Persistent Inter-Industry Wage Differences: Rent Sharing and Opportunity Costs

John M. Abowd, Cornell University, US Census Bureau, CREST, NBER and IZA
Francis Kramarz, CREST-INSEE, CEPR, and IZA
Paul Lengermann, Federal Reserve Board   Sébastien Roux, CREST-INSEE[1]

July 26, 2005

[1]The authors gratefully acknowledge the assistance of Robert Creecy, U. S. Census Bureau, who developed the computational algorithms used in this paper. They also acknowledge funding from the NSF (SBER 9618111 to the NBER and SES 9978093 to Cornell University). Joe Altonji, Orley Ashenfelter, Simon Burgess, David Card, Hank Farber, Hampton Finer, Juan Jimeno, Michael Krause, Michael Kremer, Julia Lane, David Margolis, Dale Mortensen, Walter Oi, Sébastien Perez-Duarte and seminar participants at many places provided valuable comments on earlier drafts. Melissa Bjelland, Benoit Dostie, Kaj Gittings, and Changhui Kang provided excellent research assistance. The French data used in this paper are confidential but the authors' access is not exclusive. For further information contact Francis Kramarz. This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854-02, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the authors and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see: www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov and see: lehd.dsd.census.gov). Author contact information: john.abowd@cornell.edu, kramarz@ensae.fr, paul.a.lengermann@frb.gov, and sebastien.roux@insee.fr.

Abstract We reconsider the potential for explaining inter-industry wage differences by decomposing those differences into parts due to individual and employer heterogeneity, respectively. Using longitudinally linked employer-employee data, we estimate the model for the United States and France. The part arising from individual heterogeneity can be theoretically and empirically related to the worker's opportunity wage rate. The part arising from employer heterogeneity can similarly be related to product market quasi-rents and relative bargaining power. We find that these two explanations do a good job of explaining both parts of the differential in France. Although the U.S. inter-industry wage differentials are well correlated with those in France, the decomposition is more nuanced in the American data, where the opportunity wage rate and the product market conditions are related to both the personal and employer heterogeneity.

Keywords: Inter-industry wage differentials, matched employer-employee data, quasi-rents

JEL: J31, J50, L10

1 Introduction

One of the most pervasive and difficult to explain phenomena in economics is the persistence of inter-industry wage differences. Some explanations predict that most of the variation is due to the persons employed in the industry, whose opportunity wage rates are similarly high or low. Other explanations predict that most of the variation is due to differential firm or industry compensation policies that do not follow the individual from job to job. Economists' ability to distinguish among these explanations has been hampered by the lack of appropriate matched, longitudinal employer-employee data. Recent developments in Europe and in North America have allowed researchers access to this type of data. This arrival has produced a resurgence of interest in some classic problems of labor economics for which both sides of the market – workers and firms – matter (see the survey on matched data sources and some early results in Abowd and Kramarz (1999a) or the international results contained in Haltiwanger and et al. (editors) (1999)). New techniques inspired from those used in statistical genetics have been developed to face the new econometric challenges (see Abowd and Kramarz (1999b) and more recently, Abowd, Creecy, and Kramarz (2002). See the survey in Abowd, Kramarz, and Woodcock (2004)). On the theoretical side, Mortensen (2003), Postel-Vinay and Robin (2002) developed their most recent search-based models to address the possibilities that linked data provide. Shimer (2001) developed recently an assignment model with search frictions partly inspired by some of the results contained in this literature (in particular, Abowd, Kramarz, and Margolis (1999), AKM hereafter).

In this paper, we make stock of these recent developments to reexamine this classic question of inter-industry wage differentials using new longitudinal linked employer-employee data for the United States and France. Although this topic received a flurry of attention in the 1980s, Krueger and Summers (1987) and Krueger and Summers (1988) established the consistency of these differentials over time and across countries, the fundamental question remained unresolved: are these differentials due to individual or employer components. Individual factors stay with the worker from job to job, whereas employer differences affect any worker who has a job with the firm. Because these two parts are directly interpretable in terms of economic models, it is important to apportion the inter-industry differentials into these person and employer parts. This can only be done using longitudinal linked employer-employee data (see AKM).

Krueger and Summers stressed factors related to the employer, such as compensation policy, as the primary explanation of the inter-industry differentials although their analysis showed that such factors were, at best, an incomplete explanation. Murphy and Topel (1987), on the other hand, stressed individual unmeasured differences as the primary cause of the wage differentials, although, once again, their data were incomplete. Dickens and Katz (1987) tried to explain the inter-industry wage differentials using a variety of measured individual and firm characteristics aggregated to the industry level; hence, their analysis was very much in the spirit we propose but they could not control for the unmeasured differences that we stress below. Gibbons and Katz (1992) attempted to explain the differential based on unobserved individual heterogeneity. Brown and Medoff (1989) focused their attention on the firm-size wage differential. They attempted to distinguish between explanations based on individual heterogeneity and those based on firm level compensation policy. In related work Groshen (1991) examined the role of firm and establishment compensation policy heterogeneity on wage outcomes, generally.

In two related articles AKM and Abowd, Finer, and Kramarz (1999) (AFK, hereafter) provided a basic statistical framework for decomposing inter-industry wage differentials and firm-size wage differentials into the sum of components due to individual heterogeneity (measured and unmeasured) and firm heterogeneity (measured and unmeasured). The first of these articles, AKM analyzes French data and finds that most of the inter-industry and the firm-size wage differentials are due to unmeasured individual heterogeneity. Goux and Maurin (1999) also find that most of the French inter-industry wage differential is due to individual heterogeneity using linked employer-employee data from the French Labor Force Survey (Enquête Emploi). The second of these articles, AFK, analyzes data from the State of Washington and finds that inter-industry wage differentials are due in equal proportions to individual and employer heterogeneity while firm-size wage differentials are due primarily to firm heterogeneity.

Both AKM and AFK used statistical approximations to estimate the decomposition of wage differentials into individual and employer components. Furthermore, these papers did not try to understand the origin of these differentials. In this article we apply new methods to the estimation problem, new data sources, some of them part of the recent effort of the US Bureau of the Census to assemble matched employer-employee sources, and an old model of bargaining that allows us to present economic interpretations of our results.

This paper is organized as follows. In section 2 we present a simple economic model that we use to interpret our estimates. Then, in section 3, we briefly discuss our estimation methods as well as the framework necessary to understand the estimated person and firm components of the inter-industry wage differentials and relate them to the literature. Section

4 presents the American and French data sources. Section 5 discusses the inter-industry wage differential results. In particular, and in contrast to the previous literature, we try to directly measure the economic source of these differentials. Section 6 concludes.

2 A Simple Economic Model

To give an economic interpretation of our estimates of person and firm effects in a context of inter-industry wage differentials, we need an equilibrium model of wage determination with imperfect competition. Indeed, with perfect competition in the labor market, firm-effects should not exist. Therefore, we posit a simple bargaining model of wages of which perfect competition is a special case. Let the wage be

$$w_i = x_i + \gamma \frac{QR_{j(i)}}{L_{j(i)}}$$

where x denotes the opportunity cost of time of worker i, QR and L be respectively the quasi-rent and employment of her employing firm $j(i)$, and γ be the workers' bargaining power. This formula is directly derived from a bargaining game where workers and firms bargain over employment and wages (the so-called strongly efficient bargaining; see, for example, Abowd and Lemieux (1993)). Then, write

$$E[w_i] = \mu_w = \mu_x + \gamma \mu_{\frac{QR}{L}}$$

assuming a constant γ. Next, let us rewrite worker's opportunity wage as

$$x_i = \mu_x + \xi_i$$

and firm's quasi-rent as

$$\frac{QR_{j(i)}}{L_{j(i)}} = \mu_{\frac{QR}{L}} + \rho_{j(i)}$$

then

$$\begin{aligned} w_i &= \mu_w + x_i - \mu_x + \gamma \left(\frac{QR_{j(i)}}{L_{j(i)}} - \mu_{\frac{QR}{L}} \right) \\ &= \mu_w \left(1 + \frac{x_i - \mu_x}{\mu_w} + \gamma \left(\frac{QR_{j(i)}}{\mu_w L_{j(i)}} - \frac{\mu_{\frac{QR}{L}}}{\mu_w} \right) \right) \\ &= \mu_w \left(1 + \theta_i + \gamma \frac{\rho_{j(i)}}{\mu_w} \right) \end{aligned}$$

So we may write a first-order approximation of our wage bargaining equation in the following log-separable format:

$$\ln w \approx \ln \mu_w + \theta_i + \psi_{j(i)} \tag{1}$$

where $\theta_i \equiv \frac{\xi_i}{\mu_w}$ and $\psi_{j(i)} \equiv \gamma \frac{\rho_{j(i)}}{\mu_w}$

Equation (1) has a simple format that can be directly estimated using the matched employer-employee data sources that we will use. This additive decomposition helps us identify our measures of person effects as opportunity wages and our measures of firm effects as real measures of the share of the quasi-rents that goes to workers. In this last component, two elements matter. The first γ, workers' bargaining power, is related to union behavior and power in the firms. The second $\frac{QR_{j(i)}}{L_{j(i)}} - \mu_{\frac{QR}{L}}$, the quasi-rent per worker, is related to product market competition in the various industries.

3 Statistical Models and Estimation

Our underlying statistical model can be expressed using the decomposition in AKM. Once this decomposition is estimated, we apply the formulae given therein to estimate the part of the inter-industry wage differential due to person and firm effects. A summary of the methodology is given in this section.

3.1 The AKM Decomposition

A simple linear statistical model, taken directly from AKM, is specified as:

$$\ln w_{it} = x_{it}\beta + \theta_i + \psi_{\mathrm{J}(i,t)} + \varepsilon_{it} \tag{2}$$

where x_{it} denotes the time-varying variables, θ_i the pure person effect, $\psi_{\mathrm{J}(i,t)}$ the pure firm effect, and ε_{it} the statistical residual. Note that the function $\mathrm{J}(i,t)$ gives the identifier for the dominant employer, j, of individual i at date t. In full matrix notation we have

$$y = X\beta + D\theta + F\psi + \varepsilon \tag{3}$$

where X is the $N^* \times P$ matrix of observable, time-varying characteristics (in deviations from the grand means); D is the $N^* \times N$ design matrix of indicators variables for the individual; F is the $N^* \times J$ design matrix of firm indicators variables for the firm effects for the employer at which i works at date t (J firms total); y is the $N^* \times 1$ vector of dependent data (also in deviations from the grand mean); ε is the conformable vector of residuals; and $N^* = \sum_{i=1}^{N} T_i$. We assume that ε has the following properties:

$$\mathrm{E}\left[\varepsilon \,|X, D, F\right] = 0$$

and

$$\mathrm{Cov}\left[\varepsilon_i, \varepsilon_m \,|D_i, D_m, F_i, F_m, X_i, X_m\right] = \begin{cases} \{\Sigma_{T_i}\}_i \,, \; i = m \\ 0, \text{ otherwise} \end{cases}$$

3.2 Industry Effects[1]

Industry is a characteristic of the employer. As such, the analysis of industry effects in the presence of person and firm effects can be accomplished by appropriate definition of

[1] This section is based upon the analysis in AKM.

