above	All establishments with 50 + employees complete the monthly questionnaire,	1987–1990, ongoing	Individual data: demographics, type of contract, type of entry, skill-level for all entries, seniority at exit, type of exit,	Permits monthly analysis of employment flows as well as job creation and destruction	Abowd et al. (1999b)

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
		which includes the reports on each movement for the individual employees		and skill-level for all exits; establishment data: employment, skill distribution, profit/loss		
France	Enquête Emploi (EE) and Enquête sur la Technique et l'Organisation du Travail auprès des Travailleurs Occupés (TOTTO)	Clustered probability sample of domiciles 1/300, longitudinal with 3 years in sample. Employer identifiers since 1990	1990–1996, ongoing1987 and 1993 new technologies supplement	Full complement of household-based labor force variables; periodic topical supplements; longitudinal firm data from sources keyed to Siren/Siret; new technologies supplements have a full complement of computer and computer-assisted production questions	Overlapping samples with 3-year rotation groups permit dynamic analyses; employer IDs for establishments (Siret) permit linking to establishment or firm data	Entorf and Kramarz (1997, 1999), Entorf et al. (1999), Kramarz (1997)
France	Enquête Structure des Salaires (ESS)	Probability sample of establishments with 20 or more employees; Probability sample of employees at the establishment	1967, 1978, 1986, 1992	Very detailed job and compensation descriptions, earnings; establishment level work place organization; longitudinal firm data from sources keyed to Siren/Siret	Large representative samples of employees within establishment; employer IDs for establishments (Siret)	Rotbart (1991), Kramarz et al. (1996)

France	Enquête Formation Qualification Profession (FQP)	Clustered probability sample of domiciles 1/1,000	1993	Full complement of labor force variables; detailed education and training, apprenticeships; retrospective from 1988; longitudinal firm data from sources keyed to Siren/Siret	Employer IDs for establishments (Siret) permits addition of other information	Goux and Maurin (1997)
Germany	Beschaftigungs-stichprobe (BS) matched with the Leis-tungsempfanger-datei (LD)	Probability sample of the Historikdatei (HD), 1/100th sample, of the Bundesanstalt für Arbeit (BfA)	1975–1990	Simple individual demographic variables (sex, education, nationality), gross earnings; benefits for the unemployed; employer IDs from the HD	Longitudinal information on individuals even when unemployed or on training programs; plant-level statistics from the HD	Bender et al. (1996), Dustmann and Meghir (1997)
Germany	Gehalts- und Lohnstruktur-erhebung (GLS) matched to social insurance registry	Multistage probability sample of establishments, probability sample of employees. Lower Saxony.	1990, 1995	Employer supplied detailed data on the structure of compensation and conditions of employment in October. Demographic, education, occupation for employee	Two representative cross-sections, 5 years apart with very similar structure to ESS in France	Stephan (1998)

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
Italy	Ricerche e Progetti (R&P)	Universe of private firms (industrial and service) and records of self-employed	1985–1991, ongoing	Social security earnings, wage supplements, months, weeks or days paid, occupation, employment contract type	Longitudinal information on workers and firms, transitions to self-employment can be included. Availability of universe permits different sampling schemes	Contini et al. (undated)
Japan	Establishment Census (EC) matched with Basic Survey on Wage Structure (BSWS), Census of Manufactures (CM) and Census of Commerce (CC)	EC is a census of establishments with 5 employees or more (10 + in public establishments). BSWS, see below, CC every 3 years and CM annual	1991–1994, ongoing, not all years for all sources	Wages, bonuses, seniority, occupation, employee demographics and education from BSWS, detailed business data from CM and parts of CC, varies by industry	Some longitudinal information on both workers and firms	Hayami and Abe (1998)
Japan	Basic Survey on Wage Structure (BSWS)	Probability sample of all establishments with at least 5 employees or government sector if covered by the National Enterprise Labor Relations Law or by the Local Public Labor Relations Law and at least 10 employees	1982–1994, ongoing	Simple information on the establishment; simple individual demographic and labor force variables, detailed earnings	Large sample within establishments	Abe and Sofer (1996)

Nether-lands	Wage Survey, Production Survey, RD Survey and Survey of Manufacturing Technology	Probability sample of firms, simple random sample of employees at each firm	1979, 1985, 1989, ongoing	Detailed individual and job characteristics, gross weekly earnings, hours worked per week, firm data on inputs, outputs, value added, profits, R&D activities (larger firms), computer technologies used	Very comprehensive set of repeated cross-sections with detail on both the individual and the firm	Boon (1996)
Nether-lands	Ministry of Social Affairs and Employment (AVO)	Probability sample of firms, probability sample of employees of those firms	1993, 1994	Detailed salary information, separation reasons, demographic data, seniority; firm level aggregates of these variables and employment	Two observations (successive Octobers) for each employee	Hassink (1999)
Norway	Employer–employee Register and Education Register	Universe of the Norwegian population based on the person ID used in Norwegian government registers	1986–1994, ongoing	Detailed demographic and labor force variables; employer reported earnings; employer IDs from the Norwegian Employer–employee establishment register	Not a sample; individuals who are unemployed or not in the labor force are included	Salvanes and Forre (1997)

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
Portugal	Social Security Files Sample (SSFS, Portugal)	Clustered probability sample, 1/5 of all firms	1983–1992, ongoing	Simple individual demographic variables, detailed earnings and labor force variables; establishment and firm IDs from the SSFS	Exhaustive within establishments; no longitudinal information on individuals in the sample	Ministério do Emprego e da Segurança Social (1993), Cardoso (1997)
Sweden	Register of Income Verifications, Register of Jobs and Other Activities, Register of Employment, Register of Enterprises and Register of Establishments	All registers are censuses of the relevant population.	ongoing, dates depend upon specific application	Earnings and income, job characteristics, demographics, enterprise and establishment characteristics	Surveys with more detailed information can be linked to any of the component registers	Tegsjö and Andersson (1998)
Sweden	Labor Force Survey (Arbetskraft-sundersök-ningarna, AKU) matched with the Registers of Employment, Enterprises and Establishments	Probability sample of households, census of business establishments with at least one employee	1987–1993, ongoing	Detailed employment data from the AKU, other data as described above	Makes use of the register system described above	DiPrete et al. (1998)

UK	Panel Study of Manufacturing Establishments (PSME)	Probability sample of establishments, most recently hired employee and one randomly sampled production employee	1994, 1995	Detailed employer information on the personnel, financial, investment policy; detailed demographic and earnings variables on individual	Longitudinal in the establishment, employees are not followed due to legal restrictions in the UK	Hildreth and Tremlett (1994, 1995)
UK	New Earnings Survey (NES), Joint Unemployment and Vacancies Operating System (JUVOS), Annual Census of Production (ACOP)	1/100 sample of employees enrolled in the Pay As You Earn (PAYE) tax system linked to administrative universe of unemployment system and universe of firms with 100 or more employees. Probability sample of smaller firms	1994, 1995	Employee information on weekly earnings, demographic data, unemployment spells. Firm information on inputs, production, profitability	The links can be used along several dimensions to follow individuals in and out of employment and/or to follow individuals from firm to firm	Hildreth and Pudney (1999)
UK	Workplace Employee Relations Survey	Probability sample of workplaces, 25 employees per workplace	1997, earlier surveys without the employee segment: 1980, 1984, 1990, called Workplace Industrial Relations Survey	Workplace qualitative and interval information on wages, profits, productivity, competition. Detailed employment and industrial relations data	Focused on industrial relations issues. Close collaboration with the Australian Workplace Industrial Relations Survey permits comparative analyses	Cully (1998), Blanchflower and Machin (1996)
UK	British Household Panel Study	Probability sample of households, details of employer data not available	in progress	Detailed data on individuals in the household, including labor market earnings, education and demographics.	Design permits analysis of the effects of employer variables on household outcomes	Hildreth, private communication

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
United States	Worker-Employer Characteristics Database (WECD) and New Worker-Employer Characteristics Database' (NWECD)	Manufacturing establishments from the Longitudinal Research Database (LRD), a probability sample; matched to 1990 Census of Population long form responses; NWECD establishments from the Standard Statistical Establishment List (manufacturing and non-manufacturing)	1990 match	WECD: longitudinal data on establishments income, balance sheet, investments; NWECD: employment and sales; Both datasets: full complement of labor force variables and household variables from the Census of Population	Very large samples within establishments	WECD, Troske (1998); NWECD, Bayard et al. (1998)
United States	State Unemployment Insurance Systems	Simple random samples from state unemployment insurance records Mexico, Pennsylvania, South Carolina, and Washington	various years, matched on the individual ID from Georgia, Idaho, Louisiana, Maryland, Missouri, New	Earnings and employment data required to calculate UI benefits; employer UI-related data (tax rates, taxable compensation); employer IDs from the federal employer ID system	Some states have labor force variables (sex, education, etc.) for a subsample who received UI benefits, other states have demographic data for representative samples, others have no demographic data	Jacobson et al. (1993), Anderson and Meyer (1994)

United States	Maryland Unemployment Wage Records, Current Population Survey, Standard Statistical Establishment List (SSEL)	Universe of Maryland UI records matched to CPS (probability sample); Employers matched using the SSEL	1985–1997	Individual and household data from the Current Population Survey including demographic, earnings, education; establishment data from the SSEL (employment, payroll, sales, industry)	Limited to the State of Maryland. Longitudinal information on UI earnings matched to longitudinal information on the SSEL	Lane et al. (1999)
United States	National Longitudinal Survey of Youth '79 (NLSY-79)	Clustered probability sample of individuals aged 14–21 on January 1, 1979	1986–1994	All variables from the public-use NLSY files; employer IDs from private lists, Compustat and CRSP	Unique IDs for all available employers on the NLS; employer data for publicly-held firms, some data for governments	Abowd and Finer (1998)
United States	Survey of Consumers (University of Michigan Survey Center)	Clustered probability sample	September 1991-March 1992	All variables in the Survey of Consumers; employer IDs from Dun and Bradstreet	Some employer data for 700 matches	Brown and Medoff (1996)
United States	Employment Opportunity Pilot Project (EOPP) and Multi-City Study of Urban Inequality (MCSUI)	Probability sample in metropolitan areas, data on a representative employee and the most recently hired employee	1982, 1993 (repeated the design of 1982 survey)	Detailed employer information from the human resource manager; labor force variables on individual	Heavy focus on training and training related variable	Bishop et al. (1983), Holzer and Reaser (1996).

Table 1 (*continued*)

Country	Name	Sampling plan	Dates	Main variables	Unique features	References
United States	Continuous Work History Sample (CWHS) and Longitudinal Employer–employee Database (LEED)	1/100 sample of Social Security earnings reports	1957–1972, other files continue	Social Security earnings, total employment in the firm, basic demographic variables, some schooling, hours and weeks worked information. Employer and employee identifiers	Most extensive US sample. Internal Social Security files are produced on an ongoing basis	Smith (1989), Topel and Ward (1992)
United States	Survey of Employer-Provided Training (SEPT95)	Probability sample of private establishments with 50 + employees, two employees per establishment	1995	Detailed training information at the establishment level, earnings, seniority, training and demographics for employees	Design permits analysis of both establishments and individuals for population training models	Bureau of Labor Statistics (1996)
United States	Longitudinal Research Database (LRD) and the National Labor Relations Board (NLRB) files	Probability sample of manufacturing establishments (LRD) matched with the annual NLRB election data	1977–1989	Detailed establishment data on inputs, production, costs, production and non-production employment. NLRB election data describe the proposed bargaining unit and election results	Match of union representation vote data to detailed history of the establishment permits studies of effects of new unionization on factor use and production. Establishment data available from 1963	LRD, McGuckin and Pascoe (1988); match to NLRB, Lalonde et al. (1996)

United States	White Collar Pay Survey (WCP) supplement	Probability sample of establishments, simple random sample of employees in certain white collar occupations	1989 (service), 1990 (goods)	Detailed earnings and components of compensation from employer survey, starting pay, demographics, education, seniority	Sample focuses exclusively on white collar occupations	Bronars and Famulari (1997)

tive in small units and the sampling rate is 1/24 in the largest establishments (above 5000 employees).

In the 1986 version of the survey, annual and October compensation are available for each sampled employee. The October compensation for each employee includes all employee and employer-paid benefits but excludes non-wage benefits. It can be decomposed into total wage, overtime compensation and October-specific bonuses. The total annual compensation includes all benefits and bonuses, even those not paid on a monthly basis. Finally, information on the method of pay is given (time versus piece rates, for instance). In 1992, total annual compensation, decomposed as described above, is available but the October compensation is not decomposed.

In both versions of the ESS, occupation, firm-specific seniority, age, country of origin, hour schedule (number of hours and shifts), days of absence are measured for the employee. In addition to this individual-level information, the surveyed unit gives the following information: total employment, existence of shifts and night work, existence of a firm-level agreement, of a branch-level agreement. Since some questions in the 1986 and 1992 versions of the ESS were not formulated identically, the two surveys are not always comparable.

The basic research data files for the ESS contain 16,239 establishments with 678,798 interviewed employees in 1986 and 15,858 establishments with 148,976 interviewed employees in 1992. More detailed technical information on the 1986 version of the ESS is available in Rotbart (1991). The technical report on the 1992 version of the ESS is not yet available.

Salary structure surveys with the same structure as the ESS exists in most EC countries, for instance in Germany (see Stephan, 1998) and the United Kingdom. Unfortunately, the statistical offices in charge of collecting and storing these data have been generally reluctant to let researchers access them. In France, however, the policy for non-INSEE researchers has been more generous (see Arai et al., 1997, among others). Statistics Canada is now in the process of building such a dataset called the Workplace and Employee Survey. Data collection should be completed by the end of 1997. A pilot, designed to be one-fifth of the production version, was conducted in 1996 with approximately 1000 establishments and 6000 workers (see Picot and Wannell, 1996). In the United Kingdom, the Office for National Statistics now allows contracted researchers access to these confidential data.

Salary structure data also exist in Japan, based on a annual survey called the Basic Survey on Wage Structure (see Abe and Sofer, 1996). The universe of establishments sampled every year includes all establishments of the private sector with at least 5 employees and the public sector establishments if covered by the National Enterprise Labor Relations Law or by the Local Public Labor Relations Law and at least 10 employees. Each year, approximately 70,000 establishments with 1.4 million workers are sampled. The survey is conducted during the month of July, with information recorded about the month of June (apart from annual bonuses, which come from the previous fiscal year). General information about the establishment is collected: industry, size, product, enter-

prise to which the establishment belongs, entry wage for the youngest hires. Information on individual workers includes: sex, age, education, type of contract, number of days and hours worked, experience, job position, June earnings (before taxes), and annual bonuses.

2.2. Representative cross-sections of firms with non-representative data on workers

In this type of data, a sample is designed to be representative of the cross-section of firms (or other business units) in a given year and data on some workers are collected. Some of the surveys have longitudinal or panel components but the sampling frame was, nevertheless, constructed using a universe that was fixed at a particular date. Hence, they are not dynamically representative even though they are representative over time of the business units and employees at risk to be sampled at that date.

The best example is the European Commission-sponsored research data collected in the United Kingdom, called the Panel Study of Manufacturing Establishments (PSME). The description is based on Hildreth and Tremlett (1994, 1995). The stage of the sample is based upon an establishment universe called a business location and defined as the activities of a single employer at a single address. The sample of business locations is based on British Telecom's (BT) business line records. If an establishment has a business telephone line, it is included in the population at risk to be sampled. As seems natural given the origin of the sample, the BT sample provided a contact phone number as well as an establishment name, and address. This allowed the interview to be conducted over the telephone. BT also reported the industry classification as well as size of the establishment. The sample was restricted to manufacturing establishments only (Divisions 1–4 of the 1980 Standard Industrial Classification, SIC code). Using this information, the frame was stratified according to area, size, and industry. Details of the sampling scheme can be found in Hildreth and Tremlett (1994). The initial sample comprised 881 establishments of which nearly a quarter (23%) was found to be out-of-scope for the survey.

From the original 881 establishments, 682 were in the scope for interview. Interviews were conducted between February and April 1994 using Computer Assisted Telephone Interviewing (CATI). The average interview lasted 45 min and was conducted by interviewers at the Social and Community Planning Research (SCPR) telephone interviewing unit. The questionnaire covering a range of areas of the establishment operation: ownership and control, markets and products, innovation and investment, employment and human resources, financial performance, and, finally, detailed information on two workers – the most recent hire and a randomly selected employee. There were several respondents at each establishment – the Chief Executive or Senior Manager, the Personnel or Human Resources Manager, and the Chief Accountant or Financial Director. Of the original sample of 682 establishments within scope for the survey, 430 completed interviews, of which 398 have consistent information on the establishment. Not all establishments gave complete worker information on both of the employees. The number of observations for the worker selected at random from the list of production line employees is 339 while the

number of observations for employees selected as the most recent hire is 346. Only 312 establishments have complete information for both workers.

The Employment Opportunity Pilot Projects (EOPP) employer survey for the US is based on a very similar sampling scheme as the PSME (the description is based on Bishop, 1994). The survey covers a sample of 3412 employers. It was sponsored by the National Institute on Education (NIE) and the National Center for Research in Vocational Education (NCRVE). Interviews were conducted between February and June 1982. This survey was a two-wave longitudinal survey of employers from selected geographic areas across the country. The ES-202 list of companies paying unemployment insurance taxes provided the sampling frame for the survey. Establishments in industries with a relatively high proportion of low-wage workers have been oversampled. The survey was conducted over the phone and obtained a response rate of 75%.

The second wave tried to interview all of the respondents from the first wave survey. Approximately 70% of the original respondents completed surveys for the second wave. Seventy percent of the establishments have fewer than 50 employees and 12% have more than 200 employees. In large organizations, the main respondent was most often the personnel officer in charge of hiring. Employers who received the full questionnaire were asked to select the "last new employee your company hired prior to August 1981 regardless of whether that person is still employed by your company." A total 818 employers could not provide information for a recent new hire. The employers who provided information on one new hire were also asked to provide data on a second new hire in the same job but with a different amount of vocational education. Of the 2594 employers who provided data on one new hire, 1511 had not hired anyone else in that job in the last 2 years, and 424 had not hired anyone with a different amount of vocational training for that position in the last 2 years. As a result, data are available for 659 pairs of individuals who have the same job at the same establishment. Missing data on specific questions used in the model reduce the sample to about 480. The questionnaire focused primarily on training activities on the job. See Bishop (1994) for more information on the questionnaire.

2.3. Representative cross-sections of workers matched with longitudinal data on firms

A representative cross-section of workers is often matched with longitudinal data on the employing firms. The data source for the individual workers and the source for the employing firms are not generally coordinated ex-ante, as was the case for the data described in Sections 2.1 and 2.2. In the United States, the Longitudinal Research Database (LRD) – a panel of manufacturing establishments (see McGuckin and Pascoe, 1988) – has been linked by Troske (1998) with the 1990 Decennial Census of Population. In France, the supplement to the 1987 Labor Force Survey on New Technologies contains the firm identifier and the establishment identifier number for most employed workers, which permits researchers to match with the Echantillon d'Entreprises (based on the BIC), a dynamically representative sample of French firms or the Enquête Structure des Emplois (ESE) (see Entorf and Kramarz, 1997, 1999).

We first describe the Worker-Establishment Characteristics Database (WECD) based on Troske (1998). The data for workers were extracted from the 1990 Sample Detail File (SDF), which consists of all households questionnaires from the 1990 Decennial Census of Population long form. The data for establishments come from the 1990 Standard Statistical Establishment List (SSEL), a register of all establishments active in the US in 1990. From the SSEL, a 4-digit SIC code giving the establishment primary industry and a geographic code giving location were extracted. Only manufacturing establishments were retained. Equivalent industry and location information was obtained for the individuals in the SDF through individual responses coded by the Census Bureau (using Census industry codes, however). All workers employed in manufacturing in 1990 who responded to the long form are in the sample file. The number of individual observations is 4.5 million. These individuals were at risk to be matched to an employing establishment.

The matching procedure has four steps. First, Troske standardized the geography and industry definitions across the two data sources. Second, he eliminated all establishments that are not unique in each location-industry cell. Third, he assigned a unique establishment identifier to all workers located in the same location-industry cell. Fourth, he eliminated all matches based on imputed industry or location data in the Census of Population.

To understand the first step, one must know that each Census of Population geographic code consists of a region code, a state code, and a county code. Each county code is further divided into incorporated and unincorporated areas. Each incorporated area gets a unique place code. Finally, in highly populated places, a further subdivision, blocks, is added. Since the 1990 SSEL only contains place codes, which are not the same as these Census of Population location codes, Troske used the Census Bureau's Address Reference List (ARF) to assign blocks to the 1987 SSEL which was then matched to the 1990 SSEL. In addition, the Standard Industrial Codes (4-digits) were recoded into the Census Industry Codes (3-digits).

The second step forces Troske to use only establishments that meet one of these three criteria:

- Establishments that are unique in an industry-block cell;
- Establishments in the same industry-place cell with missing block codes when all other establishments in the same industry-block cell have valid block codes;
- Establishments unique in the industry-place cell.

In the third step, Troske matched individuals using industry-block codes (first group above). Next, all remaining workers were matched to establishments with identical industry-place codes (next two groups). All matches in which the industry or the geographic code were imputed were deleted. Finally, all matches for which the total number of matched workers exceeded the establishment employment were deleted. The resulting dataset contains 200,207 workers employed in 16,197 manufacturing establishments. Troske (1998) describes various tests of the quality of the WECD. On average, 16% of an establishment's work force is included in the WECD. This match rate is correct given the sampling frame of the SDF.

Different measures of average earnings per employee result from aggregating individual data to the establishment level and calculating per employee averages directly from the LRD. These earnings measures are positively and significantly correlated. An analysis of the structure of the establishments shows that large plants and plants located in urban areas are over-represented in the WECD. This induces overrepresentation of white, male, and educated workers in comparison to the original SDF data.[4]

The techniques used to create the WECD have been extended by Bayard et al. (1999) to create a cross-sectional matched employer–employee dataset that includes both manufacturing and non-manufacturing establishments. The new dataset, which is called the New Worker-Employer Characteristic Dataset (NWECD), has not yet been as widely used in empirical analyses; however, the addition of non-manufacturing establishments greatly extends the potential of these data. The authors obtained their individual and household data from the same 1990 decennial Census of Population SDF file described above. The information on establishments was taken from the Bureau of the Census Standard Statistical Establishment List (SSEL), which provides the sampling frame for Census Bureau surveys of establishments in virtually all regions and industries. The main difference between the NWECD and the original WECD is the breadth of establishment data available. The SSEL has only limited employer information (employment, payroll, sales and industry), whereas the WECD, which permits access to all of the LRD data but for manufacturing only, contains detailed longitudinal information on establishments that appear in the LRD.

Entorf and Kramarz (1997, 1999) have constructed similar data for France by matching four different INSEE sources. The basic sources are the "Enquête Emploi," 1985–1987, a single rotation group from the French Labor Force Survey, and the"Enquête sur la Technique et l'Organisation du Travail auprès des Travailleurs Occupés" (TOTTO) from 1987, a supplement to the labor force survey, which asked questions about the diffusion of new technologies and the organization of the work place. In addition to the usual questions on labor force surveys (earnings, wage rates, tenure, age, education, etc.) the supplement contains a rich source of information on the use (e.g., intensity and experience) of microcomputers, terminals, text processing, robots and other well specified groups of "New Technologies." Likewise, questions concerning the hierarchy of labor and working-time schedules help in drawing more detailed conclusions concerning the impact of new technologies than would be possible by the analysis of usual labor force surveys.

Additional information on employing enterprises (a business unit in American terminology, not an establishment) for individuals in the EE and TOTTO was added using the standardized Siren enterprise identification number, which was coded for the first time in an INSEE survey for this particular year (1987) and survey (TOTTO). This feature of the French INSEE classification system enables the researcher to employ information from corresponding firm-level surveys (such as profits and share of sales going to exports, for instance). Entorf and Kramarz used information from the 1985–1987 period. No informa-

[4] Hildreth and Pudney (1998) consider likelihood corrections for these kinds of sampling problems.

tion on the employing firm in years 1985 and 1986 is available for workers who changed firms between these dates and 1987. Entorf and Kramarz use two additional sources: the "Bénéfices Industriel et Commerciaux" (BIC) and the "Enquête sur la Structure de l'Emploi" (ESE). From the first source, which collects annual information on balance sheets and employment, they use the measure of the annual average full-time employment, the total capital in the firm as the sum of debt and owners' equity (this sum is equal to total assets in the French accounting system), the annual operating income, and, finally, the export ratio computed as the ratio of the firm's exports to its sales. From the second source, which collects information on the employment structure, they compute a proportion of engineers, technicians and managers in the work force and a proportion of skilled workers in the work force, both expressed as ratios using the employment measure described above.

The survey "Enquête sur la Technique et l'Organisation du Travail auprès des Travailleurs Occupés" (TOTTO) was performed in March 1987. It covers a total of about 20 million individuals in civilian employment. The probability of being selected is 1/1000; thus the survey contains about 20,000 workers. Questions concerning the organization of the workplace were asked to wage-earners and salaried employees only, questions concerning the use of "New Technologies" were asked to all members (including civil-servants) of the civilian work force (according to the definition of the OECD). The sample used for cross-section estimation consists of 15,946 wage-earners and salaried employees, based on TOTTO. The longitudinal sample where individual workers are followed at least 2 years and at most 3 years has 35,567 observations. When merged with firm-level information, the cross-section dataset includes 3446 individuals and the longitudinal dataset reduces to 7965 observations. The firm-level data are based on a panel of firms covering the years 1978–1987. The firm-level information comes from an exhaustive sample for large firms (more than 500 employees) and an INSEE probability sample plan for smaller firms. The sample plan provides a weighting variable which is used in subsequent estimation in order to estimate the variance–covariance matrix that is representative of the population of individuals (such that the bias arising from the higher probability of large firms to be in the sample can be offset).

Starting in 1990, most individual-level surveys performed at INSEE contain the same firm-level identification number, the Siren; mentioned above. This means that matched worker-firm data are available on a regular basis. For instance, the "Formation, Qualification, Profession" 1993 survey on education and continuous training has the employing firm for more than 90% of the employed workers in the dataset (see Goux and Maurin, 1997). We will also examine later longitudinal uses of the French Labor Force Survey, a 3-year rotating panel for which the Siren is available in every year a worker is employed.

Other interesting examples of representative cross-section data for the employee matched to longitudinal data for the firm include the Portuguese file used by Cardoso (1997) and the British file created by Hildreth and Pudney (1997). Cardoso used Social Security files (see the discussion in Section 2.4), to construct a random sample of 20% of the firms, stratified by economic activity. For each such firm, information on workers

employed in a given year is available–sex, age, skill, occupation, schooling, tenure, earnings split into different components (base pay, bonuses, tenure-related pay, overtime pay), and hours. The sample of firms is designed to be dynamically representative of the Portuguese economy (starting in 1982). Hence, firms were initially sampled in 1983, the first year available. Then, all sampled firms were followed until their death. All new firms are at risk of being sampled at most once. Sampling frames like the one used to construct the Portuguese data make it difficult to follow the workers from firm to firm since the plan does not ensure the presence of the same workers from year to year. The Portuguese data have been used primarily to assess firm-specific wage inequality at different dates (see Cardoso, 1997). Hildreth and Pudney (1999) use the New Earnings Survey (NES), the Joint Unemployment and Vacancies Operating System (JUVOS), and the Annual Census of Production (ACOP), all for the United Kingdom. The different data sources permit dynamic links but the ACOP rules for sampling establishments changed between 1994 and 1995, creating difficulties for longitudinal analyses.

2.4. Representative matched worker-firm panels (administrative origin)

Many matched employer–employee datasets are based on administrative files. In this section we discuss some leading examples.

Every state in the US, except New York, maintains very complete information on quarterly employment and earnings so that the State Employment Security Agency (or State Unemployment Insurance Agency, depending on the state) can manage the state unemployment benefits program. The exact details of these programs may vary from state to state. However, such UI wage records cover almost all of the employment (at least 90% of the work force but more in some states). Self-employed individuals are never covered. Other categories, such as federal and military personnel, employees of the US postal service, railroad employees, employees of religious and philanthropic organizations, those who receive only commissions, and some agricultural employees may not be covered in some states (Maryland is an example; see Burgess et al. 1999).

Starting with the base UI earnings files, the different states have constructed random samples of the eligible work force. The sampling rate varies by state: 5% in Pennsylvania, 10% in Washington State to 100% in Maryland. Eight states participated in an early attempt to coordinate such data, the Continuous Wage and Benefit History Project (Georgia, Idaho, Louisiana, Missouri, New Mexico, Pennsylvania, South Carolina, and Washington, see Anderson and Meyer, 1994, who use these datasets for the period 1978–1984). Apart from the wage amount received by workers (total wages, including tips, commissions, and bonuses, up to a ceiling of $100,000 that may depend on the state), each quarterly record includes a person identifier, a firm identifier–the federal employer identification number (FEIN), and some other firm characteristics such as the industry (4-digit SIC), average monthly employment, total wages, taxable wages and tax rate as computed by the State Agency.

Unemployment benefit claim records for any worker who filed for UI are also available

in certain states (for an example, see Anderson and Meyer, 1997. These datasets contain, for each claim filed, the worker's personal identifier, the date the claim was filed, the first pay date and the exhaustion date, the total amount of benefits paid, the reason for work separation, as well as personal characteristics (age, sex, race, schooling). In addition, it is possible in some states, for some firms (mostly publicly traded firms) to merge with financial data using the FEIN. Even though this is possible for only a small fraction of firms – the largest, in general – more than half of the workers are employed in such companies. Hence, financial data, balance-sheet information may be available for a large share of the records at hand.

Lane et al. (1999) recently completed a pilot project in which the information from the State of Maryland UI wage records was matched to data from the Current Population Survey (also called the monthly household survey in the United States) and the Standard Statistical Establishment List (SSEL, Bureau of the Census). The use of data from the Current Population Survey provides demographic, educational and other individual and household data to complement the earnings history in the UI wage records. The SSEL provides longitudinal, but limited, data on the employing establishments.

Topel and Ward (1992) use the Social Security earnings reports made by employers to the Social Security administration, a Federal version of data similar to the state UI reports. The Longitudinal Employer–Employee Data (LEED) contain quarterly information for over one million individuals for the period 1957–1972. In addition to employee and employer identifiers, available individual characteristics are the age, the race, and the sex. Earnings are reported on a quarterly basis (see Smith, 1989). According to Topel and Ward (1992), top-coding problems, common with US Social Security-based data, are minimized because of the quarterly reporting. Jacobson et al. (1993) use both types of data – UI and Social Security – or a subsample of the Pennsylvanian displaced workers that they analyze.

The administrative source from which similar French data files were constructed are derived from records received by the Tax and Social Security Authorities in order to compute the wage-related taxes, to cross-check with employees' own income tax reports, to compute employers' contributions to Social Security, and to manage employees' individual accounts for entitlements to pensions and health benefits. INSEE also receives these files, called the Déclaration Annuelle de Données Sociales (DADS). As in the US, the coverage is very broad, every employer except those employing only domestic staff must report. INSEE files exclude agricultural workers as well as government employees from the statistical operations (all of whom have special social security systems). Information on the establishment consists of: Siren (firm) and Siret (establishment) identification numbers, address, 4-digit industry code (APE), work force (December 31), and total wage bill. For each individual employee, INSEE receives the name, national identity number, occupation, number of hours (since 1993), start and end dates of the employment period, employment status (full-time, part-time, home work, irregular), total compensation (before as well as after deduction of social security contributions), total benefits in kind, and total allowances for business expenses. Because of the work load that the data entry

imposes, not all of this information is accessible at all dates. For instance, the employer identifier is only available starting in 1976. The start and end dates of the employment period are not on the research files that, for instance, Abowd et al. (1999a) have used (they only have its length). Starting in 1964, only those workers born in October of an even year were kept in the research files, resulting in a 1/25 sample of the private and semi-public sector employees. The file used in Abowd et al. (1999) includes more than 1.1 million individuals and 500,000 firms for the period 1976–1987. Kramarz and Roux (1998) have extended the dataset to 1995. This new dataset includes approximately two million individuals and one million firms.

Because of the centralized nature of the French statistical system, identical identifiers (firm or individual) can be found in different data sources. It is therefore very easy to match establishments from the DADS with other firm-level or establishment-level data sources such as balance-sheet information. It is possible, subject to the approval of the "Commission Nationale Informatique et Liberté" (CNIL), to match the DADS with other individual level datasets using the person identifier. However, due to the CNIL policy, this has not been done frequently in the past. The most important example is the match between the DADS and the Echantillon Démographique Permanent (EDP). The EDP collects for 1/10th of the population, information drawn from Civil Status registers on marriage, births, deaths, as well as data from the decennial Censuses of Population (in particular, completed education).

The French Déclaration Mensuelle de Mouvements de Main d'Oeuvre (DMMO), used by Abowd et al. (1999b), is another administrative data source in which all establishments with 50 or more employees register all hiring or separations every month. Information on the workers includes age, sex, type of contract, type of entry (shortterm contract (CDD), longterm contract (CDI), or transfer from another establishment of the same firm), skill-level for all entries and age, sex, seniority at exit, type of exit (end of shortterm contract, quit, retirement, firing for cause, firing for economic reasons, transfer to another establishment of the same firm), military service, death, and skill-level for all exits. These movements are usually aggregated at the monthly level by categories of entry/exit and skill-level. Notice that no wage information is available. The data source includes the establishment identifiers required to link to other information on the establishment and enterprise, including employment structure. Thus the data are dynamically representative of establishments and of mobile workers.

Danmarks Statistik has constructed a similar database with longitudinal information on workers and their establishments (IDA, see Leth-Sørensen, 1995) based on administrative registers on individuals. All persons in the population are covered, irrespective of their labor market status, and are identified by their person ID. Starting in 1980, annual information on each person's labor situation at the end of November is available. For persons born after 1960, there are also references to the person IDs of their parents. Notice also that, since the 1970 Census of Population was the first ever to include this person number, it is possible to get information back to 1970. For all employed workers, the employing establishment identifier is known. The information available at the establishment level

consists of: years of operation, industry, and location. Many individual characteristics are collected: sex, age, family and marital status, education, employment experience, unemployment history, income, full-time or part-time job, hourly pay, seniority. There are, however, no other data on firms such as balance sheets, production, factor use or financial information. The same kind of data, based on individual registers, are also available in Norway (see Salvanes and Forre, 1997) and Sweden (Tegsjö and Andersson, 1998).

In Japan, an establishment register called the Establishment Census forms the basis of matched data that is dynamically representative. The information in the establishment census has been matched to wage information in the Basic Survey on Wage Structures, a probability sample of establishments with 5 or more employees. Other information on the firms is taken from periodic censuses of manufacturing and commercial establishments. See Hayami and Abe (1998) and Abe and Sofer (1996) for details.

In Germany, starting in January, 1973, in order to collect all the necessary information for unemployment insurance and health-retirement payments, employers have been required to report information regarding any employment relation subject to social security contributions (more specifically, at the beginning, at the termination, and on December 31st for any employee). The reporting form, known as the Historikdatei (HD), is collected by the Bundesantalt für Arbeit (BfA). A 1% sample of the HD has been used by the Institut für Arbeitsmarkt und Berufsforschung (IAB) to construct a research dataset called the Beschäftigungsstichprobe (BS), from January 1, 1975 to December 31, 1990. The information reported in every record includes sex, nationality, education, gross earnings over the spell (with both left- and right-censoring because of the floor and the ceiling in the base formula for the computation of contributions), reasons for interruption of the spell (maternity leave, military service). As in other countries, self-employed individuals as well as civil-servants are not covered by the data. The HD comprises 79% of the labor force in 1979 (see Dustmann and Meghir, 1997, for further references to this data file). In addition to the BS file, the IAB has added information from another administrative data source, the Leistungsempfangerdatei (LD). The LD provides information on all spells that resulted in benefits from the BfA: unemployment benefits, unemployment assistance, and payments while in training program. Individuals can be followed from employment to registered unemployment spells. The IAB dataset (i.e., the BS plus the LD datasets) also contain a plant and a firm identifier. Using the entire HD dataset, aggregate individual characteristics have been created at the establishment-level, making firm size and within-firm educational structure available.

Similar data are also available in Austria (see Winter-Ebmer and Zweimüller, 1997) and in Italy (see Contini et al., undated). For Belgium, the data used by Leonard and Van Audenrode (1996, 1997) are based on Social Security declarations for the national pension system of private sector workers and cover the period from 1977 to 1985.

2.5. Representative matched worker-firm panels (statistical surveys)

The French Labor Force Survey (Enquête Emploi, EE) is conducted every year by the

French National Statistical Institute (INSEE). Because this survey routinely includes the employer identifier (firm and establishment), it has become a standard for matched employer–employee database upon labor force surveys.

The universe of individuals sampled in EE includes all ordinary households in metropolitan France. In 1990, INSEE started a new series of March EEs, administered to the household sample every March for three consecutive years using a sampling frame based on the 1990 census. The sampling rate is 1/300. There are three rotation groups, so the sample is refreshed by one-third every year. Each year, a supplement (enquête complémentaire) is administered to the outgoing rotation group, one-third of the sample. Because the sampling technique is based on housing in tracts built in French territory with further inclusions or modifications in case of construction or reconstruction of buildings not known at the 1990 Census of Population, it is possible to have a dynamically representative survey (see INSEE, 1994 for all the technical details on the survey methodology).

The data collected in the EE include both standard and more unusual questions from labor force surveys–wage, country of origin, sex, marital status, number of children and their ages, region of residence, age, detailed education, age at the end of education period, occupation (4-digit classification), father's last occupation, mother's last occupation, employment status (employed, unemployed, inactive), usual number of hours, seniority in the employing firm, sector and size of the employing firm, nature of the contract (shortterm, longterm, program for young workers (stage)) for each of the individuals in the sample. Furthermore, each employed individual is asked the name and address of the employment location. This information is given to the INSEE regional agencies where the Siret (establishment identification number) is coded using the on-line SIRENE computerized system. This number is the unique establishment identifier that links the employer to the rest of the French statistical system. The first nine digits represent the enterprise to which the establishment belongs, based on an economic and not a financial definition. Employer Siret number can be coded in the EE for more than 90% of the workers. Hence, it becomes possible to use this type of dataset in the same fashion as the DADS was used in Abowd et al. (1999a) (see Goux and Maurin, 1999). In particular, the EE can be matched with other firm level datasets as the Echantillon d'Entreprises (based on the BIC), the Déclaration de Mouvements de Main d'Oeuvre (DMMO), a record of all entries and exits in all establishments with at least 50 employees (see Abowd et al., 1999b). Such matches have been performed by Entorf et al. (1999) to study New Technologies or Kramarz (1997) to analyze trade, wages, and unemployment.

In the United States, a longitudinally representative matched employer–employee data file has been created for the National Longitudinal Survey of Youth 1979 Cohort (NLS-Y).[5] The description is based on Abowd and Finer (1998). The creation required the resolution of two conceptual difficulties and one procedural problem. The conceptual difficulties were (1) defining an employer and (2) specifying the level of aggregation to

[5] As Hildreth and Pudney (1998) note, samples of this structure are representative for the target age group of the population but the resulting sample of employers is not necessarily representative of employers.

use on the employer unit. The procedural problem is to find a method for performing the analysis that is consistent with the confidentiality requirements that have been specified for NORC and the Center for Human Resource Research at Ohio State University, the two primary contractors for the survey.

The simplest and most comprehensive definition of an employer is any organization for which the respondent completed the employer questionnaire during any year of the NLS-Y. For the purposes of preparing the matched data file, this definition maximizes the number of employers for which information would be available. Employers are divided into primary employers (main job; full- and part-time employees) and secondary employers (no main job or several part-time jobs). Ultimately, all types of employers will be covered; thus, private for-profit employers (firms), public sector employers (units of government) and private not-for-profit employment (other organizations) would all be included in the file. Some summary measures about the employer (size, type) are available for all types of employers. Other measures (sales, profits, assets) are only available for some private for-profit employers. Detailed analysis of the characteristics of the employing firms, therefore, requires careful attention to the type of firm. The level of aggregation to use for the employer depends upon the purpose of the analysis and the prospects for collecting data at that level of aggregation. Three potential definitions are possible: establishment (the physical location where work occurs), business or governmental unit (the economic entity at which decisions are made concerning employment, investment, etc.) or company/governmental aggregate (the entity required to disclose information to public sources). Currently, the employer identifier file includes an ID for the company/governmental aggregate and the business/governmental unit, where possible. This level of identification permits merging information about companies and lines of business from sources like Compustat and Dun & Bradstreet. More specifically, approximately 49,000 unique employer names were checked for relevant (time period consistent) matches in a variety of public sources.

This matching process was done in several phases. First, the raw files of the NLS-Y for the years 1986–1994 were accessed to acquire the employer names for up to 5 employers per year. The first stage match attempts to match the respondent employer names with employer names in the Compustat (Standard and Poors) and CRSP header files. There were approximately 159,000 non-blank employer name fields for the years 1986–1994. Government coded employers, self-employed jobs, and employers with less than 50 workers were all eliminated. This left exactly 48,422 unique employer names eligible for match. These employer names were placed in a database with the Compustat headers and CRSP name histories. One by one, the respondents employers were checked against the Compustat headers. At the end of this process, around 8000 employer names were matched with Compustat and CRSP employers. These unique names accounted for roughly 18% of the master list of employer names. In addition to checking for matches, unmatched records were coded for additional checking, military employer, and public or non-profit. Unmatched small employers are left initially unmatched. The second stage match was used to double-check suspicious first stage matches, and further match

unmatched first stage names that may be subsidiaries of publicly-traded parent companies. A total of 9000 such names were resolved using the Directory of Corporate Affiliations for several years both in printed and machine readable formats. These second-stage names were then recoded to reflect their status as private companies, subsidiaries of public parents, franchises, help-supply services. The third phase of the match procedure was to check improperly coded government, non-profit, and military employers. Approximately 2000 employer names were checked and coded as religious organizations, military, federal government, state government, local government, and educational institutions. In addition, private and non-profit health care facilities were reserved for the future processing. The fourth stage consisted of internal matching of companies with no publicly disclosed parent that appear multiple times.

2.6. Non-representative cross-sections and panels of workers and firms

Not all datasets matching workers with their firms were designed by statistical agencies with the avowed goal of representativeness of the set of workers or firms in a country, a state, or any geographic unit of some importance. This is most apparent the matched job-firm data that have been studied by Groshen (1996), who uses employer-based salary surveys in many of her papers. In this subsection, we give examples of such datasets. Our requirement for discussion herein is that multiple firms in which multiple workers or jobs are surveyed be present.

Employers have conducted salary surveys for many years in which they collect matched worker (or job) and firm information. We base our description of the American salary surveys on Groshen (1996). Salary surveys are used by large employers as a source of information on external wage opportunities of the workers they employ. These employers are very different in nature and scope. Groshen cites the following examples: "the federal government, most of the regional Federal Reserve banks, Hay Associates, Inc., the American Hospital Association, the National Association of Business Economists, and the American Association of University Professors." Access to the data is generally granted to the members of the collecting association, which may entail a large fee or to clients (in the case of Hay Associates, Inc. for instance).

These datasets contain annual information on wages, including bonuses and incentives, of all persons with a job in predefined occupations. They also have information on the participating employers themselves: industry, total employment, and firm-specific compensation policy elements. As is obvious from the list of organizations that collect Salary Surveys, the coverage largely depends on the purpose of the user. Groshen notes that "if a survey is geographically based, then the occupations covered will be those commonly found and most comparable across industry: usually clerical, administrative, maintenance, and managerial positions." Hence, occupations such as secretaries, drivers, painters, accountants are often included. She adds that "These surveys have the advantages over industry and professional wage surveys that they allow control for regional

wage differences, they include many different industries, and they are longitudinal in establishments. While they do not cover all occupations, they do cover a broad mix...."

Industry-based surveys differ in their scopes. They generally cover a large fraction of those workers employed in a particular industry. This allows jobs and occupations to be very precisely defined. In particular, for blue-collar workers, information on training, on machines, and tools needed or used in the job is available.

Profession-based surveys focus on one narrowly-defined occupation and tend to be national in scope since professions have generally a national market, the characteristics of which the survey organizers want to know. In particular, information on the educational background and employment experience of the participants is often collected. In his chapter on executive pay, Murphy describes many of these surveys for CEOs.

All these different types of salary surveys have several common features. First, the description of the job is very detailed, "two to three paragraphs long, and specify the responsibilities, training requirements, how the job is done, what is produced, position in the corporate hierarchy, the occupation of direct supervisor, and number of supervisees," according to Groshen. Furthermore, the jobs may be classified into job families defined as all members of a career path. Finally, demographic information is usually not collected.

Although salary surveys are an important source of information in the private sector of the American economy, Groshen is one of the few to use these for research purposes. A number of researchers, however, have recently made use of similar Bureau of Labor Statistics surveys of occupational or industrial salaries matched to employer information. These are noted in Table 1 and discussed in the relevant sections below. Most of the design features noted by Groshen apply to these surveys as well.