the industry effect with respect to the firm effects. We call the properly defined industry effect a "pure industry effect." Denote the pure industry effect, conditional on the same information as in equations (2) and (3), as κ_k for some industry classification $k = 1, \ldots, K$. Our definition of the pure industry effect is simply the correct aggregation of the pure firm effects within the industry. We define the pure industry effect as the one that corresponds to putting industry indicator variables in equation (3) and then defining what is left of the pure firm effect as a deviation from the industry effects. Hence, κ_k can be represented as an employment-duration weighted average of the firm effects within the industry classification k:

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

where

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

and the function $\mathrm{K}(j)$ denotes the industry classification of firm j. If we insert this pure industry effect, the appropriate aggregate of the firm effects, into equation (2), then the equation becomes

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

or, in matrix notation as in equation (3),

$$y = X\beta + D\theta + FA\kappa + (F\psi - FA\kappa) + \varepsilon \tag{4}$$

where the matrix A, $J \times K$, classifies each of the J firms into one of the K industries; that is, $a_{jk} = 1$ if, and only if, $\mathrm{K}(j) = k$. Algebraic manipulation of equation (4) reveals that the vector κ, $K \times 1$, may be interpreted as the following weighted average of the pure firm effects:

$$\kappa \equiv (A'F'FA)^{-1} A'F'F\psi. \tag{5}$$

and the effect $(F\psi - FA\kappa)$ may be re-expressed as $M_{FA}F\psi$, where the column null space of an arbitrary matrix, Z, is denoted $M_Z \equiv I - Z\,(Z'Z)^{-}\,Z$, and $()^{-}$ is a computable g-inverse. Thus, the aggregation of J firm effects into K industry effects, weighted so as to be representative of individuals, can be accomplished directly by the specification of equation (4). Only rank$(F'M_{FA}F)$ firm effects can be separately identified using unrestricted fixed-effects methods; however, there is neither an omitted variable nor an aggregation bias in the estimates of (4), using either of the class of methods discussed below. Equation (4) simply decomposes $F\psi$ into two orthogonal components: the industry effects $FA\kappa$, and what is left of the firm effects after removing the industry effect, $M_{FA}F\psi$. While the decomposition is orthogonal, the presence of X and D in equation (4) greatly complicates the estimation using either fixed-effects or mixed-effects techniques.

When the estimation of equation (4) excludes both person and firm effects, as most of the literature has done, the estimated industry effect, κ_k^{**}, equals the pure industry effect,

κ, plus the employment-duration weighted average residual firm effect inside the industry, given X, and the employment-duration weighted average person effect inside the industry, given the time-varying personal characteristics X:

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

which can be restated as

$$\kappa^{**} = (A'F'M_X FA)^{-1} A'F'M_X F\psi + (A'F'M_X FA)^{-1} A'F'M_X D\theta. \quad (6)$$

The exact decomposition is exactly parallel to our theoretical model: the inter-industry wage differential is decomposed into two parts, a person and a firm component.

Notice that if industry effects, FA, were orthogonal to time-varying personal characteristics, X, and to non-time varying personal heterogeneity, D, so that $A'F'M_X FA = A'F'FA$, $A'F'M_X F = A'F'F$, and $A'F'M_X D = A'F'D$, the biased inter-industry wage differentials, κ^{**}, would simply equal the pure inter-industry wage differentials, κ, plus the employment-duration-weighted, industry-average pure person effect, $(A'F'FA)^{-1} A'F'D\theta$, or

$$\kappa_k^{**} = \kappa_k + \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1[\mathrm{K}(\mathrm{J}(i,t)) = k]\theta_i}{N_k}$$

where $N_k \equiv \sum_{i,t} 1[\mathrm{K}(\mathrm{J}(i,t)) = k]$.

3.3 Estimation of the Fixed-Effects Model by Direct Least Squares

The estimation methods proposed by AKM have been improved so that the statistical model can now be solved exactly for the fixed-effects case. The full solution and the associated identification analysis are reported in Abowd, Creecy, and Kramarz (2002), which is summarized in an appendix to the present paper.

4 Data Description

We constructed very similar data for both the United States and France. In particular, we used administrative earnings reports from longitudinally linked employer-employee data to estimate the inter-industry wage differentials and their decomposition into person and firm effects. Then, we assembled comparable American and French industry-level data to explain the decomposition. This section describes all of the data sources used.

4.1 United States

There are three main sources for the American data. Individual work histories come from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program infrastructure data base Abowd, Haltiwanger, and Lane (2004). The second source is the Current Population Survey. The final major source is the Bureau of Economic Analysis industry-level accounts. Each source is described below.

4.1.1 Unemployment Insurance Wage Record Information

The LEHD Program's infrastructure employer and employee database is described briefly below. The individual data were derived from the universe of unemployment insurance (UI) quarterly wage records from the following seven states: California, Florida, Illinois, Maryland, Minnesota, North Carolina, and Texas. Table 1 in Abowd, Lengermann, and McKinney (2003, ALM hereafter) details the coverage dates and number of individuals appearing in each of these states. In our analysis, by combining them into a single "pooled" file, we have information on approximately 68 million workers, accounting for over 45% of the U.S. workforce. ALM state:

> "The *Handbook of Methods* of the U.S. Department of Labor (p.42) describes UI coverage as "broad and basically comparable from state to state," and claims "over 96 percent of total wage and salary civilian jobs" were covered in 1994. The Federal Unemployment Tax Act (FUTA), first enacted in 1938, lays the ground rules for the kinds of employment which must be covered in state unemployment insurance laws. While technically mandating coverage of all employers with one or more employees in a calendar year, FUTA allows for numerous exceptions to covered employment Stevens (2000). These include workers at small agricultural cooperatives, employees of the Federal government, and certain employees of state governments, most notably elected officials, members of the judiciary, and emergency workers. According to the *Handbook*, UI wage records measure "gross wages and salaries, bonuses, stock options, tips, and other gratuities, and the value of meals and lodging, where supplied" (*Handbook of Methods*, p. 44). They do not include OASDI, health insurance, workers compensation, unemployment insurance, and private pension and welfare funds."

The LEHD data bases depend upon anonymized versions of the American Social Security Number (SSN) and identifiers in the UI account system maintained by the individual states. Both identifiers undergo extensive editing before the linked files are created. The SSN edit is discussed in Abowd and Vilhuber (2005). The UI account numbers are administratively edited by the individual states. Extensive successor-predecessor editing of the UI account numbers and the establishment identifiers occurs at LEHD (see McKinney (2002) and Benedetto (2004)).

Variable creation from the LEHD data Using Census Bureau and other LEHD infrastructure data bases, sex, date of birth, and education were integrated with the individual earnings data.[2] When a variable was created with an exact link to another database, the actual value from that data source was used. When a variable was created with a statistical link to another database or to correct for missing data, the value of the variable was imputed 10

[2] Sex, race, and date of birth are based on an exact match to administrative data. Education is based on a statistical match. See LEHD Program (2002), Abowd, Haltiwanger, and Lane (2004), Abowd and Vilhuber (2005) for details.

times by drawing from the posterior predictive distribution Rubin (1987), thereby providing information on the precision of the inferences based on incomplete data or statistical links. Upon each individual's first appearance in the data, labor force experience was recorded as potential labor force experience (age - education - 6). In subsequent periods, labor force experience was measured as the sum of observed experience and initial (potential) experience. The UI wage records connect individuals to every employer from which they received wages in any quarter of a given calendar year. Therefore, individual employment histories are constructed using the same personal identifier used in the individual data. Employers are identified by their state unemployment insurance account number (SEIN). While large employers undoubtedly operate in multiple states, their SEINs are state specific, meaning they cannot be connected. In addition, while the LEHD system matches workers to their employers, it is not possible to connect those employed in firms with multiple establishments to the specific places of work. The employer definition, therefore, resembles the one used for France.

Earnings For every year an individual appears in a state, a "dominant" employer–the employer for whom the sum of quarterly earnings is the highest–was identified in order to better approximate the individual's full-time, full-year annual wage rate using the following steps. The complete algorithm is discussed in detail in (Abowd, Lengermann, and McKinney 2003). We describe only its essential features here so that the reader can understand its principles. The concept of full-quarter employment is used extensively. Full-quarter employment occurs in quarter t in an employment history when there are positive earnings for quarters $t-1, t$, and $t+1$. Continuous employment during quarter t means having an employment history with positive earnings for either $t-1$ and t or t and $t+1$. Employment spells that are neither full-quarter nor continuous are designated discontinuous. If the individual was full-quarter employed for at least one quarter at the dominant employer, the annualized wage is computed as the appropriate multiple of the average full-quarter earnings actually observed. For example, if the individual was full-quarter employed for all four quarters at his dominant employer, then our annualized earnings measure is just the sum of those four quarters. We account for 84% of the person-year-state observations in our eventual analysis sample using this part of the algorithm. Otherwise, if the individual was continuously employed for at least one quarter at the dominant employer, the annualized wage is average earnings in all continuous quarters of employment at the dominant employer multiplied by 8 (i.e., 4 quarters divided by an expected employment duration during the continuous quarters of 0.5). Continuous employment cases account for 11% of the analysis observations. For the remaining 5%, annualized wages are average earnings in each quarter multiplied by 12 (i.e., 4 quarters divided by an expected employment duration during discontinuous quarters of 0.33).

Our analysis sample is restricted to individuals aged 18-70, employed full-time at their dominant employer. Table 2 in Abowd, Lengermann, and McKinney (2003) presents sample means for several earnings, demographic, industry, and labor force attachment variables for the period 1990-1999. The final analysis sample contains 279 million person-year-state observations for the aforementioned 68 million individuals and 2.5 million firms. See Abowd, Lengermann, and McKinney (2003) for a detailed analysis of the differences between this sample and other samples based on LEHD and CPS data.

4.1.2 Current Population Survey Variables

For our analysis of the long-run decomposition of the inter-industry wage differential, we required SIC-level estimates of characteristics of the work force. These were derived from the pooled Outgoing Rotation Group public use samples from January 1995 to December 1999, years chosen to be symmetric around 1997, which is the focus year for our analysis of the U.S. data. We used all full-time workers age 16 or over. From the pooled outgoing rotation group sample, we re-coded CPS industries into 2-digit SICs. These re-coded SICs form the basis for the industry level statistics that were computed. Variables summarized at the SIC level are discussed below.

Average age is the average of all full-time workers in the SIC. When age groups are used as a control these consist of the proportion of SIC employment in each of three categories: 16-24, 25-54, and 55 or over. Education is the proportion of SIC employment in each of five educational categories: less than high school, high school, some college, college graduate, some post-graduate. Occupation is the proportion of SIC employment in each of eight one-digit occupation codes based on aggregates of the three-digit Census Occupation Code (1990). Unionization rates were computed as the proportion of SIC employment in jobs covered by a collective bargaining agreement.

4.1.3 Bureau of Economic Analysis Variables

In order to compute the quasi-rent per worker for each SIC we used the BEA GDP by Industry and Tangible Wealth statistics from 1997 supplemented with variables from other sources discussed in the next subsection. The components of the quasi-rent per worker were estimated from the BEA SIC-based GDP data. Industry value added at the two-digit SIC level was divided by industry full-time equivalent employment to get value-added per worker. Total compensation for the SIC was estimated using the SIC-based total labor cost, inclusive of benefits, from this source. Total capital was estimated using BEA estimates for fixed tangible wealth. The capital/labor ratio was estimated using the ratio of fixed tangible wealth to total full-time equivalent employment.