Brown and Medoff (1996) describe a dataset for which individuals interviewed for the Survey of Consumers, a survey run by the Survey Research Center at the University of Michigan, between September 1991 and March 1992 were asked to complete a supplement on their employer. Supplementary questions were only asked to workers with a private-sector employer. These questions included workers' experience, seniority, occupation, and wage rate as well as information on the employer, more specifically, the collective bargaining status, the number of employees, the industry, the age of the business, fringe benefits, personnel policies, and related features of the workplace. The sample has 1410 private-sector workers of which 1168 gave information on the name and address of their employer. Brown and Medoff asked Dun & Bradstreet to locate the employer and, when located, to give the establishment and the company employment, the age of the business, and the industry. All of 863 reported matches were hand-checked, generating a set of 701 "clean matches" as described in Brown and Medoff (1996). Employers in those clean matches are larger and older (longer in business) than the employers in the original sample.

A recent study by Chennouf et al. (1997) uses matched worker-firm data for a small sample of Algerian firms in the Algiers region. This dataset comprises 42 firms from diverse manufacturing industries and 1007 employees. The available individual characteristics are the wage, the number of days worked, the education level, seniority, experience,

age, sex, and the marital status. On the firms themselves, apart from the industry and private/public status, variables are mostly defined as an aggregation of the individual characteristics of workers employed at that firm (average seniority, experience, and education).

3. Statistical models for matched employer–employee datasets

3.1. The basic linear model

Virtually all of the papers that we discuss below use a variant of the linear model that can be identified with matched employer–employee data:

$$y_{it} = x_{it}\beta + \theta_i + \psi_{J(i,t)it} + \varepsilon_{it}, \tag{3.1}$$

in which y_{it} is some measured outcome (compensation, layoff event, etc.) for the individual $i = 1, \ldots, N$ at date $t = 1, \ldots, T$; x_{it} is a vector of P time-varying exogenous characteristics of individual i; θ_i is a pure person effect; $\psi_{J(i,t)it}$ is a pure firm effect for the firm at which worker i is employed at date t (denoted by $J(i,t)$), and ψ_{it} is a statistical residual. Assume that a simple random sample of N individuals is observed for T years. The firm and person effects in Eq. (3.1) can be decomposed into components relating to seniority and non-time-varying personal characteristics as follows:

$$\psi_{jit} = \phi_j + \gamma_j s_{it}, \tag{3.2}$$

where s_{it} denotes individual i's seniority in firm $j = J(i,t)$ in year t, ϕ_j denotes the firm-specific intercept, and γ_j is the firm-specific seniority coefficient; while

$$\theta_i = \alpha_i + u_i\eta, \tag{3.3}$$

where u_i is a vector of non-time-varying measurable personal characteristics, α_i is the person-specific intercept, and η is the vector of coefficients on the non-time-varying personal characteristics.

In matrix notation we have

$$y = X\beta + D\theta + F\psi + \varepsilon, \tag{3.4}$$

where X is the $N^* \times P$ matrix of observable, time-varying characteristics, D is the $N^* \times N$ matrix of indicators for individual $i = 1, \ldots, N$, F is the $N^* \times J$ matrix of indicators for the firm at which i works at date t (J firms total), y is the $N \times 1$ vector of outcomes, ε is the conformable vector of residuals, and $N^* = NT$. Balanced samples are not necessary but simplify the discussion of the statistical models. The firm effect can also have higher dimension, as for example in Eq. (3.2), but we use this simpler form for the discussion herein.

The parameters of Eq. (3.4) are β, the $P \times 1$ vector of coefficients on the time-varying personal characteristics; θ, the $N \times 1$ vector of individual effects; ψ, the $J \times 1$ vector of

firm effects; and the error variance, σ_ε^2. The parameter θ includes both the unobservable (to the statistician) individual effect and the coefficients of the non-time-varying personal characteristics. Eqs. (3.1) and (3.4) are interpreted as the conditional expectation of individual outcomes given information on the observable characteristics, the date of observation, the identity of the individual, and the identity of the employing firm. In this section we want to make precise the interpretation of Eq. (3.4) under classical least squares when some of the effects, β, θ, and ψ are missing or are aggregated into linear combinations. The discussion draws heavily on Abowd et al. (1999a).

3.2. Aggregation and omitted variable biases

The omission or aggregation of one or more of the effects in Eq. (3.4) can change the meaning of the other effects in important and subtle ways that are not always clear from the specific equation that various authors have estimated. Variations in the set of conditioning effects, which give rise to omitted-variable biases, are one source of confusion about the interpretation of the statistical parameters. The use of different linear combinations of the effects in Eq. (3.4), which gives rise to aggregation biases, is another source of differential interpretations for the parameters. These are considered in turn.

When the estimated version of Eq. (3.4) excludes the pure firm effects (ψ), the estimated person effects, θ^*, are the sum of the pure person effects, θ, and the employment-duration weighted average of the firm effects for the firms in which the worker was employed, conditional on the individual time-varying characteristics, X:

$$\theta^* = \theta + (D'M_XD)^{-1}D'M_XF\psi, \tag{3.5}$$

where the notation $M_A \equiv I - A(A'A)^{-1}A'$ for an arbitrary matrix A. Hence, if X were orthogonal to D and F, so that $D'M_XD = D'D$ and $D'M_XF = D'F$, then the difference between θ^* and θ, which is just an omitted variable bias, would be an $N \times 1$ vector consisting, for each individual i, of the employment-duration weighted average of the firm effects ψ_j for $j \in \{J(i,1,\ldots,J(i,T)\}$:

$$\theta_i^* - \theta_i = \sum_{t=1}^{T} \frac{\psi_{J(i,t)}}{T}.$$

The estimated coefficients on the time-varying characteristics in the case of omitted firm effects, β^*, are the sum of the parameters of the full conditional expectation, β, and the omitted variable bias that depends upon the conditional covariance of X and F, given D:

$$\beta^* = \beta + (X'M_DX)^{-1}X'M_DF\psi.$$

Similarly, omitting the pure person effects (θ) from the estimated version of Eq. (3.4) gives estimates of the firm effects, ψ^{**}, that can be interpreted as the sum of the pure firm effects, ψ, and the employment-duration weighted average of the person effects of all of the firm's employees in the sample, conditional on the time-varying individual characteristics:

$$\psi^{**} = \psi + (F'M_X F)^{-1} F'M_X D\theta. \tag{3.6}$$

Hence, if X were orthogonal to D and F, so that $F'M_X F = F'F$ and $F'M_X D = F'D$, then the difference between ψ^{**} and ψ, again an omitted variable bias, would be a $J \times 1$ vector consisting, for each firm j, of the employment-duration weighted average of the person effects θ_i for $i \in \{J(i,t) = j \text{ for some } t\}$:

$$\psi_j^{**} - \psi_j = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{\theta_i 1(J(i,t) = j)}{N_j} \right],$$

where

$$N_j = \sum_{i=1}^{N} \sum_{t=1}^{T} 1(J(i,t) = j),$$

and the function $1(A)$ takes the value 1 when A is true and 0 otherwise. The estimated coefficients on the time-varying characteristics in the case of omitted individual effects, β^{**}, are the sum of the parameters of the full conditional expectation, β, and the omitted variable bias that depends upon the covariance of X and D, given F:

$$\beta^{**} = \beta + (X'M_F X)^{-1} X'M_F D\theta. \tag{3.7}$$

Almost all existing analyses of equations like (3.4) produce estimated effects that confound pure person and pure firm effects in a manner similar to that presented above. The possibility of identifying both person and firm effects thus allows users of matched employer–employee data to reexamine many important topics in labor economics using estimates that properly allocate the statistical effects associated with persons and firms. Of course, other identification issues also arise, such as in the estimation of person effects, so that longitudinal matched data are usually required.

3.3. Identification of person and firm effects

Although Eq. (3.1) is just a classical linear regression model, the full design matrix $[X \;\; D \;\; F]$ has high column dimension. The cross-product matrix

$$\begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix} \tag{3.8}$$

is patterned in the elements $D'D$ and $F'F$. The separate identification of the individual and firm effects requires the presence in the sample of individuals who move from firm to firm. The individual and firm effects are both identified whenever an individual that appears in the sample works for a firm that employs at least one individual, also in the sample, who moves to another firm, which, necessarily, also appears in the sample. The simplest

example of the complexities of identification in this model can be seen by considering an example in which there are three individual (1, 2, and 3), two firms (A and B) and two time periods. Suppose that individual 1 is continuously employed at firm A, individual 2 is continuously employed at firm B, and individual 3 moves from firm A to firm B. Then all three individual effects are identified (subject to the usual identification restriction that they sum to zero) and both firm effects are identified (again, subject to the usual identification condition that they sum to zero). If individual 3 is not mobile (stays at firm A), then firm effect B cannot be distinguished from person effect 2 and person effects 1 and 3 are entirely within firm effect A.

There are many computational difficulties associated with inverting the matrix (3.8). These computational problems are directly related to the fact that the basic statistical model is neither hierarchical nor balanced. Thus, projecting onto the columns D, the method usually called "within persons estimation," leaves a high-dimension unpatterned, non-sparse matrix to invert for the solution for β and ψ. Similarly, projecting onto the columns of F, the method usually called "within firms estimation," leaves a high-dimension unpatterned, non-sparse matrix to invert to solve for β and θ. Clearly, the usual computational methods for least squares estimation of the parameter vector $[\beta' \ \theta' \ \psi']'$ are not generally feasible. Hence, one usually cannot compute the unconstrained least squares estimates for the model (3.1). Correlated random effect models, which permit the estimation of all effects without restricting the design matrix in Eq. (3.8), also require solution of the full least squares normal equations (see Scheffé, 1959; Searle et al. 1992). See Abowd et al. (1999a) for a detailed discussion of the identification and estimation issues in models using Eq. (3.1).

3.4. Aggregation and omitted variable biases for inter-industry wage differentials[6]

Define a pure inter-industry wage differential, conditional on the same information as in Eqs. (3.1) and (3.4), as κ_k for some classification $k = 1, \ldots, K$. By definition, pure firm effects are fully nested within pure inter-industry effects so that κ_k can be represented as an employment-duration weighted average of the firm effects within the classification k:

$$\kappa_k \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{1(K(J(i,t)) = k)\psi_{J(i,t)}}{N_k} \right],$$

where

$$N_k \equiv \sum_{j=1}^{J} 1(K(j) = k) N_j$$

and the function K(j) denotes the classification of firm j. If we insert pure inter-industry effects as the appropriate aggregate of the firm effects in Eq. (3.1), then the equation

[6] This subsection draws heavily on Abowd et al. (1999a).

becomes

$$y_{it} = x_{it}\beta + \theta_i + \kappa_{K(J(i,t))} + (\psi_{J(i,t)} - \kappa_{K(J(i,t))}) + \varepsilon_{it}$$

or, in matrix notation as in Eq. (3.4),

$$y = X\beta + D\theta + FA\kappa + (F\psi - FA\kappa) + \varepsilon, \tag{3.9}$$

where the matrix A, $J \times K$, classifies each of the J firms into one of the K categories; that is, $a_{jk} = 1$ if, and only if, $K(j) = k$. The parameter vector κ, $K \times 1$, may be interpreted as the following weighted average of the pure firm effects:

$$\kappa \equiv (A'F'FA)^{-1}A'F'F\psi,$$

and the effect $(F\psi - FA\kappa)$ may be re-expressed as $M_{FA}F\psi$. Thus, the aggregation of J firm effects into K inter-industry effects, weighted so as to be representative of individuals, can be accomplished directly by estimation of Eq. (3.9). Only rank$(F'M_{FA}F)$ firm effects can be separately identified; however, there is neither an omitted variable nor an aggregation bias in the classical least squares estimates of (3.9).

Estimates of inter-industry effects, κ^*, that are computed on the basis of an equation that excludes the remaining firm effects, $M_{FA}F\psi$, are equal to the pure inter-industry effect, κ, plus an omitted variable bias that can be expressed as a function of the conditional variance of the inter-industry effects, FA, given the time-varying characteristics, X, and the person effects, D:

$$\kappa^* = \kappa + \left(A'F'M_{[D|X]}FA\right)^{-1}A'F'M_{[D|X]}M_{FA}F\psi, \tag{3.10}$$

which simplifies to $\kappa^* = \kappa$ if, and only if, the inter-industry effects, FA, are orthogonal to the subspace $M_{FA}F$, given D and X, which is generally not true even though FA and $M_{FA}F$ are orthogonal by construction. Thus, it is not possible to estimate pure inter-industry wage differentials consistently, conditional on time-varying personal characteristics and unobservable non-time-varying personal characteristics, without explicit firm-identifiers unless this conditional orthogonality condition holds. A similar argument applies to the estimates of β. Industry effects as defined by Eq. (3.10) are directly comparable to those estimated by Krueger and Summers (1988) when they include person effects.

When the estimation of Eq. (3.9) excludes both person and firm effects, the estimated inter-industry effect, $\kappa_k{}^{**}$, equals the pure inter-industry effect, κ, plus the employment-duration weighted average residual firm effect inside the category k, given X, and the employment-duration weighted average person effect inside the category, given the time-varying personal characteristics X:

$$\kappa^{**} = \kappa + (A'F'M_XFA)^{-1}A'F'M_X(M_{FA}F\psi + D\theta),$$

which can be restated as

$$\kappa^{**} = (A'F'M_XFA)^{-1}A'F'M_XF\psi + (A'F'M_XFA)^{-1}A'F'M_XD\theta. \tag{3.11}$$

Hence, the raw inter-industry effects consist of the sum of the properly-weighted average person effect and average firm effect, conditional on X. Thus, analyses that exclude person effects confound the pure inter-industry wage differential with an average of the person effects found in the category, given the measured personal characteristics, X. The inter-industry wage differentials in Eq. (3.11) are directly comparable to those studied by Krueger and Summers (1988) when person effects are omitted.

3.5. Aggregation and omitted variable biases for inter-person wage differentials

Another line of research attempts to explain inter-personal wage differentials conditional on firm effects without explicit controls for unobservable personal heterogeneity. None of the studies in this strain of the wage-determination literature includes both pure person and pure firm effects, as defined in Eq. (3.1) or (3.4) above. In our notation, studies like Groshen (1991a) estimate ψ^{**}, from Eq. (3.6), and β^{**}, from Eq. (3.7).

3.6. Firm-size wage effects

The repeated finding of a positive relation between the size of the employing firm and wage rates, even after controlling for a wealth of individual variables (see Brown and Medoff, 1989), has also generated many alternative interpretations. Properly modeled, the firm-size wage effect can also be fully decomposed using matched employee–employer data. Using our notation, a firm-size effect, δ, can be modeled using a matrix S, $J \times R$, that maps the size of firm j into R linearly independent functions of its size. Using the same methods as above, we express the wage equation, Eq. (3.4), as

$$y = X\beta + D\theta + FS\delta + M_{FS}F\psi + \varepsilon, \tag{3.12}$$

so that the pure firm-size effects are related to the underlying pure firm effects by the equation:

$$\delta \equiv (S'F'FS)^{-1}S'F'F\psi. \tag{3.13}$$

The firm-size effect is also an aggregation of the pure firm effects and can be analyzed using the same tools that we used for the inter-industry wage differential. The raw firm-size wage differential, δ^{**} (in our notation), can be represented as

$$\delta^{**} = (S'F'M_XFS)^{-1}S'F'M_XF\psi + (S'F'M_XFS)^{-1}S'F'M_XD\theta, \tag{3.14}$$

which can be interpreted as the sum of the firm-size, employment-weighted average firm effect and the similarly-weighted average person effect, conditional on personal characteristics, X, and firm size, FS.

3.7. Other methodological issues

There are a variety of technical statistical issues surrounding the use of different sampling frames to construct matched employer–employee issues. Recently, several

teams of authors have begun to examine these issues. Hildreth and Pudney (1997, 1999) examine the issues of non-random missing data, choice based sampling induced by the matching process and correlated random effects modeling of the heterogeneity. They provide full likelihoods for the hierarchical case (individual effects are fully within firm effects) and some likelihood models for non-hierarchical case (individuals move from firm to firm within the sample). Abowd et al. (1999c) address the issue of non-random missing data following the match. Dolton, Lindeboom and van den Berg (1999) address the issues of non-random missing matches (of the employer or employee), attrition and endogenous sampling. Mairesse and Greenan (1999) consider the problem of modeling employee and employer behavior when only a single employee observation is available per firm, as is common in matched training surveys.

4. From theoretical models to statistical models: potential interpretations of the descriptive models

We illustrate the relation between structural heterogeneity in the populations of workers (heterogeneous abilities or tastes) and firms (heterogeneous efficiencies or technologies) and the statistical heterogeneity in Eq. (3.1) using four economic models with very simple population structures. In each case we derive the conditional expectation of individual compensation given the identity of the employing firm and the individual. We then relate the parameters of this conditional expectation to the statistical parameterization above.

4.1. Measurement of the internal and external wage

Virtually all economic models of labor market outcomes require an estimate of the opportunity cost of the worker's time. In simple, classical equilibrium models without unmeasured person or firm heterogeneity, this generally corresponds to the measured wage rate. In models of wage determination such as quasi-rent splitting or imperfect information (efficiency wage and agency models), unmeasured statistical heterogeneity (person or firm) breaks the direct link between the observed wage rate and the opportunity cost of time. Moreover, such models usually make an explicit distinction between the compensation received and the wage rate available in the employee's next best alternative employment. The statistical model in Eq. (3.1), while not derived from an explicit labor market model, contains all the observable elements from which non-classical labor market models derive their empirical content. Indeed, the simplest definition of the components of the external and internal wage rate based a structural model leading to Eq. (3.1) is given by the following model:

$$y_{it} = x_{it}\zeta + \nu_{it},$$

where $\{x_{it}, \nu_{it}\}$ follows a general stochastic process for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ with

$$\mathrm{E}[\{x_{it}, \nu_{it}\}\{x_{ns}, \nu_{ns}\} \mid i, n, s, t, J(i,t), J(n,s)] \neq 0 \qquad \text{iff } i = n \text{ or } J(i,t) = J(n,s).$$

Then,

$$\theta_i = \mathrm{E}[x_{it}\zeta + \nu_{it} \mid i] - \mathrm{E}[x_{it}\zeta + \nu_{it}]$$

and

$$\psi_j = \mathrm{E}[x_{it}\zeta + \nu_{it} \mid J(i,t) = j] - \mathrm{E}[x_{it}\zeta + \nu_{it}].$$

4.2. A matching model with endogenous turnover

This model is based on Jovanovic (1979). Suppose that workers are homogeneous. There are two types of firms, m and n, and two periods. In type m firms a worker's marginal product and wage rate are always w^*, and employment is always available in a type m firm. In type n firms there is a matching process. Worker i's productivity is $w^* + \varepsilon_{in}$ in both periods with ε_{in} drawn from a binomial distribution $B(-H, H, 1/2)$. The matching outcome, ε_{in}, unknown to both the worker and the firm at the beginning of the first period of employment, is realized at the end of the first period and becomes public information. Workers are offered contracts at the beginning of the first period of the form (w_1, w_2) and workers may leave firm n at the end of the first period. All workers are risk-neutral and earn no rents. The equilibrium contract for firms of type n is $(w^* - H/2, w^* + \varepsilon_{in})$. All workers in type n firms with a bad matching outcome $-H$ quit to type m firms.

To simplify the model, we consider a stationary situation with nine workers who live for two periods each, three born in period 0, three born in period 1, three born in period 2. Two workers in each generation enter type n firms, one worker in each generation enters a type m firm. Of the two workers who entered type n firms, let one draw a positive matching outcome and the other draw a negative matching outcome. The worker with the negative matching outcome leaves the type n firm for a type m firm when the matching parameter is made public.

The structure of the data implied by this theoretical model is shown in Table 2. This corresponds to the following parameter values in the descriptive model:

$$\mu = w^*,$$

where μ is the overall mean;

$$\alpha_i = 0, i = 1, \ldots, 9,$$

where α_i is person i person-effect;

$$(\phi_m, \gamma_m) = (0, 0),$$

for the type m firm compensation policy; and

$$(\phi_n, \gamma_n) = \left(-\frac{H}{2}, \frac{3H}{2}\right),$$

for the type n firm compensation policy.

4.3. A rent-splitting model with exogenous turnover

Suppose there are four different individuals, two types of firms, m and n, and two time periods. Each of the two firms earns quasi-rents of q_{it}, and the quasi-rents are split by negotiation so that the workers receive a share s_j of the quasi-rent in firm j. Suppose that each firm employs two workers. With probability one, exactly one worker is randomly selected to separate from the period one employer and be re-employed at the other firm in the second period. All information about the workers and firms is known to those parties but not to the statistician. All workers are included in the data sample and the typical worker has wages of the form

$$y_{it} = x_i + s_j q_{jt},$$

where x_i is the measure of wage rate heterogeneity, i.e., the worker type, q_{jt} follows a binomial distribution $B(-Q, Q, 1/2)$, $i = 1, \ldots, 4$, $j = m, n$, and $t = 1, 2$.

Table 3 shows the relation among the theoretical parameters, x_i, s_j, and Q, and the statistical parameters of Eq. (3.1) for each worker and each period. The model cannot be solved exactly. Thus, we use these relations to solve, by least squares, the moment equations that determine the relations between the statistical parameters and the model parameters. This yields

$$\mu = \frac{1}{4}\sum_{i=1}^{4} x_i,$$

where μ is the overall mean;

$$\alpha_1 = \frac{1}{4}\left(-3s_m Q - s_n Q - \sum_{i=1}^{4} x_i\right) + x_1,$$

$$\alpha_2 = \frac{1}{4}\left(-s_m Q - 3s_n Q - \sum_{i=1}^{4} x_i\right) + x_2,$$

$$\alpha_3 = \frac{1}{4}\left(s_m Q + 3s_n Q - \sum_{i=1}^{4} x_i\right) + x_3,$$

$$\alpha_4 = \frac{1}{4}\left(3s_m Q + s_n Q - \sum_{i=1}^{4} x_i\right) + x_4,$$

where the α_i are the four person effects;

$$(\phi_m, \gamma_m) = \left(\frac{(s_n - s_m)Q}{4}, 2s_m Q\right)$$

and

Table 2
Matching model with homogeneous workers[a]

Individual	Wage period 1	Wage period 2
1	$y_{11} = \mu + \alpha_1 + \phi_m = w^*$	$y_{12} = \mu + \alpha_1 + \phi_m + \gamma_m = w^*$
2	$y_{21} = \mu + \alpha_2 + \phi_m = w^*$	
3	$y_{31} = \mu + \alpha_3 + \phi_m + \gamma_m = w^*$	
4		$y_{42} = \mu + \alpha_3 + \phi_m = w^*$
5	$y_{51} = \mu + \alpha_5 + \phi_n = w^* - (H/2)$	$y_{52} = \mu + \alpha_5 + \phi_n + \gamma_n = w^* + H$
6	$y_{61} = \mu + \alpha_6 + \phi_n + \gamma_n = w^* + H$	
7	$y_{71} = \mu + \alpha_7 + \phi_n = w^* - (H/2)$	$y_{72} = \mu + \alpha_7 + \phi_m = w^*$
8		$y_{82} = \mu + \alpha_8 + \phi_n = w^* - (H/2)$
9		$y_{92} = \mu + \alpha_9 + \phi_n = w^* - (H/2)$

[a] Individual 1 enters type m firm in period 1; individual 2 entered type m firm in period 0 (before period 1); individual 3 entered type n firm in period 0 (before period 1), had a negative matching outcome and left for a type m firm; individual 4 enters type m firm in period 2; individual 5 enters type n firm in period 1, has a positive matching outcome; individual 6 entered type n firm in period 0 (before period 1), had a positive matching outcome and remained in type n firm for period 1; individual 7 enters type n firm, has a negative matching outcome and leaves for a type m firm in period 2; individuals 8 and 9 enter type n firm in period 2.

$$(\phi_n, \gamma_n) = \left(\frac{(s_n - s_m)Q}{4}, -2s_nQ \right)$$

are respectively the type m and type n firms' policies.