The BEA reports two-digit SIC data with some aggregations. Weights were created to disaggregate the fixed asset and other BEA industry data. Reported wealth data were aggregated in groups for industries (A07, A08, A09), (C15, C16, C17), (F51, F52), (G52 to G59), and (I83, I84, I86, I87, I89). All other BEA industry data were aggregated in groups for industries (A07, A08, A09), (C15, C16, C17), (F51, F52), (G52 to G59), and (I84, I87, I89). Industry employment from the 1997 Economic Census was used to disaggregate these data.

4.1.4 Other Sources

Economic Census 1992 and 1997. Employment and sales were taken from the official Census web site for the 1992 and 1997 Economic Censuses. Employment for A07, A08, A09 was out of scope for 1997 and was obtained from the 1997 County Business Patterns data available on the Census web site. Employment for E40, also out of scope in 1997, was obtained from the 1998 OES National 2-digit SIC Industry Estimates, which are produced by the Occupational Employment Statistics Survey, Bureau of Labor Statistics.

Sales data in some industries were suppressed for 1997. When the correct value was available for 1992, the 1992 value (converted to 1997 dollars) was used to impute values in 1997. For suppressions in industries D, E, G, and H, data in 1992 were always available in those industries where data were suppressed in 1997. The imputation proceeded as follows. For 1997, calculate the total missing employment as division total employment less the sum of industry employment where available. For 1992, calculate the total employment in those industries where employment is missing in 1997. Assume the proportions are the same in 1997, and use these proportions to allocate total missing employment in 1997.

The Economic Census data for industries in division I were disaggregated between taxable and nontaxable establishments. We treated these separately. For the taxable data, employment and sales were suppressed only in industry I89 (suppressed as not comparable or not available). We computed employment and sales directly as taxable amount in I89 equals the total taxable amount in all industries less the sum of the available taxable amounts in the other industries apart from I89. This amount will therefore contain some employment and sales in other SICs due to the nature of the 1997 suppressions. Note that the imputed values for employment and sales are close to the reported values available for 1992.

Herfindahl Index for the seven states. Custom two-digit SIC-based Herfindahl indices were produced from the 1992 and 1997 Economic Censuses using establishment-level micro data for the seven states in the analysis sample.[3]

Imports and Exports. These data were taken from the U.S. International Trade Commission interactive tariff and trade data base. Missing two-digit industries were set to zero.

4.2 France

The administrative record earnings data used for France are an updated version of the data used by AKM. We describe the essential features of these updated data below. Then, we describe the sources for the industry-level variables used in our analysis of the long-term inter-industry wage differentials and their decomposition.

[3]These computations were performed at the Census Bureau using the LEHD infrastructure links to the Economic Censuses.

4.2.1 The Déclarations Annuelles des Données Sociales

We use data from the Déclarations annuelles des données sociales (DADS), a 1/25th sample of the French work force with information from 1976-1996. The DADS are matched employer-employee information collected by INSEE (Institut National de la Statistique et des Etudes Economiques) and maintained in the Division des Revenus. The data are based on a mandatory employer report of the gross earnings of each employee subject to French payroll taxes. The universe includes all statutory employed persons. Our analysis sample covers all individuals employed in French enterprises who were born in October of even-numbered years, with civil servants excluded. Our analysis sample runs from 1976 through 1996, with 1981, 1983 and 1990 excluded because the extracts from the master payroll records were not built for those years. The initial data set contained 16 millions observations. Each observation corresponds to a unique enterprise-individual-year combination. The observation includes an identifier that corresponds to the employee (called NNI below) and an identifier that corresponds to the enterprise (SIREN). For each observation, we have information on the number of days during the calendar year the individual worked in the establishment, as well as the full-time/part-time work-status of the employee. Each observation also includes, in addition to the variables listed above, the sex, year and place of birth, total net nominal earnings during the year and annualized gross nominal earnings during the year for the individual, as well as the location and industry of the employing establishment.

4.2.2 Observation selection, variable creation, and imputation

An observation is identified by a combination of two identifiers, the firm ID and the person ID. The SIREN number has an internal structure that allows a check for coding errors. But, the NNI number has no such internal control. Although 90% of current DADS information is filed by the responding firm using an electronic medium (tape or diskette), the situation in the eighties was quite different. In that era, INSEE had to perform data entry by key punch from paper forms. Entry errors in the NNI occurred (exchange of two digits of the NNI, error in one of the digits, *etc.*). This phenomenon is well-known at INSEE but, despite many attempts, no general way of solving this problem was found. As a consequence, some observations have a NNI-year-SIREN combination such that no other observation has the same NNI. As a joint product, some NNI-SIREN combinations have a unique missing year. Consider now the case of a worker with observations in, say, 1978 and 1980 in the same enterprise (SIREN) but no observation for 1979. If true, this history would mean that the worker was employed until some date in 1978 (depending on the number of days worked, December 31 most likely) and also employed after some date in 1980 (depending on the number of days worked, January 1 most likely) in this firm but not employed at all during year 1979. This is very improbable. In particular, because there is a layoff procedure in France in which workers may be recalled by their firms after some period of unemployment. Suggestions of D. Verger (head of the Division Revenus, in charge of the DADS at the beginning of the nineties) led us to adopt the following solution. Whenever an observation was missing in a given year while the same NNI-SIREN combination exists for the preceding and the following year, we created an observation for the missing year with the same NNI-SIREN combination. (This added 193,148 observations). Earnings are computed as the geometric mean of the preceding and following wages (in real terms). All other variables are taken at their preceding year value.

Because of the 1982 and 1990 Census, the 1981, 1983 and 1990 DADS data were not available. We used the same principle as the one described above to impute missing observations. Hence, imputation was performed only for those individuals that were present in the same firm in 1980 and 1982 or 1982 and 1984 or 1989 and 1991. (This added 759,017 observations to the sample). All variables were imputed as above.

As in the United States, in each year,we kept the dominant employer, i.e. the one for which total earnings were the highest, and we annualized these earnings using days of pay. Finally, as in AKM, we eliminated observations for which the logarithm of the real annualized total compensation cost was more than five standard deviations away from its predicted value based on a linear regression model of this variable on sex, region, experience, and education (see once more the data appendix in AKM).

Having done all these selections and imputations, the final data set that we use contains 13,759,302 observations, corresponding to 1,732,169 individuals and 787,795 firms.

4.2.3 Other Sources

French Labor Force Survey (1984-1989). To measure the skill, education, and age structure of the French industries in the middle of our sample period, we use the micro-data extracts from the labor force surveys for the years 1984 to 1989, inclusive. We keep all individuals in any of those years with non-missing information on industry affiliation (equivalent to 2-digits SIC codes), non-missing information on age, and non-missing information on education. We define three age groups: 15 to 29, 30 to 49, and 50 and above. We define 8 education categories similar to those defined in AKM. Then, we compute the average age-education structure for each industry over the six years. We proceed similarly to compute the average skill structure for each industry using the 2-digit French PCS (classification of occupations). This classification distinguishes 31 skill groups.

Wage Structure Survey (1986). The 1986 wage structure survey (enquête structure des salaires) is an establishment-level survey that measures a wealth of indicators pertaining to workplace organization, pay structure, unionization, hours structure (such as work in shifts), existence of bargaining agreements,...The Wage Structure Surveys (Enquête sur la Structure des Salaires, ESS hereafter), conducted jointly by the French National Statistical Institute (INSEE) and the Ministry of Labor, 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 given the usefulness and quantity of information collected during each wave. The 1986 ESS collects establishment information for a sample of establishments in the manufacturing, construction, and (some) service industries. More specifically, the universe to be sampled includes all establishments (manufacturing) or firms (construction and service) with at least ten employees. Agriculture, transportation, telecommunication and the services supplied to households are excluded from the scope of the ESS. Insurance companies, banks, and all other industries where services are supplied to businesses are in the scope of the survey. The universe is derived from the SIRENE system, a unified database recording all existing establishments and firms in France. The sampling rate is stratified according to the industry, region, and the size of the unit. Sampling rates vary from 1 (certainty) for the units above 500 employees to 1/48 for units between 10 and 20 employees. Our sample includes 16,239 observations. The resulting file therefore includes weights in order to make the data representative of the universe of establishments of the private and semi-public sectors. (More detailed technical information on the 1986 version of the ESS is available in Rotbart (1991)).

In this paper, we construct the following variables at the industry level: fraction workers in shifts, fraction blue-collar workers in 2 (*resp.*, 3, 4, 5) shifts, fraction of other workers in 2 (*resp.*, 3, 4, 5) shifts, fraction workers (total, blue-collar, other) working at night, fraction of establishments covered by an establishment or a firm-level bargaining agreement, fraction workers covered by an establishment or a firm-level bargaining agreement, fraction of establishments covered by a collective bargaining agreement (normally national or specific to the industry), fraction of workers covered by a collective bargaining agreement, fraction workers with pay increases tied to establishment or firm-level (*resp.*, collective) bargaining agreement, fraction workers with pay increases tied to individual or group performance measures, and total employment.

Ministry of Labor Data on Agreements and Union Representatives. The Ministry of Labor statistical services (DARES) collect, not on a regular basis, data on agreements as well as on union elections and representatives. The sources available at dates closest to the middle of the period of analysis that we have selected measure newly signed agreements in 1990 and results of union elections and measures of union representatives in 1989. These data sources were directly provided to us at the 2-digit industry level. More precisely, we measure the number of newly signed agreements in the industry during year 1990, the average number of ensuing general wage increases, the average number of ensuing collective wage increases at the bottom of the pay scale, the average number of ensuing collective wage increases at the top of the pay scale, and the average number of ensuing individual wage increases. From the union elections file, we use the fraction voting for CGT (confédération générale du travail, close to the Communist party), CFDT (confédération française démocratique du travail, close to Socialist party), FO (force ouvrière, strong Trotskyite tradition), CFTC (confédération française des travailleurs chrétiens, catholic tradition with no clear political affiliation), CGC (confédération générale des cadres, for managers, engineers, professionals). Notice that some "cadres" may well vote for CGT or CFDT and not necessarily for CGC. The data source also measures the average number of union representatives (délégués syndicaux) in establishments covered by the survey as well as the prevalence of CHSCT (comité d'hygiène et de sécurité et de conditions de travail) in each of the industries. These committees, present in establishments with at least 50 employees, fulfil tasks such as protection of health and safety, improving working conditions in particular for women.