4.4. An incentive model with unobserved individual heterogeneity

Following Kramarz and Rey (1995) , consider workers who are heterogeneous with respect to a parameter $q \in [0, 1]$, which is known to them but not known to the firms. Suppose, furthermore, that there are two types of firms, m and n, that differ according to their technology, and that there are two time periods. At type m firms, workers are hired for one period and have a level of productivity y^* regardless of their q. At type n firms, workers are hired in period one, produce y regardless of their q, and choose an effort level, either 0 or E, to exert during on-the-job training. At the end of the first period, workers in firm type n take a formal, verifiable test. If worker q exerts effort E, the test is passed with probability q. Otherwise, the test is passed with probability kq, where $(0 < k < 1)$. At the beginning of the second period, the firm decides which workers to keep and the workers may leave on their own. Workers who exert effort E have a level of productivity in the second period of $y + \tau_q$ if they remain in a type n firm.

There are many type m firms and two type n firms, which compete for workers in both periods. Workers in type m firms always receive a wage w^*. Workers in type n firms are offered a wage contract $(w_1(q), w_2(q), b(q))$, where $w_1(q)$ is the first period wage, $w_2(q)$ is the second period wage, and $b(q)$ is the bonus paid to those who pass the test. In equilibrium

Table 3
Rent-splitting model[a]

Individual	Wage period 1	Wage period 2
1	$y_{11} = \mu + \alpha_1 + \phi_m = x_1 - s_m Q$	$y_{12} = \mu + \alpha_1 + \phi_m + \gamma_m = x_1 + s_m Q$
2	$y_{21} = \mu + \alpha_2 + \phi_m = x_2 - s_m Q$	$y_{22} = \mu + \alpha_2 + \phi_n = x_2 - s_n Q$
3	$y_{31} = \mu + \alpha_3 + \phi_n = x_3 + s_n Q$	$y_{32} = \mu + \alpha_3 + \phi_n + \gamma_n = x_3 - s_n Q$
4	$y_{41} = \mu + \alpha_4 + \phi_n = x_4 + s_n Q$	$y_{42} = \mu + \alpha_4 + \phi_m = x_2 + s_m Q$

[a] The quasi-rent is $-Q$ in type m firm in period 1 and q in period 2. The quasi-rent is q in type n firm in period 1 and $-Q$ in period 2. Individual 1 works in type m firm in both periods. Individual 2 works in type m firm in period 1 and in type n firm in period 2. Individual 3 works in type n firm in both periods. Individual 4 works in type n firm in period 1 and in type m firm in period 2.

all firms of both types make zero profits because of the competition to attract workers. Furthermore, if $y + \delta(y + \tau_q)$ is convex in q (δ being the rate of discount of future earnings), the equilibrium contract will be such that $w_1(q) = y - qb(q)$, $w_2(q) = y + \tau_q$, and

$$b(q) = \frac{\mathrm{d}}{\mathrm{d}q}(y + \delta(y + \tau_q)).$$

All workers with type q, $q \geq p$, will choose to enter one of the type n firms and will choose to exert effort E when $b(p) \geq E/(1 - k)p$.[7] To simplify the model, we suppose that $\tau_q = \tau(Q^2/2)$ and that parameters are such that $p = 1/3$. We also suppose that there are nine workers, three of whom are employed by type m firms and the remaining six work in type n firms.

Table 4 shows the wage of every individual in each firm and in each period in terms of the theoretical model, as well as in terms of the descriptive model. These equations can be solved in order to express each parameter of the descriptive model using parameters of the theoretical model. As in the rent-splitting model, the solution is not exact–we must use least squares to express the function of the theoretical parameters that is closest to the statistical parameter. To see why, consider the workers in type n firms. Individual 7 passed the test and, consequently, received a bonus. This result generates a seniority slope for individual 7. Individual 8 did not pass the test and therefore received no bonus in period 2. Thus individual 8 has a different seniority slope in the same firm. The statistical parameter γ_n measures the average seniority slope in the firm n. Thus, the resulting estimated seniority slope will be the least squares estimate of the average of the two slopes. We illustrate these solutions for all the statistical parameters below.

The overall mean, μ, is given by the following:

$$\mu = \frac{\delta\tau}{18}\sum_{i=4}^{7} q_i\left(1 - \frac{q_i}{2}\right) - \frac{\delta\tau}{18}\sum_{i=8}^{9}\frac{q_i^2}{2} + \frac{w^*}{3} + \frac{2y}{3}.$$

[7] Proofs of all these assertions can be found in Kramarz and Rey (1995).

The individual effects, α_i, $i = 4, 5, 6, 7$ are

$$\alpha_i = \frac{\delta\tau}{24}\left[\sum_{j=4, j\neq i}^{7} q_j\left(\frac{q_j}{2} - 2\right) + 5q_i(2 - q_i) + \sum_{j=8}^{9} q_j^2\right], \qquad i = 4, 5, 6, 7$$

and those for individual $i = 8,9$ are

$$\alpha_i = \frac{\delta\tau}{24} \sum_{j=4, j\neq i}^{7} q_j(q_j - 2) + q_k^2 - 5q_i^2],$$

where $k = 8, 9$, $i \neq k$. Finally, the individual effects for $i = 1, 2, 3$ and the firm effects for m are not separately identifiable, since there are no movements between firms. We arbitrarily set

$$\alpha_i = 0, \qquad i = 1, 2, 3$$

for these individuals, implying a firm effect of

$$\phi_m = \frac{\delta\tau}{36}\left[\sum_{i=4}^{7} q_i(q_i - 2) + \sum_{i=8}^{9} q_i^2\right] + \frac{2w^*}{3} - \frac{2y}{3}.$$

For type n firms, we have

$$\phi_n = \frac{\delta\tau}{36}\left[\sum_{i=4}^{7} q_i(-5q_i - 2) - 5\sum_{i=8}^{9} q_i^2\right] - \frac{w^*}{3} + \frac{y}{3}.$$

The seniority slopes are

$$\gamma_m = 0$$

for firm m and

$$\gamma_n = \frac{\delta\tau}{12}\left[\sum_{i=4}^{7} q_i(3q_i + 2) + 3\sum_{i=8}^{9} q_i^2\right]$$

for firm n.

Notice that the α_i of the workers in the type n firm depend upon their hidden characteristics q_i as well as the characteristics of their fellow workers. Note also that the intercept in type m firms is larger than that of type n firms. Finally, as mentioned above, the seniority slope, γ_n, in type n firms is the least squares average of the career paths in the firm, depending on the success or failure of the test.

No single economic model is likely to explain a large, diverse labor markets like the ones studied in virtually all of the papers we discuss below. Nevertheless, it is important to keep in mind that it is not always possible to make a direct interpretation of the statistical parameters (for individuals or firms) in terms of simple economic parameters. In general,

Table 4
Incentive model with heterogeneous workers[a]

Individual	Wage period 1	Wage period 2
1	$y_{11} = \mu + \alpha_1 + \phi_m = w^*$	$y_{12} = \mu + \alpha_1 + \phi_m + \gamma_m = w^*$
2	$y_{21} = \mu + \alpha_2 + \phi_m = w^*$	$y_{22} = \mu + \alpha_2 + \phi_m + \gamma_m = w^*$
3	$y_{31} = \mu + \alpha_3 + \phi_m = w^*$	$y_{32} = \mu + \alpha_3 + \phi_m + \gamma_m = w^*$
4	$y_{41} = \mu + \alpha_4 + \phi_n = y - \delta\tau q_4^2$	$y_{42} = \mu + \alpha_4 + \phi_n + \gamma_n = y + (\delta\tau/2)q_4^2 + \delta\tau q_4$
5	$y_{51} = \mu + \alpha_5 + \phi_n = y - \delta\tau q_5^2$	$y_{52} = \mu + \alpha_5 + \phi_n + \gamma_n = y + (\delta\tau/2)q_5^2 + \delta\tau q_5$
6	$y_{61} = \mu + \alpha_6 + \phi_n = y - \delta\tau q_6^2$	$y_{62} = \mu + \alpha_6 + \phi_n + \gamma_n = y + (\delta\tau/2)q_6^2 + \delta\tau q_6$
7	$y_{71} = \mu + \alpha_7 + \phi_n = y - \delta\tau q_7^2$	$y_{72} = \mu + \alpha_7 + \phi_n + \gamma_n = y + (\delta\tau/2)q_7^2 + \delta\tau q_7$
8	$y_{81} = \mu + \alpha_8 + \phi_n = y - \delta\tau q_8^2$	$y_{82} = \mu + \alpha_8 + \phi_n + \gamma_n = y + (\delta\tau/2)q_8^2$
9	$y_{91} = \mu + \alpha_9 + \phi_n = y - \delta\tau q_9^2$	$y_{92} = \mu + \alpha_9 + \phi_n + \gamma_n = y + (\delta\tau/2)q_9^2$

[a] Individuals 1–3 belong to type m firm with q_i, $i = 1,2,3$ between 0 and 1/3, individuals 4–9 belong to type n firm with q_i, $i = 4$–9 above 1/3. Individuals 4–7 pass the test and receive the bonus; individuals 8 and 9 fail.

the interpretation of a given statistical parameter depends upon all the elements of the economic model under consideration.

5. New results with matched employer–employee datasets: compensation structure

5.1. Models with both person and firm effects

The papers we consider in this section all estimate a variant of the full model (3.1) and then use the results to consider related sets of questions about the links between individual heterogeneity, firm heterogeneity and observable wage differentials. We consider Abowd et al. (1999a,c), Bingley and Westergård-Nielsen (1996), Burgess et al. (1997), Goux and Maurin (1999), Finer (1997), Leonard and van Audenrode (1996, 1997), and Leonard et al. (1999). Belzil (1997) estimates the full model but is concerned primarily with worker mobility, see Section 7.5 for a discussion. Entorf et al. (1999) also estimate the full model, but their focus is on computer and wages, so this article is described in Section 7.4. Pacelli (1997) estimates the full model but is concerned primarily with seniority effects; see Section 5.3 for a discussion.

Abowd et al. (1999a) provide a very complete discussion of the statistical and economic issues surrounding estimation of Eq. (3.1) for log wage rates, much of which is summarized in Section 3. In their analysis of the French data from the DADS (see Table 1), they find that person effects, without controlling for non-time-varying personal characteristics, θ_i, or after such controls, α_i, account for 60–80% of the variation in log annual wage rates while the full firm effect (including a heterogeneous seniority effect discussed in Section 5.3) accounts for only 4–9%. The two effects are not highly correlated (0.09–0.26, depending on the statistical model).

Abowd et al. (1999a) use the estimated person and firm effects to address a number of

other questions. They show that raw interindustrial wage differentials, as defined in Eq. (3.11), can be decomposed into a part due to the industry-average person effect and a part due to the industry-average firm effect. The decomposition is exact when there is no estimation error in the relevant industry averages and the large sample sizes from the French data essentially eliminate this estimation error. These authors find that, for France, 90% of the raw inter-industry wage differential is explained by the industry-average person effects and between 7% and 25% is explained by the industry-average firm effect (according to the method used, the average effects are correlated at the industry level). They perform the same decomposition for the firm-size wage effect in France and the results are that 90% of the firm-size wage differential is due to the firm-size-average person effect and 25–40% is due to the firm-size-average firm effect (again, according to the estimation method for the basic effects and allowing for correlation among the firm-size averages). These differentials are examined in more detail in Abowd and Kramarz (1996a, 1998a).

Abowd et al. (1999c) examine the same questions as Abowd et al. (1999a) using two American data sources (Washington State UI and NLSY '79, see Table 1). As in France, individual heterogeneity (θ_i or α_i) is the most important source of variation in log wage rates for these American data, explaining about twice as much of the variance as firm-level heterogeneity. Again, as in France and the other countries discussed below, there is only a weak correlation between person and firm effects. The results concerning inter-industry wage differentials are again similar to those for France. Industry-average person effects are very important in explaining the raw differentials. In contrast to the results for France, however, the industry-average firm effect is also important in explaining the raw differentials, although less important than person effects.

In a series of papers, Leonard and Van Audenrode (1996, 1997) and Leonard et al. (1999), consider the wage determination process using longitudinal matched Belgian data that are capable of identifying both firm and individual effects as defined in the full model (3.1). Because their focus is on wage and employment mobility and, in particular, the interaction of individual and firm wage components on the subsequent wage and mobility of individuals, they model these effects differently. These authors use log wage equations of the form

$$\ln w_{it} = x_{it}\beta_{tJ(i,t)} + \psi^{**}{}_{itJ(i,t)} + \varepsilon_{it}, \tag{5.1}$$

where the firm-specific component of the wage equation contains a firm-specific effect, $\psi^{*}{}_{tJ(i,t)}$, and a within-firm person effect $\psi^{**}{}_{itJ(i,t)} - \psi^{*}{}_{tJ(i,t)}$. In the first paper (1996), these authors fix $t = 1984$. They show that there is considerable heterogeneity in β (coefficients on functions of age, seniority and sex, education is not available in the data) and in $\psi^{**}{}_{i1984J(i,1984)}$. This heterogeneity is directly related to employee mobility. Higher composite firm effects, $\psi^{**}{}_{i1984J(i,1984)}$, are associated with lower mobility, a result interpreted as supporting the hypothesis that the component of $\psi^{**}{}_{i1984J(i,1984)}$ due to the average person effect within the firm, $\bar{\theta}_{J(i,1984)}$, is less important that the effect due to the firm, $\psi_{J(i,1984)}$. Steeper profiles as a function of age or seniority are associated with lower separation rates.

In the second paper (1997), these authors estimate Eq. (5.1) using samples of non-movers and movers. The resulting parameters have the interpretation

$$\psi^{**}_{ij} = \theta_i + \psi_j, \qquad \psi^{*}_{j} = \psi_j + \bar{\theta}_j \tag{5.2}$$

$$\psi^{**}_{ij} - \psi^{*}_{j} = \theta_i - \bar{\theta}_j. \tag{5.3}$$

They find that pay is persistent, by which they mean that the components of pay, ψ^{*}_{j} and $\psi^{**}_{ij} - \psi^{*}_{j}$ estimated on the observations prior to the move are significantly related to compensation on the next job. There are two possibilities for the persistence of the effect in Eq. (5.2). Either a substantial proportion of the effect is due to unmeasured human capital, an interpretation of $\bar{\theta}_j$, or there is non-random mobility, $\psi_{J(i,t)}$ is correlated with $\psi_{J(i,s)}$ when $J(i,t) \neq J(i,s)$ because of the mobility decisions of the firms and workers. There is only one explanation for the persistence of $\psi^{**}_{ij} - \psi^{*}_{j}$, as Eq. (5.3) shows, unmeasured human capital. Leonard and Van Audenrode conclude, after considering some additional mobility evidence, that the unmeasured general human capital hypothesis is the most reasonable explanation for their results.

In the third paper, with Leonard et al. (1999), they estimate a version of Eq. (5.1) with full heterogeneity in both the observable characteristics coefficients, β, and the combined person-firm effects, ψ^{**}_{ij}. They show that there is considerable persistence in these effects by examining autocorrelation matrices. Because they do not use the longitudinal nature of the data to distinguish person and firm effects, we discuss these results in Section 5.2. They also relate the heterogeneous coefficients to productivity measures in the firm, results which we discuss in Section 7.1.

Burgess et al. (1997) analyze data from the State of Maryland unemployment insurance system (see Table 1). These authors are primarily interested in studying the effect of reallocations of workers among firms on the resulting distribution of earnings. They present several models of mobility that depend upon detailed knowledge of the parameters in Eq. (3.1). They present two methods for estimating the model. In the first, they take a subset of 4000 of the workers with 10-quarter continuous employment histories. These individuals are used to select 2426 employers who ever employed these individuals. Then, they add all of the other employees of these 2426 firms to the analysis sample but only for the quarters in which these individuals worked for the 2426 employers originally selected. The procedure is equivalent to selecting a probability sample of employers with probabilities proportional to the distribution of long term employment at a point in time. The identification of the firm effects is with respect to this sampling frame, which is not representative of the same populations as the other articles discussed in this section. This sample is used to estimate a variant of Eq. (3.1) by full least squares. They also use the methods of Abowd et al. (1999a) for comparison. Regarding the basic structure of compensation, these authors report summary statistics on the correlation between firm and worker effects (small and negative) and on the correlation between successive firm effects for movers (essentially zero). The estimated firm and worker effects are used to study the

effect of worker reallocation among firms on the distribution of earnings in the State of Maryland. The individual effects (θ_i, the person effect including permanent differences in observables like education) account for 55% of the variation in log wages. Firm effects account for 35%. The small negative correlation of the firm and worker effects is associated with relatively large changes in the distribution of earnings over the period.

Bingley and Westergard-Nielsen (1996) consider the determinants of log wages using the Danish IDA (see Table 1). They adopt the specification in Eq. (3.4) but they use a random effects, say $\theta^*_i + \phi^*_j + \psi^*_{ij}$, that permits correlation between the person and firm effects but assumes that both effects are orthogonal to all observable variables. The rationale for considering this form of model stems from the method that the authors use to sample the IDA data. They construct a 5% sample of workplaces (rather than employees) and, hence, there is no observed firm-to-firm mobility. Their person effect, θ^*_i, is, therefore, defined relative to the firm in which the worker is employed, rather than relative to the employee's entire measured work history. Their firm effect is defined in a manner that includes the average person effect within the firm ($\phi^*_j = \phi_j + \overline{\theta}_j$). Finally, their interaction measures the correlation within a firm of the random person and firm effects. Thus, the authors force a hierarchical structure on the person and firm effects (see Scheffé, 1959; Searle et al., 1992) implying that their effects relegate a part of the person effect, $\theta_i - \theta^*_i - \phi^*_j - \psi^*_{ij}$, to the model residual rather than to the person effect they estimate. Keeping these statistical qualifications in mind, they find that 38% of the variance (after controlling for x_{it}) is due to the person effect, 26% is due to the firm effect and 5.8% is due to the interaction. Their commentary indicates that the person and firm effects are of approximately equal magnitude and that the correlation between the two is not strong (due to the small contribution of the interaction term) but, because of the sample design, this conclusion is not strictly comparable to the other studies in this section.

Goux and Maurin (1999) use data from the French labor force survey (see Table 1) matched with employer information to study the influence of individual and firm factors on inter-industry wage differentials. Using Eq. (3.4), these authors estimate the underlying model, identifying about 1000 firm effects and about 10,000 individual effects (over two 3-year periods), by full least squares and by a correlated random effects method. They find that person effects are more important than firm effects as components of the variance of log wages. They also find that the correlation between firm and person effects is small and negative. Goux and Maurin use the results of their statistical analysis of the components of earnings to the decomposition of inter-industry wage differentials in Eq. (3.9), these authors find that the inter-industry differences in average person effects are the main source of inter-industry wage differences in France. The part of the inter-industry wage differential explained by the firm effects is very small. There is more firm effect variation within an industry than between industries.

Finer (1997) uses the matched employer–employee NLSY '79 data (see Table 1) to estimate Eq. (3.4) directly by least squares and by a variety of other methods proposed by Abowd et al. (1999a) and Abowd and Kramarz (1999). Their full least squares results show that the person effect θ_i, and its counterpart with observable non-time-varying effects

removed, α_i, explain about 35% of the variation in log hourly wages, while the firm effect $\psi_{itJ(i,t)}$, which includes a heterogeneous seniority effect, accounts for 5% of the variation. The correlation between the two effects is -0.049.

5.2. Models with firm effects only

In this type of work, analysts estimate a variant of Eq. (3.1) in which person fixed-effects, θ, are absent. Thus, the estimated firm effect is the sum of the true firm effect, ψ, and the firm-average of the persons effects, appropriately corrected for correlation between personal characteristics and person effects. The evidence discussed in the previous section for Danish, French, and American data, suggests that the correlation between the person effects and the individual characteristics causes a large omitted variable bias and prevents a clean interpretation of the studies discussed in this section. The introduction of plant or firm effects does not help to capture a lot of the correlation between individual effects and personal characteristics because of the low correlation between person and firm effects, again, as shown in all estimated equations discussed in the subsection above.

The papers considered in this subsection use data from a variety of countries. Two of them use French data (Kramarz et al., 1996; Pelé, 1997), two use American data (Groshen, 1991a; Troske, 1999), one uses both American and French data (Abowd et al., 1998b), one uses Belgian data (Leonard et al., 1999), one uses Portuguese data (Cardoso, 1997), and one uses German data (Stephan, 1998).

The work of Groshen (1991a, in particular, and surveyed in 1996) is an important precursor to the papers that use matched employer–employee data discussed in this section. Groshen uses employer-based salary survey data to study the role of employer effects on wages. Employer-based salary surveys contain information about the participating firms. Generally, however, the only characteristic of the employer used in the statistical analysis is the identity. Estimating Eq. (3.1) with the person effects replaced by occupation gives Groshen's primary result, which is that establishment effects are a very significant component of compensation. The papers discussed in this section try to link this finding to basic characteristics of the establishment or firm.

Kramarz et al. (1996) first document the increasing inequality in France between 1986 and 1992, the dates at which the ESS was performed. A large part of this increasing dispersion is due to firm-specific compensation policies as measured by the firm effects. Indeed, the standard deviation of this firm effect increased by almost 30% between the two dates. On the other hand, the observable characteristics explain a smaller fraction of the variance in 1992 than in 1986. Furthermore, the authors compute a specialization index proposed by Kremer and Maskin (1996) to examine whether workers with the same observed characteristics are employed increasingly in the same firms. These indices grow strongly between 1986 and 1992, implying that workers are increasingly employed in firms with other similar workers. Another important feature, also found in Cardoso (1997), is the decreasing importance of returns to seniority: the wage-setting rules rely more on experience and less on seniority.