BIC-BRN files. The BIC-BRN (Bénéfices Industriels et Commerciaux, bénéfices réels normaux) is a fiscal source compiled and controlled at INSEE. It includes the balance-sheets

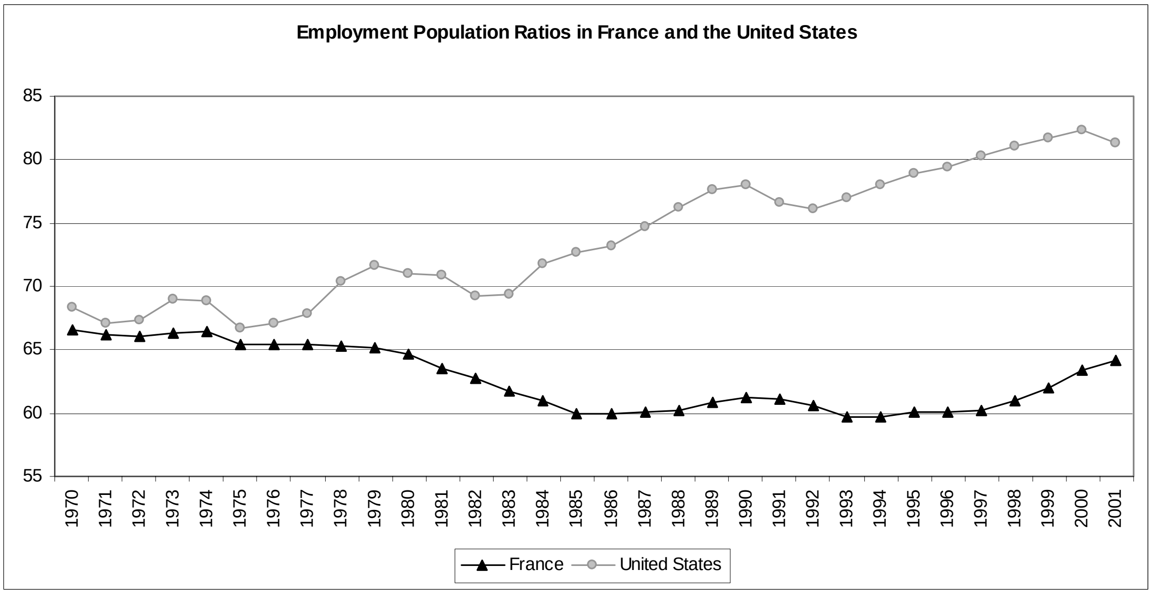

Figure 1

of almost the full universe of French firms. In particular, it includes a very large number of small firms that declare their accounts under the normal régime (as opposed to the "Forfait" used mostly by self-employed and to the "Bénéfices réels simplifiés" also employed by very small firms, often with zero or a unique employee). The normal regime is mandatory for all firms from the manufacturing industries with sales above 500,000 US dollars or from the service industries with sales above 150,000 US dollars. Hence, all measures that are computed from this file are based on virtually all firms with at least one employee. Each year, there are roughly 600,000 firms in the data set. From this source, we compute the following variables: sales concentration ratios; sales Herfindahl indices; ratio of exports to sales; total capital stock, total employment, total sales. All of these variables are for 1986, our reference year.

5 Comparison of Institutions

The Labor Markets. For most analysts, the difference in the labor market situations between France and the United States is well summarized by Figure 1. This figure shows the employment-population ratio for the two countries for our sample period. The two countries have similar employment rates in the mid-seventies (65% and 67% , respectively). But thereafter, they diverged. At the end of our sample period (the end of the nineties), the difference was close to 20 percentage points.

Since 1951, French industry has been subject to a national minimum wage (called the SMIC since the revisions to the relevant law in 1971) that is indexed to the rate of change

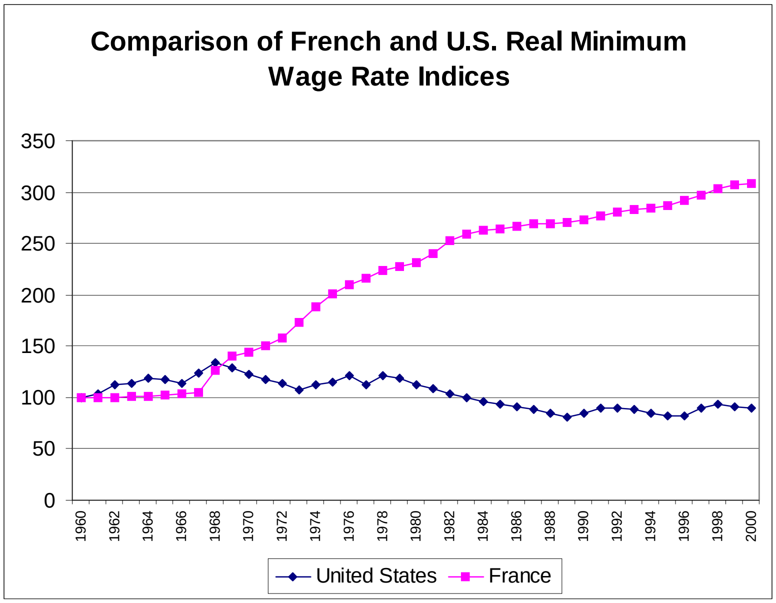

Figure 2

in consumer prices and to the average blue-collar wage rate. The United States has also had a federal minimum wage during this same period. The American federal minimum wage is superseded by state-mandated rates during some years for some states. Figure 2 depicts the changes in the two (real) minima over the sample period. Exactly when the French SMIC started its very sharp increase (beginning of the seventies), the American minimum decreased rapidly. In the rest of the sample period, the French SMIC continued its increase, partly mandated by one-shot increases and partly by formulaic increases, whereas the American minimum wage fell until 1990 when it levelled off. Notice, however, that minimum wage rates delivered to the worker do not present the firm's minimum labor costs. The structure of payroll taxes that augment wages as a part of labor cost has not changed substantially in the United States (during the period of our analysis) but it has changed in France. After a constant increase in payroll tax rates from the early 1970s, they dropped sharply in 1994 and even more so in the ensuing years (see Kramarz and Philippon (2001)) as a part of an explicit program to lower total labor costs for workers at the minimum wage.

During the sample period, both the French and the United States labor market institutions were also characterized by important changes in the bargaining institutions and environment. In the 1970s, centralized collective bargaining agreements (convention collective de branche) were the basic elements of the negotiation process in France. The different industrial sectors had collective agreements that were negotiated by groups of unions and employer associations. These agreements were binding on the negotiating parties. The complete agreement was then typically extended to cover the entire industry (or region) by the Ministry of Labor and was then made binding on workers and firms that were not party to the original negotiation (Cahuc and Kramarz (1997)). More than 95% of the work force was covered by these collective bargaining agreements at the end of the 1980s, while union membership was approximately 10%. The collective agreements specified a set of minimum wages and wage progressions for the occupational categories covered by the negotiations (sometimes called a wage grid). But, beginning in 1982, the "lois Auroux" (a set of revisions to the body of labor law named after the Minister of Labor at the time) required firms with at least 50 employees to negotiate firm-level collective agreements (accords d'entreprise). Although firms were not explicitly obligated to conclude an agreement, 65% of the work force were employed at establishments or businesses where firm-level negotiations occurred either through the union delegates or some other worker representative.[4] Among this 65% of the work force, only three-quarters of the workers ended up with an agreement as a result of these negotiations. Finally, the percentage of the work force covered by some establishment or firm-level agreement on wages is approximately 40% in 1992. The law required that the firm-level agreements could only improve the conditions stated in the industrial agreement, so that, over time, the firm-level agreements have become more important for wage determination than the industry agreements.

Although more than 90% of French workers are covered by industrial agreements throughout our analysis period (1976-1996), firm-level negotiations outpaced renegotiations of industry-wide agreements in most industries. The regular increases in the national minimum wage (in particular those driven by the indexation to the average blue-collar wage rate) resulted in the lowest categories on the collective pay scales in most industry contracts for most occupations being below the national minimum by the beginning of the 1990s. When this occurs, it is the national minimum wage, and not the collectively bargained wage, that binds.

In the United States, the bargaining environment is very different from that prevailing in France. For this description as well as for the recent changes that affected this country, we rely on the recent work by (Farber and Western 2001). Unionization in most sectors of the United States economy is governed by the National Labor Relations Act (NLRA), passed in 1935. While NLRA guaranteed the rights of employees (as distinct from workers or citizens, in general) to organize and bargain, it also established the procedure for a union to become the exclusive bargaining agent of a group of workers. The procedure starts when at least 30% of the workers in a work unit sign authorization cards. The union petitions the National Labor Relations Board (NLRB) in order to set up a representation election. The campaign takes place between the time of the petition and the election. Employers and unions both participate in the campaign. Once a union is certified and an initial contract negotiated, the jobs included in the bargaining unit become union jobs. Successor employers are normally bound to negotiate with the duly certified union unless there has been a decertification election.

Figure 4 of Farber and Western (2001) shows the huge decrease in certification elections that took place during the eighties. Furthermore, even for those elections that took place, the union win rate also decreased sharply until 1975 and then levelled off at just less than 50% (Figure 5, id.). Not surprisingly, these facts, along with the decline in manufacturing

[4] The labor market computations in this section were performed by the authors using the 1986 and 1992 wage structure surveys.

employment generally, contributed to a sharp decline in the overall unionization rate among private employers. Although the trend went in the opposite direction for public-sector employers, the overall unionization rate which went from 25% in 1974 to less than 13% in 1998. Although the private and the public sector union membership rates were equal in 1974 at 25%, the latter increased rapidly to 36% in 1980 and then stabilized but the former continuously decreased to 9.7% in 1998.

To further assess potential differences in wage setting, Abowd, Kramarz, and Margolis (1999) ran two simple wage regressions using comparable household surveys (the Enquête Emploi for France and the Current Population Survey for the U.S.)[5]. Their results show that the same set of regressors has more or less the same explanatory power for wages in both the French and American data (roughly 37% for men in both countries, 32% for women in France and 24% in the U.S.). Returns to one additional year of education were 6.1% for men and 7.2% for women in the U.S. while they are 7.7% for men and 8.8% for women in France, with the difference between the sexes being identical. Returns to experience differ slightly, with the curvature of the quartic in experience implying a more hump-shaped profile in the U.S.. Finally, the gender wage gap in the initial year is roughly equal in both countries, although it decreases over the sample period in the U.S. and is basically stable in France during the eighties.

The Product Market. Apart from labor market institutions, the intensity of competition prevailing on the product market should affect wages according to the theory developed above because wages are the sum of an opportunity wage and a rent component. These rents are the product of workers' bargaining power, an outcome of labor market institutions as described just above, and of the firms' quasi-rent, an outcome of product market competition. Regulations that affect product markets are more dispersed than labor market regulations. For instance, the American airline or banking industries were heavily regulated in the seventies. But these regulations were discarded during the 1980s and 1990s. The effects on labor market outcomes of such deregulation have been shown to be quite large (see Peoples (1998) for a review of this literature in the United States, see Card (1986) for the airline industry; Black and Strahan (2001) for the banking industry). To paraphrase the conclusion of Peoples (1998), the effect of deregulation was heavily tied to reductions in the labor costs that followed. In all the industries where labor earnings fell sharply, trucking or airline, employment increased dramatically. In industries where labor earnings fell slightly, such as telecommunications, employment was steady. And, finally, in industries where earnings did not change, such as the railroad industry, employment sharply declined. Hence, our sample period for the United States is one of intense product market competition.

This is far from the case in France. Even though France, pushed by European institutions, started in the 1990s to deregulate some industries, the process is far from completion. During our sample period, near monopolies operated in many industries. Air France (airlines), Seita (cigarettes), Electricité de France (energy), and Gaz de France (energy) are all examples of firms in which the State has a majority equity stake and there are no local competitors

[5] Similar results are also found using cross-sections of matched worker-firm data for the two countries (see Abowd, Kramarz, Margolis, and Troske (2001)).

(even though France imports cigarettes and allows foreign airlines to land in France). Entry into these industries was, and still is, heavily regulated. Surprisingly, it is also the case in many other apparently competitive industries, such as the retail trade, that entry regulations loomed and are still very important (see Bertrand and Kramarz (2002) for the detrimental effect of the Loi Royer on employment in the retail trade). Djankov, Porta, de Silanes, and Shleifer (2002) have also shown that entry regulations, as measured by requirements to starting a new business in France, are common, time-consuming and costly. This startup process takes 66 days and 16 different legal and administrative steps in France and only 7 days and 4 steps in the United States.

6 Basic results

We estimated all of the identifiable person and firm effects from equation (3) and calculated the decomposition shown in equation (6). The complete set of results is shown in Appendix Table A1 (for the United States) and Table A2 (for France). The column labeled "Raw Industry Wage Differential" is the estimated value of κ^{**}, controlling for all of the characteristics in X as well as the observable non-time-varying characteristics that are (implicitly) controlled by D, using the same specifications as in AKM but our newer estimation method. The column labeled "Industry Average Person Effect" is an estimate of $(A'F'M_XFA)^{-1}A'F'M_XD\theta$, the average of the person effects within the industry adjusted for X. The column labeled "Industry Average Firm Effect" is an estimate of $(A'F'M_XFA)^{-1}A'F'M_XF\psi$, the average of the firm effects within the industry adjusted for X. The columns of these two tables serve as the dependent variables for the long-term statistical analysis performed in section 7.[6]

An instructive way to summarize the results in these tables is to consider the industries with the largest positive and negative raw wage differentials. For the U.S. these "winners and losers" are summarized in Table 1. Notice that the industry with the largest raw differential is SIC 62 (security brokers, dealers and exchanges). This raw differential of 0.659 log points consists of two-thirds person effect and one-third firm effect. One can conclude that for whatever economic reasons, persons with high opportunity costs of time (large θ_i) accumulate in this sector and, again for whatever economic reasons, firms in this sector pay more. SIC 62 has both high-wage workers and high-wage firms. Now consider the second largest "winner," SIC 46 (pipelines, except natural gas). In this sector the large differential is is basically due to the presence of high-wage firms (large ψ_j). Consider now the "losers." The biggest negative raw differential is in SIC 88 (private households), where more of the difference is due to firm effects than to person effects but the two both contribute

[6] Note that the industry average person effect plus the industry average firm effect does not sum exactly to the industry average raw effect due to the fact that the models in the United States were estimated separately state-by-state then aggregated to the national level using weights representative of 1997 and a decomposition of the state-effect into a part due to state-average person effects, a part due to state-average firm effects, and an unexplained part (see ALM). For any given state, the decomposition is exact. For France the small discrepancy is due to the estimation of person and firm effects for all observations and the presence of a small amount of missing industry data.

Table 1: US, Winners and Losers

SIC	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
62	Security brokers, dealers, exchanges	0.659	0.393	0.258
46	Pipelines, except natural gas	0.625	0.094	0.475
29	Petroleum and coal products	0.478	0.109	0.345
81	Legal services	0.471	0.236	0.230
48	Communications	0.460	0.123	0.346
49	Electric, gas, and sanitary services	0.445	0.095	0.358
13	Oil and gas extraction	0.437	0.108	0.285
38	Instruments and related products	0.411	0.137	0.280
89	Miscellaneous services	0.409	0.136	0.227
28	Chemicals and allied products	0.407	0.126	0.303
83	Social services	-0.296	-0.199	-0.102
53	General merchandise stores	-0.303	-0.050	-0.241
79	Amusement and recreation services	-0.312	-0.079	-0.249
72	Personal services	-0.371	-0.160	-0.213
70	Hotels, rooming houses, camps, and lodging	-0.385	-0.209	-0.186
23	Apparel and other textile products	-0.402	-0.240	-0.142
58	Eating and drinking places	-0.554	-0.161	-0.397
07	Agricultural services	-0.568	-0.280	-0.316
01	Agriculture-crops	-0.570	-0.293	-0.315
88	Private households	-0.643	-0.219	-0.252

substantially to the differential. SIC 83 (social services), on the other hand, is due two-thirds to the average person effect and one-third to the average firm effect.

Table 2 shows the "winners and losers" for France. The first point to remark, as our discussion in section 5 previewed, is that there is noticeably less dispersion of the raw inter-industry wage differentials in France as compared to the U.S.–probably because of the combined effects of the SMIC (on the bottom of the wage distribution) and the government-ownership of the largest businesses (on the top of the wage distribution). Nevertheless, there are some big "winners." The biggest is NAP 05 (crude petroleum and natural gas extraction) where the raw effect consists of about equal parts person and firm effects. The story is very different for the second "winner." In NAP 27 (office and accounting machines) most of the effect comes from the average firm effect. Similarly, NAP 72 (air transport) and NAP 42 (tobacco products) and NAP 07 (distribution of gas) all derive the bulk of their raw differential from the industry average firm effect. Interestingly, all of these industries contained large government-owned firms during the sample period. Among the "losers" the most striking feature is the compression of the industry average person effects (especially in comparison with the U.S.), which is almost surely due to the SMIC. The largest "loser" NAP 92 (teaching, non-market) is a very small sector. The largest of the "loser" market sectors are NAP 67 (hotels, motels, bars and restaurants) and NAP 62 (retail specialty and neighborhood food), which employ 7% of the total workforce, have large negative raw differentials consisting of two-thirds firm effect and one-third person effect.

Krueger and Summers documented a correlation across nations of the raw inter-industry wage differentials. We show in Table 3 that both the raw differential and its components are strongly correlated between the U.S. and France. The international correlation of the

Table 2: France, Winners and Losers

NAP	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
05	Crude petroleum and natural gas extraction	0.490	0.239	0.196
27	Office and accounting machines	0.472	0.094	0.346
72	Air transportation	0.466	0.123	0.306
76	Financial holding companies	0.378	0.205	0.130
42	Tobacco products manufacture	0.337	0.016	0.256
07	Distribution of Gas	0.319	0.059	0.229
33	Aircraft and parts manufacture	0.313	0.167	0.098
04	Coal mining	0.312	0.098	0.170
94	Health care, non-market	0.301	-0.105	0.376
17	Basic chemical manufacture	0.289	0.100	0.144
67	Hotels, motels, bars and restaurants	-0.167	-0.047	-0.068
62	Retail specialty and neighborhood food	-0.177	-0.048	-0.085
90	Public administration, general	-0.218	-0.035	-0.167
38	Bakery products	-0.233	-0.019	-0.133
87	Miscellaneous commercial services	-0.239	-0.076	-0.137
95	Social services, non-market	-0.241	-0.099	-0.150
96	Recreational, cultural, and sporting, non-market	-0.252	-0.024	-0.214
82	Commercial education services	-0.259	0.044	-0.321
97	Miscellaneous public services, non-market	-0.288	-0.025	-0.255
92	Teaching, non-market	-0.414	0.000	-0.431

effects does not depend upon which country we select to supply the base weights although it is slightly stronger when weighted by French industry shares than when we use the U.S. industry shares as weights. The smallest correlations on the table are between the industry average person effect and the industry average firm effect (within or between countries). This correlation is positive but it is not very large, suggesting that the forces that sort person effects are not strongly correlated with the forces that sort firm effects.

7 Long Term Factors

In order to implement the wage determination model embodied in equation (1), we estimated normalized quasi-rents per worker using the industry aggregate data described in section 4. The exact definition of the normalized quasi-rent per worker is for industry k is:

$$\text{Normalized}\left(\frac{\text{Quasi-Rent}}{\text{Worker}}\right)_k$$

$$= \frac{\left(\frac{\text{Value Added}}{\text{FTE Employment}}\right)_k - 0.03\left(\frac{\text{Fixed Tangible Assets}}{\text{FTE Employment}}\right)_k - \left(\frac{\text{Total Compensation}}{\text{FTE Employment}\times\exp(\psi)}\right)_k}{\sum_k \frac{\text{Total Compensation}}{\text{FTE Employment}}}$$

where the appropriate, country-specific, value has been used for each of the variables noted in the general formula. Notice that the division of the total compensation per FTE employee by $\exp(\psi_k)$ removes the non-portable part of the compensation from the opportunity cost of labor. As is clear from the formula, we assumed a 3% real opportunity cost of capital.

Table 3: Correlation of Industry Effects						
	France			*US*		
	Raw Industry Wage Differential	*Industry Average Person Effect*	*Industry Average Firm Effect*	*Raw Industry Wage Differential*	*Industry Average Person Effect*	*Industry Average Firm Effect*
France			***French Industry Weights***			
Raw Industry Wage Differential	1.0000	0.6110	0.9046	0.5783	0.4444	0.5689
Industry Average Person Effect	0.6110	1.0000	0.2549	0.4810	0.6337	0.2880
Industry Average Firm Effect	0.9046	0.2549	1.0000	0.3792	0.1655	0.4564
US			***US Industry Weights***			
Raw Industry Wage Differential	0.4914	0.3652	0.3630	1.0000	0.8167	0.9214
Industry Average Person Effect	0.2662	0.4631	0.0630	0.8167	1.0000	0.5382
Industry Average Firm Effect	0.5500	0.2024	0.5065	0.9214	0.9214	1.0000

Table 4 presents a summary of the variables we use to explain the raw inter-industry wage differentials and their component industry-average person and firm effects in our statistical analysis of the long run determinants.

Table 5 shows the results with no controls other than the variables listed. The first six columns present estimates for the United States. The last six columns present equivalent results for France. For each country, we estimate three specifications of the explanatory variables. For industry k, the industry-average person and firm effects, θ_k and ψ_k constitute our two endogenous variables in each specification. These average person and firm effects are estimated using the methods described in section 3. In each case, we include a measure of the (normalized) industry quasi-rent per worker.

In all specifications and as expected, the quasi-rent explains much more of the variance in the firm effects than in the person effects. This difference is more marked in the French regressions than in the American regressions. This result supports the idea that bargaining is more prevalent in France than in the United States, an unsurprising feature when comparing the wage-setting institutions. Indeed, union strength helps explain the firm effect and not the person effect in both countries.

The two regressions contained in the second set of columns have different specifications because union presence takes different forms in the two countries. In the US, we use employment in a job covered by a collective bargaining agreement as the measure of unionization. In France, we use various measures of the bargaining outcomes in the industry: types of wage increases (industry-wide, individual-based, targeted to the low-wage, and targeted to the high-wage). We also use measures of union presence in the firms under its various legal guises, existence of union representatives, and existence of a health and security commission (CHSCT) as unionization measures. Finally, to measure the type of orientation of unions that were in charge of negotiating with the employers associations, we include fractions of votes received by the nationally representative unions: the CGT, close to the communist party; the CFDT, reformist and closer to the socialist party; FO, which went from reformist to more extremist and strike-prone; and the CGC, the union for professionals, managers, and engineers. In the United States, even though unions were quite weak during our sample

TABLE 4: Descriptive Statistics of Indpendent Variables

	United States		France	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
Normalized quasi-rent	1.048	1.512	0.708	0.908
Herfindahl index	0.009	0.022	0.116	0.227
Export-sales ratio	0.062	0.097	0.168	0.162
Capital-labor ratio	193.84	445.85	511.99	679.81
Union	0.149	0.131		
Wage increases (industry-wide agreement)			2.532	0.860
Wage increases (individual-based)			1.357	1.112
Wage increases (targeted to the low-wage)			2.512	1.420
Wage increases (targeted to the high-wage)			1.777	1.325
Votes for the CGT union			0.868	0.846
Votes for the CFDT union			0.770	0.795
Votes for the FO union			0.652	0.794
Votes for the CGC union			0.524	0.608
Union delegates			3.548	4.461
Health and Security Commissions (CHSCT)			0.984	0.672

TABLE 5: Regressions of Person and Firm Effects on Economy-Wide Factors (no controls)