For each firm, the authors estimate the fixed firm effects. These estimates are then used in a second set of regressions that tries to explain the level and the growth of firm-effects for all firms that are present in 1986 or 1992. In all these regressions, the establishment or firm-level variables used as independent variables are the size, the existence of a firm-level collective agreement, existence of an industry-level collective bargaining agreement, the proportion of workers employed at different skill-levels, the existence and number of shifts (in level at each date for the first two regressions and in difference (1992 minus 1986) for the growth regression). Indeed, most variables matter both in 1986 and 1992, with the size of the establishment, the proportion of highly skilled workers, and the existence of 4 or 5 shifts being the most important. Interestingly, these same variables are also best (positively) correlated the growth of the fixed-effects between 1986 and 1992. Finally, in order to investigate the firm by firm compensation policies, the authors concentrate on a subsample of 132 establishments or firms for which they have a sufficient number of observations both in 1986 and 1992 and perform firm by firm wage equations for both dates. They use the estimated coefficients (on experience, measured as the experience prior to entry in the firm, and seniority) to examine how they relate one to the other as well as their correlation with mean experience and seniority at the firm. They show that the firm-specific intercept is negatively correlated to the seniority coefficient, a feature also found in Abowd et al. (1999a). They also find that the seniority coefficient is also negatively related to the mean seniority; high-seniority firms do not reward seniority very highly. Finally, the authors show that the evolution of the mean seniority at the firm (which increased by 3 years between 1986 and 1992) is negatively correlated to evolution in the mean experience (which only slightly increased) which shows that firms reduced drastically their hiring of young workers and separated mostly from workers with little seniority.

Cardoso (1997) used a very similar dataset (described above) to examine related issues. More specifically, she tried to understand the origin of the increase in wage inequality in Portugal, an increase that started between 1983 and 1986, a timing that is identical to what was observed for France (Kramarz et al., 1996). In addition, she showed that most of this increasing inequality occurred within firms rather than between firms. Therefore, she tried to identify the dimensions along which this within-firm inequality developed. First, she computed a specialization index as in Kramarz et al. but, in contrast with what was found for France, specialization decreased in Portugal between 1983 and 1992, i.e., workers with different attributes have been working more and more together. Then, the author estimate a hierarchical model of the following form:

$$y_j = X_j\beta_j + e_j,$$

where y_j is a $(n_j \times 1)$ vector with n_j being the size of the employing firm, where X_j is a $(n_j \times K)$ matrix of workers observables, β_j is the $(K \times 1)$ vector of coefficients, and e_j is the $(n_j \times 1)$ statistical residual vector which is assumed to be distributed $N(0, \sigma^2 I_{n_j})$ with I_k being the identity matrix of size k. Furthermore, each coefficient β_j is modelled as the sum of a fixed component, β_0, and a random component, α_j, normally distributed with zero mean, Γ variance matrix, and independent of e_j.

The author estimated this model both in 1983 and in 1992. She displayed the distribution of these coefficients at both dates for the following variables: experience, tenure, tenure smaller than 1 year, sex, and schooling. She also tested the equality of the two distributions (using a Kolmogorov–Smirnov test). The main features that emerge from this statistical analysis are the following. Apart from experience, all other returns changed between 1983 and 1992. In particular, the gender gap increased, returns to schooling increased very strongly, while returns to tenure decreased. Such conclusions are directly related to the process of modernization that was taking place during this period, and still is, in the Portuguese economy.

Troske (1999) examines the employer size-wage premium using the WECD (see Table 1). As in most earnings function, the author starts by estimating the following equation

$$\ln w_i = X_i\beta + Z_{j(i)}\gamma + u_i,$$

where X_i is a vector of individual i's characteristics, $Z_{j(i)}$ is a vector of the employer j of individual i's characteristics, and u_i is a residual term. Among those employer characteristics, Troske uses the logarithm of the establishment employment as well as the logarithm of the firm employment. First, he shows that all results obtained using the WECD, without employers characteristics, are identical to those obtained using the 1990 SDF (Sample Detail File, from the Census, used to construct the WECD). Then, Troske shows that the size of the employing establishment or firm generates large returns (for instance, workers in plants with log employment one standard deviation above mean log employment receive 13% higher wages than workers in plants with log employment one standard deviation below mean log employment, the equivalent number for firms is 11%). After having established these basic facts, the author tries to find potential explanations for these large returns. First, to check if these returns come from the fact that large firms hire more skilled labor than other firms, he introduces measures of skills of the work force in the above regression. More specifically, the added variables are the mean years of potential experience of workers in the plant, the percentage of workers who are scientists, engineers, or technical workers, the percentage of workers who have some post-secondary education (but no college degree), and the percentage of workers with at least a college degree. The returns to size of the establishment fall from 13% (see above) to 11%, and from 11% to 9%. Second, the author examines the capital-skill complementarity hypothesis by introducing the capital–labor ratio of the plant. Once the skills of the work force variables are in the equation, the introduction of the capital–labor ratio reduces the coefficient of the firm-size variable (yielding a 6% premium). But, the introduction of this capital–labor ratio does not reduce the establishment size-wage premium. Then, the plant age is shown to be uncorrelated to the wage, once workers' characteristics are controlled for. To assess the rent-sharing hypothesis as an explanation of the firm-size wage premium, Troske uses the proportion of the total value of a seven-digit product produced by the plant and an Herfindahl index of concentration computed at the primary five-digit product of the plant. None of these variables affect the firm-size wage premium. The same diagnostic applies to measures of the managerial skills at the plant, the proportion of supervisors at

the plant (as a measure of the cost of monitoring). Finally, the inclusion of the logarithm of total new investment in computers (in 1987) per employee adds no information as soon as the size and the labor–capital ratio are present in the regression.

Troske's conclusion is consistent with the results reported by Abowd et al. (1998b). Large employers appear to employ better workers. Most establishment or firm-level variables do not explain a large fraction of the firm-size wage premium.

Stephan (1998) uses a German dataset to examine similar questions to those we just analyzed. This dataset, the GLS (see Table 1), has matched employer–employee information for two cross-sections (1990 and 1995) of firms active in Lower Saxony, one of the largest German Länder. Each wave contains approximately 65,000 employees and 1500 firms. The sampling frame is such that for small firms all employees are included in the data while less than 10% of employees are in firms with 1000 employees or more. In addition to sex, tenure, age, contractual and effective working hours, shift or night work, and wages (with information on overtime, taxes and social security contributions) reported directly by the firm, schooling and occupation come from social insurance data, matched to this survey by the German statistical office. Indeed, the structure of the dataset is very similar to the French ESS. Stephan uses the hourly wage rate excluding overtime pay as the dependent variable. For blue-collar workers, between-plant dispersion accounts for a large fraction of the variance in wages (80% for females but less so for males) while this is the reverse for white-collar workers (60% for males and 40% for females). Then, Stephan notes that the inclusion of fixed-effects modifies the estimated coefficients in wage regressions. For firms with more than one sampled worker, the author computes establishment fixed-effects from the first-stage regressions for the above four groups of workers. Stephan finds that the dispersion of these fixed-effects is not different from those observed in other countries and that the standard deviation of these fixed-effects has increased between 1990 and 1995. Since Stephan estimates at most four effects per firm, it becomes possible to look at their correlation. He finds positive correlation between these effects. Finally, Stephan also performs firm by firm regressions and analyzes correlations between various estimated coefficients. His results show that returns to age and returns to tenure are negatively correlated. The author's results give the impression that pay determination in Germany does not differ widely from what is observed in other countries including the United States.

Leonard et al. (1999) use a Belgian dataset (see Table 1) to examine productivity in relation to firm compensation policy. To do that, they start their analysis by performing the same kind of regressions as done by Kramarz et al. (1996) and Stephan (1998)-firm by firm regressions of individual wages on observed characteristics in each year of their sample period. This results in 695 (number of firms) times 8 (years) of firm-specific estimates of a constant, age profiles, sex differentials, and white-collar/blue-collar differentials. They find that, as in all other countries, pay dispersion between and within-firms has increased over the period. By examining correlations across time of the estimated coefficients, as in Kramarz et al., they find evidence of large persistence of pay policies (see Section 5.2). These pay policies differ widely from firm to firm.

Chennouf et al. (1997) estimate the same type of equation using matched worker-firm data for a sample of establishments of the Algiers region (see Table 1). As in Cardoso (1997), the authors try to control for the group effects when estimating Mincer's model that includes both years of education and potential experience. Their results show that returns to education decrease when firm-effects are introduced.

Abowd et al. (1998b) compare the relative importance of employer and employee effects in compensation in France and in the United-States. For the US, they use the 1990 WECD while, for France, they use the ESS for 1986 and 1992. The basic statistical model used throughout the paper is identical to the one described in the statistical section in which the wage is regressed on worker's characteristics and a firm (or establishment) specific constant as follows:

$$\ln w_{it} = X_{it}\beta + \phi_{j(i,t)} + \varepsilon_{it}.$$

Then, the authors use the estimates to decompose the average wage at the firm into a part due to the average observed characteristics of the workers employed at the firm and a part due to the firm-effects, $\phi_{j(i,t)}$, to analyze the impact of compensation structure on firm's productivity and profitability.

The wage equations give the following results. First, coefficients are not very dissimilar across countries. A first noticeable difference is the shape of returns to experience which are steeper and never turn down in the US whereas the French profile peaks at 34 years of potential experience. A second interesting difference are the respective R^2 which are larger in France (around 0.80) than in the US (around 0.60). Therefore, firm fixed-effects obviously explain a larger fraction of wages in France than in the US.

Then, the authors present a table of correlation among the components of individual compensation. Strikingly, the correlation structure is very similar in the two countries. In particular, individual characteristics and firm fixed-effects are comparable in terms of their contribution to the variation of annual wages (approximately from 0.6 to 0.7) with an inter-correlation of 0.25 in the two countries. None of the above results are inconsistent with those of Abowd et al. (1999a,c) since the firm fixed-effect as estimated by Abowd et al. (1998b) are a mixture of individual and firm fixed-effects (see Eqs. (3.4) and (3.7)).

Finally, the authors estimate the impact of the compensation structure on firm's productivity and profits. They show that firms who employ workers with higher predicted wage rates (based on observed characteristics) are more productive (both in terms of log value-added per employee and of log sales per employee) in the two countries. The same is true for firms with higher fixed-effects. However, none of these two components have an impact on profitability.

This study confirms the above findings: the structure of compensation is very similar across countries. In addition, the effects of the compensation structure on firm outcomes appear roughly identical in two apparently different countries, France and the US.

Pelé (1997) also uses the French ESS to examine the effect on compensation of the coexistence of different methods of pay in the same firm for the same detailed occupation. This type of dataset, by matching firm and workers, enables him to compare within a firm

and within an occupation (in a 4-digit classification) workers paid under time-rates or piece-rate (measured under different rules). There has been a great body of literature dealing with the choice of method of pay and the wage differentials due to various ways of payment. For example, Seiler (1984) showed that incentive workers receive more disperse and higher earnings than time workers. Brown (1992) found that piece-rate workers receive higher earnings than time-rate workers but when compensation is linked to merit (an evaluation by supervisors), it is lower than time-rate.

The data come from the Wage Structure Survey (Enquête Structure des Salaires) carried out by INSEE in 1986 (see above). For each worker, the age, seniority in the firm, sex, method of pay, conditions of work (especially work in shifts), occupation are used as control variables in the wage equation. Besides, for each firm or establishment, a the identification number is used to identify workers employed at the same establishment or firm. Four methods of pay are possible. The first one is time-rate, which consists in a salary. The three other ones are bonus payments. The bonus is based either on individual output, or on collective output or on both kinds of output. A worker can receive a bonus of exclusively one of those three types. Pelé used the total wage in October 1986, the amount and the type of the bonus (if it is the case) and the payment for overtime. He corrected for differences in hours worked in order to compare wages for a same duration (which is equivalent to use an hourly wage). To estimate the wage equation, he added indicator variables for each bonus method. To control for the exact occupation in the firm, Pelé introduced an indicator for each 4-digit occupation within each firm or establishment. Only those couples (occupation-establishment) with two workers employed in the same occupation within the same establishment under different methods of pay contribute to the identification of the coefficients of the bonus methods variables. Therefore, the occupation-establishment fixed-effects are nested within the pure establishment fixed-effects.

The results are the following. First, he found that bonus payments lead to higher compensation, result which is consistent with a selection of workers among the different payment schemes. It is profitable to give incentives through a bonus payment only to the best workers. But beyond this first conclusion, Pelé also showed that workers who get a bonus also receive a higher base wage, when comparing within homogeneous groups of workers of the same occupation in the same firm. High French minimum wages may partly explain such an observation, by preventing firms to set a low value to the base wage.

Other recent papers that include employer effects in wage equations and use matched data to study the resulting estimates include Bronars et al. (1999) and Vainiomäki (1999).

5.3. Models of the wage-seniority relation

How large are returns to seniority? This question has generated many important articles in the last 20 years. Some authors argue that returns to seniority are large and pervasive (on the order of 5% a year) while others find these returns to be small.

Although Topel and Ward's data were based on a matched employee–employer dataset,

the LEED file constructed from the Social Security reports by employers, these authors did not use this aspect of their file. Indeed, they restricted their initial sample of over one million individuals to a final sample of 872 persons. Therefore, they lost the potential of looking at the inter-firm variability in the returns to seniority.

The possibility that returns to seniority might vary between firms has only been examined recently. Abowd et al. (1999a) allow for such variation in an exogenous mobility framework using French data. Margolis (1996) reexamines Topel's two stage estimator on the same French dataset. At least five other papers examine this issue of firm-specific compensation policies, two on US data (Bronars and Famulari, 1997; Finer, 1997), two on Norwegian data (Barth, 1997; Barth and Dale-Olsen, 1999), one on Italian data (Pacelli, 1997), and one on Portuguese data (Cardoso, 1997). Another paper (Dustmann and Meghir, 1997), based on German data, allows for the possibility of firm-specific returns to tenure even though the authors do not introduce firm fixed effects.

Abowd et al. (1999a) provide different estimation techniques for firm-specific returns to seniority. These different estimation methods (see above for brief description) provide estimates that differ in their levels but which are largely correlated across methods.

The consistent methodology, which uses first differences for workers who do not move between two consecutive years, gives the largest estimates and, indeed, the closest to the OLS results. Notice however that this methodology assumes that mobility is exogenous. All other techniques examined in this paper give returns to seniority that are close to zero. But all methodologies show that there is considerable between-firm variation in these returns. And, furthermore, all of these estimated firm-specific returns are strongly correlated across estimation techniques. Unfortunately, these authors do not examine the same question when mobility is endogenous.

Margolis (1996) tries to address this issue by allowing firm-specific compensation policies to vary by entry-cohort (the cohort of entry refers to entry in the firm). The data he uses are identical to Abowd et al. (1999a). Margolis compares OLS estimates other techniques. First, he examines on French data the results using Topel's two-stage techniques. He shows that based on French data, the returns to tenure are much lower than those estimated by Topel, 2% against 5% using US data. But, Margolis also notes that unobserved heterogeneity may well bias these results. Hence, he goes one step ahead of Abowd et al. (1999a) by introducing within-firm cohort-effects. Although the value of the mean value of the estimated seniority slopes is close to zero, Margolis (1996) finds even more variance in the returns to seniority than previously found in Abowd et al. (1999b).

Bronars and Famulari (1997) examine similar questions use a US dataset, the WCP (see Table 1), matching roughly 1700 workers and 241 firms. In addition, for 736 workers employed in 130 establishments, retrospective information on the starting pay is available. That allows the authors, first, to estimate returns to seniority based on a cross-section equation that includes firm or establishment fixed-effects. Then, they look at within-firm wage growth with, once more, establishment fixed-effects. Hence, the first estimates are directly comparable to the OLS with firm fixed-effects given in Abowd et al. (1999a) while the second are also directly comparable to the consistent estimates from the same authors.

At least for men, the authors find that estimated returns to tenure are roughly equal to 1% a year and approximately invariant across estimation techniques. This result does not hold for women. In addition, women's wage growth is larger than man's wage growth. This result, consistent with Abowd et al.'s (1999a) findings, demonstrates that exogenous mobility is not a very reasonable assumption and that all these estimated returns to seniority are likely to be biased upwards. Finally, Bronars and Famulari find important variation across firms in returns to tenure; standard deviation of the estimated firm-specific slopes is equal to 0.022, consistent with the French estimates giving a larger number based on a much greater number of firms.

Barth (1997) estimates related coefficients for Norway. He uses a cross-section of 2321 workers employed in 549 firms. His base equation is identical to the one estimated by Bronars and Famulari (1997). As found by Abowd et al. as well as Bronars and Famulari, the estimated coefficients are identical across estimation techniques, i.e., OLS, firm random effects, and firm fixed-effects. One additional year of seniority adds approximately 0.3–0.4% to the individual's wage, a much lower number than the one found in the US or in France. Interestingly, Barth finds no evidence of correlation between seniority and firm fixed-effects while there is a positive correlation between education or age and these fixed-effects. These final results are consistent with those given in Kramarz et al. (1996) (see Section 5.2). Barth and Dale-Olsen (1999) examine the relation between wage-seniority profiles and worker turnover using a heterogeneous firm effect model as in Eq. (3.2). They show that employee lower turnover is associated with having higher initial wages (ϕ_j) and higher slopes (γ_j).

Cardoso (1997) examines identical issues with the same type of dataset, i.e., cross-sections of matched employer–employee data (see Section 5.2 for other results using the same dataset). Using a multilevel model with the associated estimation method, the author confirms the heterogeneity of the returns to seniority across firms. Most estimated returns are inferior to 1%. She also estimates the distribution of firm-specific starting wages (hence, the firm-specific component of wage is either this latter part for the entrants or the former part for those with 1 year or more seniority), negative indeed for most of the distribution. Notice once more that these results are not widely different from those estimated in all other European countries.

The technique used in Dustmann and Meghir (1997) for Germany is completely different. Even though their data is based on Social Security reports of firms, these authors do not have full access to the matched employer–employee component of their data source. But, the principle of their technique – instrumental variables – is directly applicable and conceived for matched employer–employee data.[8] Their instruments are firm closure and information on the job held two jobs ago (available only for workers with at least three jobs in their dataset). Indeed, allowing for heterogeneous returns across firms yield surprisingly high estimates (which jump from 4.5% to 9%), casting some doubt on these exact

[8] That they should eventually obtain.

values based on approximate data, for instance sector-specific firm closures instead of firm-specific firm closure, on a selected sample of young apprentices, but showing the fruitfulness of the general approach.

Finer (1997) uses the same models as Abowd et al. (1999a). He finds that the average return to 1 year of additional seniority is 5% for the first 5 years and zero thereafter in the NLSY '79 data. The standard deviation of the estimated seniority coefficient is of the same order of magnitude, which indicates that in this younger sample there is still considerable heterogeneity in the return to seniority.

Pacelli (1997) examines identical issues using a longitudinal sample of 1737 young Italian workers under a period of 5 years with information on the employing firm constructed from the R&P dataset (see Table 1). The methodology adopted resembles Topel and Ward's. All estimates show that returns to seniority are, once more, smaller in Europe than in the US, even when controlling for firm-specific variables in a wage growth equation (see also Entorf and Kramarz, 1999, for similar results), but still significant.

6. New results with matched employer–employee datasets: wage and employment mobility

In this section we consider models of the changes in wage rates and employment that have been estimated using matched employer–employee data. All of the papers make use of large longitudinal administrative data sources that are dynamically representative of the target populations.

Jacobson et al. (1993) study the earnings losses of displaced workers in the State of Pennsylvania. A 5% random sample of the quarterly earnings reports from the State's unemployment insurance tax records for the period from 1974 to 1986 were matched to employer data from the State's ES-202 files, which are also administrative data from the unemployment insurance system. These authors define large involuntary worker displacements using the matched data. In particular, using the matched information about the employer, these authors are able to define a mass-layoff displacement, where there is a large reduction in the employment of the firm surrounding the displacement, and a non-mass-layoff displacement. They find that the earnings losses for mass-layoff displacements are very large: initially 25% of predisplacement earnings, rapid recovery during the first two post-displacement years to losses of around 15% of predisplacement earnings, followed by many years of stable earnings with no further recovery. The non-mass-layoff displacement sample has an initial loss of about 15% followed by a rapid recovery during the 2 years following displacement in which the full earnings loss is recovered. For all of the comparisons it is possible to use the earnings histories of workers who do not suffer displacements as the comparison group, including the possibility of comparing workers who were not displaced during the mass layoff, but who worked for the firm that incurred the mass layoff, with those who were displaced. Using this information, these authors find that individuals who are going to suffer a mass layoff also experience an earnings decline

in the 3 years prior to the mass-layoff displacement. Workers who are going to be displaced in a non-mass-layoff displacements do not experience a predisplacement earnings decline.

Topel and Ward (1992) use the quarterly LEED data for 1957 to 1972 (see Table 1) to study the early career wage and employment mobility of young male workers. Although these authors use the employer identifying information only to identify the within-job wage growth, they are among the very few researchers to have used the LEED data in this manner. They find that the typical young male worker holds seven full-time job during his first 10 years in the labor force. Within job wage growth is one-third of total wage growth over this period and between job wage growth accounts for the remainder. Wages grow at an annual rate of 11%. Although the authors do not directly implement Eq. (3.1), their basic model is consistent with this formulation and their method for identifying within and between job wage growth.

Burgess et al. (1997, 1999) examine earnings dispersion using a decomposition similar to those described previously. They first decompose wages into a person fixed-effect, a firm fixed effect, a time trend, an unemployment effect, and a residual. Then, they examine the share of earnings dispersion that can be attributed to the different components. In particular, they focus on the share attributed to individual fixed-effects, the share attributed to establishment fixed-effects, the share due to the correlation between individual and firm fixed-effects, and, finally, the residual unexplained variance. Using a sample and a technique described above, they estimate individual and establishment fixed-effects for 2000 individuals and their 1432 employers based on a dataset with more than 2,700,000 quarterly observations. Their estimates show that 55% of the variance in log wages can be attributed to individual heterogeneity while firm-effects account for 35% and the correlation between person and firm effects is virtually nil. Then, they try to assess the share in the increased earnings dispersion over the 1980s that can be attributed to reshuffling of jobs between and within employers. Their first estimates seem to support the idea that, indeed, reallocation of jobs across firms was an important source of increase in the dispersion of wages under the sample period.

Abowd et al. (1998a) examine job and wage mobility in France and in the US using comparable matched employee–employer panel datasets for both countries. Most of their analysis focuses on employment durations and wage changes both between and within firms. The employment spells can be constructed because of the matched nature of the data; the employer identifier is a crucial component when constructing the individual careers across time and firms. Even though the analysis is still preliminary, the authors findings show that, in contrast with the usual view that the French labor market is inflexible i.e., little employment mobility and considerable real wage stability, in France there is substantial employment mobility, although the most mobile groups in France are not the same as those in the United States, and there is substantial real wage mobility on changes of employers.