	United States						France					
Variable	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect
Normalized quasi-rent	0.0750	0.1632	0.0739	0.1576	0.1158	0.2077	0.0272	0.0913	0.0309	0.0693	0.0218	0.0531
	(0.0164)	(0.0237)	(0.0166)	(0.0231)	(0.0261)	(0.0369)	(0.0101)	(0.0157)	(0.0088)	(0.0121)	(0.0102)	(0.0137)
Herfindahl index					-2.9045	1.0575					-0.0221	0.1362
					(2.5039)	(3.5346)					(0.0465)	(0.0621)
Export-sales ratio					0.2137	0.3395					0.0658	0.1715
					(0.1520)	(0.2146)					(0.0485)	(0.0648)
Capital-labor ratio					-0.2730	-0.3840					0.0208	0.0034
					(0.1291)	(0.1823)					(0.0111)	(0.0149)
Union			0.0739	0.3919	0.1209	0.3622						
			(0.0166)	(0.1668)	(0.1343)	(0.1895)						
Wage increases									-0.0309	-0.0436	-0.0289	-0.0324
(industry-wide agreement)									(0.0090)	(0.0125)	(0.0094)	(0.0125)
Wage increases									0.0140	0.0109	0.0140	0.0149
(individual-based)									(0.0058)	(0.0080)	(0.0060)	(0.0080)
Wage increases									-0.0014	0.0147	-0.0032	0.0261
(targeted to the low-wage)									(0.0069)	(0.0095)	(0.0085)	(0.0113)
Wage increases									0.0020	-0.0406	0.0034	-0.0381
(targeted to the high-wage)									(0.0058)	(0.0080)	(0.0058)	(0.0077)
Votes for the CGT union									0.0015	0.0871	-0.0107	0.0084
									(0.0303)	(0.0418)	(0.0355)	(0.0475)
Votes for the CFDT union									0.0508	0.0407	0.0480	0.0704
									(0.0199)	(0.0275)	(0.0214)	(0.0287)
Votes for the FO union									0.0127	-0.0609	0.0091	-0.0639
									(0.0364)	(0.0503)	(0.0361)	(0.0483)
Votes for the CGC union									0.0720	0.0167	0.0649	0.0316
									(0.0188)	(0.0259)	(0.0200)	(0.0267)
Union delegates									-0.0113	-0.0109	-0.0083	-0.0050
									(0.0056)	(0.0077)	(0.0058)	(0.0078)
Health and Security									-0.0516	0.0069	-0.0492	0.0031
Commissions (CHSCT)									(0.0108)	(0.0149)	(0.0109)	(0.0145)
R-Square	0.2242	0.4015	0.2168	0.4388	0.2697	0.4831	0.0628	0.2593	0.4090	0.6333	0.4206	0.6626
Number of Obs	70	70	70	70	66	66	95	95	93	93	93	93

Note: Results are employment-weighted least squares estimates with standard errors in parentheses.

period, relatively strong unions still existed in some industries. For instance, Table 1 in Peoples (1998) shows that, as late as 1996, the union membership rate was as high as 23% in the Trucking industry, 36% in the Airline industry, and 29% in the Telecommunications industry. Our results show that, in the United States, stronger unions are associated with larger firm effects in the industry. In France, the type of agreement is often positively related to the size of the firm effect. Industries with few agreements targeted at the high-wage workers have relatively high firm effects. But, because of the minimum wage which sets the minimum when no agreement for the low-wage exists, agreements targeted at low-wage workers have no impact on person or firm effects. Individual-based agreements are positively associated with large person effects. Industry-wide agreements, which affect every worker in the industry, have a negative impact on both person and firm effects in the industry–a potential reflection that a compromise taking many aspects of the industry into consideration was signed but also a reflection that the industry has many low-wage workers.

The orientation of the union is also quite revealing. The CGT affects solely and positively the firm effect. Interestingly enough, when the Herfindahl index, the capital-labor ratio, and the export to sales ratio of the industry are included, as shown in the last two columns, the coefficient becomes insignificant. Therefore, CGT appears to benefit from product market aspects of the industries in which this union is strong. The CFDT and, above all, the CGC are apparently strong among the more educated which translates their presence into large person effects. Finally, a strong presence of union representatives, often present in low-wage manufacturing industries, translates into lower person effects. In the United States, the product market variables appear to have little role on the person and firm effects, except for the capital-labor ratio which has a negative coefficient, potentially because of the substitution of capital for labor.

Table 6 presents our second specification in which we control for occupations or the age and education structure in the industry. Results change much more for the United States than for France. In particular, the coefficient of the quasi-rent is divided by a factor of three in the firm effect regression. The other variables have little impact on the firm or the person effects in the presence of age and education controls. A first conclusion is that most of the person and firm effects in the United States reflect educational or occupational capital, specific to the industry. Little is left to the quasi-rent, even though its coefficients remain significant. This is far from being so in France: the firm effect remains strongly related to the quasi-rent, but not the person effect. The impact of the product market variables on the firm effect is unchanged. Finally, the union variables have an unchanged effect on the firm component of the inter-industry differential but have, now, virtually no effect on the person component.

This analysis clearly shows that although the person and the firm components of the inter-industry wage differentials are highly correlated between the United States and France, this similarity hides strong differences in their origin. The firm component of the differentials appears to be tightly related to the presence and structure of unions in France, it is also significantly related to the product market conditions prevailing in the industry. The person component of the differentials is mostly related to skills of the workforce, be it education or

TABLE 6: Regressions of Person and Firm Effects on Economy-Wide Factors (with controls)

	United States						France					
	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect	Person effect	Firm effect
	education and age		occupations		education and age		education and age		occupations		education and age	
Normalized quasi-rent	0.0471	0.0623	0.0475	0.1177	0.0816	0.0780	0.0003	0.0678	0.0128	0.0594	0.0093	0.0536
	(0.0170)	(0.0180)	(0.0126)	(0.0192)	(0.0251)	(0.0269)	(0.0073)	(0.0139)	(0.0088)	(0.0123)	(0.0085)	(0.0128)
Herfindahl Index					-2.3524	2.2784					-0.0466	0.0981
					(1.9821)	(2.1186)					(0.0395)	(0.0590)
Export-Sales ratio					0.1985	0.1626					0.0248	0.1191
					(0.1299)	(0.1389)					(0.0437)	(0.0652)
Capital-Labor ratio					-0.1903	-0.1295					0.0037	0.0169
					(0.1055)	(0.1128)					(0.0094)	(0.0141)
Union					0.1484	-0.2125						
					(0.1344)	(0.1437)						
Wage increases											-0.0165	-0.0347
(industry-wide agreement)											(0.0083)	(0.0123)
Wage increases											0.0046	0.0066
(individual-based)											(0.0054)	(0.0080)
Wage increases											-0.0067	0.0185
(for the low-wage)											(0.0069)	(0.0102)
Wage increases											0.0019	-0.0232
(for the high-wage)											(0.0050)	(0.0075)
Votes for the CGT union											-0.0129	0.0616
											(0.0314)	(0.0469)
Votes for the CFDT union											0.0039	0.0466
											(0.0201)	(0.0301)
Votes for the FO union											0.0269	-0.0631
											(0.0302)	(0.0451)
Votes for the CGC union											0.0343	0.0507
											(0.0179)	(0.0267)
Union delegates											-0.0037	-0.0130
											(0.0050)	(0.0075)
Health and Security											-0.0230	0.0111
Commissions (CHSCT)											(0.0103)	(0.0154)
R-Square	0.5255	0.8041	0.6114	0.6706	0.5583	0.8208	0.5905	0.5159	0.5076	0.6866	0.6406	0.7393
Number of Obs	70	70	70	70	66	66	95	95	95	95	93	93

Note: Results are employment-weighted least squares estimates with standard errors in parentheses.

the lack of. In the United States, the person and the firm components of the inter-industry wage differentials appear to reflect some elements of bargaining and unionization, but they mostly reflect occupational or educational capital specific to the industry.

8 Conclusion

We have specified and tested a model of long-term interindustry wage differences that decomposes the measure into a part due to unobservable individual heterogeneity and a part due to employer heterogeneity. We find that for both the United States and France this decomposition reveals much about the sources of the differences. A single aggregate measure, the quasi-rents per worker, explains a remarkable percentage of the variation. It is more strongly related to the employer heterogeneity than to the individual heterogeneity, particularly in France. Controls for the capital/labor ratio, product market concentration, and measurable demographic differences within sectors do not reduce the strength of this conclusion.

Predictably, the wage rate decomposition and the subsequent analysis of its components raise some new issues for the labor economist. The fundamental decomposition is identifiable because longitudinally linked employer-employee data, whether sampled or from universes, provide a sufficiently rich description of the connectedness of the labor market. This connectedness, which we exploit to estimate the person and firm effects, can also be exploited to model the temporal variation in functions of these effects. In our present analysis, this temporal variation is entirely attributed to changes in the composition of the aggregates (industries) and not to changes in the structure of the compensation within aggregates (time-varying person and firm effects). Such a pursuit is one important extension of our work.

9 Appendix[7]

The normal equations for least squares estimation of fixed person, firm, and characteristic effects are of very high dimension. Estimation of the full model by fixed-effect methods requires special algorithms to deal with the high dimensionality of the problem. After completing work on AFK and AKM, which use statistical approximations, ACK developed new algorithms that permit the exact least squares estimation of all the effects in equation (2). These algorithms, which are based on the iterative conjugate gradient method, deal with the high dimensionality of the data by using sparse matrices. These methods have some similarity to those used in the animal and plant breeding literature[8]. Because of the way

[7] This appendix draws directly from Abowd, Creecy, and Kramarz (2002).

[8] See Abowd and Kramarz (1999b) for a longer dicussion of the relation of these models to those found in the breeding literature. The techniques are summarized in Robinson (1991) and the random-effects methods are thoroughly discussed in Neumaier and Groeneveld (1996). The programs developed for breeding applications cannot be used directly for the linked employer-employee data application because of the way the breeding effect that is equivalent to our employer effects is parameterized.

these algorithms work, conventional methods for assuring that the effects are identified (estimable) do not work. Thus, ACK also developed appropriate new methods for computing the estimable functions of interest based on equation (7) below.

Least Squares Normal Equations. The full least squares solution to the estimation problem for equation (2) solves the following normal equations for all estimable effects:

$$\begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix} \begin{bmatrix} \beta \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} X'y \\ D'y \\ F'y \end{bmatrix} \tag{7}$$

In both of our estimation samples, the cross-product matrix on the left-hand side of the equation is too high-dimensional to use conventional algorithms (*e.g.*, those implemented in SAS, Stata, and other general purpose linear modeling software based on variations of the sweep algorithm for solving (7)). Abowd, Creecy, and Kramarz (2002) solve the system directly.

Identification of Individual and Firm Effects. Many interesting economic applications of equation (2) make use of the estimated person and firm effects. Estimation requires a method for determining the identified effects[9]. The usual technique of sweeping out singular row/column combinations from the normal equations (7) is not applicable to the method described in this paper because equation (7) is solved without the computation of a generalized inverse. Hence, identification of the person and firm effects for estimation by direct least squares requires finding the conditions under which equation (7) can be solved exactly for some estimable functions of the person and firm effects. In this sub-section we ignore the problem of identifying the coefficients β because in practice this is rarely difficult.

The identification problem for the person and firm effects can be solved by applying methods from graph theory to determine groups of connected individuals and firms. Within a connected group of persons/firms, identification can be determined using conventional methods from the analysis of covariance. Connecting persons and firms requires that some of the individuals in the sample be employed at multiple employers. When a group of persons and firms is connected, the group contains all the workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed. In contrast, when a group of persons and firms is not connected to a second group, no firm in the first group has ever employed a person in the second group, nor has any person in the first group ever been employed by a firm in the second group. From an economic perspective, connected groups of workers and firms show the realized mobility network in the economy. From a statistical perspective, connected groups of workers and firms block-diagonalize the normal equations and permit the precise statement of identification restrictions on the person and firm effects.

The following algorithm constructs G mutually-exclusive groups of connected observations from the N workers in J firms observed over the sample period.

[9]Standard statistical references, for example Searle, Casella, and McCulloch (1992), provide general methods for finding the estimable functions of the parameters of equation (3). These methods also require the solution of a very high dimension linear system and are, therefore, impractical for our purposes.

For g = 1, ..., repeat until no firms remain:

The first firm not assigned to a group is in group g.

Repeat until no more firms or persons are added to group g:

Add all persons employed by a firm in group g to group g.

Add all firms that have employed a person in group g to group g.

End repeat.

End for.

At the conclusion of the algorithm, the persons and firms in the sample have been divided into G groups. The number of individuals in each group is N_g. The number of employers in each group is J_g.

Within each group g, the group mean of y and $N_g - 1 + J_g - 1$ person and firm effects are identified. Some groups contain a single employer and, possibly, only one individual. For groups that contain more than one employer, every employer in the group is connected (in the graph-theoretic sense) to at least one other employer in the group. This algorithm finds all of the maximally connected sub-graphs of a graph. The relevant graph has a set of vertices that is the union of the set of persons and the set of firms and edges that are pairs of persons and firms. An edge (i, j) is in the graph if person i has worked for firm j. After the construction of the G groups, exactly $N + J - G$ effects are estimable. See the proof in Appendix 1 of ACK[10].

Normal Equations after Group Blocking. The identification argument can be clarified by considering the normal equations after reordering the person and firm effects so that those associated with each group are placed in the design matrix in ascending order. For simplicity, let the arbitrary equation determining the unidentified effect simply set that effect equal to zero, *i.e.*, set one person or firm effect equal to zero in each group. Thus, the column associated with this effect can be removed from the reorganized design matrix and the column associated with the group mean is suppressed. The resulting normal equations are:

[10] The grouping algorithm that is used identifies the "main effect" contrasts due to persons and firms in our model within each group. In the linear models literature the "groups" are called "connected data." See (Searle 1987), chapter 5, section 3, pp. 139-149 for a discussion of connected data. See Weeks and Williams (1964) for the general algorithm in analysis of variance models.

$$
\left[\begin{array}{c|cc|cc|c|cc}
X'X & X'D_1 & X'F_1 & X'D_2 & X'F_2 & \cdots & X'D_G & X'F_G \\
\hline
D_1'X & D_1' & D_1 & 0 & 0 & \cdots & 0 & 0 \\
F_1'X & F_1' & D_1 & 0 & 0 & \cdots & 0 & 0 \\
\hline
D_2'X & 0 & 0 & D_2'D_2 & D_2'F_2 & \cdots & 0 & 0 \\
F_2'X & 0 & 0 & F_2'D_2 & F_2'F_2 & \cdots & 0 & 0 \\
\hline
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\hline
D_G'X & 0 & 0 & 0 & 0 & \cdots & D_G'D_G & D_G'F_G \\
F_G'X & 0 & 0 & 0 & 0 & \cdots & F_G'D_F & F_G'F_G
\end{array}\right]
\left[\begin{array}{c}
\beta \\
\hline
\theta_1 \\
\psi_1 \\
\hline
\theta_2 \\
\psi_2 \\
\hline
\cdots \\
\hline
\theta_G \\
\psi_G
\end{array}\right]
=
\left[\begin{array}{c}
X'y \\
\hline
D_1'y \\
F_1'y \\
\hline
D_2'y \\
F_2'y \\
\hline
\cdots \\
\hline
D_G'y \\
F_G'y
\end{array}\right]
\quad (8)
$$

The normal equations have a sub-matrix with block diagonal components. This matrix is of full rank and the solution for the parameter vector is unique. ACK do not solve equation (8) directly; they apply the technique discussed below to estimate the identifiable effects.

Estimation by Direct Solution of the Least Squares Problem. Appendix 2 in ACK shows the exact algorithm used to solve equation (7), a variant of the conjugate gradient algorithm for which ACK customized the sparse representation of equation (7) so that very large problems with many X variables would be practical. In practice, ACK apply this algorithm to the full set of persons, firms and characteristics shown in the design matrices of equations (2) and (7). Unlike equation (8), the design matrix in equation (7) is not of full rank. Although the algorithm ACK use converges to a least squares solution, the parameter estimates are not unique. The output from the algorithm provides a non-unique set of effects to which they subsequently apply the identification procedure. To make the effects unique for each group, they eliminate one person effect by setting the group mean person effect to zero. ACK also set the overall mean person and firm effects equal to zero. This procedure identifies the grand mean of the dependent variable (or the overall regression constant if X and y have not been standardized to mean zero) and a set of $N + J - G - 1$ person and firm effects measured as deviations from the grand mean of the dependent variable. The computer software is available from the authors for both the direct least squares estimation of the two-factor analysis of covariance and the grouping algorithm. Computer software that implements both the random and fixed effects versions of these models used in breeding applications can be found in Groeneveld (1998). The specific algorithm we use can be found in (Dongarra, Duff, Sorensen, and der Vorst 1991) p. 146.

References

Abowd, J. M., R. Creecy, and F. Kramarz (2002). Computing person and firm effects using linked longitudinal employer- employee data. *LEHD Technical Paper 2002-06*.

Abowd, J. M., H. Finer, and F. Kramarz (1999). Individual and firm heterogeneity in compensation: An analysis of matched longitudinal employer- employee data for the state of washington. In J. Haltiwanger, J. Lane, and P. Theeuwes (Eds.), *The Creation and Analysis of Employer-Employee Matched Data*, pp. 3–24. Amsterdam: North Holland.

Abowd, J. M., J. Haltiwanger, and J. I. Lane (2004). Integrated longitudinal employee-employer data for the united states. *American Economic Review, Papers and Proceedings 94*(2), 224–229.

Abowd, J. M. and F. Kramarz (1999a). The analysis of labor markets using matched employer-employee data. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics, Volume 3*, Volume 3B, Chapter 40, pp. 2629–710. Amsterdam: North Holland.

Abowd, J. M. and F. Kramarz (1999b, March). Econometric analyses of linked employer–employee data. *Labour Economics 6*, 53–74.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

Abowd, J. M., F. Kramarz, D. N. Margolis, and K. R. Troske (2001, December). The relative importance of employer and employee effects on compensation: A comparison of France and the United States. *Journal of the Japanese and International Economies 15*(4), 419–436.

Abowd, J. M., F. Kramarz, and S. Woodcock (2004). Econometric analyses of linked employer-employee data. mimeo.

Abowd, J. M. and T. 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, J. M., P. Lengermann, and K. McKinney (2003). The measurement of human capital in the U.S. economy. *LEHD Technical Paper TP 2002-09*.

Abowd, J. M. and L. Vilhuber (2005). The sensitivity of economic statistics to coding errors in personal identifiers. *Journal of Business and Economic Statistics 23*(2), 133–152.

Benedetto, Gary, e. (2004). Using worker flows in the analysis of firm data. LEHD Technical Paper No. TP-2003-09.

Bertrand, M. and F. Kramarz (2002). Does entry regulation hinder job creation ?: Evidence from the french retail industry. *Quarterly Journal of Economics 117*(4), 1369–1414.

Black, S. and P. Strahan (2001). The division of spoils: Rent-sharing and discrimination in a deregulated industry. *American Economic Review 91*, 814–831.

Brown, C. and J. L. Medoff (1989). The employer size-wage effect. *Journal of Political Economy 97*, 1027–1059.

Cahuc, P. and F. Kramarz (1997). Voice and loyalty as a delegation of authority: A model and a test on matched worker-firm panel. *Journal of Labor Economics 15*(4), 258–288.

Card, D. (1986). Efficient contracts with costly adjustments: Short-run employment determination for airline mechanics. *American Economic Review 76*, 1045–1071.

Dickens, W. T. and L. Katz (1987). Inter-industry wage differences and industry characteristics. In K. Lang and J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets*. Oxford: Basil Blackwell.

Djankov, S., R. L. Porta, F. L. de Silanes, and A. Shleifer (2002). The regulation of entry. *Quarterly Journal of Economics 117*(1), 1–37.

Dongarra, J. J., I. S. Duff, D. C. Sorensen, and H. A. V. der Vorst (1991). *Solving Linear Systems on Vector and Shared Memory Computers*. Philadelphia: SIAM.

Farber, H. S. and B. Western (2001). Accounting for the decline of unions in the private sector, 1973-1998. *Journal of Labor Research 22*, 459–486.

Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *Review of Economic Studies 59*, 515–535.

Goux, D. and E. Maurin (1999). Persistence of inter-industry wage differentials: A reexamination on matched worker-firm panel data. *Journal of Labor Economics 17*, 492–533.

Groeneveld, E. (1998). *VCE4 User's Guide and Reference Manual* (2 ed.). Höltystrass, Germany: Institute of Animal Husbandry and Animal Behavior.

Groshen, E. (1991). Sources of intra-industry wage dispersion: How much do employers matter? *Quarterly Journal of Economics 106*, 869–884.

Haltiwanger, J. and et al. (editors) (1999). *The Creation and Analysis of Employer-Employee Matched Data*. North Holland.

Kramarz, F. and T. Philippon (2001). The impact of differential payroll tax subsidies on minimum wage employment. *Journal of Public Economics 82*, 115–146.

Krueger, A. and L. H. Summers (1987). Reflections on the inter-industry wage structure. In K. Lang and J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets*. Oxford: Basil Blackwell.

Krueger, A. and L. H. Summers (1988, March). Efficiency wages and the inter-industry wage structure. *Econometrica 56*(2), 259–293.

LEHD Program (2002). The Longitudinal Employer-Household Dynamics Program: Employment Dynamics Estimates Project Versions 2.2 and 2.3. *LEHD Technical Paper 2002-05*.

McKinney, K. (2002). Successor/predecessor firms. *LEHD Technical Paper 2002-04*.

Mortensen, D. (2003). *Wage Dispersion: Why Are Similar Workers Paid Differently?* MIT Press. ISBN 0262134330.

Murphy, K. M. and R. H. Topel (1987). Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. In K. Lang and J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets*. Oxford: Basil Blackwell.

Neumaier, A. and E. Groeneveld (1996). Restricted maximum likelihood estimation of covariance in sparse linear models. *Institut für Mathematik, Wien University Working Paper*.

Peoples, J. (1998). Deregulation and the labor market. *Journal of Economic Perspectives XII*, 112–130.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Robinson, G. K. (1991). That blup is a good thing: The estimation of random effects. *Statistical Science 6*, 15–51.

Rotbart, G. (1991). *Enquête sur la Structure des Salaires (INSEE-Méthodes)*. INSEE.

Rubin, D. (1987). *Multiple Imputation for Non-Response in Surveys* (2 ed.). Wiley-Interscience.

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York: John Wiley and Sons.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. New York: John Wiley and Sons.

Shimer, R. (2001). The assignment of workers to jobs in an economy with coordination frictions. *NBER Working Paper No. w8501*.

Stevens, D. (2000). Employment that is not covered by state unemployment insurance laws. *LEHD Technical Paper 2002-01*.