Margolis (1998) uses several sources of matched employer–employee data (the DADS, BIC and BRN) to consider the effect of firm closure on workers in France. He find that a

large share (almost 60%) of workers displaced by firm closure find new jobs without experiencing any interruption in their employment histories. In addition, falling into nonemployment appears to be a relatively transitory phenomenon for displaced workers, with over three-quarters finding a new job within the year following displacement and essentially all of them being reemployed 6 years after displacement. Workers who separate for reasons other than firm closure, on the other hand, have a much harder time, with 25% still without a job 6 years after separation. Wage changes for displaced workers in France reflect a major difference between those who find new jobs quickly and those who do not, with a wage penalty of over 20 percentage points for displaced workers who do not find new jobs in the year following their separation. The pre-separation pattern of wages shows a drop in the year preceding separation. Controlling for seniority differences causes wages for displaced workers to be consistently below those of continuously employed workers, even in the pre-separation period, and the penalty for finding a job slowly drops to under 10%.

7. New results with matched employer–employee datasets: firm outcomes and worker characteristics

7.1. Productivity

In this subsection we consider studies that relate individual characteristics of the employees and of their compensation to the productivity of the enterprise or establishment. The employer-level measures of productivity come either from direct measures of production or value-added per worker or from full production function specifications.

Hellerstein et al. (1996) use the WECD (see Table 1) to study the relative productivity of employee characteristics, estimated directly from a production function, that they compare to the relative pay earned by these different characteristics, estimated directly from a wage equation. They use a variety of production function specifications to capture the marginal productivity associated with employee characteristics like sex, race, marital status, age, and education. They use standard cross-sectional wage equations (estimated for the US census year 1990, which is the only year of individual data available in the WECD) to capture the market compensation associated with these factors. The find that workers who have ever been married are paid more than never-married workers and that there is a corresponding productivity difference of the same magnitude. On the other hand, prime-age workers (35–54) are as productive as younger workers but earn a wage premium. Wage premia for older workers (55–64) exceed all estimated productivity premia for this group. The same technique is used to conclude that wage differentials unfavorable to blacks are also associated with productivity differentials of the same magnitude. Wage differentials favoring men are not associated with productivity differentials of the same magnitude. Bayard et al. (1999) extend the analysis of Hellerstein et al.

(1996, 1997) to include non-manufacturing establishments using the NWECD (see Table 1). Their results are discussed below under sex segregation in the workplace.

Haegeland and Klette (1999) estimate similar productivity and wage models using Norwegian data. They find that education, except those with the lowest education, premia are directly and appropriately related to productivity differentials. Workers with highest experience have wage premia that exceed their productivity while the opposite is true for those in lower experience categories. The lower wages of females correspond to productivity differences of equal size.

Hayami and Abe (1998) use the Japanese matched data to estimate a full set of labor demand equations for age and sex categories for retail trade to study the deregulation of this labor market and the resulting effects on wages, employment and productivity.

Abowd et al. (1999a) develop a method for relating the firm effect, ψ_j, and the average value of the person effects (θ_j and its components $\bar{\alpha}_j$ and $\bar{u}_j\eta$) to firm-level productivity, profitability and factor use. The profitability results are discussed in Section 7.1. To measure productivity, they use firm-level value-added and sales per worker, averaged over the period 1978–1988. They find that all firm-level compensation components related to personal and enterprise heterogeneity are positively related to both productivity measures.

Abowd et al. (1998b), also take the firm effects from their analysis of compensation (see Section 5.2) and relate these to value-added and sales per worker for both the US and France firm effects inclusive of the average value of person effects within the firm, $\psi_j + \bar{\alpha}_j$, are positively related to the productivity measures.

Finer (1997) estimates an equation relating productivity, measured as ln(sales/employee), to the components of compensation structure. He finds that all compensation structure components are strongly related to sales/employee.

Using the estimated coefficients of firm by firm regressions across time, Leonard et al. (1999) examine the relation between firm-specific compensation policies and productivity (measured by value-added per worker). Since they have multiple estimates of the same coefficient for the same firm but different years, these authors are able to estimate this relation with fixed firm effects. They find that firms with high wage levels (i.e., the firm-specific constant of the firm by firm regression), returns to age, white-collar pay premium, or male pay premium also have high productivity.

7.2. Productivity and seniority

Kramarz and Roux (1998) use the DADS for the period 1976–1995 to examine the relationship between within-firm seniority structure and firm performance. Hence, these authors provide one of the first analysis of the impact of hiring and separations decisions on firm-specific outcomes such as productivity or profitability as well as employment or capital structures.

They first measure the seniority at the end of all job spells (either censored or non-censored). Then, these seniorities are aggregated at the firm-level for three subperiods

(1977–1982, 1983–1988, 1989–1994) in order to compute firm-specific descriptive statistics of the seniority structure. Using the Echantillon d'Entreprises (see Table 1) that gives information on balance-sheet, skill structure, and employment, Kramarz and Roux estimate various equations relating tenure structure and firm performance. The use of three subperiods allows these authors to perform instrumental variable techniques, in particular, they estimate their coefficients with equations in first difference (subperiod 3 minus subperiod 2) instrumented by the levels of subperiod 1. Their results show that a low turnover rate is associated with higher productivity but a high turnover rate slightly favors profitability. In addition, an increase in the within-firm variance of the seniority of the stayers (i.e., those workers who stay employed at the firm until the end of the subperiod, and, therefore, have censored spells) boosts profitability. Finally, capital, firm size, and the capital–labor ratio are all positively related to low turnover rates.

7.3. Profits

Four papers consider the relation between profits and firm-specific compensation or hiring and separation policies using matched employer–employee data. In all cases, the profit measure is operating income divided by total assets. Two papers focus on France (Kramarz and Roux, 1998; Abowd et al., 1999a), one uses American data (Finer, 1997), and one compares the French and the American case (Abowd et al., 1998b). The equation relating firm level profits, π_j, to firm level compensation policies, ψ_j and $\bar{\theta}_j$, and other firm level measures, z_j, is given by

$$\pi_j = z_j\gamma + \alpha\psi_j + \beta\bar{\theta}_j + \varepsilon_j. \tag{7.1}$$

This equation is estimated directly in Abowd et al. (1999a) and Finer (1997) while Abowd et al. (1998b) cannot separate person from firm-effects since they use cross-sections (WECD and ESS for the US and France, respectively). Kramarz and Roux replace ψ_j and θ_j by the within-firm seniority structure (see Section 5.3).

Abowd et al. find that those firms with higher wages because of observed characteristics, $\bar{x}_j\beta$ (in z_j), or with a larger firm effect, ψ_j, (thus, high-wage firms) are more profitable, while those employing high-wage workers (large values of $\bar{\theta}_j$) are not. Finer finds no effects on profits, although his data are from a sample representative of younger workers. Interestingly, Abowd et al. (1998b) find no effect for the same equation as soon as ψ_j and θ_j are confounded, a result which is fully consistent with the previous one. Finally, Kramarz and Roux find that a higher turnover rate as well as a larger variance of within-firm seniority tend to induce a larger profitability.

7.4. New technologies

We know that changes in the structure of wages were dramatic along the 1980s in the US while unemployment increased in Western Europe. Many analysts have blamed the same technical shocks that affected differentially the two continents. The role of computers has been central in the indictment, in particular after Krueger's seminal work (Krueger, 1993)

that showed that computer users were better paid than non-users. On both sides of the Atlantic, researchers have tried to understand the nature of the computer wage premium. Two sets of studies, one for the US and one for France, are of particular interest for us since they use matched employee–employer datasets. They both demonstrate that new technology (NT) workers were better paid than non-users even before using NT (Entorf and Kramarz, 1997, 1999; Entorf et al., 1999) or that NT firms employed high-wage workers even before implementing NT (Doms et al., 1997).

Doms et al. (1997) use the WECD (see Table 1) in conjunction with the 1988 Survey of Manufacturing Technologies (SMT) among many data sources. The 1988 SMT contains plant-level information on NT use in American manufacturing plants. The techniques surveyed are production technologies such as robots, computer-aided design (CAD), lasers, networks, automatic systems, or computers used on the factory floor. To assess the technical development of the plant, the authors use the count of different techniques used at the plant. The SMT is matched to the WECD. This allows Doms et al. to examine the relation between the spread of techniques and education or the occupational mix of the work force. To perform this analysis, since they do not know if individual workers use a given technique, they create various plant-level measures of the educational or occupational structures. Results demonstrate that plants that use more advanced technologies employ a more educated or a more skilled work force. Using the same framework, they examine the relation between wages, once more averaged at the plant-level, and NT. The analysis is performed for different subgroups (production workers, managers and professionals, other non-production workers) and include average characteristics of the group under consideration together with plant-level employment and capital–labor ratio. These results show that, as in Krueger (1993), technology use is associated with a premium even after inclusion of workers characteristics. Then, using the LRD and the 1993 SMT, a longitudinal analysis demonstrates that the most technology advanced plants paid their workers higher wages prior to adoption of NT.

The same pattern emerges from the three studies performed on French matched employee–employer data. However, since the datasets used in these studies are built from the supplements on NT of the1987 and the 1993 waves of the French Labor Force Survey (LFS) in which workers can be followed at most three times (from 1985 to 1987 and from 1991 to 1993, respectively), it is possible to perform an individual-level longitudinal analysis while controlling for the employing firm.

The data used in Entorf and Kramarz (1997, 1999) come from four different INSEE sources. The basic sources are the French LFS, 1985–1987, a 3 year rotating panel, and the "Enquête sur la Technique et l'Organisation du Travail auprès des Travailleurs Occupés" (TOTTO) from 1987, an appendix to the labor force survey that asked questions about the diffusion of new technologies and the organization of the work place. Besides usual questions from labor force surveys (salary, tenure, age, education, etc.) the appendix contains information on the use (e.g., intensity, experience) of microcomputers, terminals, text processing, robots and other well specified groups of "New Technology" labor. The use of computers is described in more detail than in other surveys (see Krueger, 1993, for

instance). The questionnaire provides explicit categories for using microcomputer for text processing only, data entry and use of listings. "Terminal" even covers a distinction between "reception only," "emission only" and both reception and emission while information on production techniques are also present.

In the first version of the TOTTO survey, only the 1987 employing firm is known (using the standardized Siren enterprise identification number). This feature of the French INSEE classification system enables the authors to employ information from corresponding firm-level surveys (BIC, which collects annual information on balance sheets and employment and ESE, which collects information on the employment structure).

In the cross-section, the approach is identical to Krueger's (1993). Entorf and Kramarz regressed the log of monthly wage on a vector of characteristics of the individual X_i and a vector of indicator variables for workers using one (or more) of the various NT groups. These variables were supplemented with firm-level characteristics $Z_{j(i)}$ (where $j(i)$ denotes the firm at which i is employed), some of which are available from the complement to the labor force survey (working time schedules, sector, size) and the others from the firm-level panel dataset (size, assets, profits, skill structure, export ratio). In all regressions, they control for the usual observable variables. Their results show that, in 1987, a worker receives a 16% bonus for using modern computer-related NT. This premium can be decomposed into two parts: for a worker with no NT-experience, a NT worker receives a premium of approximately 6%. Returns from experience with NT add 10% to the above premium (when estimated at the average level of experience in the population of modern computer-related NT users). When firm-level variables are introduced, some of the above results seem to be attenuated: the coefficient of the modern computer-related NT dummy is smaller (5%) and the standard error is larger. However, the role of experience with modern computer-related NT is increased. The firm-level variables that are used, even if they do not seem to be correlated with the individual NT variables, are important and increase significantly the explanatory power of the regression. Most important is the skill structure : the more skilled the structure is, in terms of larger shares of skilled workers and of managers, professionals and technicians, the larger is the influence on the wage. This effect is particularly strong for the latter category: a 1% increase in this share entails a 0.27% increase in the individual wage. The profits (profits/assets) also have a positive impact on wages. Finally, total employment has no significant influence on earnings. Finally, if firm fixed-effects are introduced, results are unchanged.

In the longitudinal dimension, all the above effects of NT almost completely disappear. The coefficients on the NT indicator variables are never significantly different from zero. However, even though NT use per se does not yield an immediate wage gain, coefficients of the experience with modern computer-related NT variables are significantly different from zero. In Entorf and Kramarz (1997), another version of the same equation in which a dummy for each year of experience (1,2,...,9 and more) is included is estimated and results are quite similar: returns increase until workers have 5–6 years of experience and then slightly decrease. The introduction of the firm-level variables do not change these results.[9]

In addition, these firm-level variables that represent the firm-specific policy have little impact on the individual wage once individual fixed effects are introduced. Coefficients are either not significantly different from zero or small (assets).

Most of the results that we have described for the 1985–1987 period also hold between 1991 and 1993. Most datasets are identical. A new feature of the LFS is the inclusion of the employing firm identifier in every year while only the 1987 employing firm was known previously (see above). In addition, the authors use a newly available dataset, the "Déclarations de Mouvements de Main d'Oeuvre (DMMO)," an establishment-based survey on hiring and separations. Entorf et al. are therefore able to follow the workers across firms in the 3 years of the panel.

Entorf et al. (1999) estimate wage equations with NT indicator variables without and with individual fixed-effects as well as without or with firm fixed-effects. Returns to computer use in 1993 are not different from those observed in 1987. The introduction of individual fixed-effects has the same impact as obtained in Entorf and Kramarz (1997, 1999). Returns are maximal, 2%, after 2 or 3 years experience with NT. The introduction of firm fixed-effects has no impact on the estimated coefficients, both in the cross-section dimension and in the longitudinal dimension.[10] This is consistent with the Abowd et al. (1999a) findings for France as well as those of Abowd et al. (1999c) for the US-firm compensation policies (as captured by the firm fixed-effects) are not highly correlated with individual observables and individual fixed-effects. To test other explanations of the results (in particular, to control for firm-level idiosyncratic shocks), the authors use the matched worker–establishment information on hiring, quits, and terminations coming from the DMMO. Results are identical to those described above. Finally, Entorf et al. use the quarterly LFS where workers are followed for three quarters after the TOTTO survey to test whether NT workers are protected from unemployment. Indeed, they find that in the short-run, NT users are protected from job losses. This result is stable, even when using the DMMO information on quits and terminations to measure the business conditions at the firm-level.

7.5. Creation and destruction of jobs

In this part, we do not intend to describe the whole "creation-destruction" vision of the labor market. Davis and Haltiwanger's chapter in this Handbook is fully devoted to this task. In this subsection we concentrate on the new types of results that matched worker-firm data have helped to bring to researchers' attention. Many of the papers that are discussed below use the basic definitions and analysis techniques that initiated by Leonard

[9] Since only the 1987 employing firm is known, the 1985 and 1986 firm is unknown for workers who changed firm at one of these dates. Entorf and Kramarz (1999) use the 1987 firm also for the movers.

[10] As indicated in Abowd et al. (1999a), in the longitudinal dimension, firm fixed-effects can only be separately identified from worker fixed-effects when at least one worker in the firm quits for another firm in the sample. Here, the authors are able to identify 494 of the 1045 firm dummies.

(1987) and Davis and Haltiwanger (1992, 1996), neither of whom used matched employee–employer data.

The analysis of worker flows, in contrast to the study of job flows, has been made possible by the use of matched worker-firm data. The researchers who started this vein, which is flourishing now, were Anderson and Meyer (1994). Using the CWBH dataset for the years 1978–1984 (see Table 1), they compare worker turnover defined as the sum of total accessions – recalls plus new hires – and total separations – temporary layoffs plus permanent separations – to job creation and job destruction measures as promoted by Davis and Haltiwanger. In addition, they use firm-level measures computed from their individual-level data in relation with their measures of job turnover. They compute firm-size, quarterly payroll per worker, and tenure at the firm to create categories such as high- or low-paying firm, high- or low-tenure firm. Then, they present a tabulation of job creation, destruction, and turnover statistics for every of the above firm-level categories (Table 2, p. 191).[11] For instance, Anderson and Meyer (1994) show that high and low-tenure firms do not differ in their temporary separation rates but widely differ in their permanent separation rates. Indeed, the same pattern is exhibited for high and low-paying firms. The same type of analysis is pursued on the number of earnings weeks lost after separations followed by reemployment (Table 9, p. 212). They are able to show that the distribution of weeks lost is extremely skewed (the mean is roughly 13 weeks as the median is equal to 2 for total separations). In addition, they show that mean weeks lost after a temporary separation are a decreasing function of firm size while mean weeks lost after a permanent separation are an increasing function of firm size. Similar computations are provided for high- and low-paying firms. These statistics being computed from individual-level data, the authors also regress the above separations variables onto the (time-varying) firm-level variables and individual fixed-effects taking advantage of the structure of the dataset in which workers can be followed from firm to firm (Table 10, p. 214).

The main disadvantage of the CWBH dataset lies in the absence of individual characteristics of the employed workers. Even though the states may have collected such information for the beneficiaries of unemployment insurance, these complementary datasets are inaccessible to the researchers. Of course, for each individual, it is always possible to compute a date of appearance and a firm-specific tenure, which is left-censored for all observations in the first-quarter of 1978. But no information on age, sex, education is used.

The same problem affects the recent analyses of Lane et al. (1997b) and Burgess et al. (1999). These articles have mostly focused on churning, the hires and separations in excess of total job reallocation using the Maryland quarterly employment and earnings information from the unemployment insurance dataset (see above). The period of analysis, 1985–1994, is the only difference between the data used in these papers and those used in Anderson and Meyer (1994). Lane et al. (1997b) provide an description of hiring and exit flows. The individual data on the characteristics of the movers have been aggregated

[11] Hamermesh et al. (1996) also document the importance of worker flows as compared to job creation and destruction for data from the Netherlands, although they do not use matched employer–employee data.

to the establishment-level and used as explanatory variables in the churning regression (Eq. (1) in their paper). The longitudinal component of the dataset allows the authors to include firm fixed-effects in this regression. One striking result, also found in Abowd et al. (1999b) for France, is that most of the changes in employment are accommodated through changes in the hiring rate.

Abowd et al. (1999b), who use an administrative dataset of all entries and exits in French establishments (see Table 1), perform most of their flow analysis at the establishment-level. Their empirical analyses distinguished between flows of workers, directly measured, and job creation and destruction, again, directly measured, using a representative sample of all French establishments for 1987–1990 (with more than 50 employees). The most important findings were that (a) annual job creation can be characterized as hiring three persons and separating two for each job created in a given year; (b) annual job destruction can be characterized as hiring one person and separating two for each job destroyed in a given year; (c) when an establishment is changing employment, the adjustment is made primarily by reducing entry and not by changing the separation rates; (d) for the highest skill groups, 10% of months with firm-initiated exits also have new hiring in the same skill group and, for the lowest skill groups, 25% of the months with firm-initiated separations also have new hiring in that skill group; (e) the rate of internal promotion into higher skilled positions is about three times the size of net employment changes inside the job category; (f) two-thirds of all hiring is on short term contracts and more than half of all separations are due to the end of these short term contracts; (g) approximately one-third of all short term employment contracts are converted to longterm contracts at their termination; (h) controlling for between-establishment heterogeneity and common trends, entry and exit of workers are both countercyclical.

Other studies that use matched employer–employee data to analyze these issues of job creation and job destruction are Norwegian (Salvanes and Forre, 1997), Austrian (Winter-Ebmer and Zweimüller, 1997), Danish (Belzil, 1997; Bingley and Westergård-Nielsen, 1998, Vejrup-Hansen, 1998; Albaek and Sørensen, 1999), Swedish (Persson, 1998), and Finnish (Laaksonen et al., 1998).

Albaek and Sørensen (1999) examine the relation between worker flows and job flows using the Danish IDA (see Table 1). In that respect, the type of analysis they perform is close to Hamermesh et al. (1996) and even closer to Abowd et al. (1998). These authors find that annual rates of hires and separations are much higher than the job creation or job destruction rates–28% and 12% respectively for Danish manufacturing. They also find that separations from existing jobs are dominated by quits. Another issue studied at length by these authors is cyclicality of the flows. They show that worker flows are strongly asymmetric over the business cycle.

Bingley and Westergård-Nielsen (1998) use the Danish IDA to show that the Danish labor market is dynamic and flexible. Among growing establishments two hires and one separation are required for each net job creation. Among shrinking establishments they find that one hire and two separations are required for each net job destroyed. Vejrup-

Hansen (1998) finds that workers separated from establishments with job destructions have unemployment incidence that is comparable to the general Danish population.

Salvanes and Forre (1997) use the individual information on the education-level of the employed workers from their registers (see above) to examine creation, destruction, entry, exit, and churning for three groups of education. They find an asymmetric and inverse U-shaped churning curve (the churning rate, i.e., the entries and exits in excess of job creation or job destruction, is larger for medium-education workers).

Winter-Ebmer and Zweimüller (1997) examine the relationship between firm-level measures of earnings dispersion and employment growth. To examine this issue they use a firm-based random sample of the Social Security files for the period 1975–1991 (see Table 1). With the resulting 130 firms, for which they have all employed workers (with their earnings – top-coded for 9% of them – and most other individual characteristics but education), they compute within firm measures of earnings dispersion as follows. Because of top-coding, they run a Tobit regression for each year and each firm. The resulting standard error of the residual of this regression is used as a first measure of dispersion due to all unobserved factors. They also perform the same regression using only male workers (they do not have hours worked, hence working with males reduces the part-time probability). Then, they use these variables in their firm-level employment growth regressions (1236 observations). These regressions are estimated without and with firm fixed-effects. While there is some evidence with OLS that an increased earnings variance reduces employment growth, the introduction of fixed-effects wipes out any such effects.

Interestingly, Belzil (1997) and Burgess et al. (1997) do the same type of analysis, but in the reverse direction. They both use measures of employment growth (creation, destruction, reallocation, or churning) as additional regressors for explaining wage structure of wage changes.[12] Belzil (1997) use a subsample of the IDA dataset to perform individual-level wage regressions. Since the dataset is longitudinal and contains both employee and employer identifier, it is possible to control for person fixed-effects as well as firm fixed-effects. Even though none of the reported regressions include firm fixed-effects, the author states that the introduction of these effects does not affect the coefficients of interest, a feature consistently found in France (Abowd et al., 1999a; Entorf et al., 1999), in the US (Abowd et al., 1999c), or in Denmark (Bingley and Westergard-Nielsen, 1996). Belzil also finds that employment creation, destruction, or reallocation affects wages, even though no systematic pattern seems to emerge across the different subsamples that he analyzes.