U.S. Department of Labor, Bureau of Labor Statistics (2000). *Handbook of Methods*.

Weeks, D. L. and D. R. Williams (1964). A note on the determination of connectedness in n-way cross classification. *Technometrics 6*, 319–324.

Table A1: Inter-industry Wage Differentials for the US

SIC	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
01	Agriculture-crops	-0.570	-0.293	-0.315
02	Agriculture-livestock	-0.229	-0.218	-0.034
07	Agricultural services	-0.568	-0.280	-0.316
08	Forestry	-0.201	-0.147	-0.098
09	Fishing	-0.165	-0.212	0.014
10	Metal mining	0.382	-0.129	0.500
12	Coal mining	0.396	-0.194	0.636
13	Oil and gas extraction	0.437	0.108	0.285
14	Nonmetallic minerals, except fuels	0.244	-0.027	0.277
15	Building construction--general contractors	0.045	-0.035	0.047
16	Heavy construction other than buildings	0.185	0.015	0.142
17	Construction--special trade contractors	0.007	-0.043	0.030
20	Food and kindred products	0.004	-0.097	0.096
21	Tobacco products	0.167	0.110	0.213
22	Textile mill products	-0.044	-0.068	0.118
23	Apparel and other textile products	-0.402	-0.240	-0.142
24	Lumber and wood products	-0.098	-0.125	0.031
25	Furniture and fixtures	-0.120	-0.084	0.024
26	Paper and allied products	0.221	0.017	0.232
27	Printing and publishing	0.106	0.074	0.040
28	Chemicals and allied products	0.407	0.126	0.303
29	Petroleum and coal products	0.478	0.109	0.345
30	Rubber and miscellaneous plastics products	0.010	-0.083	0.105
31	Leather and leather products	-0.292	-0.215	-0.070
32	Stone, clay, and glass products	0.086	-0.082	0.183
33	Primary metal industries	0.148	-0.071	0.255
34	Fabricated metal products	0.055	-0.063	0.126
35	Industrial machinery and equipment	0.284	0.072	0.233
36	Electronic and other electric equipment	0.288	0.072	0.228
37	Transportation equipment	0.289	-0.008	0.292
38	Instruments and related products	0.411	0.137	0.280
39	Miscellaneous manufacturing industries	-0.087	-0.088	0.009
40	Railroad transportation	-0.100	-0.145	0.081
41	Local and interurban passenger transportation	-0.101	-0.110	0.023
42	Motor freight transportation and warehousing	-0.008	-0.055	0.044
44	Water transportation	0.166	0.103	0.052
45	Transportation by air	0.227	0.199	0.024
46	Pipelines, except natural gas	0.625	0.094	0.475
47	Transportation services	0.098	0.036	0.060
48	Communications	0.460	0.123	0.346
49	Electric, gas, and sanitary services	0.445	0.095	0.358
50	Wholesale trade - durable goods	0.255	0.102	0.162
51	Wholesale trade - nondurable goods	0.121	0.044	0.085
52	Building materials, hardware, garden supply	-0.131	-0.042	-0.078
53	General merchandise stores	-0.303	-0.050	-0.241
54	Food stores	-0.236	-0.025	-0.195
55	Automotive dealers and gasoline service stations	0.015	-0.001	0.019
56	Apparel and accessory stores	-0.248	0.016	-0.252
57	Home furniture, furnishings, and equipment	-0.039	0.015	-0.042
58	Eating and drinking places	-0.554	-0.161	-0.397
59	Miscellaneous retail	-0.229	-0.004	-0.212

Table A1: Inter-industry Wage Differentials for the US

SIC	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
60	Depository institutions	0.175	0.120	0.069
61	Nondepository credit institutions	0.396	0.189	0.213
62	Security brokers, dealers, exchanges	0.659	0.393	0.258
63	Insurance carriers	0.366	0.146	0.230
64	Insurance agents, brokers, and services	0.228	0.142	0.099
65	Real estate	-0.086	-0.049	-0.042
67	Holding and other investment offices, exc. trusts	0.292	0.229	0.080
70	Hotels, rooming houses, camps, and lodging	-0.385	-0.209	-0.186
72	Personal services	-0.371	-0.160	-0.213
73	Business services	-0.168	-0.125	-0.071
75	Automotive repair, services, and parking	-0.182	-0.137	-0.044
76	Miscellaneous repair services	-0.008	-0.050	0.049
78	Motion pictures	-0.196	0.068	-0.256
79	Amusement and recreation services	-0.312	-0.079	-0.249
80	Health services	0.112	0.051	0.064
81	Legal services	0.471	0.236	0.230
82	Educational services	-0.016	0.158	-0.183
83	Social services	-0.296	-0.199	-0.102
84	Museums, art galleries, and botanical and zoos	-0.118	0.043	-0.157
86	Membership organizations	-0.208	-0.049	-0.160
87	Engineering, accounting, research, management	0.317	0.142	0.167
88	Private households	-0.643	-0.219	-0.252
89	Miscellaneous services	0.409	0.136	0.227
91	Executive, legislative and general	0.114	0.076	0.043
92	Justice, public order	0.207	0.024	0.198
93	Finance, taxation and monetary policy	0.068	0.006	0.057
94	Administration of human resources	0.062	0.026	0.057
95	Environmental quality and housing	0.072	-0.006	0.090
96	Administration of economic programs	0.045	0.045	0.037
97	National security	-0.229	-0.152	-0.052

Table A2: Inter-industry Wage Differentials for France

NAP	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
01	Agriculture	-0.135	0.038	-0.187
04	Coal mining	0.312	0.098	0.170
05	Crude petroleum and natural gas extraction	0.490	0.239	0.196
06	Electricity production and supply	0.283	0.018	0.235
07	Distribution of Gas	0.319	0.059	0.229
08	Water and city-heating supply	0.201	0.092	0.081
09	Ferrous metal mining	0.246	-0.207	0.370
10	Iron and steel foundries	0.208	-0.078	0.218
11	Primary metal manufacturing	0.054	-0.064	0.078
12	Mineral and non-metal mining	0.262	0.032	0.218
13	Primary nonmetallic manufacturing	0.228	0.033	0.148
14	Miscellaneous mineral production	0.126	-0.006	0.108
15	Cement, stone, and concrete products	0.027	-0.022	0.020
16	Glass and glass products	0.148	-0.015	0.139
17	Basic chemical manufacture	0.289	0.100	0.144
18	Allied chemical products, soaps, cosmetics	0.132	0.074	0.034
19	Pharmaceuticals	0.283	0.035	0.218
20	Foundries and smelting works	0.057	-0.054	0.080
21	Metal works	0.036	-0.006	0.033
22	Farm machinery and equipment	0.019	-0.019	0.033
23	Metalworking machinery manufacture	0.116	0.002	0.091
24	Industrial machinery manufacture	0.103	0.026	0.061
25	Material handling machines and equipment	0.123	-0.008	0.105
26	Ordnance	0.212	0.013	0.180
27	Office and accounting machines	0.472	0.094	0.346
28	Electrical machinery equipment	0.100	0.009	0.071
29	Electronic computing equipment	0.148	0.032	0.102
30	Household appliances	0.057	-0.071	0.108
31	Motor vehicles, trains, land transport man.	0.116	-0.034	0.119
32	Ship and boat building	0.157	-0.022	0.143
33	Aircraft and parts manufacture	0.313	0.167	0.098
34	Professional and scientific equipment man.	0.062	0.012	0.044
35	Meat products	-0.031	-0.043	0.025
36	Dairy products	0.100	-0.017	0.108
37	Canned and preserved products	-0.001	-0.049	0.071
38	Bakery products	-0.233	-0.019	-0.133
39	Grain mill and cereal products	0.089	-0.003	0.083
40	Miscellaneous food preparations	0.082	-0.034	0.108
41	Beverage industries	0.127	-0.027	0.124
42	Tobacco products manufacture	0.337	0.016	0.256
43	Knitting mills, threads and artificial fibers	0.238	-0.030	0.176
44	Textile products	-0.038	-0.089	0.033
45	Leather products except footwear	-0.096	-0.078	-0.025
46	Footwear	-0.059	-0.086	0.021
47	Apparel, clothing and allied products	-0.065	-0.072	0.015
48	Lumber mills	-0.093	-0.064	-0.026
49	Furniture and fixtures manufacture	-0.100	-0.034	-0.059
50	Pulp and paper mills and packaging prod.	0.118	-0.016	0.112
51	Printing and publishing	0.034	0.078	-0.057
52	Rubber products	0.098	-0.090	0.149
53	Plastic products	0.021	-0.025	0.043

Table A2: Inter-industry Wage Differentials for France

NAP	Industry	Raw Industry Wage Differential	Industry Average Person Effect	Industry Average Firm Effect
54	Miscellaneous manufacturing industries	-0.041	-0.035	-0.008
55	Construction	-0.100	-0.020	-0.083
56	Waste product management	-0.053	-0.028	-0.041
57	Wholesale food trade	-0.039	-0.035	-0.005
58	Wholesale non-food trade	0.026	0.024	-0.005
59	Inter-industry wholesale trade	0.106	0.047	0.038
60	Commercial intermediaries	0.013	0.058	-0.057
61	Retail food and supermarkets	0.071	-0.040	0.151
62	Retail specialty and neighborhood food	-0.177	-0.048	-0.085
63	Retail general merchandise and non food	0.017	0.014	0.013
64	Retail specialty non food trade	-0.084	0.027	-0.103
65	Automobile dealers, auto parts and repair	-0.107	0.019	-0.095
66	Miscellaneous repair services	-0.118	-0.005	-0.101
67	Hotels, motels, bars and restaurants	-0.167	-0.047	-0.068
68	Railroad transportation	0.186	-0.023	0.158
69	Bus, taxicab and other urban transit	-0.045	-0.020	-0.035
70	Inland water transportation	0.000	-0.020	0.049
71	Marine transport and coastal shipping	0.215	0.106	0.088
72	Air transportation	0.466	0.123	0.306
73	Allied transportation and warehousing	0.113	0.018	0.088
74	Travel agencies	0.031	0.009	0.018
75	Telecommunications and postal	0.063	-0.010	0.103
76	Financial holding companies	0.378	0.205	0.130
77	Advertising and consulting services	0.024	0.033	0.019
78	Brokers, credit agencies, and insurance	0.125	0.090	0.020
79	Commercial real estate development, sales	-0.025	0.018	-0.067
80	Nonresidential goods rental services	-0.018	0.046	-0.060
81	Real estate renting and leasing	-0.099	-0.023	-0.092
82	Commercial education services	-0.259	0.044	-0.321
83	Commercial research services	0.266	0.134	0.108
84	Commercial health services	0.022	-0.037	0.044
85	Commercial social services	-0.100	-0.146	0.035
86	Commercial entertainment and recreation	0.214	0.091	0.115
87	Miscellaneous commercial services	-0.239	-0.076	-0.137
88	Insurance carriers	0.134	0.063	0.048
89	Banks and financial institutions	0.279	0.132	0.125
90	Public administration, general	-0.218	-0.035	-0.167
91	Social security administration	0.167	0.103	0.029
92	Teaching, non-market	-0.414	0.000	-0.431
93	Research, non-market	0.119	0.154	-0.030
94	Health care, non-market	0.301	-0.105	0.376
95	Social services, non-market	-0.241	-0.099	-0.150
96	Recreational, cultural, and sporting, non-market	-0.252	-0.024	-0.214
97	Miscellaneous public services, non-market	-0.288	-0.025	-0.255