DiPrete et al. (1998) use matched employee–employer longitudinal data from France (LFS matched with firm-level information using the SIREN number, see Table 1) and Sweden (LFS matched with establishment registers, see Table 1) to examine the relation between the dynamics of employment of the employing establishment and job mobility. They model simultaneously unemployment, exit from an establishment, job mobility within an establishment, and entry into an establishment and estimate jointly five probit

[12] Other studies use employment growth as a regressor in their analysis of earnings (Kramarz et al., 1996; Entorf and Kramarz, 1997). But, they do not focus on the resulting estimates of employment growth coefficients.

equations. In particular, they try to examine the age of the mobile workers and how the selection process of such age category is determined in each country by the specific labor market institutions that prevail. Even though their results are only preliminary, the estimation methodology and the way the different datasets are matched constitute an excellent example of the potentialities of the use of matched employee–employer longitudinal data.

Hassink (1999) uses longitudinal matched Dutch data (see Table 1) to examine the effects of firm and employee characteristics on the probability of layoffs. He conducts parallel analyses using the firm's lay off rate and the individual's layoff event as the two dependent variables. Using a specification that includes firm effects, firm characteristics and individual characteristics. The effect of seniority is negative and essentially linear in both equations. The minimum layoff probability occurs at age 32, a result that is interpreted as supporting Lazear-style compensation models.

7.6. Training

The potential of matched employer–employee data to address issues surrounding training is enormous. Indeed, we believe that questions such as the identification of general versus specific knowledge can only be addressed with such longitudinal datasets. And movements of workers between firms, workers for which we measure most individual characteristics would help isolate those firms which provide firm-specific training on one side and those firms which provide general training on the other. Unfortunately, there are few datasets that provide information on training of individuals together with employer information. Even though some are being built now.

Bishop (1994) uses the EOPP which provides retrospective longitudinal data on training and productivity of two new hires at 659 firms. Using this pair, Bishop is able to estimate all equations of interest by doing within-firm difference, therefore eliminating all firm-specific unobserved heterogeneity. The dependent variables are respectively the logarithm of training time, the productivity at the end of first week, the starting wage, the current productivity, the current wage, and the profit in the first months. The results can be summarized as follows. New hires with relevant previous work experience, relevant employer-sponsored formal training, and relevant vocational education tend to require less training, to be more productive, and to receive higher starting wages and higher wages after 1 year of seniority.

Similar questions are examined in a group of papers based on a newly available dataset, the Multi-City Study of Urban Inequality for Holzer and Reaser (1996), the 1995 Survey of Employer-Provided Training for Frazis et al. (1997) (see Table 1). Unfortunately, the first of these two datasets only has one observation per person or per firm, while the paper which uses the second dataset does not use the full potentiality of matched employer–employee data which makes most estimated coefficients difficult to interpret since they are likely to be biased due to unobserved person or worker heterogeneity. Frazis et al. (1998) extend this analysis to show that those establishments that encourage long term relationships, using pension plans and other employee benefits, also provide more training.

Finally, Goux and Maurin (1997) use the French FQP dataset (see Table 1) to examine the impact of training on wages and mobility. Interestingly, they show that having been trained in the past years is associated with a higher wage (approximately 6%) in a simple OLS regression. However, the introduction of firm fixed-effects reduces this effect to less than 3%. Furthermore, when a correction for the selection bias induced by participation in a training program is introduced, all effects of training on wages disappear. Hence, they conclude that the higher wage associated to training is partly due to firm-specific compensation policies and partly due to unobserved worker heterogeneity.

7.7. Unions and collective bargaining

This section discusses the use of matched employer–employee data to study the behavior of unionized firms and negotiated wage rates. We consider, in sequence, Abowd and Allain (1996), Cahuc and Kramarz (1997), Lalonde et al. (1996), Hildreth (1996), Hildreth and Pudney (1997) and Margolis (1993).

Abowd and Allain (1996) use data from the French DADS and BIC (see Table 1) to model the division of the quasi-rent per worker in collectively bargained French wage rate.[13] They fit an equation of the form

$$w_j = x_j + \gamma_j q_j + \varepsilon_j, \tag{7.2}$$

where w_j is the negotiated wage rate, x_j is the opportunity cost of the worker's time, γ_j is the bargaining power of the union, and q_j is the expected quasi-rent per worker. The heterogeneity in γ_j is modeled using q_j and other variables z_j. Using the decomposition in Eq. (3.1), the opportunity cost of the workers is modeled as

$$x_j = \bar{\theta}_j + \psi_{10},$$

where ψ_{10} is the firm effect at the 10th percentile of the French labor force.[14] The quasi-rent per worker has two components: an expected part, which is related to international competition using export prices (from France or from the United States), and a measurement error, which is eliminated by the instrumental variable procedure (see Abowd and Lemieux, 1993). Two empirical measures of the quasi-rent per worker were used–one which eliminated only the opportunity cost of the workers' time and the other which also eliminated an estimate of the opportunity cost of capital. The interpretation of the coefficient on the quasi-rent per worker is, therefore, the average part of the expected quasi-rent per worker that goes to the workers. Abowd and Allain estimate that this coefficient is 0.4 in the French economy.

Hildreth (1996) investigates the same question as Abowd and Allain using the British PSME, a panel of manufacturing establishments (see Table 1) and the British Household

[13] In an earlier effort, Abowd and Kramarz (1993) use the firm data from the BIC combined with occupation data from the ESE and aggregated wage data from the DADS to fit models similar to those in Abowd and Allain. This earlier paper does not make direct use of matched employer–employee data.

[14] Approximately 90% of French jobs are covered by collective bargaining agreements.

Panel Study (BHPS). The basic wage equation is essentially the same as Eq. (7.2), except that Hildreth specifies the relation using log wage as the dependent variable, which means that the coefficient on q_j cannot be interpreted as the bargaining power of the union. The method of calculating the quasi-rent per worker is also different. Hildreth defines the opportunity cost of the worker's time using a table of the usual weekly earnings cross-classified by education, age and sex, with further refinements for the location and industry of the establishment. The appropriate value from this table was subtracted from the value-added per worker to get the quasi-rent per worker. There was no correction for the opportunity cost of capital. Hildreth gets estimates that are much smaller than those of Abowd and Allain, but of the same order of magnitude as those found in other studies using British data (e.g., Hildreth and Oswald, 1993). The main difference appears to be the interpretation of the bargaining power parameter. Abowd and Allain (following Abowd and Lemieux) interpret this parameter as applying to the expected quasi-rent per worker, that is, the part related to the price instruments, and not to the realized profit per worker, which is much more variable.

Cahuc and Kramarz (1997) use the French ESS of 1986 and 1992 (see Table 1) to examine the impact of the signature of a firm-level agreement on the stability of the work force. In some sense, they try to find an exchange of voice, as approximated by the existence of an agreement, against stability. Their analysis uses both the cross-sectional dimension of the ESS, i.e., individual information on multiple employees in each establishment, and the longitudinal dimension, i.e., the same establishments can be found in 1986 and 1992. Cahuc and Kramarz start by examining the probability of signature of an agreement at the firm-level. They show that this probability is positively affected by most variables that increase the cost of turnover, more particularly by training expenses and by the presence of workers with intermediary skills. Then, they examine the relation between workers' seniority and the impact of the signature of an agreement between 1986 and 1992. The relevant regressions have approximately 50,000 observations and more than 250 firm fixed-effects. Results show that the signature of an agreement induces an increase in the average seniority of the work force of roughly one month for every additional year of the agreement.

LaLonde et al. (1996) use an unusually well-conceived matched dataset containing longitudinal information on American manufacturing establishments and employee information on the conduct and results of union representation elections. The establishment data come from the Longitudinal Research Database while the union election data come from the National Labor Relations Board (see Table 1). A union representation election is necessary in the United States before the employees of an ongoing business can negotiated collectively over wages and working conditions. By following establishments over a period of 4 years prior and 9 years after the election, these authors were able to measure the effects of the newly formed union on total output, employment, other factor utilization, wage rates, and productivity. LaLonde et al. present both short and long term evidence comparing the profiles of establishments where the representation election was successful (union wins) with those which had an unsuccessful (union losses) representation election.

Following a successful representation election, establishments reduce their output, material purchases and employment levels permanently (an effect that lasts at least 9 years). The establishments do not, however, experience higher wage rates.

Margolis (1993) uses data from the French Enquête Structure des Salaires (see Table 1) matched with detailed information on the collective bargaining agreements supplied by the Ministry of Labor to study the consequences of mandatory extension of the collective agreements to firms and workers who did not participate in the negotiations, a common practice in France. He finds that the willingness of employers to join the negotiation is strongly affected by the probability that the agreement will be extended. He also finds that the possibility of non-compliance with the collective bargaining agreement influences the behavior of the firms during the negotiations.[15]

Hildreth and Pudney (1997) use the British Panel Study of Manufacturing Establishments (PSME, see Table 1), which includes information on two workers: the most recently hired and one randomly selected, cooperating, individual, to study the effects of union recognition on firm outcomes. In the United Kingdom, there is no statutory requirement that an employer recognize and bargain with a union. Employees can choose whether or not to join a union independent of the employers negotiating stance towards that union. Any collective agreement applies to all workers in the covered jobs regardless of the employees' union status. Hildreth and Pudney model the two-sided decision process that determines the union status of the employee and of the job using the matched data. Their statistical models correct for a variety of sampling and self-selection problems. They find that firms that recognize unions have lower quit rates and higher wage rates. Interestingly, the union wage premium is higher for individuals who are covered by the collective agreement but who do not join the union. The results also suggest that applicant workers do not find the jobs covered by a collective agreement more attractive than non-covered jobs, so there are not increased applicant rates for these jobs. These statistical results generally allow for firm effects in all equations.

7.8. Other firm outcomes

The new data sources matching workers and their firms have allowed American researchers to re-examine classical issues of American labor economics: race discrimination and sex segregation. All these new analyses have been based on the WECD (see Table 1 and Section 2.3). The matched data using state unemployment insurance records have also permitted the examination of the effects of changes in the tax system on layoffs and other employment decisions. We discuss these applications, as well as those that do not have an obvious place in other sections, in this subsection.

[15] Non-compliance with collective bargaining agreements in France is accomplished by reclassifying jobs into lower pay categories and gambling that the labor inspector will not force a higher classification.

7.8.1. Segregation of the work force

Carrington and Troske (1998b) examine the extent to which blacks and whites are integrated at work. In addition to the WECD (see Table 1), the authors use the Characteristics of Business Owners (CBO) database which give demographic information on owners, employees, and customers of small businesses (hence complementing the WECD which is particularly strong on large businesses). Then, the authors propose different measures of segregation and assess their adequacy in a multifirm context. First, they define the Gini coefficient as follows:

$$G = 1 - \sum_{i-1}^{T} s_{bi}\left(s_{wi} + 2\sum_{j=i+1}^{T} s_{wj}\right),$$

where T is the number of firms, s_{bi} and s_{wi} are firm i's share of the black and white sample populations, respectively, and where firms are sorted in ascending order of s_{bi}/s_{wi}. Then, they define the Gini coefficient of random segregation in order to take into account the fact that random allocation of black and white workers will never generate a zero Gini coefficient. Based on the comparison of the two Gini indices, they create a Gini coefficient of systematic segregation.

Their results suggest that the national distribution of black and white employees across employers is far from even, as some employers have mostly white employees while others have mostly black. They also show that this segregation is due to black-white differences in MSA residence. Most of the remaining interfirm segregation come from racial differences in occupation, industry, or by simple random allocation which can almost never be rejected. Then, using more classical tools, they regress the black share of non-supervisory employment in the establishment on the share of black supervisors, the black sample share within each MSA, the log of establishment employment, the average age and education of non-supervisory employees, and indicator variables for industry and region. They show in particular that, using the WECD, black workers tend to be supervised by black managers (and vice-versa for whites). While, using the CBO, they show that black workers are more likely to work for firms with black owners and customers.

Finally, they decompose the black-white wage gap into a between and a within-plant component. In particular, they use the same type of techniques already described at many places in the preceding subsections, i.e., they introduce establishment fixed-effects. Carrington and Troske' s results demonstrate that the wage gap is mostly a within-plant phenomenon. Very little of the black-white wage gap comes from the allocation of black and white workers into firms that pay systematically different wages. Moreover, a large fraction of the within-plant gap is explained by the observable characteristics of the workers even though a significant fraction cannot be explained. In addition, when wages are regressed on the racial structure of the employing firm, it appears that black-majority plants pay their black employees less than black-minority plants. But these black-majority plants also pay their white employees more than their black employees.

The same authors use the same database, the WECD, and the same techniques to

examine sex segregation (Carrington and Troske, 1998a). They find that the distribution of men and women across plants is far from even. But, they also find that much of this apparent segregation appears to be due to random allocation. Similarly to the race analysis, they examine the plant female share of non-supervisory employees and regress it on variables similar to those described above. Of interest are the following results that female managers tend to supervise female employees and that women have higher employment shares in large establishments. The analysis of the male-female wage gap proceeds along the same lines as those presented for the black-white wage gap. The authors show that there is an important, even dominant in the case of blue-collar workers, role played by between-plant segregation in explaining this wage gap. Therefore, men work in relatively high paying plants while women work in relatively low paying plants even after controlling for observable characteristics of the workers. In addition, they demonstrate that workers, either men or women, are paid less if they work in largely female plants.

Hellerstein et al. (1997) continue the analysis of sex discrimination. They match the WECD with information from the Longitudinal Research Database (LRD) to get information on the employing establishment or firm. They find that large firms or large establishments make more profits if they employ more women. No such relation exist for small firms. In the present version of the paper at least, the authors are not able to provide a definitive explanation of this phenomenon. Bayard et al. (1999) also study sex segregation and male-female wage differentials. In the current version of the paper, they find that a substantial portion of the male-female wage gap takes the form of wage differentials within narrowly-defined occupations within establishments, results that stand in marked contrast to Groshen (1991b), who found that sex-segregation into occupations within establishments explained most of the gap.

7.8.2. Unemployment insurance and layoffs

In Anderson and Meyer (1994), these authors begin a long series of papers that used state-level unemployment insurance system matched employer–employee data to study the effects of the unemployment insurance tax and benefit system on a variety of outcomes: layoffs, employment, wages, and UI benefit takeup. The 1994 paper is discussed in section 7.5. The data structure in the other papers is very similar and is not discussed again. Anderson and Meyer (1996a,b, 1997) use the state unemployment insurance data and the establishment employment information to establish a number of basic features of the UI system and its effects on labor market outcomes.

Anderson and Meyer (1996a) studies the effects of firm-level experience rating on layoff probabilities. Experience rating is the system of UI financing that increases a firm's UI tax payments as the firm imposes benefit liabilities on the system. The effect of such financing systems on a the firm's propensity to use layoffs and on its wage structure is an old an important question in the labor economics literature. They use the same States as the 1994 paper. They use a form of Eq. (3.1) in which the dependent variable is the event that a worker is laid off during the quarter. Both person and firm effects are included. The firm's UI tax rate is included in the model and instrumental variables are used to correct for the

endogeneity that experience-rating induces. They find that the elasticity of the layoff rate with respect to the firm-specific tax cost is −0.3 and the corresponding fraction of temporary layoff unemployment that can be attributed to incomplete experience-rating is 20%.

Anderson and Meyer (1996b) studies the adoption of experience rating in the State of Washington UI system. Because the State of Washington adopted experience rating in 1985 to avoid a massive surplus in the system, these authors, who have data from 1979 to 1993 are able to provide direct evidence on the changes surrounding the adoption. They examine in detail the changes between the last half of 1984 (third and fourth quarters) and the last half of 1985 (again, third and fourth quarters). The change in the in the UI tax rates that was induced by the adoption of experience ratings was based on layoff rates over the period 1980:3–1984:2; however, the firm's only learned of the adoption of experience rating in 1984:3. Thus, this period represents an essentially exogenous change in the UI tax rates. The authors report that the full amount of the market-level change in tax rates is passed on to the workers in the form of lower earnings but that the firm-specific component is borne completely by the firm. They report mixed results of the effects of the experience rating on layoffs.

Anderson and Meyer (1997) use the CWBH data discussed in conjunction with their 1994 paper to study the effect of UI benefit levels and benefit tax treatment on the take-up rate for UI. They explain the late 1980s–early 1990s decline in UI receipts. According to these authors there is a strong positive relation between benefit levels and take-up rates. There are smaller, but still important effects arising from the tax treatment and potential duration of benefits. The inclusion of UI income in the US income tax base, therefore, accounts for most of the recent decline in UI receipts.

Abowd and Allain (1997) use the State of Washington UI data (see Table 1) to study the role played by workers and firms observable characteristics, as well as unobserved heterogeneity, in the probability that an individual participates in a short-time UI compensation (STC) program. Short-time compensation programs allow firms to pay UI benefits to workers whose hours have been reduced to avoid layoffs. These authors show that both types of unobserved heterogeneity are strongly correlated with this probability, with the individual effect having stronger correlation than the firm effect. In the context of Eq. (3.1), the dependent variable is the incidence of short-time compensation. A person-effect means that the individual has experienced short-time UI compensation and a firm effect means that the firm has used short-time compensation. Thus, the results are interpreted as meaning that some individuals have a greater propensity to be employed in short-time compensation jobs than others and some firms have a greater propensity to use this form of UI compensation. Firms with higher experience ratings were more likely to use short-time compensation.

Needels and Nicholson (1998) also study short time compensation systems using UI data from the states of California, Florida, Kansas, New York and Washington for the period 1991–1993 for 3300 establishments. They use a statistical matching algorithm to pair establishments with short-time compensation programs to those without. The statistical analyses are all conducted by differencing the paired establishments. Establishments

with STC programs have higher layoff levels than those without, which they interpret as evidence of unmeasured heterogeneity among the establishments.

7.8.3. International trade and other topics

Kramarz (1997) examines the impact of international trade on wages and mobility of French workers using the matching of the French labor force survey with a unique dataset on all imports (to France) and exports (from France) of goods during the period 1986–1990. Origins of the imports and destination of the exports are known at the firm-level and are disaggregated into eight groups of countries. It comprises all movements of goods since it is an administrative data source from the customs administration. Matching is performed using the Siren number present in both files. This import-export dataset is also matched to the Echantillon d'Entreprises to measure total sales and total purchases. Hence, the independent variables on trade in the regressions are the ratio of imports to total purchases – a way to measure the reorientation of purchases from local markets to outside suppliers – and the ratio of exports to total sales. Then, Kramarz computes the change in these ratios between 1986 and 1990 for firm j and relates them to the probability of being unemployed at date $t + 1$ conditional on being employed in the same firm j at date t (t going from 1990 to 1993). He also examines the impact on the level of wage at date t in firm j. All these regressions include individual characteristics from the LFS that one expects to find in this type of analysis. In addition, to assess the impact of the competitive pressure, he also includes the change of the ratio of imports to purchases of all firms in the same 4-digit sector as well as in firms from the trade (retail or wholesale) industries. Results are the following. The unemployment probability is positively affected by increasing imports from the 4-digit competitors of the firm; a best response for the firm being to increase its imports. Hence, importing protects workers from unemployment when most other firms in the same sector increase their imports. The impact of imports and exports on wages have a similar structure. An increase of the share of firm-level purchases coming from outside France between 1986 and 1990 negatively affect the level of future wages while the opposite is true both for exports and imports from the firms of the same 4-digit sector. Origins of the imports appear to matter. For instance if these imports come from Germany, the impact on wage is positive, while if they come from developing countries or from the UK, the impact is large and negative. Similarly, workers employed at date t in firms that increase their share of exports to Japan have higher wages. Of course, these results could be seen as evidence of the impact of international trade on prices or, on the contrary, as showing nothing on international trade but capturing worker unobserved heterogeneity.

Abowd and Kramarz (1998b) analyze the costs of separating from French workers using the 1992 ESS (see Table 1). In this study, the authors used the individual-level variables: total annual compensation inclusive of all employee- and employer-paid benefits and bonuses but exclusive of non-wage-benefits, firm seniority, type of contract (permanent, CDI, or temporary, CDD), number of days of employment in the establishment in 1992, sex, age, nationality (French or non-French), skill-level (in 4 groups), bonuses for retire-

ment and severance payments for workers that retired or were fired in 1992. They present estimates of the structure of retirement, termination, and hiring costs using, representative establishment-level data matched with individual-level information. These costs are directly reported by the sampled establishments. Both retirement and termination costs are increasing and mildly concave in the number of retired or terminated workers. The fixed costs are very large. Hence, these costs act as fixed adjustment costs, giving the firm an incentive to group exits instead of adjusting gradually. Termination costs are largest for collective terminations as opposed to individual ones. These costs are largest for highly skilled employees. Hiring costs also exhibit the same structure; concave adjustment costs with a strong fixed component. But these hiring costs do not have the same structure for all skill levels. Only hires of managers on longterm contracts (CDI) have an increasing and concave impact on the cost. For all other skill levels and types of contract, hiring costs do not depend upon the number of entries. Thus, for hiring costs, the firms have an incentive to group the managerial hiring but no adjustment costs for other hiring. The costs of hiring are much less important in France than the costs of separations (retirements and terminations).

Abowd et al. (1996) consider the following question: Are high-quality products produced by high-quality workers? To do this they use the decomposition of wages from Abowd et al. (1999a) and price measures of product quality. To measure the quality of a product, they use prices for very detailed products (8-digit classification) collected at the firm-level. Each basic product is allocated to a 6-digit basic commodity group. Therefore, quality can be measured as either the relative price of a 8-digit elementary product within a 6-digit basic commodity group or as the price change of each elementary product within the basic commodity group. Abowd et al. find little relation between worker quality and product quality within basic products. Hence, technological differences among firms, given basic products, seem very small. They conclude that, if worker quality and product quality are positively related, the effect is apparently more important for sorting workers among diverse and non-substitutable products than for explaining variation within groups of imperfectly substitutable detailed products.

7.9. Specialized applications

There are a variety of specialized uses of matched employer–employee data that we have not discussed in detail in this chapter because they figure prominently in other surveys. Pension data collected from the employers have been matched to several nationally representative cross-sectional and longitudinal databases in the United States including the Health and Retirement Survey, Mature Cohorts from the National Longitudinal Surveys, and the Survey of Consumer Finances. See the chapter on retirement issues by Mitchell and Lumsdaine for a discussion of these applications. Health researchers have also used administrative matched data to study productivity issues in the health service industry (see Dunn et al., 1998, and the chapter on health issues by Currie).

8. Conclusion

As the beginning of our chapter makes clear, new economic and statistical problems will emerge as new types of labor market questions are investigated with detailed data concerning both the worker and the firm. Even though the analysis of matched employer–employee data is relatively new, we are already confronted with some puzzling new results: the lack of correlation between person and firm effects in wage determination and the enormous employment flows associated with job creation and destruction, among others. To model such new facts, standard models of the allocation of workers among firms must be modified.

New statistical problems have also emerged: the analysis of duration models with correlated person and firm effects and the design of statistical models for non-random matches, for example. To estimate such statistical models, some solutions already exist, based on simulations, but they are extremely computer-intensive. Some simpler ones will surely be implemented in the near future. These statistical models could also be useful in other areas, such as health economics (doctors and hospitals–on one side–and their clients–on the other) or education economics (schools and professors–on one side–and their students–on the other) to resolve the same type of identification questions that the analysis of matched employer–employee data have helped resolve.

An area that will be even more demanding is the formulation and estimation of structural economic models. As we show in our discussion of the relation between different theoretical models and the simplest wage equation with correlated person and firm effects and firm-specific returns to seniority, an enormous amount of detail is required to assign the statistical effects to an economic model. Recovering the deep structural parameters from statistical models that include such effects will surely be difficult. In addition, one can argue that it is very unlikely that all firms follow the same model. Hence, the estimation of structural models will force the researcher to address structural heterogeneity problems, for instance, is rent-sharing more important than agency problems for a particular firm?.

Matched longitudinal employer–employee datasets should constitute the basis for further refinements of the theory of production and of the theory of the workplace organization. The possibility of evaluating the various combination of workers, jobs, and machines within a firm should allow labor economists to delve deeper into the internal organization of the firm. Indeed, data collected in the future should give information on each job in conjunction with each individual job holder in each individual firm. We are back to the "get more data" conclusion, so that we can play the role of Rosen and Willis for this volume of the *Handbook*.

References

Abe, Masahiro and Catherine Sofer (1996), "Effets de l'ancienneté sur le salaire: une comparaison France/Japon", Comparaisons internationales de salaires (INSEE-DARES, Paris) pp. 198–222.

Abowd, John M. and Laurence Allain (1996), "Compensation structure and product market competition", Annales d'économie et de statistique 41–42: 207–217.

Abowd, John M. and Laurence Allain (1997), "The Washington State short-time compensation program and its implication for European work share programs", Working paper (Cornell University).

Abowd, John M and Hampton Finer (1998), "Incorporating employer data into the National Longitudinal Survey Of Youth 1979 cohort: methods and results", Working paper (Cornell University).

Abowd, John M. and Francis Kramarz. (1993), "A test of negotiation and incentive compensation models using longitudinal French enterprise data", in: J.C. van Ours, G.A. Pfann and G. Ridder, eds., Labour demand and equilibrium wage formation (North Holland, Amsterdam) pp. 111–146.

Abowd, John M. and Francis Kramarz (1996a), "Les politiques salariales: individus et entreprises", Revue Economique 47: 611–622.

Abowd, John M. and Francis Kramarz (1996b), "The microeconometrics of human resource management: international studies of firm practices, introduction and overview", Annales d'économie et de statistique 41–42: 11–19.

Abowd, John M. and Francis Kramarz (1998a), "Internal and external labor markets: an analysis of matched employer-employee data", in: J. Haltiwanger, M. Manser and R. Topel, eds., Labor statistics measurement issues (University of Chicago Press, Chicago, IL).

Abowd, John M. and Francis Kramarz (1998b), "The cost of hiring and separations", Working paper no. 6110, revised (NBER, Cambridge, MA).

Abowd, John M. and Francis Kramarz (1999), "Econometric analysis of linked employer-employee data", Labour Economics 6: 53–74.

Abowd, John M. and Thomas Lemieux (1993), "The effects of product market competition on collective bargaining agreements: the case of foreign competition in Canada", Quarterly Journal of Economics 108: 983–1014.

Abowd, John M., Francis Kramarz and Antoine Moreau (1996), "Product quality and worker quality", Annales d'économie et de statistique 41–42: 299–322.

Abowd, John M., Hampton Finer, Francis Kramarz and Sébastien Roux (1998a), "Job and wage mobility: an analysis of the dynamics of employment durations using matched employee and employer data from the U.S. and France", Working paper (Cornell University and INSEE-CREST).

Abowd, John M., Francis Kramarz, David N. Margolis and Kenneth R. Troske (1998b), "The relative importance of employer and employee effects on compensation: a comparison of France and the United States", Mimeo.

Abowd, John M., Francis Kramarz and David N. Margolis (1999a), "High wage workers and high wage firms", Econometrica 67: 251–333.

Abowd, John M., Patrick Corbel and Francis Kramarz (1999b), "The entry and exit of workers and the growth of employment: an analysis of French establishments", Review of Economics and Statistics 81: in press.

Abowd, John M., Hampton Finer and Francis Kramarz (1999c), "Individual and firm heterogeneity in compensation: an analysis of matched longitudinal employer-employee data for the State of Washington", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 3–24.

Albaek, Karsten and Bent E. Sørensen (1999), "Worker flows and job flows in Danish manufacturing, 1980–91", Economic Journal, in press.

Alexander, Michael and Alison Morehead (1999), "The 1995 Australian Workplace Industrial Relations Survey – a discussion of the methodology used to create linked employer-employee data", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 535–552.

Anderson, Patricia M. and Bruce D. Meyer (1994), "The extent and consequences of job turnover", Brookings Papers Microeconomics: 177–248.

Anderson, Patricia M. and Bruce D. Meyer (1996a), "The effects of unemployment insurance taxes and benefits on layoffs using firm and individual data", Working paper (Dartmouth College).

Anderson, Patricia M. and Bruce D. Meyer (1996b), "Using a natural experiment to estimate the effects of the unemployment payroll tax on layoffs, employment and wages", Working paper (Dartmouth College).

Anderson, Patricia M. and Bruce D. Meyer (1997), "Unemployment insurance takeup rates and the after-tax value of benefits", Quarterly Journal of Economics 112: 913–937.

Arai, Mahmood, Gérard Ballot and Ali Skalli (1997), "Différentiels intersectoriels de salaire et caractéristiques des employeurs en France", Economie et Statistique 299.

Baker, Michael and Gary Solon (1997), "Earnings dynamics and inequality among Canadian men, 1976–1992: evidence from longitudinal income tax records", Mimeo.

Barth, Erling (1997), "Firm-specific seniority and wages", Journal of Labor Economics 15 (3 part 1): 495–506.

Barth, Erling and Harald Dale-Olsen (1999), "The employer's wage policy and worker turnover", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 285–312.

Bayard, Kimberly, Judith Hellerstein, David Neumark and Kenneth Troske (1999), "Why are racial and ethnic wage gaps larger for men than for women? Exploring the role of segregation using the new worker-establishment characteristics database", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 175-204.

Belzil, Christian (1997), "Job creation and destruction, worker reallocation and wages", Working paper (Department of Economics, Concordia University).

Bender, S., J. Hilzendegen, G. Rohwer and H. Rudolph (1996), "Die IAB Beshaeftigungsstichprobe 1975–1990", Beitraege zur Arbeitsmarkt- und Berufsforschung (BeitrAB 197, Nuernberg).

Bingley, Paul and Niels Westergård-Nielsen (1996), "Individual wages within and between establishments", Working paper (University of Århus).

Bingley, Paul and Niels Westergård-Nielsen (1998), "Establishment tenure and worker turnover", Working paper (University of Århus).

Bishop, John (1994), "The impact of previous training on productivity and wages", in: Lisa Lynch, ed., Training and the private sector (University of Chicago Press for NBER, Chicago, IL).

Blanchflower, David and Stephen Machin (1996), "Product market competition, wages and productivity: international evidence from establishment-level data", Annales d'économie et de statistique 41–42: 219–254.

Boon, Martin (1996), "Effects of firm performance and technology on wages: evidence from cross-sectional matched worker-firm data", Working paper no. 4692-96-RSM (Department of Statistical Methods, Statistics Netherland).

Bronars, Stephen G. and Melissa Famulari (1997), "Wage, tenure and wage growth variation within and across establishments", Journal of Labor Economics 15 (2): 285–317.

Bronars, Stephen G., Paul Bingley, Melissa Famulari and Niels Westergård-Nielsen (1999), "Employer wage differentials in the United States and Denmark", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 205–230.

Brown, Charles and James L. Medoff (1989), "The employer size-wage effect", Journal of Political Economy 97: 1027–1059.

Brown, Charles and James L. Medoff (1996), "Employer characteristics and work environment", Annales d'économie et de statistique 41–42: 275–298.

Bureau of Labor Statistics (1996), "BLS reports on the amount of formal and informal training received by employees", Press release (December 19) pp. 1–21.

Burgess, Simon, Julia Lane and David Stevens (1997), "Jobs, workers and changes in earnings dispersion", Discussion paper no. 1714 (CEPR).

Burgess, Simon, Julia Lane and David Stevens (1999), "Job flows, worker flows and churning", Journal of Labor Economics 17, in press.

Cahuc, Pierre and Francis Kramarz (1997), "Voice and loyalty as a delegation of authority: a model and a test on a matched worker-firm panel", Journal of Labor Economics 15: 658–688.

Callus, R., A. Morehead, M. Cully and J. Buchanan (1991), Industrial relations at work. The Austrialian workplace industrial relations survey (Australian Government Publishing Service, Canberra, Australia).

Cardoso, Ana Rute (1997), "Company wage policies: do employer wage effects account for the rise in labor market inequality?" Working paper (European University Instititute, Italy).

Carrington, William J. and Kenneth R. Troske (1998a), "Sex segregation in U.S. manufacturing", Industrial and Labor Relations Review 51: 445–464.

Carrington, William J. and Kenneth R. Troske (1998b), "Interfirm segregation and the black-white wage gap", Journal of Labor Economics 16 (2): 231–260.

Chennouf, S., L. Lévy-Garboua and C. Montmarquette (1997), "Les effets de l'appartenance à un groupe de travail sur les salaires individuels", Working paper (LAMIA, University of Paris-I).

Contini, Bruno, Riccardo Revelli et al. (undated), "The R&P longitudinal sample of workers and firms based on the Social Security archives", Working document (Ricerche e Progetti, Turin, Italy).

Cully, Mark (1998), "Guvnors, employees and brothers: triangulation and noise in workplace surveys", Working paper (UK Department of Trade and Industry).

Davis, Steven J. and John Haltiwanger (1992), "Gross job creation, gross job destruction and employment reallocation", Quarterly Journal of Economics 107: 819–863.

Davis, Steven J. and John Haltiwanger (1996), "Employer size and the wage structure in the United States", Annales d'économie et de statistique 41–42: 323–368.

DiPrete,Thomas A., Dominique Goux, Eric Maurin and Michael Tåhlin (1998), "Establishment dynamics and job mobility: a comparative analysis of the age-mobility relationship in France and Sweden using linked employer-employee data", Working paper (Duke University, CREST, Stockholm University).

Dolton, Peter, Maarten Lindeboom and Gerard J. van den Berg (1999), "A taxonomy of survey nonresponse and its relation to labor market behavior", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 439–460.

Doms, M., T. Dunne and K.R. Troske (1997), "Workers, wages and technology", Quarterly Journal of Economics CXII: 253–290.

Dunn, Rodney L., Ronald J. Ozminkowski, Ron Z. Goetzel and Teresa B. Gibson (1998), "Issues and challenges when merging health and productivity management (HPM) and insurance claims data", Working paper (MEDSTAT Group).

Dustmann, Christian and Costas Meghir (1997), "Seniority and wage growth", Working paper (IFS and University College London).

Entorf, Horst and Francis Kramarz (1997), "Does unmeasured ability explain the higher wage of new technology workers", European Economic Review 41: 1489–1510.

Entorf, Horst and Francis Kramarz (1999), "The impact of new technologies on wages: lessons from matching panels on employees and on their firms", Economics of Innovation and New Technology, in press.

Entorf, Horst, Michel Gollac and Francis Kramarz (1999), "New technologies, wages and worker selection", Journal of Labor Economics 17: in press.

Finer, Hampton (1997), "Firm and individual wage components in the national longitudinal survey of youth", Working paper (Department of Labor Economics, Cornell University).

Frazis, Harley, Maury Gittleman, Michael Horrigan and Mary Joyce (1997), "Formal and informal training: evidence from a matched employee-employer survey", in: G.D. Libecap, ed., Advances in the study of entrepreneurship, innovation and economic growth, Vol. 9 (JAI Press, Greenwich, CT) pp. 9–46.

Frazis, Harley, Maury Gittleman and Mary Joyce (1998), "Determinants of training: an analysis uing both employer and employee characteristics", Working paper (Bureau of Labor Statistics).

Goux, Dominique and Eric Maurin (1997), "Train or pay: does it reduce inequalities to encourage firms to train their workers", Working paper no. G9703 bis (INSEE, Direction des Études et Synthèses Économique).

Goux, Dominique and Eric Maurin (1999), "Inter-industry wage differentials: the role of employers", Journal of Labor Economics 17: in press.

Groshen, Erica (1991a), "Sources of intra-industry wage dispersion: how much do employers matter?" Quarterly Journal of Economics 106: 869–884.

Groshen, Erica (1991b), "The structure of the female/male wage differential: is it who you are, what you do, or where you work?" Journal of Human Resources 26: 457–472.

Groshen, Erica (1996), "American employer salary surveys and labor economics research: issues and contributions", Annales d'économie et de statistique 41–42: 413–442.

Haegeland, Torbjörn and Tor Jakob Klette (1999), "Do higher wages reflect higher productivity? Education, gender and experience premia in a matched plant-worker data set", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 231–260.

Hamermesh, Daniel, Wolter Hassink and Jan Van Ours (1996), "Job turnover and labor turnover: a taxonomy of employment dynamics", Annales d'économie et de statistique 41–42: 21–40.

Hassink, Wolter (1999), "On the incidence of layoffs", in J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 329–344.

Hayami, Hitoshi and Masahiro Abe (1998), "Labor demand by age and gender: evidence from linked micro data", Working paper (Keio University).

Hellerstein, Judith K., David Neumark and Kenneth R. Troske (1996), "Wages, productivity and worker characteristics: evidence from plant-level production functions and wage equations", Working paper no. 5626 (NBER, Cambridge, MA).

Hellerstein, Judith K., David Neumark and Kenneth R. Troske (1997), "Market forces and sex discrimination", Working paper no. 6321 (NBER, Cambridge, MA).

Hildreth, Andrew K. (1996), "Rent-sharing and wages: product demand or technology driven premia?" Working paper (University of Essex).

Hildreth, Andrew K. and Andrew J. Oswald (1993), "Rent-sharing and wages: evidence form company and establishment panels", Working paper no. 154 (Institute of Economics and Statistics, Oxford University).

Hildreth, Andrew K. and Stephen E. Pudney (1997), "Employers, workers and unions: an analysis of a firm-worker panel with endogenous sampling, attrition and missing data",Working paper (University of Essex).

Hildreth, Andrew K. and Stephen Pudney (1999), "Econometric issues in the analysis of linked cross-section employer-worker surveys", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 461–488.

Hildreth, Andrew K. and N. Tremlett (1994), "Panel study of manufacturing establishments - first stage", Technical paper no. 8 (ESRC Research Centre on Micro-Social Change, University of Essex).

Hildreth, Andrew K. and N. Tremlett (1995), "Panel study of manufacturing establishments - second stage", Technical paper no. 10 (ESRC Research Centre on Micro-Social Change, University of Essex).

Holzer, Harry J. and Jess Reaser (1996), "Firm-level training for newly hired workers: its determinants and effects", Working paper (Michigan State University).

Jacobson, Louis S., Robert J. LaLonde and Daniel G. Sullivan (1993), "Earnings losses of displaced workers", American Economic Review 83: 685–709.

INSEE (1994), Enquêete sur l'emploi de 1993: résultats détaillés, no. 59-60 (INSEE, Paris).

Jovanovic, Bojan (1979), "Job matching and the theory of turnover", Journal of Political Economy 87: 972–990.

Kramarz, Francis (1997), "International competition, employment and wages: the microeconometrics of international trade", Working paper (CREST-INSEE).

Kramarz, Francis and Patrick Rey (1995), "The dynamics of wages", Working paper (CREST-INSEE).

Kramarz, Francis and Sébastien Roux (1998), "Within-firm seniority structure and firm performance", Working paper (CREST-INSEE).

Kramarz, F., S. Lollivier and L.P. Pelé (1996), "Wage inequalities and firm–specific compensation policies in France", Annales d'economie et de statistique 41–42: 369–386.

Kremer, Michael and Eric Maskin (1996), "Specialization by skill and the rise in inequality", Working paper (MIT).

Krueger, Alan B. (1993), "How computers have changed the wage structure: evidence from microdata, 1984–1989", Quarterly Journal of Economics 108: 33–60.

Krueger, Alan B. and Lawrence H. Summers (1988), "Efficiency wages and the inter-industry wage structure", Econometrica 56: 259–293.

Laaksonen, Seppo, Mika Maliranta, Hannu Piekkola, Jarl Vainiomäki, Pertti Haaparanta and Pekka Ilmakunnas (1998), "Linking of employee and plant characteristics for the analysis of Finnish manufacturing", Working paper (Statistics Finland).

Lalonde, Robert, J. Gérard Marschke and Kenneth Troske (1996), "Using longitudinal data on establishments to

analyze the effects of union organizing campaigns in the United States", Annales d'économie et de statistique 41–42: 155–186.

Lane, Julia, Simon Burgess and Jules Theeuwes (1997a), "The uses of longitudinal matched worker/employer data in labor market analysis", American Statistical Association Papers and Proceedings 249–254.

Lane, Julia, D. Stevens and Simon Burgess (1997b), "Churning dynamics: an analysis of hires and separations at the employer level", Mimeo.

Lane, Julia, Javier Miranda, James Spletzer and Simon Burgess (1999), "The effect of worker reallocation on the earnings distribution: longitudinal evidence from linked data", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 345–374.

Leonard, Jonathan (1987), "In the wrong place at the wrong time: the extent of structural and frictional unemployment", in: K. Lang and J. Leonard, eds., Unemployment and the structure of labor markets (Basil Blackwell, Oxford, UK), pp. 141–163.

Leonard, Jonathan S. and Marc Van Audenrode (1996), "Worker's limited liability, turnover and employment contracts", Annales d'économie et de statistique 41–42: 41–78.

Leonard, Jonathan S. and Marc Van Audenrode (1997), "The persistence of pay", Working paper (University of California at Berkeley).

Leonard, Jonathan S., Benoit Mulkay and Marc Van Audenrode (1999), "Compensation policies and firm productivity", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 79–114.

Leth-Sørensen, Søren (1995), The IDA database - a longitudinal database of establishments and their employees (Statistics Denmark, IDA Project).

Mairesse, Jacques and Nathalie Greenan (1999), "Using employee-level data in a firm-level econometric study", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 489–514.

Margolis, David N. (1993), "Compensation practices and government policies in western European labor markets", PhD dissertation (Cornell University).

Margolis, David N. (1996), "Cohort effects and returns to seniority in France", Annales d'économie et de statistique 41–42: 443–464.

Margolis, David (1999), "Past-year employment, slow reemployment and earnings losses: the case of worker displacement in France", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 375–418.

McGuckin, Robert and George Pascoe (1988), "The longitudinal research database (LRD), status and research possibilities", Survey of Current Business: 30–37.

Ministério do Emprego e da Segurança Social (1993), Quadros de Pessoal, 1992, Colecçao Relatórios e Análises, Série estatísticas, 32 (Departamento de Estatística, Lisbon, Portugal).

Morehead, Alison, Mairi Steele, Michael Alexander, Kerry Stephen and Linton Duffin (1997), Changes at work: the 1995 Australian Workplace Industrial Relations Survey (Addison Wesley Longman, Melbourne, Australia).

Morissette, Rene and Charles Berube (1996), "Longitudinal aspects of earnings inequality in Canada", Working paper (Statistics Canada).

Needels, Karen and Walter Nicholson (1998), "The value of employer-based, linked UI data sets", Working paper (Mathematica Policy Research).

Pacelli, Lia (1997), "Wage growth of young Italian workers: does the firm matter?" Working paper (Ricerche e Progetti and University College London).

Pelé, Louis-Paul (1997), "Method of pay, worker selection and minimum wages", Working paper no. 9721 (CREST).

Persson, Helena (1998), "Job and worker flows in Sweden 1990—1995: who gets the new jobs and who loses the old jobs?" Working paper (Stockholm University).

Picot, Garnett and Ted Wannell (1996), "An experimental Canadian survey that links workplace practices and

employee outcomes: why it is needed and how it works", in: Comparaisons internationales de salaires (Ministère du travail et des affaires sociales and INSEE, France) pp. 17–34.

Rosen, Sherwin (1986), "The theory of equalizing differences", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North Holland, Amsterdam), pp. 641–692.

Rotbart, Gilbert (1991), Enquête sur la structure des salaires (INSEE-Méthodes, Paris).

Salvanes, Kjell G. and Svein E. Forre (1997), "Employment policies at the plant-level: job and worker flows for heterogeneous labor in Norway", Mimeo.

Scheffé, H. (1959), The analysis of variance (Wiley, New York).

Searle, Shayle R., George Casella and Charles E. McCulloch (1992), Variance components (Wiley, New York).

Seiler, Eric (1984), "Piece rate vs. time rate: the effect of incentives on earnings", Review of Economics and Statistics 66 (August) 363–376.

Smith, Creston M. (1989), "The social security administration's continuous work history sample", Social Security Bulletin 52: 20–27.

Stephan, Gesine (1998), "Establishment effects on wages in West Germany", Working paper (University of Hanover).

Tegsjö, Björn and Jan Andersson (1998), "Labor satistics based on administrative registers: the Swedish system linking individual data with enterprise information", Working paper (Statistics Sweden).

Topel, Robert and Michael Ward (1992), "Job mobility and the careers of young men", Quarterly Journal of Economics 107: 441–479.

Troske, Kenneth R. (1998), "The worker-establishment characteristics database", in: J. Haltiwanger, M. Manser and R. Topel, eds., Labor statistics measurement issues (University of Chicago Press for the NBER, Chicago, IL).

Troske, Kenneth R. (1999), "Evidence on the employer size-wage premium from worker-establishment matched data", Review of Economics and Statistics, in press.

Vainiomäki, Jari (1999), "Technology and skill upgrading: results from linked worker-plant data for Finnish manufacturing", in: J. Haltiwanger et al., eds., The creation and analysis of employer-employee matched data (North-Holland, Amsterdam) pp. 115–146.

Vejrup-Hansen, Per (1998), "The effect of job destruction on worker mobility and unemployment", Working paper (Copenhagen Business School).

Willis, Robert (1986), "Wage determinants: a survey", in: Orley Ashenfelter and Richard Layard, eds., Handbook of labor economics (North Holland, Amsterdam), pp. 525–602.

Winter-Ebmer, R. and J. Zweimüller (1997), "Intra-firm wage dispersion and firm performance", Working paper no. 9708 (Universität Linz).